Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Computer Networks 46 (2004) 797–816

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

On signaling performance bounds of location management in Next Generation Wireless Networks

Eylem Ekici [*]

Department of Electrical and Computer Engineering, The Ohio State University, 2015 Neil Ave., 320 Dreese Labs, Columbus, OH 43210, USA

Received 16 December 2003; received in revised form 8 May 2004; accepted 14 June 2004
Available online 31 July 2004

Responsible Editor: B. Baykal

Abstract

Next Generation Wireless Networks (NGWN) are proposed to achieve the goal of ubiquitous broadband networking by utilizing multiple wireless access subnetworks serving overlapping areas. Location management schemes play a very important role in NGWN since mobile users roam in coverage areas of these subnetworks simultaneously. The performance of location management schemes directly affect the overall performance of NGWN at large. In this paper, signaling performance bounds achievable by a location management scheme in the wireless portion of NGWN are presented. Assuming complete knowledge about user mobility, call arrival patterns, detailed maps of subnetworks coverage areas, and other NGWN parameters, equations for signaling performance achievable by an idealized location management scheme over the wireless interface has been derived. The performance bounds presented in this paper serve as an upper bound for the performance of any other location management scheme designed for NGWN. These bounds are intended to serve as a benchmark to determine how well a proposed location management scheme operates in NGWN and will help to determine how much room for improvement exists.

Keywords: Next Generation Wireless Networks; Location management; Location registration; Paging; Performance bounds

1. Introduction

Location management is one of the key factors determining the performance of wireless networks. Location management in wireless networks involves tracking the location of the mobile terminals (MTs) in the service area. Existing cellular

---

[*] Tel.: +1 614 292 0495; fax: +1 614 292 7596.
E-mail address: ekici@ece.osu.edu

1389-1286/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2004.06.019

systems such as IS-41 [1] and GSM [2] utilize location management techniques primarily based on tracking users at the location area (LA) granularity, which is a predetermined set of cells where MTs are not required to refresh their location information unless they cross LA boundaries. MTs are located by paging them in all cells in the last known LA. In literature, other location management techniques have also been proposed to achieve higher performance. The location update techniques based on the time elapsed between successive location registrations, number of cells crossed, and the physical distance between last registered location and current location have been analyzed in [3]. In [4], a movement-based location update mechanism and a selective paging procedure have been introduced. An improved version of the movement-based location registration scheme has been presented in [5]. The movement-based location update approach was also extended from cells to location areas in [6] and [7]. An example of the distance-based location management technique where the paging is performed based on MT's estimated location is presented in [8]. In [9], an adaptive distance-based location update algorithm is proposed aiming to compute optimal location update boundary for individual MTs.

Next Generation Wireless Networks (NGWN) are proposed to achieve the goal of ubiquitous broadband networking. NGWN architectures proposed in the literature [10–13] all envision varying levels of collaborative use of multiple wireless access techniques ranging form high capacity picocells to satellite-based cells to provide world-wide coverage and high data rates. To provide seamless integration with the Internet, packet-based access systems are proposed for data communications in second and third generation (2G and 3G) [14] cellular systems, which will eventually evolve into all-IP wireless systems [15]. Location management issues for next generation wireless networks are presented in [16] and [17]. Among the specific solutions targeting the location management problem in next generation wireless, methods presented in [18–20] propose using boundary location registers and boundary interworking units to allow the movement of MTs among tiers of multi-tier wireless systems. Forming dynamic boundary location areas between the tiers of the wireless system, efficient transitions from one tier to the next is accomplished. These solutions focus on improving location management performance near the boundary of different tiers of the network and do not address the overall location management problem in NGWN.

We aim to look at NGWN beyond the generations as they are currently defined and propose an alternative NGWN architecture that serves as a superset of all proposed NGWN architectures. The new NGWN architecture assumes coexistence of multiple cellular systems of possibly different characteristics, such as coverage, access techniques, and bandwidth, connected to each other over a common backbone network and a set of management nodes. The cellular systems that constitute the NGWN is referred to as *subnetworks*. Our new NGWN architecture does not assume any predefined relationship between the subnetworks. This will enable our NGWN architecture to include all intermediary systems that will be generated by combining existing systems and adding new wireless systems to the existing ones. Within the NGWN framework, we assume that mobile terminals can directly communicate with any set of subnetworks.

In this paper, we study the performance bounds of location management schemes in the wireless portion of NGWN. Using our new proposed network architecture as the basis, we will explore how different system parameters affect the location management signaling cost on the wireless link. Following an analytical approach, we derive equations to assess the highest location management signaling performance achievable in NGWN considering the wireless interface between the NGWN infrastructure and the wireless devices. These performance bounds are intended to serve as benchmarks for new proposed schemes and will help to comprehend how well a proposed scheme performs or if it can be further improved.

The remainder of the paper is organized as follows: in Section 2.1, the new Next Generation Wireless Network Architecture (NGWN) and the idealized location management system are introduced. The signaling performance measures for location management in NGWN are presented in

Section 3. Location management performance bounds and effect of a set of parameters are presented in Section 4 for sample system parameters. Finally, Section 5 concludes this paper.

2. System description

2.1. NGWN architecture

In current wireless systems, the applications enabled in the system are determined by the bandwidth availability and the coverage area, which are in general conflicting characteristics. Although improvements in bandwidth are expected in next generation wireless systems, designing NGWN as a heterogeneous architecture that will include all existing as well as all future wireless systems is a more viable solution. By defining NGWN as a collection of subnetworks, the flexibility in gradual network deployment and in network expansion is introduced.

Our new NGWN architecture can be viewed as a collection of various wireless access networks that are connected over a common backbone network, i.e., the Internet, and share a common resource and network management system. In Fig. 1, the components of NGWN are depicted. The system consists of $\mathscr{N}$ subnetworks $\mathbf{S}_i$, $i = 1, \ldots, \mathscr{N}$, and the Internet as the backbone. Although all subnetworks in this figure are different, subnetworks of same technology are permitted even in the case of overlapping coverage. Every subnetwork $\mathbf{S}_i$ is served by a *Subnet Location System* SLS_i. SLSs perform the location management functions for the mobile terminals for which they carry the location management responsibility. SLSs can be compared to current location management systems consisting of HLRs and VLRs and can be replaced by them. NGWN also employs another entity located in the backbone that can be regarded as the counter part of SLS. The *Global Location System* GLS acts as a coordinator that collects information from SLSs and tracks MTs across the system in all subnetworks. Detailed topology information of the NGWN coverage area is maintained in a global database called *Topology Database* TDB. GLS uses the topology information stored in TDB to pinpoint an MT precisely across all subnetwork coverage areas.

The coverage area of each subnetwork is divided into cells. The entire coverage area of NGWN, denoted as $\mathbf{C}^T$ is the set union of the coverage areas in of the individual subnetworks, i.e., $\mathbf{C}_T = \bigcup_{i=1,\ldots,\mathscr{N}} \mathbf{C}_i$, where $\mathbf{C}^i$ is the coverage area of the subnetwork $\mathbf{S}_i$. We assume that the coverage area of subnetworks heavily overlap. Overlapping coverage areas increase the overall network availability as well as the diversity in available resources. Every subnetwork coverage area $\mathbf{C}^i$ is composed of a set of cells C^i_j, $j = 1, \ldots, N^C_i$, where N^C_i is the number of cells of the subnetwork $\mathbf{S}_i$. The cell definition used in current PCS and satellite systems is extended to include the WLAN coverage served by

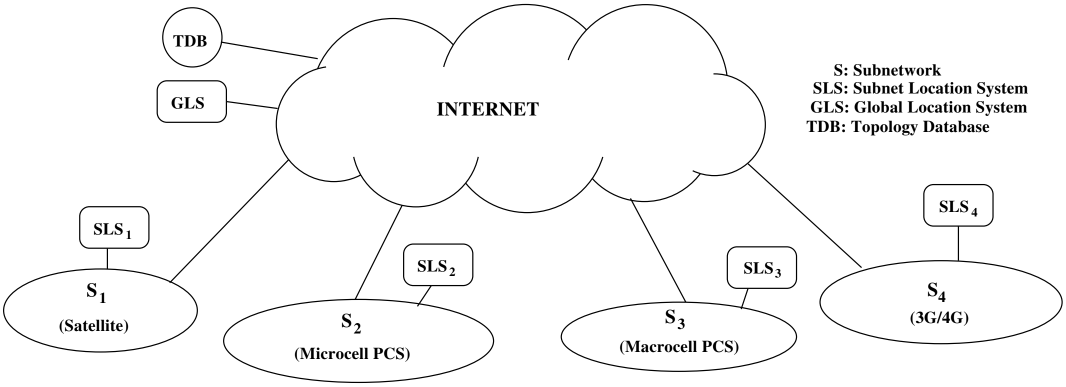

Fig. 1. Next Generation Wireless Network architecture.

individual WLAN access points. In general, the system can be expanded to include all wireless systems where MTs can reach the fixed network over a single wireless hop. The cells of the four subnetworks of Fig. 1 covering the same service are depicted in Fig. 2. Note that no special relationship between cell sizes or orientation is assumed according to the principle of NGWN being a collection of subnetworks.

Location tracking of mobile terminals is preformed at the *location area* (LA) level. A location area is composed of a set of neighboring cells belonging to the same subnetwork. In NGWN, every subnetwork $\mathbf{S}_i$ is partitioned into mutually exclusive location areas $LA^i_j,\ j = 1, \ldots, N^{LA}_i$, where N^{LA}_i is the number of location areas in the subnetwork $\mathbf{S}_i$. The location areas belonging to different subnetworks may (fully or partially) overlap.

MTs are assumed to be equipped with wireless interfaces enabling them to communicate with multiple subnetworks simultaneously. Furthermore, MTs are also assumed to be able to maintain multiple simultaneous data flows over the same or different subnetworks. However, MTs only need to perform very basic location update operations and need to respond to simple paging queries. Effectively, the bulk of the location management functions are shifted from MTs to the network.

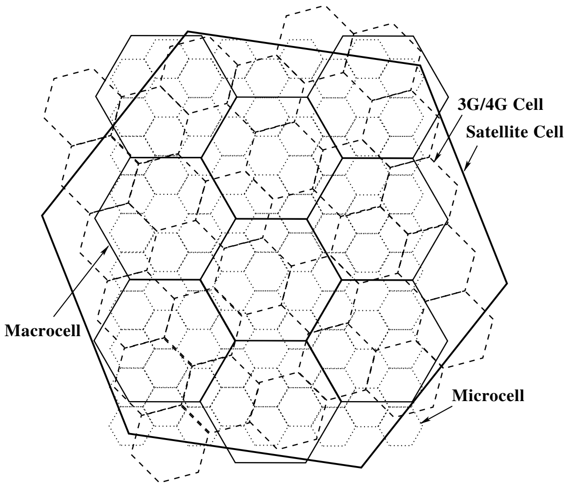

Fig. 2. Sample cellular coverage area of four subnetworks.

