Dynamic Power Saving Mechanism for 3G UMTS System

Shun-Ren Yang

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

This paper investigates the power saving mechanism of Universal Mobile Telecommunications System (UMTS). UMTS discontinuous reception (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 t_I and the DRX cycle t_D. Analytic analysis and simulation model are proposed to study the optimal t_I and t_D selections that maximize the MS power saving under the given mean packet waiting time constraint. We also devise an adaptive algorithm called dynamic DRX (DDRX). This algorithm dynamically adjusts the t_I and t_D values to enhance the performance of UMTS DRX. Our study quantitatively shows how to select the best inactivity timer and DRX cycle values for various traffic patterns. We also show that DDRX nicely captures the user traffic patterns, and always adjusts the t_I and t_D close to the optimal values.

Keywords: Adaptive algorithm, Discontinuous reception (DRX), Power saving, Universal Mobile Telecommunications System (UMTS)

1. Introduction

The third generation (3G) systems such as *Universal Mobile Telecommunications System* (UMTS) [9] offer wireless broadband access, and therefore can support mobile multimedia applications with high data transmission rates. As shown in Figure 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 the access to the PSTN/ISDN, while the PS domain provides the 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 Figure 1(a)) are provided by the *Serving GPRS Support Node* (SGSN; see Figure 1(d)) and the *Gateway GPRS Support Node* (GGSN; see Figure 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 Figure 1(b)) and *Radio Network Controllers* (RNCs; see Figure 1(c)) that are 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 the SGSNs. An MS communicates with Node Bs through the radio interface based on the *Wideband CDMA* (WCDMA) technology [9].

1.1. Discontinuous Reception

In UMTS, MS power consumption is a serious problem for wireless data transmission. The data bandwidth is significantly limited by the battery capacity [14]. 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* t_D) instead of continuously listening to the radio channel. Two types of DRX-based power saving protocols are proposed in the literature: *synchronous* (or *centrally controlled*) and *asynchronous* (or *distributedly controlled*) power saving protocols.

The synchronous power saving protocols are exercised in MOBITEX [17], CDPD [5,15] and IEEE 802.11 [10]. In MOBITEX, 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 these ⟨SVP6⟩ frame transmissions 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 and IEEE 802.11 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

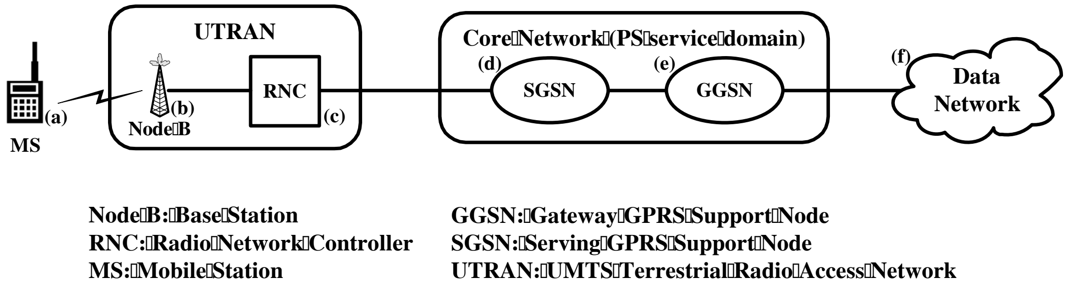


Figure 1. A simplified network architecture for the UMTS PS domain

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.

In contrast to the synchronous power saving protocols, asynchronous power saving protocols proposed in [18–20] do not require any MS wake-up synchronization. In these protocols, an MS is allowed to manage its power budget independently. At high power level, the MS may wake up more frequently to reduce the packet delay; on the other hand, at low power level, the MS may wake up less frequently to save the power budget. Packets for MSs with unknown reception state are temporarily stored in the buffers, and the network continuously broadcasts the addresses of these sleeping MSs through the paging procedure. Each MS may wake up at any time to check for pending packets. The paging process repeats until these MSs have received the paging messages. The asynchronous power saving protocols offer the flexibility of MS power management, while increasing the network signaling load for paging operation. Furthermore, these protocols can not guarantee an upper bound for the packet delay, and the packet delay variance at high traffic condition is expected to be considerably large.

1.2. UMTS Discontinuous Reception

UMTS DRX [2,4] combines the advantages of both synchronous and asynchronous power saving protocols. It allows 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 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 Figure 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 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 [22]. 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*.

