# Modeling Prepaid Application Server of VoIP and Messaging Services for UMTS

Sok-Ian Sou, *Student Member, IEEE*, Yi-Bing Lin, *Fellow, IEEE*, Quincy Wu, and Jeu-Yih Jeng

***Abstract*—Universal mobile-telecommunication system supports IP multimedia services through the IP multimedia core-network subsystem (IMS). This paper proposes a prepaid application server (PAS) to handle both the prepaid calls and messaging services in the IMS. When both voice and messaging are simultaneously offered, a potential problem is that the delivery of a message during a call may result in force-termination of that call due to credit depletion. To address this issue, we describe a strategy to determine if a prepaid message can be sent out during a call session. We propose an analytic model to investigate the performance of this strategy. This paper provides guidelines to select appropriate input parameters for the PAS.**

***Index Terms*—Billing, diameter protocol, IP multimedia core-network subsystem (IMS), session initiation protocol (SIP), VoIP.**

## I. Introduction

The universal mobile-telecommunication system (UMTS) [1], [2], [17] is one of the major standards for the third-generation (3G) mobile telecommunications (details for UMTS can be found in [16, Chs. 2, 6, 7, and 9]. In UMTS, the IP multimedia core-network subsystem (IMS) provides multimedia services by utilizing session initiation protocol (SIP) [3], [19]. In this subsystem, the call-session control function (CSCF) is basically an SIP server, which is responsible for call control. Other nodes in the IMS include media gateway control function (MGCF), media gateway (MGW), and transport signaling gateway (T-SGW). These nodes are typically used in a VoIP network and will be elaborated upon in the next section (more details for the IMS can be found in [5], [11], and [16]). With the growing demand for instant messaging services, the authors of RFC 3428 [12] propose an SIP extension, which is the MESSAGE method that allows transfer of instant messages over the Internet. With this extension, Internet services (such as mail and instant messaging [18]) can be integrated with Short Message Service (SMS)/Multimedia Messaging Service (MMS) through the IMS. It is clear that billing is an essential part of service provision. Billing mechanisms for messaging (both SMS and MMS) and VoIP (particularly for the VoIP calls toward the fixed-line telephones) are typically deployed for postpaid services. On the contrary, the prepaid mechanisms for combining these services have seldom been studied in the literature. This paper proposes a prepaid application server (PAS) that can simultaneously process SMS/MMS messages and VoIP calls. This PAS supports VoIP calls toward the public switched-telephone network (PSTN) and messaging service toward UMTS. We show how the PAS can accommodate both IMS-to-PSTN calls and instant messaging services through SIP. Then, we propose an analytic model to investigate the performance of the PAS.

Manuscript received October 22, 2005; revised April 12, 2006 and May 24, 2006. This work was supported in part by the NSC Excellence Project under Grants NSC 94-2752-E-009-005-PAE, NSC 94-2219-E-009-001, and NSC 94-2213-E-009-104, by the NTP VoIP Project under Grant NSC 94-2219-E-009-002, by the NTP Service IOT Project under Grant NSC 94-2219-E-009-024, by Intel, by Chung Hwa Telecom, by IIS/Academia Sinica, by the ITRI/NCTU Joint Research Center, and by MoE ATU. The review of this paper was coordinated by Dr. Q. Zhang.

S.-I. Sou and Y.-B. Lin are with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan, R.O.C. (e-mail: sisou@csie.nctu.edu.tw; liny@csie.nctu.edu.tw).

Q. Wu is with the Graduate Institute of Communication Engineering, National Chi Nan University, Nantou 545, Taiwan, R.O.C. (e-mail: solomon@ncnu.edu.tw).

J.-Y. Jeng is with Information Technology Laboratory, Telecommunication Laboratories, Chunghwa Telcom Company, Ltd., Taiwan, R.O.C. (e-mail: jyjeng@cht.com.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2007.895493

## II. PAS of SIP-Based Services

SIP-based prepaid service over the Internet has been recently investigated (e.g., [22] and [23]). However, SIP-based prepaid service in the UMTS/IMS network has seldom been studied. In 3G Partnership Project (3GPP) Release 6, the online-charging system (OCS) is proposed to provide a real-time convergent charging solution for next-generation services [6]. It also provides the protocol to support prepaid services (see the Appendix). However, how to manage and control the allocated credit for the SIP-based prepaid services is not mentioned in the 3GPP specification. Therefore, a network entity is required to provide these functionalities without modifying the existing network entities such as the OCS and the CSCF. This section proposes a PAS to manage credit allocation for both the prepaid SIP calls and messaging services in the IMS.

Fig. 1 illustrates the PAS that interacts with several network entities in the IMS network. In this figure, the CSCF [Fig. 1(a)] utilizes SIP to provide control signaling for IP-based multimedia services. The MGCF, MGW, and T-SGW [Fig. 1(b)] interwork the IMS network with the PSTN. The user equipment [UE; Fig. 1(c)] supports SIP-based VoIP call services and SIP-based instant messaging services. The IP-Message Gateway [Fig. 1(d)] interworks with the SIP-based messaging service and SMS/MMS delivery [9]. The SMS-Interworking Mobile Switching Center [SMS-IWMSC; Fig. 1(e)] sends/receives SMS to/from the IP-Message Gateway using standard Mobile

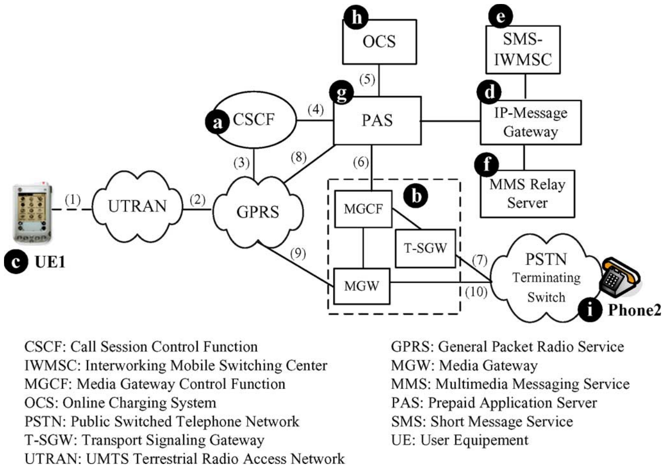



Fig. 1. IMS environment for SIP-based prepaid services.

Application Part signaling [10], while the MMS relay/server [Fig. 1(f)] sends/receives MMS to/from the IP-Message Gateway via the MM1 interface. The PAS [Fig. 1(g)] interacts with the OCS [Fig. 1(h)] to obtain the amount of prepaid user credits and process the online-charging information. The PAS has the capability to set up prepaid IP–PSTN calls through the MGCF, MGW, and T-SGW. The PAS also supports prepaid SMS/MMS services through interaction with the IP-Message Gateway.