2.2. Idealized location management scheme

In the new proposed NGWN architecture, the location management systems of the subnetworks are abstracted to SLSs. SLS_i of a subnetwork $\mathbf{S}_i$ is responsible for processing the location registration messages as well as paging MTs in the coverage area of $\mathbf{S}_i$. We assume that all registration messages are collected at the GLS, which processes the paging requests, as well. Let us consider a case of a mobile terminal MT_i moving in the coverage area of NGWN. Let us also assume that MT_i is tracked at LA granularity in a subset of subnetworks $\mathbf{TS}_i = \{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$ where P is the number of subnetworks in the set. Note that $\{\mathbf{S}_t\}_{t=1}^{\mathcal{N}} \supseteq \mathbf{TS}_i$, i.e., there may be subnetworks in NGWN which can deliver calls to MT_i, but which do not track MT_i's location. When MT_i crosses an LA boundary of a subnetwork $\mathbf{S}_j$, it sends a registration message to SLS_j only if $\mathbf{S}_j$ is tracking its location, i.e., $\mathbf{S}_j \in \{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$. All location registration messages sent to SLS_j are directly forwarded to GLS. Note that MT_i updates its location information only when it does not have any active connections over any subnetwork. While a connection is in progress, we assume that system can calculate the location of MT_i in $\mathbf{TS}_i$. Upon completing all active connections, MT_i resumes with updating its location in subnetworks listed in $\mathbf{TS}_i$.

If MT_i has active connections over a set subnetworks $\mathbf{AS}_i = \{\mathbf{S}_{c_1}, \ldots, \mathbf{S}_{c_A}\}$ and a call over another subnetwork $\mathbf{S}_n \notin \mathbf{AS}_i$ is received, then MT_i can be queried about its location in $\mathbf{S}_n$ over one of the subnetworks $\mathbf{S}_x \in \mathbf{AS}_i$ without requiring paging. Paging is required when MT_i has no active connections when a call is received. In this case, GLS is responsible for creating a paging set in which MT_i should be paged. The paging set **PS** is constructed by intersecting the last known location areas of MT_i where it is tracked and mapping it onto the coverage area of the subnetwork $\mathbf{S}_t$ where the call is received. Formally, paging set $\mathbf{PS}_i$ can be defined as $\mathbf{PS}_i = \mathscr{M}_t(\bigcap_{q=1}^{P} LA^{k_q}_{s_q})$, where $\{LA^{k_1}_{s_1}, \ldots, LA^{k_P}_{s_P}\}$ is the set of most recent LAs in subnetworks $\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}$ where MT_i has registered its location, and $\mathscr{M}_n(\cdot)$ is a function returning the set of cells in subnetwork $\mathbf{S}_n$ that covers the area

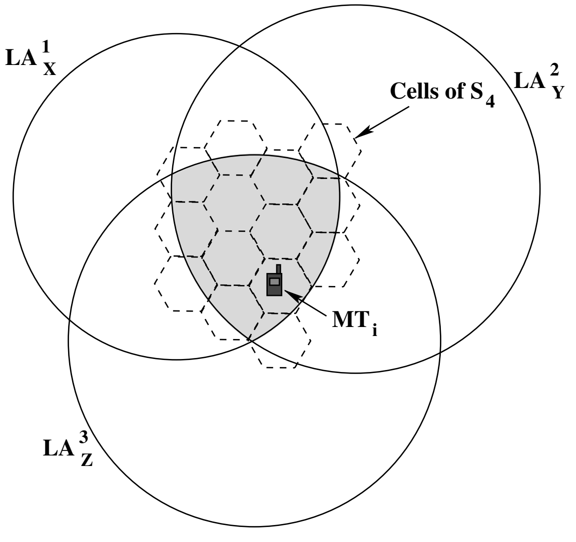

Fig. 3. Creation of the paging set.

passed as the argument. Note that $\mathbf{S}_n$ does not necessarily track MT_i's location. After GLS determines the paging set $\mathbf{PS}_i$, it sends the paging request to SLS_n, which, in turn, pages MT_i in cells specified $\mathbf{PS}_i$. An example scenario is shown in Fig. 3. In this figure, MT_i is tracked in subnetworks $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$, and has registered its location most recently in LA_X^1, LA_Y^2, and LA_Z^3, respectively. When a call arrives for MT_i in subnetwork $\mathbf{S}_4$, GLS creates the paging set **PS** by calculating the intersection area of LA_X^1, LA_Y^2, and LA_Z^3 (shaded region), and determining the cells of subnetwork $\mathbf{S}_4$ that fully covers the intersection area (dashed hexagons). The paging is then performed only in the cells belonging to $\mathbf{PS}_i$.

One of the questions that remains unanswered is how the set of subnetworks tracking a particular MT is determined. In the ideal case, GLS should be able to accurately predict the movement behavior of all MTs individually. Similarly, the call arrival patterns and call characteristics are also tracked for individual MTs. Based on these inputs, GLS calculates the set of subnetworks tracking a particular MT such that the overall signaling cost is minimized. [1] As will be presented in Section 3, this set will be determined by considering all non-empty subsets of subnetworks of NGWN. Hence, to determine the ideal subset, one needs to consider $2^{\mathcal{N}} - 1$ subnetwork combinations, where $\mathcal{N}$ is the total number of subnetworks in NGWN. Note that this selection must be performed for all MTs to achieve the lowest location management signaling cost and must be updated whenever call arrival and mobility rates change.

Although the location registration and paging methods described above provide lowest signaling cost, it is extremely difficult, if not impossible, to realize these solutions in practice. The main reasons for this can be summarized as follows:

1. GLS is responsible for all location management functions and therefore constitutes a bottleneck. Although distributed implementations can be adopted, an entity that supervises all location management functions is required to achieve the same degree of performance as the system described above.
2. In practice, it is impossible to maintain a high precision global map of all subnetworks at cell granularity. Furthermore, since all paging requests will create queries to TDB, TDB is another potential bottleneck of the system.
3. Determining the set of subnetworks where MTs will be tracked is very hard to accomplish given the amount and detail level of information needed. Although utilizing user mobility patterns with varying degrees of regularity have been proposed in literature [21], it is potentially infeasible to implement high precision real-time user tracking systems in networks with very high user population, very large coverage area, and multiple subnetworks of picocells.

Due to reasons presented above, the Idealized Location Management Scheme is not suitable for implementation. However, the focus of this paper is *not* to introduce a new location management scheme, either. In this paper, we intend to introduce a very generalized and flexible network architecture and present the bounds of location management signaling performance in this new environment for the purpose of serving as a benchmark for new location management schemes.

---

[1] It is assumed that all call deliveries have strict delay constraints and MTs must be located within one paging cycle.

Should new advancements enable more feasible implementations of concepts introduced in this section, the performance of the location management schemes employing these concepts will approach the bounds that will be introduced in Section 3.

3. Performance bounds of NGWN

In this section, we will study the performance measures of the idealized location management scheme in NGWN and derive the equations to calculate these measures. We are interested in the cost incurred by the location management scheme in the wireless portion of NGWN as the wireless links are more likely to constitute bottlenecks that may limit the performance of NGWN. The signaling and processing costs inside the wired portion of NGWN are excluded from the calculations.

3.1. Call arrivals

The connections destined to a mobile terminal MT_i can arrive from any of the subnetworks in NGWN. The matching of the subnetworks and the connections depends on various factors such as characteristics of mobile terminal, connection requirements, subnetwork characteristics, and location and connectivity of the other party. The choice of the subnetwork for individual connection requests is beyond the scope of this paper. We will, however, assume that the connection requests over different subnetworks are independent from each other. Furthermore, it is assumed that connections over different subnetworks arrive at different rates and have different call holding times. We also assume that all subnetworks can support an arbitrary number of connections to the same mobile terminal. For the sake of simplicity of notation, the calls generated by the mobile are not included in the calculations. However, the outlined calculations can be used to incorporate the mobile-generated calls, as well. Without loss of generality, connection arrival and holding processes can formally be described as follows.

Let the connection arrivals for MT_i over a subnetwork $\mathbf{S}_j$ follow a Poisson distribution with rate $\lambda^c_{i,j}$, where $\mathbf{S}_j \in \{\mathbf{S}_i,\ i = 1,\ldots,\mathcal{N}\}$. Consequently, the time between call arrivals $t^c_{i,j}$ for MT_i over subnetwork $\mathbf{S}_j$ is exponentially distributed with $f^c_{i,j}(t) = \lambda^c_{i,j}\mathrm{e}^{-\lambda^c_{i,j}t}$. Furthermore, let $f^h_{i,j}(t) = \mu^h_{i,j}\mathrm{e}^{-\mu^h_{i,j}t}$ be the connection holding time $t^h_{i,j}$ of MT_i in subnetwork $\mathbf{S}_j$, where the mean connection holding time is $E[t^h_{i,j}] = 1/\mu^h_{i,j}$. The connection state $(N_i^1,\ldots,N_i^{\mathcal{N}})$ represents the number of active connections N_i^j of MT_i in all subnetworks $\mathbf{S}_j$, $j = 1,\ldots,\mathcal{N}$. In Fig. 4, Markov chain for the isolated connection state N_i^j of MT_i in subnetwork $\mathbf{S}_j$ is shown. The steady-state probability $\Pr[N_i^j = k]$ that the number of active connections of MT_i in subnetwork $\mathbf{S}_j$ is k is calculated as follows:

$$\Pr[N_i^j = k] = \Pr[N_i^j = 0]\frac{(\rho_{i,j})^k}{k!}, \tag{1}$$

where $\Pr[N_i^j = 0] = \mathrm{e}^{-(\rho_{i,j})}$ and $\rho_{i,j} = \lambda^c_{i,j}/\mu^h_{i,j}$.

