

A Mobility Management Strategy for GPRS

Yi-Bing Lin* and Shun-Ren Yang

Abstract

In *General Packet Radio Service* (GPRS), a mobile station (MS) is tracked at the cell level during packet transmission, and is tracked at the routing area (RA) level when no packet is delivered. A READY Timer (RT) mechanism was proposed in 3GPP 23.060 to determine when to switch from cell tracking to RA tracking. In this mechanism, a threshold interval T is defined. If no packet is delivered within T , the MS is tracked at the RA level. When a packet arrives, the MS is tracked at the cell level again. However, the RT mechanism has a major fallacy that the READY timers in both the MS and the SGSN may lose synchronization. This paper considers another mechanism called READY Counter (RC) to resolve this problem. In this approach, a threshold K is used. Like the RT approach, the MS is tracked at the cell level during packet transmission. If no packets are delivered after the MS has made K cell crossings, the MS is tracked at the RA level. We also devise an adaptive algorithm called dynamic RC (DRC). This algorithm dynamically adjusts the K value to reduce the location update and paging costs. We propose analytic and simulation models to investigate RC, RT, and DRC. Our study indicates that RC may outperform RT. We also show that DRC nicely captures the traffic/mobility patterns, and always adjusts the K threshold close to the optimal values.

Keywords: GPRS, mobile network, mobility management, wireless data.

*Corresponding Author: Yi-Bing Lin, Dept. Comp. Sci. & Info. Engr., National Chiao Tung University, Hsinchu, Taiwan, R.O.C.; Fax:+886-3.5724176; Email: liny@csie.nctu.edu.tw

1 Introduction

General Packet Radio Service (GPRS) provides packet switched data services for existing mobile telecommunication networks such as GSM and Digital AMPS [10]. GPRS core network has also evolved into 3G network (i.e., UMTS [11]). Most GSM-based mobile operators are deploying GPRS for wireless Internet services. The network architecture of GSM/GPRS is shown in Figure 1. In this figure, the dashed lines represent signaling links, and the solid lines represent data and signaling links. The core network consists of two service domains, a *circuit-switched* (CS) service domain (i.e., PSTN/ISDN) and a *packet-switched* (PS) service domain (i.e., IP). GPRS is evolved from GSM by introducing two new core network nodes: *Serving GPRS Support Node* (SGSN) and *Gateway GPRS Support Node* (GGSN). Existing GSM nodes including *Base Station Subsystem* (BSS), *Visitor Location Register* (VLR), and *Home Location Register* (HLR) are upgraded. GPRS BSS consists of *Base Transceiver Stations* (BTSs) and *Base Station Controller* (BSC) where the BSC is connected to the SGSN through frame relay link. The BTS communicates with the MS through the radio interface *Um* based on the TDMA technology.

The cells (i.e., radio coverages of BTSs) in a GPRS 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 link. Location change of an MS is detected as follows. The cells broadcast their cell identities. The MS periodically 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 CS domain, cells are partitioned into *location areas* (LAs). The LA of an MS is tracked by the VLR. In the PS domain, the cells are partitioned into *routing areas* (RAs). An RA is typically a subset of an LA. The RA of an MS is tracked by the SGSN. The SGSN also tracks the cell of an MS when packets are delivered between the MS and the SGSN.

In GPRS, the mobility management activities for an MS are characterized by a Mobility Management (MM) finite state machine exercised in both the SGSN and the MS. There are three states in the machine. In the **IDLE** state, the MS is not known (i.e., not attached) to GPRS. In the

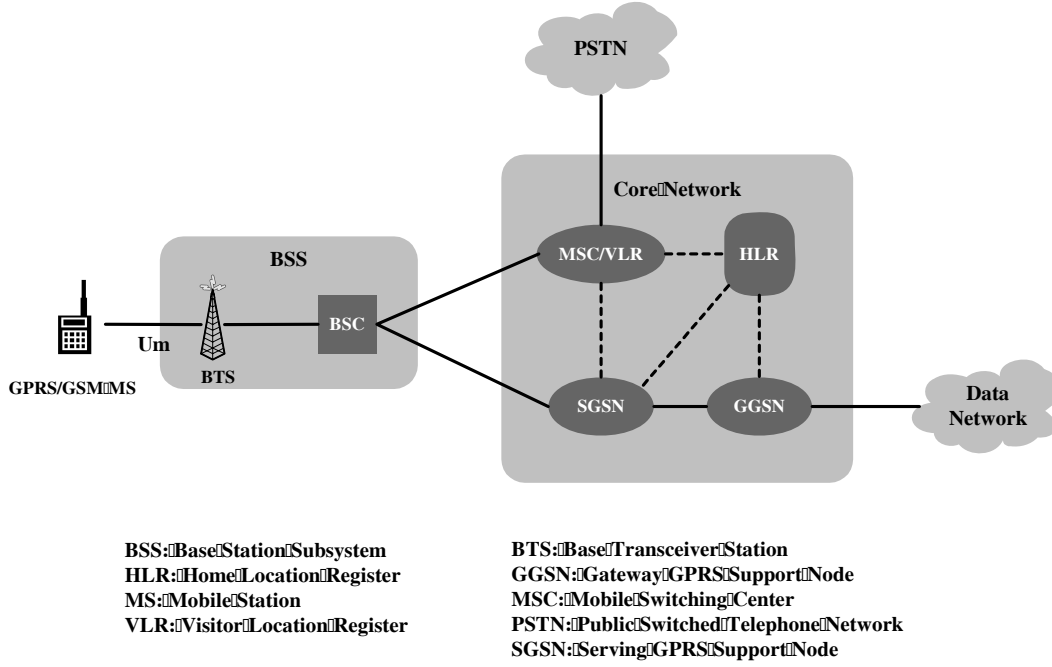


Figure 1: The network architecture of GSM/GPRS

STANDBY state, the MS is attached to GPRS and the MS is tracked by the SGSN at the RA level. In the **READY** state, the SGSN tracks the MS at the cell level. Packet data units can only be delivered in this state. Descriptions of transitions among the MM states can be found in [11] and are briefly described as follows.

T1. IDLE → READY. This transition is triggered by an MS when the MS performs GPRS attach.

T2. READY → IDLE. This transition is triggered by the MS or the SGSN when the MS is detached from the GPRS network.

T3. STANDBY → READY. This transition occurs when the MS sends a packet data unit to the SGSN, possibly in response to a page from the SGSN.

T4. READY → STANDBY. This transition is triggered by either the SGSN or the MS. In GPRS, a READY timer is maintained in the MS and the SGSN. If no packet data unit is transmitted before the timer expires, then this MM transition occurs. The length of the READY timer can only be changed by the SGSN. The MS is informed of the READY timer value change

through messages such as **Attach Accept** and **Routing Area Update Accept**. This MM transition may also occur when the SGSN forces to do so, or when abnormal condition is detected during radio transmission.

T5. STANDBY → IDLE. This transition is triggered by the SGSN when tracking of MS is lost. This transition may also be triggered by SGSN when the SGSN receives a **Cancel Location** message from the HLR, which implies that the MS has moved to the service area of another SGSN.