The prepaid-call function in PAS works as follows. By utilizing the B2BUA technique [23], the prepaid-call function is inserted into an SIP connection session by breaking it into two subsessions. In this way, the PAS can monitor and terminate the call session when the user credit is depleted. To set up an IMS-to-PSTN call, the signaling message is first routed from the UE [Fig. 1(c)] to the CSCF [Fig. 1(a)]. The CSCF identifies the charging type (i.e., prepaid or postpaid) of the service. For a postpaid SIP request, the call is set up following standard IMS procedures [16]. For a prepaid SIP request, the CSCF forwards the request to the PAS [Fig. 1(g)] for further authorization and call-session control.

### A. Prepaid IMS-to-PSTN Call Setup and Termination

Consider the scenario where an IMS UE [UE1; Fig. 1(c)] makes a VoIP call to a PSTN phone [Phone2 in Fig. 1(i)]. In this case, the PAS (i.e., the prepaid B2BUA) breaks the SIP session between UE1 and the MGCF into one subsession (**Subsession 1**) between UE1 and the PAS [Fig. 1(f)] and another subsession (**Subsession 2**) between the PAS and the MGCF [Fig. 1(b)]. When the PAS receives an SIP message, such as INVITE/200 OK/ACK and BYE/200 OK, it sends the Diameter Credit Control Request message with parameter "INITIAL_REQUEST" or "TERMINATE_REQUEST" to the OCS [7], [13] and terminates the prepaid call session when the prepaid credit of the caller is depleted. The prepaid-call setup and force-termination message flows are described as follows.

In Fig. 1, the SIP INVITE request from UE1 (a prepaid user) is first routed to the PAS through the CSCF [path (1) → (2) → (3) → (4)]. Then, the PAS interacts with the OCS to reserve the subscriber's credits [path (5)]. When the user's authorized time (i.e., the prepaid credits) is obtained, the PAS generates a new INVITE message for Subsession 2 and then forwards it to the MGCF through path (6). The MGCF instructs the T-SGW to deliver the SS7 ISUP IAM message (the SS7 call-setup message [15]) to Phone2 via path (7). When the called party answers, the SS7 ISUP ANM message is sent to the T-SGW through path (7). The MGCF sends a final response 200 OK to the PAS through path (6). Then, the PAS responds SIP 200 OK to UE1 through path (4) → (3) → (2) → (1). Finally, the PAS starts an authorized session timer with the value (the available prepaid credit) obtained from the OCS. UE1 then sends the ACK message (for Subsession 1) to the PAS through path (1) → (2) → (8). The PAS sends the ACK message (for Subsession 2) to the MGCF through path (6). At this point, the MGW opens a real-time transport-protocol (RTP) [21] connection so that UE1 can send/receive voice packets to/from Phone2 through the MGW. The media path for the prepaid call is (1) ↔ (2) ↔ (9) ↔ (10).

If the prepaid credit is exhausted (i.e., the authorized session timer expires) before the conversation is complete, the call is forced to terminate by the PAS. The authorized session timer expires, and the PAS sends the BYE messages to both UE1 [path (8) → (2) → (1)] and Phone2 through the MGCF [path (6)]. Specifically, the MGCF instructs the MGW to release the RTP connection, and the subsequent voice packets delivered between UE1 and Phone2 are not allowed to pass through the MGW. The MGCF instructs the T-SGW to send the SS7 ISUP Release message to Phone2 via path (7). Finally, the PAS triggers the OCS to debit the user account through path (5). The prepaid message flow for prepaid messaging service is similar to that for VoIP calls (details can be found in [22]).

## III. Modeling the PAS

The PAS for pure voice services or pure message services can be easily implemented, as described in Section II. However, when both voice and messaging are simultaneously offered, an important issue must be resolved for the PAS. That is, during a prepaid IMS-to-PSTN call, the user may attempt to send messages. Deduction of prepaid credit for sending out this message may result in insufficient credit left for the in-progress prepaid call (note that a message can be a short text or a huge multimedia data file, which can causes large deduction). In this case, the IMS-to-PSTN call is forced to terminate. Therefore, a strategy is required to determine if the prepaid message can be sent out without causing force-termination of an ongoing call. To avoid unnecessary force-termination (UFT), we set a threshold amount $X_T$ of credit units (CU) to protect the in-progress IMS-to-PSTN call (consider the timing diagram in Fig. 2). A prepaid call arrives at $t_1$ and completes at $t_3$. Without loss of generality, we assume that each time unit (TU) of the call is charged for 1 CU. A prepaid authorized timer for this call starts at $t_1$ and expires at $t_4$. That is, the amount of the prepaid credit is $t_4 - t_1$. Upon receipt of the prepaid message request at $t_2$, where $t_1 < t_2 < t_3$, the PAS estimates if the remaining credit $x^* = t_4 - t_2$ suffices to support both the remaining call and the message service. Assume that the prepaid message service is charged for a fixed amount of $T_m$

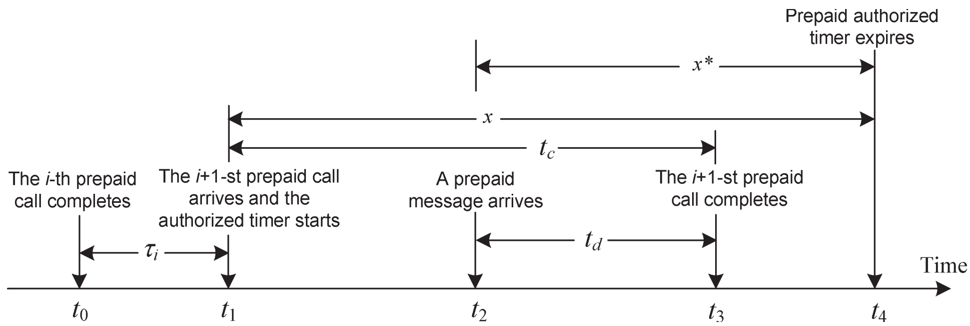



Fig. 2. Timing diagram for prepaid services.

CUs. The strategy used in our PAS is described as follows: 1) If the remaining prepaid credit $x^* \geq X_T + T_m$, the message is sent immediately, and the amount of remaining prepaid credits is reduced to $x^* - T_m$, and 2) if not, the message is stored and will be processed after the IMS-to-PSTN call is completed.