The steady-state probability of the connection state of MT_i over all subnetworks $\Pr[(N_i^1,\ldots,N_i^{\mathcal{N}}) = (k_1,\ldots,k_{\mathcal{N}})]$ is the product of individual steady-state probabilities $\Pr[N_i^j = k_j]$, $j = 1,\ldots,\mathcal{N}$:

$$\Pr[(N_i^1,\ldots,N_i^{\mathcal{N}}) = (k_1,\ldots,k_{\mathcal{N}})] = \prod_{j=1}^{\mathcal{N}} \Pr[N_i^j = k_j]. \tag{2}$$

Then, the probability that an MT is idle, i.e., that it has no active connections, is given by the following equation:

$$\begin{aligned}\Pr[\mathrm{MT}_i \text{ is idle}] &= \prod_{j=1}^{\mathcal{N}} \Pr[N_i^j = 0] \\ &= \exp\left(-\sum_{j=1}^{\mathcal{N}} \rho_{i,j}\right),\end{aligned} \tag{3}$$

where $\rho_{i,j} = \lambda^c_{i,j}/\mu^h_{i,j}$.

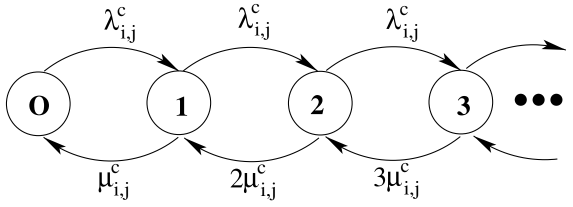

Fig. 4. Isolated connection state for MT_i in subnetwork $\mathbf{S}_j$.

3.2. Location registration

As described in Section 2.2, a mobile terminal MT_i's location is tracked at LA granularity in a subset of subnetworks $\mathbf{TS}_i = \{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$. When MT_i moves in the NGWN coverage area, it generates location registration messages when it crosses the location area boundaries of subnetworks in **TS** when it is idle. As shown in Fig. 2, there is no predefined relationship between sizes and positions of the location areas and cells of different subnetworks. Therefore, we can model the movement of MT_i in different subnetworks as independent events, i.e., as MT_i moves in overlapping coverage areas of subnetworks in **TS**, it generates mutually independent registration messages. Based on this observation, we will derive the distribution of number of registration messages between two active periods of MT_i.

Let $f^m_{i,j}(t)$ be the probability density of the residence time $t^m_{i,j}$ of MT_i in the location areas of the subnetwork $\mathbf{S}_j$ with the mean LA residence time $E[t^m_{i,j}] = 1/\lambda^m_{i,j}$. To calculate the distribution of the number of registration messages $K_{i,j}$ generated by MT_i in subnetwork $\mathbf{S}_j$, we need to consider the overall connection arrival process since MT_i generates location registration messages only when it is idle. A call arrival over any subnetwork pauses the location registration process of MT_i. Since individual connection arrival processes are independent Poisson processes with rates $\lambda^c_{i,j}$, the resulting overall connection arrival process is also a Poisson process with rate equal to the sum of the original rates. The overall connection arrival rate $\lambda^c_{i,\Sigma}$ is calculated as follows:

$$\lambda^c_{i,\Sigma} = \sum_{j=1}^{\mathscr{N}} \lambda^c_{i,j}, \tag{4}$$

where $\mathscr{N}$ is the total number of subnetworks in NGWN. The movement of an MT in the coverage area of a single network has been analyzed in [22]. Using the same approach, the probability distribution $\alpha_{i,j}(K)$ of the number of registration messages $K_{i,j}$ created by MT_i in $\mathbf{S}_j$ is calculated as

$$\alpha_{i,j}(K) = \begin{cases} 1 - (1 - a_{i,j})/\gamma_{i,j}, & K = 0, \\ (1/\gamma_{i,j})[1 - a_{i,j}]^2 [a_{i,j}]^{K-1}, & K > 0, \end{cases} \tag{5}$$

In Eq. (5), $\gamma_{i,j} = \lambda^c_{i,\Sigma}/\lambda^m_{i,j}$ is the *call-to-mobility ratio* defined in [23]. Furthermore, $a_{i,j}$ is calculated as follows:

$$a_{i,j} = f^{m*}_{i,j}(\lambda^c_{i,\Sigma}), \tag{6}$$

where $f^{m*}_{i,j}(\cdot)$ is the Laplace transform of the residence time density $f^m_{i,j}(t)$ of MT_i in location areas of the subnetwork $\mathbf{S}_j$. The z-transform $\alpha^z_{i,j}(z)$ of the distribution $\alpha_{i,j}(K)$ is calculated as

$$\begin{aligned} \alpha^z_{i,j}(z) &= \sum_{q=0}^{\infty} z^q \alpha_{i,j}(q) \\ &= 1 - \frac{1 - a_{i,j}}{\gamma_{i,j}} + \frac{(1 - a_{i,j})^2}{\gamma_{i,j}} \frac{z}{1 - z a_{i,j}}, \end{aligned} \tag{7}$$

where $a_{i,j}$ is as defined in Eq. (5) and $|z a_{i,j}| < 1$. The derivative of $\alpha^z_{i,j}(z)$ with respect to z is also given by

$$\frac{\mathrm{d}}{\mathrm{d}z} \alpha^z_{i,j}(z) = \frac{(1 - a_{i,j})^2}{\gamma_{i,j}(1 - z a_{i,j})^2}. \tag{8}$$

The total number of registration messages $K_{i,\Sigma}$ generated during an idle period of MT_i is the sum of the number of messages $K_{i,j}$ generated in the subnetworks that track MT_i's location, where $S_j \in \mathbf{TS}_i$. Let $K_{i,\Sigma}$ follow the distribution $\alpha_{i,\Sigma}(K)$. The z-transform $\alpha^z_{i,\Sigma}(z)$ of the distribution $\alpha_{i,\Sigma}(K)$ is calculated as

$$\alpha^z_{i,\Sigma}(z) = \prod_{j=k_1}^{k_P} \alpha^z_{i,j}(z), \tag{9}$$

where $\alpha^z_{i,j}(z)$ is defined in Eq. (7) and the indices $j = k_1, \ldots, k_P$ are taken from the set of subnetworks $\mathbf{TS}_i = \{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$ where MT_i is tracked. The expected value $E[K_{i,\Sigma}]$ calculated by taking the derivative of $\alpha^z_{i,\Sigma}(z)$ with respect to z and evaluating it at $z = 1$:

$$\begin{aligned} E[K_{i,\Sigma}] &= \frac{\mathrm{d}}{\mathrm{d}z} \alpha^z_{i,\Sigma}(z)\bigg|_{z=1} = \frac{\mathrm{d}}{\mathrm{d}z} \prod_{q=k_1}^{k_P} \alpha^z_{i,q}(z)\bigg|_{z=1} \\ &= \sum_{j=k_1}^{k_P} \left[\left(\frac{\mathrm{d}}{\mathrm{d}z} \alpha^z_{i,j}(z) \right) \prod_{q=k_1, q \neq j}^{k_P} \alpha^z_{i,q}(z) \right]_{z=1}. \end{aligned} \tag{10}$$

Evaluating Eqs. (7) and (8) at $z = 1$ we obtain

$$\alpha^z_{i,j}(z = 1) = 1 - \frac{1 - a_{i,j}}{\gamma_{i,j}} + \frac{1 - a_{i,j}}{\gamma_{i,j}} = 1, \tag{11}$$

$$\frac{\mathrm{d}}{\mathrm{d}z} \alpha^z_{i,j}(z)\bigg|_{z=1} = \frac{1}{\gamma_{i,j}}. \tag{12}$$

Substituting Eqs. (11) and (12) into Eq. (10), $E[K_{i,\Sigma}]$ can be calculated

$$E[K_{i,\Sigma}] = \frac{\mathrm{d}}{\mathrm{d}z}\alpha^{z}_{i,\Sigma}(z)\bigg|_{z=1} = \sum_{j=k_1}^{k_P}\frac{1}{\gamma_{i,j}}, \tag{13}$$

where $\gamma_{i,j} = \lambda^{c}_{i,\Sigma}/\lambda^{m}_{i,j}$ and P is the number of subnetworks in $\mathbf{TS}_i$. Note that $E[K_{i,\Sigma}]$ calculated in Eq. (13) gives the expected value of the number of registration messages generated only by a particular mobile terminal MT_i that is tracked in subnetworks $\mathbf{TS}_i$ when it is idle. We aim to achieve the best location management signaling performance by optimizing the tracking sets $\mathbf{TS}$ for individual mobile terminals.

3.3. Paging

Paging under idealized location management scheme relies extensively on the location information gathered by the subnetworks and on the detailed map of LAs system-wide. The basic principle is to calculate the smallest area in which an MT is known to reside in and then to send paging signal in the cells of the subnetwork over which the connection request has been received. Paging has significance only when a connection request is received during the idle periods of an MT. In case the MT has already active connections, its location is known to the system. Even though the known position of the MT may be too general to locate the exact cell location in the subnetwork of interest, it is possible to pinpoint the necessary information by sending a small query packet in one the cells where MT is known to maintain connections.

Let a mobile terminal MT_i maintain active connections in a set of subnetworks $\mathbf{AS}_i = \{\mathbf{S}_{c_1}, \ldots, \mathbf{S}_{c_A}\}$. When a new connection request $R^{c}_{i,n}$ over subnetwork $\mathbf{S}_n$ for MT_i is received, it is handled as follows:

1. If MT_i is not idle, i.e., $\mathbf{AS}_i \neq \emptyset$:
   (a) If $\mathbf{S}_n \in \mathbf{AS}_i$, i.e., new connection request arrives over a subnetwork in which MT_i already maintains an active connection, then GLS knows exactly in which cell MT_i resides in subnetwork $\mathbf{S}_n$, hence, no paging is necessary.
   (b) If $\mathbf{S}_n \notin \mathbf{AS}_i$, i.e., new connection request arrives over a subnetwork which does not carry an active connection for MT_i:
      - GLS instructs SLS_x of subnetwork $\mathbf{S}_x \in \mathbf{AS}_i$ to send a paging request to MT_i in the cell in which it resides.
      - MT_i replies with the cell-level location information in subnetwork $\mathbf{S}_n$.
2. If MT_i is idle, i.e., $\mathbf{AS}_i = \emptyset$:
   - GLS forms the paging set $\mathbf{PS}_i$ as described in Section 2.2.
   - GLS instructs SLS_n to page MT_i in the cells $\mathscr{M}_n(\mathbf{PS}_i)$, where $\mathscr{M}_n(\cdot)$ is the function returning the set of cells in subnetwork $\mathbf{S}_n$ covering a given area.
   - Upon receiving paging signal, MT_i replies with its cell-level location information in subnetwork $\mathbf{S}_n$.