Transition **T4** merits further discussion. In the **READY** state, the MS expects to receive packets in short intervals. Therefore when the MS moves to a new cell, it should inform the SGSN of the movement immediately. In this way, the SGSN can deliver the next packet to the destination cell without paging the whole RA. On the other hand, if the communication session between the MS and the SGSN completes, the SGSN may not send the next packet (the first packet of the next session) to the MS in a long period. In this case, tracking the MS at the cell level is too expensive. Thus the MM state should be switched to **STANDBY** and the MS is tracked at the RA level. To conclude, in the **READY** state, no paging is required (the packets are sent directly to the MS) while the location update cost is high (location update is performed for every cell movement). In the **STANDBY** state, the paging cost is high (all cells in the RA are paged) while the location update cost is low (location update is performed for every RA movement). The **T4** transition can be implemented by two approaches. In the *READY Timer (RT)* approach [1], an RT threshold T is defined. At the end of a packet transmission, the RT timer is set to the T value, and is decremented as time elapses. Transition **T4** occurs if the MS does not receive the next packet before the RT timer expires. However, the RT approach has a major fallacy that the RT timers in both the MS and the SGSN may lose synchronization (i.e., when the SGSN moves to **STANDBY**, the MS may be still in **READY**). To resolve this problem, we consider the *READY Counter (RC)* approach. In the RC approach, an RC counter counts the number of cell movements in the packet idle period between two packet transmissions to an MS. If the number of movements reaches a threshold K , then the MM state switches from **READY** to **STANDBY**. To capture user mobility more accurately, one may dynamically adjust the K value to further reduce the net cost of cell/RA updates and paging.

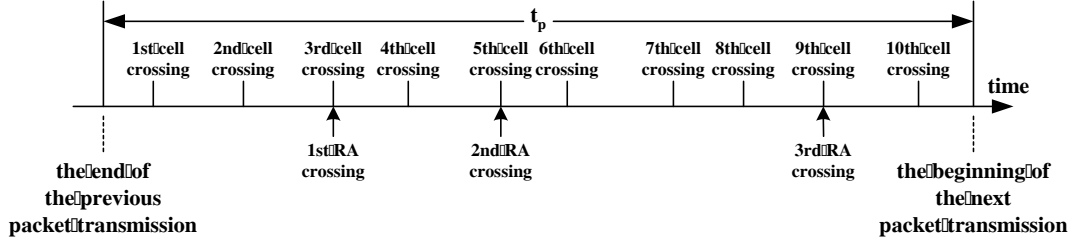


Figure 2: Cell and RA crossings in an idle period

Consider the interval t_p between the end of a packet transmission and the beginning of the next packet transmission. If we know the number of cell crossings in t_p and the distribution of RA crossings among these cell crossings, then we can find the optimal K value such that the net cost is minimized. Let N_c be the number of cell crossings during t_p . Let $N_r(j)$ be the number of RA crossings occurring between the $j + 1$ -st cell crossing and the N_c -th cell crossing, where $j < N_c$. By convention, $N_r(K) = 0$ for $K \geq N_c$. Figure 2 illustrates the cell and RA crossings in an idle period. In this example, $N_c = 10$. If $K = 2$, then $N_r(2) = 3$. If $K = 4$, then $N_r(4) = 2$. Let U be the cost for a cell/RA update, and V be the cost for paging in a cell. Let S be the number of cells in an RA. Consider the RC algorithm with threshold K . The net cost $C_T(K)$ in an idle period can be expressed as

$$C_T(K) = \begin{cases} UN_c, & \text{for } K > N_c \\ U[K + N_r(K)] + SV, & \text{for } K \leq N_c \end{cases} \quad (1)$$

In the following theorem we show how to find the optimal threshold value K^* for RC in an idle period such that the net cost is minimized.

Theorem 1. Consider an idle period where no packet is delivered. Let N_c be the number of cell crossings in this period. In the RC algorithm, let K^* be the optimal threshold value that minimizes the net cost $C_T^* = C_T(K^*)$ in the idle period. Then $K^* = 0$ or $K^* = N_c + 1$.

Proof. As previously defined, $N_r(K)$ is the number of RA updates in RC with threshold K . If $N_c \geq K$, then

$$C_T(K) = U[K + N_r(K)] + SV \quad (2)$$

The number of RA crossings within the first K cell crossings is $N_r(0) - N_r(K)$. It is clear

that

$$N_r(0) - N_r(K) \leq K \quad (3)$$

From (2) and (3), we have

$$C_T(K) \geq U[N_r(0)] + SV = C_T(0) \quad \text{for } K \leq N_c \quad (4)$$

For $K > N_c$, we have $C_T(K) = C_T(N_c + 1)$. Therefore,

$$C_T^* = \min_{0 \leq K \leq \infty} C_T(K) = \min_{0 \leq K \leq N_c + 1} C_T(K) \quad (5)$$

From (4) and (5), we have

$$C_T^* = \min [C_T(0), C_T(N_c + 1)] \quad (6)$$

In other words, $K^* = 0$ or $K^* = N_c + 1$.

Based on Theorem 1, we devise an algorithm to select K as follows. Let $t_p(i)$ be the interval between the end of the $i - 1$ st packet transmission and the beginning of the i th packet transmission. Let $K(i)$ be the optimal K value for $t_p(i)$. The K value can be dynamically adjusted using the following algorithm.

Dynamic READY Counter (DRC) Algorithm

Initialization. Assign an arbitrary value to $K(0)$. Exercise the RC approach with threshold $K(0)$ before the first packet arrives.

When the i th packet transmission is completed. Compute the optimal $K(i)$ that minimizes the net cost for the period $t_p(i)$. Based on Theorem 1, $K(i)$ is either 0 or $N_c + 1$ in $t_p(i)$, which can be quickly computed with very low cost. Exercise the RC approach with threshold $\bar{K}$ during $t_p(i + 1)$, where

$$\bar{K} = \begin{cases} \left\lceil \sum_{j=i-M+1}^i \frac{K(j)}{M} \right\rceil, & \text{for } i \geq M \\ \left\lceil \sum_{j=1}^i \frac{K(j)}{i} \right\rceil, & \text{for } i < M \end{cases} \quad (7)$$

In other words, the threshold $\bar{K}$ between the i th and the $i + 1$ st packet transmissions is selected as the average of the previous M optimal K values. Simulation experiments show that $M \geq 5$ is appropriate when using (7) to compute $\bar{K}$. In this paper, we consider $M = 25$.

Since the MS has the complete information on the numbers of cell and RA crossings in t_p , the dynamic RC (DRC) algorithm is implemented in MS. In DRC, the state transition of SGSN (i.e., from **READY** to **STANDBY**) is triggered by the MS. When the MS crosses the $\bar{K}$ -th cell boundary during t_p , it sends an RA update message to the SGSN instead of the cell update message. Once the SGSN receives the first RA update message from the MS, it switches the MM state from **READY** to **STANDBY**. No extra message is introduced to switch the MM states in DRC. Thus, the network signaling cost is the same as that of the static RC mechanism. The only cost incurred by DRC is the computation of $\bar{K}$ in the MS. As shown in (7), the computation can be done in micro seconds. This cost is not significant and can be ignored.

This paper proposes analytic and simulation models to study the RT, RC, and DRC approaches. We investigate how the RA size and the location update/paging costs affect the performance of these approaches. Specifically, we show that RC is better than RT, and that DRC can automatically adjust the K value to minimize the net cost. The notation used in this paper is listed in Appendix A.

