Modeling UMTS Discontinuous Reception Mechanism

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

***Abstract*—This paper investigates the discontinuous reception (DRX) mechanism of universal mobile telecommunications system (UMTS). DRX is exercised between the network and a mobile station (MS) to save the power of the MS. The DRX mechanism is controlled by two parameters: the inactivity timer threshold and the DRX cycle. Analytic and simulation models are proposed to study the effects of these two parameters on output measures including the expected queue length, the expected packet waiting time, and the power saving factor. Our study quantitatively shows how to select appropriate inactivity timer and DRX cycle values for various traffic patterns.**

***Index Terms*—Discontinuous reception (DRX), power saving, universal mobile telecommunications system (UMTS).**

I. Introduction

UNIVERSAL mobile telecommunications system (UMTS) [10] supports mobile multimedia applications with high data transmission rates. Fig. 1 illustrates a simplified UMTS architecture, which consists of the *core network* and the *UMTS terrestrial radio access network* (UTRAN). The core network is responsible for switching/routing calls and data connections to the external networks, while the UTRAN handles all radio-related functionalities. The UTRAN consists of *radio network controllers* (RNCs) and *node Bs* (i.e., base stations) that are connected by an *asynchronous transfer mode* (ATM) network. A *mobile station* (MS) communicates with node Bs through the radio interface based on the WCDMA (Wideband CDMA) technology [10].

In UMTS, MS power consumption is a serious problem for wireless data transmission. The data bandwidth is significantly limited by the battery capacity [15]. Therefore, power saving mechanisms are typically exercised to reduce power consumption. Most existing wireless mobile networks (including UMTS) employ *discontinuous reception* (DRX) to conserve the power of MSs. DRX allows an idle MS to power off the radio receiver for a predefined period (called the *DRX cycle*) instead of continuously listening to the radio channel. In MOBITEX [17], the network periodically transmits a specific ⟨SVP6⟩ frame to announce the list of the MSs that have pending packets. All MSs are required to synchronize with every ⟨SVP6⟩ frame and wake up immediately before the transmission starts. When some MSs experience high traffic loads, the network may decide to shorten the announcement interval to reduce the frame delay. As a consequence, the low traffic MSs will consume extra unnecessary power budget. In CDPD [5], [16], and IEEE 802.11 [11] standards, similar DRX mechanisms are utilized except that an MS is not forced to wake up at every announcement instant. Instead, the MS may choose to omit some announcements to further reduce its power consumption. A wake-up MS has to send a receiver ready (RR) frame to inform the network that it is ready to receive the pending frames. However, such RR transmissions may collide with each other if the MSs tend to wake up at the same time. Thus, RR retransmissions are likely to occur and extra power is unnecessarily consumed. UMTS DRX [2], [4] enhances the above mechanisms by allowing an MS to negotiate its own DRX cycle length with the network. Therefore, the network is aware of sleep/wake-up scheduling of each MS, and only delivers the paging message when the MS wakes up.

The CDPD DRX mechanism has been investigated through simulation models [16]. In [13], an analytic model was proposed to investigate CDPD DRX mechanism. This model does not provide close-form solution. Furthermore, the model was not validated against simulation experiments. This paper investigates the performance of the UMTS DRX with a variant of the *M/G/1* vacation model. We derive the close-form equations for the output measures and validate the results against simulation. Based on the proposed model, the DRX performance is investigated by numerical examples.

II. UMTS DRX Mechanism

The UMTS DRX mechanism is realized through the *radio resource control* (RRC) finite state machine exercised between the RNC and the MS [1]. There are two modes in this finite state machine (see Fig. 2). In the **RRC Idle** mode, the MS is tracked by the core network without involving the UTRAN. When an RRC connection is established between the MS and its serving RNC, the MS enters the **RRC Connected** mode. This mode consists of four states. If the MS obtains a dedicated traffic channel for the RRC connection, it enters the **Cell_DCH** state. On the

Manuscript received June 10, 2003; revised October 3, 2003; accepted November 18, 2003. The editor coordinating the review of this paper and approving it for publication is Z. Zhang. The work of S.-R. Yang was supported by the Mediatek Fellowship and the National Science Council under Contract NSC 93–2213–E-007–112. The work of Y.-B. Lin was supported in part by the National Science Council Excellence Project NSC93–2752–E-0090005–PAE, Chair Professorship of Providence University, IIS/Academia Sinica, FarEastone, CCL/ITRI, and by the Lee and MTI Center for Networking Research, National Chiao Tung University.

S.-R. Yang was with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan 30050, R.O.C. He is now with the Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan 30050, R.O.C. (e-mail: sryang@cs.nthu.edu.tw).

Y.-B. Lin was with the Department of Computer Science and Information Management, Providence University, Taichung, Taiwan 30050, R.O.C. He is now with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan 30050, R.O.C. (e-mail: liny@csie.nctu.edu.tw).

Digital Object Identifier 10.1109/TWC.2004.840259

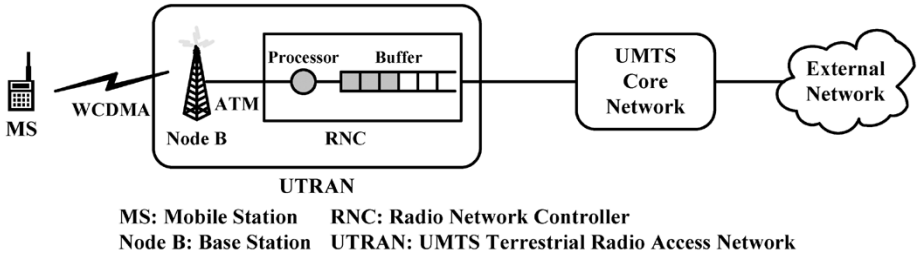


Fig. 1. Simplified UMTS network architecture.