In the paging scheme described above, the cost of paging for Case 1a is 0 since no paging signal is being sent. The probability that a connection request $R^{c}_{i,n}$ is received over subnetwork $\mathbf{S}_n \in \mathbf{AS}_i$ is given by

$$\Pr[\mathbf{S}_n \in \mathbf{AS}_i] = \Pr[N^{n}_{i} > 0] = 1 - \mathrm{e}^{-\rho_{i,n}}, \tag{14}$$

where N^{n}_{i} is the number of active connections of MT_i over subnetwork $\mathbf{S}_n$ and $\rho_{i,n} = \lambda^{c}_{i,n}/\mu^{h}_{i,n}$. Extending this result to a general case, the probability that any call arrival is received over a subnetwork that already has an active connection for MT_i is calculated as follows:

$$\Pr[\text{Paging 0 cells}] = \sum_{q=1}^{\mathscr{N}} \Pr[\mathbf{S}_n = \mathbf{S}_q]\Pr[N^{q}_{i} > 0]$$

$$= \sum_{q=1}^{\mathscr{N}} \frac{\lambda^{c}_{i,q}}{\lambda^{c}_{i,\Sigma}}\left(1 - \mathrm{e}^{-\rho_{i,q}}\right), \tag{15}$$

where $\mathscr{N}$ is the number of subnetworks in NGWN and $\lambda^{c}_{i,\Sigma}$ is given in Eq. (4).

For Case 1b, the paging cost is equal to the cost of paging in a single cell of any subnetwork $\mathbf{S}_x \in \mathbf{AS}_i$. The probability that a new connection request arrives over a subnetwork $\mathbf{S}_n \notin \mathbf{AS}_i$ and $\mathbf{AS}_i \neq \emptyset$ can be calculated as

$$\Pr[\mathbf{S}_n \notin \mathbf{AS}_i, \mathbf{AS}_i \neq \emptyset] = \mathrm{e}^{-\rho_{i,n}}\left(1 - \prod_{r=1,r\neq n}^{\mathcal{N}} \mathrm{e}^{-\rho_{i,r}}\right), \tag{16}$$

where $\rho_{i,r} = \lambda^c_{i,r}/\mu^h_{i,r}$. Once again, this probability can be generalized to obtain the probability that a single cell paging is required to learn about the cell location of MT$_i$ in subnetwork $\mathbf{S}_n$

$$\Pr[\text{Paging 1 cell}] = \sum_{q=1}^{\mathcal{N}} \Pr[\mathbf{S}_n = \mathbf{S}_q] \Pr[\mathbf{S}_q \notin \mathbf{AS}_i, \mathbf{AS}_i \neq \emptyset] = \sum_{q=1}^{\mathcal{N}} \left[\frac{\lambda^c_{i,q}}{\lambda^c_{i,\Sigma}} \mathrm{e}^{-\rho_{i,q}}\left(1 - \prod_{r=1,r\neq q}^{\mathcal{N}} \mathrm{e}^{-\rho_{i,r}}\right)\right]. \tag{17}$$

Finally, Case 2 involves multiple-cell paging in subnetwork $\mathbf{S}_n$. The probability of such an occurrence is equal to the probability of MT$_i$ being idle, which is given in Eq. (3). However, in this case, the calculation of the number of cells to be paged is not trivial. In the following, we will present the equations necessary to calculate the number of cells to be paged.

3.3.1. Average size of the paging set

An idle mobile terminal MT$_i$ can be located using the location area information $\{LA^{k_1}_{s_1}, \ldots, LA^{k_P}_{s_P}\}$ in the subnetworks $\mathbf{TS}_i = \{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$ where it is tracked. The intersection area of known LAs is calculated with the exact mapping of LAs to coverage areas maintained in TDB. Let $\mathrm{A}^R_i = \bigcap_{q=1}^P LA^{k_q}_{s_q}$ be the reduced location area for MT$_i$. If A^R_i is modeled as a circular region, its radius $r_{1\ldots P}$ is given by the following equation:

$$r_{1\cdots P} = \sqrt{\frac{\mathrm{A}^R_i}{\pi}}. \tag{18}$$

If the connection request is received over the subnetwork $\mathbf{S}_n$, MT$_i$ is paged in the cells of $\mathbf{S}_n$. Considering that the cells of $\mathbf{S}_n$ can fully or partially fall into A^R_i, the number of cells $N_C(r_{1\ldots P}|\mathbf{S}_n)$ belonging to $\mathbf{S}_n$ is approximated for an intersection area of radius $r_{1\ldots P}$ as follows:

$$N_C(r_{1\cdots P}\,|\,\mathbf{S}_n) = \frac{(r_{1\cdots P} + r^c_n)^2}{(r^c_n)^2}, \tag{19}$$

where r^c_n is the radius of the cells of the subnetwork $\mathbf{S}_n$.

3.3.2. Radius of the reduced location area

Eq. (19) is useful only if we can calculate the radius $r_{1\ldots P}$ of the reduced location area. For this purpose, we will assume that the location areas are circular regions. Furthermore, it is assumed that the subnetworks in $\mathbf{TS}_i$ are ordered according to increasing size of the location area radius, i.e., given $\mathbf{TS}_i\{\mathbf{S}_{k_1}, \ldots, \mathbf{S}_{k_P}\}$, $k_x < k_y \iff r_x < r_y$, where r_x and r_y are radii of LAs of subnetworks $\mathbf{S}_{k_x}$ and $\mathbf{S}_{k_y}$, respectively. Let $g(r, r_1, r_2)$ be defined as the density of the radius r of the intersection area given that circles with radii r_1 and r_2 intersect. We also denote the radius of the intersection area of b location areas as $r_{1\ldots b}$, where $1 < b \leqslant P$. We define $f_{i,r_{1\ldots P}}(r)$ as the probability density of radii of the intersection areas of LAs belonging to subnetworks in $\mathbf{TS}_i$, where P is the number of the subnetworks in $\mathbf{TS}_i$. The probability density function $f_{i,r_{1\ldots P}}(r)$ can be calculated as follows:

$$f_{i,r_{12}}(r) = g(r, r_{k_1}, r_{k_2}), \tag{20}$$

$$f_{i,r_{123}}(r) = \int_{r_{12}=0}^{r_{k_1}} g(r, r_{12}, r_{k_3})\, g(r_{12}, r_{k_1}, r_{k_2})\, \mathrm{d}r_{12}, \tag{21}$$

$$f_{i,r_{1234}}(r) = \int_{r_{123}=0}^{r_{12}} \int_{r_{12}=0}^{r_{k_1}} g(r, r_{123}, r_{k_4})\, g(r_{123}, r_{12}, r_{k_3}) \times g(r_{12}, r_{k_1}, r_{k_2})\, \mathrm{d}r_{12}\, \mathrm{d}r_{123}. \tag{22}$$

Eqs. (20)–(22) can be extended to include arbitrary number of location areas, i.e., the radius $r(\mathrm{A}^R_i) = r_{1\cdots P}$ of the reduced location area A^R_i can be calculated for any $\mathbf{TS}_i$. Note that Eqs. (20)–(22) are based on the assumption that the LAs as well as the intersection areas of arbitrary number of LAs are circular. The derivation of the pdf $g(r, r_1, r_2)$ of the radius r of the intersection area of two circles of radii r_1 and r_2 is presented in Appendix A.

3.4. Signaling cost generated by a mobile terminal

In previous two sections, we have presented how the expected value of the location registrations between two busy periods and the estimated number of cells to be paged are calculated. The best selection of subnetworks $\mathbf{TS}_i$ in which a mobile terminal MT_i is tracked is determined considering the average cost of location management signaling per unit time for all possible selections of subnetworks to be included in $\mathbf{TS}_i$. Below, we present the location management signaling cost $\mathrm{LMC}_i(\mathbf{TS}_i)$ calculation caused by MT_i given a tracking set $\mathbf{TS}_i$.

A mobile terminal MT_i attempts to register its location only when it crosses an LA boundary of a subnetwork $\mathbf{S}_j \in \mathbf{TS}$ while MT_i is idle. The probability that MT_i is idle is given in Eq. (3). The expected value of location registrations $E[K_{i,\Sigma}]$ during an idle period is calculated in Eq. (13). In order to find the number of registration attempts per unit time, we approach the problem as follows: $E[K_{i,\Sigma}]$ corresponds to the expected number of location registrations given that MT_i was idle upon connection request arrival, which marks the end of the period in which the LA crossings are counted. Let us assume that an observer does not actually count LA boundary crossings, but tries to guess how many LA crossings have occurred since the previous connection request arrival. To create a guess, the observer checks the condition of MT_i every time a new connection request arrives. When receiving a new connection request, if MT_i does not have any active connections, i.e., if MT_i is idle, then this means that MT_i could have potentially been sending registration messages. In this case, the guess of the observer will be $E[K_{i,\Sigma}]$. If MT_i has active connections when the new connection request arrives, this indicates that MT_i was not in idle state, and therefore, was not sending any new location registration messages since the previous connection request arrival. Considering that total connection arrival rate is $\lambda^c_{i,\Sigma}$, the rate $\mathrm{LRR}_i(\mathbf{TS}_i)$ at which MT_i sends location registration messages can be computed as follows:

$$\begin{aligned}\mathrm{LRR}_i(\mathbf{TS}_i) &= \lambda^c_{i,\Sigma} \cdot \Pr[\mathrm{MT}_i \text{ is idle}] \cdot E[K_{i,\Sigma}] \\ &= \lambda^c_{i,\Sigma} \cdot \mathrm{e}^{-\rho_{i,\Sigma}} \cdot \sum_{q=k_1}^{k_P} \frac{1}{\gamma_{i,q}} \\ &= \lambda^c_{i,\Sigma} \cdot \mathrm{e}^{-\rho_{i,\Sigma}} \cdot \sum_{q=k_1}^{k_P} \frac{\lambda^m_{i,q}}{\lambda^c_{i,\Sigma}} \\ &= \mathrm{e}^{-\rho_{i,\Sigma}} \cdot \sum_{q=k_1}^{k_P} \lambda^m_{i,q}, \qquad (23)\end{aligned}$$

where P is the number of subnetworks in $\mathbf{TS}_i$ and $\rho_{i,\Sigma}$ is given in Eq. (24):

$$\rho_{i,\Sigma} = \sum_{j=1}^{\mathcal{N}} \rho_{i,j}, \qquad (24)$$

where $\mathcal{N}$ is the total number of subnetworks in NGWN and $\rho_{i,j} = \lambda^c_{i,j}/\mu^h_{i,j}$.

