A Weakly Consistent Scheme for IMS Presence Service

Whai-En Chen, *Member, IEEE,* Yi-Bing Lin, *Fellow, IEEE,* and Ren-Huang Liou

***Abstract*—IP Multimedia Core Network Subsystem (IMS) provides presence service for Universal Mobile Telecommunications System (UMTS). In IMS, the presence server is responsible for notifying an authorized watcher of the updated presence information. If the updates occur more frequently than the accesses of the watcher, the presence server will generate many notifications. This paper uses a weakly consistent scheme (called delayed update) to reduce the notification traffic. In this scheme, a delayed timer is defined to control the notification rate. We propose an analytic model and simulation experiments to investigate the performance of delayed update. The study indicates that delayed update can effectively reduce the notification traffic without significantly degrading the valid access probability.**

***Index Terms*—Delayed update, IP multimedia core network subsystem (IMS), Presence Service.**

I. Introduction

PRESENCE service allows an Internet user to access another user's presence information including the user status, the activities (e.g., working, playing, etc.), the email/phone addresses, and so on. This service is typically utilized together with the instant messaging applications such as Windows Live Messenger.

In *Universal Mobile Telecommunication System* (UMTS), presence service is provided through *IP Multimedia Core Network Subsystem* (IMS) [2], [8]. Fig. 1 illustrates a simplified UMTS network architecture for presence service [1]. In this architecture, a user with a *User Equipment* (UE; Fig. 1 (1)) accesses presence service in IMS (Fig. 1 (b)) via the UMTS (Fig. 1 (a)). The user plays the role as a *presentity* when he/she provides information to the presence server (Fig. 1 (3)). On the other hand, the user is called a *watcher* if he/she accesses other users' information from the presence server. In IMS, *Call Session Control Function* (CSCF; Fig. 1 (2)) is responsible for carrying out the control signaling. As an IMS application server, the presence server stores the information of the presentities, and determines the authorized watchers who can access the presence information.

Manuscript received October 28, 2008; revised January 19, 2009; accepted February 27, 2009. The associate editor coordinating the review of this paper and approving it for publication was D. Wu.

W.-E. Chen is with the Institute of Computer Science and Information Engineering, National I-Lan University, Taiwan (e-mail: wechen@niu.edu.tw).

Y.-B. Lin is with the Department of Computer Science, National Chiao Tung University, Taiwan. He is also with the Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan (e-mail: liny@cs.nctu.edu.tw).

R.-H. Liou is with the Department of Computer Science, National Chiao Tung University, Taiwan (e-mail: rhliou@cs.nctu.edu.tw).

This work was sponsored in part by NSC 97-2221-E-009-143-MY3, NSC 97-2219-E009-016, NSC 97-2219-E197-001, NSC 97-2219-E197-003, Chung Hwa Telecom, ITRI/NCTU Joint Research Center and MoE ATU.