1.3. Literature Review

In the literature, the CDPD DRX mechanism has been investigated through simulation models [15]. In [12], 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. In [23], we proposed a variant of the $M/G/1$ vacation model to investigate the performance of the UMTS DRX. We derived the close-form equations for the output measures and validated the results against simulation. Our

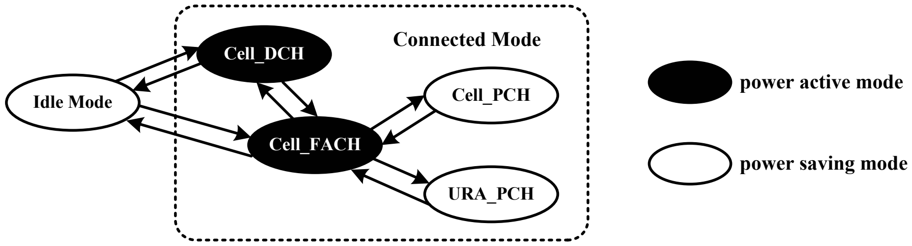


Figure 2. The RRC state diagram

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;
- whenever the server becomes idle, the inactivity timer is activated;
- the server enters vacation mode only if no packets arrive at the server before the inactivity timer expires.

In this paper, we further examine the optimal inactivity timer threshold t_I and DRX cycle t_D selection based on the proposed queueing model. The outline of this paper is as follows. (Some of the background material in this paper is repeated from our earlier paper [23] in order to keep this paper self-contained.) In Section 2, we review the proposed $M/G/1$ vacation model for UMTS DRX. Section 3 provides an analytic analysis for the best t_I and t_D selection. We also devise an adaptive algorithm called dynamic DRX (DDRX). This algorithm dynamically adjusts the t_I and t_D values to maximize the MS power saving under the given mean packet waiting time constraint. In Section 4, we investigate the performance of the UMTS DRX mechanism and the DDRX algorithm by numerical examples. Section 5 concludes this paper.

2. An $M/G/1$ Vacation Model for UMTS DRX

In UMTS, the MS receiver can be in the power active mode (**Cell_DCH** or **Cell_FACH** state) or the power saving mode (**RRC Idle** mode, **Cell_PCH** or **URA_PCH** state). The MS receiver activities are illustrated in Figure 3, and are described in terms of three periods:

The busy period. During packet transmission (i.e., the "server" is "busy"; see Figure 3(a)), 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 (see Figure 4). 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 (see Figure 3(b)). 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 on.

The sleep period. If no packet arrives before the inactivity timer expires (see Figure 3(c)), the MS enters the power saving mode (see Figure 3(d)) and the MS receiver is turned off. The MS sleep period contains at least one DRX cycles t_D. 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 have proposed an analytic model for the UMTS DRX mechanism in [23]. As illustrated in Figure 5, the UMTS core network sends the

