

Performance Evaluation of Location Management in UMTS

Shun-Ren Yang and Yi-Bing Lin, *Fellow, IEEE*

Abstract—The universal mobile telecommunications system utilizes a three-level location-management strategy to reduce the net costs of location update and paging in the packet-switched service domain. Within a communication session, a mobile station (MS) is tracked at the cell level during packet transmission. In the idle period of an ongoing session, the MS is tracked at the UTRAN registration area (URA) level to avoid frequent cell updates while still keeping the radio connection. If the MS is not in any communication session, the MS is tracked at the routing-area (RA) level. The inactivity counter mechanism was proposed in 3GPP 25.331 to determine when to switch between the three location-tracking modes. In this mechanism, two inactivity counters are used to count the numbers of cell and URA updates in an idle period between two packet transmissions. If the number of cell updates reaches a threshold K_1 , the MS is switched from cell tracking to URA tracking. After that, if the number of URA updates reaches a threshold K_2 , the MS is tracked at the RA level. This paper proposes analytical and simulation models to investigate the performance of the inactivity counter mechanism. Our study provides guidelines for K_1 and K_2 selection to achieve lower net costs of location update and paging.

Index Terms—Mobile network, mobility management, universal mobile telecommunications system (UMTS), wireless data.

I. INTRODUCTION

EXISTING second-generation (2G) mobile-communications systems [such as global system for mobile communications (GSM)] are designed for voice services, which only have limited capabilities for offering data services. On the other hand, the third-generation (3G) systems [such as universal mobile telecommunications system (UMTS)] [10] support mobile multimedia applications with high data-transmission rates. As shown in Fig. 1, the UMTS infrastructure includes the core network (CN) and the UMTS terrestrial radio-access network (UTRAN). The CN is responsible for switching/routing calls and data connections to the external networks, while the UTRAN handles all radio-related functionalities. The CN consists of two service domains: the

circuit-switched (CS) service domain and the packet-switched (PS) service domain. The CS domain provides access to the PSTN/ISDN, while the PS domain provides access to the IP-based networks. In the remainder of this paper, we will focus on the UMTS packet-switching mechanism. In the PS domain of the CN, the packet data services of a mobile station [MS; see Fig. 1(a)] are provided by the serving GPRS support node [SGSN; see Fig. 1(d)] and the gateway GPRS support node [GGSN; see Fig. 1(e)]. The SGSN connects the MS to the external data network through the GGSN. The UTRAN consists of Node Bs [the 3G term for base stations; see Fig. 1(b)] and radio network controllers [RNCs; see Fig. 1(c)] connected by an asynchronous transfer mode (ATM) network. The connection between the UTRAN and the CN is achieved via the ATM links between the RNCs and SGSNs. The MS communicates with Node Bs through the radio interface based on the wide-band CDMA (WCDMA) technology.

The cells (i.e., radio coverages of Node Bs) in a UMTS service area are partitioned into several groups. To deliver services to an MS, the cells in the group covering the MS will page the MS to establish the radio connection. The location change of an MS is detected as follows. The cells periodically broadcast their cell identities. The MS listens to the broadcast cell identity and compares it with the cell identity stored in the MS's buffer. If the comparison indicates that the location has been changed, then the MS sends the location-update message to the network. In the UMTS PS domain, the cells are grouped into routing areas (RAs). The RA of an MS is tracked by the SGSN. The cells in an RA are further grouped into UTRAN registration areas (URAs). The URA and the cell of an MS are tracked by the UTRAN. Fig. 2 illustrates an example layout of cells, URAs, and RAs.

In UMTS, the mobility-management activities for an MS are characterized by two finite state machines: mobility management (MM) and radio resource control (RRC). The packet MM (PMM) state machine for the UMTS PS domain is exercised between the SGSN and the MS for CN-level tracking, while the RRC state machine is executed between the UTRAN and the MS for UTRAN-level tracking. Incomplete state diagrams for these two machines are illustrated in Fig. 3. Specifically, the figure only considers the states after the MS has attached to the PS domain. The PMM state diagram in the MS is slightly different from that in the SGSN. To simplify the presentation, we only show the common portions of the MS and the SGSN state transitions that are used in this paper. The state diagrams in Fig. 3 are described as follows. After an MS is attached to the PS service domain, the PMM state machine will be in one of the two states: PMM idle and PMM connected. In the RRC state machine, there are three states: RRC idle mode, RRC cell-connected mode, and

Manuscript received September 26, 2002; revised February 7, 2003. The work of S.-R. Yang was supported by the MediaTek Fellowship. The work of Y.-B. Lin was supported in part by the MOE Program for Promoting Academic Excellence of Universities under Grant 89-E-FA04-1-4, Chair Professorship of Providence University, IIS/Academia Sinica, FarEastone, CCL/ITRI, and the Lee and MTI Center for Networking Research/NCTU. This paper was presented in part at the International Conference on Information Networking 2003, Jeju Island, Korea, February 2003.

S.-R. Yang is with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. (e-mail: sjyoun@csie.nctu.edu.tw).

Y.-B. Lin is with Institute of Information Science, Academia Sinica, Taiwan, R.O.C. (e-mail: liny@csie.nctu.edu.tw)

Digital Object Identifier 10.1109/TVT.2003.819616

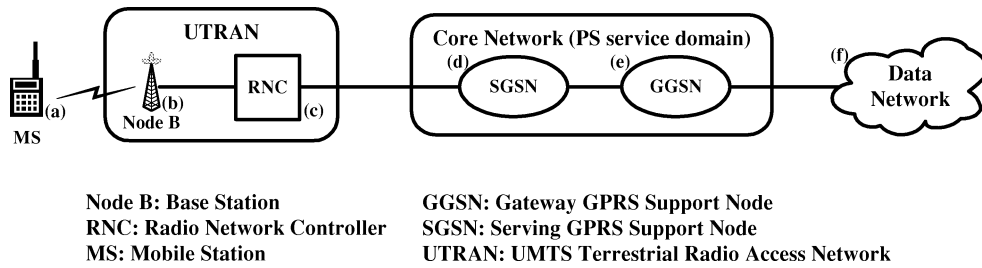


Fig. 1. Simplified network architecture for the UMTS PS domain.

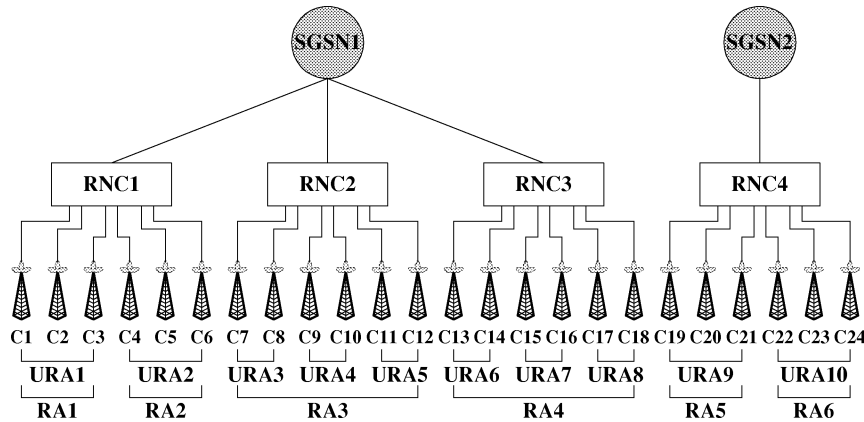


Fig. 2. Cells, URAs, and RAs.

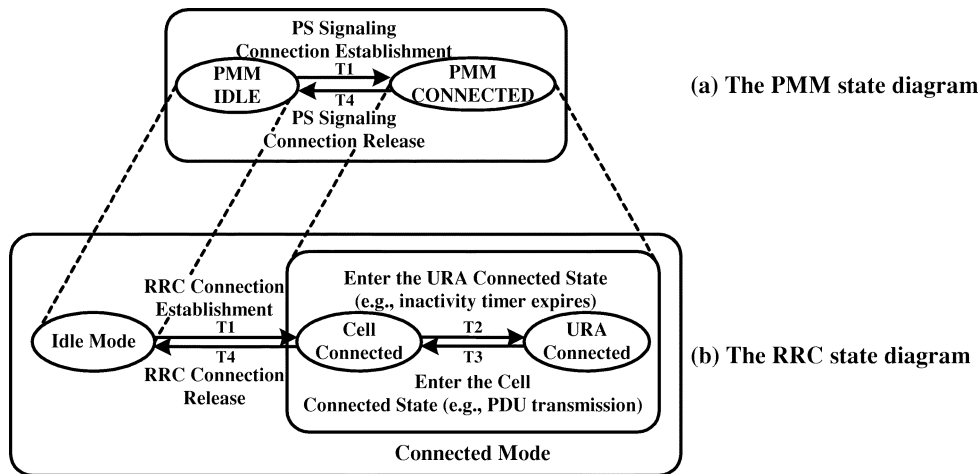


Fig. 3. State diagrams for UMTS mobility management. (a) Incomplete PMM state diagram for PS domain. (b) Simplified RRC state diagram.

RRC URA connected mode. We will briefly elaborate on the PMM and RRC state transitions. See [1] and [2] for complete descriptions of the PMM and RRC state machines.

When there is no data transmission between the MS and the core network, the MS is in the PMM idle state and RRC idle mode. In this case, UTRAN has no information about the idle MS and the MS is tracked by the SGSN at the RA level.

