



# An Improved GGSN Failure Restoration Mechanism for UMTS

PHONE LIN \* and GUAN-HUA TU

Dept. Comp. Sci. & Info. Engr., National Taiwan University, Taipei, R.O.C.

**Abstract.** *Universal mobile telecommunications system (UMTS)* provides packet-switched data services for mobile users. To efficiently deliver packets in the UMTS core network, the PDP contexts (i.e., the routing information) are maintained in the volatile storage (e.g., memory) of SGSN, GGSN, and UE. The GGSN routes packets between the UMTS core network and external data networks, and thus has heavy traffic and computation loading, which may result in PDP contexts lost or corrupted, and the QoS of the UMTS network may degrade significantly. To resolve this issue, 3GPP 23.007 proposes a mechanism for GGSN failure restoration. In this mechanism, the corrupted PDP contexts can be restored through the PDP Context Activation procedure. However, this incurs extra signaling cost to the network. To reduce the network signaling cost and delay for restoration of the corrupted PDP contexts, this paper proposes an improved mechanism “GGSN Failure Restoration” (GFR) with different backup algorithms. The analytic models and simulation experiments are conducted to evaluate GFR. Our study indicates that the GFR mechanism can significantly reduce the cost for the PDP context restoration.

**Keywords:** failure restoration, GGSN, UMTS

## 1. Introduction

3GPP proposed the 3rd generation wireless system, *Universal mobile telecommunications system (UMTS)* [6,8,10], to provide high speed transmission services for *packet-switched (PS)* domain and *circuit-switched (CS)* domain services for mobile users. Figure 1 illustrates the UMTS network architecture. UMTS consists of the *UMTS terrestrial radio access network (UTRAN)* and the core network. In this architecture, a *user equipment (UE)* communicates with the UTRAN through the air interface *Uu* [2]. In the core network, the *serving GPRS support node (SGSN)* delivers the packets between the UEs and their counter-parts in the external data network. The *gateway GPRS support node (GGSN)* interworks with the external data network using connectionless network protocols (e.g., Internet protocol) or connection-oriented network protocols (e.g., X.25). The GGSN is connected with SGSNs via an IP-based GPRS backbone network. The mobility databases HLR and VLR maintain the location information for mobile users. The MSC provides the circuit-switched services for mobile users. To efficiently deliver the packets in the core network, the *packet data protocol (PDP)* contexts (i.e., the routing information) are maintained in the volatile storage (e.g., memory) of SGSN, GGSN, and UE. The details of the contents in a PDP context can be found in [5,6]. Before a UE starts a connection to the external data network, the PDP Context Activation procedure is invoked to establish a session from the UE to the application server in the external data network through SGSN and GGSN. At this moment, the PDP contexts are created in GGSN, SGSN, and UE for the session, respectively. If the

QoS requirement for the activated session is changed, the PDP Context Modification procedure is executed to modify the PDP contexts. When the user terminates the connection, the network exercises the PDP Context Deactivation procedure to deactivate the PDP contexts in GGSN, SGSN, and UE. For the details of the three procedures, readers may refer to [6].

In UMTS Release 4 [3] and UMTS Release 5 [7], 3GPP evolved the UMTS system to the all-IP architecture which integrates the IP and the wireless technologies. In this architecture, the traffic for the CS domain services are packetized and delivered through the UMTS core network so that the network can be utilized more efficiently. Obviously, the traffic and computation loading of the GGSN in the core network increases in the UMTS system, and its stability affects the performance of the UMTS network significantly. As mentioned previously, the PDP contexts in the GGSN are referred to efficiently route the packet traffic. Due to the increasing reference rate of the PDP context, an activated PDP context is more likely to be lost or corrupted. When an activated PDP context is lost or corrupted, the corresponding session is interrupted, and the application (run between the UE and the server in the external data network) is terminated. The QoS of the UMTS network degrades significantly. To resolve this issue, 3GPP 23.007 [4] defines the necessary fault tolerance mechanism for GGSN, which is known as the “Basic” mechanism in this paper. In the Basic mechanism, the corrupted PDP contexts can be restored through the basic restoration procedure or the GGSN restart procedure, whose details are illustrated in Section 2. In the restoration procedure or the GGSN restart procedure, the PDP Context Activation procedure is invoked to reactivate the corrupted PDP context, which results in extra signaling cost to the network, or unexpected delay. Specifically, at least 8 signaling messages should be exchanged in

\* Corresponding author.  
E-mail: plin@csie.ntu.edu.tw

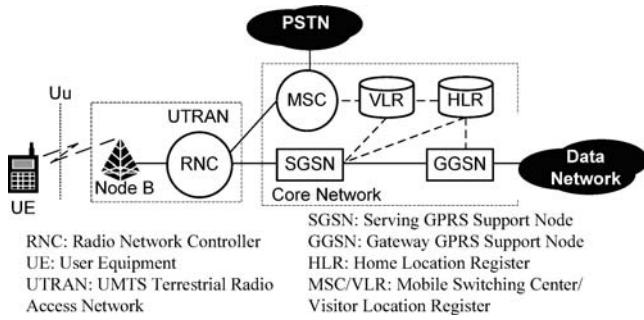


Figure 1. The UMTS network architecture.

the UMTS network for the restoration of a corrupted PDP context. Typically, a GGSN can support more than 10,000 simultaneously activated data sessions; e.g., Lucent’s GGSN implementation [9] can accommodate 40,000 ongoing data sessions. In other words, a GGSN failure may cause more than 10,000 ongoing data sessions to be interrupted, and at least 80,000 message exchanges are required to restore the corrupted PDP contexts.

To reduce the network signaling cost and delay for the restorations of the corrupted PDP contexts, this paper proposes an improved mechanism GFR (GGSN Failure Restoration) with three backup algorithms, GBA1, GBA2, and GBA3. The paper is organized as follows. In Section 2, the Basic mechanism in 3GPP 23.007 is briefly described. Section 3 describes the GFR mechanism, and the GBA1, GBA2, and GBA3 backup algorithms. Section 4 proposes analytical mod-

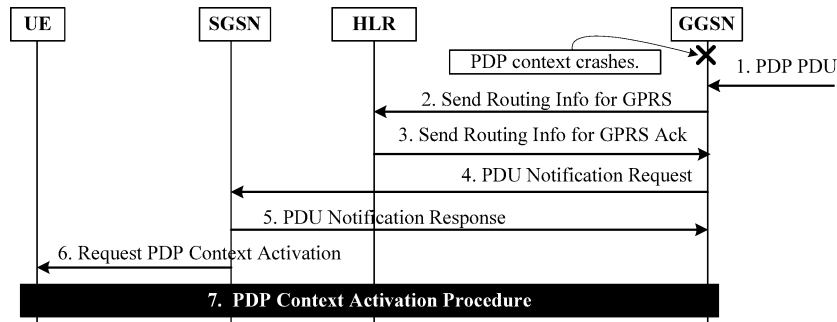
els for the GFR mechanism with different backup algorithms. Section 5 evaluates the performances for the Basic mechanism in 3GPP and the GFR mechanism with different backup algorithms. Section 6 gives a concluding remark.

## 2. The basic mechanism

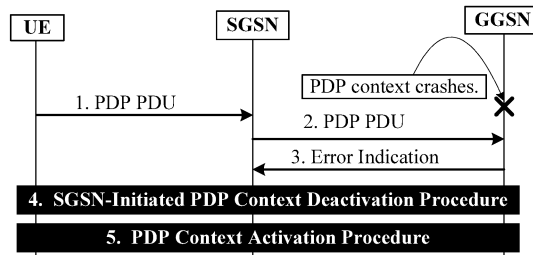
This section describes the GGSN failure restoration mechanism proposed in 3GPP 23.007, where the corrupted PDP contexts are restored through the restoration procedure or the GGSN restart procedure. The details are given as follows.

**The Restoration Procedure:** Figure 2 illustrates the message flow for this procedure. When an activated PDP context in the GGSN is corrupted, it is restored when one of the two following events occurs in the GGSN:

**Event 1.** The GGSN receives a mobile terminated *protocol data unit* (PDU; from the external data network) for which no valid PDP context exists. If the GGSN has the static PDP information for the PDP address of the PDU (i.e., the GGSN can retrieve the IP address of the SGSN that currently serves the mobile user; see Steps 2–3 in figure 2(a)), the GGSN initiates the Network-Requested PDP Context Activation procedure [6] before delivering the PDU (see Steps 4–7 in figure 2(a)). Otherwise (i.e., the GGSN does not have the static PDP information), the GGSN discards the received PDU and returns an appropriate Error message depending on the protocol used between the GGSN and the external data network. Note that in the latter case, the corrupted PDP context cannot be restored, and the application is terminated.



(a) Evnet 1



(b) Evnet 2

Figure 2. The message flow for the restoration procedure.