If $X_T$ is set too large, the PAS may reserve too many prepaid CUs, and therefore, message deliveries are unnecessarily delayed. If $X_T$ is set too small, the PAS may reserve too few prepaid CUs for the remaining prepaid call and, therefore, results in UFT calls. We present an analytic model to derive the UFT probability $P_{\mathrm{UFT}}$ and the expected unnecessary delay $E[t_d]$ with the following assumptions.

1) Let $X$ be the amount of the initial prepaid credit.
2) The prepaid message arrivals are a Poisson stream with rate $\lambda_m$. Let $N(t)$ be the number of prepaid messages occurring in period $t$. The probability mass function of $N(t)$ is

$$\Pr[N(t) = n] = \left[\frac{(\lambda_m t)^n}{n!}\right] e^{-\lambda_m t}. \tag{1}$$

3) Let $\tau_i$ be the interval between when the $i$th prepaid call completes and when the $i+1$th prepaid call starts ($i > 0$). In Fig. 2, $\tau_i = t_1 - t_0$. We define $\tau_0$ as the interval between when the prepaid account is first activated and when the first prepaid call arrives. Let $\tau_i$ be exponentially distributed with the mean $1/\lambda_c$.
4) The prepaid-call holding time $t_c$ (in Fig. 2, $t_c = t_3 - t_1$) is exponentially distributed with the density function

$$f_c(t_c) = \mu e^{-\mu t_c}. \tag{2}$$

## A. Derivation for $f_x(x)$

Consider the timing diagram in Fig. 2, where a prepaid call arrives at $t_1$, and the remaining credit left is $x = t_4 - t_1$. To derive $P_{\mathrm{UFT}}$ and $E[t_d]$, we first derive the density function $f_x(x)$ of $x$. Let $N_c$ be the number of prepaid calls completed during $[0, t_1]$. Let $X_c$ be the amount of CUs that are consumed by the prepaid calls arriving before $t_1$. Since we assume that 1 TU of the call is charged for 1 CU, the amount of the consumed prepaid credit (i.e., the accumulated call holding times) for these $N_c$ prepaid calls is $X_c$. For $N_c > 0$, $X_c$ has the Erlang distribution with the mean $N_c/\mu$ and the shape parameter $N_c$. The conditional probability mass function of $N_c$, given that $X_c = x_c$, can be expressed as

$$\begin{aligned}\Pr[N_c = n_c | X_c = x_c] &= \frac{\left[\frac{(\mu x_c)^{n_c-1}}{(n_c-1)!}\right]\mu e^{-\mu x_c}}{\sum_{i=1}^{\infty}\left[\frac{(\mu x_c)^{i-1}}{(i-1)!}\right]\mu e^{-\mu x_c}} \\ &= \frac{e^{-\mu x_c}(\mu x_c)^{n_c-1}}{(n_c-1)!}.\end{aligned} \tag{3}$$

Let $T_i = \sum_{i=0}^{N_c} \tau_i$. Since $\tau_i$ has the exponential distribution with the mean $1/\lambda_c$, $T_i$ has the Erlang distribution with the mean $(N_c+1)/\lambda_c$ and the shape parameter $N_c + 1$. Therefore, the density function $f_T(T_i)$ of $T_i$ is

$$f_T(T_i) = \left[\frac{\lambda_c(\lambda_c T_i)^{N_c}}{N_c!}\right] e^{-\lambda_c T_i}. \tag{4}$$

Let $N_m$ be the number of prepaid messages arriving before $t_1$. In other words, there are $N_m$ prepaid messages occurring in period $X_c + T_i$. Then, $N_m$ is a Poisson random variable with the mean $\lambda_m(X_c + T_i)$. From (1), the probability mass function of $N_m$ can be expressed as

$$\begin{aligned}\Pr[N_m = n_m] &= \Pr[N(X_c + T_i) = n_m] \\ &= \frac{[\lambda_m(X_c+T_i)]^{n_m}}{n_m!} e^{-\lambda_m(X_c+T_i)}.\end{aligned} \tag{5}$$

Note that $N(t)$ [see (1)] represents the number of prepaid messages occurring in any arbitrary period $t$. Therefore, $N_m = N(X_c + T_i)$ [see (5)] represents the number of prepaid messages occurring before the prepaid call arriving at $t_1$. From (3)–(5), the conditional probability mass function of $N_m$, given that $X_c = x_c$, is derived as

$$\begin{aligned}&\Pr[N_m = n_m | X_c = x_c] \\ &\quad = \sum_{n_c=1}^{\infty} \Pr[N_c = n_c | X_c = x_c] \\ &\qquad \times \int_{t_i=0}^{\infty} \Pr[N_m = n_m | X_c = x_c, N_c = n_c, T_i = t_i] f_T(t_i) dt_i\end{aligned}$$

$$= \sum_{n_c=1}^{\infty} \frac{e^{-\mu x_c}(\mu x_c)^{n_c-1}}{(n_c-1)!} \times \int_{t_i=0}^{\infty} \left\{ \frac{[\lambda_m(x_c+t_i)]^{n_m} e^{-\lambda_m(x_c+t_i)}}{n_m!} \right\} \times \left[ \frac{(\lambda_c t_i)^{n_c} \lambda_c e^{-\lambda_c t_i}}{n_c!} \right] dt_i$$

$$= \left[ \frac{\lambda_c \lambda_m^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \int_{t_i=0}^{\infty} e^{-(\lambda_c+\lambda_m)t_i} \sum_{k=0}^{n_m} \frac{t_i^k x_c^{n_m-k-1}}{k!(n_m-k)!} \times \sum_{n_c=1}^{\infty} \frac{(\mu\lambda_c x_c t_i)^{n_c}}{(n_c-1)! n_c!} dt_i$$

$$= \left[ \frac{\lambda_c {\lambda_m}^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \sum_{k=0}^{n_m} \sum_{n_c=1}^{\infty} \left[ \frac{(n_c+k)!}{(\lambda_c+\lambda_m)^{n_c+k+1}} \right] \times \left[ \frac{x_c^{n_m-k-1}}{k!(n_m-k)!} \right] \left\{ \frac{(\mu\lambda_c x_c)^{n_c}}{(n_c-1)! n_c!} \right\}$$

$$= \left[ \frac{\lambda_c {\lambda_m}^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \sum_{k=0}^{n_m} \left[ \frac{x_c^{n_m-k-1}}{(\lambda_c+\lambda_m)^{k+1} k!(n_m-k)!} \right] \times \sum_{n_c=1}^{\infty} \left( \frac{\mu\lambda_c x_c}{\lambda_c+\lambda_m} \right)^{n_c} \left[ \frac{(n_c+k),\ldots,(n_c+1)}{(n_c-1)!} \right]. \quad (6)$$