When a PS signaling connection is established between the MS and the SGSN (possibly in response to a page from the SGSN), the MS enters the PMM connected state [see T1 in Fig. 3(a)]. Since the establishment of the PS signaling connection triggers the establishment of the RRC connection between the MS and its serving RNC, the RRC state of the MS is switched to RRC cell-connected mode [see T1 in Fig. 3(b)]. In this case, the SGSN tracks the MS with accuracy of the

serving RNC and the serving RNC is responsible for tracking the cell where the MS resides. Packets can only be delivered in this state.

In the PMM connected/RRC cell-connected mode, if the MS has not transmitted/received packets for a period, the RRC state of the MS is switched to RRC URA connected mode [see T2 in Fig. 3(b)]. In this case, the RRC connection is still maintained, while the URA of the MS is tracked by the serving RNC. In this transition, the PMM state of the MS remains unchanged, i.e., the state is PMM connected.

In the PMM connected/RRC URA connected mode, if the MS transmits/receives a packet, the RRC state is moved back to RRC cell-connected mode [see T3 in Fig. 3(b)]. On the other hand, if the PS signaling connection and the RRC connection are released (e.g., a communication session is completed) or if

TABLE I
UMTS LOCATION-UPDATE COSTS

location update modes	cell update	URA update	RA update
location update cost	high	medium	low
paging cost	low	medium	high

no packet is transmitted for a long time, the RRC state is first switched to RRC cell-connected mode and then to RRC idle mode [see T3 and T4 in Fig. 3(b)]. In this case, the PMM state is also changed to PMM idle [see T4 in Fig. 3(a)].

The above three-level location-management strategy is designed to reduce the net costs of location update and paging. Within a communication session (i.e., the MS is in the RRC cell-connected mode and the PMM connected state), the MS expects to continuously transmit/receive packets. In this case, the MS performs a location update whenever it moves to a new cell. With a cell update, no cell is paged if there are dedicated channels allocated to the MS. If the MS has no dedicated channel, the currently visited cell is requested to page the MS for packet delivery. Without loss of generality, we assume one cell-paging cost in this paper. On the other hand, during an ongoing session, there may exist some idle periods where no packets are delivered. To avoid frequent cell updates while still keeping the RRC connection in this situation, the MS is switched to RRC URA connected mode to perform a URA update for every URA crossing. When the communication session between the MS and the SGSN is completed, the MS will not transmit/receive any packet for a long period. In this case, the RRC connection should be released for efficient wireless bandwidth utilization. In addition, the MS should be tracked at the RA level because cell- or URA-level tracking is too expensive in terms of signaling overhead and power consumption. Therefore, the RRC and PMM states are switched to RRC idle mode and PMM idle state, respectively, and the MS is tracked at the RA level. Based on the above description, the costs of the three location-update modes are summarized in Table I.

In the PMM and RRC state machines, the mechanism that triggers transitions T2 and T4 has significant impacts on the signaling traffic of the UMTS system. This mechanism can be implemented by two approaches. The first makes use of two inactivity timers X_1 and X_2 . At the end of a packet transmission, timer X_1 is set to a predefined threshold value and is decremented as time elapses. Transition T2 occurs if the MS does not transmit/receive any packets before timer X_1 expires. When timer X_1 expires, the second timer X_2 is set to a predefined threshold value and is decremented. Timer X_2 is used to determine the time when transition T4 occurs. In the second approach, two inactivity counters Y_1 and Y_2 are employed. Counter Y_1 counts the number of cell updates in the idle period between two packet transmissions. If the number of cell updates reaches a threshold K_1 , then the MS is switched to perform URA updates through transition T2. After T2 has occurred, counter Y_2 is used to count the number of URA updates in the observed idle period. If the number of URA updates reaches a threshold K_2 , then the MS is switched to perform RA updates (i.e., transition T4 occurs).

As noted in our previous work [12], the timer approach may have a synchronization problem. That is, the peer-state machines in the MS and the UTRAN or the SGSN may stay in different states at the same time due to the errors of the clock rates. Besides, the counter approach may significantly outperform the timer approach because the timer approach uses two timers of fix-length thresholds. When the mobility rate and/or packet-transmission patterns change, the fixed thresholds of the timers do not adapt to the changes. On the other hand, the thresholds K_1 and K_2 of the counter approach always capture the K_1 th cell update and the K_2 th URA update of an MS no matter how the mobility rate and packet-transmission patterns change. Therefore, this paper will not elaborate on the inactivity timer approach and will focus only on the inactivity counter mechanism. In the following sections, we propose analytical and simulation models to investigate the performance of the inactivity counter mechanism. Specifically, given any mobility and traffic patterns, we determine the net costs of location update and paging under various K_1 and K_2 threshold values. Our study provides guidelines for K_1 and K_2 selection that results in lower net costs.

II. ANALYTICAL MODEL FOR INACTIVITY COUNTER MECHANISM

This section proposes an analytical model to study the UMTS inactivity counter mechanism. We first describe a two-dimensional (2-D) random walk model for user movement based on a hexagonal cellular configuration. We then show how to use this model to investigate the performance of the UMTS inactivity counter mechanism. For specific thresholds K_1 and K_2 , we derive the expected number of location updates (including cell, URA, and RA updates) performed in the idle period between two packet transmissions and the expected number of cells that need to be paged for packet delivery.

A. Random Walk Model

For demonstration purposes, we consider the hexagonal cell layout in Fig. 4. In this configuration, the cells are clustered into several URAs, which are in turn clustered into several RAs. An n -layer URA covers $S(n) = 3n^2 - 3n + 1$ cells. Fig. 4 illustrates a three-layer URA. The cell at the center of the URA is referred to as the layer-0 cell. The cells surrounding layer $x - 1$ cells are referred to as the layer- x cells. There are $6x$ cells in layer x except for layer 0, which contains exactly one cell. An n -layer URA consists of cells from layer 0 to $n - 1$. The structure of an n -layer RA is similar to that of an n -layer URA, except that the basic elements are URAs instead of cells. Therefore, an n -layer RA covers $S(n)$ URAs. Fig. 4 illustrates seven two-layer RAs (A, B, C, D, E, F, and G). Each consists of seven three-layer URAs. Based on this RA/URA/cell structure, we derive the number of cells and URAs that are visited before a user moves out of an RA.

In [3], we proposed a cell-type classification algorithm based on the random walk model. The work showed that an n -layer hexagonal cellular network can be modeled by a state transition diagram with $n(n + 1)/2$ states. Based on our model, Tseng *et*

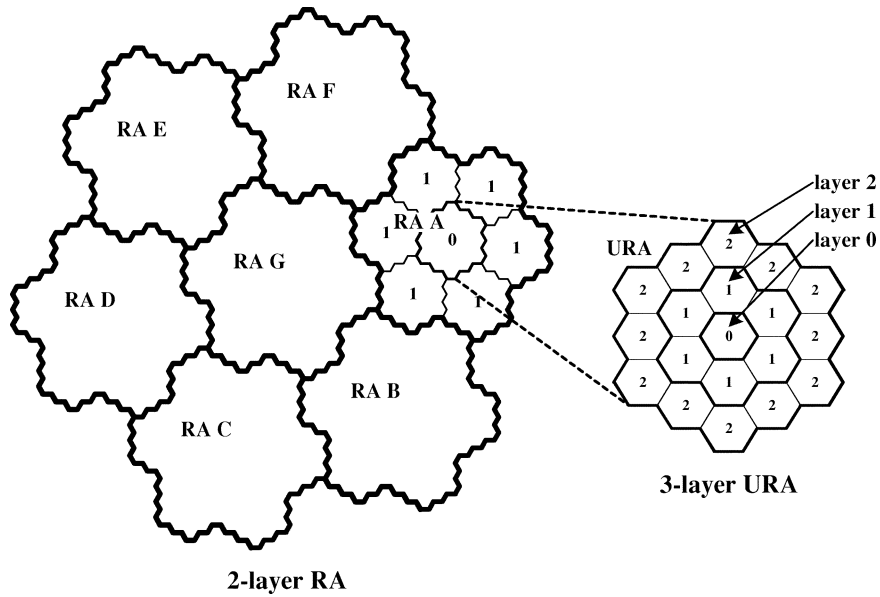


Fig. 4. Cell/URA/RA layout in a UMTS network.

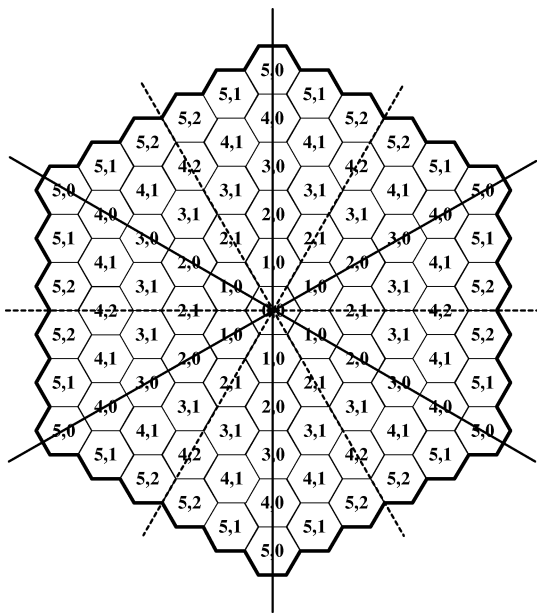


Fig. 5. Type classification for a six-layer UTRAN registration area.

al. [16] further reduced the number of states to $A(n)$, where for $n > 1$

$$A(n) = \begin{cases} \frac{(n+1)(n+3)}{4}, & n \text{ is odd} \\ \frac{n(n+4)}{4}, & n \text{ is even.} \end{cases} \quad (1)$$

According to the type-classification algorithm in [3] and [16], we classify the cells in a URA into several cell types. For $x \geq 0$ and $y \geq 0$, a cell type is of the form $\langle x, y \rangle$, where x represents that the cell is in layer x and y represents the y th type in layer x . Cells of the same type are indistinguishable in terms of movement pattern, because they are at the symmetrical positions (with respect to the solid and dashed lines in Fig. 5) on the hexagonal URA. Based on the derivation in [16], Fig. 5 labels the types of cells for a six-layer URA. In this example, the

cell in layer 0 is of the type $\langle 0, 0 \rangle$. The six cells in layer 1 are grouped together and assigned to the same type $\langle 1, 0 \rangle$. A layer-2 cell may have three or two neighbors in layer 3 and is assigned to types $\langle 2, 0 \rangle$ and $\langle 2, 1 \rangle$, respectively.