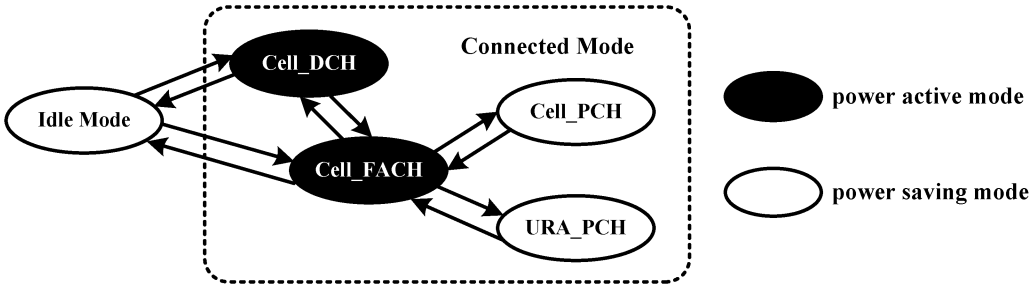


Fig. 2. RRC state diagram.

other hand, if the MS is allocated a common or shared traffic channel (i.e., the channel is shared by several MSs), it enters the **Cell_FACH** state. The data communication activities can only be performed in these two states. In the **Cell_PCH** state, no uplink access is possible, and the MS selects a paging channel (PCH) to monitor paging messages from the RNC. In the above three RRC states, the MS performs location update whenever it moves to a new cell (i.e., the radio coverage of a Node B). If the MS receives packets infrequently, the UTRAN may eliminate the cell update overhead by instructing the MS to move to the **URA_PCH** state. In this state, the MS performs location update for every *UTRAN registration area* (URA) crossing. Details of cell and URA updates can be found in [20].

In the **Cell_DCH** and **Cell_FACH** states, the MS receiver is always turned on to receive packets. These states correspond to the *power active mode*. In the **RRC Idle** mode, **Cell_PCH** and **URA_PCH** states, the DRX is exercised to reduce the MS power consumption. These states/mode correspond to the *power saving mode*. The MS receiver activities are described in terms of three periods.

The busy period: During packet transmission (i.e., the "server" is "busy"), the UMTS core network sends the packets to an MS through the RNC and Node B. The incoming packets are first stored in the RNC buffer before they are delivered to the MS. Since the MS is in the power active mode, the RNC processor immediately transmits packets in the *first in–first out* (FIFO) order. Due to high error-rate and low bit-rate nature of radio transmission, the *stop-and-wait hybrid automatic repeat reQuest* (SAW-Hybrid ARQ) flow control algorithm [3] is exercised between the Node B and the MS to guarantee successful radio packet delivery. The SAW-Hybrid ARQ algorithm works as follows. When the Node B sends a packet to the MS, it waits for a positive acknowledgment (ack) from the MS before it can transmit the next packet. The Node B may receive negative acknowledgments (naks) from the MS, which indicate that some errors have occurred (e.g., the transmitted packet is damaged). In this case, the Node B re-transmits the packet until an ack is received.

The inactivity period: If the RNC buffer becomes empty, the RNC inactivity timer is activated. If any packet arrives at the RNC before the inactivity timer expires, the timer is stopped. The RNC processor starts to transmit packets, and another busy period begins. Note that the MS is in the power active mode in both the busy and inactivity periods, where the MS receiver is turned off.

The sleep period: If no packet arrives before the inactivity timer expires, the MS enters the power saving mode and the MS receiver is turned off. The MS sleep period contains at least one DRX cycles. At the end of a DRX cycle, the MS wakes up to listen to the PCH. If some packets have arrived at the RNC during the last DRX cycle (i.e., the paging indicator for this MS is set), the MS starts to receive packets and the sleep period terminates. Otherwise, the MS returns to sleep until the end of the next DRX cycle. In the power saving mode, the RNC processor will not transmit any packets to the MS.

Based on the above description, we propose an analytic model for the UMTS DRX mechanism. As illustrated in Fig. 1, the UMTS core network sends the packets to an MS through the RNC and Node B. We assume that packet arrivals to the RNC form a Poisson stream with rate λ_a. The RNC processor sends the packets to the Node B through an ATM link. The Node B then forward the packets to the MS by the WCDMA radio link. Compared with WCDMA radio transmission, ATM is much faster and more reliable. Therefore the ATM transmission delay is ignored in our analytic model, and the RNC and the Node B are treated as a FIFO server. Let t_x denote the time interval between when a packet is transmitted by the RNC processor and when the corresponding ack is received by the RNC processor. Let t_I be the threshold of the RNC inactivity timer, and t_S be the MS sleep period. The UMTS DRX is modeled as a variant of

the *M/G/1* queue with multiple vacations [19], where t_x represents the service time (the period between when a packet is sent from the UTRAN to the MS and when the UTRAN receives the ack from the MS), and t_S corresponds to the server vacations. Our model is different from the existing *M/G/1* vacation model due to the introduction of the inactivity timer threshold t_I. In our model, the server can not enter vacation mode immediately after the queue is empty. The following output measures are derived:

1) mean queue length: the expected number of packets buffered in the UTRAN, including the one in delivery and those waiting in the RNC buffer;
2) mean packet waiting time: the expected waiting time of a packet in the RNC buffer before it is transmitted to the MS;
3) power saving factor: the probability that the MS receiver is turned off when exercising the UMTS DRX mechanism; this factor indicates the percentage of power saving in the DRX (compared with the case where DRX is not exercised).