When paging a mobile terminal, a paging packet is broadcast in all cells in the paging set. The paging packets sent per unit time $\mathrm{PR}_i(\mathbf{TS}_i)$ for MT_i when tracked by subnetworks in $\mathbf{TS}_i$ consists of two components. The first component is the single cell paging packet caused by Case 1b of Section 3.3. The probability of Case 1b is given in Eq. (17). The second component is the multiple cell paging cost caused by Case 2 of Section 3.3. Eq. (19) gives an approximation for the number of cells to be paged given that MT_i is idle, tracking set is $\mathbf{TS}_i$, and the connection request is received over the subnetwork $\mathbf{S}_n$. Let the expected value of the radius $r_{1\ldots P}$ of the intersection of the LAs of subnetworks in $\mathbf{TS}_i$ be $E[r_{1\ldots P}]$, where $r_{1\ldots P}$ is distributed with $f_{i,r_{1\ldots P}}(r)$. Considering that new connection requests arrive at a combined rate of $\lambda^c_{i,\Sigma}$, paging packets sent per unit time PR_i for MT_i when tracked by subnetworks in $\mathbf{TS}_i$ is calculated as follows:

$$\begin{aligned}&\mathrm{PR}_i(\mathbf{TS}_i) \\ &\quad = \lambda^c_{i,\Sigma} \cdot 1 \cdot \Pr[\text{Paging 1 cell}] \\ &\qquad + \lambda^c_{i,\Sigma} \cdot \Pr[\mathrm{MT}_i \text{ is idle}] \sum_{q=1}^{\mathcal{N}} N_C(E[r_{1\cdots P}]|\mathbf{S}_n) \\ &\qquad \times \Pr[\mathbf{S}_n = \mathbf{S}_q] \\ &\quad = \lambda^c_{i,\Sigma} \cdot \sum_{q=1}^{\mathcal{N}} \left[\frac{\lambda^c_{i,q}}{\lambda^c_{i,\Sigma}} \mathrm{e}^{-\rho_{i,q}} \left(1 - \prod_{r=1,r\neq q}^{\mathcal{N}} \mathrm{e}^{-\rho_{i,r}} \right) \right] \\ &\qquad + \lambda^c_{i,\Sigma} \cdot \mathrm{e}^{-\rho_{i,\Sigma}} \sum_{q=1}^{\mathcal{N}} \frac{(E[r_{1\cdots P}] + r^c_q)^2}{(r^c_q)^2} \frac{\lambda^c_{i,q}}{\lambda^c_{i,\Sigma}} \\ &\quad = \sum_{q=1}^{\mathcal{N}} \left[\lambda^c_{i,q} \mathrm{e}^{-\rho_{i,q}} \left(1 - \prod_{r=1,r\neq q}^{\mathcal{N}} \mathrm{e}^{-\rho_{i,r}} \right) \right] \\ &\qquad + \mathrm{e}^{-\rho_{i,\Sigma}} \sum_{q=1}^{\mathcal{N}} \frac{\lambda^c_{i,q}(E[r_{1\cdots P}] + r^c_q)^2}{(r^c_q)^2}, \qquad (25)\end{aligned}$$

where $\mathcal{N}$ is the number of subnetworks in NGWN, and $\rho_{i,\Sigma}$ is defined in Eq. (24).

The overall location management signaling cost $\mathrm{LMC}_i(\mathbf{TS}_i)$ per unit time generated by MT_i when tracked in subnetworks of $\mathbf{TS}_i$ can be defined in many ways depending on the definition of the "cost" of a location registration message sent by an MT_i and the "cost" of broadcasting a paging packet in a cell. These costs usually depend on the procedures used to process the location registration messages and to generate and broadcast the paging messages. The term "cost" can also be define to reflect the bandwidth usage as well as processing power required. To serve a wide range of cost definitions, we normalize the cost of a location registration message to 1 and introduce a scaling factor β to adjust the relative cost of paging an MT in a single cell, assuming the total contribution of the paging cost is directly proportional to the number of cells a paging message is broadcast in. Consequently, we define the location management cost $\mathrm{LMC}_i(\mathbf{TS}_i)$ per unit time as follows:

$$\mathrm{LMC}_i(\mathbf{TS}_i) = \mathrm{LRR}_i(\mathbf{TS}_i) + \beta \cdot \mathrm{PR}_i(\mathbf{TS}_i), \qquad (26)$$

$\mathrm{LRR}_i(\mathbf{TS}_i)$ and $\mathrm{PR}_i(\mathbf{TS}_i)$ are given in Eqs. (23) and (25), respectively, and β is the scaling factor for the paging cost.

Note that the total registration cost given in Eq. (26) is calculated for a particular MT_i and for a particular choice of the tracking set $\mathbf{TS}_i$. To determine the optimal $\mathbf{TS}_i$ selection, all subsets of the subnetworks in NGWN must be considered and the subset with the lowest registration cost must be determined. For the lowest signaling cost in NGWN, this operation must be repeated for all MTs and must be refreshed when the connection arrival and mobility characteristics of MTs change. Hence, given the mobility parameters $\lambda^m_{i,1}, \ldots, \lambda^m_{i,\mathcal{N}}$, the connection parameters $\lambda^c_{i,1}, \ldots, \lambda^c_{i,\mathcal{N}}$ and $\mu^h_{i,1}, \ldots, \mu^h_{i,\mathcal{N}}$ for mobile terminal MT_i, and the network parameters for $\mathbf{S}_1, \ldots, \mathbf{S}_{\mathcal{N}}$, the minimum location management cost $\mathrm{LMC}^{\min}_i$ is calculated as follows:

$$\mathrm{LMC}^{\min}_i = \min_{\mathbf{TS}_i \subset \{\mathbf{S}_1,\ldots,\mathbf{S}_{\mathcal{N}}\}} \{\mathrm{LMC}_i(\mathbf{TS}_i)\}, \qquad (27)$$

where $\mathrm{LMC}_i(\mathbf{TS}_i)$ is given in Eq. (26) and $\mathcal{N}$ is the total number of subnetworks of NGWN.

4. Numerical examples

In this Section, we present several numerical examples based on the performance measures derived in Section 3. Overall performance of the location management schemes in NGWN depends on several factors including the number of subnetworks, subnetwork-related parameters such as size and distribution of the cells and location areas, mobility characteristics, and connection request arrival rates for the users. Instead of trying to cover the entire parameter space, we will consider two example cases of NGWN architectures consisting of three subnetworks each and present the effect of changing connection request arrival and mobility rates. The NGWN architecture of the first case correspond to an urban environment covered by three subnetworks with similar cell but different location area sizes. The mobile terminal we consider in this example receives connection requests evenly distributed among the three subnetworks. The second case considers an architecture composed of three subnetworks utilizing pico-, micro-, and macrocells, respectively. The mobile terminal in this scenario utilizes primarily the microcellular system. In both cases, we assume uninterrupted service area for all subnetworks and assume that the scaling factor $\beta = 0.1$, where β was introduced in Eq. 26. The effects of β is also presented based on the second scenario. The base parameters for these experiments are given in Table 1.

Note that the choice of parameters for the numerical examples presented in this section reflects two particular scenarios. The performance measures will clearly be different for different choices of both NGWN composition, e.g., number of subnetworks, and individual subnetwork and user parameters.

4.1. Case I: three similar subnetworks

The first set of numerical examples focus on the performance bounds in a system composed of three similar subnetworks. The subnetworks utilize approximately same size cells and have different LA sizes. Here, we aim to mimic the behaviour of a user that commutes to work, remains stationary for longer periods of time before moving to

Table 1
Base parameters for the numerical examples

Case	Cell radii			LA radii			Mobility rate			Conn. arrival			Conn. holding			β
	r_1^c	r_2^c	r_3^c	r_1	r_2	r_3	λ_1^m	λ_2^m	λ_3^m	λ_1^c	λ_2^c	λ_3^c	μ_1^h	μ_2^h	μ_3^h	
I	1	1	1	10	15	20	$\frac{1}{20,000}$	$\frac{1}{20,000}$	$\frac{1}{20,000}$	$\frac{1}{5200}$	$\frac{1}{3600}$	$\frac{1}{5200}$	$\frac{1}{600}$	$\frac{1}{600}$	$\frac{1}{600}$	0.1
II	1	5	15	10	50	150	$\frac{1}{1200}$	$\frac{1}{6000}$	$\frac{1}{30,000}$	$\frac{1}{72,000}$	$\frac{1}{1800}$	$\frac{1}{72,000}$	$\frac{1}{60}$	$\frac{1}{300}$	$\frac{1}{500}$	0.1

another location area. The user we consider for this scenario is assumed to be located at the vicinity of the LA boundaries of all subnetworks in an urban area, spending approximately 5.5 h on the average before switching to another location area in the same subnetwork. The user also receives connection requests from all subnetworks at comparable rates as shown in Table 1. The average connection inter-arrival time in the first and third subnetworks is approximately 1 h 25 min, and in the second subnetwork 1 h. In all subnetworks, the average connection duration is 5 min.

Location management cost, paging and, location registration rates have been studied for varying levels of connection arrival rates in subnetwork $\mathbf{S}_2$. The connection arrival rate is uniformly increased from $\lambda_2^c = \frac{1}{36000}$ to $\lambda_2^c = \frac{1}{620}$. In Fig. 5, the overall location management cost LMC is depicted for all choices of tracking sets **TS**. For all values of λ_2^c, tracking the user in all subnetworks yields the best overall performance. This behavior is better understood when the paging packet rate PR presented in Fig. 6 and the location registration rate LRR presented in Fig. 7 are considered. Since the location registration events occur relatively infrequently when compared with connection request arrivals, the paging rate becomes the dominating factor in the calculation of the overall location management cost. Tracking the mobile terminal in a higher number of subnetworks decreases the size of the paging set, and consequently, the number of cells paged upon connection request arrival. The numerical

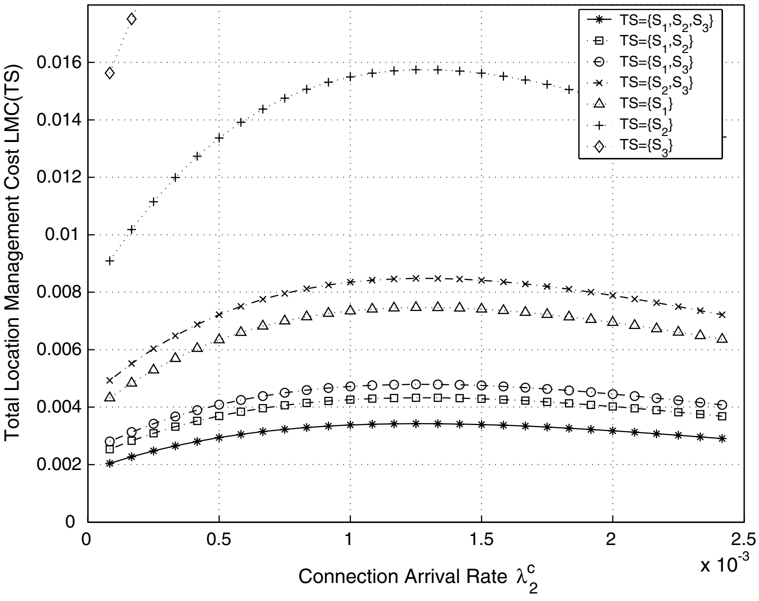

Fig. 5. Total location management cost for Case I.