2 Analytic Model for READY Counter Mechanism

This section develops an analytic model to study the GPRS READY Counter (RC) mechanism. We first describe a uniform random walk model for user movement. Then we show how to use this model to investigate the GPRS RC performance. For a specific threshold K , we derive the number of cell updates and RA updates between the end of a packet transmission and the beginning of the next packet transmission. Our model considers a GPRS network with hexagonal cell layout. Such a layout with small cells is reasonable for mobile services in big cities. In this configuration, the cells are grouped into several n -layer routing areas (RAs). Every RA covers $S(n) = 3n^2 - 3n + 1$ cells as shown in Figure 3 (where $n = 4$). The figure plots seven RAs (A, B, C, D, E, F and G) and

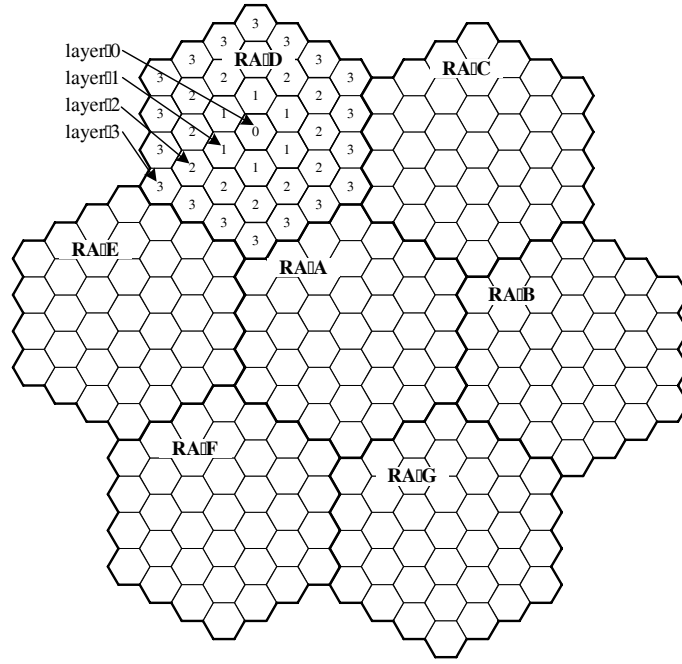


Figure 3: Cell/routing area layout in a GPRS network

the cells within the RAs. The cell at the center of an RA is referred to as the *layer 0* cell. The cells surrounding layer $x-1$ cells are referred to as *layer x* cells. There are $6x$ cells in layer x except for layer 0 which contains exactly one cell. An n -layer RA consists of cells from layer 0 to layer $n-1$. Based on this RA/cell structure, we derive the number of cell crossings before a user crosses an RA boundary.

Based on the equal routing probability assumption (i.e., the MS moves to each of the neighboring cells with probability $1/6$), we classify the cells in an RA into several cell types [2]. 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 $y + 1$ st 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 lines 1, 2, and 3 in Figure 4) on the hexagonal RA. According to the type classification algorithm in [2], Figure 4 illustrates the types of cells for a 4-layer RA. We develop a random walk model to compute when an MS crosses the boundary of an n -layer RA. A state of this random walk is of the form (x, y) . For $0 \leq x < n$ and $0 \leq y \leq x - 1$, the state (x, y) is transient, which represents that the MS is in

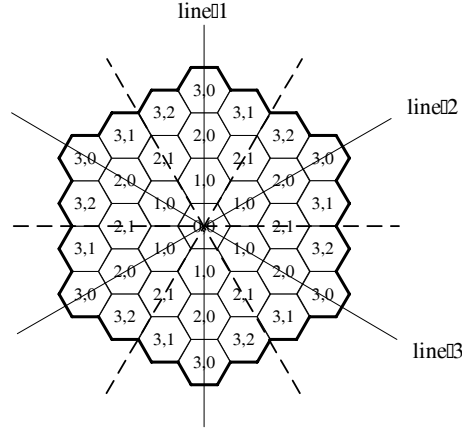


Figure 4: Type classification for a 4-layer routing area

one of the cells of type $\langle x, y \rangle$. For $x = n$ and $0 \leq y < n - 1$, the state (n, y) is absorbing, which represents that the MS crosses the boundary of the RA from a cell of type $\langle n - 1, y \rangle$. The total number of states for an n -layer RA random walk is equal to $A(n) = n(n + 1)/2$. 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. From [2], the transition probability matrix $P = (p_{(x,y),(x',y')})$ of the random walk is given below.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1/6 & 1/3 & 1/6 & 1/3 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/3 & 1/6 & \cdots & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}_{A(n) \times A(n)} \quad (8)$$

Note that the number of states for the random walk can be further reduced if we consider the symmetry along the dashed lines in Figure 4. The details will not be presented in this paper.

Let $p_{(x,y),(x',y')}^{(k)}$ be the probability that the random walk moves from state (x, y) to state (x', y') with exact k steps. Let $p_{k,(x,y),(n,j)}$ be the probability that an MS initially resides at a $\langle x, y \rangle$ cell, moves into a $\langle n - 1, j \rangle$ cell at the $k - 1$ st step, and then moves out of the RA at the k th step. Then

$$p_{k,(x,y),(n,j)} = \begin{cases} p_{(x,y),(n,j)}, & \text{for } k = 1 \\ p_{(x,y),(n,j)}^{(k)} - p_{(x,y),(n,j)}^{(k-1)}, & \text{for } k > 1 \end{cases} \quad (9)$$

Equation (9) can be solved using the transition probability matrix (8).

Suppose that an MS is in any cell of an RA with equal probability. In other words, the MS is in cell $\langle 0, 0 \rangle$ with probability $1/S(n)$, and is in a cell of type $\langle x, y \rangle$ ($0 < x < n, 0 \leq y \leq x - 1$) with probability $6/S(n)$, where $S(n) = 3n^2 - 3n + 1$ is the number of cells covered by an n -layer RA. Consider the example where a 4-layer RA covers $S(4) = 37$ cells (see Figure 4). Since there are one cell of type $\langle 0, 0 \rangle$ and six cells of type $\langle 2, 1 \rangle$ in a 4-layer RA, the MS is in cell $\langle 0, 0 \rangle$ with probability $1/37$ and in a cell of type $\langle 2, 1 \rangle$ with probability $6/37$. Let $\theta_1(k)$ be the probability that the MS will leave the RA at the k th step. Then

$$\theta_1(k) = \left[\frac{1}{S(n)} \right] \left[\sum_{j=0}^{n-2} p_{k, \langle 0, 0 \rangle, \langle n, j \rangle} \right] + \left[\frac{6}{S(n)} \right] \left[\sum_{x=1}^{n-1} \sum_{y=0}^{x-1} \sum_{j=0}^{n-2} p_{k, \langle x, y \rangle, \langle n, j \rangle} \right] \quad (10)$$

Let $\theta_2(k)$ be the probability that after an MS enters an RA, it moves out of the RA at the k th step. Probability $\theta_2(k)$ is derived as follows. It can be shown [12] that after entering the RA, the MS is in a boundary cell with probability in proportion to the number of boundary edges for that boundary cell. Under the condition that an MS is moving into a boundary cell, the MS is in a boundary cell of type $\langle n - 1, 0 \rangle$ with probability $3 \cdot 6/B(n)$, and is in a boundary cell of type $\langle n - 1, y \rangle$ ($1 \leq y \leq n - 2$) with probability $2 \cdot 6/B(n)$, where $B(n) = 6[3 + 2(n - 2)]$ is the number of boundary edges in an n -layer RA. In Figure 4, $B(n) = 42$ for a 4-layer RA. In this example, there are three boundary edges for each of the six $\langle 3, 0 \rangle$ cells, and the MS is in a boundary cell of type $\langle 3, 0 \rangle$ with probability $18/42$. Similarly, there are two boundary edges for each of the six $\langle 3, 1 \rangle$ cells, and the MS is in a boundary cell of type $\langle 3, 1 \rangle$ with probability $12/42$. Based on the above discussion, we have

$$\theta_2(k) = \left[\frac{3 \cdot 6}{B(n)} \right] \left[\sum_{j=0}^{n-2} p_{k, \langle n-1, 0 \rangle, \langle n, j \rangle} \right] + \left[\frac{2 \cdot 6}{B(n)} \right] \left[\sum_{y=1}^{n-2} \sum_{j=0}^{n-2} p_{k, \langle n-1, y \rangle, \langle n, j \rangle} \right] \quad (11)$$