Based on the above cell labeling, we compute the number of movements an MS will take to cross the boundary of an n -layer URA. A state of this random walk is of the form (x, y) . For $0 \leq x < n$ and $0 \leq y \leq \lfloor x/2 \rfloor$, the state (x, y) is transient, which represents that the MS is in one of the cells of type $\langle x, y \rangle$. For $x = n$ and $0 \leq y \leq \lfloor (n-1)/2 \rfloor$, the state (n, y) is absorbing, which represents that the MS crosses the boundary of the URA from a cell of type $\langle n-1, y \rangle$. Details of the random walk and derivation of its steady-state probabilities are given in [3]. The results are summarized as follows. Let $p_{(x,y),(x',y')}$ be the one-step transition probability from state (x, y) to state (x', y') , i.e., the probability that the MS moves from a $\langle x, y \rangle$ cell to a $\langle x', y' \rangle$ cell in one step. The transition probability matrix $P = (p_{(x,y),(x',y')})$ of the random walk is

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & \frac{1}{3} & \frac{1}{6} & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}_{A(n) \times A(n)} \quad (2)$$

As an example, the element $p_{12} = 1/6$ in matrix P represents that the MS moves from a $\langle 1, 0 \rangle$ cell to a $\langle 2, 0 \rangle$ cell in one step with probability $1/6$. Define $P^{(k)}$ as

$$P^{(k)} = \begin{cases} P, & \text{for } k = 1 \\ P^{(k-1)}P, & \text{for } k > 1. \end{cases} \quad (3)$$

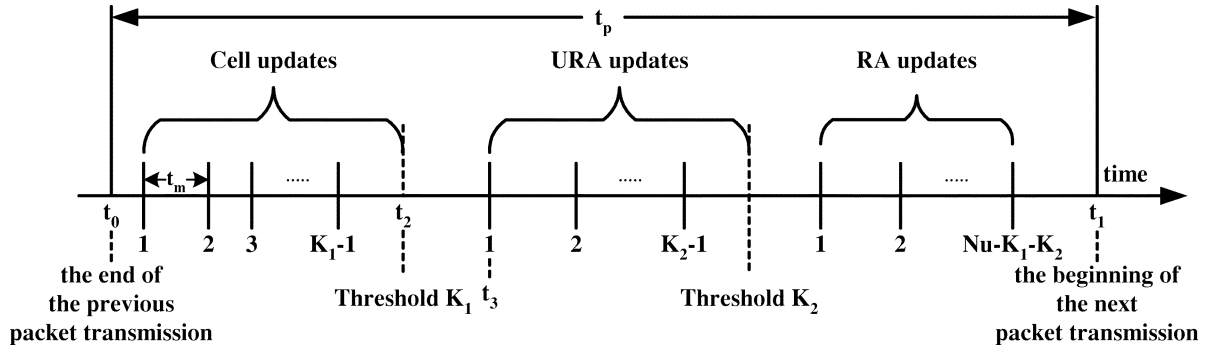


Fig. 6. Timing diagram for cell, URA, and RA updates.

An element $p_{(x,y),(x',y')}^{(k)}$ in $P^{(k)}$ is the probability that the random walk moves from state (x, y) to state (x', y') with k steps [it is possible that state (x', y') is revisited several times during these k steps]. Let $p_{k,(x,y),(x',y')}$ be the probability that an MS initially stays at state (x, y) and takes its first entrance into state (x', y') at the k th step. Then, $p_{k,(x,y),(n,j)}$ is expressed as

$$p_{k,(x,y),(n,j)} = \begin{cases} p_{(x,y),(n,j)}, & \text{for } k = 1 \\ p_{(x,y),(n,j)}^{(k)} - p_{(x,y),(n,j)}^{(k-1)}, & \text{for } k > 1 \end{cases} \quad (4)$$

which can be solved by using the transition probability matrices (2) and (3). Equation (4) gives the probability that an MS will move out of a URA at the k th cell crossing. Note that the number of URA movements before the user leaves an RA can also be derived using the same approach where the cells are replaced by the URAs.

B. Location Update and Paging Costs

Fig. 6 shows the timing diagram of the location-update activities for an MS in an idle period between two packet transmissions. Suppose that the previous packet transmission of the MS ends at time t_0 and the next packet transmission begins at time t_1 . Let $t_p = t_1 - t_0$. Consider the execution of the inactivity counter mechanism with specific thresholds K_1 and K_2 . For the first K_1 cell crossings during period t_p , the MS performs a cell update for each of the cell crossings. After the K_1 th cell update, the MS will not perform any cell updates and begins performing the URA updates. After the K_2 th URA update, the MS will not perform any URA updates and begins performing the RA updates. Assume that the K_1 th cell update occurs at time t_2 and that the first URA update occurs at time t_3 . Based on the equal-routing probability assumption of the random walk at time t_2 , the MS will stay at any cell of the visited URA with equal probability. On the other hand, after any URA update, the MS will only stay at one of the boundary cells of the visited URA. Thus, to derive the number of URA updates in t_p , we need to consider the following two cases.

Case I. Starting from an arbitrary cell, we compute the number of cell crossings before the MS moves out of the URA [i.e., the number of cell crossings in (t_2, t_3) in Fig. 6]. As we mentioned before, at time t_2 , the MS can be in any cell of the URA with the same probability. Therefore, at

t_2 , the MS is at the $\langle 0, 0 \rangle$ cell with probability $1/S(n)$, at a cell of type $\langle x, 0 \rangle$ ($1 \leq x < n$) with probability $6/S(n)$, at a cell of type $\langle 2x, x \rangle$ ($1 \leq x \leq \lfloor (n-1)/2 \rfloor$) with probability $6/S(n)$ and at a cell of type $\langle x, y \rangle$ ($3 \leq x < n$, $1 \leq y \leq \lfloor (x-1)/2 \rfloor$) with probability $12/S(n)$, where

$$S(n) = 3n^2 - 3n + 1 \quad (5)$$

is the number of cells covered by an n -layer URA. Consider a six-layer URA that covers $S(6) = 91$ cells (see Fig. 5). Since there are six cells of type $\langle 2, 1 \rangle$ and 12 cells of type $\langle 3, 1 \rangle$ in the URA, the MS is in a cell of type $\langle 2, 1 \rangle$ with probability $6/91$ and in a cell of type $\langle 3, 1 \rangle$ with probability $12/91$. Let $\theta_1(n, k)$ be the probability that after t_2 , the MS will leave the n -layer URA at the k th cell movement. Then

$$\begin{aligned} \theta_1(n, k) = & \left[\frac{1}{S(n)} \right] \left[\sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k,(0,0),(n,j)} \right] \\ & + \left[\frac{12}{S(n)} \right] \left[\sum_{x=3}^{n-1} \sum_{y=1}^{\lfloor \frac{(x-1)}{2} \rfloor} \sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k,(x,y),(n,j)} \right] \\ & + \left[\frac{6}{S(n)} \right] \left[\sum_{x=1}^{n-1} \sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k,(x,0),(n,j)} \right. \\ & \left. + \sum_{x=1}^{\lfloor \frac{(n-1)}{2} \rfloor} \sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k,(2x,x),(n,j)} \right]. \quad (6) \end{aligned}$$

Case II. Starting from a boundary cell of an URA at time t_3 , we compute the number of cell crossings before the MS moves out of the URA. In [13], we showed that after entering the URA, the MS is in a boundary cell with probability proportional to the number of boundary edges for that boundary cell. Under the condition that an MS is moving into a boundary cell, the MS enters a type $\langle n-1, 0 \rangle$ cell with probability $3 \cdot 6/B(n)$, enters a type $\langle n-1, (n-1)/2 \rangle$ cell (when n is odd) with probability $2 \cdot 6/B(n)$, and enters a type $\langle n-1, y \rangle$ cell ($1 \leq y \leq \lfloor n/2 \rfloor - 1$) with probability $2 \cdot 12/B(n)$, where $B(n) = 6[3 + 2(n-2)]$ is the number of boundary edges in an n -layer URA. In Fig. 5,