Digital Object Identifier 10.1109/TWC.2009.081433

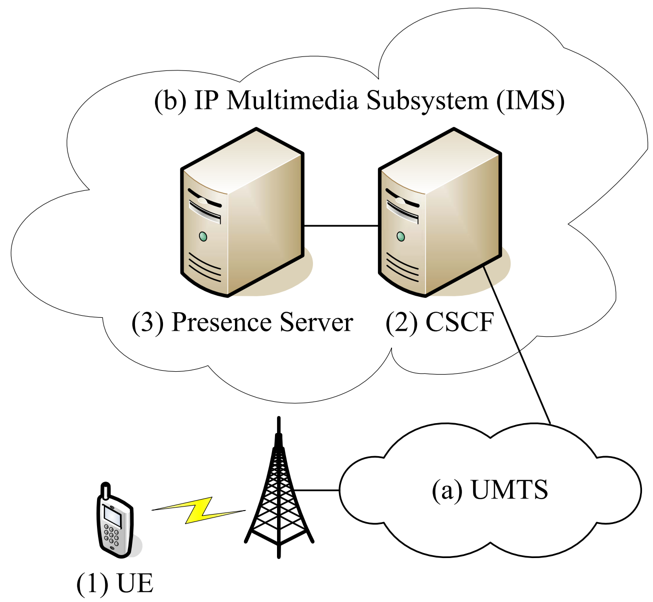

Fig. 1. The UMTS Network Architecture for Presence Service

3rd Generation Partnership Project (3GPP) defines the presence service procedures for IMS. These procedures may incur large traffic that significantly consumes the limited radio resources and UE power. To resolve this issue, we consider a weakly consistent scheme called *delayed update* exercised at the presence server to reduce the presence traffic between a watcher and the presence server. In this scheme, the UE needs not be modified.

This paper is organized as follows. Section II describes the message flows for the 3GPP IMS presence service, and then introduces delayed update. Section III describes the input parameters and output measures for modeling delayed update. Sections IV-V propose an analytic model to study the performance of delayed update. Section VI investigates delayed update by numerical examples, and the conclusions are given in Section VII.

II. Message Flows for IMS Presence Service

This section describes the message flows for the 3GPP presence service procedures including the subscription, the publication, and the notification [1], [3]. These procedures are implemented by utilizing *Session Initiation Protocol* (SIP) [9]–[11], [15]. Then we introduce delayed update for reducing the notification traffic.

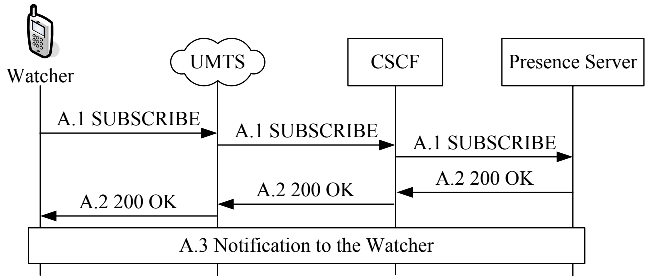

Fig. 2. The Subscription Procedure

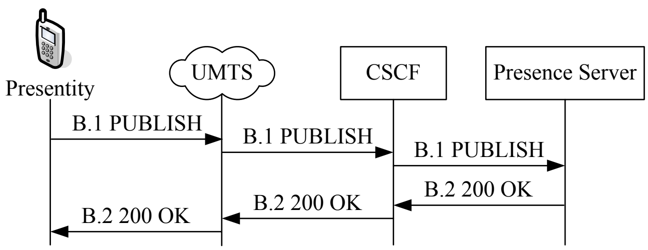

Fig. 3. The Publication Procedure

A. Subscription Procedure

When a watcher attempts to access the presence information of other users (i.e., the presentities), the subscription procedure is executed. Fig. 2 illustrates the subscription message flow between the watcher and the presence server. Detailed steps are described as follows:

Step A.1: The watcher (i.e., a UE) issues a SIP SUBSCRIBE message to the presence server through the UMTS and the CSCF.

Step A.2: After verifying the identify of the watcher [4], the presence server sends a 200 OK message to the watcher. The message indicates that the watcher is authorized to subscribe to the presence information.

Step A.3: The presence server sends the presence information to the watcher through the notification procedure to be described in Fig. 4.

B. Publication Procedure

The publication procedure is executed when a presentity changes his/her presence information (e.g., the activity is changed from "playing" to "working"). Fig. 3 illustrates the publication message flow with the following steps:

Step B.1: A SIP PUBLISH message is issued by the presentity to the presence server through the UMTS and the CSCF. This message contains the updated presence information.

Step B.2: After the presence server received this SIP PUBLISH message, it stores the presence information in association with this presentity, and returns the 200 OK message. This message indicates that the publication is successful.