III. Queue Length and Packet Waiting Time

This section derives the generating function for the queue length distribution and the Laplace Transform for the packet waiting time. Denote L_n as the queue length of the RNC buffer immediately after the nth packet completes service and departs (i.e., at the time when the nth ack is received by the RNC processor). $L_n = 0$ implies that the RNC buffer is empty after the nth packet has been received by the MS. In this case, the RNC inactivity timer is activated. Suppose that Q_n packets arrive at the RNC before the inactivity timer expires. If $Q_n = 0$, the MS enters the power saving mode. Let R_n be the number of packets that arrive during the sleep period. The probability mass functions for Q_n and R_n are denoted as $q_m = \Pr[Q_n = m]$ (where $m = 0, 1, 2, \ldots$) and $r_l = \Pr[R_n = l]$ (where $l = 1, 2, 3, \ldots$), respectively. Let A_{n+1} be the number of packets that arrive during the service time of the $n+1$st packet. The relationship between $L_{n+1}, L_n, A_{n+1}, Q_n$, and R_n can be expressed as

$$L_{n+1} = \begin{cases} L_n + A_{n+1} - 1, & \text{for } L_n \geq 1 \\ A_{n+1}, & \text{for } L_n = 0, \quad Q_n > 0 \\ R_n + A_{n+1} - 1, & \text{for } L_n = 0, \quad Q_n = 0 \end{cases}. \quad (1)$$

The above equation indicates that L_{n+1} is independent of L_m for $m < n$, and the sequence of the random variables $\{L_n; n = 1, 2, \ldots\}$ constitutes a *Markov chain* [18]. Let L be the queue length in the steady state. The steady state distribution for this Markov chain is defined as

$$\pi_k = \Pr[L = k] = \lim_{n\to\infty} \Pr[L_n = k], \quad \text{for } k = 0, 1, 2, \ldots$$

which can be solved by using $\sum_{k=0}^{\infty} \pi_k = 1$ and the balance equations

$$\pi_k = \sum_{j=0}^{\infty} \pi_j \Pr[L_{n+1} = k \,|\, L_n = j] \quad (2)$$

where $\Pr[L_{n+1} = k \,|\, L_n = j]$ is the state transition probability. Denote $a_k = \Pr[A_n = k]$ for $n \geq 1$ and $k \geq 0$. Based on (1), $\Pr[L_{n+1} = k \,|\, L_n = j]$ is expressed as (3), shown at the bottom of the page. Substituting (3) into (2), π_k is expressed as

$$\pi_k = \pi_0 \left[(1 - q_0) a_k + q_0 \left(\sum_{l=1}^{k+1} r_l a_{k-l+1} \right) \right] + \sum_{j=1}^{k+1} \pi_j a_{k-j+1}, \quad k \geq 0. \quad (4)$$

Let $\Pi(z) = \sum_{k=0}^{\infty} \pi_k z^k$ be the generating function for the π_k distribution. From (4), we have

$$\begin{aligned} \Pi(z) &= \sum_{k=0}^{\infty} z^k \pi_0 \left[(1 - q_0) a_k + q_0 \left(\sum_{l=1}^{k+1} r_l a_{k-l+1} \right) \right] + \sum_{k=0}^{\infty} z^k \sum_{j=1}^{k+1} \pi_j a_{k-j+1} \\ &= (1 - q_0)\pi_0 \sum_{k=0}^{\infty} a_k z^k + q_0 \pi_0 \sum_{l=1}^{\infty} r_l z^{l-1} \sum_{k'=0}^{\infty} a_{k'} z^{k'} + \left[\frac{\Pi(z) - \pi_0}{z} \right] \sum_{k'=0}^{\infty} a_{k'} z^{k'}. \end{aligned} \quad (5)$$

Let $A(z) = \sum_{k=0}^{\infty} a_k z^k$ be the probability generating function for a_k (the probability that k packets arrive during a packet service time t_x). These packet arrivals form a Poisson process with rate λ_a, which is independent of t_x. Suppose that the t_x distribution has the Laplace Transform $f_x^*(s)$, mean $1/\lambda_x$ and variance V_x. From [7, Th. 4.2], we have

$$A(z) = f_x^*(\lambda_a - \lambda_a z). \quad (6)$$

$$\Pr[L_{n+1} = k \,|\, L_n = j] = \begin{cases} a_{k-j+1}, & \text{for } j \geq 1, \quad k \geq j - 1 \\ 0, & \text{for } j \geq 1, \quad 0 \leq k < j - 1 \\ \left[(1 - q_0) a_k + q_0 \times \left(\sum_{l=1}^{k+1} r_l a_{k-l+1} \right) \right], & \text{for } j = 0, \quad k \geq 0 \end{cases} \quad (3)$$

Let $R(z) = \sum_{l=1}^{\infty} r_l z^l$ be the generating function for the r_l distribution. From (6), (5) is rewritten as

$$\Pi(z) = \left\{(1-q_0)\pi_0 + q_0\pi_0\left[\frac{R(z)}{z}\right] + \left[\frac{\Pi(z)-\pi_0}{z}\right]\right\} f_x^*(\lambda_a - \lambda_a z). \quad (7)$$

By rearranging the terms in (7), we obtain

$$\Pi(z) = \frac{\pi_0[1 - z(1-q_0) - q_0 R(z)] f_x^*(\lambda_a - \lambda_a z)}{f_x^*(\lambda_a - \lambda_a z) - z}. \quad (8)$$

Let $R = \lim_{n\to\infty} R_n$. Since $\Pi(1) = 1$, we derive π_0 from (8) as

$$\pi_0 = \frac{1-\rho}{1 - q_0 + q_0 E[R]} \quad (9)$$

where $\rho = \lambda_a/\lambda_x$. Substituting (9) into (8), we have

$$\Pi(z) = \frac{[1 - z(1-q_0) - q_0 R(z)](1-\rho) f_x^*(\lambda_a - \lambda_a z)}{(1 - q_0 + q_0 E[R])[f_x^*(\lambda_a - \lambda_a z) - z]}. \quad (10)$$