$B(n) = 66$ for a six-layer URA. In this example, there are three boundary edges for each of the six $\langle 5, 0 \rangle$ cells and the MS enters a boundary cell of type $\langle 5, 0 \rangle$ with probability $18/66$. Similarly, there are two boundary edges for each of the 12 $\langle 5, 2 \rangle$ cells and the MS enters a boundary cell of type $\langle 5, 2 \rangle$ with probability $24/66$. Let $\theta_2(n, k)$ be the probability that after an MS enters an n -layer URA, it moves out of the URA at the k th cell movement. If n is odd, then

$$\begin{aligned} \theta_2(n, k) = & \left[\frac{3 \cdot 6}{B(n)} \right] \left[\sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k, (n-1, 0), (n, j)} \right] \\ & + \left[\frac{2 \cdot 6}{B(n)} \right] \left[\sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k, (n-1, \frac{(n-1)}{2}), (n, j)} \right] \\ & + \left[\frac{2 \cdot 12}{B(n)} \right] \left[\sum_{y=1}^{\lfloor \frac{n}{2} \rfloor - 1} \sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k, (n-1, y), (n, j)} \right]. \quad (7) \end{aligned}$$

If n is even, then

$$\begin{aligned} \theta_2(n, k) = & \left[\frac{3 \cdot 6}{B(n)} \right] \left[\sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k, (n-1, 0), (n, j)} \right] \\ & + \left[\frac{2 \cdot 12}{B(n)} \right] \left[\sum_{y=1}^{\lfloor \frac{n}{2} \rfloor - 1} \sum_{j=0}^{\lfloor \frac{(n-1)}{2} \rfloor} p_{k, (n-1, y), (n, j)} \right]. \quad (8) \end{aligned}$$

Suppose that an MS is in an arbitrary cell of an n -layer URA. Let $P(n, N_m, k)$ be the probability that after N_m cell movements, the MS crosses k URA boundaries. Similarly, consider an MS initially residing at a boundary cell of an n -layer URA. Let $P^*(n, N_m^*, k^*)$ be the probability that after N_m^* cell movements, the MS crosses k^* URA boundaries. From (6)–(8), we have

$$\begin{aligned} P(n, N_m, k) & \begin{cases} 1, & \text{for } k = N_m = 0 \\ \sum_{j=N_m+1}^{\infty} \theta_1(n, j), & \text{for } k = 0, N_m > 0 \\ \sum_{j=1}^{N_m} [\theta_1(n, j) \\ \times P^*(n, N_m - j, k - 1)], & \text{for } k \geq 1, N_m \geq k \\ 0, & \text{for } N_m < k. \end{cases} \quad (9) \end{aligned}$$

Four cases are considered in (9).

Case I: $k = N_m = 0$. If there is no cell movement during an idle period, there will be no URA boundary crossing during this period and $P(n, N_m = 0, k = 0) = 1$.

Case II: $k = 0, N_m > 0$. Assume that the MS crosses the first URA boundary at the j th cell movement [with

probability $\theta_1(n, j)$]. For $N_m > 0$, if $k = 0$ (i.e., the MS does not cross any URA boundary during these N_m cell movements), j must be larger than N_m . Therefore, in this case, $P(n, N_m, k) = \Pr[j > N_m] = \sum_{j=N_m+1}^{\infty} \theta_1(n, j)$. *Case III:* $k \geq 1, N_m \geq k$. In this case, there are k URA boundary crossings during the N_m cell movements. If the MS crosses the first URA boundary at the j th cell movement [with probability $\theta_1(n, j)$], then the MS must cross the remaining $k - 1$ URA boundaries during the subsequent $N_m - j$ cell movements [with probability $P^*(n, N_m - j, k - 1)$]. Therefore, $P(n, N_m, k) = \sum_{j=1}^{N_m} \theta_1(n, j) P^*(n, N_m - j, k - 1)$.

Case IV: $N_m < k$. During an idle period, it is impossible that the number of URA boundary crossings is larger than the number of cell movements. Thus, $P(n, N_m, k) = 0$ in this case.

Similarly, we have

$$\begin{aligned} P^*(n, N_m^*, k^*) & \begin{cases} 1, & \text{for } k^* = N_m^* = 0 \\ \sum_{j=N_m^*+1}^{\infty} \theta_2(n, j), & \text{for } k^* = 0, N_m^* > 0 \\ \sum_{j=1}^{N_m^*} [\theta_2(n, j) \\ \times P^*(n, N_m^* - j, k^* - 1)], & \text{for } k^* \geq 1, N_m^* \geq k^* \\ 0, & \text{for } N_m^* < k^*. \end{cases} \quad (10) \end{aligned}$$

Equations (9) and (10) can be effectively computed by using the dynamic programming technique [14]. Note that (6)–(10) can also be used to derive the number of RA boundary crossings for an n -layer RA layout, given that the number of URA movements is known.

With (9) and (10), we derive the number of cell/URA/RA updates in the idle period between two packet transmissions as follows. Assume that $t_p = t_1 - t_0$ in Fig. 6 has a general distribution with the density function $f_p(t_p)$, the expected value $1/\lambda_p$, and the Laplace transform

$$f_p^*(s) = \int_{t_p=0}^{\infty} e^{-st_p} f_p(t_p) dt_p.$$

For the inactivity counter mechanism with specific thresholds K_1 and K_2 , let N_u be the number of location updates (including the cell, URA, and RA updates) during period t_p . Based on the aforementioned random walk model, the distribution of N_u can be derived as follows. Suppose that the cell residence time t_m has an Erlang distribution with mean $1/\lambda_m = m/\lambda$, variance $V_m = m/\lambda^2$, and density function

$$f_m(t_m) = \frac{\lambda e^{-\lambda t_m} (\lambda t_m)^{m-1}}{(m-1)!}, \quad \text{for } t_m \geq 0 \quad (11)$$

where $m = 1, 2, 3, \dots$. We select the Erlang distribution because this distribution can be easily extended into a hyper-Erlang distribution, which has been proven to be a good

approximation to many other distributions as well as measured data [7], [8], [11].

Since the cell crossings of an MS can be modeled as an equilibrium Erlang-renewal process [15], the probability mass function of the number of cell crossings N_c within t_p is

$$\begin{aligned} \Pr[N_c = k] &= \int_{t_p=0}^{\infty} \left(\frac{e^{-\lambda t_p}}{m} \right) \\ &\quad \times \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j)(\lambda t_p)^j}{j!} \right] \right. \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m)(\lambda t_p)^j}{j!} \right] \right\} \\ &\quad \times f_p(t_p) dt_p \\ &= \left(\frac{1}{m} \right) \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j)\lambda^j}{j!} \right] \right. \\ &\quad \times \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m)\lambda^j}{j!} \right] \right. \\ &\quad \left. \times \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \right\} \\ &= \left(\frac{1}{m} \right) \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j)(-\lambda)^j}{j!} \right] \right. \\ &\quad \times \left[\left. \frac{d^j f_p^*(s)}{ds^j} \right] \right|_{s=\lambda} \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m)(-\lambda)^j}{j!} \right] \right. \\ &\quad \left. \times \left[\left. \frac{d^j f_p^*(s)}{ds^j} \right] \right|_{s=\lambda} \right\} \end{aligned}$$

for $k = 1, 2, \dots$. For $k = 0$, we have

$$\begin{aligned} \Pr[N_c = 0] &= \int_{t_p=0}^{\infty} \left(\frac{e^{-\lambda t_p}}{m} \right) \\ &\quad \times \sum_{j=0}^{m-1} \left[\frac{(m-j)(\lambda t_p)^j}{j!} \right] f_p(t_p) dt_p \\ &= \left(\frac{1}{m} \right) \left\{ \sum_{j=0}^{m-1} \left[\frac{(m-j)\lambda^j}{j!} \right] \right. \\ &\quad \left. \times \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \right\} \\ &= \left(\frac{1}{m} \right) \left\{ \sum_{j=0}^{m-1} \left[\frac{(m-j)(-\lambda)^j}{j!} \right] \right. \\ &\quad \left. \times \left[\left. \frac{d^j f_p^*(s)}{ds^j} \right] \right|_{s=\lambda} \right\}. \quad (12) \end{aligned}$$

Consider a UMTS network with n' -layer URA and n'' -layer RA structure. From (9), (10), and (12), the probability mass function for N_u is shown in (13) at bottom of the page. Equation (13) is explained as follows.

Case I: $j < K_1$. In this case, the MS only performs cell updates in the idle period. Therefore, the probability of N_u location updates is equal to the probability of N_c cell crossings.

Case II: $K_1 \leq j < K_1 + K_2$. In this case, there are k cell crossings in the idle period (with probability $\Pr[N_c = k]$). Note that $k \geq j \geq K_1$. Therefore, during these k cell crossings, the MS performs K_1 cell updates for the first K_1 cell crossings (with probability 1). After the MS enters the URA update mode, the MS performs $j - K_1$ URA updates during the subsequent $k - K_1$ cell crossings [with probability $P(n', k - K_1, j - K_1)$].