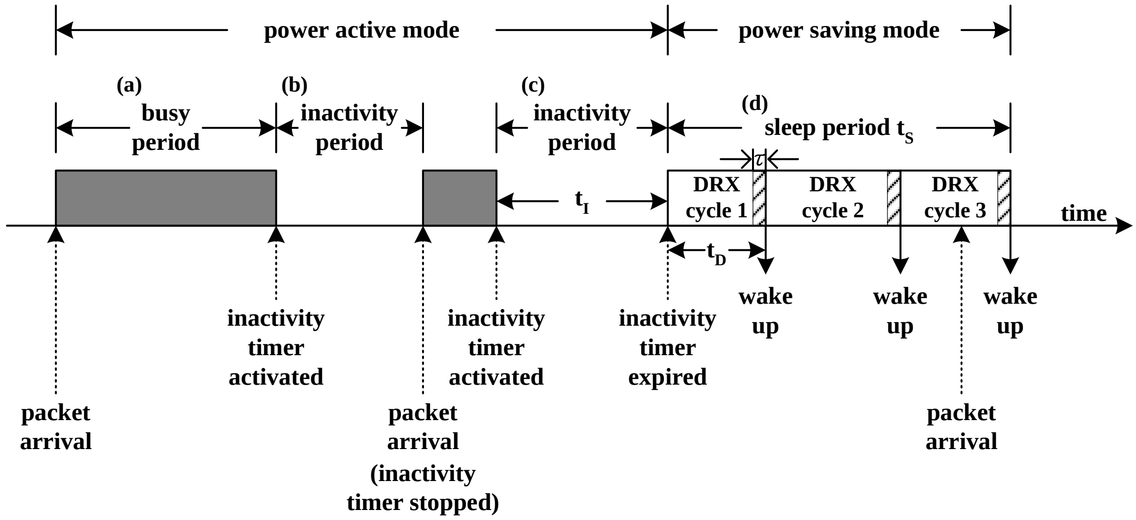


Figure 3. The timing diagram for the UMTS DRX mechanism

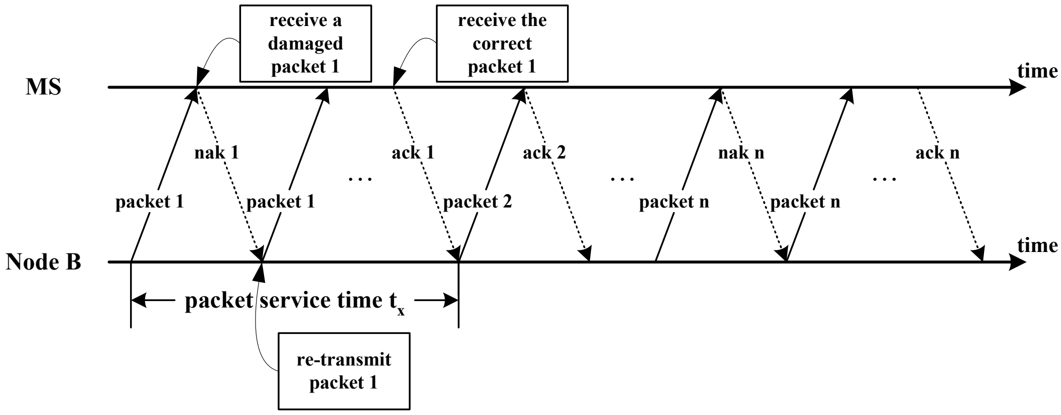


Figure 4. The timing diagram for SAW-Hybrid ARQ

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 forwards 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 (see the packet service time in Figure 4). 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 [21], 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. We have derived the following output measures [23]:

- mean packet waiting time $E[t_w]$: the expected waiting time of a packet in the RNC buffer before it is transmitted to the MS
- power saving factor P_s: 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)

Suppose that the t_x distribution has mean $1/\lambda_x$ and variance V_x. For fixed $t_I = 1/\lambda_I$ and fixed $t_D = 1/\lambda_D$, we have shown that

$$E[t_w] = \frac{\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}}{2[\lambda_D^2(1-e^{-\frac{\lambda_a}{\lambda_I}})(1-e^{-\frac{\lambda_a}{\lambda_D}})+\lambda_a\lambda_D e^{-\frac{\lambda_a}{\lambda_I}}]}$$

$$+\frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2} \quad (2.1)$$

$$P_s = \frac{\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}(1-\rho)(1-\lambda_D\tau)}{\lambda_D(1-e^{-\frac{\lambda_a}{\lambda_I}})(1-e^{-\frac{\lambda_a}{\lambda_D}})+\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}} \quad (2.2)$$

where $\rho = \lambda_a/\lambda_x$ and τ is the cost of wakeup[1].

3. Analytic Analysis for the Best t_I and t_D Selection

In this section, we investigate the effects of t_I and t_D on power saving under a given mean packet waiting time constraint. We first show in the following theorem that for fixed $E[t_w]$, P_s has the maximum value when the length of the DRX cycle t_D is set to 2τ (i.e., $\lambda_D = 1/2\tau$).

Theorem 1. In our $M/G/1$ vacation model, for any given mean packet waiting time $E[t_w] = C_w$, we have the maximum power saving factor P_s^* when $t_D = 2\tau$.

Proof. From (2.1) we have

$$C_w = \frac{\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}}{2[\lambda_D^2(1-e^{-\frac{\lambda_a}{\lambda_I}})(1-e^{-\frac{\lambda_a}{\lambda_D}})+\lambda_a\lambda_D e^{-\frac{\lambda_a}{\lambda_I}}]} + \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2} \quad (3.1)$$

Subtract $\frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}$ from both sides of (3.1) to yield

$$\frac{\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}}{2[\lambda_D^2(1-e^{-\frac{\lambda_a}{\lambda_I}})(1-e^{-\frac{\lambda_a}{\lambda_D}})+\lambda_a\lambda_D e^{-\frac{\lambda_a}{\lambda_I}}]} = C_w - \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2} \quad (3.2)$$

Multiplying both sides of (3.2) by $2(1-\rho)\lambda_D(1-\lambda_D\tau)$, we obtain

$$\frac{\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}(1-\rho)(1-\lambda_D\tau)}{\lambda_D(1-e^{-\frac{\lambda_a}{\lambda_I}})(1-e^{-\frac{\lambda_a}{\lambda_D}})+\lambda_a e^{-\frac{\lambda_a}{\lambda_I}}} = 2(1-\rho)\left[C_w - \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}\right] \times\lambda_D(1-\lambda_D\tau) \quad (3.3)$$

Then, from (2.2) and (3.3), we have

$$P_s = 2(1-\rho)\left[C_w - \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}\right]\lambda_D(1-\lambda_D\tau) = 2\tau(1-\rho)\left[C_w - \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}\right] \times\left[\frac{1}{4\tau^2} - \left(\lambda_D - \frac{1}{2\tau}\right)^2\right] \quad (3.4)$$

From (3.4), the maximum power saving factor P_s^* is achieved when $\lambda_D = 1/2\tau$ and is expressed as

$$P_s^* = \frac{1}{2\tau}(1-\rho)\left[C_w - \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}\right] \quad (3.5)$$

■

Let $\theta = \frac{\lambda_a(1+V_x\lambda_x^2)}{2(1-\rho)\lambda_x^2}$, which is the mean waiting time of a pure $M/G/1$ queue. The range of possible C_w values in Theorem 1 is discussed in the following two cases.