The Laplace Transform $f_w^*(s)$ for the packet waiting time t_w is derived as follows. Following the FIFO scheduling policy, the RNC queue length seen by a departing packet is precisely the number of packets that arrived during the response time t_r of the packet, where

$$t_r = t_w + t_x. \quad (11)$$

Let $f_r^*(s)$ be the Laplace Transform of t_r. Similar to the derivation of $A(z)$ (in [7, Th. 4.2]),

$$\Pi(z) = f_r^*(\lambda_a - \lambda_a z). \quad (12)$$

Since the service time for a packet does not depend on the time it spends in the RNC buffer, t_w and t_x are independent. From (11) and the convolution property of the Laplace Transform, we have

$$f_r^*(\lambda_a - \lambda_a z) = f_w^*(\lambda_a - \lambda_a z) f_x^*(\lambda_a - \lambda_a z). \quad (13)$$

From (10), (12), and (13), we have

$$f_w^*(s) = \frac{\Pi(1 - s/\lambda_a)}{f_x^*(s)} = \left\{\frac{\lambda_a q_0[1 - R(1 - s/\lambda_a)] + s(1-q_0)}{1 - q_0 + q_0 E[R]}\right\} \times \left[\frac{(1-\rho)}{\lambda_a f_x^*(s) - \lambda_a + s}\right]. \quad (14)$$

Assume that the length of each DRX cycle t_D in a sleep period is independent and identically distributed with mean $1/\lambda_D$, variance V_D, and the Laplace Transform $f_D^*(s)$. From [21]

$$R(z) = \frac{f_D^*(\lambda_a - \lambda_a z) - f_D^*(\lambda_a)}{1 - f_D^*(\lambda_a)} \quad (15)$$

and

$$E[R] = \frac{\lambda_a}{[1 - f_D^*(\lambda_a)]\lambda_D}. \quad (16)$$

Assume that the inactivity timer threshold t_I has the density function $f_I(t)$, mean $1/\lambda_I$, variance V_I, and Laplace Transform $f_I^*(s)$. From the Poisson distribution

$$q_0 = \int_{t=0}^{\infty} \left[\frac{e^{-\lambda_a t}(\lambda_a t)^0}{0!}\right] f_I(t)\, dt = f_I^*(\lambda_a). \quad (17)$$

Substituting (15), (16), and (17) into (10) and (14), we obtain (18) and (19), as shown at the bottom of the page.

From (18) and (19), the expected number of packets $E[L]$ and the mean packet waiting time $E[t_w]$ are expressed as

$$E[L] = \left.\frac{d\Pi(z)}{dz}\right|_{z=1} = \frac{\lambda_a^2 f_I^*(\lambda_a)(1 + V_D\lambda_D^2)}{2\{\lambda_D^2[1 - f_I^*(\lambda_a)][1 - f_D^*(\lambda_a)] + \lambda_a\lambda_D f_I^*(\lambda_a)\}} + \frac{\lambda_a^2(1 + V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2} + \rho \quad (20)$$

and

$$E[t_w] = -\left.\frac{df_w^*(s)}{ds}\right|_{s=0} = \frac{\lambda_a f_I^*(\lambda_a)(1 + V_D\lambda_D^2)}{2\{\lambda_D^2[1 - f_I^*(\lambda_a)][1 - f_D^*(\lambda_a)] + \lambda_a\lambda_D f_I^*(\lambda_a)\}} + \frac{\lambda_a(1 + V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}. \quad (21)$$

$$\Pi(z) = \left\{\frac{\lambda_D(1 - z[1 - f_I^*(\lambda_a)])[1 - f_D^*(\lambda_a)] - \lambda_D f_I^*(\lambda_a)[f_D^*(\lambda_a - \lambda_a z) - f_D^*(\lambda_a)]}{\lambda_D[1 - f_I^*(\lambda_a)][1 - f_D^*(\lambda_a)] + \lambda_a f_I^*(\lambda_a)}\right\} \times \left[\frac{(1-\rho) f_x^*(\lambda_a - \lambda_a z)}{f_x^*(\lambda_a - \lambda_a z) - z}\right] \quad (18)$$

and

$$f_w^*(s) = \left\{\frac{\lambda_D\lambda_a f_I^*(\lambda_a)[1 - f_D^*(s)] + \lambda_D s[1 - f_I^*(\lambda_a)][1 - f_D^*(\lambda_a)]}{\lambda_D[1 - f_I^*(\lambda_a)][1 - f_D^*(\lambda_a)] + \lambda_a f_I^*(\lambda_a)}\right\} \times \left[\frac{(1-\rho)}{\lambda_a f_x^*(s) - \lambda_a + s}\right] \quad (19)$$

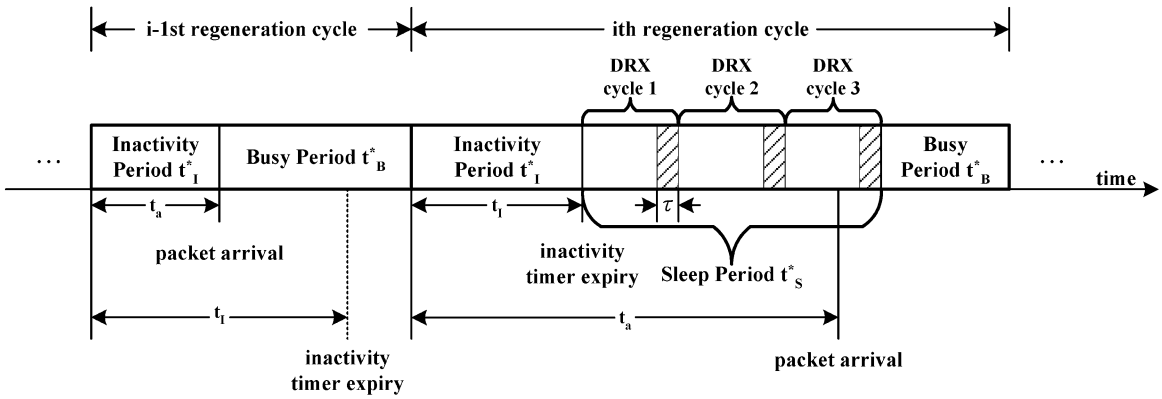


Fig. 3. Regeneration cycles for MS activities.