C. Notification Procedure

The presence server pushes the subscribed presence information to an authorized watcher through the notification procedure when the information is updated. Fig. 4 illustrates the notification message flow with the following steps:

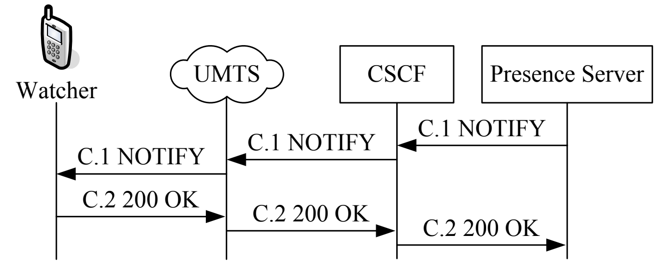

Fig. 4. The Notification Procedure

Step C.1: The presence server sends a SIP NOTIFY message to the watcher through the CSCF and the UMTS. This message contains the modified presence information.

Step C.2: The watcher returns the 200 OK message to indicate that the presence information is successfully received.

Note that the notification is a "push" action. A watcher may access the presence information through "pull" action such as the poll-each-read mechanism. The details can be found in [7].

D. Delayed Update

The presence server immediately notifies a watcher of the presence information updates. If the updates occur more frequently than the accesses of the watcher, then the presence server will generate many notifications. A weakly consistent scheme called delayed update can be used to reduce the notification traffic between the presence server and the watcher (i.e., Steps C.1 and C.2 in Fig. 4).

In delayed update, when the presence server receives the updated presence information (i.e., Step B.1 in Fig. 3), the watcher is not notified of the updated information immediately. Instead, the presence server starts a delayed timer with a period T. This period is referred to as the *delayed threshold*. If the presence information is updated again within period T, the old information in the presence server is replaced by the new one. When the timer expires, the presence server notifies the watcher of the presence information (i.e., Step C.1 in Fig. 4 is executed). Therefore, the notifications for the updates in T are saved. On the other hand, if an access occurs at any time in T, the watcher may access the obsolete information. RFC 3856 describes a mechanism similar to delayed update, where $T = 5$ seconds [12]. However, this default delayed threshold value may not fit all user activities, and it is important to select an appropriate T value such that notification traffic is reduced while the watcher only occasionally accesses the obsolete presence information.

III. Input Parameters and Output Measures

This section describes the input parameters and output measures for modeling delayed update. Fig. 5 illustrates a timing diagram where the watcher accesses the presence information at t_0, t_3, and t_8. The presence information is updated at t_1, t_4, t_5, t_6, and t_9. At t_2 and t_7, the presence server notifies the watcher of modified presence information.

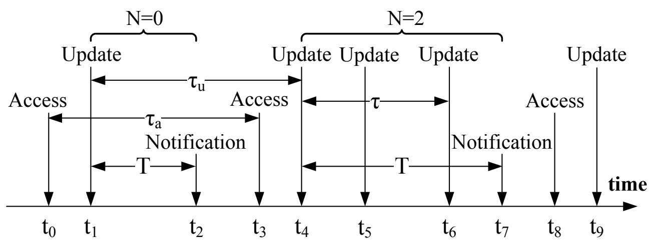

Fig. 5. Timing Diagram for Delayed Update

In this figure, the delayed threshold $T = t_2 - t_1$ (also $t_7 - t_4$) is a random variable with the mean $1/\gamma$. Let N be the number of update messages occurring during T. In this example, $N = 0$ in $[t_1, t_2]$, and $N = 2$ in $[t_4, t_7]$. The inter-access interval $\tau_a = t_3 - t_0$ (also $t_8 - t_3$) is a random variable with the mean $1/\mu$ and the variance V_a. Let the inter-update interval $\tau_u = t_4 - t_1$ (also $t_5 - t_4$, $t_6 - t_5$, and $t_9 - t_6$) be a random variable which has density function $f_u(\tau_u)$, the mean $1/\lambda$, the variance V_u, and Laplace transform $f_u^*(s)$. Let $\tau = t_6 - t_4$ be the interval between the first update message and the last update message occurring in T. We define an *observation interval* as a period between when the delayed threshold begins and when the first update message occurs after the end of the delayed threshold. In Fig. 5, $[t_1, t_4]$ and $[t_4, t_9]$ are the observation intervals.