Case I. When $t_I \to \infty$ (or $\lambda_I \to 0$), the MS never enters the power saving mode, and the system is the same as an $M/G/1$ queue. In this case, we have the shortest mean packet waiting time $E[t_w] = \theta$. Consequently,

$$C_w \ge \theta.$$

Case II. For a given t_D, when $t_I \to 0$ (or $\lambda_I \to \infty$), the MS enters the power saving mode immediately after the RNC buffer is empty, and the system degenerates into a conventional $M/G/1$ queue with multiple vacations. In this case, we have the longest mean packet waiting time $E[t_w] = \theta + t_D/2$. As a result,

$$C_w \le \theta + t_D/2. \quad (3.6)$$

Therefore, from Cases I and II, we have

$$\theta \le C_w \le \theta + t_D/2.$$

We proceed to examine the selection of the best t_D values. Since the "power saving" period in a DRX cycle is $t_D - \tau$, we have the following inequality for t_D

$$t_D \ge \tau \quad (3.7)$$

In addition, from (3.6), we have another inequality for t_D

$$t_D \ge 2(C_w - \theta) \quad (3.8)$$

From (3.7) and (3.8), the range of possible t_D values is $[\max(\tau, 2(C_w - \theta)), \infty]$. We consider the following two cases for the optimal t_I and t_D setting.

Case III: $2\tau \ge 2(C_w - \theta)$. In this case, the maximum power saving factor P_s^* could be obtained when t_D is set to 2τ. The optimal t_I value t_I^* is derived from (3.2) and is expressed as:

$$t_I^* = \frac{\ln\left[\frac{\frac{1}{2\tau^2}(C_w-\theta)(1-e^{-2\tau\lambda_a})}{\lambda_a - \frac{\lambda_a}{\tau}(C_w-\theta)+\frac{1}{2\tau^2}(C_w-\theta)(1-e^{-2\tau\lambda_a})}\right]}{-\lambda_a} \quad (3.9)$$

[1] 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 $t_D - \tau$.

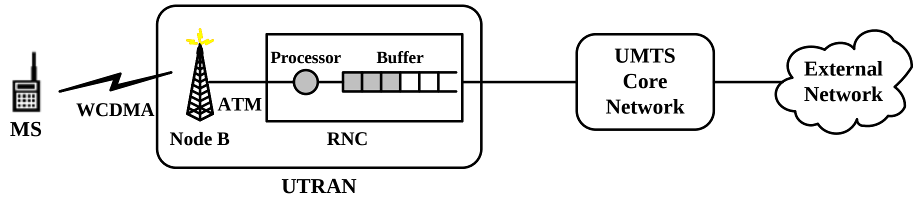


Figure 5. A queueing model for UMTS DRX

Case IV: $2\tau < 2(C_w - \theta)$**.** From (3.4), we know the largest power saving factor is achieved when t_D is set to $2(C_w - \theta)$. In this case, t_I is set to 0, and the corresponding C_w is the maximum mean packet waiting time for the given $t_D = 2(C_w - \theta)$ (see the explanation in Case II).

Based on the above discussion, the best $[t_I, t_D]$ pair that maximizes the power saving factor P_s under the given mean packet waiting time constraint C_w is either $[t_I^*, 2\tau]$ or $[0, 2(C_w - \theta)]$.

In a real UMTS network, the traffic pattern of a mobile user usually changes from time to time. In order to capture user traffic pattern more accurately, one may dynamically adjust the t_I and t_D values to enhance the $E[t_w]$ and P_s performance. According to the above analytic analysis of the best t_I and t_D selection, we devise an adaptive algorithm to dynamically adjust t_I and t_D values by using the iterative technique [16]. Our algorithm employs a two-level nested loop structure. The outer loop is controlled by index j and adjusts the t_I value based on the power saving factor. On the other hand, the inner loop is controlled by index i and adjusts the t_D value according to the mean packet waiting time. Let $t_D(i)$ be the i-th iterate of the t_D sequence, and $t_I(j)$ be the j-th iterate of the t_I sequence. The t_D and t_I values can be dynamically adjusted using the following algorithm.

Dynamic Discontinuous Reception (DDRX) Algorithm

Step 0 - Initialization. Assign an initial value to $t_D(1)$ and $t_I(1)$, respectively. Set i to 0, j to 1, D to 1 and $P_s(0)$ to 0. Go to Step 1.

Step 1 - Execution of the DRX mechanism. $i \leftarrow i + 1$. Exercise the DRX mechanism with parameters $t_D(i)$ and $t_I(j)$ from the $[(i-1) \times M + 1]$-st packet to the $(i \times M)$-th packet and calculate the mean packet waiting time $E[t_w](i)$ of these M packets. Go to Step 2.