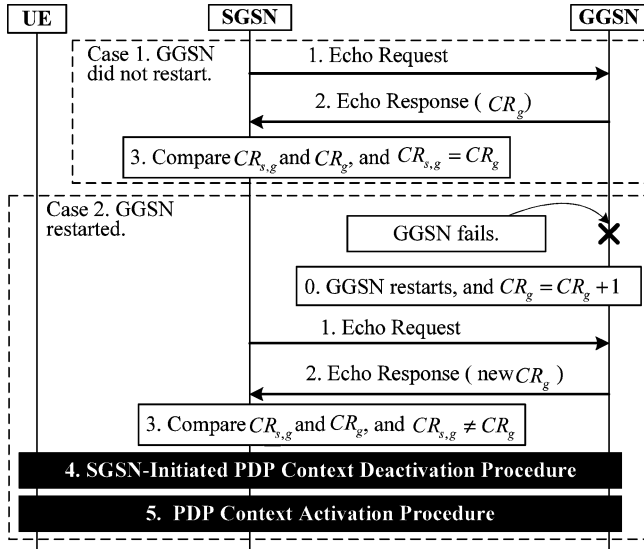


Figure 3. The message flow for the GGSN restart procedure.

**Event 2.** The GGSN receives a tunnel PDU (from the SGSN) for which no PDP context exists. The GGSN discards the tunnel PDU and sends an Error Indication message [5] to the originating SGSN (see Step 3 in figure 2(b)). Then the originating SGSN deactivates the PDP context and sends an Error Indication message to the originating UE (see Step 4 in figure 2(b)). The UE performs the PDP Context Activation procedure [6] to reactivate the PDP context (see Step 5 in figure 2(b)).

**The GGSN Restart Procedure:** Figure 3 illustrates the message flow for this procedure. In this procedure, the SGSN detects if the GGSN has been restarted through the polling function. As shown in Case 1 in figure 3, if no restart has been performed by the GGSN, then the SGSN takes no actions. Otherwise (i.e., the SGSN detects a restart in a GGSN with which the SGSN has one or more activated PDP contexts; see Case 2 in figure 3), the SGSN deactivates all these PDP contexts and then requests the UEs to reactivate them. In the GGSN, a GGSN Restart counter (denoted as  $CR_g$ ) is maintained in the non-volatile storage (i.e., when the GGSN is powered off, the value of  $CR_g$  is not destroyed) to count the number of restart actions that have been performed on the GGSN. After the GGSN is restarted, the  $CR_g$  counter is incremented by one (see Step 0, Case 2, figure 3). Each SGSN (with which the GGSN is in contact) maintains a GGSN Restart counter (denoted as  $CR_{s,g}$ ) to record the number of restarts that have been exercised on the GGSN. The polling function consists of three steps:

**Step 1.** The SGSN sends an Echo Request message [5] to the GGSN, and starts a timer  $T_p$ . It expects to receive an Echo Response message [5] from GGSN before  $T_p$  expires. If the  $T_p$  timer expires, the SGSN recognizes that the polled GGSN fails.

**Step 2.** Upon receipt of Echo Request, the GGSN responds the SGSN an Echo Response message in which the  $CR_g$  value is included.

**Step 3.** When the SGSN receives Echo Response, it stops the  $T_p$  timer and compares the values of  $CR_g$  and  $CR_{s,g}$ . If  $CR_{s,g} = CR_g$  (see Step 3, Case 1, figure 3), the SGSN recognizes that the GGSN has not restarted before the GGSN received the Echo Request message, and the SGSN takes no action. Otherwise (i.e.,  $CR_{s,g} \neq CR_g$ ; see Step 3, Case 2, figure 3), the SGSN recognizes that the GGSN has restarted before the GGSN received the previous Echo Request message from SGSN. The SGSN deactivates all the PDP contexts related to the GGSN and sets  $CR_{s,g} \leftarrow CR_g$ . The SGSN waits for a time period, and then returns to Step 1.

### 3. An improved GGSN failure restoration mechanism

This section proposes an improved mechanism GFR (GGSN Failure Restoration) with three different backup algorithms, GBA1, GBA2, and GBA3. The details of GFR are given below.

During an activated session, when the GGSN receives or sends a PDU (which may contain user data or network signaling message) from or to the SGSN, the corresponding fields in the PDP context are modified. In GFR, the GGSN backs up the PDP context into the non-volatile storage (e.g., hard disk) at some checkpoints by using the backup algorithms (to be elaborated later). A flag  $D$  (that is backed up into the non-volatile storage when being changed) is maintained in the volatile storage to indicate if the backed up PDP context is valid. If  $D = 0$ , the PDP context in the non-volatile storage is consistent with that in the volatile storage. Otherwise (i.e.,  $D = 1$ ), the PDP context in the non-volatile storage is inconsistent with that in the volatile storage. Initially,  $D$  is set to 0. During an activated session,  $D$  is maintained as follows.

- When the PDP context is modified (i.e., new PDU arrives),  $D$  is set to 1.
- When the PDP context is backed up into the non-volatile storage,  $D$  is set to 0.
- When the  $D$  value is changed from 1 to 0 or changed from 0 to 1, the GGSN writes the  $D$  flag into the non-volatile storage.

The basic restoration procedure and the GGSN restart procedure are modified as follows.

**The Restoration Procedure** is similar to that in the Basic mechanism. The difference is that when the GGSN receives a mobile terminated PDU (from the external network) or a tunnel PDU (from the SGSN) for which no valid PDP context exists in the volatile storage, the GGSN restores the backed up PDP context and its corresponding  $D$  flag into the volatile storage. If  $D = 0$ , the corrupted PDP context

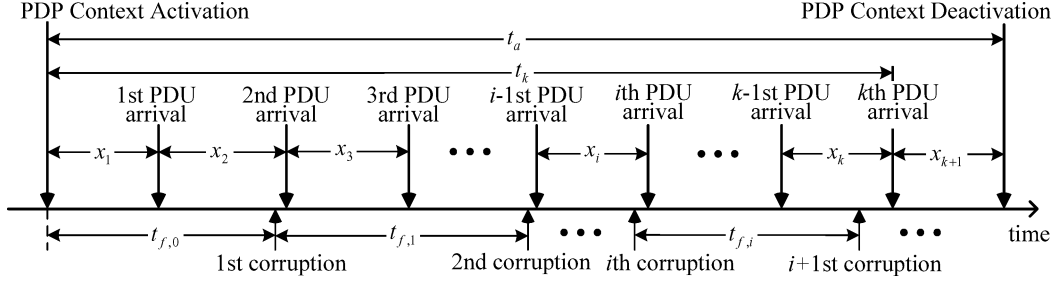


Figure 4.  $x_i$ ,  $t_{f,i}$ ,  $t_a$ , and  $t_k$ .

is restored. Otherwise (i.e.,  $D = 1$ ), the UE is requested to reactivate the PDP context.

**The GGSN Restart Procedure** is similar to that in the Basic mechanism, except that after the GGSN restarts, instead of incrementing the  $CR_g$  counter, it restores all the backed up PDP contexts and the corresponding  $D$  flags into the volatile storage.

To determine the checkpoints when the GGSN backups the PDP context into the non-volatile storage, we propose three backup algorithms, GBA1, GBA2, and GBA3. Consider the timing diagram in figure 4. Let  $t_a$  denote the duration time of a session. For  $i \geq 1$ , let  $x_i$  be the period between the time when the  $i - 1$ st PDU arrives and the time when the  $i$ th PDU arrives. By convention,  $x_0 = 0$ , and the 0th PDU arrival is the PDU that activates the PDP context for the session. Figure 5 illustrates the timing diagrams for the three backup algorithms.

**GBA1:** See figure 5(a). The GGSN backups the PDP context into the non-volatile storage for every time interval  $T_{G1}$ .

**GBA2:** See figure 5(b). The GGSN maintains a counter  $CR_{G2}$  for the PDP context. Initially,  $CR_{G2}$  is set to 0. When the PDP context is modified (i.e., a PDU arrives),  $CR_{G2}$  is incremented by one, and then compared with a threshold value  $K_{G2}$ . If  $CR_{G2} < K_{G2}$ , GGSN takes no actions. If  $CR_{G2} = K_{G2}$ , the PDP context is backed up into the non-volatile storage, and  $CR_{G2}$  is reset to 0.

**GBA3:** See figure 5(c). When the PDP context is created for a session, a timer  $T_{G3}$  starts. When the PDP context is modified, the GGSN restarts  $T_{G3}$ . When the  $T_{G3}$  timer expires (i.e., no PDU arrives during the time period of  $T_{G3}$ ), the PDP context is backed up to the non-volatile storage.

The proposed GFR mechanism can be deployed in the UMTS system without introducing any new component and modifying the protocol of the UMTS network. This mechanism is considered practical and cost-effective.

#### 4. Analytical models

This section proposes analytic models for the Basic mechanism and the GFR mechanism with GBA1, GBA2, and GBA3. The notation is listed in Appendix A.

Consider the timing diagram in figure 4. In a session, suppose that in a GGSN, when a PDU arrives, with probability  $1 - \alpha$ , the arrival PDU deactivates the PDP context, and the session ends, and with probability  $\alpha$ , the session continues. Assume that the PDU inter-arrival times  $x_i$  are i.i.d. with exponential density function  $f_x(x_i) = \mu_x e^{-\mu_x x_i}$  and mean  $1/\mu_x$ . Exponential periods are used in the analytic models to provide the mean value analysis, and indicate the performance trends for our mechanism and the Basic mechanism. The effects of higher moments for general distribution are studied based on our simulation experiments.