We consider two output measures:

- p: the probability that the watcher accesses the valid presence information. In Fig. 5, an access is valid if it occurs in $[t_2, t_4]$ and $[t_7, t_9]$.
- $E[N]$: the expected number of update messages occurring in T.

It is clear that the update messages occurring in T are the saved notification messages. Therefore, the larger the p and the $E[N]$ values, the better the performance of delayed update.

The notation used in this paper is summarized below.

- T: the delayed threshold.
- p: the probability that the watcher accesses the valid presence information.
- N: the number of update messages occurring in T.
- τ_u: the inter-update interval of the presence information.
- τ: the interval between the first update message and the last update message occurring in T.
- τ_a: the inter-access interval of the watcher.
- $1/\lambda = E[\tau_u]$: mean inter-update interval.
- $1/\mu = E[\tau_a]$: mean inter-access interval.
- $1/\gamma = E[T]$: mean delayed threshold.
- V_a: the variance for the τ_a distribution.
- V_u: the variance for the τ_u distribution.
- $f_u(\cdot)$: density function for the τ_u distribution.
- $f_u^*(\cdot)$: Laplace transform for the τ_u distribution.

IV. Derivation for $E[N]$ and p

This section describes an analytic model for deriving $E[N]$ and p. If the inter-update interval τ_u has the exponential distribution with the mean $1/\lambda$, then due to the Poisson property of the update arrivals, for a delayed threshold T with an arbitrary distribution with the mean $1/\gamma$, the expected number $E[N]$ of update messages occurring in this period is

$$E[N] = \lambda E[T] = \frac{\lambda}{\gamma} \tag{1}$$

If τ_u is not exponentially distributed, (1) does not hold, and $E[N]$ is derived as follows. Suppose that T is exponentially distributed, and τ_u has a general distribution. Assume that the first update message in an observation interval occurs at t_0, and the ith subsequent update message occurs at t_i. Let $\tau_{u,i} = t_i - t_{i-1}$ for $i \geq 1$. Define $t_{u,n} = \sum_{i=1}^{n} \tau_{u,i}$. Let $\Pr[N = n]$ be the probability that there are n update messages occurring in T (excluding the update at t_0). Then

$$\Pr[N = n] \quad = \quad \Pr[t_{u,n} < T < t_{u,n+1}] \tag{2}$$

For exponential delayed threshold, (2) can be re-written as (3). From (3), $E[N]$ is derived as

$$\begin{aligned} E[N] &= \sum_{n=0}^{\infty} \{n \Pr[N = n]\} \\ &= \sum_{n=1}^{\infty} \{n\,[f_u^*(\gamma)]^n\,[1 - f_u^*(\gamma)]\} \\ &= \frac{f_u^*(\gamma)}{1 - f_u^*(\gamma)} \end{aligned} \tag{4}$$

If τ_u has the Gamma distribution with the mean $1/\lambda$ and the variance V_u, then its Laplace transform is

$$f_u^*(s) \quad = \quad \left(\frac{1}{V_u \lambda s + 1}\right)^{\frac{1}{V_u \lambda^2}} \tag{5}$$

It has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions [6], and is therefore selected in our study to represent the inter-update interval distribution. Substitute (5) into (4) to yield

$$E[N] \quad = \quad \frac{1}{(V_u \lambda \gamma + 1)^{\frac{1}{V_u \lambda^2}} - 1} \tag{6}$$

Following the past experience [5], [14], [16], [17], we can measure the inter-update intervals of the presence information in a real mobile network, and generate the Gamma distribution from the measured data. Then (6) can be used to compute $E[N]$.