IV. Power Saving Factor

This section derives the power saving factor P_s. We draw the timing diagram of MS receiver activities in Fig. 3. In this figure, the MS receiver activities are characterized by a *regenerative process* [18], where a regeneration cycle consists of an inactivity period t_I^*, a sleep period t_S^*, and a busy period t_B^*. We note that at the end of every DRX cycle, the MS must wake up for a short period τ so that it can listen to the paging information from the network. Therefore, the "power saving" period in a DRX cycle is $t_D - \tau$. Suppose that there are N DRX cycles in a sleep period. From [18, Th. 3.7.1]

$$\begin{aligned} P_s &= \lim_{t\to\infty} \Pr[\text{the } MS \text{ receiver is turned off at time } t] \\ &= \frac{E[t_S^*] - E[N]\tau}{E[t_I^*] + E[t_S^*] + E[t_B^*]}. \end{aligned} \tag{22}$$

$E[t_I^*], E[t_S^*], E[N]$, and $E[t_B^*]$ are derived as follows. In Fig. 3, t_a is the time interval between when the RNC inactivity timer is activated and when the next packet arrives. Thus, $t_I^* = \min(t_a, t_I)$. If the next packet arrives before the inactivity timer expires (i.e., $t_a < t_I$), then $t_I^* = t_a$, and the next busy period follows (see the $i-1$st cycle in Fig. 3). Otherwise, (the next packet arrives after the inactivity timer has expired; i.e., $t_a \geq t_I$) $t_I^* = t_I$, and the next sleep period follows as illustrated in the ith cycle in Fig. 3. From the memoryless property of Poisson process [18], t_a is exponentially distributed with rate λ_a. Therefore

$$\begin{aligned} E[t_I^*] &= E[\min(t_a, t_I)] \\ &= \int_{x=0}^{\infty} \Pr[\min(t_a, t_I) > x]\, dx \\ &= \int_{x=0}^{\infty} \Pr[t_a > x] \Pr[t_I > x]\, dx \\ &= \left(\frac{1}{\lambda_a}\right)[1 - f_I^*(\lambda_a)]. \end{aligned} \tag{23}$$

$E[t_S^*]$ is derived as follows. If $t_a \geq t_I$, then $t_S^* = t_S$ and R packets arrive during the t_S period. Let $f_S^*(s)$ be the Laplace Transform of t_S. From [7, Th. 4.2]

$$R(z) = f_S^*(\lambda_a - \lambda_a z) \Leftrightarrow f_S^*(s) = R(1 - s/\lambda_a) \tag{24}$$

From (15) and (24), we have

$$f_S^*(s) = \frac{f_D^*(s) - f_D^*(\lambda_a)}{1 - f_D^*(\lambda_a)}. \tag{25}$$

From (25), the mean sleep period $E[t_S^* \,|\, t_a \geq t_I]$ is

$$\begin{aligned} E[t_S^* \,|\, t_a \geq t_I] = E[t_S] &= -\frac{df_S^*(s)}{ds}\bigg|_{s=0} \\ &= \frac{1}{[1 - f_D^*(\lambda_a)]\lambda_D}. \end{aligned} \tag{26}$$

Note that $t_a < t_I$ with probability $1 - q_0$. From (26), we have

$$\begin{aligned} E[t_S^*] &= (1 - q_0) \times 0 + q_0 \times E[t_S^* \,|\, t_a \geq t_I] \\ &= \frac{q_0}{[1 - f_D^*(\lambda_a)]\lambda_D}. \end{aligned} \tag{27}$$

Substitute (17) into (27) to yield

$$E[t_S^*] = \frac{f_I^*(\lambda_a)}{[1 - f_D^*(\lambda_a)]\lambda_D}. \tag{28}$$

The expected value $E[N]$ is derived as follows. Since a t_S^* period consists of N DRX cycles t_D, from Wald's theorem [18], we have

$$E[t_S^*] = E[N]E[t_D] = E[N]/\lambda_D. \tag{29}$$

Substitute (28) into (29) to yield

$$E[N] = \frac{f_I^*(\lambda_a)}{1 - f_D^*(\lambda_a)}. \tag{30}$$

We derive $E[t_B^*]$ as follows. If $t_a < t_I$ (with probability $1 - q_0$), $t_S^* = 0$ and t_B^* is exactly the same as the busy period t_B of an *M/G/1* queue. From [7]

$$E[t_B^* \,|\, t_a < t_I] = E[t_B] = \frac{\rho}{\lambda_a(1 - \rho)} \tag{31}$$

where $\rho = \lambda_a/\lambda_x$. If $t_a \geq t_I$ (with probability q_0), $t_S^* > 0$, and R packets arrive during the t_S^* period. In this case, t_B^* is the sum of R t_B periods. From Wald's theorem, we have

$$E[t_B^* \,|\, t_a \geq t_I] = E[R]E[t_B]. \tag{32}$$

TABLE I
COMPARISON OF THE ANALYTIC AND SIMULATION RESULTS ($V_x = 10/\lambda_x^2, t_I = 60E[t_x], t_D = 5E[t_x]$, AND $\tau = 0.1E[t_x]$)