Let $Z = \mu\lambda_c x_c/(\lambda_c+\lambda_m)$. Since $(d/dZ)(Z^{n_c+k})^{(k)} = (n_c+k),\ldots,(n_c+1)Z^{n_c}$, then (6) can be rewritten as

$$\Pr[N_m = n_m | X_c = x_c] = \left[ \frac{\lambda_c \lambda_m^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \sum_{k=0}^{n_m} \left[ \frac{x_c^{n_m-k-1}}{(\lambda_c+\lambda_m)^{k+1} k!(n_m-k)!} \right] \times \sum_{n_c=1}^{\infty} \left[ \frac{d}{dZ} \frac{(Z^{n_c+k})^{(k)}}{(n_c-1)!} \right]$$

$$= \left[ \frac{\lambda_c \lambda_m^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \times \sum_{k=0}^{n_m} \left[ \frac{x_c^{n_m-k-1}}{(\lambda_c+\lambda_m)^{k+1} k!(n_m-k)!} \right] \frac{d}{dZ} \left( Z^{k+1} e^Z \right)^{(k)}. \quad (7)$$

According to the Leibniz's rule for higher derivatives of products, we can expand (7) as

$$\Pr[N_m = n_m | X_c = x_c] = \left[ \frac{\lambda_c \lambda_m^{n_m} e^{-(\mu+\lambda_m)x_c}}{\mu} \right] \sum_{k=0}^{n_m} \left[ \frac{x_c^{n_m-k-1}}{(\lambda_c+\lambda_m)^{k+1} k!(n_m-k)!} \right] \times \sum_{l=0}^{k} \binom{k}{l} \frac{d}{dZ} (Z^{k+1})^{(l)} \frac{d}{dZ} (e^Z)^{(k-l)}$$

$$= \left( \frac{\lambda_c \lambda_m^{n_m}}{\mu} \right) \sum_{k=0}^{n_m} \left[ \frac{(k+1!)}{(\lambda_c+\lambda_m)^{k+1}(n_m-k)!} \right] \times \sum_{l=0}^{k} \left[ \frac{\left( \frac{\mu\lambda_c}{\lambda_c+\lambda_m} \right)^{k-l}}{l!(k-l)!(k-l+1)!} \right] e^{-\left(\mu+\lambda_m-\frac{\mu\lambda_c}{\lambda_c+\lambda_m}\right)x_c} x_c^{n_m-l}. \quad (8)$$

When $T_m = 0$, it is clear that $X_c > 0$ is uniformly distributed over $(0, X]$ with the density $1/X$. When $T_m > 0$ and $x > 0$, the number of prepaid messages that arrived before $t_1$ is at most $\lfloor (X - X_c)/T_m \rfloor$. From (8)

$$\begin{aligned} &\Pr[N_m = n_m | X_c = x_c, T_m > 0, x > 0] \\ &= \frac{\frac{1}{X} \Pr[N_m = n_m | X_c = x_c]}{\Pr[X > X_c + N_m T_m > 0]} \\ &= \frac{\frac{1}{X} \Pr[N_m = n_m | X_c = x_c]}{\frac{1}{X} \int_{t=0}^{X} \sum_{n=0}^{\lfloor \frac{X-t}{T_m} \rfloor} \Pr[N_m = n | X_c = t] dt} \\ &= \frac{1}{C} \Pr[N_m = n_m | X_c = x_c] \end{aligned} \quad (9)$$

where $C$ is expressed in (10). In Fig. 2, suppose that there are $N_m$ prepaid messages that are delivered before $t_1$, and assume that $X_c = X - x - N_m T_m$. When a prepaid call arrives at $t_1$, it is clear that the remaining prepaid credit left is $x$. Therefore

$$\begin{aligned} C &= \int_{t=0}^{X} \sum_{n=0}^{\lfloor \frac{X-t}{T_m} \rfloor} \Pr[N_m = n | X_c = t] dt \\ &= \sum_{n=0}^{\lfloor \frac{X}{T_m} \rfloor} \int_{t=0}^{X-nT_m} \Pr[N_m = n | X_c = t] dt \\ &= \frac{\lambda_c}{\mu} \sum_{n=0}^{\lfloor \frac{X}{T_m} \rfloor} \lambda_m^n \left[ \sum_{k=0}^{n} \frac{(k+1)!}{(\lambda_c+\lambda_m)^{k+1}(n-k)!} \right] \\ &\quad \times \sum_{l=0}^{k} \left[ \frac{\left( \frac{\mu\lambda_c}{\lambda_c+\lambda_m} \right)^{k-l} (n-l)!}{l!(k-l)!(k-l+1)! \left( \mu+\lambda_m-\frac{\mu\lambda_c}{\lambda_c+\lambda_m} \right)^{n-l+1}} \right] \\ &\quad \times \left\{ 1 - e^{-\left(\mu+\lambda_m-\frac{\mu\lambda_c}{\lambda_c+\lambda_m}\right)(X-iT_m)} \times \sum_{j=0}^{n-i} \frac{\left[ \left( \mu+\lambda_m-\frac{\mu\lambda_c}{\lambda_c+\lambda_m} \right)(X-nT_m) \right]^j}{j!} \right\} \end{aligned} \quad (10)$$

and from (9), the density function $f_x(x)$ of $x$ can be derived as

$$f_x(x)dx = \sum_{n_m=0}^{\lfloor \frac{X-x}{T_m} \rfloor} \Pr[N_m = n_m | T_m > 0, X_c = X - x - n_m T_m, x > 0] \quad (11)$$

where $\Pr[N_m = n_m | X_c = x_c, T_m > 0, x > 0]$ can be obtained from (8)–(10).