Next, we derive p. Note that probability $1-p$ is proportional to the expected delayed threshold $E[T]$, and is inversely proportional to the expected observation interval $(1+E[N])E[\tau_u]$ [13]. Therefore, p can be expressed as

$$p \quad = \quad 1 - \frac{E[T]}{(1 + E[N])E[\tau_u]} \tag{7}$$

When τ_u is exponentially distributed, (7) is re-written as

$$p \quad = \quad \frac{\gamma}{\lambda + \gamma} \tag{8}$$

In Appendix A, we also derive p by actually using the exponential τ_u distribution for fixed and exponential delayed thresholds. The results are the same as (8). When the inter-update interval τ_u is not exponentially distributed, (8) does not hold, and its variance V_u has significant impact on p.

$$\begin{aligned}\Pr[N=n] &= \int_{\tau_{u,1}=0}^{\infty} f_u(\tau_{u,1})\cdots\int_{\tau_{u,n}=0}^{\infty} f_u(\tau_{u,n})\int_{\tau_{u,n+1}=0}^{\infty} f_u(\tau_{u,n+1})\int_{T=t_{u,n}}^{t_{u,n+1}} \gamma e^{-\gamma T} dT d\tau_{u,n+1} d\tau_{u,n}\cdots d\tau_{u,1} \\ &= [f_u^*(\gamma)]^n\,[1-f_u^*(\gamma)] \end{aligned} \quad (3)$$

Assume that T is exponentially distributed and τ_u has a general distribution, then from (4) and (7)

$$p = \frac{\gamma-\lambda}{\gamma} + \left(\frac{\lambda}{\gamma}\right) f_u^*(\gamma) \quad (9)$$

If τ_u has the Gamma distribution with the mean $1/\lambda$ and the variance V_u, then (9) is re-written as

$$p = \frac{\gamma-\lambda}{\gamma} + \left(\frac{\lambda}{\gamma}\right)\left(\frac{1}{V_u\lambda\gamma+1}\right)^{\frac{1}{V_u\lambda^2}} \quad (10)$$

V. Simulation Validation

The purpose of the analytic model in Section IV is two folds. First, it provides mean value analysis ((1) and (8)) to show the trends of the impacts of λ and γ on the output measures. Second, the analytic model is used to validate the simulation experiments. We validate the $E[N]$ values of the simulation experiments for the exponential τ_u and arbitrary T distributions (by using (1)), and the Gamma τ_u and exponential T distributions (by using (6)). We validate the p values of the simulation experiments for the exponential τ_u and arbitrary T distributions (by using (8)), and the Gamma τ_u and exponential T distributions (by using (10)). In this section, we assume that τ_a is exponentially distributed with the mean $1/\mu$. In Section VI, we will investigate the relationship between p and the distribution of the inter-access interval τ_a. Tables I and II compare p and $E[N]$ values for the analytic and simulation results under different μ and V_u values. Table I lists the errors between the analytic and simulation results, where $\lambda = \gamma$, $V_a = 1/\mu^2$, and $V_u = 1/\lambda^2$ with various μ values. The table indicates that the errors between the analytic and simulation results are within 0.5%. Table II lists the errors between the analytic and simulation results, where $\lambda = \gamma = \mu$ and $V_a = 1/\mu^2$ with various V_u values. For fixed T, we can not derive analytic equation for $E[N]$. We use the $E[N]$ values from simulation to compute (7), which is consistent with the p values directly obtained from the simulation experiments. The table shows that the errors between the analytic and simulation results are within 0.3%. Therefore, the analytic analysis is consistent with the simulation results.

TABLE I

Comparison of Analytic and Simulation Results under Different μ Values ($\lambda = \gamma$, $V_a = 1/\mu^2$, and $V_u = 1/\lambda^2$)

Fixed delayed threshold			
μ	0.1γ	γ	10γ
p (Analytic)	0.5	0.5	0.5
p (Simulation)	0.500146	0.499889	0.500036
Error	0.0292%	0.0222%	0.0071%
$E[N]$ (Analytic)	1	1	1
$E[N]$ (Simulation)	0.999317	1.00076	1.00233
Error	0.0682%	0.0759%	0.2326%
Exponential delayed threshold			
μ	0.1γ	γ	10γ
p (Analytic)	0.5	0.5	0.5
p (Simulation)	0.499937	0.499549	0.500286
Error	0.0126%	0.0902%	0.0571%
$E[N]$ (Analytic)	1	1	1
$E[N]$ (Simulation)	0.999864	1.00187	0.994636
Error	0.0135%	0.1867%	0.5364%