Let  $t_a$  be the elapsed time of a session. Then the density function  $f_a(t_a)$  for  $t_a$  can be expressed as

$$f_a(t_a) = \sum_{k=0}^{\infty} \left[ \int_{x_1=0}^{t_a} \int_{x_2=0}^{t_a-x_1} \int_{x_3=0}^{t_a-x_1-x_2} \dots \right. \\ \times \int_{x_k=0}^{t_a-x_1-x_2-x_3-\dots-x_{k-1}} \alpha^k (1-\alpha) \left( \prod_{i=1}^k \mu_x e^{-\mu_x x_i} \right) \\ \left. \times \mu_x e^{-\mu_x (t_a-x_1-x_2-\dots-x_k)} dx_k dx_{k-1} dx_{k-2} \dots dx_1 \right] \quad (1)$$

and its Laplace transform is

$$f_a^*(s) = \sum_{k=0}^{\infty} \left( \frac{\alpha \mu_x}{s + \mu_x} \right)^k \left[ \frac{(1-\alpha) \mu_x}{s + \mu_x} \right] \\ = \frac{(1-\alpha) \mu_x}{s + (1-\alpha) \mu_x} \quad (2)$$

As shown in figure 4, for  $i \geq 1$ , let  $t_{f,i}$  denote the time interval between the time when the  $i$ th PDP context corruption occurs and the time when the  $i + 1$ st PDP context corruption occurs,  $t_{f,0}$  denotes the interval between the time when the PDP context is activated and the time when the first PDP context corruption occurs in the session. For all  $i \geq 0$ ,  $t_{f,i}$  are assumed to be i.i.d. exponential random variables with the exponential density function  $f_f(t_{f,i}) = \eta e^{-\eta t_{f,i}}$ . Let  $N_f$  be the number of PDP context corruptions in a session. Then

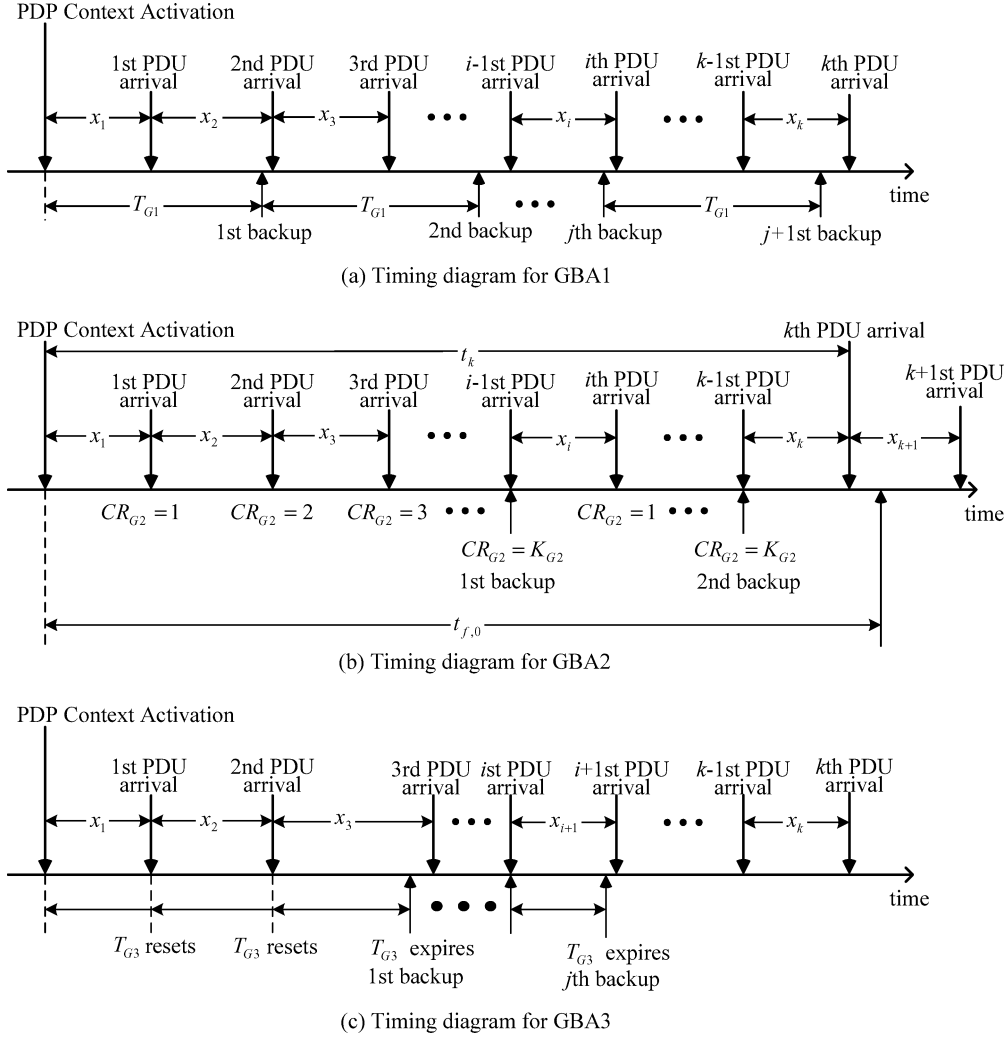


Figure 5. Timing diagrams for GBA1, GBA2, and GBA3.

we have

$$\begin{aligned} \Pr[N_f = n] &= \int_{t_a=0}^{\infty} \left[ \frac{(\eta t_a)^n}{n!} \right] e^{-\eta t_a} f_a(t_a) dt_a \\ &= \left( \frac{\eta^n}{n!} \right) \left[ (-1)^n \frac{d^n f_a^*(s)}{ds^n} \right] \Big|_{s=\eta} \end{aligned} \quad (3)$$

From Appendix B, we have

$$\frac{d^n f_a^*(s)}{ds^n} = \frac{(-1)^n n! (1 - \alpha) \mu_x}{[s + (1 - \alpha) \mu_x]^{n+1}}$$

and (3) is rewritten as

$$\Pr[N_f = n] = \left[ \frac{(1 - \alpha) \mu_x}{\eta + (1 - \alpha) \mu_x} \right] \left[ \frac{\eta}{\eta + (1 - \alpha) \mu_x} \right]^n$$

The expected value of  $N_f$  can be obtained as follows.

$$\begin{aligned} E[N_f] &= \sum_{n=1}^{\infty} n \Pr[N_f = n] \\ &= \sum_{n=1}^{\infty} \left\{ n \left[ \frac{(1 - \alpha) \mu_x}{\eta + (1 - \alpha) \mu_x} \right] \left[ \frac{\eta}{\eta + (1 - \alpha) \mu_x} \right]^n \right\} \\ &= \frac{\eta}{(1 - \alpha) \mu_x} \end{aligned} \quad (4)$$

For an activated session, there are at least two PDU arrivals. One PDU is for the PDP context activation, and the other is for the PDP context deactivation. Let  $N_m$  be the number of the PDU arrivals (including the PDU arrival that activates the PDP context but excluding the last PDU arrival that deactivates the PDP context) in a session. Then

$$\Pr[N_m = n] = \alpha^{n-1} (1 - \alpha) \quad \text{where } n \geq 1$$

and its expected value is

$$E[N_m] = \sum_{n=1}^{\infty} [n\alpha^{n-1}(1-\alpha)] = \frac{1}{1-\alpha} \quad (5)$$

For  $k \geq 1$ , let  $t_k$  be the duration time of the first  $k$  PDU arrivals (i.e.,  $t_k = x_1 + x_2 + x_3 + \dots + x_k$ ) [11]. By convention,  $t_0 = 0$ . The density function  $f_k(t_k)$  of  $t_k$  is expressed as

$$f_k(t_k) = \begin{cases} 0, & \text{if } k = 0 \\ \int_{x_1=0}^{t_k} \int_{x_2=0}^{t_k-x_1} \int_{x_3=0}^{t_k-x_1-x_2} \dots \\ \int_{x_k=0}^{t_k-x_1-x_2-x_3-\dots-x_{k-1}} \times \alpha^k \left( \prod_{i=1}^k \mu_x e^{-\mu_x x_i} \right) dx_k dx_{k-1} dx_{k-2}, \dots dx_1, & \text{if } k > 0 \end{cases} \quad (6)$$

and its Laplace transform is

$$f_k^*(s) = \left( \frac{\alpha \mu_x}{s + \mu_x} \right)^k \quad (7)$$

Let  $P_{G,f}$  be the probability that the PDP context is corrupted in a session. As shown in figure 4,  $P_{G,f}$  can be derived as follows:

$$\begin{aligned} P_{G,f} &= \sum_{k=0}^{\infty} \Pr[t_k < t_{f,0} < t_k + x_{k+1}] \\ &= \sum_{k=0}^{\infty} \left[ \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{x_{k+1}=t_{f,0}-t_k}^{\infty} \right. \\ &\quad \left. \times f_k(t_k) \eta e^{-\eta t_{f,0}} \mu_x e^{-\mu_x x_{k+1}} dx_{k+1} dt_{f,0} dt_k \right] \\ &= \left( \frac{\eta}{\eta + \mu_x} \right) \sum_{k=0}^{\infty} \left[ \int_{t_k=0}^{\infty} f_k(t_k) e^{-\eta t_k} dt_k \right] \\ &= \left( \frac{\eta}{\eta + \mu_x} \right) \sum_{k=0}^{\infty} f_k^*(\eta) \\ &= \frac{\eta}{\eta + \mu_x - \alpha \mu_x} \quad (8) \end{aligned}$$