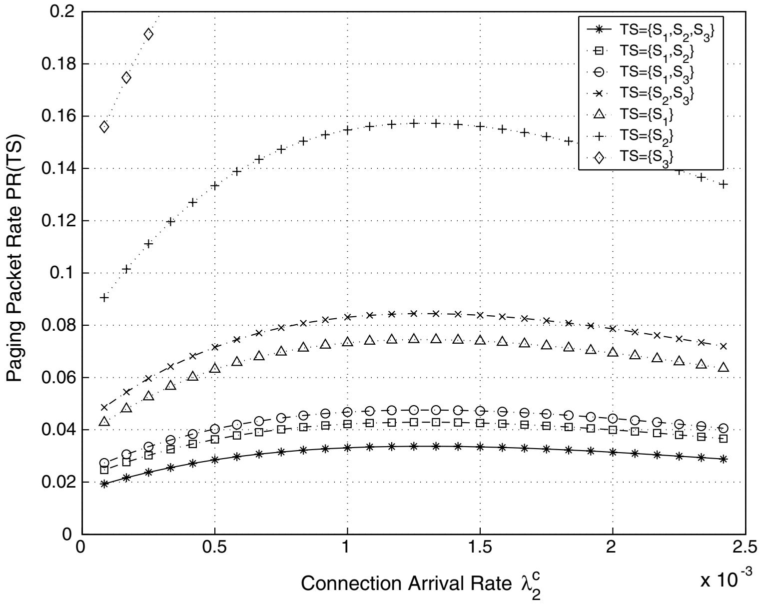

Fig. 6. Paging packet rate for Case I.

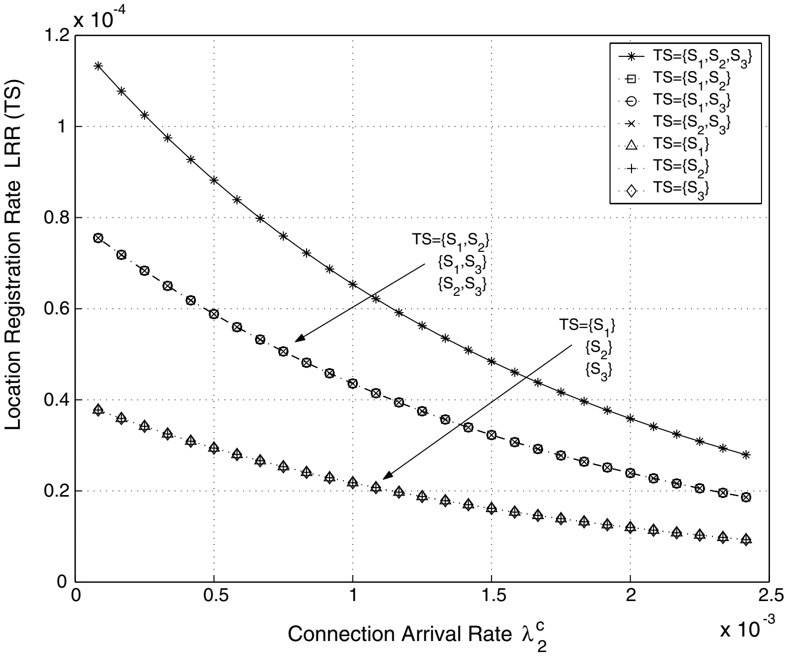

Fig. 7. Location registration rate for Case I.

evaluation of the expected value of intersection area radii of two and three LAs are presented in Appendix A. As shown in Fig. 6, the subnetworks with smaller LAs contribute more to the reduction of paging sets. This becomes more apparent when the paging packet rates PR for different tracking sets **TS** are compared. For example, keeping $\mathbf{S}_1$ fixed in **TS**, we can see that a subnetwork with smaller LA size (in this case $\mathbf{S}_2$) reduces the location management cost more than a subnetwork with larger LA size (in this case $\mathbf{S}_3$) because $(\mathrm{PR}(\{\mathbf{S}_1\})-\mathrm{PR}(\{\mathbf{S}_1,\mathbf{S}_2\})) > (\mathrm{PR}(\{\mathbf{S}_1\})-\mathrm{PR}(\{\mathbf{S}_1,\mathbf{S}_3\}))$. Including all subnetworks in the tracking set, the minimum packet paging rate and total location management cost is achieved.

Another important point to note in Fig. 6 is the shape of all curves. The number of paging packets per unit time starts increasing as the offered load increases. However, at the same time, the ratio of non-idle times also increases, which reduces the need for paging upon connection request arrival. The increase in the number of paging packets per unit time stops approximately between $\lambda_2^c = 1\times 10^{-3}$ and $\lambda_2^c = 1.25\times 10^{-3}$ for different paging sets, and the rate declines for higher λ_2^c values. Similar results are obtained when the mobility of the user is changed.

The location registration signaling heavily depends on the mobility of the user. However, the rate of connection request arrivals also affects the number of registrations per unit time. As shown in Fig. 7, the average number of location registrations per unit time are grouped according to the size of the tracking set **TS** since all subnetworks share the same average LA residence times. As more subnetworks are included in the tracking set, the number of location registrations per unit time increases. This average is a decreasing function of λ_2^c because increasing λ_2^c decreases the probability of the mobile being idle, and also the need for location registration when crossing LA boundaries.

4.2. Case II: pico-, micro-, and macrocellular subnetworks

The second set of numerical examples refers to an environment that includes pico-, micro-, and macrocellular subnetworks. The cell sizes of these subnetworks are chosen as 1, 5, and 15 units, respectively. The LA radii of three subnetworks are 10 times their respective cell radii. The mobile terminal we consider in this scenario remains in a location area on the average 20 min, 1 h 40 min, and 6 h 20 min in $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$, respectively. The connection request arrival rate is negligible in $\mathbf{S}_1$ and $\mathbf{S}_3$, and twice every hour in $\mathbf{S}_2$. A connection lasts 2.5 min on the average. The movement behavior considered in this example is similar to the movement patterns of a person who is always on the move, such as a delivery person.

We have considered the effects of both connection arrival rate as well as the mobility on the performance metrics of the system. In all experiments, the connection arrival rate in $\mathbf{S}_2$ is varied between $\lambda_2^c = 0.11\times 10^{-3}$ and $\lambda_2^c = 3.2\times 10^{-3}$. Furthermore, we have analyzed the effect of mobility wherein the LA residence rate λ_3^m for $\mathbf{S}_3$ was kept constant and the base values λ_1^m and λ_2^m for $\mathbf{S}_1$ and $\mathbf{S}_2$ given in Table 1 was multiplied with the mobility factor k_m, $0.5 \leqslant k_m \leqslant 2$. Keeping λ_3^m constant, we create a scenario where the movement patterns within a locality changes and the long term movement patterns remain the same. In Fig. 8, the minimum location management costs for the above mentioned parameter ranges are shown. As opposed to the case presented in Section 4.1, the minimum location management cost is not achieved by a particular tracking set alone. As the mobility and connection arrival rates change, different tracking sets provide better performance. The points on the surface correspond to the minimum location management cost introduced in Eq. (27), and are marked with circles, squares, and triangles according to the tracking set that generates those values. In this example, when λ_2^c is low, tracking the mobile terminal in $\mathbf{S}_1$ gives the best performance. As the connection arrival rate increases, $\mathbf{TS}=\{\mathbf{S}_1,\mathbf{S}_2\}$ and $\mathbf{TS}=\{\mathbf{S}_1,\mathbf{S}_3\}$ provide better performance depending on the mobility and connection arrival rates. This shows that utilizing a fixed tracking set may not always provide the best performance.

Fixing the mobility factor $k_m=1$, the total location management cost LMC is plotted for four tracking sets with lowest total location management costs [2] in Fig. 9. The total location

---

[2] Remaining subsets have very high total cost values and increase the *y*-range of the plot, negatively affecting clarity of lower values.

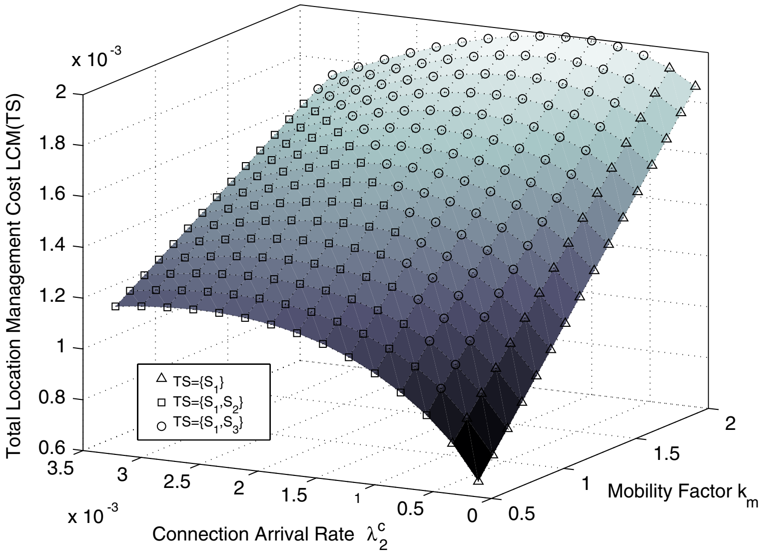

Fig. 8. Total location management cost for Case II.

