Location Management in 3G/4G Wireless Systems

Ian F. Akyildiz Wenye Wang Young-Jun Lee
Broadband Wireless Networking Laboratory
School of Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA 30332, USA
Email: {ian,wenye,young}@ece.gatech.edu

ABSTRACT

This paper presents current and latest schemes for location management in next generation wireless systems (3G and 4G). First, the hierarchical reference model of 3G wireless systems is introduced. Then a dynamic inter-system location management technique is developed based on new concepts of boundary location area and boundary location register, which are suitable for inter-system roaming scenarios. Moreover, the corresponding mobility application part (MAP) protocol is designed for inter-system location registration and call delivery procedures. The performance is evaluated in terms of signaling costs, latency of location registration and call delivery, as well as call losses due to the inter-system roaming. For the 4G systems, a generic system model is introduced to incorporate the existing wireless local networks and cellular systems and so on. In particular, a new scheme for reducing costs in route optimization is introduced to solve the above problems. In this new scheme, route optimization is performed only when it minimizes the total cost function, which provides the optimal result from the viewpoint of link and signaling costs. The simulation results show that the proposed scheme provides the better performance.

Keywords: L ocation Management, Location Update, Paging, Mobile IP, Routing.

1. INTRODUCTION

The foreseeable employment of next generation (NG) wireless systems, such as the International Mobile Telecommunication System 2000 (IMT-2000) and the Universal Mobile Telecommunications System (UMTS), as well as 4G systems (IP-based) will lead to a tremendous increase in both the number of mobile users and mobile services. In the mobile environment, the mobile subscribers use mobile terminals (MTs) to communicate with others through base stations (BSs) via radio links when they change their locations. Currently proposed NG systems provide global roaming, higher capacity, and increased flexibility for multimedia services, such as audio, video, teleconferencing, and so on. [1–3] The most distinguished features of NG wireless systems, can be highlighted as reliable quality of service (QoS) for various applications and global roaming. Among the challenges for the NG wireless systems, the location management is the focus of the this paper.

Location management is a two-stage process that enables the network to discover the current attachment point of the mobile user for call delivery. The first stage is *location update or registration*. In this stage, the MT periodically notifies the network of its new access point, allowing the network to authenticate the user and revise the user's location profile. The second stage is *call delivery*. Here the network is queried for the user location profile and the current position of the mobile terminal is found.

Current techniques for location management involve database architecture design and the transmission of signaling messages between various components of a signaling network. As the number of mobile subscribers increases, new or improved schemes are needed to effectively support a continuously increasing subscriber populations. Other issues include; *security, dynamic database updates, querying delays, terminal paging methods, and paging delays*.

In this paper, we focus on the inter-system location management for 3G wireless systems since the inter-system roaming is different from the existing personal communication service (PCS) systems. Also, the optimal routing is emphasized and discussed because most of the current proposals about 4G systems are referred to as IP-based architecture. The rest of this paper is organized as follows. In Section 2, the system architecture and location management schemes of 3G wireless systems will be described, and the performance evaluation will be presented. In Section 3, the location management issue of 4G wireless systems will be addressed, a new location management scheme will be described, and the performance of the new scheme will be evaluated. Finally, the paper is concluded in Section 4.

Modeling and Design of Wireless Networks, Edwin K. P. Chong, Editor, Proceedings of SPIE Vol. 4531

2. LOCATION MANAGEMENT IN 3G WIRELESS SYSTEMS

There is an increasing demand for a wide variety of high bandwidth services beyond voice and data, including video, Internet access, and tele-conferencing.[3,4] Much as existing cellular systems brought mobile telephony capabilities to the impressive market, NG wireless systems will introduce value that extends beyond basic telephony services. The widespread growth of the Internet has created a mass market base for multimedia and information services. The challenge is to merge mobile telephony coverage and the associated user base with the Internet and other multimedia applications.[4] Consequently, there are several driving forces for NG wireless systems: (i) the increasing demand of bandwidth and multimedia services, (ii) the cost-effective radio access for service delivery, and (iii) the need of integration between wireless systems with each other and with the wired infrastructure.

Thus NG wireless systems, are mobile multimedia, personal services, and the convergence of mobility and Internet. They are able to provide QoS approaching that of the traditional broadband wireline networks for a variety of traffic classes over wireless links and global roaming across diverse wireless network backbones.

2.1. System Architecture

One of the most distinguished features of 3G wireless systems is that they are visualized as seamless worldwide radio systems. They will enable data transfer, image transfer, video conferencing, and video delivery, regardless of locations. It is predicted that by 2005 there will be more than one billion mobile phone subscribers around the world.[5] However, to make this prevision come true, the wireless systems must be able to locate mobile users roaming across different network service providers, network operators, as well as geographical regions.[6–8] The NG wireless network is comprised of many systems using different standards such as GSM/GPRS, IS-95, IS-54/136 and so on. Each system in a given geographical area will typically be supported by a specific network operator. Thus, the wireless service area will be covered by similar and dissimilar systems. It is desirable to consider some factors which will influence the radio connections of the MTs roaming between different systems or tiers with similar cell architecture, e.g., the system architecture shown in Figure 1.[9] In the satellite network, the MT will communicate with fixed earth stations (FES), which govern wireless traffic for satellite terminals or with satellite itself, and a cell site switch (CSS) will handle one or more BSs.

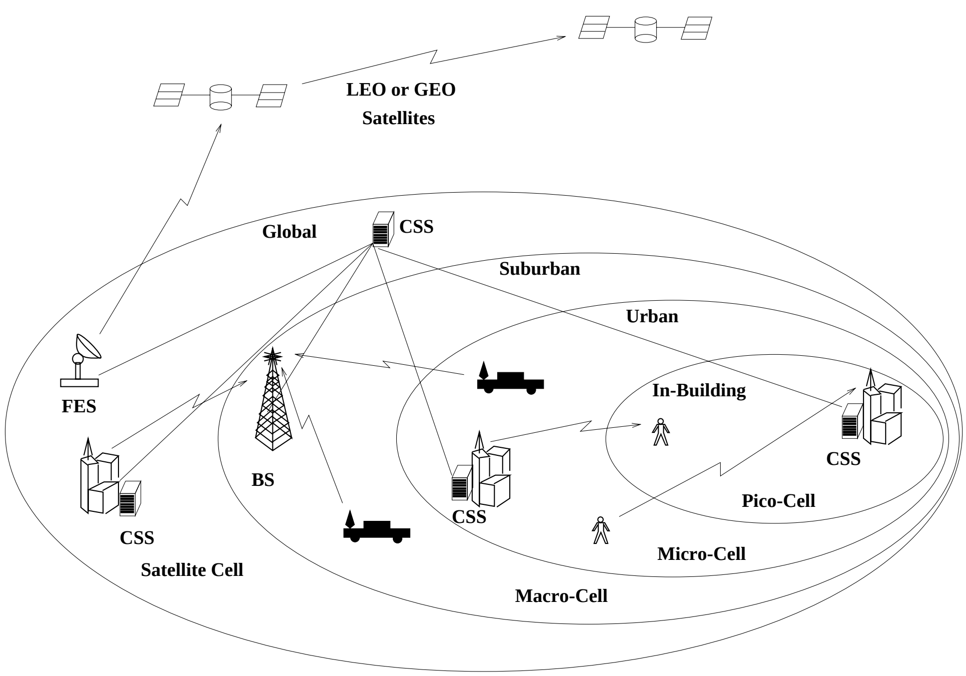


Figure 1. System Architecture of Intra-system and Inter-system Roaming.