TABLE II

Comparison of Analytic and Simulation Results under Different V_u Values ($\lambda = \gamma = \mu$ and $V_a = 1/\mu^2$)

Fixed delayed threshold				
V_u	$0.001/\lambda^2$	$0.1/\lambda^2$	$10/\lambda^2$	$1000/\lambda^2$
p (Analytic + Simulation)	0.335125	0.35292	0.798605	0.993404
p (Simulation)	0.334833	0.352609	0.799029	0.993215
Error	0.0871%	0.0882%	0.0531%	0.0189%
Exponential delayed threshold				
V_u	$0.001/\lambda^2$	$0.1/\lambda^2$	$10/\lambda^2$	$1000/\lambda^2$
p (Analytic)	0.368063	0.385543	0.786793	0.993115
p (Simulation)	0.367212	0.385587	0.785982	0.993004
Error	0.2312%	0.0113%	0.1031%	0.0111%
$E[N]$ (Analytic)	0.582437	0.627454	3.69029	144.244
$E[N]$ (Simulation)	0.58043	0.627413	3.69701	144.723
Error	0.3445%	0.0064%	0.1822%	0.3319%

VI. Numerical Examples

This section investigates the performance of delayed update.

Effects of γ/λ and V_u on probability p: Fig. 6 plots p against $\gamma = \frac{1}{E[T]}$ (normalized by λ) and the variance V_u of the inter-update intervals. From the mean value analysis (8), it is clear that p increases as γ/λ increases (i.e., the delayed threshold T decreases). When $V_u \le \frac{1}{\lambda^2}$, p is not sensitive to the change of γ/λ. This figure also indicates that p increases as V_u increases and large p is observed. This phenomenon is explained as follows. When the inter-update interval τ_u becomes more irregular (i.e., V_u increases), more long and short inter-update intervals are observed. Since access events are more likely to fall in long inter-update intervals and are not in T, larger p is observed. When V_u is very large, p is not sensitive to the γ/λ values, and large p is always observed.

Effects of γ/λ and V_u on $E[N]$: Fig. 7 plots $E[N]$ against γ/λ and V_u. From the mean value analysis (1), it is clear that $E[N]$ increases as γ/λ decreases. This figure also indicates that $E[N]$ increases as V_u increases. For a fixed $E[\tau_u]$ value, when V_u increases, there are more short τ_u periods than long τ_u periods. Therefore, it is likely to find many consecutive update messages occurring in T (i.e., larger $E[N]$). When V_u is very small, $E[N]$ is not sensitive to the V_u values.

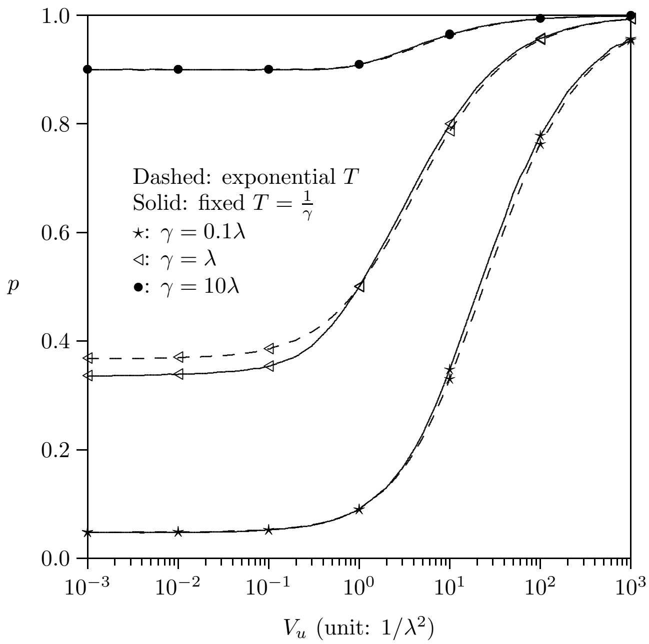

Fig. 6. Effects of γ/λ and V_u on p

TABLE III
EFFECT OF THE τ_a DISTRIBUTION ON p ($\lambda = \gamma$ AND $V_u = 100/\lambda^2$)