Step 2 - Adjustment of t_D.

if $(E[t_w](i) > C_w + \Delta_w)$ **then** $t_D(i+1) \leftarrow t_D(i) - \delta_D$ and go to Step 1.
else if $(E[t_w](i) < C_w - \Delta_w)$ **then** $t_D(i+1) \leftarrow t_D(i) + \delta_D$ and go to Step 1.
else go to Step 3.

Step 3 - Adjustment of t_I.

Compute the power saving factor $P_s(j)$ for the period when $t_I(j)$ is used to exercise the DRX mechanism.
if $P_s(j) < P_s(j-1)$ and $D = 1$ **then** $D \leftarrow 0$.
else if $P_s(j) < P_s(j-1)$ and $D = 0$ **then** $D \leftarrow 1$.
if $D = 1$ **then** $t_I(j+1) \leftarrow t_I(j) + \epsilon_I$.
else if $D = 0$ **then** $t_I(j+1) \leftarrow t_I(j) - \epsilon_I$.
$j \leftarrow j + 1$ and go to Step 1.

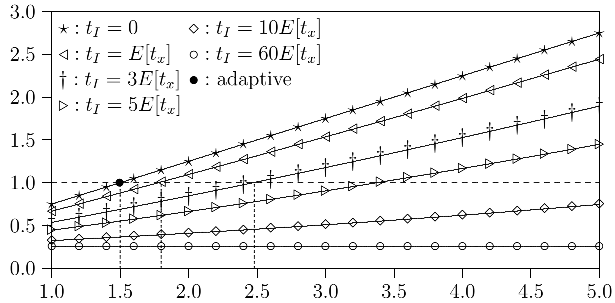
At Step 1, the algorithm observes the previous M consecutive packets and calculates their mean packet waiting time $E[t_w](i)$ for the i-th iteration. Step 2 adjusts the t_D value so that the $E[t_w](i)$ value converges to the given mean packet waiting time constraint C_w. Note that at Step 2, Δ_w is used to test if the computation converges. At Step 3, the algorithm attempts to enhance the power saving factor by adjusting the t_I value. Step 3 uses the variable D to determine the changing direction (i.e., increasing or decreasing) of t_I value.

4. Numerical Examples

Based on the analytic results presented in the last section, this section examines the effects of input parameters λ_a, τ and V_x, and mixed traffic pattern on the best t_I and t_D selection. Then by conducting discrete-event simulation experiments, we investigate the performance of our DDRX algorithm and show that DDRX can capture the best $[t_I, t_D]$ pair to maximize the MS power saving. This simulation model is similar to the one we developed in [13], 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

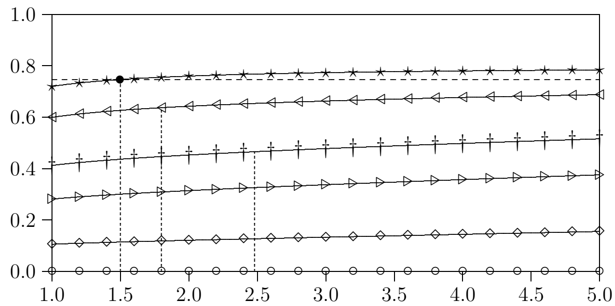
$$f_x^*(s) = \left(\frac{\lambda_x \gamma}{s + \lambda_x \gamma}\right)^{\gamma} \quad \text{where} \quad \gamma = \frac{1}{V_x \lambda_x^2}.$$

The Gamma distribution is often used in mobile telecommunications network modeling [6–8]. It has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions (see



t_D (unit: $E[t_x]$)

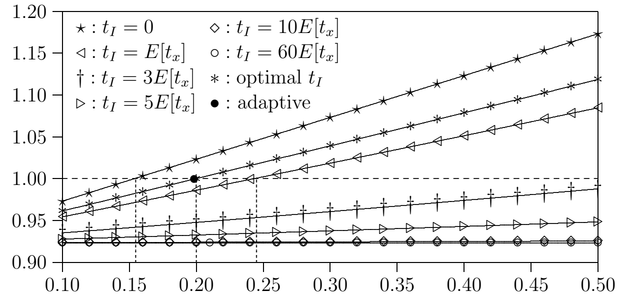
(a) $E[t_w]$ (unit: $E[t_x]$)



t_D (unit: $E[t_x]$)

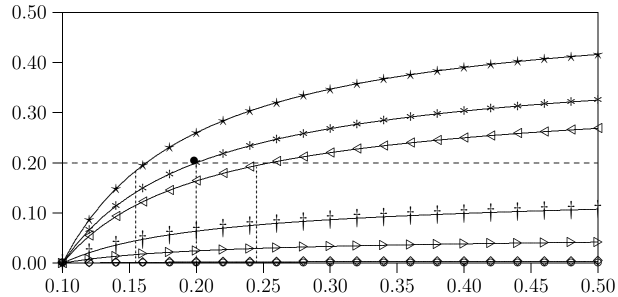
(b) P_s

Figure 6. $\lambda_a = 0.2\lambda_x$, $\tau = 0.1E[t_x]$ and $V_x = \frac{1}{\lambda_x^2}$



t_D (unit: $E[t_x]$)

(a) $E[t_w]$ (unit: $E[t_x]$)



t_D (unit: $E[t_x]$)

(b) P_s

Figure 7. $\lambda_a = 0.48\lambda_x$, $\tau = 0.1E[t_x]$ and $V_x = \frac{1}{\lambda_x^2}$

Lemma 3.9 in [11]). 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.