Suppose that an MS is in an arbitrary cell of an RA. Let $P(N_m, k)$ be the probability that after N_m cell movements, the MS crosses k RA boundaries. Similarly, consider an MS that is initially at a boundary cell of an RA. Let $P^*(N_m^*, k^*)$ be the probability that after N_m^* cell movements, the MS

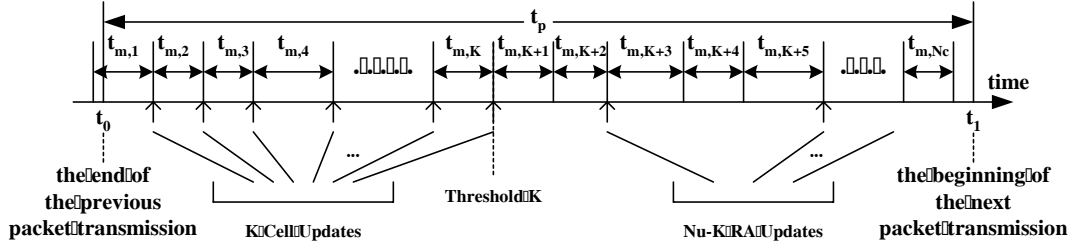


Figure 5: Timing diagram for cell and RA crossings

crosses k^* RA boundaries. From (10) and (11), we have

$$P(N_m, k) = \begin{cases} 1, & \text{for } k = N_m = 0 \\ \sum_{j=N_m+1}^{\infty} \theta_1(j), & \text{for } k = 0, N_m > 0 \\ \sum_{j=1}^{N_m} \theta_1(j) P^*(N_m - j, k - 1), & \text{for } k \geq 1, N_m \geq k \\ 0, & \text{for } N_m < k \end{cases} \quad (12)$$

$$P^*(N_m^*, k^*) = \begin{cases} 1, & \text{for } k^* = N_m^* = 0 \\ \sum_{j=N_m^*+1}^{\infty} \theta_2(j), & \text{for } k^* = 0, N_m^* > 0 \\ \sum_{j=1}^{N_m^*} \theta_2(j) P^*(N_m^* - j, k^* - 1), & \text{for } k^* \geq 1, N_m^* \geq k^* \\ 0, & \text{for } N_m^* < k^* \end{cases} \quad (13)$$

Equations (12) and (13) can be effectively computed using dynamic programming technique [13]. Note that the above derivations are based on the uniform movement assumption. To accommodate non-uniform random walk, we need to modify the state diagram and the transition probability matrix. With (12) and (13), we derive the number of cell/RA updates between the end of a packet transmission and the beginning of the next packet transmission as follows. Figure 5 shows the timing diagram of the activities for an MS. 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 $t_p = t_1 - t_0$, which has a general distribution with density function $f_p(t_p)$, expected value $1/\lambda_p$ and Laplace Transform

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

For RC with a specific threshold K , let N_u be the number of location updates (cell updates plus RA updates) during the period t_p . Based on the random walk model mentioned above, the distribution

of N_u can be derived as follows. Suppose that the cell residence time $t_{m,j}$ at cell C_j has an Erlang distribution with mean $1/\lambda_m = m/\lambda$, variance $V_m = m/\lambda^2$, and density function

$$f_m(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^{m-1}}{(m-1)!} \quad \text{for } t \geq 0 \quad (14)$$

where $m = 1, 2, 3, \dots$. We select Erlang distribution because this distribution can be easily extended into hyper-Erlang distribution. Hyper-Erlang distribution has been proven as a good approximation to many other distributions as well as measured data [5, 9].

The probability mass function of the number of cell crossings N_c within t_p is

$$\begin{aligned} \Pr[N_c = k] &= \int_{t_p=0}^{\infty} \frac{e^{-\lambda t_p}}{m} \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j)(\lambda t_p)^j}{j!} \right] \right. \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m)(\lambda t_p)^j}{j!} \right] \right\} f_p(t_p) dt_p \\ &= \frac{1}{m} \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j) \lambda^j}{j!} \right] \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \right. \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m) \lambda^j}{j!} \right] \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \right\} \\ &= \frac{1}{m} \left\{ \sum_{j=km}^{km+m-1} \left[\frac{(km+m-j) (-\lambda)^j}{j!} \right] \left[\frac{d^j f_p^*(s)}{ds^j} \right] \Big|_{s=\lambda} \right. \\ &\quad \left. - \sum_{j=km-m}^{km-1} \left[\frac{(j-km+m) (-\lambda)^j}{j!} \right] \left[\frac{d^j f_p^*(s)}{ds^j} \right] \Big|_{s=\lambda} \right\}, \quad k = 1, 2, \dots \\ \Pr[N_c = 0] &= \int_{t_p=0}^{\infty} \frac{e^{-\lambda t_p}}{m} \sum_{j=0}^{m-1} \left[\frac{(m-j)(\lambda t_p)^j}{j!} \right] f_p(t_p) dt_p \\ &= \frac{1}{m} \left\{ \sum_{j=0}^{m-1} \left[\frac{(m-j) \lambda^j}{j!} \right] \left[\int_{t_p=0}^{\infty} t_p^j f_p(t_p) e^{-\lambda t_p} dt_p \right] \right\} \\ &= \frac{1}{m} \left\{ \sum_{j=0}^{m-1} \left[\frac{(m-j) (-\lambda)^j}{j!} \right] \left[\frac{d^j f_p^*(s)}{ds^j} \right] \Big|_{s=\lambda} \right\} \end{aligned} \quad (15)$$

Within the time interval t_p , the MS performs cell updates for the first K cell crossings and then performs RA updates whenever it crosses RA boundaries. From (12), (13) and (15), the probability mass function for N_u is

$$\Pr[N_u = j] = \begin{cases} \Pr[N_c = j], & \text{for } j < K \\ \sum_{k=j}^{\infty} \Pr[N_c = k] P(k-K, j-K), & \text{for } j \geq K \end{cases} \quad (16)$$