### B. Derivations for $P_{\mathrm{UFT}}$ and $E[t_d]$

Based on (11), we first derive the output measure $P_{\mathrm{UFT}}$ for $X_T = 0$ (that is, a prepaid message is immediately served if $x^* > T_m$). In this case, a call is UFT if $N(t_c) > \lfloor (x - t_c/T_m) \rfloor$. Based on (1), (2), and (11), if $X_c > 0$, the UFT probability $P_{\mathrm{UFT}}$ can be derived as (12), where $f_x(x)$ is expressed in (11) as

$$
\begin{aligned}
P_{\mathrm{UFT}} &= \int_{x=0}^{X}\int_{t_c=0}^{x} \Pr\left[N(t_c) > \left\lfloor \frac{x-t_c}{T_m} \right\rfloor\right] f_c(t_c) f_x(x) dt_c dx \\
&= \int_{x=0}^{X}\int_{t_c=0}^{x} \left[1 - \sum_{i=0}^{\left\lfloor \frac{x-t_c}{T_m} \right\rfloor} \frac{(\lambda_m t_c)^i e^{-\lambda_m t_c}}{i!}\right] \mu e^{-\mu t_c} dt_c f_x(x) dx \\
&= 1 - \int_{x=0}^{X} \Bigg\{ e^{-\mu x} + \mu \sum_{i=0}^{\left\lfloor \frac{x}{T_m} \right\rfloor} \Bigg\{ 1 - e^{-(\mu+\lambda_m)(x - iT_m)} \\
&\qquad \times \sum_{j=0}^{i} \frac{[(\mu+\lambda_m)(x-iT_m)]^j}{j!} \Bigg\}\Bigg\} \\
&\qquad \times f_x(x) dx. \qquad (12)
\end{aligned}
$$

The output measure $E[t_d]$ for $X_T \geq X$ is derived as follows. A prepaid message that arrives during an in-progress call must be stored and will be processed after the prepaid call is completed. Since the prepaid message arrivals are random-observation points of the prepaid-call holding interval [20], the delay for the message can be considered as the residual life of the prepaid call. Consider the situation in which a prepaid call with the call holding time $t_c$ arrives at $t_1$, where $t_c = t_3 - t_1 \leq x - T_m$. If a prepaid message arrives at $t_2$ during interval $[t_1, t_3]$, the message is sent at $t_3$ with the delay $t_d = t_3 - t_2$. According to the residual life theorem [14] and for $X_c > 0$, the expected unnecessary delay $E[t_d]$ for the SMS delivery can be expressed as

$$
\begin{aligned}
E[t_d] &= \frac{E[{t_c}^2 | x > t_c + T_m]}{2E[t_c | x > t_c + T_m]} \\
&= \frac{\int_{x=T_m}^{X} \int_{t_c=0}^{x-T_m} t_c^2 f_c(t_c) f_x(x) dt_c dx}{\int_{x=T_m}^{X} \int_{t_c=0}^{x-T_m} t_c f_c(t_c) f_x(x) dt_c dx} \\
&= \frac{X - T_m - \int_{x=T_m}^{X} e^{-\mu x}\left[1 + \mu x + \frac{(\mu x)^2}{2}\right] f_x(x) dx}{\mu[X - T_m - \int_{x=T_m}^{X} e^{-\mu x}(1+\mu x) f_x(x) dx]}
\end{aligned} \qquad (13)
$$

where $f_x(x)$ is expressed in (11).

We also developed a simulation model for the PAS. The simulation experiments are validated against the analytic model developed in this section. Based on (12) and (13), Table I lists the analytic and the simulation results for the UFT probability $P_{\mathrm{UFT}}$ and the expected unnecessary delay $E[t_d]$. Table I indicates that, for all cases considered in this paper, the errors are within 0.6%. Therefore, the analytic and the simulation results are consistent.

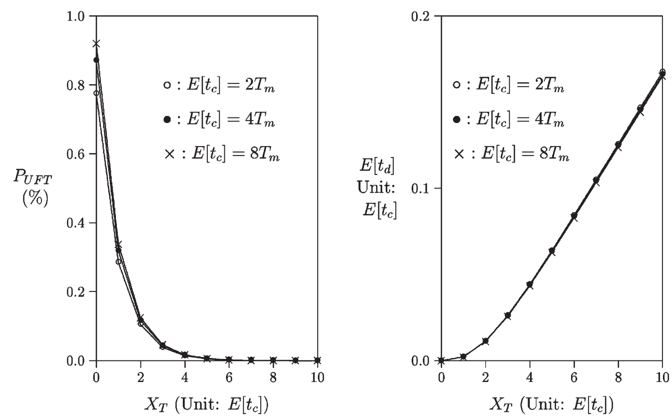
TABLE I
COMPARISON OF THE ANALYTIC AND SIMULATION RESULTS FOR $X_c > 0$ ($T_m = 5$ CU, $X = 20\,T_m$, $1/\lambda_c = 50$ TU).
(a) $P_{\mathrm{UFT}}(X_T = 0, E[t_c] = \frac{1}{5\lambda_c})$. (b) $E[t_d](X_T \geq X, \lambda_m = \lambda_c)$

| $1/\lambda_m$ (Unit: TU) | 1 | 5 | 10 | 50 | 100 |
|---|---|---|---|---|---|
| Simulation | 43.086% | 13.172% | 6.409% | 1.079% | 0.520% |
| Analytic | 42.960% | 13.196% | 6.419% | 1.078% | 0.520% |
| Error | −0.29% | 0.18% | 0.16% | −0.09% | −0.00% |

(a)

| $E[t_c]$ (Unit: TU) | 1 | 5 | 10 | 50 | 100 |
|---|---|---|---|---|---|
| Simulation (Unit: TU) | 0.990 | 4.678 | 8.520 | 18.128 | 20.056 |
| Analytic (Unit: TU) | 0.985 | 4.681 | 8.534 | 18.172 | 20.050 |
| Error | −0.51% | 0.06% | 0.16% | 0.24% | −0.03% |

(b)




(a) (b)

Fig. 3. (a) Effect on $P_{\mathrm{UFT}}$. (b) Effect on $E[t_d]$. Effects of $E[t_c]$ on $P_{\mathrm{UFT}}$ and $E[t_d]$ ($X = 50E[t_c]$ and $\lambda_m = 5\lambda_c$).

## IV. NUMERICAL EXAMPLES

This section uses the simulation experiments to investigate the performance of the PAS. The input parameter threshold $X_T$ and the output measure $E[t_d]$ are normalized by the mean of the call holding time $E[t_c] = 1/\mu$. For the purposes of demonstration, we assume that the prepaid message service is charged for $T_m = 5$ CUs and the expected intercall arrival time $1/\lambda_c = 50$ TUs.

1) *Effects of the expected call holding time $E[t_c]$.* Fig. 3 plots $P_{\mathrm{UFT}}$ and $E[t_d]$ against the threshold $X_T$ and $E[t_c]$, where $X = 50E[t_c]$, and $\lambda_m = 5\lambda_c$. This figure shows that $E[t_c]$ has insignificantly effects on $P_{\mathrm{UFT}}$ and $E[t_d]$ when the proportion of $X$ and $E[t_c]$ is fixed.

   In the remainder of this section, the expected intermessage arrival time $1/\lambda_m$ and the initial prepaid credit $X$ are normalized by $E[t_c] = 4T_m$.