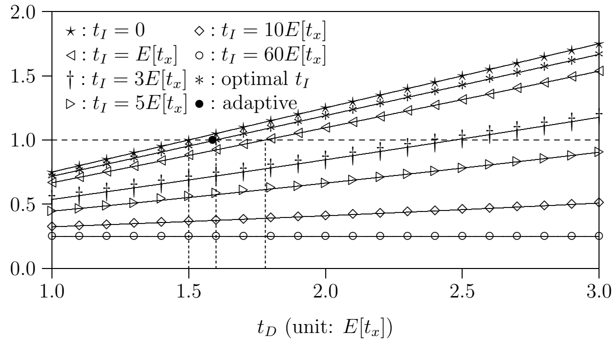
Figures 6-10 plot the $E[t_w]$ and P_s curves. In these figures, t_x has the Gamma distribution with variance V_x. The parameter settings are described in the captions of the figures. Specifically, we consider the mean packet waiting time constraint $C_w = E[t_x]$.

Effects of λ_a. Figures 6 and 7 plot the $E[t_w]$ and P_s curves for $\lambda_a = 0.2\lambda_x$ and $\lambda_a = 0.48\lambda_x$, respectively. When $\lambda_a = 0.2\lambda_x$, 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 the resulting $E[t_w]$ is small. To achieve the constant mean packet waiting time $C_w = E[t_x]$ under such circumstances, the DRX cycle length t_D has to be set to at least $2(C_w - \theta) = 1.5E[t_x]$. From Figure 6(b), we observe that the largest power saving factor P_s is achieved when $t_D = 1.5E[t_x]$ and $t_I = 0$ (see Case IV: $2\tau < 2(C_w - \theta)$ in the last section). On the other hand, as λ_a increases to $0.48\lambda_x$, t_D could be set to $2\tau = 0.2E[t_x]$ to obtain the maximum power saving factor P_s^* (see Case III: $2\tau \geq 2(C_w - \theta)$ in the last section).

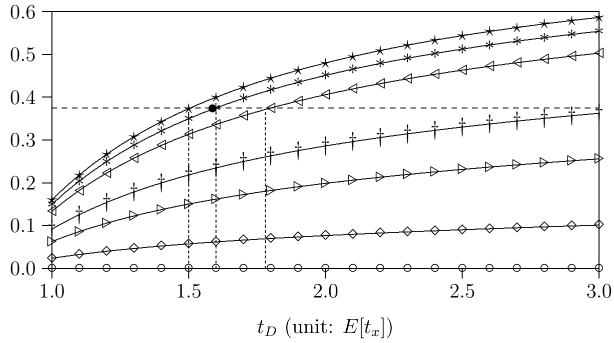
Effects of τ. Figures 6 and 8 show how the cost of wakeup τ affects the selection of t_I and t_D. When τ is small, e.g. $\tau = 0.1E[t_x]$ (see Figure 6), $t_D = 2\tau$ can not satisfy the mean packet waiting time constraint $C_w = E[t_x]$. In this case, the largest power saving factor P_s results from the setting $t_I = 0$ and $t_D = 1.5E[t_x]$. On the other hand, when τ is increased to $0.8E[t_x]$ (see Figure 8), we have $2\tau = 1.6E[t_x] > 1.5E[t_x]$. In this case, t_D could be set to $2\tau = 1.6E[t_x]$ to maximize the power saving factor P_s.

Effects of V_x. Figures 6 and 9 show how the variance V_x of the packet transmission delay t_x affects the selection of t_I and t_D. Note the well known effect in queueing systems that the mean waiting time θ of an $M/G/1$ queue is an increasing function of V_x. When V_x is small, e.g. $V_x = \frac{1}{\lambda_x^2}$ (see Figure 6), θ is small as well and we have $2(C_w - \theta) > 2\tau$. In this case, the best setting is $t_I = 0$ and $t_D = 2(C_w - \theta) = 1.5E[t_x]$. In contrast, when V_x is large, e.g. $V_x = \frac{6.5}{\lambda_x^2}$ (see Figure 9), the setting $t_D = 2\tau = 0.2E[t_x]$ results in the maximum power saving factor P_s^*.