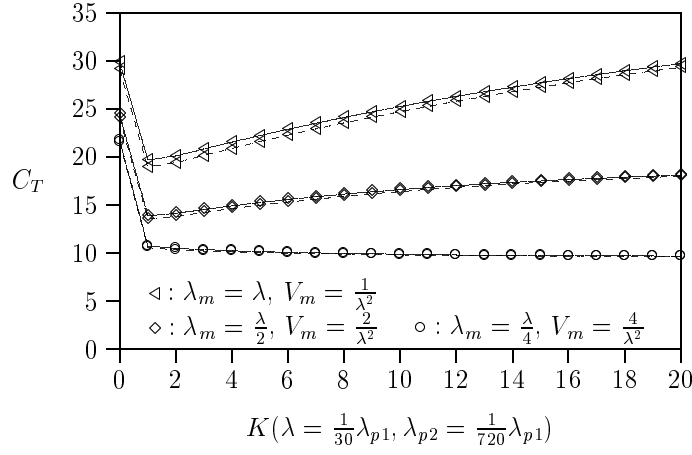


Figure 6: Comparison between the analytic and simulation results

Based on (15) and (16), we derive the expected total cost C_T for location update and paging during t_p . Assume that the cost for performing a location update is U and the cost for paging at one cell is V . Let C_u be the expected location update cost during t_p . From (16), we have

$$C_u = U \sum_{j=0}^{\infty} j \Pr[N_u = j] \quad (17)$$

If the number of cell crossings $N_c < K$, then no cell needs to page the MS. Otherwise, all cells of the RA should page the MS. Let C_v be the expected terminal paging cost during t_p . From (15), we have

$$C_v = VS(n) \sum_{j=K}^{\infty} \Pr[N_c = j] \quad (18)$$

From (17) and (18), the expected total cost C_T for location update and terminal paging during t_p is

$$C_T = C_u + C_v \quad (19)$$

We have developed a discrete simulation model to validate against our analytic analysis. The simulation actually simulates the movement of an MS on the hexagonal plane. The $\theta_1, \theta_2, P(N_m, k), P^*(N_m^*, k^*)$ and C_T values produced by the analytic and simulation models show that both models are consistent. For example, Figure 6 plots the C_T curves for analytic and simulation results based on the 3-layer RA configuration. The t_p intervals are exponentially distributed, and the cell residence times have Erlang distribution given in (14). Other parameters such as λ_{p1} and λ_{p2} will

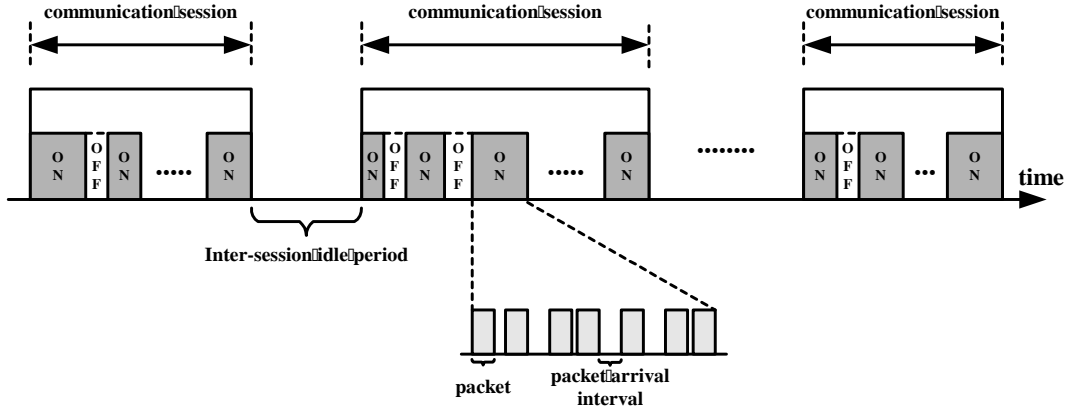


Figure 7: Packet data traffic

be explained in the next section. In this figure, the dashed curves represent the analytic model and the solid curves represent the simulation experiments. The results indicate that the discrepancy between analytic analysis and simulation is within 1% in most cases. The comparisons for various input parameters and t_p , t_m distributions show similar results (i.e., both models are consistent) and will not be elaborated in this paper.

3 Numerical Results

This section investigates the performance of the GPRS RC mechanism. Then based on the simulation experiments, we compare RC with RT, and show that DRC can capture an appropriate K value that reduces the net signaling cost for RC. In this study, we combine the ETSI packet data model [4] with the ON/OFF source model (also known as a packet train model) [7]. As shown in Figure 7, we assume that the packet data traffic consists of communication sessions, where the inter-session idle period has an exponential distribution with mean $1/\lambda_{p2}$. For general inter-session idle periods, the variances of the idle period distributions have similar effects as that of the variances for cell residence times (to be elaborated later in Figure 11), and the details will not be presented in this paper. Within a communication session, packet traffic is modeled by the ON/OFF source model. In an ON-period a burst of data packets are sent. In an OFF-period no packets are sent. 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 (in the same session) with probability α , and is followed by the inter-session idle period (for the next session) with probability $1 - \alpha$. The OFF-periods t_{p1} are drawn from a Pareto distribution [8] with mean $1/\lambda_{p1}$ and infinite variance, which has been found to match very well with the actual data traffic measurements [14]. 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

$$f_p(t_{p1}) = \left(\frac{\beta}{l}\right) \left(\frac{l}{t_{p1}}\right)^{\beta+1} \quad \text{and the expected value is} \quad E[t_{p1}] = \left(\frac{\beta}{\beta - 1}\right) l$$

If β is between 1 and 2, the variance for the distribution becomes infinity. The typical parameter values obtained in [14] are $E[t_{p1}] = 10.5$ seconds and $\beta = 1.2$ for OFF-periods. Our study follows the above $E[t_{p1}]$ and β values. We also assume that the cell residence time is Gamma distributed with mean $1/\lambda_m$ and variance V_m . The Gamma distribution with shape parameter η and scale parameter λ (i.e., mean $1/\lambda_m = \eta/\lambda$ and variance $V_m = \eta/\lambda^2$) has the following density function

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

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 (see Lemma 3.9 in [9]). The Gamma distribution was used to model MS movement in many studies [3, 5, 6], and is used in this paper to investigate the impact of variance for cell residence times. (The impact of V_m is shown in Figure 11. In Figures 8-10 and 12, $V_m = 1/\lambda_m^2$ for cell residence times.) Note that the Erlang distribution used in Figure 6 is a special case of the Gamma distribution where $V_m = 1/(m\lambda_m^2)$. The impacts of several input parameters are discussed as follows.

Effects of $\frac{U}{V}$ ratio. Figure 8 plots C_T as functions of $\frac{U}{V}$ where U is the cost of a cell/RA update, and V is the cost for paging at a cell. The figure indicates that for a large $\frac{U}{V}$ (e.g., $\frac{U}{V} > 20$), the C_T cost for a small RA layout is higher than that for a large RA layout. It is apparent that for the same number of cell crossings within a period, small RA layout has more RA crossings. If a location update operation is expensive (i.e., $\frac{U}{V}$ is large), then a small RA layout will result in large C_T . On the other hand, the RA paging cost increases as the RA size increases.

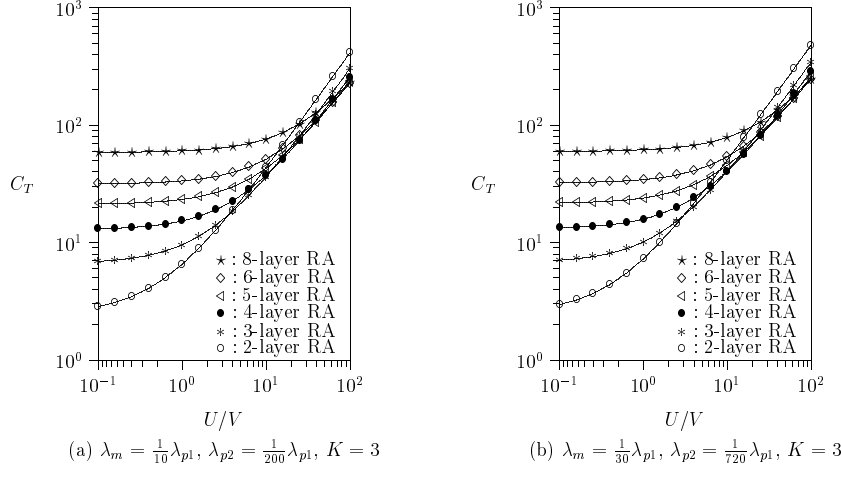


Figure 8: Effects of $\frac{U}{V}$ on C_T ($V_m = \frac{1}{\lambda_m^2}$, $\alpha = 0.6$)

Therefore, when $\frac{U}{V}$ is small (i.e., the paging cost dominates), C_T is an increasing function of the RA size. In the following discussions, we assume that $\frac{U}{V} = 4$. Other $\frac{U}{V}$ ratios show similar results and will not be presented in this paper.