Case III: $j \geq K_1 + K_2$. Similar to Case II, k represents the number of cell crossings in the idle period, where $k \geq j \geq K_1 + K_2$. The MS first performs K_1 cell updates for the first K_1 cell crossings. Then, it enters the URA update mode, where there are i URA crossings in the remaining $k - K_1$ cell crossings with probability $P(n', k - K_1, i)$,

$$\Pr_{(n', n'')} [N_u = j]$$

$$= \begin{cases} \Pr[N_c = j], & \text{for } j < K_1 \\ \sum_{k=j}^{\infty} \{ \Pr[N_c = k] \\ \quad \times P(n', k - K_1, j - K_1) \}, & \text{for } K_1 \leq j < K_1 + K_2 \\ \sum_{k=j}^{\infty} \sum_{i=K_2}^{k-K_1} \{ \Pr[N_c = k] \times P(n', k - K_1, i) \\ \quad \times P(n'', i - K_2, j - K_1 - K_2) \}, & \text{for } j \geq K_1 + K_2 \end{cases} \quad (13)$$

TABLE II
COMPARISON BETWEEN THE ANALYTICAL AND SIMULATION RESULTS

Erlang $\lambda_m = \lambda$, $V_m = 1/\lambda^2$ ($K_2 = 5$, $\lambda = \lambda_{p1}/30$, $\lambda_{p2} = \lambda_{p1}/720$, $\alpha = 0.6$, $U = 4$, $V = 1$)						
K_1	1	5	9	13	17	21
C_T (Analytical)	25.2514	27.2976	29.1329	30.6918	32.0157	33.1403
C_T (Simulation)	25.5038	27.5381	29.3305	30.7652	32.1216	33.2821
Error	0.9995%	0.8810%	0.6783%	0.2392%	0.3308%	0.4279%
Erlang $\lambda_m = \lambda/5$, $V_m = 5/\lambda^2$ ($K_1 = 5$, $\lambda = \lambda_{p1}/30$, $\lambda_{p2} = \lambda_{p1}/720$, $\alpha = 0.6$, $U = 4$, $V = 1$)						
K_2	0	4	8	12	16	20
C_T (Analytical)	13.9573	9.14372	8.26156	8.09384	8.06196	8.05589
C_T (Simulation)	14.0157	9.12147	8.23973	8.05265	8.03802	8.03023
Error	0.4184%	0.2439%	0.2649%	0.5115%	0.2978%	0.3195%
Erlang $\lambda_m = \lambda/10$, $V_m = 10/\lambda^2$ ($K_2 = 5$, $\lambda = \lambda_{p1}/30$, $\lambda_{p2} = \lambda_{p1}/720$, $\alpha = 0.6$, $U = 4$, $V = 1$)						
K_1	1	5	9	13	17	21
C_T (Analytical)	5.61760	4.99601	4.87692	4.85365	4.84910	4.84822
C_T (Simulation)	5.58651	4.99591	4.88063	4.84336	4.84785	4.85471
Error	0.5565%	0.0020%	0.0761%	0.2125%	0.0258%	0.1339%

where $i \leq k - K_1$. During these i URA crossings, the MS performs K_2 URA updates for the first K_2 URA crossings (with probability 1). Then the MS performs $j - K_1 - K_2$ RA updates in the subsequent $i - K_2$ URA crossings with probability $P(n'', i - K_2, j - K_1 - K_2)$, where $i \geq K_2$.

Based on (13), we derive the net cost C_T of location update and paging during the expected period $E[t_p]$. Assume that the cost for performing a location update is U and that the cost for paging at one cell is V . Let C_u be the expected location-update cost during t_p . From (13), we have

$$C_u = U \sum_{j=0}^{\infty} j \Pr_{(n', n'')} [N_u = j]. \quad (14)$$

For the paging operation, there are three possibilities.

- If the MS stays at the cell-update mode, then only one cell needs to page the MS.
- If the MS stays at the URA update mode, then all cells of the URA should page the MS.
- If the MS stays at the RA update mode, then all cells of the RA should page the MS.

Let C_v be the expected paging cost during t_p . Based on (13), we have

$$C_v = V \left\{ \sum_{j=0}^{K_1-1} \Pr_{(n', n'')} [N_u = j] + S(n') \sum_{j=K_1}^{K_1+K_2-1} \Pr_{(n', n'')} [N_u = j] + S(n')S(n'') \sum_{j=K_1+K_2}^{\infty} \Pr_{(n', n'')} [N_u = j] \right\}. \quad (15)$$

From (14) and (15), the net cost C_T for location update and paging during t_p is

$$C_T = C_u + C_v. \quad (16)$$

The analytical analysis has been validated by a discrete event-simulation model. The simulation simulates the movement of an MS on the hexagonal plane, where the UMTS network consists of two-layer URAs and RAs. The t_p interval is exponentially distributed and the cell residence times have the Erlang distribution given in (11). Note that our models can be applied to general t_p distributions and that we only show the exponential case for demonstration. Table II compares the analytical and simulation results. The parameters λ_{p1} , λ_{p2} , and α used in Table II will be explained in detail later. The table indicates that the errors between the analytical and simulation models are within 1%. Also, the errors for the $\theta_1(n, k)$, $\theta_2(n, k)$, $P(n, N_m, k)$, and $P^*(n, N_m^*, k^*)$ values are less than 1% in most cases. Details of these results will not be presented. It is clear that the analytical analysis is consistent with the simulation results.

III. NUMERICAL EXAMPLES

Based on the simulation validated by the analytical model described in the previous section, this section investigates the performance of the UMTS inactivity counter mechanism. The experiment settings in this study are similar to that of our previous study on the GPRS ready-counter mechanism [12]. We combine the ETSI packet data model [6] with the ON/OFF source model (also known as a packet train model) [4]. As shown in Fig. 7, we assume that the packet data traffic consists of communication sessions. Within a communication session, packet traffic is characterized by ON/OFF periods. In an ON period, a burst of data packets are transmitted. In an OFF period, no packets are delivered. Other assumptions are summarized as follows.

- The OFF period t_{p1} is drawn from a Pareto distribution with mean $1/\lambda_{p1}$ and infinite variance. It has been shown that the Pareto distribution with infinite variance can match with the actual data traffic measurements very well [17]. A Pareto distribution has two parameters, β and l , where β describes the ‘‘heaviness’’ of the tail of the distribution. The probability density function is $f_{p1}(t_{p1}) = (\beta/l)(l/t_{p1})^{\beta+1}$ and the expected value is $E[t_{p1}] = (\beta/\beta - 1)l$. If β is between 1 and 2, the

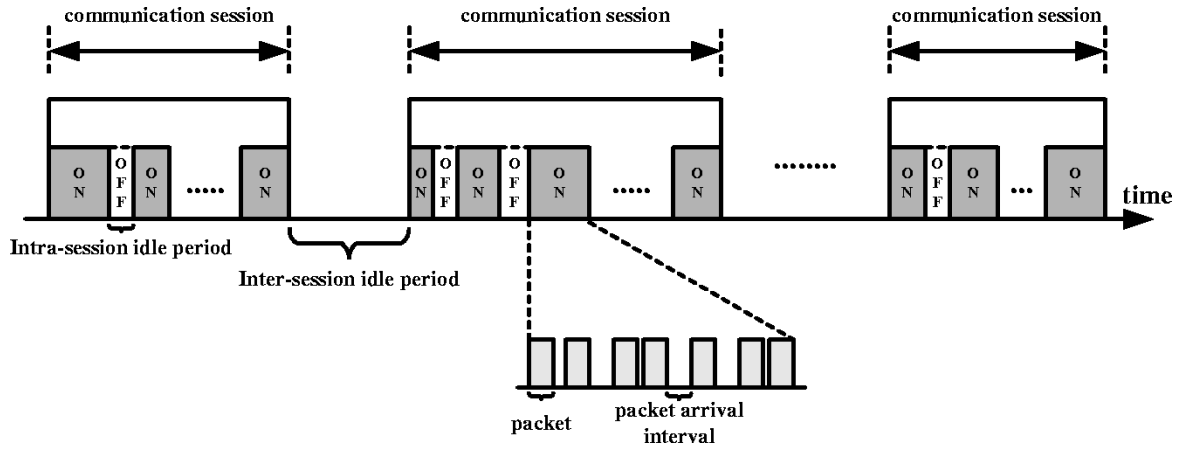
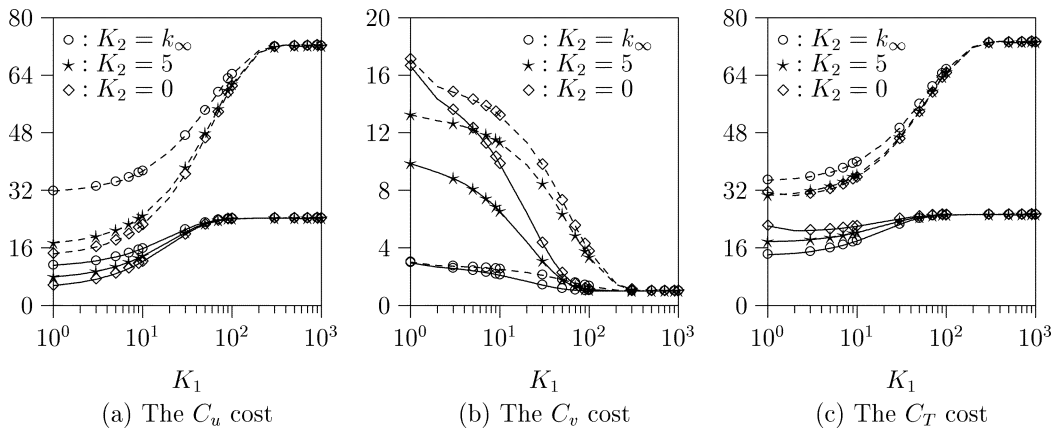


Fig. 7. Packet data traffic.


 Fig. 8. Effects of K_1 , K_2 , and λ_{p2} on C_u , C_v , and C_T (solid: $\lambda_{p2} = (1/200)\lambda_{p1}$; dashed: $\lambda_{p2} = (1/600)\lambda_{p1}$; $\lambda_m = (1/10)\lambda_{p1}$, $V_m = (1/\lambda_m^2)$, $V_{p2} = (1/\lambda_{p2}^2)$, $(U/V) = 4$, and $\alpha = 0.7$).

variance for the distribution becomes infinite. The typical parameter values obtained in [17] are $E[t_{p1}] = 10.5$ s and $\beta = 1.2$ for OFF periods. Our study uses the above $E[t_{p1}]$ and β values.