Effects of mixed traffic. In general, the packet traffic to an MS is not homogeneous. In a typical scenario, an MS may experience with, for example, two or three multimedia messages in three or four hours, and then intensive Internet accesses for 20 minutes. In Figure 10, we consider a mixed traffic during an observation period where

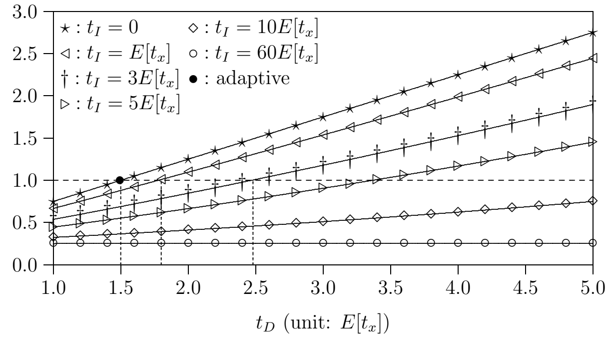


(a) $E[t_w]$ (unit: $E[t_x]$)

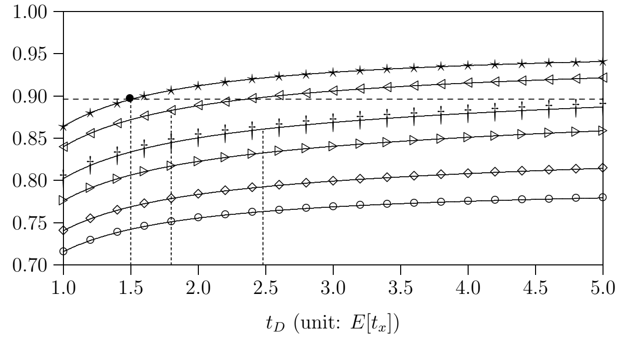


(b) P_s

Figure 8. $\lambda_a = 0.2\lambda_x$, $\tau = 0.8E[t_x]$ and $V_x = \frac{1}{\lambda_x^2}$

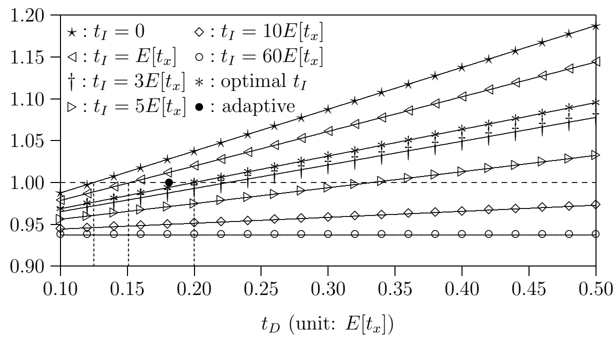


(a) $E[t_w]$ (unit: $E[t_x]$)

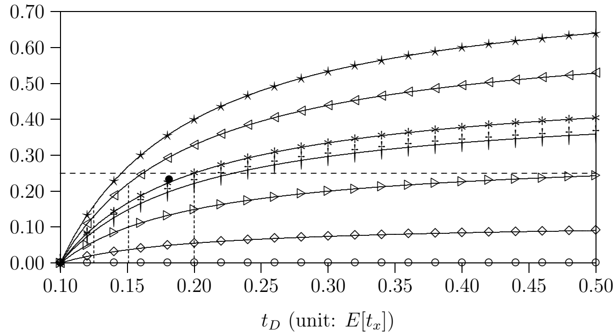


(b) P_s

Figure 10. Effects of mixed traffic ($\tau = 0.1E[t_x]$ and $V_x = \frac{1}{\lambda_x^2}$)



(a) $E[t_w]$ (unit: $E[t_x]$)



(b) P_s

Figure 9. $\lambda_a = 0.2\lambda_x$, $\tau = 0.1E[t_x]$ and $V_x = \frac{6.5}{\lambda_x^2}$

$\lambda_a = 10^{-4}\lambda_x$ for 80% of the period and $\lambda_a = 0.2\lambda_x$ for 20% of the period. We note that for the maximum power saving factor P_s^* to exist at $t_D = 2\tau$, θ must be no less than $(C_w - \tau) = E[t_x] - 0.1E[t_x] = 0.9E[t_x]$ (see Case III: $2\tau \geq 2(C_w - \theta)$ in the last section). Both of the high traffic load and low traffic load cases considered in Figure 10 result in θ less than $0.9E[t_x]$. Therefore, from Case IV: $2\tau < 2(C_w - \theta)$, the largest MS power saving for this mixed traffic is when t_I value is set to 0, as indicated in Figure 10.

Performance of DDRX. The $E[t_w]$ and P_s curves in Figures 6-10 indicate that optimal $[t_I, t_D]$ pairs exist to maximize the MS power saving. However, the optimal t_I and t_D values are not the same for different traffic patterns. Therefore, a mechanism that automatically selects the optimal t_I and t_D values in real time is required. The DDRX algorithm proposed in this paper serves for this purpose. In Figures 6-10, the "•" points represent the ($E[t_D(i)]$,$E[t_w]$) and ($E[t_D(i)]$,P_s) pairs of DDRX. These points indicate that DDRX nicely captures the user traffic patterns and always adjusts the t_I and t_D close to the optimal values.