λ_a(unit: λ_x)	0.010	0.012	0.014	0.016	0.018	0.020
$E[L]$(Analytical)	0.02443	0.02763	0.03050	0.03312	0.03555	0.03783
$E[L]$(Simulation)	0.02421	0.02747	0.03026	0.03313	0.03531	0.03759
Error	0.9005%	0.5791%	0.7869%	0.0302%	0.6751%	0.6344%
λ_a(unit: λ_x)	0.010	0.012	0.014	0.016	0.018	0.020
$E[t_w]/E[t_x]$(Analytical)	1.44298	1.30233	1.17876	1.07024	0.97502	0.89158
$E[t_w]/E[t_x]$(Simulation)	1.44223	1.30221	1.17911	1.07249	0.97474	0.89200
Error	0.0520%	0.0092%	0.0297%	0.2102%	0.0287%	0.0471%
λ_a(unit: λ_x)	0.010	0.012	0.014	0.016	0.018	0.020
P_s(Analytical)	0.53843	0.47852	0.42542	0.37833	0.33652	0.29939
P_s(Simulation)	0.53864	0.47843	0.42554	0.37872	0.33654	0.29928
Error	0.0390%	0.0188%	0.0282%	0.1031%	0.0059%	0.0367%

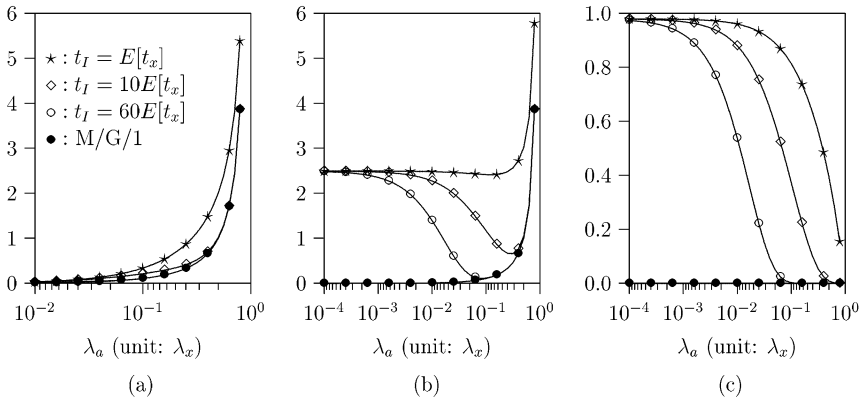


Fig. 4. Effects of λ_a and t_I ($V_x = (1/\lambda_x^2), t_D = 5E[t_x], \tau = 0.1E[t_x]$). (a) $E[L]$. (b) $E[t_w]$ (unit: $E[t_x]$). (c) P_s.

Substitute (31) and (16) into (32) to yield

$$E[t_B^* \,|\, t_a \geq t_I] = \frac{\rho}{(1-\rho)[1-f_D^*(\lambda_a)]\lambda_D}. \tag{33}$$

From (31) and (33), we have

$$\begin{aligned} E[t_B^*] &= (1-q_0)E[t_B^* \,|\, t_a < t_I] + q_0 E[t_B^* \,|\, t_a \geq t_I] \\ &= (1-q_0)\left[\frac{\rho}{\lambda_a(1-\rho)}\right] \\ &\quad + q_0\left\{\frac{\rho}{(1-\rho)[1-f_D^*(\lambda_a)]\lambda_D}\right\}. \end{aligned} \tag{34}$$

Substitute (17) into (34) to yield

$$E[t_B^*] = \frac{\rho\{\lambda_D[1-f_I^*(\lambda_a)][1-f_D^*(\lambda_a)] + \lambda_a f_I^*(\lambda_a)\}}{\lambda_a\lambda_D[1-f_D^*(\lambda_a)](1-\rho)}. \tag{35}$$

From (23), (28), (30), and (35), (22) is rewritten as

$$P_s = \frac{\lambda_a f_I^*(\lambda_a)(1-\rho)(1-\lambda_D\tau)}{\lambda_D[1-f_I^*(\lambda_a)][1-f_D^*(\lambda_a)] + \lambda_a f_I^*(\lambda_a)}. \tag{36}$$

V. NUMERICAL EXAMPLES

Our analytic model has been validated against a discrete event simulation model. This simulation model is similar to the one we developed in [14], and the details are omitted. For demonstration purpose, we assume t_x to have a Gamma distribution with mean $1/\lambda_x$, variance V_x, and the Laplace Transform

$$f_x^*(s) = \left(\frac{\lambda_x\gamma}{s+\lambda_x\gamma}\right)^\gamma \quad \text{where} \quad \gamma = \frac{1}{V_x\lambda_x^2}.$$

The Gamma distribution is often used in mobile telecommunications network modeling [6], [8], [9]. It has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions (see [12, Lemma 3.9]). One may also measure the t_x periods in a real UMTS network, and the measured data can be approximated by a Gamma distribution as the input to our models. Table I compares the analytic and simulation results. We consider fixed t_I and t_D. In this table, $V_x = 10/\lambda_x^2, t_I = 60E[t_x], t_D = 5E[t_x]$, and $\tau = 0.1E[t_x]$. The table indicates that the errors between the analytic and simulation models are less than 0.05% in most cases. Based on the analytic model, we investigate the DRX performance.

Figs. 4–6 plot the $E[L], E[t_w]$, and P_s curves. In these figures, t_x has the Gamma distribution with variance V_x, and t_I and t_D are fixed. The parameter settings are described in the captions of the figures.