There are two types of roaming in Figure 1: *intra-system* and *inter-system* roaming. Intra-system roaming refers to an MT's movement between the different tiers of the same systems, i.e., between the pico, micro, macro cells of Figure 1. For the current cellular network architecture, the service area is divided into many location areas (LAs), and each LA consists of a group of cells for microcell systems while one LA may be one cell for macrocell systems. Thus, when an MT moves from one LA to another within a system, it is experiencing the intra-system roaming. Many location management schemes for intra-system roaming have been investigated and designed as described in.[6]

Inter-system roaming refers to the MTs that move between backbones using protocols, technologies, or service providers. The inter-system roaming can be either the MT moving between different systems within the same tier or in different tiers. In the presence of inter-system roaming, the MTs may change from a North America system such as IS-95/IS-136 to a European system such as GSM/GPRS. In this case, new techniques must be devised to retrieve the mobility profiles between networks, and to prepare the requested services before the MTs enter the new systems. Of particular concern is how to reduce set-up delays, processing time, extra overhead, and call loss rates due to inter-system roaming.

Although many efforts have been taken on the location management for stand-alone systems,[6] little attention has been done on inter-system location management so far. The mobile users in NG wireless systems require that the reliable QoS characteristics be maintained throughout the duration of a call as they travel not only from cell to cell, but also from one system to another that use different technologies.[10,11] However, location management of inter-system roaming has not been paid much attention compared to that of intra-system roaming.[12,13] The current schemes for location update and paging are not efficient for future wireless networks in terms of connection setup time, signaling costs, and call loss rates. This means that the calls in progress are very likely to be dropped or interrupted when the MT moves from one system to another since the MT needs additional time for location registration, user identity authorization, and resource allocation. In addition, the MT's behavior is not well researched such as how to model an MT's movement pattern during an inter-system roaming and how to analyze and improve overall system performance. Thus, location management schemes, which mainly deal with the inter-system location update and paging, must be considered and enhanced to meet the requirements of inter-system roaming.

2.2. Inter-System Location Update and Paging

As the objective of NG wireless systems has been proposed in the middle of 1990s, some research activities have been conducted on this inter-system roaming. However, most of them are concerned with finding the ways to enable an external user to initiate a call or receive a call in a new system. For example, the registration, call delivery, and handset identity for the heterogeneous PCS systems were discussed in.[14] The registration protocols were modified to accommodate systems using different technologies such as GSM and IS-41. In,[15] the authors suggested to "homogenize" the service areas of two adjacent systems by filling the bigger size cells with smaller ones. Accordingly, the location update and paging would be implemented as in one network.

Moreover, it has been demonstrated in[16] that the roaming across systems imposes a significant increase in signaling traffic. The traffic analysis was based on the interworking of PCS1900 and IS-41 systems by using a dual mode home location registration. In,[17] the interworking between cordless and cellular systems, e.g., DECT and GSM, was thoroughly investigated. The signaling traffic of inter-system location registration and paging was analyzed based on different system architectures. In,[18] the inter-system location registration was studied on GSM/PDC roaming and additional interworking units were proposed to carry out signaling format transformation and authentication. However, in each of the above papers, the signaling costs were not computed based on any specific location update and paging algorithm and QoS requirements such as call loss and paging delay were not considered in the performance analysis.

In the following context, the location management in multi-tier PCS systems tackles the inter-system roaming, which includes inter-system location update and paging.[19] The inter-system location update is concerned with updating the location information of an MT performing inter-system roaming. The inter-system paging is about searching for the called terminal roaming between different service areas. The goal of inter-system location management is to reduce the signaling cost while maintaining QoS requirements. For example, reducing call loss is one of the key issues in maintaining the call connection; and decreasing the paging delay is critical to reduce the call set-up time.

2.2.1. Boundary Location Area and Inter-System Location Update

For inter-system location update, we consider a region called boundary location area (BLA) existing at the boundary between two systems, X and Y, as illustrated in Figure 2. The BLA is controlled by a boundary interworking unit (BIU) which is connected to mobile switching centers (MSCs) in both systems. The BIU is also responsible for querying the user's service information and transforming message formats. Also, the BIU is assumed to handle some other issues; such as, the compatibility of the air interfaces and the authentication of the mobile users.

The BLA of an MT is defined as a region in which the MTs can send the location registration request to the new system toward which the MT is moving. A distance based location update mechanism is designed such that the MT reports its location when its distance from the boundary is less than an *update distance*, d_{xy}, which is computed by (1). This location update

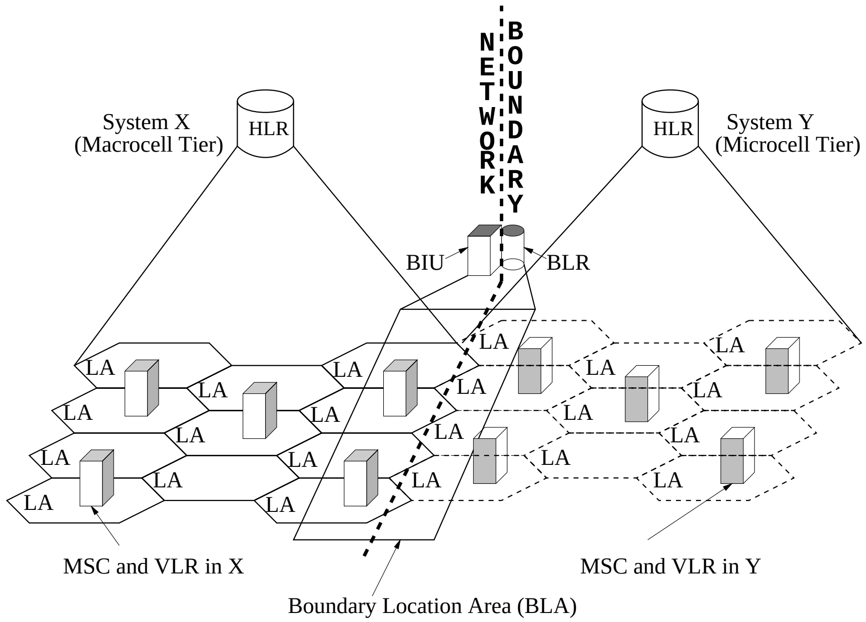


Figure 2. Boundary Location Area (BLA) and Boundary Location Register (BLR)

scheme guarantees that the MT updates its location information in an area that is within a distance threshold away from the boundary of two systems, X and Y. The *update distance* d_{xy} is determined by:

$$d_{xy} = \lceil \eta \cdot v_r \rceil \cdot d_o \qquad (1)$$

where η is a QoS factor relating to the bandwidth requirement and v_r is velocity ratio describing the MT's movement. d_o is *basic threshold* which is assumed to be the minimum distance away from the boundary that an MT must send its registration request for authenticating the identification and transforming signaling formats. d_o is a system parameter depending on the situation in system Y, such as system resources, service specifications, and architecture. This inter-system location update scheme is dynamic in the sense that v_r and η are variable over time depending on the network load and each user's mobility characteristics.