Effects of RA size. Figure 9(a) shows that the 5-layer RA layout has the lowest C_T for mobility rate $\lambda_m = \lambda_{p1}$ (where $\lambda_{p2} = \frac{1}{200}\lambda_{p1}$, $K = 3$ and $\alpha = 0.6$). Both the 2-layer and the 3-layer RA layouts have smaller C_T when λ_m is less than $\frac{1}{10}\lambda_{p1}$. Figure 9(b) shows that the 4-layer RA layout has the lowest C_T when $\lambda_{p2} = \frac{1}{720}\lambda_{p1}$ (where $\lambda_m = \frac{1}{10}\lambda_{p1}$, $K = 3$ and $\alpha = 0.6$). Both the 2-layer and the 3-layer RA layouts have lower C_T when λ_{p2} is larger than $\frac{1}{200}\lambda_{p1}$. Figures 9(c) and (d) show that both the 2-layer and the 3-layer RA layouts have smaller C_T for various K and α values, where $\lambda_m = \frac{1}{10}\lambda_{p1}$ and $\lambda_{p2} = \frac{1}{200}\lambda_{p1}$. For the packet traffic and mobility patterns considered in Figure 9, the 3-layer RA layout is likely to produce lower network operation cost (i.e., C_T) in most cases. The above discussion shows that the C_T curves in Figures 8 and 9 provide guidelines for mobile service operators to deploy optimal cell layout. In the remainder of this paper, we assume 3-layer RA configuration.

Effects of K . Figure 10 shows how K affects C_T . As K increases, the location update cost increases while the paging cost decreases. Therefore, paging and location update are two conflict factors which result in concave C_T curves, and optimal K values exist. This obser-

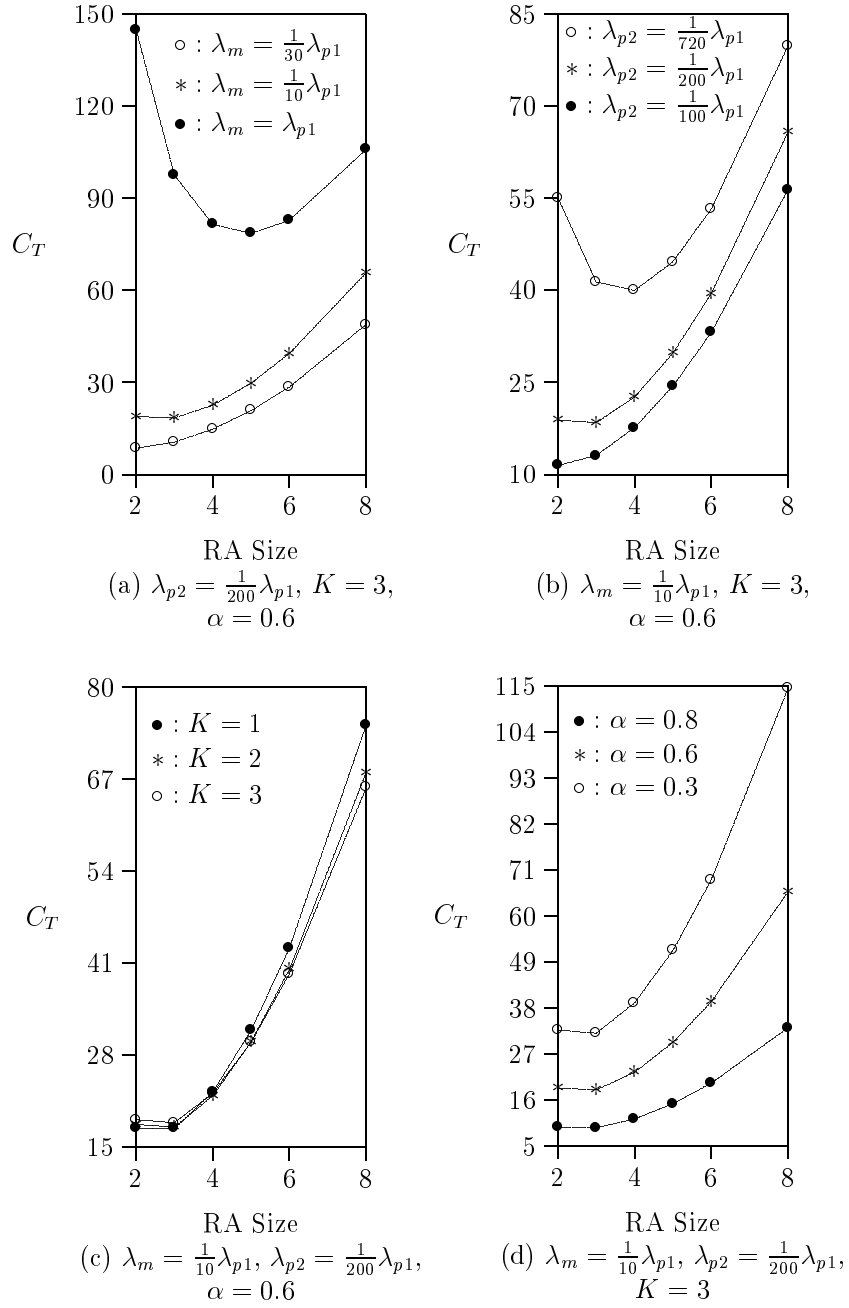


Figure 9: Effects of RA size on C_T ($\frac{U}{V} = 4$, $V_m = \frac{1}{\lambda_m^2}$)

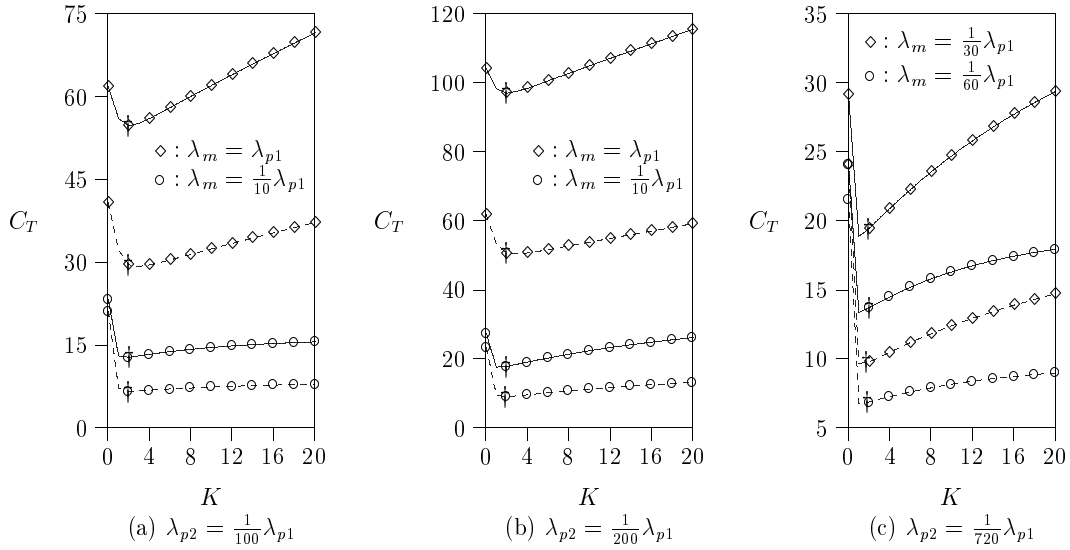


Figure 10: Effects of K , λ_m , λ_{p2} and α on C_T (solid: $\alpha = 0.6$; dashed: $\alpha = 0.8$; $\frac{U}{V} = 4$, $V_m = \frac{1}{\lambda_m^2}$)

variation justifies the motivation to develop a mechanism that dynamically selects appropriate K values when the traffic/mobility patterns change.