Effects of λ_a: Fig. 4(a) indicates that the mean queue length $E[L]$ increases as λ_a increases. For $\lambda_a < 0.1\lambda_x, E[L]$ is insignificantly affected by the change of λ_a. When λ_a approaches λ_x, the resulting queueing system becomes unstable, and $E[L]$ grows without bound. Fig. 4(b) plots the mean packet waiting time $E[t_w]$ normalized by the mean

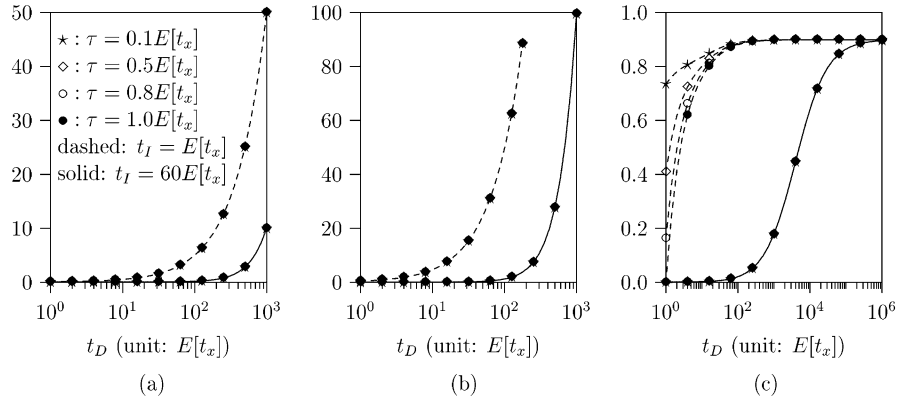


Fig. 5. Effects of t_D and τ ($V_x = (1/\lambda_x^2), \lambda_a = 0.1\lambda_x$). (a) $E[L]$. (b) $E[t_w]$ (unit: $E[t_x]$). (c) P_s.

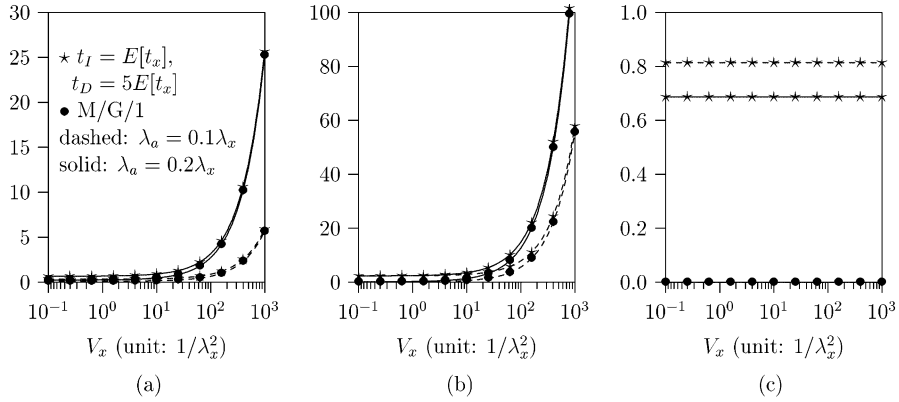


Fig. 6. Effects of V_x ($\tau = 0.1E[t_x]$). (a) $E[L]$. (b) $E[t_w]$ (unit: $E[t_x]$). (c) P_s.

packet service time $E[t_x]$. The "●" curve shows intuitive result that for an *M/G/1* queue, $E[t_w]$ is an increasing function of λ_a. For DRX, $E[t_w]$ curves decrease and then increase as λ_a increases. This phenomenon is explained as follows. For $\lambda_a < 0.1\lambda_x$, $E[t_w]$ is affected by the power saving mode operation. Specifically, when λ_a approaches 0, the MS is always in the power saving mode. From (21), every arrival packet is expected to wait for

$$\lim_{\lambda_a \to 0} E[t_w] = \frac{t_D}{2}. \tag{37}$$

In Fig. 4(b), $t_D = 5E[t_x]$. From (37), $\lim_{\lambda_a \to 0} E[t_w] = 2.5E[t_x]$, which is the value we observe in the figure. As λ_a increases, it is more likely that the MS is in the power active mode when packets arrive. In this case, more packets are processed without experiencing the sleep periods, and $E[t_w]$ decreases as λ_a increases. On the other hand, if λ_a approaches λ_x, the packet traffic load exceeds the transmission capability, and $E[t_w]$ increases as λ_a increases. Fig. 4(c) shows the intuitive result that the power saving factor P_s is a decreasing function of λ_a.

Effects of t_I: Fig. 4 indicates that by increasing the inactivity timer threshold t_I, $E[L]$, $E[t_w]$, and P_s decrease. When $t_I \to \infty$, the MS never enters the power saving mode, and the system is the same as an *M/G/1* queue. When λ_a is small (e.g., $\lambda_a < 10^{-3}\lambda_x$), the inactivity timer always expires before the next packet arrives. Therefore, the output measures are insignificantly affected by the change of t_I.

Effects of t_D: Fig. 5 indicates that $E[L]$, $E[t_w]$, and P_s are increasing functions of t_D. We observe that when t_D is small [e.g., $t_D < 10E[t_x]$ for the dashed curves in Figs. 5(a) and (b)], decreasing t_D will not improve the $E[L]$ and $E[t_w]$ performance. On the other hand, when t_D is large [e.g., $t_D > 100E[t_x]$ for the dashed curves in Fig. 5(c)], increasing t_D will not improve the P_s performance. Therefore, for $t_I = E[t_x]$ (i.e., the dashed curves), t_D should be selected in the range $[10E[t_x], 100E[t_x]]$.

Effects of τ: It is clear that $E[L]$ and $E[t_w]$ are not affected by τ. Fig. 5(c) illustrates the impact of τ on P_s. When t_D is large, τ (the cost of wakeup) is a small portion of a DRX cycle, which only has insignificant impact on P_s. When t_I is large and t_D is small, the MS receiver is almost always turned on and $P_s \simeq 0$. In this case, the τ impact is also insignificant. When both t_I and t_D are small, P_s is significantly increased as τ decreases.