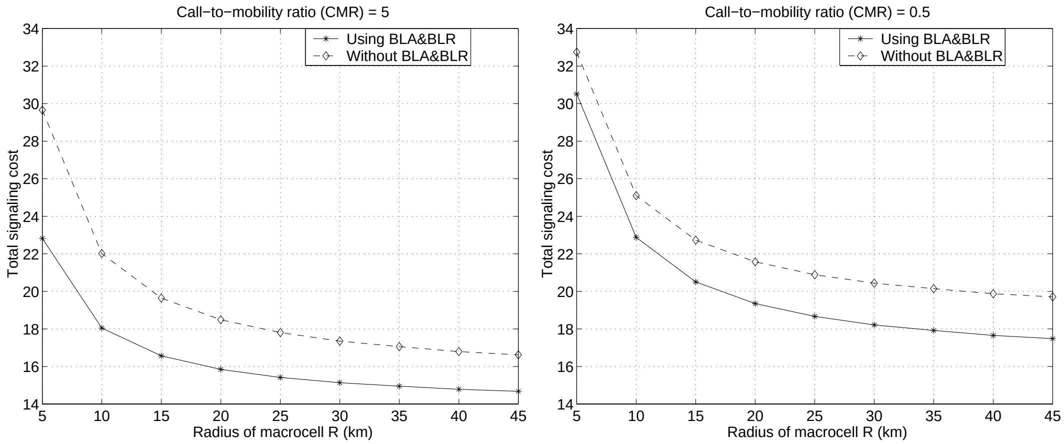
2.2.2. Boundary Location Register and Inter-System Paging

In addition to the concept of BLA, a database called *boundary location register* (BLR) is designated, which is a location information cache to maintain the roaming information of MTs between different systems. The BLR enables the inter-system paging to be implemented within a system that an MT is currently residing so that the paging costs can be reduced. Even though the BLR behaves in a similar fashion as a visitor location register (VLR) which is used to store the roaming information of MTs in the wireless systems, they have different features. First, the BLR provides MT's roaming information to two different systems; i.e., it is transparent to both adjacent systems. In contrast, VLR can only provide roaming information within one system. Second, there is only one BLR between a pair of neighboring systems while there may be many VLRs inside a stand-alone system with distributed architecture.

During the inter-system paging process, only one system (X or Y) is searched in the paging process for inter-system roaming terminals. This approach significantly reduces the signaling cost caused by inter-system paging. In particular, it is very suitable for the high traffic environment because it avoids searching in two adjacent systems. Moreover, since the BLR is an additional level of cache database, it will not affect the original database architecture. Another advantage of the BLR is that it reduces the zigzag effect caused by inter-system roaming. For example, when an MT is moving back and forth on the boundary, it only needs to update the information in the BLR instead of contacting the home location registers (HLRs). [20]

2.2.3. Estimation of Roaming Probabilities

Inter-system location update and paging costs are associated with the roaming probability from one network to another. The roaming probability can be either experimental results obtained from practical circumstances or numerical results estimated



(a) (b)

Figure 3. Total signaling costs of inter-system roaming from macrocell to microcell tier.

from theoretical models. Thus, we analyze the roaming probability based on a two-dimensional cellular configuration. There are two commonly used mobility models in the literature: fluid flow model [21–23] and random walk model.[24,25] Of these two models, fluid flow model is more suitable for users with high mobility, infrequent speed and direction changes. For pedestrian movements in which mobility is generally confined to a limited geographical area such as residential and business building, random walk model is more appropriate. Since we are considering inter-system roaming in multi-tier PCS systems, both fluid flow and random walk model are taken into account in the analysis. [26]

Under the fluid flow model, the direction of an MT's movement in the LA is uniformly distributed in the range of $(0, 2\pi)$.[21,27] The roaming probability which is denoted by $Pr_f[X, Y]$ is calculated by following expression:

$$Pr_f[X, Y] = \int_0^\infty f_Z(z)dz \cdot \int_0^\infty \lambda_n t \cdot e^{-\lambda_n t} \cdot \tilde{f}(t)dt, \tag{2}$$

where $f_Z(z)$ is the probability density function (pdf) of $Z \triangleq t_2 - \widetilde{T}$, t_2 is the call duration time and $\widetilde{T}$ is the residual sojourn time within an LA. λ_n is the mean arrival rate (Poisson) and $\tilde{f}(t)$ is the pdf of residual sojourn time the LA.

On the other hand, the most important characteristic of a random walk model is that the next position an MT occupies is equal to the previous position plus a random variable whose value is drawn independently from an arbitrary distribution. Correspondingly, the roaming probability, $Pr_r[X, Y]$, is obtained by

$$Pr_r[X, Y] = \sum_{k=0}^{K} P_K(k) \cdot Pr[K|k] \cdot \alpha_{K,K+1}, \tag{3}$$

where state k represents the mobile user's position and $P_K(k)$ is the probability that an incoming call arrives at state k; K is the total number of state and $\alpha_{K,K+1}$ is transition probability from K to $K + 1$. $Pr[K|k]$ is the probability that a call starts in state k and ends up in state K.

2.2.4. Performance Evaluation

Some experiments are conducted to show the performance of inter-system roaming supported by BLA&BLR concepts and the corresponding strategy. The total signaling costs, call loss rates, and paging delays depend on various parameters in multi-tier PCS systems. These parameters include update distance d_{xy} or d_{yx} explained in [19]; the size of a location area A(K) in terms of ring K, radius R of the cell, and registration cost; call arrival rate λ_n; the mean and variance of the cell residence time as well as the mean of call duration time; the number of available channels for inter-system roaming users, Q; paging costs and paging delays in systems X and Y, respectively and so forth.