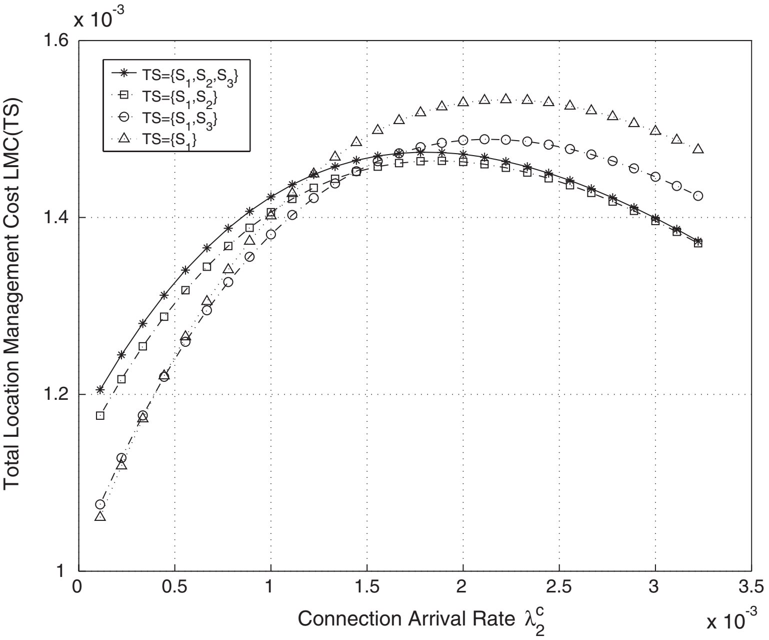

Fig. 9. Total location management cost for Case II, $k_m = 1$.

management cost difference between $\mathbf{TS} = \{\mathbf{S}_1\}$ and $\mathbf{TS} = \{\mathbf{S}_1, \mathbf{S}_3\}$ is very small for lower values of λ_2^c. For $\lambda_2^c > 0.5 \times 10^{-3}$, $\mathbf{TS} = \{\mathbf{S}_1, \mathbf{S}_3\}$ starts resulting in lower total cost, which last until $\lambda_2^c = 1.45 \times 10^{-3}$. For larger λ_2^c, lowest overall cost is achieved by $\mathbf{TS} = \{S_1, S_2\}$. It is also worth noting that tracking the mobile in all subnetworks becomes increasingly close to the lowest total cost,

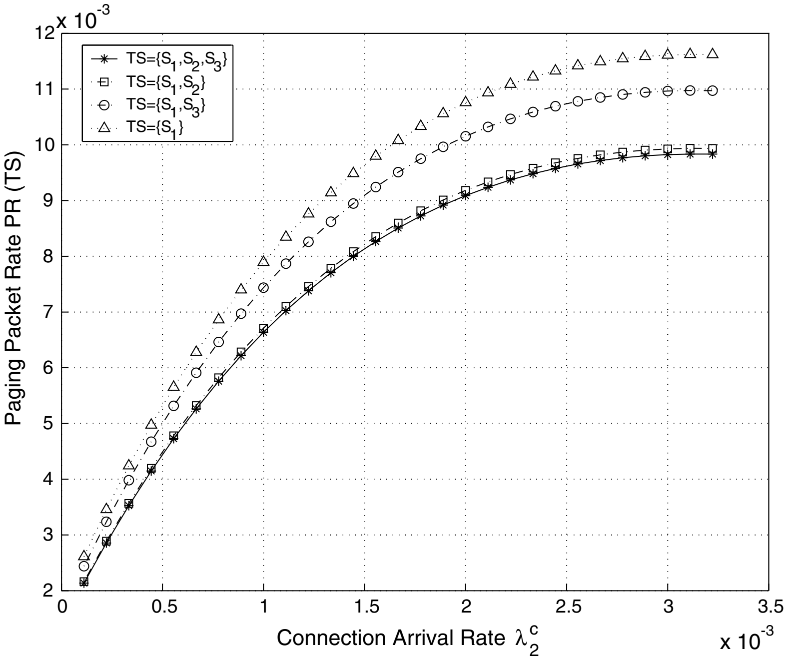

Fig. 10. Paging packet rate for Case II, $k_m = 1$.

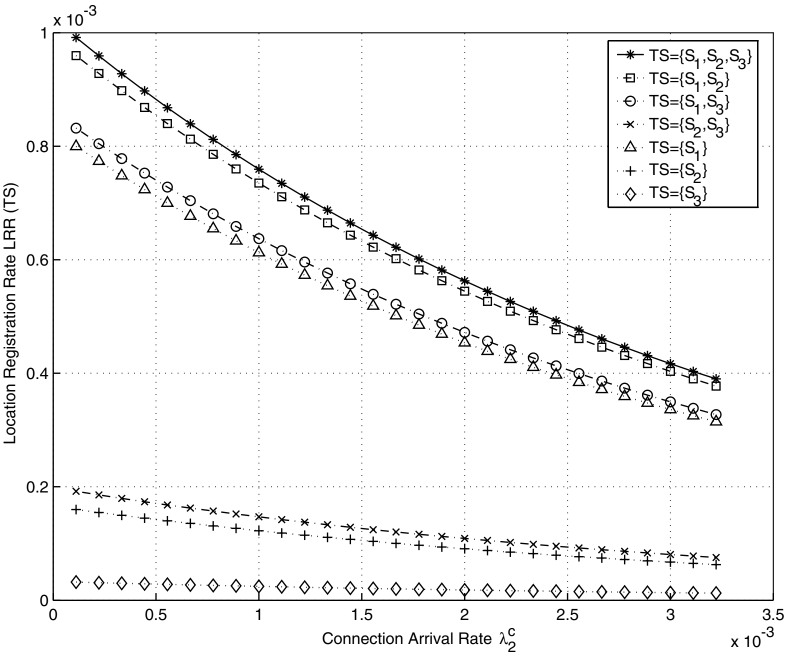

Fig. 11. Location registration rate for Case II, $k_m = 1$.

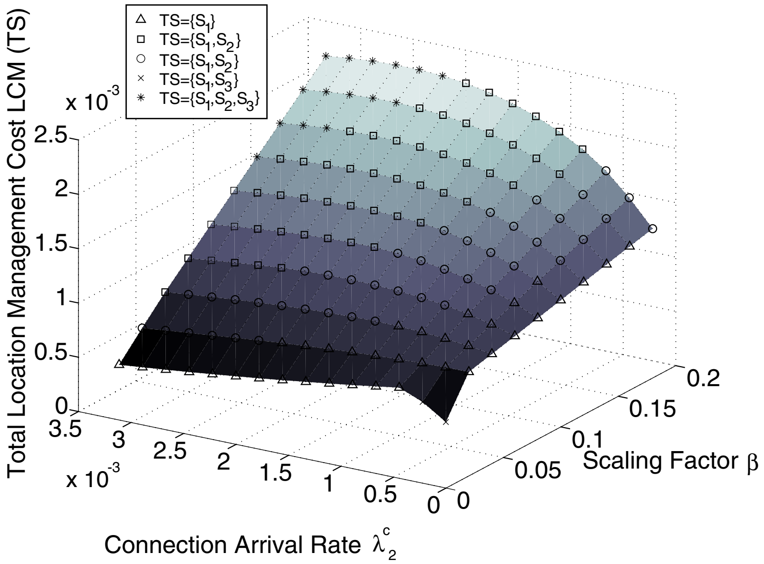

Fig. 12. Effect of β on total location management cost for Case II, $k_m = 1$.

although it never provides the lowest cost in the given λ_2^c or k_m ranges.

The number of paging packets generated for and the number of registration messages generated by the mobile terminal in this scenario is shown in Figs. 10 and 11. Unlike in the case presented in Section 4.1, the paging packet rate does not dominate the overall location management cost. In Fig. 10, we observe an increasing packet paging rate (with negative second derivative) shown for the same tracking sets as in Fig. 9. The decreasing idle period lengths reduce growth the paging packet rate for the higher values of the λ_2^c range. Similarly, the same effect is observed in Fig. 11 for the location registration rate. As more connections arrive, the need for location registration upon crossing LA boundaries decreases monotonically. It is also important to note that tracking sets $\mathbf{TS} = \{S_1, S_2\}$ and $\mathbf{TS} = \{S_1, S_2, S_3\}$ have very similar values in terms of LRR, PR, and LMC.

4.3. Effect of the scaling factor β

One of the most important parameters of the total location management cost calculation is the selection of the scaling factor β. As used in Eq. (26), taking the cost of a location registration as the unit cost, β determines the relative cost of broadcasting a paging packet in a cell. Since the cost of paging as well as location registration can vary from implementation to implementation, we have chosen to employ a relative scaling factor in the total location management cost calculation.

In Fig. 12, the total location management cost of the case presented in Section 4.2 is plotted for $k_m = 1$ while changing λ_2^c and β, $0.01 \leqslant \beta \leqslant 0.2$. The range of β values is between 0.01 and 0.2. The surface in this figure contains the points marking the minimum location management cost $\mathrm{LMC}^{\min}$ for the given set of parameters. The data points in this surface are marked with circles, squares, triangles, and crosses according to the tracking set that generates those values. It can be observed that LMC is highly sensitive to changes in β. To achieve correct location management signaling cost bounds, determining the relative cost of registration and paging plays an important role. As an example, choosing β as 0.13 instead of 0.1 in experiments of Section 4.2 would have included the tracking set $\mathbf{TS} = \{S_1, S_2, S_3\}$ into the performance graphics. Therefore, the selection of β must be done in such a way that the actual anticipated relative overheads are reflected in the calculations.

5. Conclusion

In this paper, we studied the performance bounds of location management schemes in the wireless portion of NGWN. The NGWN architecture we consider is composed of multiple subnetworks serving overlapping coverage areas. Assuming complete and perfect knowledge about the network parameters, user mobility information, and connection patterns, we have presented an idealized location management scheme and derived equations to calculate the location registration, paging, and overall location management signaling costs over the wireless medium. The signaling performance bound equations presented in this paper are intended to be used as benchmarks for the performance of other location management systems developed for NGWN. We also presented the performance results of two sample scenarios which correspond to two architectures composed of subnetworks with similar and different cell sizes, respectively. Using these network configurations, effects of user related connection arrival and mobility parameters on the performance metrics as well as the effect of paging cost weight are studied. These studies show that tracking a user in a fixed predetermined set of subneworks does not necessarily provide the best location management signaling cost in the wireless portion. Another important observation is that the relative weight of location registration and paging signaling has a big influence on the overall location management cost, and hence, must be selected carefully.