2) *Effects of the variance $V_c$ of the call holding time $t_c$.* Fig. 4 plots $P_{\mathrm{UFT}}$ and $E[t_d]$ against $X_T$ and the variance $V_c$ of the call holding time $t_c$ with Gamma distribution, where $X = 25E[t_c]$, and $1/\lambda_m = 0.5E[t_c]$. This figure shows that both $P_{\mathrm{UFT}}$ and $E[t_d]$ increase as $V_c$ increases. This phenomenon is explained as follows: As the variance $V_c$ of $t_c$ increases, more long and short $t_c$ periods are observed. The prepaid message (i.e., the random observer) is more likely to fall in the long $t_c$ periods than the short $t_c$ periods, and larger residual call holding

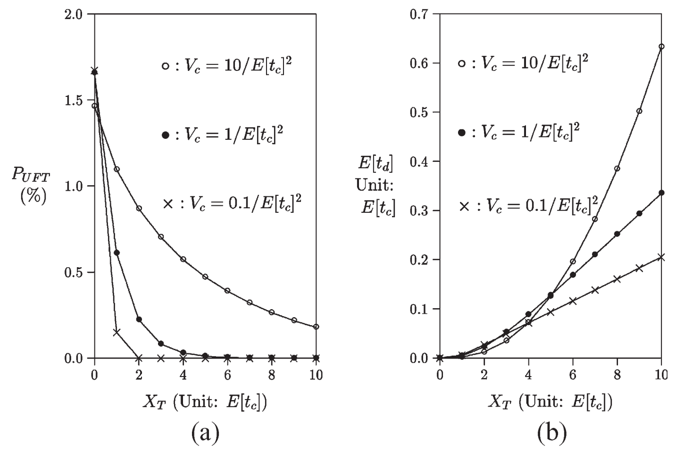



Fig. 4. (a) Effect on $P_{\mathrm{UFT}}$. (b) Effect on $E[t_d]$. Effects on $V_c$ on $P_{\mathrm{UFT}}$ and $E[t_d]$ ($X = 25E[t_c]$ and $1/\lambda_m = 0.5E[t_c]$).

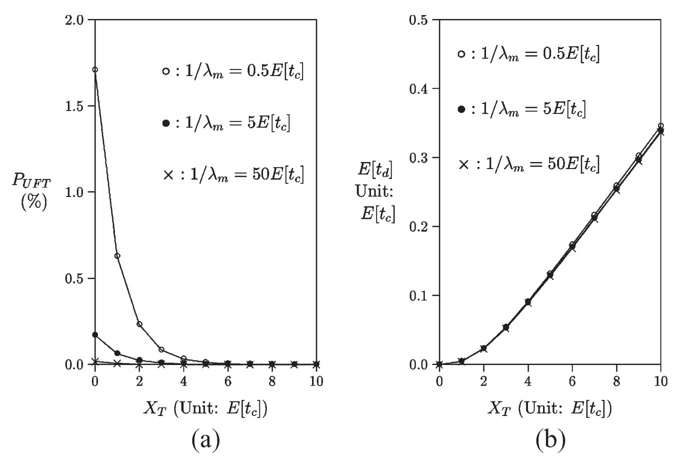



Fig. 5. (a) Effect on $P_{\mathrm{UFT}}$. (b) Effect on $E[t_d]$. Effects on $1/\lambda_m$ on $P_{\mathrm{UFT}}$ and $E[t_d]$ ($X = 25E[t_c]$).

times $t_d$ are expected for larger variance $V_c$. Therefore, the performance of both $P_{\mathrm{UFT}}$ and $E[t_d]$ degrade as $V_c$ increases.

3) *Effects of the expected intermessage arrival time* $1/\lambda_m$. Fig. 5 plots $P_{\mathrm{UFT}}$ and $E[t_d]$ against $X_T$ and $\lambda_m$, where $X = 25E[t_c]$. Fig. 5(a) shows that $P_{\mathrm{UFT}}$ increases as $\lambda_m$ increases. When $\lambda_m$ increases, more message deliveries are likely to occur during an in-progress call. Therefore, the UFT probability $P_{\mathrm{UFT}}$ increases. For $X_T = 2E[t_c]$, when $1/\lambda_m$ decreases from 50 $E[t_c]$ to 5 $E[t_c]$ and from $5E[t_c]$ to $0.5E[t_c]$, $P_{\mathrm{UFT}}$ increases by 9.13 and 9.01 times, respectively. This effect becomes insignificant when $X_T$ is large (e.g., $X_T \geq 6E[t_c]$). $P_{\mathrm{UFT}}$ is not significantly affected by $\lambda_m$ when $1/\lambda_m > 10E[t_c]$. Fig. 5(b) shows that $E[t_d]$ is insignificantly affected by $\lambda_m$. This phenomenon can be explained as follows. Since the prepaid messages are random-observation points of the prepaid-call holding interval, the delivery delays are not significantly affected by the message arrival rate $\lambda_m$.
4) *Effects of the initial prepaid credit* $X$. Fig. 6 plots $P_{\mathrm{UFT}}$ and $E[t_d]$ against $X_T$ and $X$, where $1/\lambda_m = 0.5E[t_c]$. This figure shows that both $P_{\mathrm{UFT}}$ and $E[t_d]$ decrease as $X$ increases. When the initial prepaid credit $X$ increases,

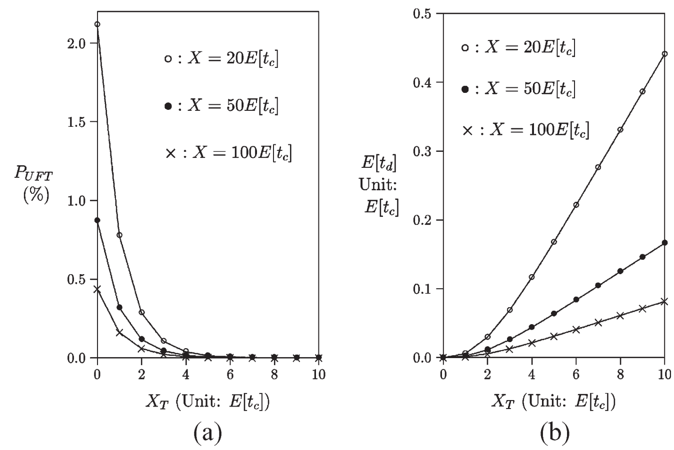



Fig. 6. (a) Effect on $P_{\mathrm{UFT}}$. (b) Effect on $E[t_d]$. Effects of $X$ on $P_{\mathrm{UFT}}$ and $E[t_d]$ ($1/\lambda_m = 0.5E[t_c]$).