Our results in Figure 3 demonstrate that the total cost, C_T, is reduced by up to 18% using BLA&BLR for the MTs moving from macrocell to microcell tier; and up to 14% for the MTs moving from microcell to macrocell tier. [28] The difference between the cost reductions in two movement directions is caused by the different mobility patterns in macrocell and microcell tiers. We

Radius R (km)		5	10	15	20	25	30	35	40	45
Case A	Pr_{loss}(%)	24.48	20.72	17.94	15.88	14.30	12.90	11.90	10.95	10.11
$\lambda/\xi = 0.8, Q = 2$	$\widehat{Pr}_{loss}$(%)	50	50	50	50	50	50	50	50	50
Case B	Pr_{loss}(%)	13.88	11.38	9.66	8.51	7.53	6.82	6.22	5.68	5.26
$\lambda/\xi = 2.0, Q = 5$	$\widehat{Pr}_{loss}$(%)	20	20	20	20	20	20	20	20	20
Case C	Pr_{loss}(%)	4.78	3.91	3.30	2.89	2.58	2.31	2.10	1.93	1.78
$\lambda/\xi = 3.0, Q = 8$	$\widehat{Pr}_{loss}$(%)	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5

Table 1. Call loss rates of moving from macrocell to microcell tier.

observe that the reduction in call loss rate is very noticeable given that the available channels for external users is scarce and the volume of traffic is high. The call loss rates can be reduced by up to 88% when BLA&BLR are used. The delays are evaluated under the sequential paging scheme, in which the cells are searched one by one in decreasing order of location probabilities. The paging delays for call-to-mobility ratio (CMR), $CMR = 5$ and $CMR = 0.5$ are investigated, which show that the paging delays are reduced by using BLR even though the size of LA increases, yet they are less sensitive to the changes of CMR. More detailed results can be found in.[28]

2.3. Inter-System Location Registration and Call Delivery

In wireless networks, the mobility application part (MAP) protocol is related to a set of signaling messages that communicate between the mobile entities such as BSs, mobile switching centers (MSCs), HLR, and VLRs. The MAP protocol is closely related to location update and paging schemes and they consume radio bandwidth as well as cause processing latency. The existing MAP protocols are not appropriate to be used in NG systems because they are not able to support mobile users to inform the systems of their new positions in different networks. Moreover, the signaling costs in terms of number of messages required for information exchange are increased, impacting the bandwidth utilization efficiency. In the meantime, the delay of services delivery is prolonged because searching process involves more than one network. Thus, a new MAP protocol is developed in accordance with the inter-system location update and paging schemes proposed in the previous section.

2.3.1. GLR-MAP and BLR-MAP Protocols

To support the open environment of inter-system roaming, the *gateway location register* (GLR) is used, which conforms to both commonly used existing location management standards IS-41 and GSM.[29] Besides, this gateway converts the signaling and data format from one network to the other. Therefore, the protocols and interfaces in different networks are similar to the mobility support, while the air-interfaces and physical entities are different. According to the UMTS standard ,[30] the VLR sees the GLR as an HLR, and the HLR sees the GLR as a VLR, which is referred as GLR-MAP in the remaining of this proposal.

The GLR-MAP protocol is able to identify MTs appearing in a new system, and it allows for an MT to initiate a call after it finishes location registration in the new system. However, GLR-MAP protocol is not designed for ongoing call connection during inter-system roaming. When an MT has an active call while crossing the boundary of two systems, the MT must request location registration after it receives signals from the new system. As a result, the existing connection is very likely to be interrupted or lost. In addition, the incoming calls are always delivered to the old system regardless the MT's current location. It is not clearly shown how to avoid this waste of signaling in GLR-MAP.[30,29] Finally, it gives rise to a triangular call routing when the incoming call to a roaming MT is originated by an MT in the same new system. If it is the case, the incoming call is routed to the old system first; then, it is delivered to the roaming MT in the new system. As a result, the call delivery time is prolonged.

In order to solve the above problems, an active location management mechanism is proposed, in which the location registration can be finished prior to the arrival of an MT in the new network. Moreover, the call delivery is performed by querying a cache database BLR introduced in Section 2.2.2. This method is named BLR-MAP protocol, which is a set of signaling messages to implement corresponding to the inter-system location update and paging schemes described in Section 2.2. The basic idea of BLR-MAP is that the MT can request location registration of inter-system roaming when it is in an LA adjacent to another system. As a result, the MT may finish signaling transformation and authentication before it arrives at the new system. This is an *active* mechanism compared to the *passive* scheme of GLR-MAP in which the MT requests location registration after it arrives at the new system.[26]

2.3.2. Procedure of Inter-System Location Registration

When an MT moves into a peripheral location area (PLA) of system X, which is an LA adjoining to another network, it receives the location information via the broadcast channel. The MT can move to another system only through the PLAs of a system. If an inter-system roaming occurs, the detailed procedure of location registration is designed. Each step shown in Figure 4 is described as follows:

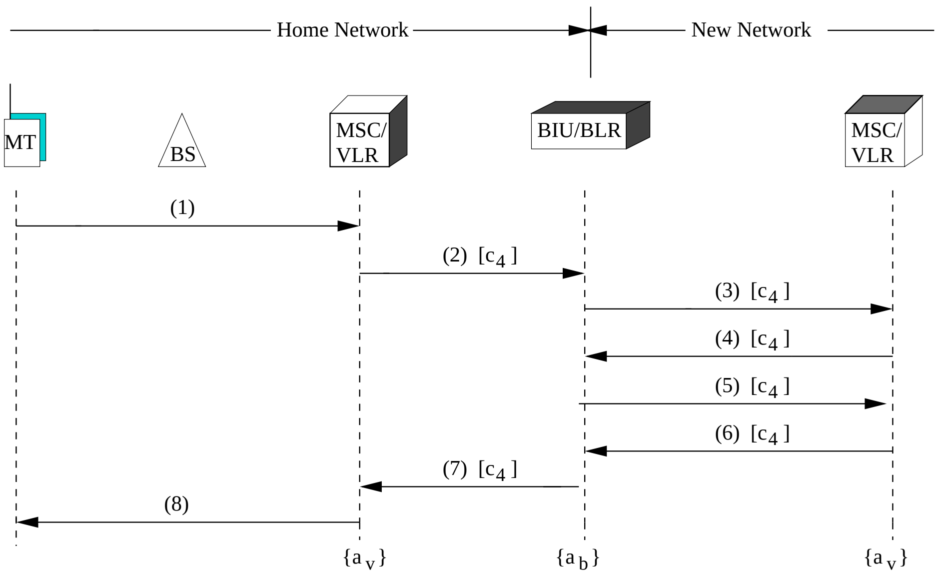


Figure 4. Process of Location Registration/Update Using BLR-MAP.