- The idle period t_{p2} between two consecutive communication sessions has a Gamma distribution with mean $1/\lambda_{p2}$ and variance V_{p2} . The Gamma distribution with shape parameter η and scale parameter λ (i.e., mean $1/\lambda_{p2} = \eta/\lambda$ and variance $V_{p2} = \eta/\lambda^2$) has the following density function:

$$f_{p2}(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^{\eta-1}}{\Gamma(\eta)}, \quad \text{for } t \geq 0$$

where $\Gamma(\eta) = \int_{z=0}^{\infty} z^{\eta-1} e^{-z} dz$ is the Gamma function. It has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions ([11], see Lemma 3.9). In this paper, we use the Gamma distribution to investigate the impact of variance for intersession idle periods.

- Following the ETSI packet data model, the number of OFF periods in a session has a geometric distribution with mean $\alpha/(1-\alpha)$ where $0 \leq \alpha < 1$. In other words, an ON period is followed by an OFF period with probability α and

is followed by an intersession idle period with probability $1-\alpha$.

- The cell-residence times have a Gamma distribution with the mean $1/\lambda_m$ and variance V_m . The Gamma distribution was employed to model MS movement in many studies [5], [8], [9] and is used in this paper to investigate the impact of variance for cell-residence times.

To simplify our discussion, we consider the two-layer URA and the two-layer RA cell layout. The effects of the input parameters are investigated as follows.

Fig. 8 shows how K_1 affects C_u , C_v , and C_T . For a fixed K_2 , it is clear that if K_1 increases, the location-update (LU) cost C_u increases while the paging cost C_v decreases. When K_1 is very large, the MS always performs cell updates and no URA or RA update is executed. In this case, the C_u and C_v costs are not affected by the change of K_1 [see the C_u and C_v curves in Figs. 8(a) and (b), where $K_1 \geq 10^2$]. In Fig. 8(c), the C_T is computed directly from the C_u and C_v using (16), where $(U/V) = 4$. For the input parameters selected in Fig. 8(c), the lowest C_T costs are observed when $K_1 = 1$ or 2 . The selection of K_1 is affected by t_{p1} and t_{p2} , which are described in the following two cases.

Case I: Since the OFF periods t_{p1} are short [specifically $E[t_{p1}] = (1/10\lambda_m)$ in Fig. 8], there are few cell updates

TABLE III
 C_u AND C_v COSTS IN A t_{p2} PERIOD ($\lambda_{p1} = 10\lambda_m$; COSTS ARE
 NORMALIZED BY ONE LU COST)

K_1 and K_2	$\lambda_m = 20\lambda_{p2}$	$\lambda_m = 60\lambda_{p2}$
$K_1 = k_\infty, K_2 = k_{any}$	$C_T = 20.25$	$C_T = 60.25$
$K_1 = 1, K_2 = k_\infty$	$C_T = 10.8$	$C_T = 28.0$
$K_1 = 1, K_2 = 0$	$C_T = 16.1$	$C_T = 23.9$

during t_{p1} and the paging cost C_v dominates the C_T cost. Therefore, to maintain low C_v cost (i.e., to ensure that only one cell is paged for the next packet delivery), the MS should stay at the cell-update mode. For all experiments considered in this paper, the probability that there are less than two cell crossings during t_{p1} is larger than 99%. Therefore, if $K_1 \geq 2$ is selected, then it is likely that the MS will stay at the cell-update mode in the t_{p1} period. *Case II:* On the other hand, the MS crosses many cell boundaries during an intersession idle period t_{p2} . For example, the average number of cell crossings is 20 for $\lambda_m = 20\lambda_{p2}$ and 60 for $\lambda_m = 60\lambda_{p2}$. Consider the case where $\lambda_m = 20\lambda_{p2}$. Table III lists C_u and C_v normalized by one LU cost. When $K_1 = k_\infty^1$ and $K_2 = k_{any}^2$, the C_T cost is 20.25 LUs. When $K_1 = 1$ and $K_2 = k_\infty$, the C_T cost is 10.8 LUs. Based on the above discussion, we have

$$C_T(K_1 = 1) < C_T(K_1 = k_\infty). \quad (17)$$

Since Theorem 1 in the Appendix indicates that the lowest C_T value occurs when $K_1 = 1$ or $K_1 = k_\infty$, (17) implies that $K_1 = 1$ is the optimal threshold value in a t_{p2} period.

To obtain low C_T values by considering both Cases I and II, it is appropriate to select $K_1 = 1$ or 2.

Figs. 8 and 9 show how K_2 affects C_u , C_v , and C_T . The effects of K_2 on the location update cost C_u and paging cost C_v are similar to that of K_1 : given a K_1 value, the C_u cost is an increasing function of K_2 while the C_v cost is a decreasing function of K_2 . Fig. 8 also indicates that when K_1 is large, C_u and C_v are not affected by the change of K_2 . Fig. 9 shows that if λ_{p2} is small, C_T is an increasing function of K_2 . On the other hand, if λ_{p2} is large, C_T is a decreasing function of K_2 . This phenomenon is explained in the following two cases.

Case I: $\lambda_{p2} = (1/600)\lambda_{p1}$ (the t_{p2} period is long). When $K_1 = 1$ and $K_2 = k_\infty$, Table III indicates that the C_T cost is 28 LUs. When $K_1 = 1$ and $K_2 = 0$, the C_T cost is 23.9 LUs. Thus

$$C_T(K_2 = 0) < C_T(K_2 = k_\infty). \quad (18)$$

From Theorem 1, the lowest C_T value occurs when $K_2 = 0$ or $K_2 = k_\infty$, and (18) implies that the lowest C_T cost is expected when $K_2 = 0$.

Case II: $\lambda_{p2} = (1/200)\lambda_{p1}$ (the t_{p2} period is short). The C_T cost is 10.8 LUs when $K_1 = 1$ and $K_2 = k_\infty$ and

is 16.1 LUs when $K_1 = 1$ and $K_2 = 0$. In this case, Theorem 1 indicates that the lowest C_T cost is observed when $K_2 = k_\infty$.

In a real mobile-communications network, the mobility and traffic patterns of a mobile user usually change dynamically. To obtain lower C_T cost in such an environment, it is required to consider both high (i.e., large λ_{p2}) and low (i.e., small λ_{p2}) packet-arrival patterns. Fig. 10 illustrates the discrepancies $D(K_2 = j)$ of C_T for various λ_{p2} values when $K_2 = j$. The discrepancy $D(K_2 = j)$ is defined as

$$D(K_2 = j) = \frac{C_T(K_2 = j) - C_T^*}{C_T^*}$$

where $C_T(K_2 = j)$ is the C_T cost for $K_2 = j$ and C_T^* is the lowest C_T cost. Under a given λ_{p2} value, Fig. 9 indicates that there exists an optimal value K_2^* that results in the lowest C_T^* . Fig. 10 plots the $D(K_2 = j)$ curves for various λ_{p2} . When $\lambda_{p2} = (1/200)\lambda_{p1}$, Fig. 9 shows that the optimal K_2^* is k_∞ and that the corresponding C_T^* is 14.46. Since $C_T(K_2 = 0)$ is 20.76, the discrepancy $D(K_2 = 0)$ is $((20.76 - 14.46)/14.46) * 100 \simeq 43.57\%$.

Fig. 10 indicates that when $\lambda_{p2} < (1/440)\lambda_{p1}$, C_T is an increasing function of K_2 . In this case, the discrepancies $D(K_2 = j)$ for $j < 30$ are less than 10%. On the other hand, when $\lambda_{p2} > (1/440)\lambda_{p1}$, C_T is a decreasing function of K_2 . In this case, $D(K_2 = j)$ are less than 10% for $j > 20$. Therefore, for all λ_{p2} values considered in our experiments, if we want to maintain reasonably low C_T values such that $D(K_2 = j) < 10\%$, the K_2 value should be selected from the range [20,30].

Effects of λ_{p2} . Fig. 8 plots the C_u , C_v , and C_T curves for $\lambda_{p2} = (1/200)\lambda_{p1}$ and $\lambda_{p2} = (1/600)\lambda_{p1}$. The figures show that C_u , C_v , and C_T increase as λ_{p2} decreases. A small λ_{p2} implies a long intersession idle period and more cell movements during this period. Therefore, the location-update cost C_u will increase accordingly. For fixed K_1 and K_2 values, increasing the number of cell movements implies increasing the probability that the MS will enter the URA update mode or even the RA update mode when the next packet arrives. Thus, a high paging cost is expected. We also notice that when both K_1 and λ_{p2} are large, increasing K_1 only has an insignificant effect on C_T . This phenomenon is explained as follows. For large λ_{p2} and K_1 , it is likely that K_1 is larger than the number of cell crossings during the idle period. Therefore, the MS will only perform cell updates and increasing K_1 only insignificantly increases the C_T value.

Effects of α . A smaller α implies more intersession idle periods. Since more cell crossings are observed in an intersession idle period than in the OFF periods of a session, C_T increases as α decreases. Similar to the discussion for the interaction between K_1 and λ_{p2} , C_T is more sensitive to the change of K_1 for a small α than a large α .