5. Conclusions

This paper investigated the UMTS discontinuous reception (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. Based on our previous $M/G/1$ queueing model with vacations for UMTS DRX, analytic analysis and simulation model were proposed to study the optimal t_I and t_D selections that maximize the MS power saving under the given mean packet waiting time constraint. We also devised an adaptive algorithm called dynamic DRX (DDRX). This algorithm dynamically adjusts the t_I and t_D values to enhance the performance of UMTS DRX. Several numerical examples were presented to quantitatively show how to select the best t_I and t_D values for various traffic patterns. We also showed that DDRX nicely captures the user traffic patterns, and always adjusts the t_I and t_D close to the optimal values.

Acknowledgment

The author would like to express his sincere thanks to Prof. Hui-Nien Hung for his valuable discussions on this paper. He would also like to thank the three anonymous reviewers. Their comments have significantly enhanced the quality of this paper. This work was supported in part by the Mediatek Fellowship, the National Science Council (NSC) under Contract NSC-93-2219-E-007-002, and by the NSC Program for Promoting Academic Excellence of Universities NSC-94-2752-E-007-003-PA.

References

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

[2] 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.

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

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

[5] CDPD Forum, Cellular Digital Packet Data System Specification: Release 1.1. Technical report, CDPD Forum, Inc., January 1995.

[6] I. Chlamtac, Y. Fang, and H. Zeng, Call Blocking Analysis for PCS Networks under General Cell Residence Time. *IEEE Wireless Communications and Networking Conference (WCNC), New Orleans*, September 1999.

[7] Y. Fang and I. Chlamtac, Teletraffic Analysis and Mobility Modeling for PCS Networks. *IEEE Transactions on Communications*, 47(7):1062–1072, July 1999.

[8] 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 Transactions on Parallel and Distributed Systems*, 11(6):615–624, June 2000.

[9] H. Holma and A. Toskala, *WCDMA for UMTS*. John Wiley & Sons, Inc., 2000.

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

[11] F.P. Kelly, *Reversibility and Stochastic Networks*. John Wiley & Sons, 1979.

[12] S.J. Kwon, Y.W. Chung, and D.K. Sung, Queueing Model of Sleep-Mode Operation in Cellular Digital Packet Data. *IEEE Transactions on Vehicular Technology*, 52(4):1158–1162, July 2003.

[13] Y.-B. Lin, Estimating the Likelihood of Success of Lazy Cancellation in Time Warp Simulations. *International Journal in Computer Simulation*, 6(2):163–174, 1996.

[14] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. John Wiley & Sons, Inc., 2001.

[15] Y.-B. Lin and Y.-M. Chuang, Modeling the Sleep Mode for Cellular Digital Packet Data. *IEEE Communications Letters*, 3(3):63–65, March 1999.

[16] M.J. Maron and R.J. Lopez, *Numerical Analysis A Practical Approach , 3rd ed.* Wadsworth Publishing Company, 1991.

[17] RAM Mobile Data, Mobitex Interface Specification. Technical report, RAM Mobile Data, 1994.

[18] A.K. Salkintzis and C. Chamzas, An In-Band Power-Saving Protocol for Mobile Data Networks. *IEEE Transactions on Communications*, 46(9), September 1998.

[19] A.K. Salkintzis and C. Chamzas, Performance Analysis of a Downlink MAC Protocol with Power-Saving Support. *IEEE Transactions on Vehicular Technology*, 49(3), May 2000.

[20] A.K. Salkintzis and C. Chamzas, An Outband Paging Protocol for Energy-Efficient Mobile Communications. *IEEE Transactions on Broadcasting*, 48(3), September 2002.

[21] H. Takagi, *Queueing Analysis - Volume 1: Vacation and Priority Systems, Part 1*. Elsevier Science Publishers B.V., 1991.

[22] S.-R. Yang and Y.-B. Lin, Performance Evaluation of Location Management in UMTS. *IEEE Transactions on Vehicular Technology*, 52(6):1603–1615, November 2003.

[23] S.-R. Yang and Y.-B. Lin, Modeling UMTS Discontinuous Reception Mechanism. *IEEE Transactions on Wireless Communications*, 4(1):312–319, January 2005.

Shun-Ren Yang was born in Taichung, Taiwan, R.O.C., in 1976. He received the BSCSIE, MSCSIE, and Ph.D. degrees from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1998, 1999, and 2004, respectively.

From April 1, 2004 to July 31, 2004, he was

appointed as a Research Assistant in the Department of Information Engineering, the Chinese University of Hong Kong. Since August 2004, he has been with the Department of Computer Science and Institute of Communications Engineering, National Tsing Hua University, Taiwan, where he is an Assistant Professor. His current research interests include design and analysis of personal communications services networks, computer telephony integration, mobile computing, and performance modeling. E-mail: sryang@cs.nthu.edu.tw.