Gamma distribution				
V_a	$0.001/\mu^2$	$0.1/\mu^2$	$10/\mu^2$	$1000/\mu^2$
p (fixed T)	0.958406	0.957799	0.958335	0.957274
p (exponential T)	0.955652	0.955313	0.953606	0.955345
Weibull distribution				
V_a	$0.001/\mu^2$	$0.1/\mu^2$	$10/\mu^2$	$1000/\mu^2$
p (fixed T)	0.957609	0.957626	0.957932	0.957613
p (exponential T)	0.954867	0.954526	0.954771	0.954768
Lognormal distribution				
V_a	$0.001/\mu^2$	$0.1/\mu^2$	$10/\mu^2$	$1000/\mu^2$
p (fixed T)	0.957667	0.957586	0.957652	0.957407
p (exponential T)	0.954767	0.954666	0.954458	0.954906

Effects of the T distribution on p and $E[N]$: Fig. 6 indicates that the exponential delayed threshold is better than the fixed delayed threshold when V_u is small. On the other hand, the fixed delayed threshold is slightly better than the exponential delayed threshold when V_u is large. For the $E[N]$ performance, Fig. 7 indicates that the fixed delayed threshold slightly outperforms the exponential delayed threshold when V_u is large.

Effect of the τ_a distribution on p: Simulation experiments indicate that the distribution of the inter-access interval τ_a does not affect p. Assume that τ_u has the Gamma distribution with the mean $1/\lambda$ and the variance V_u. We use the Gamma, Weibull, and lognormal distributions for τ_a with the mean $1/\mu$ and the variance V_a. Table III shows the effect of the τ_a distributions on p, where $\lambda = \gamma$ and $V_u = 100/\lambda^2$. From this table, p is independent of the τ_a distributions.

VII. CONCLUSIONS

This paper investigated the performance of delayed update for IMS presence service. Both the fixed and the exponential

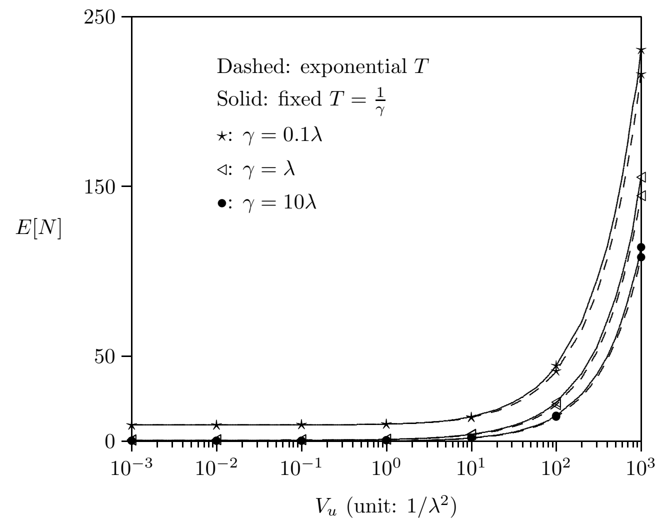

Fig. 7. Effects of γ/λ and V_u on $E[N]$

delayed thresholds are considered in our study. The performance is measured by the valid access probability p and the expected number $E[N]$ of the update messages occurring in T. Our study indicated that delayed update can effectively improve p and $E[N]$ when the variance V_u of the inter-update interval is large (when the update behavior is irregular). Furthermore, it is appropriate to select the exponential delayed threshold when V_u is small, and the fixed delayed threshold should be selected when V_u is large.