Let  $P_f$  be the probability that the PDP context is corrupted, and the backed-up PDP context is invalid, and  $P_{(f|G,f)}$  be the conditional probability that under the condition when the PDP context is corrupted, the backed-up PDP context is invalid, and then we have

$$\begin{aligned} P_{(f|G,f)} &= [\text{Pr Backup is invalid PDP context is corrupted}] \\ &= \frac{\text{Pr[PDP context is corrupted, and backup is invalid]}}{\text{Pr[PDP context is corrupted]}} \\ &= \frac{P_f}{P_{G,f}} \quad (9) \end{aligned}$$

Let  $N_r$  be the number of the PDP context reactivations executed due to PDP context corruption in a session. Then  $N_r$  is obtained as follows

$$N_r = N_f P_{(f|G,J)} \quad \text{and} \quad E[N_r] = E[N_f] P_{(f|G,J)} \quad (10)$$

By applying  $E[N_f]$  in (4) into (10), (10) is rewritten as

$$E[N_r] = \left[ \frac{\eta}{(1-\alpha)\mu_x} \right] P_{(f|G,f)} \quad (11)$$

Let  $N_b$  be the number of the backups executed for a PDP context in a session. Suppose that the  $d_r$  cost is introduced to the network to reactivate a PDP context, and the  $d_b$  cost is required for each backup. The total cost  $d_t$  introduced in a session is expressed as

$$d_t = d_b E[N_b] + d_r E[N_r] \quad (12)$$

In the following, we derive  $E[N_b]$ ,  $P_f$ ,  $E[N_r]$ , and  $d_t$  for the Basic mechanism, and the GFR mechanism with GBA1, GBA2 and GBA3.

#### 4.1. The analytic model for the Basic mechanism

In the Basic mechanism, since no backup is exercised by the GGSN, we have

$$E[N_b] = 0 \quad (13)$$

When the PDP context is corrupted, the PDP context becomes invalid. Then

$$P_{(f|G,f)} = 1 \quad (14)$$

From (10), (14), and (4), the  $E[N_r]$  for the Basic mechanism is

$$E[N_r] = E[N_f] = \frac{\eta}{(1-\alpha)\mu_x} \quad (15)$$

By applying (13) and (15) into (12), we have the  $d_t$  for the Basic mechanism as

$$d_t = d_b E[N_b] + d_r E[N_r] = d_r \left[ \frac{\eta}{(1-\alpha)\mu_x} \right] \quad (16)$$

#### 4.2. The analytic model for GFR with GBA1

Suppose that in GFR with GBA1, the value  $y$  of the  $T_{G1}$  timer has the exponential distribution with the density function  $f_{G1}(y) = \gamma_{G1} e^{-\gamma_{G1}y}$  and mean  $1/\gamma_{G1}$ . As shown in figure 5(a), similar to the derivation of  $\Pr[N_f = n]$  in (3),  $\Pr[N_b = n]$  for the GFR with GBA1 mechanism is obtained as follows:

$$\begin{aligned} \Pr[N_b = n] &= \int_{t_a=0}^{\infty} \left[ \frac{(\gamma_{G1} t_a)^n}{n!} \right] e^{-\gamma_{G1} t_a} f_a(t_a) dt_a \\ &= \left[ \frac{(1-\alpha)\mu_x}{\gamma_{G1} + (1-\alpha)\mu_x} \right] \left[ \frac{\gamma_{G1}}{\gamma_{G1} + (1-\alpha)\mu_x} \right]^n \quad (17) \end{aligned}$$

and

$$\begin{aligned} E[N_b] &= \sum_{n=1}^{\infty} n \Pr[N_b = n] \\ &= \sum_{n=1}^{\infty} \left\{ n \left[ \frac{(1-\alpha)\mu_x}{\gamma_{G1}} + (1-\alpha)\mu_x \right] \right\} \end{aligned}$$

$$\begin{aligned}
 & \times \left[ \frac{\gamma_{G1}}{\gamma_{G1} + (1 - \alpha)\mu_x} \right]^n \Big\} \\
 & = \frac{\gamma_{G1}}{(1 - \alpha)\mu_x} \quad (18)
 \end{aligned}$$

For the analysis of  $P_f$ ,  $E[N_r]$ , and  $d_t$  for the GFR with GBA1, the derivations are too complicated and are not conducted in this paper. Instead, we use simulation experiments to analyze the three values.

#### 4.3. The analytic model for GFR with GBA2

In GFR with GBA2, in a session, the GGSN backups the PDP context for every  $K_{G2}$  PDU arrivals (see figure 5(b)), and then

$$N_b = \left\lfloor \frac{N_m}{K_{G2}} \right\rfloor$$

In this study, we obtain an approximated value for the mean of  $N_b$  by using the following equation.

$$E[N_b] = \left\lfloor \frac{E[N_m]}{K_{G2}} \right\rfloor \quad (19)$$

From (5), we have  $E[N_m] = 1/(1 - \alpha)$ , and (19) can be rewritten as

$$E[N_b] = \left\lfloor \frac{1}{(1 - \alpha)K_{G2}} \right\rfloor \quad (20)$$

Consider the timing diagram in figure 5(b). The  $P_f$  probability for GFR with GBA2 is

$$\begin{aligned}
 P_f &= \sum_{k=0}^{\infty} \Pr [\text{the PDP context is corrupted in the interval} \\
 & \quad x_{k+1}, \text{ and the PDP context is not backed up} \\
 & \quad \text{for the } k\text{th PDU arrival}] \\
 &= \sum_{k=0}^{\infty} \Pr[t_k < t_{f,0} < t_k + x_{k+1} \text{ and } k \bmod K_{G2} \neq 0] \quad (21)
 \end{aligned}$$

Then (21) can be approximately obtained by using the following equation:

$$P_f = \left(1 - \frac{1}{K_{G2}}\right) \sum_{k=0}^{\infty} \Pr[t_k < t_{f,0} < t_k + x_{k+1}] \quad (22)$$

From (8), (22) is rewritten as

$$P_f = \left(1 - \frac{1}{K_{G2}}\right) P_{G,f} \quad (23)$$

By applying (23) into (9), the  $P_{(f|G,f)}$  probability for GFR with GBA2 is

$$P(f | G, f) = P_f / P_{G,f} = 1 - \frac{1}{K_{G2}} \quad (24)$$

Apply (24) into (11), and we have

$$\begin{aligned}
 E[N_r] &= \left[ \frac{\eta}{(1 - \alpha)\mu_x} \right] \left(1 - \frac{1}{K_{G2}}\right) \\
 &= \frac{\eta(K_{G2} - 1)}{(1 - \alpha)\mu_x K_{G2}} \quad (25)
 \end{aligned}$$

By applying (20) and (25) into (12), we have the  $d_t$  cost for GFR with GBA2 as follows.

$$d_t = d_b \left[ \frac{1}{(1 - \alpha)K_{G2}} \right] + \left[ \frac{\eta(K_{G2} - 1)}{(1 - \alpha)\mu_x K_{G2}} \right] \quad (26)$$

#### 4.4. The analytic model for GFR with GBA3

In GBA3, assume that the value  $y$  of the  $T_{G3}$  timer has an exponential distribution with density function  $f_{G3}(y) = \gamma_{G3}e^{-\gamma_{G3}y}$ . Consider the timing diagram in figure 5(c). After the  $i$ th PDU arrives, the PDP context is backed up if  $x_{i+1}$  is larger than the value of  $T_{G3}$ . Let  $\beta$  be the probability that  $x_{i+1} > y$ , and then

$$\begin{aligned}
 \beta &= \Pr[x_{i+1} > y] \\
 &= \int_{y=0}^{\infty} \int_{x_{i+1}=y}^{\infty} \mu_x e^{-\mu_x x_{i+1}} \gamma_{G3} e^{-\gamma_{G3} y} dx_{i+1} dy \\
 &= \frac{\gamma_{G3}}{\mu_x + \gamma_{G3}} \quad (27)
 \end{aligned}$$

By using the GBA3 backup algorithm, for each PDU arrival (which causes the PDP context to be modified) in a session, with probability  $\beta$ , the PDP context is backed up. Then we have  $N_b = \beta N_m$  and  $E[N_b] = \beta E[N_m]$ . From (5) and (27),

$$E[N_b] = \frac{\gamma_{G3}}{(1 - \alpha)(\mu_x + \gamma_{G3})} \quad (28)$$

The  $P_f$  for GFR with GBA3 is derived as follows. As shown in figure 6, consider the time intervals  $t_k$ ,  $x_{k+1}$ ,  $y$ , and  $t_{f,0}$ . If  $t_k < t_{f,0} < t_k + x_{k+1}$  and  $x_{k+1} < y$  (see figure 6(a)), or if  $t_k < t_{f,0} < t_k + y$  and  $y < x_{k+1}$  (see figure 6(b)), then as the PDP context is corrupted, the backed up PDP context is