Effects of V_x: Fig. 6 illustrates that $E[L]$ and $E[t_w]$ are increasing functions of the variance V_x of the packet transmission delay t_x. This effect is well known in queueing systems. On the other hand, P_s is not affected by V_x. Therefore, the effect of V_x can be ignored when tuning the DRX parameters for power saving.

VI. Conclusion

This paper investigated the UMTS DRX mechanism for MS power saving. The DRX mechanism is controlled by two parameters: the inactivity timer threshold t_I and the DRX cycle t_D. We proposed a variant *M/G/1* queueing model with vacations to study the effects of t_I and t_D on output measures including the expected queue length, the expected packet waiting time, and the power saving factor. Our analytic model is different from the existing *M/G/1* vacation model due to the introduction of t_I (and therefore, the server can not enter the vacation mode immediately after the queue is empty). The analytic approach was validated against the simulation experiments. Several numerical examples were presented to quantitatively show how to select appropriate t_I and t_D values for various traffic patterns. Our study indicated that with proper parameter settings, UMTS DRX can effectively reduce the MS power consumption.

Acknowledgment

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

References

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

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

[3] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UTRA High Speed Downlink Packet Access*, 2001. 3GPP, Technical Specification 3G TR 25.950 ver. 4.0.0 (2001–03).

[4] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; UE Procedures in Idle Mode and Procedures for Cell Reselection in Connected Mode*, 2002. 3GPP, Technical Specification 3G TS 25.304 version 5.1.0 (2002–06).

[5] "Cellular Digital Packet Data System Specification: Release 1.1," CDPD Forum, Inc., Tech. Rep., Jan. 1995. CDPD Forum.

[6] I. Chlamtac, Y. Fang, and H. Zeng, "Call blocking analysis for PCS networks under general cell residence time," in *IEEE Wireless Commun. Netw. Conf.*, vol. 2, New Orleans, LA, Sep. 21–24, 1999, pp. 550–554.

[7] J. N. Daigle, *Queueing Theory for Telecommunications*. Reading, MA: Addison-Wesley, 1992.

[8] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1062–1072, Jul. 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, no. 6, pp. 615–624, Jun. 2000.

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

[11] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1996. IEEE, Draft Standard 802.11 D3.1.

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

[13] S. J. Kwon, Y. W. Chung, and D. K. Sung, "Queueing model of sleep-mode operation in cellular digital packet data," *IEEE Trans. Veh. Technol.*, vol. 52, no. 4, pp. 1158–1162, Jul. 2003.

[14] Y.-B. Lin, "Estimating the likelihood of success of lazy cancellation in time warp simulations," *Int. J. Comput. Simulation*, vol. 6, no. 2, pp. 163–174, 1996.

[15] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley, 2001.

[16] Y.-B. Lin and Y.-M. Chuang, "Modeling the sleep mode for cellular digital packet data," *IEEE Commun. Lett.*, vol. 3, no. 3, pp. 63–65, Mar. 1999.

[17] "Mobitex Interface Specification," RAM Mobile Data, Tech. Rep., 1994. RAM Mobile Data.

[18] S. M. Ross, *Stochastic Processes*, 2nd ed. New York: Wiley, 1996.

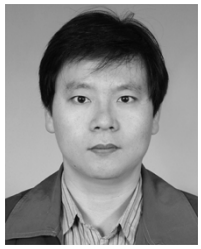
[19] H. Takagi, *Queueing Analysis—Volume 1: Vacation and Priority Systems, Part 1*. New York: Elsevier, 1991.

[20] S.-R. Yang and Y.-B. Lin, "A mobility management strategy for UMTS," in *Int. Conf. Inf. Netw.*, South Korea, Feb. 2003.

[21] ——, "Performance Analysis of UMTS Power Saving Mechanism," National Chiao Tung University, Hsinchu, Taiwan, Tech. Rep., 2003.

Shun-Ren Yang received the BSCSIE, MSCSIE, and Ph.D. degrees from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1998, 1999, and 2004, respectively.

Since August 2004, he has been with the Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University, Taiwan. His current research interests include 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 at Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed as a Professor of Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed Deputy Director of Microelectronics and Information Systems Research Center, NCTU. During 1997–1999, he was elected as Chairman of CSIE, NCTU. He is an Adjunct Research Fellow of Academia Sinica, and is Chair Professor at Providence University, Taichung, Taiwan, R.O.C. He serves as a consultant to many telecommunications companies, including FarEasTone and Chung Hwa Telecom. His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling. He has published over 150 journal articles and more than 200 conference papers. He is coauthor of the book *Wireless and Mobile Network Architecture* (New York: Wiley, 2001).

Dr. Lin is a Senior Technical Editor of IEEE NETWORK, an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, an Associate Editor of IEEE COMMUNICATIONS SURVEY AND TUTORIALS, an Editor of IEEE PERSONAL COMMUNICATIONS MAGAZINE, an Editor of Computer Networks, an Area Editor of *ACM Mobile Computing and Communication Review*, a columnist of *ACM Simulation Digest*, an Editor of *International Journal of Communications Systems*, an Editor of *ACM/Baltzer Wireless Networks*, an Editor of *Computer Simulation Modeling and Analysis*, an Editor of *Journal of Information Science and Engineering*, Program Chair for the 8th Workshop on Distributed and Parallel Simulation, General Chair for the 9th Workshop on Distributed and Parallel Simulation. Program Chair for the 2nd International Mobile Computing Conference, Guest Editor for the *ACM/Baltzer MONET* special issue on Personal Communications, a Guest Editor for IEEE TRANSACTIONS ON COMPUTERS special issue on Mobile Computing, a Guest Editor for IEEE TRANSACTIONS ON COMPUTERS special issue on Wireless Internet, and a Guest Editor for IEEE COMMUNICATIONS MAGAZINE special issue on Active, Programmable, and Mobile Code Networking. 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 a ACM Fellow.