Effects of λ_m . As λ_m increases, more cell crossings are expected and C_T increases (see the $\diamond$ and the $\circ$ curves in Figure 10). Note that when K is large and λ_m is small, increasing K only has insignificant effect on C_T . This phenomenon is due to the fact that $C_T(K) = C_T(N_c + 1)$ for $K > N_c$. When λ_m is small and K is very large, it is likely that $K > N_c$, and increasing K only insignificantly increases the C_T value.

Effects of λ_{p2} . Figures 10(a), (b) and (c) plot the C_T curves for $\lambda_{p2} = \frac{1}{100}\lambda_{p1}$, $\frac{1}{200}\lambda_{p1}$, and $\frac{1}{720}\lambda_{p1}$, respectively. The figures show that C_T increases as λ_{p2} decreases. A small λ_{p2} implies large inter-session idle periods, and more cell movements will occur during this period. Thus, a large C_T is expected. As in our previous discussion for the interaction between K and λ_m , if K is sufficiently large, C_T is not sensitive to the change of K when λ_{p2} is large (increasing λ_{p2} has similar effect as decreasing λ_m).

Effects of α . Figure 10 shows the effect of α on C_T (the solid curves are for $\alpha = 0.6$ and the dashed curves are for $\alpha = 0.8$). A smaller α implies more inter-session idle periods. Since

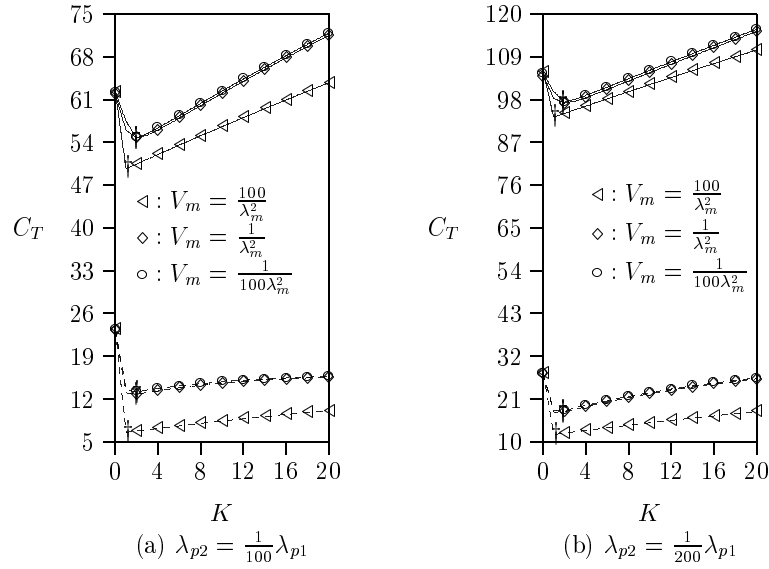


Figure 11: Effects of V_m on C_T (solid: $\lambda_m = \lambda_{p1}$; dashed: $\lambda_m = \frac{1}{10}\lambda_{p1}$; $\frac{U}{V} = 4$, $\alpha = 0.6$)

inter-session idle periods are longer than the OFF-periods in the sessions, C_T increases as α decreases. The figure also indicates that C_T is more sensitive to the mobility rate λ_m for a small α than a large α .

Effects of variance V_m . Figure 11 plots the C_T curves where the cell residence times have a Gamma distribution with mean $1/\lambda_m$ and variance V_m . The figure shows that a large V_m (e.g., $V_m = \frac{100}{\lambda_m^2}$) results in a small C_T . This phenomenon is explained as follows. When V_m increases, more short and long cell residence times are observed. Long cell residence times imply small number of cell crossings in t_p , which result in a small C_T . On the other hand, short cell residence times imply large N_c values (more cell crossings), which increases C_T . However, when $N_c > K$, the number of RA crossings among $N_c - K$ cell crossings does not increase as fast as N_c does. Therefore, the (negative) effect of short cell residence times are not as significant as the (positive) effect of long cell residence times, and the net effect is that C_T decreases as V_m increases. Figure 11 also indicates that C_T is not affected by the change of V_m when $V_m \leq \frac{1}{\lambda_m^2}$.

Comparison of RC and RT. Both RC and RT have similar performance in examples shown in

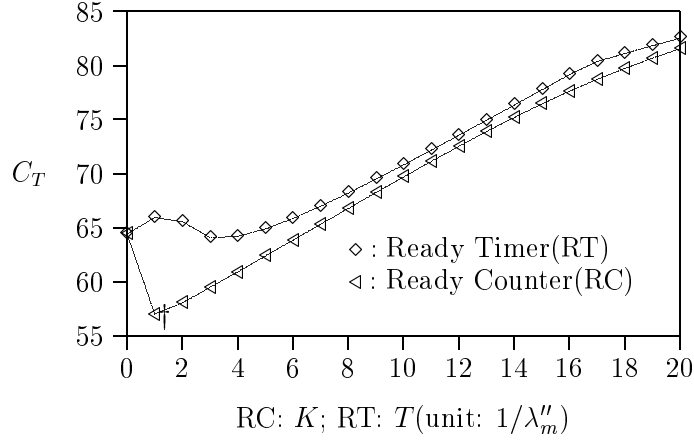


Figure 12: Comparison between RC and RT ($\frac{U}{V} = 4$)

Figures 9-11. However, when the MS mobility rate changes from time to time, RC may significantly outperform RT. The reason is given below. The threshold T of RT is a fixed time interval. When λ_m changes, T cannot adapt to the change. On the other hand, the threshold K of RC always captures the K -th cell movement of an MS no matter how λ_m changes. Consider the example in Figure 12. In this experiment, we mix two types of packet traffic and mobility patterns.

Type I pattern: mean OFF-period time $E[t'_{p1}] = \frac{1}{\lambda'_{p1}} = 100\text{s}$, mobility rate $\lambda'_m = \frac{1}{1000\text{s}}$ and $V'_m = \frac{1}{\lambda_m'^2}$.

Type II pattern: mean OFF-period time $E[t''_{p1}] = \frac{1}{\lambda''_{p1}} = 1000\text{s}$, mobility rate $\lambda''_m = \frac{1}{10\text{s}}$ and $V''_m = \frac{1}{\lambda_m''^2}$.