1. The MT sends a location update message to the MSC/VLR for inter-system roaming through its serving BS.
2. The MSC/VLR sends a location registration request to the BLR with the user information.
3. The BLR stores the MT's location indicating its current MSC/VLR in X and sends location registration message to the MSC/VLR in Y.
4. The MSC/VLR in Y sends a message of the insert subscriber data to the BLR.
5. The BLR sends the user profile of the MT to the MSC/VLR in Y.
6. The MSC/VLR in Y responds to the BLR with a confirmation message.
7. Then the BLR sends an location update acknowledgment message to the MSC/VLR in the PLA of system X.
8. The registration confirmation message is sent from the serving MSC/VLR t o the calling MT through the BS.

where $[c_{\{\cdot\}}]$ denotes the transmission cost of messages between two functional entities, and $a_{\{\cdot\}}$ represents the access cost of each database such as HLR, BLR, VLR, and GLR.

2.3.3. Procedure of Inter-System Call delivery

In order for the NG wireless system to establish the call connection for an MT during the inter-system roaming, the signaling messages involve the HLRs, VLRs, and the BLR associated with the two systems that the MT moves from and into. The detailed procedure of delivering a call to a called MT of home network X (e.g., mobile-to-mobile) is described as follows.

1. A call is initiated by an MT in system X and it is sent to the MSC/VLR through the serving BS.
2. The MSC/VLR sends a location information request message to the HLR in X.

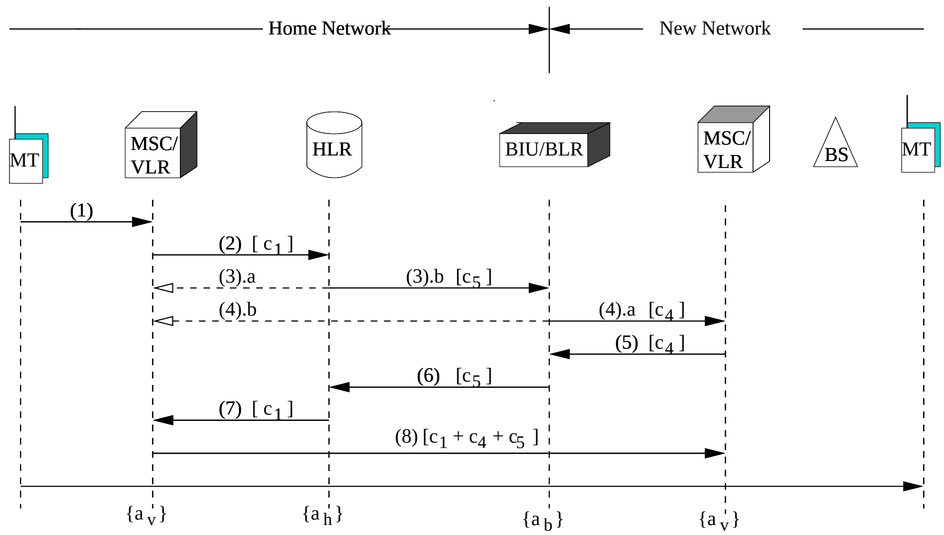


Figure 5. Process of Call Delivery Using BLR-MAP.

3. The procedure of locating the called MT depends on the location information indicated in the HLR in X.

   (a) If the HLR shows the last LA to which the called MT registers is an ordinary LA, i.e., non-PLA, it means that the call delivery follows the procedure in a stand-alone system. We call this case as *intra-system call delivery* in which the call connection is established between the two MSCs serving the called and calling MTs.

   (b) If the HLR shows the last LA to which the called MT registers in X is a PLA adjoining to Y, then the HLR sends a query message to the BIU/BLR associated with systems X and Y. In Figure 5, the detailed procedure of delivering a call to a called MT of home network X (e.g., mobile-to-mobile) is described as follows.

4. The BLR is queried first and it shows the serving MSC/VLR of the called MT. The information of the MSC/VLR is available due to the registration process.

   (a) If the BLR indicates that the called MT has moved to a PLA in system Y, the BLR sends a routing request message to the serving MSC/VLR of the called MT in system Y.

   Otherwise,

   (b) If the BLR indicates that the called MT has not moved to network Y, it means that the called MT is still in X. Then the HLR of X will be queried to find the serving MSC/VLR of the called MT as an intra-system call delivery.

5. The MSC/VLR of called MT responds the routing number to the BLR.

6. The BLR forwards the routing number of the called MT to the HLR of the calling MT.

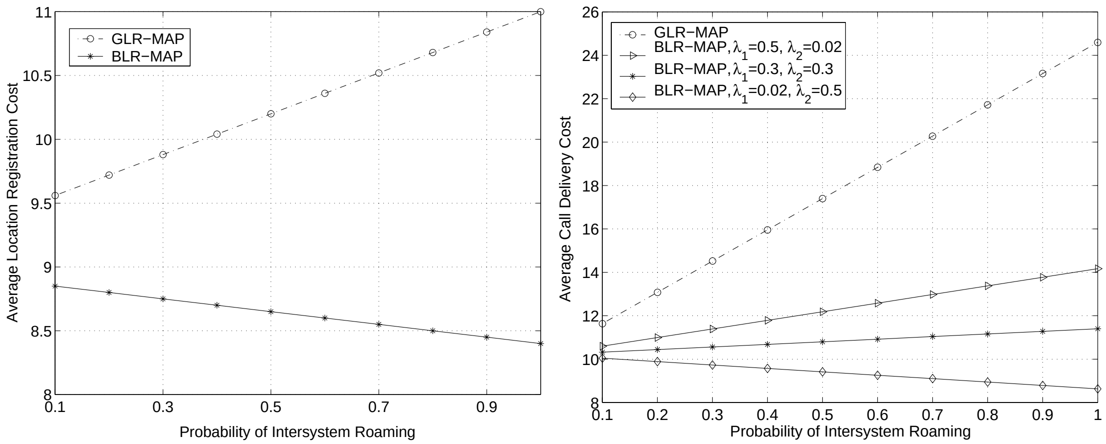
7. The HLR of calling MT forwards the routing number to the MSC/VLR of the calling MT.

8. The call connection is setup between two MSC/VLRs.

9. The call is delivered to the called MT.

2.3.4. Performance Evaluation

Figure 6(a) shows the average location registration cost as a function of inter-system roaming probability. We can observe that the average cost of location registration of GLR-MAP is always higher than that of BLR-MAP. Also, as the inter-system roaming probability increases, the location registration cost of BLR-MAP decreases slightly. The reason for decreasing registration cost is that the inter-system location registration only involves BLR and VLR. Furthermore, the call loss rates and latency of call delivery are analyzed, which also reveal that the BLR-MAP protocol outperforms the GLR-MAP protocol. More results and analysis can be found in.[31]



(a) (b)

Figure 6. Average Location Registration and Call Delivery Cost vs. Inter-system Roaming Probability.