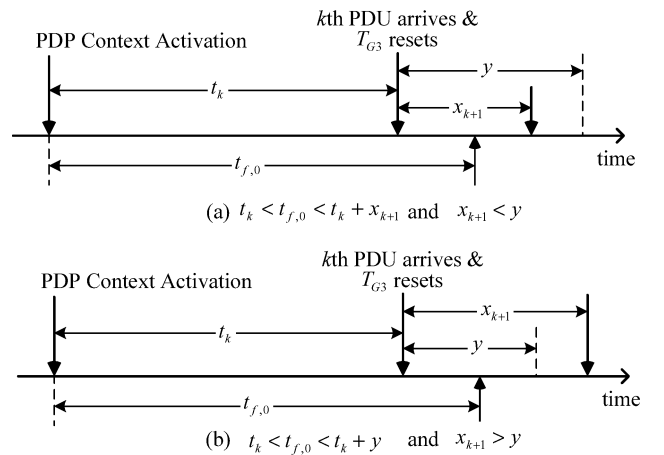


Figure 6. The timing diagram for the derivation of  $P_f$  for GBA3.

invalid.

$$\begin{aligned}
P_f &= \sum_{k=0}^{\infty} \{ \Pr[t_k < t_{f,0} < t_k + x_{k+1} \text{ and } x_{k+1} < y] \\
&\quad + \Pr[t_k < t_{f,0} < t_k + y \text{ and } y < x_{k+1}] \} \\
&= \sum_{k=0}^{\infty} \Pr[t_k < t_{f,0} < t_k + \min(x_{k+1}, y)] \quad (29)
\end{aligned}$$

Let  $z = \min(x_{k+1}, y)$  and with density function  $f_z(z)$ . We have

$$\begin{aligned}
f_z(z) &= d \left\{ 1 - \left[ \int_{x_{k+1}=z}^{\infty} \mu_x e^{-\mu_x x_{k+1}} dx_{k+1} \right] \right. \\
&\quad \left. \times \left[ \int_{y=z}^{\infty} \gamma_{G3} e^{-\gamma_{G3} y} dy \right] \right\} / dz \\
&= \mu_x e^{-(\mu_x + \gamma_{G3})z} + \gamma_{G3} e^{-(\mu_x + \gamma_{G3})z} \quad (30)
\end{aligned}$$

Replace  $\min(x_{k+1}, y)$  in (29) with  $z$ , and we obtain

$$\begin{aligned}
P_f &= \sum_{k=0}^{\infty} \Pr[t_k < t_{f,0} < t_k + z] \\
&= \sum_{k=0}^{\infty} \left[ \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \right. \\
&\quad \left. \times f_k(t_k) f_f(t_{f,0}) f_z(z) dz dt_{f,0} dt_k \right] \\
&= \sum_{k=0}^{\infty} \left\{ \left[ \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} f_k(t_k) \eta e^{-\eta t_{f,0}} [\mu_x e^{-(\mu_x + \gamma_{G3})z} \right. \right. \\
&\quad \left. \left. + \gamma_{G3} e^{-(\mu_x + \gamma_{G3})z}] dz dt_{f,0} dt_k \right] \right\} \\
&= \sum_{k=0}^{\infty} (C + D) \quad (31)
\end{aligned}$$

where

$$\begin{aligned}
C &= \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \\
&\quad \times f_k(t_k) \eta e^{-\eta t_{f,0}} \mu_x e^{-(\mu_x + \gamma_{G3})z} dz dt_{f,0} dt_k
\end{aligned}$$

and

$$\begin{aligned}
D &= \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \\
&\quad \times f_k(t_k) \eta e^{-\eta t_{f,0}} \gamma_{G3} e^{-(\mu_x + \gamma_{G3})z} dz dt_{f,0} dt_k
\end{aligned}$$

From Appendix C, we have

$$C = \left[ \frac{\eta \mu_x}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k$$

and

$$D = \left[ \frac{\eta \gamma_{G3}}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k$$

Then (31) is rewritten as

$$\begin{aligned}
P_f &= \sum_{k=0}^{\infty} \left\{ \left[ \frac{\eta \mu_x}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k \right. \\
&\quad \left. + \left[ \frac{\eta \gamma_{G3}}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k \right\} \\
&= \frac{\eta(\eta + \mu_x)}{(\eta + \mu_x + \gamma_{G3})(\eta + \mu_x - \alpha \mu_x)} \quad (32)
\end{aligned}$$

By applying  $P_f$  (obtained in (32)) and  $P_{G,f}$  (obtained in (8)) into (9), the  $P_{(f|G,f)}$  for GFR with GBA3 is obtained as follows.

$$\begin{aligned}
P_{(f|G,f)} &= P_f / P_{G,f} \\
&= \left[ \frac{\eta(\eta + \mu_x)}{(\eta + \mu_x + \gamma_{G3})(\eta + \mu_x - \alpha \mu_x)} \right] / \left( \frac{\eta}{\eta + \mu_x - \alpha \mu_x} \right) \\
&= \frac{\eta + \mu_x}{\eta + \mu_x - \gamma_{G3}} \quad (33)
\end{aligned}$$

Apply (33) into (11), and we have  $E[N_r]$  for GFR with GBA3:

$$E[N_r] = \left[ \frac{\eta}{(1 - \alpha)\mu_x} \right] \left[ \frac{\eta + \mu_x}{\eta + \mu_x + \gamma_{G3}} \right] \quad (34)$$

The  $d_t$  cost for GFR with GBA3 can be obtained by applying  $E[N_r]$  in (34) and  $E[N_b]$  in (28) into (12), that is,

$$\begin{aligned}
d_t &= d_b \left[ \frac{\gamma_{G3}}{(1 - \alpha)(\mu_x + \gamma_{G3})} \right] \\
&\quad + d_r \left[ \frac{\eta(\eta + \mu_x)}{(1 - \alpha)\mu_x(\eta + \mu_x + \gamma_{G3})} \right] \quad (35)
\end{aligned}$$

#### 4.5. Simulation validation

From (16), (26), and (35), the  $d_t$  cost for the Basic, GFR with GBA2 and GBA3 mechanisms can be computed. From (18), (20), and (28), we calculate the  $E[N_b]$  values for the GFR with GBA1, GBA2, and GBA3 mechanisms. We also conduct simulation experiments for the four mechanisms. The simulation technique used in this paper is similar to that used in [11], and the details are omitted. Table 1 lists the  $P_{(f|G,f)}$  for GBA2 and GBA3,  $d_t$  (for Basic, GBA2, and GBA3) and  $E[N_b]$  (for GBA1, GBA2, and GBA3) obtained from simulation and analysis. In this table, the errors between the simulation experiments and analytic data are within 1%, which indicates that the simulation results match closely with the analytic data.

## 5. Numerical results

In this section, we investigate the  $d_t$  and  $P_{(f|G,f)}$  performances for the Basic mechanism, the GFR mechanism with different backup algorithms. To simplify our description, we use GBA1, GBA2, and GBA3 to denote the GFR with the GBA1, GBA2, and GBA3 algorithms, respectively. As described in



Table 1

The  $P_{f|G,f}$ ,  $E[N_b]$  and  $d_t$  values: simulation vs. analysis ( $\alpha = 99.93\%$ ,  $n = 2 \times 10^{-4} \mu_x$ ,  $d_r = 10^5 d_b$ ,  $\gamma_{G1} = \frac{1}{10} \mu_x$ ,  $K_{G2} = 10$ , and  $\gamma_{G3} = \frac{1}{9} \mu_x$ ).

	Outputs	Simulation	Analysis	Error
Basic	$d_t$	$8.5356 \times 10^4 d_b$	$8.5714 \times 10^4 d_b$	$4.17 \times 10^{-3}$
GBA1	$E[N_b]$	142.1	142.857	$5.3 \times 10^{-3}$
GBA2	$d_t$	$7.7079 \times 10^4 d_b$	$7.7285 \times 10^4 d_b$	$2.67 \times 10^{-3}$
	$E[N_b]$	141.62	142	$2.68 \times 10^{-3}$
	$P_{f G,f}$	90.19%	90%	$2.1 \times 10^{-3}$
GBA3	$d_t$	$7.6960 \times 10^4 d_b$	$7.7287 \times 10^4 d_b$	$4.23 \times 10^{-3}$
	$E[N_b]$	142.05	142.857	$5.65 \times 10^{-3}$
	$P_{f G,f}$	90.14%	90%	$1.53 \times 10^{-3}$

Section 4, the GBA2 mechanism uses the  $K_{G2}$  counter to determine the checkpoints to backup the PDP context, and the GBA1 and GBA3 mechanisms maintain the  $T_{G1}$  and  $T_{G3}$  timers to determine the checkpoints, where the values of  $T_{G1}$  and  $T_{G3}$  are exponentially generated with means  $1/\gamma_{G1}$  and  $1/\gamma_{G3}$ , respectively. For GBA1, we have a variant algorithm where the  $T_{G1}$  timer is set up with a fixed value  $1/\gamma_{G1}$ . This variant algorithm is denoted as GBA1.f.