Consider 1,000,000 OFF-periods. Type I pattern is exercised for the first 500,000 OFF-periods, and Type II pattern is exercised for the remaining 500,000 OFF-periods. Let C'_T be the expected cost of an OFF-period for Type I pattern and C''_T be the expected cost for Type II pattern. In RC, the lowest C'_T is expected when $K = 1$ and the lowest C''_T is expected when $K = 0$. Thus, by choosing $K = 1$, good performance can be expected for both Type I and II patterns. In RT, the best threshold value for C'_T occurs when $T > 100\text{s}$. On the other hand, the best threshold value for C''_T occurs when $T < 10\text{s}$. Figure 12 plots the C_T

curve against K for RC and the C_T curve against T (measured by $1/\lambda_m''$) for RT. The figure indicates that selection of any T value will not satisfy both Type I and II patterns, and the resulting C_T cost for RT is higher than that of RC.

Performance of DRC. The C_T curves in Figures 10-12 are concave with respect to K , which indicate that optimal K values exist to minimize C_T . However, the optimal K values are not the same for different traffic/mobility patterns. Therefore, a mechanism that automatically selects the optimal K values in real time is required. The DRC algorithm proposed in Section 1 serves for this purpose. In Figures 10-12, the “†” points represent the $(C_T, E[\bar{K}])$ pairs of DRC. These points indicate that DRC nicely captures the traffic/mobility patterns and always adjusts the RC threshold close to the optimal values.

4 Conclusions

In GPRS, an MS is tracked at the cell level during packet transmission, and is tracked at the RA level when no packet is delivered. A READY Timer (RT) mechanism was proposed in 3GPP 23.060 [1] to determine when to switch from cell tracking to RA tracking. In this mechanism, a threshold interval T is defined. If no packet is delivered within T , the MS is tracked at the RA level. When a packet arrives, the MS is tracked at the cell level again. However, the RT mechanism has a major fallacy that the READY timers in both the MS and the SGSN may lose synchronization. This paper considered another mechanism called READY Counter (RC) to resolve this problem. In this approach, a threshold K is used. Like the RT approach, the MS is tracked at the cell level during packet transmission. If no packets are delivered after the MS has made K cell crossings, the MS is tracked at the RA level. We also devised an adaptive algorithm called dynamic RC (DRC). This algorithm dynamically adjusts the K value to reduce the location update and paging costs. We proposed analytic and simulation models to investigate RC, RT, and DRC. Our study indicated the following.

- If the location update cost dominates, then a large RA layout should be chosen. On the other hand, if paging cost dominates, a small RA layout is appropriate.

- We quantitatively showed how the network operation cost (C_T) increases as inter-session idle times and user mobility increase.
- The variance V_m of cell residence times affect the network operation cost C_T . As V_m increases, C_T decreases.
- RC and RT have similar performance when the mobility rate λ_m is fixed. RC may significantly outperform RT when λ_m changes from time to time.
- DRC nicely captures the traffic/mobility patterns, and always adjusts the K threshold close to the optimal values.

Acknowledgments

The authors would like to thank the three anonymous reviewers. Their comments have significantly improved the quality of this paper. This work was sponsored in part by MOE Program of Excellence Research under contract 92-E-FA04-4, FarEastone, the Lee and MTI Center for Networking Research, NCTU.

A Notation

- α : the probability that an ON-period is followed by an OFF-period in the same session
- $A(n)$: the total number of states for an n -layer RA random walk
- $B(n)$: the number of boundary edges in an n -layer RA
- C_T : the expected total cost for location update and terminal paging during t_p
- $C_T(K)$: the net cost in an idle period with threshold K
- C_u : the expected location update cost during t_p
- C_v : the expected terminal paging cost during t_p

- $f_m(t_{m,j})$: the density function for the $t_{m,j}$ distribution
- $f_p(t_p)$: the density function for the t_p distribution
- $f_p^*(s)$: the Laplace Transform for the t_p distribution
- K : the READY Counter threshold
- $1/\lambda_m$: the expected value for the $t_{m,j}$ distribution
- $1/\lambda_p$: the expected value for the t_p distribution
- $1/\lambda_{p1}$: the expected value for the OFF-periods t_{p1} distribution
- $1/\lambda_{p2}$: the expected value for the inter-session idle periods t_{p2} distribution
- N_c : the number of cell crossings during t_p
- $N_r(j)$: the number of RA crossings occurring between the $j + 1$ -st cell crossing and the N_c -th cell crossing during t_p
- N_u : the number of location updates (cell updates plus RA updates) during t_p
- $P(N_m, k)$: the probability that after N_m cell movements, an MS crosses k RA boundaries provided that the MS is initially in an arbitrary cell of an RA
- $P^*(N_m^*, k^*)$: the probability that after N_m^* cell movements, an MS crosses k^* RA boundaries provided that the MS is initially at a boundary cell of an RA
- $S(n)$: the number of cells in an n -layer RA
- T : the READY Timer threshold
- $t_{m,j}$: the cell residence time of an MS at cell C_j
- t_p : the time interval between the end of a packet transmission and the beginning of the next packet transmission
- U : the cost for a cell/RA update

- V : the cost for paging in a cell
- V_m : the variance for the $t_{m,j}$ distribution
- $\theta_1(k)$: the probability that an MS will leave an RA at the k th step provided that the MS is initially in an arbitrary cell of the RA
- $\theta_2(k)$: the probability that after an MS enters an RA, it moves out of the RA at the k th step

References

- [1] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2. Technical Specification 3G TS 23.060 version 3.6.0 (2001-01), 2000.
- [2] Akyildiz, I.F., Lin, Y.-B., Lai, W.-R., and Chen, R.-J. A New Random Walk Model for PCS Networks. *IEEE Journal on Selected Areas in Communications*, 18(7):1254–1260, 2000.
- [3] Chlamtac, I., Fang, Y., and Zeng, H. Call Blocking Analysis for PCS Networks under General Cell Residence Time. *IEEE WCNC, New Orleans*, September 1999.
- [4] ETSI. UMTS Terrestrial Radio Access (UTRA); Concept Evaluation, Version 3.0.0. Technical Report UMTS 30.06, December 1997.
- [5] Fang, Y. and Chlamtac, I. Teletraffic Analysis and Mobility Modeling for PCS Networks. *IEEE Transactions on Communications*, 47(7):1062–1072, July 1999.
- [6] Fang, Y., Chlamtac, I., and Fei, H.-B. Analytical Results for Optimal Choice of Location Update Interval for Mobility Database Failure Restoration in PCS networks. *IEEE Transactions on Parallel and Distributed Systems*, 11(6):615–624, 2000.
- [7] Jain, R. and Routhier, S.A. Packet Trains: Measurements and a New Model for Computer Network Traffic. *IEEE Journal on Selected Areas in Communications*, 4(6):986–995, 1986.
- [8] Johnson, N.L. *Continuous Univariate Distributions-1*. John Wiley & Sons, 1970.

- [9] Kelly, F.P. *Reversibility and Stochastic Networks*. John Wiley & Sons, 1979.
- [10] Lin, Y.-B. and Chlamtac, I. *Wireless and Mobile Network Architectures*. John Wiley & Sons, Inc., 2001.
- [11] Lin, Y.-B., Haung, Y.-R., Chen, Y.-K., and Chlamtac, I. Mobility Management: From GPRS to UMTS. *Accepted and to appear in Wireless Communications and Mobile Computing*.
- [12] Lin, Y.-B., Lai, W.-R., and Chen, R.-J. Performance Analysis for Dual Band PCS Networks. *IEEE Transactions on Computers*, 49(2):148–159, 2000.
- [13] Manber, Udi. *Introduction to Algorithms - A Creative Approach*. Addison Wesley., 1989.
- [14] Willinger, W., Taqqu, M.S., Sherman, R., and Wilson, D.V. Self-Similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level. *IEEE/ACM Trans. Networking*, 5(1):71–86, 1997.