The new BLR-MAP protocol is specifically designed for the MTs with ongoing connections during inter-system roaming. The BLR-MAP protocol enables an MT to update its location and information actively prior to their arrivals at the new systems. Since the BLR is used to provide MTs' up-to-date location information, the incoming calls can be delivered to the inter-system roaming MTs directly, rather than always being delivered to the old system first. Thus, the latency of call delivery and call loss as well as signaling costs are reduced. The latency of call delivery includes retrieval time of database and the waiting time for service in each database, e.g., HLR/VLR/GLR/BLR. For the signal costs, both radio resource and database access are taken into account. The numerical results demonstrate that the proposed protocol improves the overall system performance when an MT is residing in an LA adjoining to another system.

3. LOCATION MANAGEMENT IN 4G WIRELESS SYSTEMS

In order to enable the use of multimedia services, higher bandwidths need to be provided at a lower cost compared to those in 3G systems. 4G wireless systems are proposed to be an extension of the capacity of 3G wireless systems by at least one order of magnitude. 4G wireless systems are supposed to support interactive multimedia services like teleconferencing, wireless Internet, etc and to provide wider bandwidths, even higher data rates, global mobility, service portability, and quality approaching what the traditional broadband wireline network can provide.

There are some innovative features proposed in 4G. First, it starts with entirely packet-switched networks. All network elements are digital. Second, higher bandwidth with up to 100Mbps is provided to support interactive multimedia services at low cost. Third, terminal battery capacity/weight/volume ratios are increased by one order of magnitude compared with mobile terminals used for 3G systems. In addition, tight security features will protect the services from all sorts of security threats such as identity masquerade, unauthorized access, trojan horse, denial of service attack, etc. Moreover, 4G systems will be based on IP technology. Therefore, they will be part of the Internet infrastructure.

4G systems will cover all of the proposed operating environments of the mobile user. It will support radio environments that range from high capacity picocells, to urban terrestrial micro and macrocells which are considered as PCS networks. It also include wireless local area networks (WLANs) and satellite Internet protocol (IP) networks as shown in Figure 7. The mobile IP gateways (MIGs) are used to enable the mobile users to access the IP backbone. For example, an MT in a WLAN will be able to communicate with an MT in the PCS wireless networks through the MIGs, in which the necessary link layer protocols provide the accessibility to the IP backbone. In addition, the protocols should be capable of handling high bit error rate resulted from wireless links and real-time multimedia traffic.

3.1. A New Scheme for Reducing Link and Signaling Costs in Mobile IP

IP mobility support is becoming very important as the Internet is growing fast, and the wireless communication technology is advancing. The basic Mobile IP protocol [32] was proposed to provide IP mobility support. Although the basic Mobile IP protocol proposes a simple and elegant mechanism to provide IP mobility support, there is a major drawback, where each packet destined to the MN must be routed through the HA along an indirect path. This is known as the *triangle routing problem*.

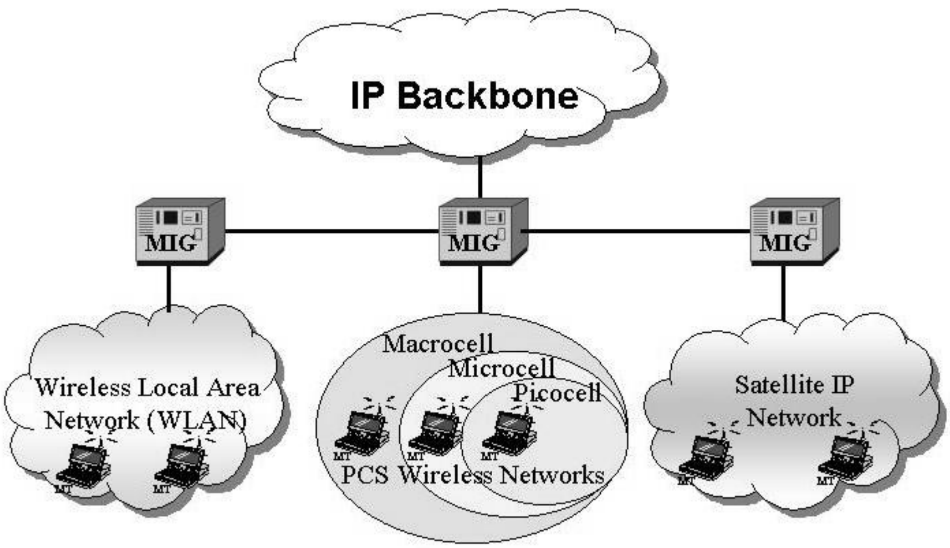


Figure 7. The System Architecture for 4G Wireless Systems.

The so-called *Route Optimization Protocol* (IETF RO)[33] was proposed by the IETF to solve the triangle routing problem. The major drawback of the IETF RO[33] is that there are additional control messages such as binding warning and binding update, which cause communication overhead and introduce high signaling and processing load on the network and on certain nodes. Some mechanisms such as local anchoring scheme,[34] regional registration,[35] and hierarchical management scheme[36] have been proposed to reduce signaling costs and communication overhead recently.

If the route optimization is not performed as often as it is in the IETF RO,[33] signaling and processing load will be reduced. This naturally leads to two issues:

1. How to guarantee that the packets are routed temporarily along a sub-optimal path without performing route optimization.
2. When to perform route optimization.

For the first issue, the FA smooth handoff scheme[33] gives an answer. In this work, we focus on the second issue. Although the route optimization increases the network utilization by allowing packets to be routed along an optimal path from the CN to the MN, it will also increase the signaling load of the network and the processing load of certain nodes. We know from this fact that there is a trade-off between the network resources consumed by the routing path and the signaling and processing load incurred by the route optimization.

We develop a mathematical model to determine when to perform route optimization. Link cost and signaling cost functions are introduced to capture the trade-off. Our objective is to find a cost efficient scheme for route optimization which minimizes the total cost function defined as the sum of the link and signaling cost functions.

The decision must be made in the time interval $[t_i,\ t_{i+1})$ between the current handoff at t_i and the next handoff at t_{i+1} whether to perform route optimization or not. We introduce two cost functions to capture the trade-off: the link and signaling cost functions. The link cost function is denoted by $g(x_i(\alpha_i))$ where $x_i(\alpha_i)$ is the number of links in the routing path between the CN and the FA during the i-th period and $\alpha_i \in \{RO, NRO\}$ where RO is the action which performs route optimization, and NRO is the action without route optimization. The signaling cost function is denoted by $h(y_i, \alpha_i)$ where y_i is the number of links in the shortest path between the current FA and the previous FA during the i-th period.