there are more CUs left for a prepaid call, and it is more likely that there are enough CUs for both the remaining call and the message (i.e., the amount of the prepaid credits left is larger than $X_T + T_m \geq t_c + T_m$). Therefore, the performance of both $P_{\mathrm{UFT}}$ and $E[t_d]$ improve as $X$ increases. For $X_T = 3E[t_c]$, when $X$ increases from 20 $E[t_c]$ to 50 $E[t_c]$, $P_{\mathrm{UFT}}$ decreases by 58.64%, and $E[t_d]$ decreases by 36.71%. When $X$ increases from 50 $E[t_c]$ to 100 $E[t_c]$, $P_{\mathrm{UFT}}$ decreases by 49.96%, and $E[t_d]$ decreases by 50.87%. When $X > 150E[t_c]$ (i.e., the initial prepaid credit can support more than 150 voice calls), increasing $X$ only has an insignificant impact on $P_{\mathrm{UFT}}$ and $E[t_d]$.
5) *Effects of the threshold* $X_T$. From Figs. 3–6, when $X_T$ is small, increasing $X_T$ reduces $P_{\mathrm{UFT}}$ significantly. When $X_T \geq 5E[t_c]$, the effects of $X_T$ on $P_{\mathrm{UFT}}$ become insignificant. On the other hand, increasing $X_T$ always increases $E[t_d]$. Therefore, in these scenarios, it is appropriate to choose $X_T = 5E[t_c]$ in the PAS.

## V. Conclusion

This paper proposed an SIP-based PAS to handle both the prepaid IMS-to-PSTN calls and messaging services in UMTS. When both voice and messaging are simultaneously offered, a strategy is required to determine if a prepaid message can be sent out during an in-progress call without force-terminating this call. To avoid UFT, a threshold amount $X_T$ of prepaid credit is set to protect the in-progress IMS-to-PSTN call. This paper provided guidelines to select an appropriate $X_T$. The output measures are the UFT probability $P_{\mathrm{UFT}}$ and the expected unnecessary delay $E[t_d]$. We investigated how these two output measures are affected by input parameters, including the expected call holding time $E[t_c]$, the variance $V_c$ of the call holding time, the message arrivals rate $\lambda_m$, $X_T$, and the initial prepaid credit $X$. We make the following observations.

1) When the proportion of $X$ and $E[t_c]$ is fixed, $P_{\mathrm{UFT}}$ and $E[t_d]$ are not affected by the change of $E[t_c]$.
2) The performance of both $P_{\mathrm{UFT}}$ and $E[t_d]$ degrade as $V_c$ increases.

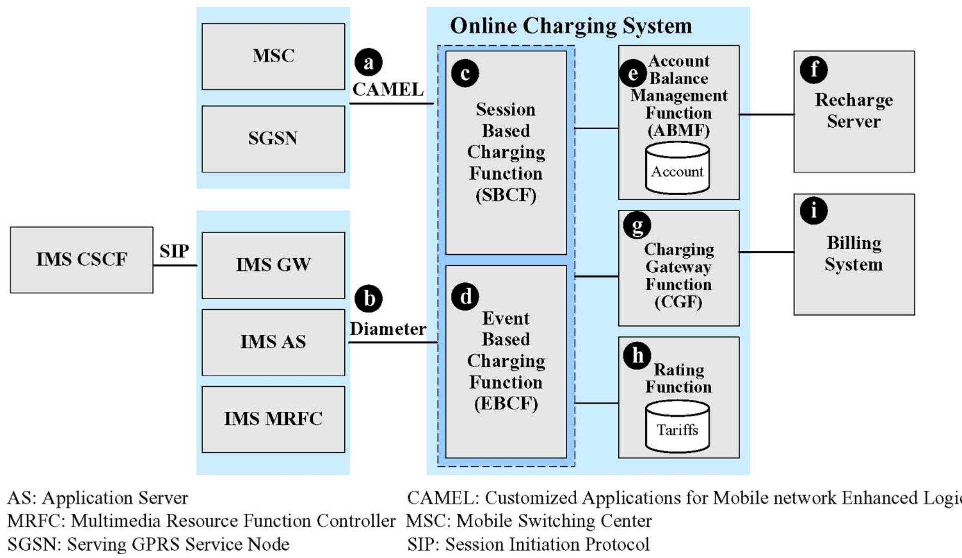



Fig. 7. OCS architecture.

3) $P_{\mathrm{UFT}}$ increases as $\lambda_m$ increases. This effect becomes insignificant when $X_T$ is large (e.g., $X_T \geq 6E[t_c]$ in our examples). On the other hand, $E[t_d]$ is insignificantly affected by $\lambda_m$.
4) Both $P_{\mathrm{UFT}}$ and $E[t_d]$ decrease as $X$ increases.
5) When $X_T$ is small, increasing $X_T$ reduces $P_{\mathrm{UFT}}$ significantly. When $X_T$ is large (e.g. $X_T > 5E[t_c]$ in our examples), increasing $X_T$ only has insignificant impact on $P_{\mathrm{UFT}}$. On the other hand, increasing $X_T$ always increases $E[t_d]$.

Based on the above discussion, the network operator can select the appropriate $X_T$ values for various traffic conditions based on our study.

## Appendix
## Online-Charging System (OCS)

This Appendix provides details of the OCS shown in Fig. 1(h). Fig. 7 shows the architecture of the OCS defined in 3GPP TS 32.296 [6]. Online charging for Global System for Mobile Communications and General Packet Radio Service (GPRS) is performed by using CAMEL [see Fig. 7(a)] [4]. Online charging for the IMS services is performed by using the DCC protocol [see Fig. 7(b)] [8]. In Fig. 7, the OCS consists of two charging functions: the session-based charging function [SBCF; Fig. 7(c)] and the event-based charging function [EBCF; Fig. 7(d)]. The SBCF is responsible for online charging of network bearer and user sessions, such as voice calls, GPRS packet data protocol contexts, and IMS sessions. The EBCF performs event-based online charging (also referred to as "content charging") for the IMS network elements [e.g., the CSCF, which sends through the IMS GW, AS, and MRFC in Fig. 7(b)].

In the OCS, the account-balance-management function [ABMF; Fig. 7(e)] keeps a user's balance and other account data. When the prepaid user's credit depletes, the ABMF connects the recharge server [Fig. 7(f)] to trigger the recharge-account function. The charging functions (i.e., the SBCF and EBCF) interact through the rating function [Fig. 7(h)] to determine the price of the requested service. The rating function handles a wide variety of rateable instances, such as data volume, session connection time, and event service (e.g., for web-content charging). The charging functions interact through the ABMF to query and update the user's account. The charging-data records generated by the charging functions are transferred to the charging gateway function [CGF; Fig. 7(g)] immediately. The CGF acts as a gateway between the 3GPP network and the billing system [Fig. 7(i)].

## Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments, which have significantly improved the quality of this paper, and for their efforts.

## References

[1] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Architectural Requirements for Release 1999*, Tech. Rep. 3G TS 23.121 Version 3.4.0 (2000-10), 3GPP, 2000.
[2] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2*, Tech. Specification 3G TS 23.060 Version 4.1.0 (2001-06), 2001.
[3] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; IP Multimedia Subsystem Stage 2*, Tech. Specification 3G TS 23.228 Version 6.2.0 (2003-06), 2003.
[4] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Core Network; Customised Applications for Mobile Network Enhanced Logic (CAMEL) Phase 4; Stage 2*, Tech. Specification 3G TS 23.078 Version 5.10.0 (2005-06), 2005.
[5] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Core Network; IP Multimedia (IM) Session Handling; IM Call Model; Stage 2*, Tech. Specification 3G TS 23.218 Version 5.8.0 (2005-03), 2005.
[6] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication Management; Charging Management; Online Charging System (OCS): Applications and Interfaces. 3rd Generation Partnership Project*, Tech. Specification 3G TS 32.296 Version 6.1.0 (2005-06), 2005.

[7] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication Management; Charging Management; Charging Data Description for the IP Multimedia Subsystem (IMS)*, Tech. Specification 3G TS 32.225 Version 5.8.0 (2005-03), 2005.

[8] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Service and System Aspects; Telecommunication Management; Charging Management; Diameter Charging Applications*, Tech. Specification 3G TS 32.299 Version 6.3.0 (2005-06), 2005.

[9] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Support of SMS and MMS Over Generic 3GPP IP Access*, Tech. Specification 3G TR 23.804 Version 7.1.0 (2005-09), 2005.

[10] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Terminals; Technical Realization of the Short Message Service (SMS)*, Tech. Specification 3G TS 23.040 Version 6.5.0 (2004-09), 2005.

[11] 3GPP, *3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; IP Multimedia Call Control Protocol Based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3*, Tech. Specification 3G TS 24.229 Version 5.16.0 (2006-03), 2006.

[12] B. Campbell, J. Rosenberg, H. Schulzrinne, C. Huitema, and D. Gurle, *Session Initiation Protocol (SIP) Extension for Instant Messaging*, IETF RFC 3428, Dec. 2002.

[13] H. Hakala, L. Mattila, J.-P. Koskinen, M. Stura, and J. Loughney, *Diameter Credit-Control Application*, IETF RFC 4006, Aug. 2005.

[14] F. P. Kelly, *Reversibility and Stochastic Networks*. Hoboken, NJ: Wiley, 1979.

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

[16] Y.-B. Lin and A.-C. Pang, *Wireless and Mobile All-IP Networks*. Hoboken, NJ: Wiley, 2005.

[17] Y.-B. Lin, Y.-R. Huang, A.-C. Pang, and I. Chlamtac, “All-IP approach for third generation mobile networks,” *IEEE Netw.*, vol. 16, no. 5, pp. 8–19, Sep./Oct. 2002.

[18] H. C.-H. Rao, Y. H. Cheng, K. S. Chang, and Y.-B. Lin, “iMail: A WAP mail retrieving system,” *Inf. Sci.*, vol. 151, pp. 71–91, 2003.

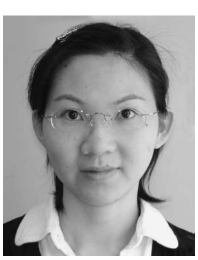
[19] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, *SIP: Session Initiation Protocol,* IETF RFC 3261, Jun. 2002.

[20] S. M. Ross, *Stochastic Processes*. Hoboken, NJ: Wiley, 1996.

[21] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications,* IETF RFC 1889, Jan. 1996.

[22] S.-I. Sou, Q. Wu, Y.-B. Lin, and C.-H. Yeh, “Prepaid mechanism of VoIP and messaging services,” in *Proc. IEEE Int. Conf. ITRE*, 2005, pp. 255–259.

[23] *Vovida*. [Online]. Available: http://www.vovida.org

**Sok-Ian Sou** (S’06) received the B.S.CSIE. and M.S.CSIE. degrees from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1997 and 2004, respectively, where she is currently working toward the Ph.D. degree at the Department of Computer Science.

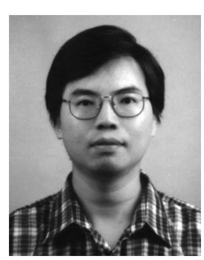
Her current research interests include design and analysis of personal communications services networks, mobile computing, and performance modeling.

**Yi-Bing Lin** (M’96–SM’96–F’03) is currently a Chair Professor and the Vice President of Research and Development, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. His current research interests include wireless communications and mobile computing. He has published over 190 journal articles and more than 200 conference papers. He is the coauthor of the book *Wireless and Mobile Network Architecture* (with I. Chlamtac, New York: Wiley, 2001).

Dr. Lin is an ACM Fellow, an AAAS Fellow, and an IEE Fellow.

**Quincy Wu** received the B.S. degree in mathematics and the Ph.D. degree in computer science and information engineering from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1992 and 2000, respectively.

He was with National Center for High-Performance Computing with the National Broadband Experimental Network Project, where he successfully designed and established the first island-wide IPv6 network among universities. In 2002, he began serving as a Research Assistant Professor with National Chiao Tung University, Hsinchu, and helped the National Telecommunications Program Office to deploy a session-initiation-protocol (SIP)-based VoIP platform across several universities. Since 2004, he has been cochairing the SIP-H323 Working Group of Asia-Pacific Advanced Network and has helped the Taiwan Academic Network to design and deploy VoIP services. In 2005, he was appointed Assistant Professor with the Graduate Institute of Communication Engineering, National Chi Nan University, Nantou, Taiwan, where he is currently helping to initiate the VoIP over WiMAX Project. His current research interests include session initiation protocols, open-service architecture, Internet protocol version 6, design and analysis of approximation algorithms, and service creation on the third-generation mobile network.



**Jeu-Yih Jeng** received the B.S. degree in mathematics from Fu-Jen University, Taipei, Taiwan, R.O.C., in 1983 and the M.S. degree in applied mathematics and the Ph.D. degree in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1985 and 1998, respectively.

Since 1985, he has been with Information Technology Laboratory, Telecommunication Laboratories, Chunghwa Telcom Company, Ltd., where he is currently a Distinguished Researcher and a Project Manager. His research interests include the design and analysis of the personal communications services network, the development of telecommunication-operation support systems, and performance modeling.