In our study, the input parameters are normalized by  $\mu_x$ . For example, if the mean of the PDU inter-arrival times is  $1/\mu_x = 1$  second, then  $\eta = 10^{-3} \mu_x$  implies that averagely 16.6 minutes, the PDP context is corrupted. When the PDU inter-arrival times are exponentially distributed, the mean values  $E[N_b/N_m]$  of GBA1, GBA2, and GBA3 can be calculated. For a fair comparison for the performances of GBA1, GBA2, and GBA3, we set up the  $T_{G1}$ ,  $K_{G2}$  and  $T_{G3}$  values based on an equal expected value  $E[N_b/N_m] = 10\%$ . Similar results are observed for various  $E[N_b/N_m]$  values and will not be presented in the paper. Therefore, we set up  $\gamma_{G1} = \mu_x/10$  for GBA1 and GBA1.f,  $K_{G2} = 10$  for GBA2, and  $\gamma_{G3} = \mu_x/9$ .

**Effects of  $\alpha$  on  $P_{f|G,f}$  and  $d_t$ :** Figure 7 plots  $P_{f|G,f}$  and  $d_t$  as functions of  $1 - \alpha$ , where  $\eta = 2 \times 10^{-4} \mu_x$  and  $d_r = 3 \times 10^5 d_b$ . Figure 7(a) shows that  $1 - \alpha$  insignificantly affects  $P_{f|G,f}$  for the five mechanisms. From (14), (24), and (33), it is obvious that  $\alpha$  is not the factor that affects  $P_{f|G,f}$  Basic, GBA2, and GBA3. Figure 7(a) also indicates that the  $P_{f|G,f}$  for GBA1, GBA1.f, GBA2, and GBA3 are almost identical with 90%. In this figure, we set the expected value  $E[N_b/N_m] = 10\%$  for GBA1, GBA1.f, GBA2, and GBA3, and averagely the PDP context is backed up for 10% PDU arrivals.

In figure 7(b), we observe that  $d_t$  is a decreasing function of  $1 - \alpha$ . As  $1 - \alpha$  increases, a session is more likely to end, and shorter time period of the session is observed. Thus the probability that PDP context is corrupted during a session becomes small. The network cost caused by the PDP context restoration (i.e.,  $d_r E[N_r]$ ) drops, and lower total cost  $d_t$  is observed. Figure 7(b) indicates that the  $d_t$  values of GBA1, GBA1.f, GBA2, and GBA3 are almost identical,

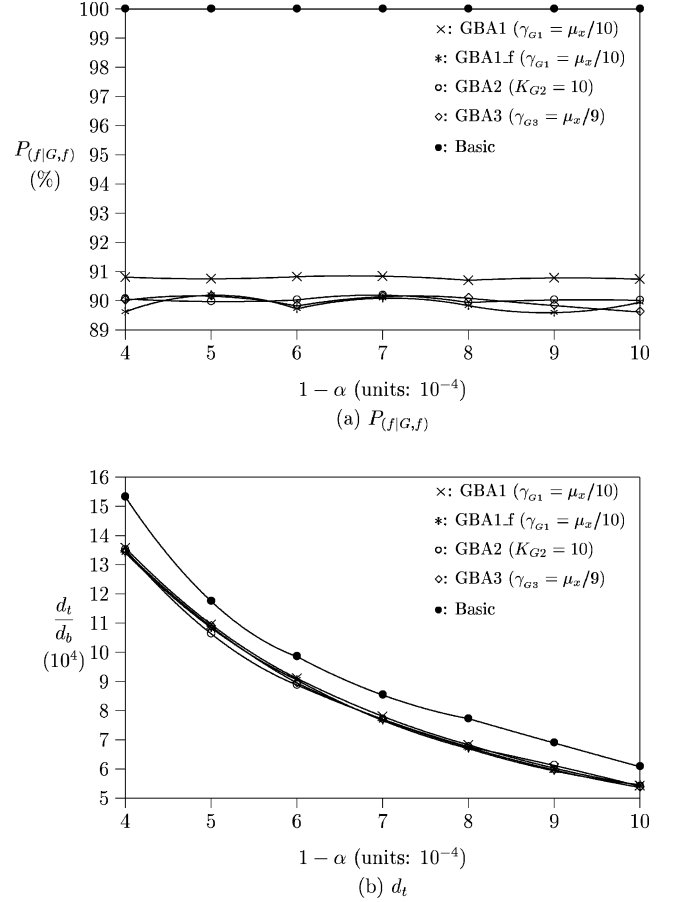


Figure 7. Effects of  $\alpha$  on  $P_{f|G,f}$  and  $d_t$  ( $\eta = 2 \times 10^{-4} \mu_x$ ;  $d_r = 3 \times 10^4 d_b$ ).

and about 12% improvement over the Basic mechanism is observed.

**Effects of  $\eta$  on  $P_{f|G,f}$  and  $d_t$ :** Figure 8 plots  $P_{f|G,f}$  and  $d_t$  as functions of  $1/n$ , where  $\alpha = 99.93\%$  and  $d_r = 3 \times 10^5 d_b$ . Figure 8(a) indicates that the values of  $P_{f|G,f}$  for the five mechanisms change slightly as  $1/\eta$  increases. Equations (14) and (24) indicate that  $\eta$  is not the factor affecting  $P_{f|G,f}$  Basic and GBA2. For GBA3, from equation (33), we know that  $\eta$  is shown in the numerator and the denominator of equation (33), and the effects of  $\eta$  on  $P_{f|G,f}$  for GBA3 is considered minor. Similar to the effects of  $\alpha$  on  $P_{f|G,f}$ , we see the same phenomenon caused by the effects of  $\eta$ .

In Figure 8(b),  $d_t$  is a decreasing function of  $1/\eta$ . As  $1/\eta$  increases, the time period between two PDP context corruption becomes longer. Thus the probability that the PDP context is corrupted during a session becomes small, and lower network cost  $d_r E[N_r]$  is expected. From (12), it is clear that  $E[N_r]$  is a major factor dominating the  $d_t$  cost. Therefore, with a larger  $1/\eta$ ,  $E[N_r]$  decreases, and smaller  $d_t$  values are observed.

**Effects of the Variance of PDU Inter-Arrival Times:** Based on the simulation experiments, we investigate the  $P_{f|G,f}$

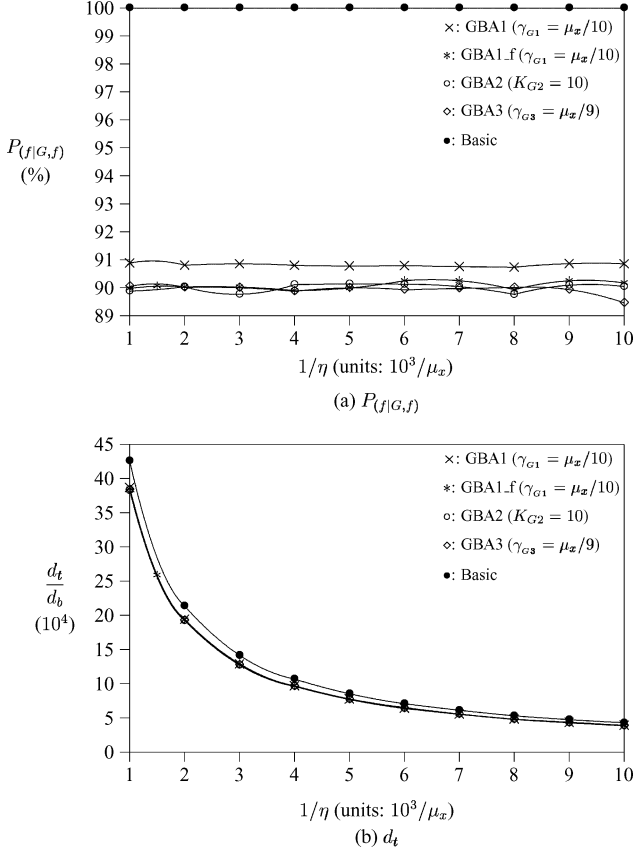


Figure 8. Effects of  $\eta$  on  $P_{(f|G,f)}$  and  $d_t$  ( $\alpha = 99.93\%$ ;  $d_r = 3 \times 10^5 d_b$ ).