Let $\pi(i)$ denote a sequence of actions, $(\alpha_1, \alpha_2, \ldots, \alpha_i)$, which are taken sequentially during the occurrence of i handoffs. We call this sequence $\pi(i)$ as a *Route Optimization Sequence*. Let $F_i^{\pi(i)}$ denote the total accumulative cost under the route optimization sequence $\pi(i)$. Then,

$$F_i^{\pi(i)} = \sum_{j=1}^{i} f(x_j(\alpha_j), y_j, \alpha_j)$$

$$= \sum_{j=1}^{i} \{g(x_j(\alpha_j)) + h(y_j, \alpha_j)\} \quad (4)$$

where $\pi(i) = (\alpha_1, \ldots, \alpha_i)$. Here $x_j(\alpha_j)$ and y_j are network parameters.

The objective of this work is to find the optimal sequence which we denote as $\pi_{opt}(i)$ which minimizes the expected value of total cost $F_i^{\pi(i)}$ in (4).

$$\begin{aligned} E[F_i^{\pi_{opt}(i)}] &= \min_{\pi(i)\in\Pi} E[F_i^{\pi(i)}] \quad (5) \\ &= \min_{\pi(i)\in\Pi} \sum_{j=1}^{i} (E[g(x_j(\alpha_j)) + h(y_j, \alpha_j)]) \end{aligned}$$

where $\pi(i) = (\alpha_1, \ldots, \alpha_i)$, and Π is the set of all possible sequences of $\pi(i)$.

We restrict our model within intra-domain (intra-subnet) handoff and make a reasonable assumption, i.e., *if handoffs occur in the same domain (subnet), the length of the shortest path between the CN and any FA is the same, i.e.,* $x'_i = x'_j$ *for* $i \neq j$.

The expected value of total accumulative cost, $E[F_i^{\pi_{opt}(i)}]$, becomes

$$E[F_i^{\pi_{opt}(i)}] = E[F_{i-1}^{\pi_{opt}(i-1)} + \min(f(x_i(NRO), y_i, NRO), f(x_i(RO), y_i, RO))]. \quad (6)$$

From (6) a decision rule can be found.

Decision Rule:
if $E[f(x_i(NRO), y_i, NRO)] < E[f(x_i(RO), y_i, RO)]$ **then**
$\alpha_i = NRO$;
else
$\alpha_i = RO$;
end if

Thus, the optimal sequence, $\pi_{opt}(i)$, can be obtained by following the above decision rule in each decision stage.

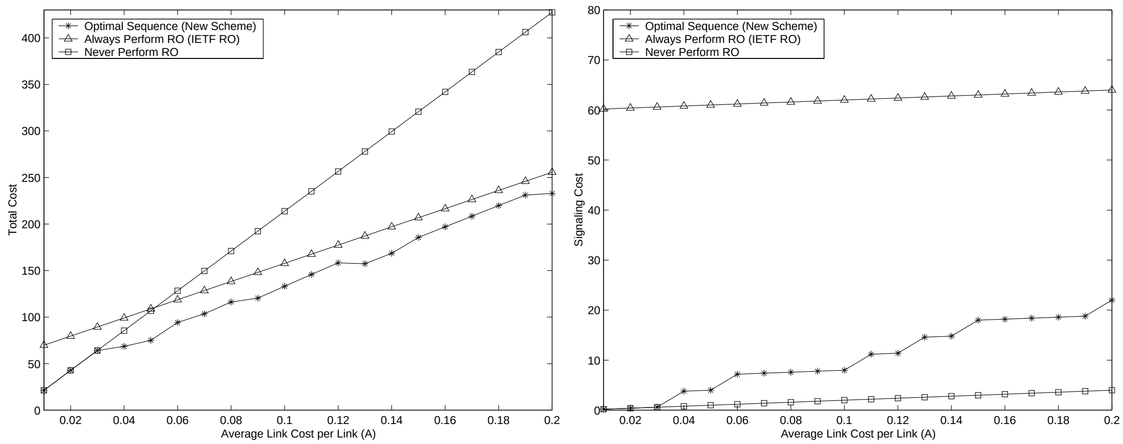
Finally we evaluate the performance of the proposed scheme, $\pi_{opt}(\tilde{N})$, for route optimization and compare it with other schemes which are explained below:

- Scheme 1: Always perform route optimization.
  $\pi_{ARO}(\tilde{N}) = (\alpha_1, \ldots, \alpha_{\tilde{N}})$ where $\alpha_i = RO$ for $i = 1, \ldots, \tilde{N}$.
- Scheme 2: Never perform route optimization.
  $\pi_{NRO}(\tilde{N}) = (\alpha_1, \ldots, \alpha_{\tilde{N}})$ where $\alpha_i = NRO$ for $i = 1, \ldots, \tilde{N}$.

In this work, we proposed a cost efficient scheme for route optimization to reduce the signaling cost caused by the route optimization. A decision rule is derived from this model. The optimal sequence π_{opt} is obtained by following the decision rule in each decision stage. The performance of the optimal sequence π_{opt} is compared with the other sequences π_{ARO} and π_{NRO}. The simulation results show that the optimal sequence π_{opt} provides the lowest total costs among the given sequences.

4. CONCLUSIONS

There are many ongoing research activities on 3G and 4G systems. This paper presents the system architecture of 3G and 4G systems from the perspective of location management. Especially, we put our efforts on the inter-system roaming because it is a new topic compared to well-developed location management techniques for stand-alone systems. A dynamic location management scheme for inter-system roaming is developed, which combines inter-system location update and paging schemes. The inter-system location update is implemented based on the concept of BLA, which is dynamically determined by the MT's velocity and network load. This process greatly reduces the call loss rates. The inter-system paging is realized through a cache



(a) $B = A$, K = 3, N = 20 (b) $B = A$, K = 3, N = 20

Figure 8. Total Cost and Signaling Cost vs. Average Link Cost per Link A.

database named BLR which maintains the roaming information of mobile terminals moving across two systems. Numerical results demonstrate that the proposed scheme results in significant performance improvements for multi-tier PCS systems.

An enhanced MAP protocol for NG Wireless systems is proposed to reduce signaling costs and delays of inter-system roaming. The detailed procedure of location registration and call delivery are devised to resolve the problems in the current GLR-MAP protocol. The overall system performance is analyzed with respect to the signaling cost of location registration and call delivery, the delay of location management, and the call loss rate due to the location registration from one system to another for various scenarios.