Appendix A. Derivation of $g(r, r_1, r_2)$

Let us consider two circles C_1 and C_2 with radii r_1 and r_2, where $r_1 \leqslant r_2$, as shown in Fig. A.1. The intersection area $A(d)$ is a function of the distance d between the circle centers M_1 and M_2. The intersection area is calculated as follows:

$$A(d) = \begin{cases} \pi r_1^2, \quad 0 \leqslant d < r_2 - r_1, \\ \dfrac{r_1^2}{2}(2\alpha_1 - \sin(2\alpha_1)) + \dfrac{r_2^2}{2}(2\alpha_2 - \sin(2\alpha_2)), \\ \qquad r_2 - r_1 \leqslant d \leqslant r_2 + r_1, \end{cases} \tag{A.1}$$

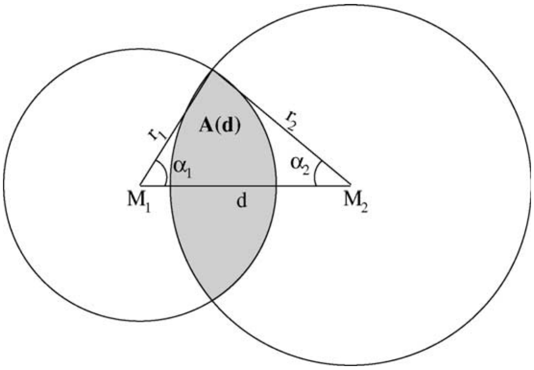

Fig. A.1. Intersection of two circles.

where

$$\alpha_1 = \cos^{-1}\left(\frac{d^2 + r_1^2 - r_2^2}{2dr_1}\right), \tag{A.2}$$

$$\alpha_2 = \cos^{-1}\left(\frac{d^2 + r_2^2 - r_1^2}{2dr_2}\right). \tag{A.3}$$

In a general case, it is complicated to use $A(d)$ as it is. Instead, we approximate the region $r_2 - r_1 \leqslant d \leqslant r_2 + r_1$ with a linear function. This approach yields the approximation function $A'(d)$

$$A'(d) = \begin{cases} \pi r_1^2, & 0 \leqslant d < r_2 - r_1, \\ \pi r_1\left(\dfrac{r_1 + r_2 - d}{2}\right), & r_2 - r_1 \leqslant d \leqslant r_2 + r_1. \end{cases} \tag{A.4}$$

The approximation function simplifies the computation of the intersection area with negligible amount of error. Fig. A.2 shows three examples of $A(d)$ and $A'(d)$ for $r_2 = 2r_1$, $r_2 = 5r_1$, and $r_2 = 10r_1$. As shown in this figure, the error introduced by using $A'(d)$ is small even for circles of similar sizes.

The next step is the calculation of the pdf of the intersection area A_{12}. Using Eq. (A.4), the pdf $f_{A_{12}}(a)$ of A_{12} can be computed as follows:

$$f_{A_{12}}(a) = \begin{cases} \dfrac{2}{\pi r_1(r_1 + r_2)}, & 0 \leqslant a < \pi r_1^2, \\ \dfrac{r_2 - r_1}{r_2 + r_1}, & a = \pi r_1^2. \end{cases} \tag{A.5}$$

We will approximate the intersection area with a circle of the same area. This approach reduces the complexity of calculation of the intersection

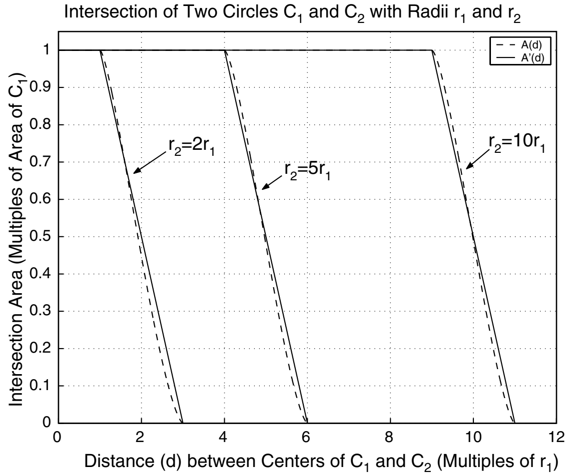

Fig. A.2. Actual and approximated intersection areas of two circles.

area of multiple circular regions. Considering that an intersection area a of two circles with radii r_1 and r_2 corresponds to a circle with radius $r = \sqrt{a/\pi}$ and using Eq. (A.5), we obtain the pdf $g(r, r_1, r_2)$

$$g(r, r_1, r_2) = \begin{cases} \dfrac{4r}{r_1(r_1 + r_2)}, & 0 \leqslant r < r_1, \\ \dfrac{r_2 - r_1}{r_2 + r_1}, & r = r_1. \end{cases} \tag{A.6}$$

The expected value $E[r]$ of the radius r of the intersection area is calculated as

$$E[r] = \frac{r_1(r_1 + 3r_2)}{3(r_1 + r_2)}. \tag{A.7}$$

If three circles with radii r_1, r_2, and r_3 are intersected, the pdf $f_{r_{123}}(r)$ of the radius of the intersection area defined in Eq. (21) can be expressed as follows:

$$f_{r_{123}}(r) = \begin{cases} \dfrac{4r}{r_1(r_1 + r_2)}\left(4\ln\dfrac{r_1 + r_3}{r + r_3} + \dfrac{r_3 - r}{r_3 + r}\right), & 0 \leqslant r < r_1, \\ \dfrac{r_2 - r_1}{r_1 + r_2}, & r = r_1. \end{cases} \tag{A.8}$$

The expected value $E[r_{123}]$ of the intersection area radius r_{123} is given by

$$E[r_{123}] = \frac{r_1(r_2 - r_1)}{r_2 + r_1} + \frac{4\left(r_1^2(r_1 + 3) + r_3(1 - 6r_1r_3) + 6r_3^3\ln\dfrac{r_1 + r_3}{r_3}\right)}{9r_1(r_1 + r_2)}. \tag{A.9}$$

References

[1] EIA/TIA, Cellular radio-telecommunications intersystem operations, EIA/TIA Technical Report IS-41 Revision B.

[2] ETSI, Digital cellular telecommunications system (phase 2+): mobile application part (MAP) specification, GSM 09.02 version 7.51 Release.

[3] A. Bar-Noy, I. Kessler, M. Sidi, Mobile users: to update or not to update? Wireless Networks 1 (2) (1994) 175–186.

[4] I. Akyildiz, J. Ho, Y.-B. Lin, Movement-based location update and selective paging for PCS networks, IEEE/ACM Transactions on Networking 4 (4) (1996) 629–638.

[5] J. Ho J. Xu, History-based location tracking for personal communications networks, in: Proceedings of IEEE Vehicular Technologies Conference, vol. 1, 1998, pp. 244–248.

[6] Y.-B. Lin, Reducing location update cost in a PCS network, IEEE/ACM Transactions on Networking 5 (1) (1997) 25–33.

[7] P. Escalle, V. Giner, J. Oltra, Reducing location update and paging cost in a PCS network, IEEE Transactions on Wireless Communications 1 (1) (2002) 200–209.

[8] B. Liang, Z. Haas, Predictive distance-based mobility management for PCS networks, in: Proceedings of IEEE INFOCOM '99, vol. 3, 1999, pp. 1377–1384.

[9] V. Wong, V. Leung, An adaptive distance-based location update algorithm for next generation PCS networks, IEEE Journal on Selected Areas in Communications 19 (10) (2001) 1942–1952.

[10] I. Akyildiz, J. McNair, J. Ho, H. Uzunalioglu, W. Wang, Mobility management in next generation wireless systems, Proceedings of the IEEE 87 (8) (1999) 1347–1384.

[11] G. Wu, M. Mizuno, P. Havinga, MIRAI architecture for heterogeneous network, IEEE Communications Magazine 40 (2) (2002) 126–134.

[12] R. Berezdivin, R. Breinig, R. Topp, Next generation wireless communications concepts and technologies, IEEE Communications Magazine 40 (3) (2002) 108–116.

[13] P. Conforto, C. Tocci, G. Losquadro, A. Spazio, R. Sheriff, M. Chan, Y. Hu, Ubiquitous Internet in an integrated satellite-terrestrial environment: the SUITED solution, IEEE Communications Magazine 40 (1) (2002) 98–107.

[14] J.-H. Park, Wireless internet access for mobile subscribers based in the GPRS/UMTS network, IEEE Communications Magazine 40 (4) (2002) 38–49.

[15] J.D. Vriendt, P. Laine, C. Lerouge, X. Xu, Mobile network evolution: A revolution on the move, IEEE Communications Magazine 40 (4) (2002) 104–111.

[16] P. Chan, R. Sheriff, Y. Hu, P. Conforto, C. Tocci, Mobility management incorporating fuzzy logic for a heterogeneous IP environment, IEEE Communications Magazine 39 (12) (2001) 42–51.

[17] T. Zahariadis, K. Vaxevanakis, C. Tsantilas, N. Zervos, N. Nikolaou, Gloabal roaming in next-generation networks, IEEE Communications Magazine 40 (2) (2002) 145–151.

[18] W. Wang, I. Akyildiz, Intersystem location update and paging schemes for multitier wireless networks, in: Proceedings of ACM Mobicom 2000, pp. 99–109.

[19] W. Wang, I. Akyildiz, A new signaling protocol for intersystem roaming in next generation wireless systems, IEEE Journal on Selected Areas in Communications 19 (10) (2001) 2040–2052.

[20] I. Akyildiz, W. Wang, A dynamic location management scheme for next generation multitier PCS systems, IEEE Transactions on Wireless Communications 1 (1) (2002) 178–189.

[21] E. Cayirci, I. Akyildiz, User mobility pattern scheme for location update and paging in wireless systems, IEEE Transactions on Mobile Computing 1 (3) (2002) 236–247.

[22] Y.-B. Lin, Reducing location update cost in a PCS network, IEEE/ACM Transactions on Networking 5 (1) (1997) 25–33.

[23] R. Jain, Y. Lin, N. Lo, S. Mohan, A caching strategy to reduce network impacts of PCS, IEEE Journal on Selected Areas of Communications 12 (8) (1994) 1434–1445.

Eylem Ekici has received his BS and MS degrees in Computer Engineering from Bogazici University, Istanbul, Turkey, in 1997 and 1998, respectively. He received his Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, GA, in 2002. Currently, he is an assistant professor in the Department of Electrical and Computer Engineering of the Ohio State University, Columbus, OH. Dr. Ekici's current research interests include wireless sensor networks, next generation wireless systems, space-based networks, and vehicular communication systems, with a focus on routing and medium access control protocols, resource management, and analysis of network architectures and protocols. He also conducts research on interfacing of dissimilar networks.