Effects of variance V_m . Fig. 11 plots the C_T curves where the cell-residence times have a Gamma distribution with the mean $1/\lambda_m$ and the variance V_m . This figure indicates that C_T is a decreasing function of V_m . When V_m increases, more short and long cell-residence times are observed. Long cell-residence times imply a small number of cell crossings N_c , which result in a small C_T . On the contrary, short cell-residence times imply

¹In computer simulation, it is impossible to generate an infinite number. Thus, k_∞ is defined as a number larger than the number of cell crossings in any idle period.

² k_{any} is an arbitrary integer number.

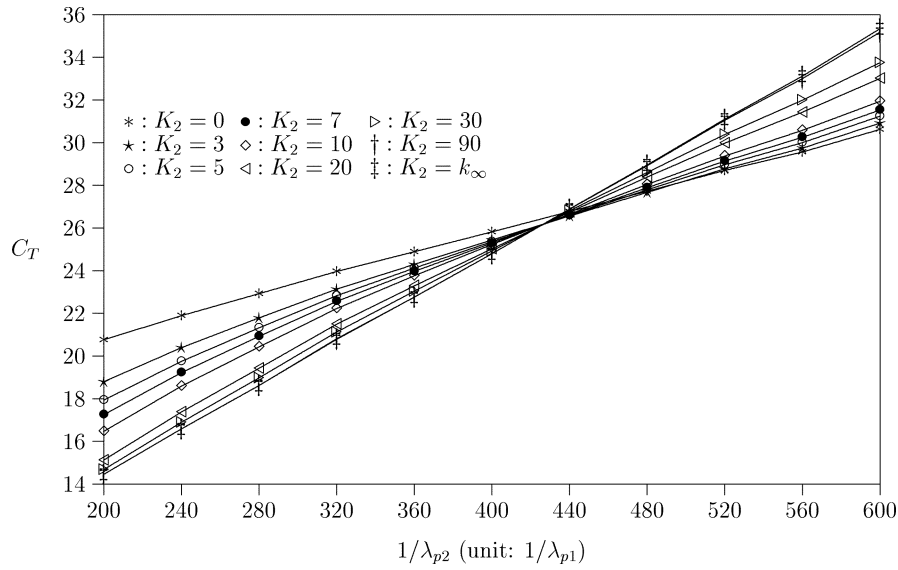


Fig. 9. Effects of K_2 on C_T ($K_1 = 2$, $\lambda_m = (1/10)\lambda_{p1}$, $V_m = 1/\lambda_m^2$, $V_{p2} = 1/\lambda_{p2}^2$, $(U/V) = 4$, and $\alpha = 0.7$).

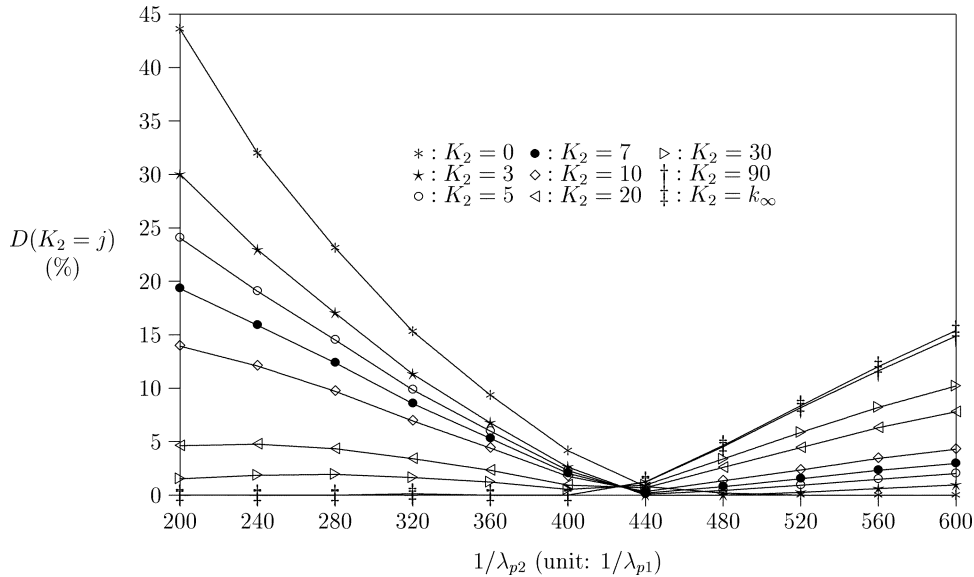


Fig. 10. Discrepancy $D(K_2 = j)$ of C_T ($K_1 = 2$, $\lambda_m = (1/10)\lambda_{p1}$, $V_m = 1/\lambda_m^2$, $V_{p2} = 1/\lambda_{p2}^2$, $(U/V) = 4$, and $\alpha = 0.7$).

large number of cell crossings N_c , which increases C_T . We observe that when $N_c > K_1$, the numbers of URA and RA crossings do not increase as quickly as N_c does. The result is that the negative effect of short cell-residence times are not as significant as the positive effect of long cell-residence times. Therefore, the combined effect is that C_T decreases as V_m increases. Fig. 11 also indicates that C_T is not sensitive to the change of V_m when $V_m \leq (1/\lambda_m^2)$.

Effects of variance V_{p2} . Fig. 12 plots C_T as a function of V_{p2} , where V_{p2} is the variance of the intersession idle periods t_{p2} . This figure shows that C_T decreases as V_{p2} increases. When V_{p2} increases, more long and short t_{p2} periods are observed. For a short t_{p2} , few cell movements occur during this period (i.e., N_c is small) and a small C_T is expected. On the other hand, for a long t_{p2} , many cell movements occur during this period (i.e., N_c is large) and a large C_T is expected. Nevertheless, when $N_c >$

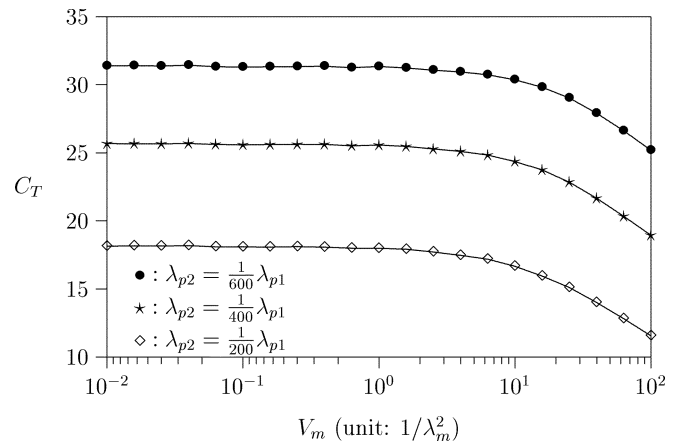


Fig. 11. Effects of V_m on C_T ($K_1 = 2$, $K_2 = 5$, $\lambda_m = (1/10)\lambda_{p1}$, $V_{p2} = 1/\lambda_{p2}^2$, $(U/V) = 4$, and $\alpha = 0.7$).

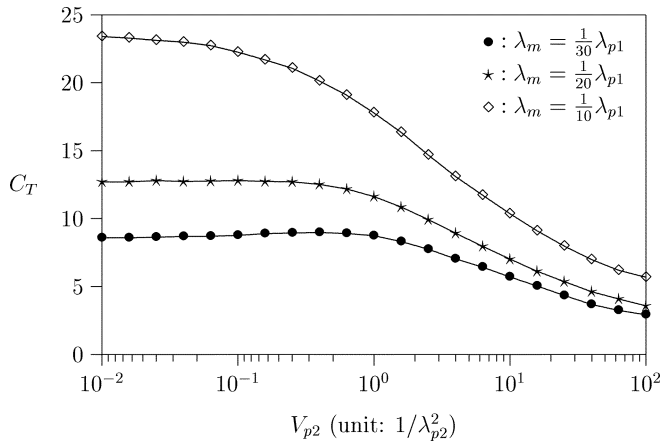


Fig. 12. Effects of V_{p2} on C_T ($K_1 = 2$, $K_2 = 5$, $\lambda_{p2} = (1/200)\lambda_{p1}$, $V_m = (1/\lambda_m^2)$, $(U/V) = 4$, and $\alpha = 0.7$).

K_1 , the numbers of URA and RA crossings do not increase as rapidly as N_c does. Consequently, the net effect is that C_T decreases as V_{p2} increases.

IV. CONCLUSION

This paper investigated the location-management strategy for UMTS PS service domain. When an MS is not in any communication session, the system tracks the RA where the MS resides. Within a communication session, the MS is tracked at the cell level during packet transmission. In the idle period of an ongoing session, the MS is tracked at the URA level to avoid frequent cell updates while still keeping the radio connection. The inactivity counter mechanism was proposed in 3GPP 25.331 [1] to determine when to switch between the three location-tracking modes (cell, URA, or RA). In this mechanism, two inactivity counters are used to count the numbers of cell and URA updates in an idle period between two packet transmissions. If the number of cell updates reaches a threshold K_1 , the MS is switched from the cell tracking to the URA tracking. After that, if the number of URA updates reaches a threshold K_2 , the MS is tracked at the RA level. We utilized analytical and simulation models to investigate the performance of the inactivity counter mechanism. It is clear that as K_1 and K_2 increase, the location-update cost increases while the paging cost decreases. There exists optimal K_1 and K_2 that minimize the net cost C_T of location update and paging. For the input parameters considered in this paper, the lowest C_T costs are observed when $K_1 = 1$ or 2. If the intersession idle periods t_{p2} are long, C_T is an increasing function of K_2 . On the other hand, if the t_{p2} are short, C_T is a decreasing function of K_2 . We quantitatively showed how C_T increases as intersession idle periods and user mobility increase. In addition, both the variances V_m of cell-residence times and V_{p2} of intersession idle periods affect C_T . Our study indicated that as V_m and V_{p2} increase, C_T decreases.