and  $d_t$  performances for GBA1, GBA1.f, GBA2, and GBA3 when the PDU inter-arrival times are gamma distributed with mean  $1/\mu_x$  and variance  $v_{\mu_x} = 1/(\mu_x^2 \omega)$ , where  $\omega$  is the shape parameter and  $\omega > 0$ . Figure 9 plots  $P_{(f|G,f)}$  and  $d_t$  as functions of  $v_{\mu_x}$ , where  $\eta = 2 \times 10^{-4} \mu_x$ ,  $\alpha = 99.93\%$  and  $d_r = 3 \times 10^5 d_b$ . In this figure, the expected values of  $E[N_b/N_m]$  for GBA1, GBA1.f, GBA2, and GBA3 are set to 10%. In figure 9, we observe that when  $v_{\mu_x} > 1/\mu_x^2$ , GBA1, GBA1.f, and GBA3 significantly outperform GBA2 in terms of  $P_{(f|G,f)}$  and  $d_t$ . As  $v_{\mu_x}$  increases, the improvement is more significant. This is due to the fact that when  $v_{\mu_x}$  becomes larger, more small PDU inter-arrival times are observed. On the other hand, it is more likely to observe an extremely long PDU inter-arrival time in a session. If the PDP context can be backed up immediately at the beginning of the long PDU inter-arrival period, the backed up PDP context is valid until the end of the long PDU inter-arrival period. Consider the long PDU inter-arrival time (which is between when  $j$ th PDU arrives and when  $j+1$ st PDU arrives). When GBA1, GBA1.f, or GBA3 is adopted as the backup algorithms, since they are both timer-based, the PDP context is likely to be backed up for the  $j$ th PDU arrival. On the other hand, for GBA2 (counter-based), if  $j \bmod K_{G2} \neq 0$ , then the PDP context is not backed up for the

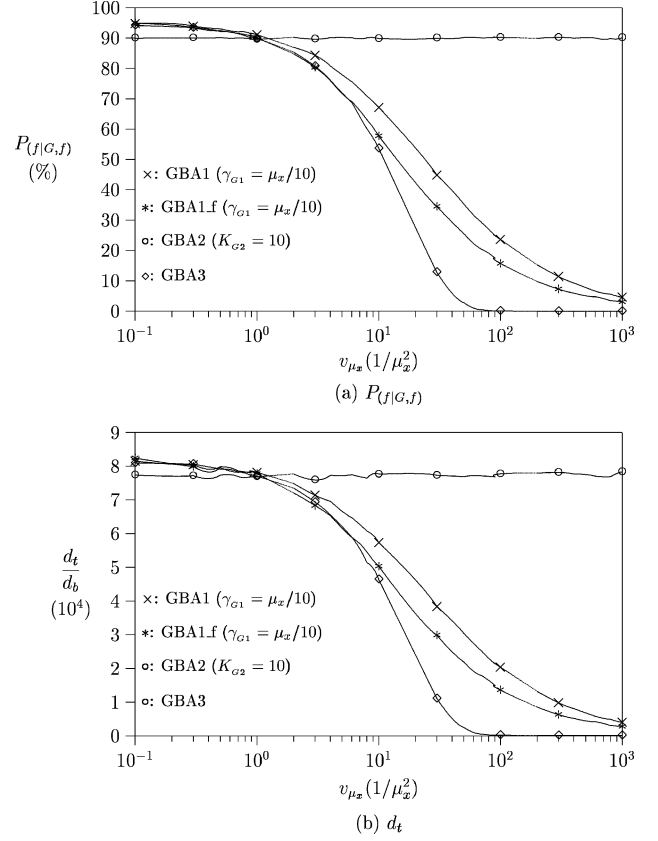


Figure 9. Effects of variance of PDU inter-arrival times on  $P_{(f|G,f)}$  and  $d_t$  ( $\eta = 2 \times 10^{-4} \mu_x$ ;  $\alpha = 99.93\%$ ;  $d_r = 3 \times 10^5 d_b$ ).

$j$ th PDU arrival. When the PDU inter-arrival time becomes long, the PDP context corruption is more likely to occur in this interval. Thus, we observe that as  $v_{\mu_x}$  increases, the  $P_{(f|G,f)}$  improvement of GBA1, GBA1.f, or GBA3 over GBA2 becomes significant. For the  $d_t$  performance, figure 9(b) shows that the similar trend as that in figure 9(a).

**Effects of Pareto PDU Inter-Arrival Times:** Based on the simulation experiments, we also evaluate the  $P_{(f|G,f)}$  and  $d_t$  performances for Basic, GBA1, GBA2 and GBA3 when the PDU inter-arrival times have a Pareto distribution with parameters  $\beta$  and  $l$ . It has been shown that Pareto distribution can approximate the packet traffic very well [1]. The  $\beta$  parameter describes the “heaviness” of the tail of the distribution. The Pareto density function is

$$f_x(x) = \left(\frac{\beta}{l}\right) \left(\frac{l}{x}\right)^{\beta+1}$$

and its expected value is

$$\frac{1}{\mu_x} = E[x] = \left(\frac{\beta}{\beta-1}\right) l \quad (36)$$

If  $\beta$  is between 1 and 2, the variance for the distribution becomes infinite. Once a suitable value for  $\beta$  is selected to

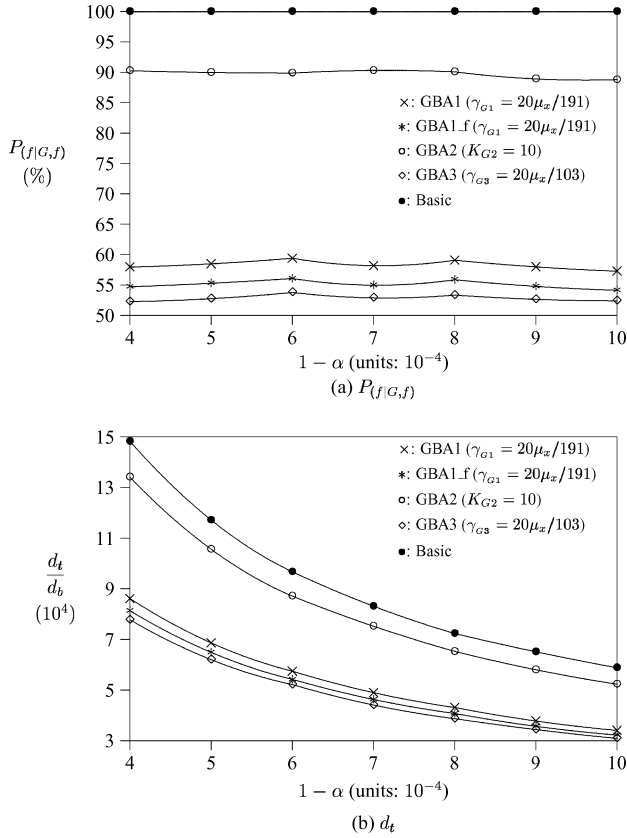


Figure 10. Effects of Pareto PDU inter-arrival times on  $P_{(f|G,f)}$  and  $d_t$  ( $\beta = 1.2$ ;  $\eta = 2 \times 10^{-4}\mu_x$ ;  $d_r = 3 \times 10^5 d_b$ ).

describe the traffic characteristics, the  $l$  value is determined by the mean of the distribution (using (36)). In our study,  $\beta$  is set to 1.2. The mean values  $E[N_b/N_m]$  of GBA1, GBA1.f, GBA2 and GBA3 are set to 10%, and we set up the  $\gamma_{G1}$ ,  $K_{G2}$  and  $\gamma_{G3}$  as  $20\mu_x/191$ , 10 and  $20\mu_x/103$ , respectively. Figure 10 plots  $P_{(f|G,f)}$  and  $d_t$  as functions of  $1 - \alpha$ , where  $\eta = 2 \times 10^{-4}\mu_x$  and  $d_r = 3 \times 10^5 d_b$ . Figure 10 shows that the  $P_{(f|G,f)}$  and  $d_t$  values for the five mechanisms have the following order:

$$\begin{aligned}
 P_{(f|G,f)}(\text{Basic}) &> P_{(f|G,f)}(\text{GBA2}) > P_{(f|G,f)}(\text{GBA1}) \\
 &> P_{(f|G,f)}(\text{GBA1.f}) > P_{(f|G,f)}(\text{GBA3}) \\
 d_t(\text{Basic}) &> d_t(\text{GBA2}) > d_t(\text{GBA1}) > d_t(\text{GBA1.f}) > d_t(\text{GBA3})
 \end{aligned}$$

The Pareto density function is widely used to simulate the characteristics of the burst-ness for packet data. As the packet traffic becomes bursty (i.e., the PDP context is modified more frequently in a short period), it is more likely to observe a very long PDU inter-arrival time during a session. Therefore, as we explain for the phenomenon when  $v_{\mu_x} > 1/\mu_x^2$  (see figure 9), the timer-based mechanisms (i.e., GBA1, GBA1.f, and GBA3) outperform the counter-based mechanism (i.e., GBA2). To conclude, GBA3 has the best performance among the four mechanisms.

## 6. Conclusions

We proposed an improved GGSN Failure Restoration (GFR) mechanism with different backup algorithms, GBA1, GBA2, and GBA3, to reduce the network signaling cost and delay for the restorations of the corrupted PDP contexts. The GFR backups the PDP context (in volatile storage) into non-volatile storage in GGSN at some checkpoints by using GBA1, GBA2, or GBA3. In GBA1 (a timer-based algorithm), the GGSN backups the PDP context into the non-volatile storage for every time interval  $T_{G1}$ . In GBA2 (a counter-based algorithm), the PDP context is backed up into the non-volatile storage for every  $K_{G2}$  PDU arrivals. In GBA3 (a timer-based algorithm), during a session, when the PDP context is modified (i.e., a PDU arrives), GGSN restarts the  $T_{G3}$  timer. When  $T_{G3}$  expires (i.e., no PDU arrives during the time period of  $T_{G3}$ ), the PDP context is backed up into the non-volatile storage. The GFR mechanism can be applied in the UMTS system without introducing any new component and modifying the protocol of the UMTS network. This mechanism is considered practical and cost-effective, which is very suitable to be deployed in the real system.