The main drawback of the basic Mobile IP protocol is that the packets must be routed along the paths longer than the optimal one, which is known as the triangle routing problem.[32] We introduced a new scheme for reducing costs in route optimization to solve the triangle routing problems. In this new scheme, route optimization is performed only when it minimizes the total cost function, which provides the optimal result from the viewpoint of link and signaling costs. The simulation results show that the proposed scheme provides better performance than the basic scheme.[37]

REFERENCES

1. S. Dixit and Z. Antoniou, "Resource Management and Quality of Service in Third-Generation Wireless Networks," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 125–133, February 2001.
2. W. Groenen, "Towards the Needs of Third-Generation Systems: Planning the Evoluation of the GSM Platform," *British Telecom Engineering*, vol. 18, no. 2, pp. 21–24, August 1999.
3. T. Nilsson, "Toward Third-Generation Mobile Multimedia Communication," *Ericsson Review*, , no. 3, pp. 122–131, 1999.
4. M. Nilsson, "Third-Generation Radio Access Standards," *Ericsson Review*, , no. 3, pp. 110–121, 1999.
5. M. Gallagher and W. Webb, "TUMTS–the Next Generation of Mobile Radio," *IEE Review*, vol. 45, no. 2, pp. 59–63, March 1999.
6. I.F. Akyildiz, J. Mcnair, J.S.M. Ho, H. Uzunalioğlu, and W. Wang, "Mobility Management in Next-Generation Wireless Systemes," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1347–1384, August 1999.
7. A. Ganz, C.M. Krishna, D. Tang, and Z.J. Haas, "On Optimal Design of Multitier Wireless Cellular Systems," *IEEE Communications Magazine*, vol. 35, no. 2, pp. 88–93, February 1997.
8. R. Pandya, D. Grillo, E.Lycksell, P. Mieybégué, and M. Yabusaki, "IMT-2000 Standards: Network Aspect," *IEEE Personal Communications*, vol. 4, no. 4, pp. 20–29, August 1997.
9. I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioğlu, and W. Wang, "Mobility Management in Current and Future Communications Networks," *IEEE Network*, vol. 12, no. 4, pp. 39–49, July/August 1998.
10. T. Brown and S. Mohan, "Mobility Management for Personal Communications Systems," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 2, pp. 269–278, May 1997.
11. S. Tabbane, "Location Management Methods for Third-Generation Mobile Systems," *IEEE Communications Magazine*, vol. 35, no. 8, pp. 72–84, August 1997.

12. B-N. Amotz, I. Kessler, and M. Sidi, “Mobile Users: To Update or not to Update?,” in *IEEE INFOCOM’94*, vol. 2, pp. 570–576, June 1994.
13. A. Bhattacharya and S. K. Das, “LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks,” in *ACM/IEEE MobiCom’99*, pp. 1–12, August 1999.
14. Y-B. Lin and I. Chlamtac, “Heterogeneous Personal Communications Services: Integration of PCS Systems,” *IEEE Communications Magazine*, vol. 34, no. 9, pp. 138–145, September 1996.
15. M. Vudali, “The Location Area Design Problem in Cellular and Persoanl Communications System,” in *IEEE ICUPC’95*, pp. 591–595, November 1995.
16. V. Garg and J. E. Wilkes, “Interworking and Interoperability Issues for North American PCS,” *IEEE Communications Magazine*, vol. 34, no. 3, pp. 94–99, March 1996.
17. S.G. Niri and R. Tafazoli, “Cordless - Cellular Network Integration for the 3rd Generation Personal Communication Systems,” in *IEEE VTC’98*, pp. 402–408, May 1998.
18. A. Yamaguchi, S. Ota, Y. Ito, M. Ohashi, and F. Watanabe, “Inter-system Mobility and Service Management in GSM/PDC Roaming,” in *IEEE GLOBECOM’97*, pp. 694–698, November 1997.
19. W. Wang and I.F. Akyildiz, “Intersystem Location Update and Paging Schemes for Multitier Wireless Networks,” in *ACM MobiCom’00*, pp. 99–109, August 2000.
20. W. Wang and I.F. Akyildiz, “A Predictive Paging Scheme for IMT-2000 Systems,” in *Proc. of the IEEE Vehicular Technology Conference (VTC)’00*, September 2000.
21. S. Tabbane H. Xie and D.J. Goodman, “Dynamic Location Area Management and Performance Analysis,” in *IEEE 42nd VTC’92*, pp. 536–539, May 1992.
22. G. Wan and E. Lin, “Cost Reduction in Location Management Using Semi-realtime Movement Information,” *ACM-Baltzer Journal of Wireless Networks (WINET)*, vol. 5, no. 4, pp. 245–256, 1999.
23. G.L. Lyberopoulos J.G. Markoulidakis and M.E. Anagnostou, “Traffic Model for Third Generation Cellular Mobile Telecmmunications Systems,” *ACM-Baltzer Journal of Wireless Networks (WINET)*, vol. 4, pp. 389–400, August 1998.
24. J.S.M. Ho and I.F. Akyildiz, “Mobile User Location Update and Paging Under Delay Constraints,” *ACM-Baltzer Journal of Wireless Networks (WINET)*, vol. 1, no. 4, pp. 413–425, December 1995.
25. Y-B. Lin, “Reducing Location Update Cost in a PCS Network,” *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 25–33, February 1997.
26. I.F. Akyildiz and W. Wang, “A Dynamic Location Management Scheme for Next Generation Multi-tier PCS Systems,” *to appear in IEEE Journal on Selected Areas in Communications (JSAC)*, 2001.
27. R. Thomas, H. Gilbert, and G. Mazziotto, “Influence of the Movement of the Mobile Station on the Performance of the Radio Cellular Network,” in *3rd Nordic Sem.*, p. 9.4, September 1988.
28. W. Wang and I.F. Akyildiz, “A Dynamic Inter-System Location Management Scheme for Next Generation Multi-Tier PCS Systems,” .
29. A. Bertrand, “Jambala Mobility Gateway – Convergence and Inter-System Roaming,” *Ericsson Review*, vol. 76, no. 2, pp. 89–93, 1999.
30. ETSI TS 129 120 V3.0.0 (2000-03), “Universal Mobile Telecommunications Systems (UMTS); Mobile Application Part (MAP) Specification for Gateway Location Register (GLR); Stage 3,” *3GPP/ETSI*, 2000.
31. W. Wang and I.F. Akyildiz, “An Cost-Efficient Mobility Application Part (MAP) Protocol for IMT-2000 Systems,” in *ACM MobiCom’01*, July 2001.
32. C. Perkins, “IP Mobility Support,” in *RFC 2002*, October 1996.
33. C. Perkins and D. Johnson, “Route Optimization in Mobile IP (work in progress),” in *Internet Draft, IETF*, November 2000.
34. J. Ho and I. Akyildiz, “Local Anchor Scheme for Reducing Signaling Costs in Personal Communications Networks,” *IEEE/ACM Trans. Networking*, vol. 4, no. 5, pp. 709–725, October 1996.
35. E. Gustafsson, A. Jonsson, and C. Perkins, “Mobile IP Regional Registration (work in progress),” in *Internet Draft, IETF*, July 2000.
36. H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier, “Hierarchical MIPv6 mobility management (work in progress),” in *Internet Draft, IETF*, November 2000.
37. I. Akyildiz and Y. Lee, “A New Scheme for Reducing Link and Signaling Costs in Mobile IP,” in *WPMC’01*, September 2001.