APPENDIX PROOF FOR THEOREM 1

Consider an idle period where no packet is delivered. Let N_c be the number of cell crossings in this idle period. In the inactivity counter algorithm, let $C_T(K_1, K_2)$ be the net cost

of location update and paging in the idle period with cell-update threshold K_1 and URA-update threshold K_2 . Let $K^* = (K_1^*, K_2^*)$ be the optimal threshold value pair that minimizes the net cost C_T . Define k_∞ as a number larger than the N_c value in the idle period. Let k_{any} be an arbitrary integer number.

Theorem 1: For an idle period, $K^* = (k_\infty, k_{\text{any}})$, $(1, k_\infty)$ or $(1, 0)$.

Proof: Let $N_{ur}(j)$ and $N_r(j)$ be the numbers of URA and RA crossings between the $j + 1$ st cell crossing and the N_c th cell crossing, respectively. It is clear that $j < N_c$. We consider three cases.

Case I: $N_c < K_1$. In this case, there are no URA and RA updates during the idle period and the net cost C_T is not affected by the value of K_2 . Therefore, for an arbitrary K_2 value k_{any} , we have

$$\begin{aligned} C_T(N_c + 1, k_{\text{any}}) &= C_T(N_c + 2, k_{\text{any}}) \\ &= \dots \\ &= C_T(k_\infty, k_{\text{any}}). \end{aligned} \quad (19)$$

Case II: $N_c \geq K_1$ and $N_{ur}(K_1) < K_2$. In this case, there are no RA updates during the idle period. The net cost C_T is

$$C_T(K_1, K_2) = U[K_1 + N_{ur}(K_1)] + S_{\text{URA}}V \quad (20)$$

where S_{URA} is the number of cells in an URA [e.g., $S(n)$ in (5) for an n -layer URA layout]. Note that the number of URA crossings between the second and the K_1 th cell crossing is $N_{ur}(1) - N_{ur}(K_1)$. Since the number of URA crossings is no more than the number of cell crossings in the same time period, we have

$$\begin{aligned} N_{ur}(1) - N_{ur}(K_1) &\leq K_1 - 1 \text{ or } K_1 + N_{ur}(K_1) \\ &\geq 1 + N_{ur}(1). \end{aligned} \quad (21)$$

From (20) and (21), we have

$$\begin{aligned} C_T(K_1, K_2) &\geq U[1 + N_{ur}(1)] + S_{\text{URA}}V \\ &= C_T(1, N_{ur}(1) + 1) \\ &= C_T(1, N_{ur}(1) + 2) \\ &= \dots \\ &= C_T(1, k_\infty). \end{aligned} \quad (22)$$

Case III: $N_c \geq K_1$ and $N_{ur}(K_1) \geq K_2$. Let the K_2 th URA update correspond to the i th cell crossing in this idle period; then, $N_r(i)$ is the number of RA updates and the net cost is

$$C_T(K_1, K_2) = U[K_1 + K_2 + N_r(i)] + S_{\text{RA}}V \quad (23)$$

where S_{RA} is the number of cells in an RA. The number of RA crossings between the second and the i th cell crossing is $N_r(1) - N_r(i)$. Similar to the derivation of (21)

$$\begin{aligned} N_r(1) - N_r(i) &\leq K_1 - 1 + K_2 \text{ or } K_1 + K_2 + N_r(i) \\ &\geq 1 + N_r(1). \end{aligned} \quad (24)$$

From (23) and (24), we have

$$C_T(K_1, K_2) \geq U[1 + N_r(1)] + S_{\text{RA}}V = C_T(1, 0). \quad (25)$$

From (19), (22), and (25), for all K_1 and K_2 values, we have

$$C_T(K_1, K_2) \geq \min [C_T(k_\infty, k_{\text{any}}), C_T(1, k_\infty), C_T(1, 0)].$$

In other words, $K^* = (k_\infty, k_{\text{any}}), (1, k_\infty),$ or $(1, 0)$.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers. Their comments have significantly improved the quality of this paper.

REFERENCES

[1] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; RRC Protocol Specification for Release 1999, 2000. Technical Specification 3G TS 25.331 version 3.5.0 (2000-12).

[2] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2, 2000. Technical Specification 3G TS 23.060 version 3.6.0 (2001-01).

[3] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J. Chen, "A new random walk model for PCS networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1254–1260, July 2000.

[4] M. Cheng and L. F. Chang, "Wireless dynamic channel assignment performance under packet data traffic," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1257–1269, July 1999.

[5] I. Chlamtac, Y. Fang, and H. Zeng, "Call blocking analysis for PCS networks under general cell residence time," in *IEEE Wireless Communications and Networking Conf. (WCNC)*, New Orleans, LA, Sept. 1999.

[6] ETSI, "UMTS Terrestrial Radio Access (UTRA)," Concept Evaluation, Version 3.0.0., Tech. Rep. UMTS 30.06, Dec. 1997.

[7] Y. Fang, "Hyper-Erlang distribution model and its application in wireless mobile networks," *ACM-Baltzer J. Wireless Networks*, vol. 7, no. 3, pp. 211–219, 2001.

[8] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Trans. Commun.*, vol. 47, pp. 1062–1072, July 1999.

[9] Y. Fang, I. Chlamtac, and H.-B. Fei, "Analytical results for optimal choice of location update interval for mobility database failure restoration in PCS networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 11, pp. 615–624, June 2000.

[10] H. Holma and A. Toskala, *WCDMA for UMTS*. New York: Wiley, 2000.

[11] F. P. Kelly, *Reversibility and Stochastic Networks*. New York: Wiley, 1979.

[12] Y.-B. Lin and S.-R. Yang, "A mobility management strategy for GPRS," *IEEE Trans. Wireless Commun.*, vol. 2, Nov.

[13] Y.-B. Lin, W.-R. Lai, and R.-J. Chen, "Performance analysis for dual band PCS networks," *IEEE Transactions Comput.*, vol. 49, pp. 148–159, Feb. 2000.

[14] U. Manber, *Introduction to Algorithms: A Creative Approach*. Reading, MA: Addison-Wesley, 1989.

[15] R. Nelson, *Probability, Stochastic Processes, and Queueing Theory*. New York: Springer-Verlag, 1995.

[16] Y.-C. Tseng and W.-N. Hung, "An improved cell type classification for random walk modeling in cellular networks," *IEEE Commun. Lett.*, vol. 5, pp. 337–339, Aug. 2001.

[17] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level," *IEEE/ACM Trans. Networking*, vol. 5, pp. 71–86, Feb. 1997.



Shun-Ren Yang received the B.S. and M.S. degrees in computer science and information engineering from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1998 and 1999, respectively. He is currently pursuing the Ph.D. degree at NCTU.

His research interests include the design and analysis of personal communications services networks, computer telephony integration, mobile computing, and performance modeling.



Yi-Bing Lin (M'95–SM'95–F'03) received the B.S.E.E. degree from National Cheng Kung University, Tainan, Taiwan, R.O.C., in 1983 and the Ph.D. degree in computer science from the University of Washington, Seattle, in 1990.

From 1990 to 1995, he was with the Applied Research Area, Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed Professor, Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C. In 1996, he was appointed Deputy Director of the Microelectronics and Information Systems Research Center at NCTU. From 1997 to 1999, he was elected Chairman of CSIE at NCTU. He has published over 150 journal articles and more than 200 conference papers. He is the author of the book *Wireless and Mobile Network Architecture* (coauthored with Imrich Chlamtac; Wiley: New York, 2001). His current research interests include the design and analysis of personal communication-services networks, mobile computing, distributed simulation, and performance modeling. He serves as a consultant to many telecommunications companies, including FarEasTone and Chung Hwa Telecom.

Dr. Lin is a Senior Technical Editor of IEEE NETWORK. He is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE PERSONAL COMMUNICATIONS MAGAZINE, *Computer Networks*, *International Journal of Communications Systems*, *ACM/Baltzer Wireless Networks*, *Computer Simulation Modeling and Analysis*, and *Journal of Information Science and Engineering*. He is an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and IEEE COMMUNICATIONS SURVEY AND TUTORIALS. He is an Area Editor of *ACM Mobile Computing and Communication Review* and a Columnist for *ACM Simulation Digest*. He has been a Guest Editor for the *ACM/Baltzer MONET* special issue on personal communications, the IEEE TRANSACTIONS ON COMPUTERS special issue on mobile computing, the IEEE TRANSACTIONS ON COMPUTERS special issue on wireless internet, and the IEEE COMMUNICATIONS MAGAZINE special issue on active, programmable, and mobile code networking. He served as Program Chair for the Eighth Workshop on Distributed and Parallel Simulation and for the Second International Mobile Computing Conference. He also served as General Chair for the Ninth Workshop on Distributed and Parallel Simulation. He received the 1998, 2000, and 2002 Outstanding Research Awards from the National Science Council, R.O.C., and the 1998 Outstanding Youth Electrical Engineer Award from the Chinese Institute of Electrical Engineering (CIEE), R.O.C. He also received the NCTU Outstanding Teaching Award in 2002. He is an Adjunct Research Fellow of Academia Sinica and is Chair Professor at Providence University, Taichung, Taiwan, R.O.C.