The analytic models (for GBA1, GBA2, GBA3, and the Basic mechanism proposed by 3GPP) and simulation experiments were conducted to evaluate the performances. Our study indicated that when the PDU inter-arrival times are exponentially distributed, GBA1, GBA2, and GBA3 have almost identical  $P_{(f|G,f)}$  and  $d_t$  performances, and outperform the Basic mechanism significantly. When the PDU inter-arrival times have a gamma distribution with large variance or a Pareto distribution (i.e., the packet traffic is bursty), GBA1 and GBA3 outperform GBA2 significantly in terms of  $P_{(f|G,f)}$  and  $d_t$ . Our study also showed that GBA3 can capture the packet traffic pattern precisely and has the best performance.

## Acknowledgments

The authors would like to thank the two anonymous reviewers. Their comments have significantly improved the quality of this paper. P. Lin's work was sponsored in part by the National Science Council (NSC), R.O.C., under the contract number NSC93-2213-E-002-095, Computer and Communications Research Labs/Industrial Technology Research Institute (CCL/ITRI), Chungghwa Telecom Labs, Microsoft, Telcordia Applied Research Center and Taiwan Network Information Center (TWNIC).

## Appendix A: Notation

The notation used in this paper is listed as follows.

- $1 - \alpha$ : the probability that an arrival PDU deactivates the PDP context
- $1/\gamma_{G1}$ : the expected value of the  $T_{G1}$  timer

- $1/\gamma_{G3}$ : the expected value of the  $T_{G3}$  timer
- $\eta$ : the PDP context corruption rate
- $\mu_x$ : the PDU arrival rate
- $d_b$ : the cost for each backup
- $d_r$ : the network cost to reactivate a PDP context
- $d_t$ : the total network cost introduced to a session
- $E[N_b]$ : the expected number of the backups executed for a PDP context in a session
- $E[N_f]$ : the expected number of PDP context corruptions in a session
- $E[N_m]$ : the expected number of PDU arrivals in a session
- $E[N_r]$ : the expected number of the PDP context reactivations in a session executed by the network
- $K_{G2}$ : the threshold number of PDU arrivals for which the GGSN backups the activated PDP context in GBA2
- $P_f$ : the probability that a PDP context corruption occurs, and the backup of the PDP context is invalid
- $P_{(f|G,f)}$ : the conditional probability that under the condition when the PDP context is corrupted, the backup is invalid
- $P_{G,f}$ : the probability that the PDP context is corrupted in a session

### Appendix B: Derivation of $d^n f_a^*(s)/ds^n$

Let  $f_a^*(S) = [(1 - \alpha)\mu_x]/[s + (1 - \alpha)\mu_x]$ . We prove that the following hypothesis is true.

$$\frac{d^n f_a^*(S)}{ds^n} = \frac{(-1)^n n! (1 - \alpha)\mu_x}{[s + (1 - \alpha)\mu_x]^{n+1}}$$

*Proof.* We prove by induction on  $n$ .

**Basis:** Consider the case when  $n = 1$ .

$$\frac{f_a^*(S)}{ds} = \frac{(-1)(1 - \alpha)\mu_x}{[s + (1 - \alpha)\mu_x]^2}$$

and the hypothesis holds.

**Inductive Step:** Suppose that the hypothesis holds when  $n = k$ . For  $n = k + 1$ , we have

$$\begin{aligned} \frac{d^{k+1} f_a^*(s)}{ds^{k+1}} &= d \left[ \frac{d^k f_a^*(s)}{ds^k} \right] / ds \\ &= d \left\{ \frac{(-1)^k k! (1 - \alpha)\mu_x}{[s + (1 - \alpha)\mu_x]^{k+1}} \right\} / ds \\ &= \frac{(-1)(k+1)(-1)^k k! (1 - \alpha)\mu_x}{[s + (1 - \alpha)\mu_x]^{k+2}} \\ &= \frac{(-1)^{(k+1)} (k+1)! (1 - \alpha)\mu_x}{[s + (1 - \alpha)\mu_x]^{(k+1)+1}} \end{aligned}$$

Thus the hypothesis holds for all cases.  $\square$

### Appendix C: Derivations of $C$ and $D$

$$\begin{aligned} C &= \int_{t_k=0}^{\infty} \int_{t_f=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \\ &\quad \times f_k(t_k) \eta e^{-\eta t_{f,0}} \mu_x e^{-(\mu_x + \gamma_{G3})z} dz dt_{f,0} dt_k \\ &= \eta \mu_x \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \\ &\quad \times f_k(t_k) e^{-\eta t_{f,0}} \mu_x e^{-(\mu_x + \gamma_{G3})z} dz dt_{f,0} dt_k \\ &= \frac{\eta \mu_x}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \int_{t_k=0}^{\infty} f_k(t_k) e^{-\eta t_k} dt_k \\ &= \left[ \frac{\eta \mu_x}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] f_k^*(\eta) \end{aligned}$$

Then from (7), we obtain

$$C = \left[ \frac{\eta \mu_x}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k$$

Similar to the derivation of  $C$ , we have  $D$  as follows.

$$\begin{aligned} D &= \int_{t_k=0}^{\infty} \int_{t_{f,0}=t_k}^{\infty} \int_{z=t_{f,0}-t_k}^{\infty} \\ &\quad \times f_k(t_k) \eta e^{-\eta t_{f,0}} \gamma_{G3} e^{-(\mu_x + \gamma_{G3})z} dz dt_{f,0} dt_k \\ &= \left[ \frac{\eta \gamma_{G3}}{(\mu_x + \gamma_{G3})(\eta + \mu_x + \gamma_{G3})} \right] \left( \frac{\alpha \mu_x}{\eta + \mu_x} \right)^k \quad (37) \end{aligned}$$

### References

- [1] M. Cheng and L.-F. Chang, Wireless dynamic channel assignment performance under packet data traffic, *IEEE Journal on Selected Areas in Communications* 17(7) (1999) 1257–1269.
- [2] 3GPP, 3rd Generation Partnership Project; Radio Interface Protocol Architecture. Technical Report Technical Specification 3G TS 25.301 version 3.4.0 (2000-03) (2000).
- [3] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architecture Principles for Release 2000; Release 2000. Technical Report Technical Specification 3G TR 23.821 version 1.0.1 (2000-07) (2000).
- [4] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Core Network; Restoration Procedures; Release 5. Technical Report Technical Specification 3G TS 23.007 version 4.0.0 (2001-03) (2001).
- [5] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Core Network; General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp Interface; Release 5. Technical Report Technical Specification 3G TS 29.060 version 5.0.1 (2002-01) (2002).
- [6] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2; Release 5. Technical Report Technical Specification 3G TS 23.060 version 5.0.0 (2002-01) (2002).
- [7] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS);

- Stage 2 (Release 5). Technical Report Technical Specification 3G TS 23.228 version 5.11.0 (2003-12) (2003).
- [8] H. Holma and A. Toskala, *WCDMA for UMTS* (John Wiley & Sons 2000).
- [9] <http://www.lucent.com>.
- [10] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures* (John Wiley & Sons, 2001).
- [11] P. Lin, Y.-B. Lin and I. Chlamtac, Modeling frame synchronization for UMTS high-speed downlink packet access, *IEEE Transactions on Vehicular Technology* 52(1) (2003).
- [12] P. Lin, Y.-B. Lin and J.-Y. Jeng, Improving GSM call completion by call re-establishment, *IEEE Journal on Selected Areas in Communications* 17(7) (1999) 1305–1317.



**Phone Lin** (M'02) received his BSCSIE degree and Ph.D. degree from National Chiao Tung University, Taiwan, R.O.C. in 1996 and 2001, respectively. From August 2001 to July 2004, he was an Assistant Professor in Department of Computer Science and Information Engineering (CSIE), National Taiwan University, R.O.C. Since August 2004, he has been an Associate Professor in Department of Computer Science and Information Engineering (CSIE), National Taiwan University, R.O.C. His current research interests include personal communications services, wireless In-

ternet, and performance modeling. Dr. Lin is a Guest Editor for IEEE Wireless Communications special issue on Mobility and Resource Management. He is also an Associate Editorial Member for the WCMC Journal. P. Lin's email and website addresses are [plin@csie.ntu.edu.tw](mailto:plin@csie.ntu.edu.tw) and <http://www.csie.ntu.edu.tw/~plin>, respectively.



**Guan-Hua Tu** received his B.S.C.S.I.E degree from National Central University, Taiwan, R.O.C., in 2001 and his Master degree in Computer Science from National Taiwan University, Taiwan, R.O.C., in 2003. He is currently a software engineer in MediaTek Inc. His research interests include personal communication services, mobile computing, and performance modeling.