As a final remark, the amount of notification traffic to be reduced heavily depends on the operation status of the mobile networks, and is determined in network planning of mobile operators. On the other hand, the valid access probability p is determined by the *Quality of Service* (QoS) policy. By considering both network planning and the QoS policy, a Taiwan's mobile operator, for example, decides that it is acceptable if inaccuracy of the presence information caused by delayed update is less than 10%. For a VIP subscriber, $T = 0$ (delayed update is not exercised). For a flat-rate subscriber, T is set such that p is around 90%-95%. Also, in a heterogeneous wireless network environment (e.g., in the Taipei city), delayed update is triggered through tier-switch (from WiFi to 2.5G/3G). When a subscriber is connected to WiFi (where wireless resources are abundant), delayed update is not exercised ($T = 0$). When the subscriber is switched from WiFi to 2.5G/3G (with higher wireless cost), T is set such that p is around 90%-95%.

APPENDIX
AN ALTERNATIVE DERIVATION FOR p

This appendix shows an alternative for deriving p to double check that (8) is correct. To compute p, two cases are considered.

Case I. Under the condition that no update message occurs in T, the expected delayed threshold and the expected observation interval are $E[T|T < \tau_u]$ and $E[\tau_u|T < \tau_u]$, respectively.

Case II. Under the condition that one or more update messages occur in T, the expected delayed threshold and the

$$p = 1 - \frac{E[T|T<\tau_u]\Pr[T<\tau_u] + E[T|\tau<T<\tau+\tau_u]\Pr[\tau<T<\tau+\tau_u]}{E[\tau_u|T<\tau_u]\Pr[T<\tau_u] + E[\tau+\tau_u|\tau<T<\tau+\tau_u]\Pr[\tau<T<\tau+\tau_u]} \tag{11}$$

expected observation interval are $E[T|\tau < T < \tau+\tau_u]$ and $E[\tau+\tau_u|\tau<T<\tau+\tau_u]$, where τ is the interval between the first update message and the last update message occurring in T.

We note that probability $1-p$ is proportional to the expected delayed threshold $E[T]$, and is inversely proportional to the expected observation interval. Therefore, p can be expressed as (11).

Since $E[X|Y] = E[X \ \& \ Y]/\Pr[Y]$, (11) is re-written as

$$p = 1 - \frac{E[T \ \& \ T<\tau_u] + E[T \ \& \ \tau<T<\tau+\tau_u]}{E[\tau_u \ \& \ T<\tau_u] + E[\tau+\tau_u \ \& \ \tau<T<\tau+\tau_u]} \tag{12}$$

Based on (12), we derive p for fixed and exponential T assuming that τ_u is exponentially distributed.

A. Fixed Delayed Threshold

When T is fixed, we have

$$E[T \ \& \ T<\tau_u] = \int_{\tau_u=T}^{\infty} T\lambda e^{-\lambda\tau_u} d\tau_u = \left(\frac{1}{\gamma}\right) e^{-\frac{\lambda}{\gamma}} \tag{13}$$

$$E\left[\begin{array}{c} T \ \& \\ \tau<T<\tau+\tau_u \end{array}\right] = \sum_{N=1}^{\infty}\left\{\int_{\tau=0}^{T} T\left[\frac{(\lambda\tau)^{N-1}}{(N-1)!}\right]\lambda e^{-\lambda\tau} \times \int_{\tau_u=T-\tau}^{\infty}\lambda e^{-\lambda\tau_u} d\tau_u d\tau\right\} = \left(\frac{1}{\gamma}\right)(1-e^{-\frac{\lambda}{\gamma}}) \tag{14}$$

$$E[\tau_u \ \& \ T<\tau_u] = \int_{\tau_u=T}^{\infty} \tau_u\lambda e^{-\lambda\tau_u} d\tau_u = \left(\frac{1}{\gamma}+\frac{1}{\lambda}\right) e^{-\frac{\lambda}{\gamma}} \tag{15}$$

and

$$E\left[\begin{array}{c} \tau+\tau_u \ \& \\ \tau<T<\tau+\tau_u \end{array}\right] = \sum_{N=1}^{\infty}\left\{\int_{\tau=0}^{T}\left[\frac{(\lambda\tau)^{N-1}}{(N-1)!}\right]\lambda e^{-\lambda\tau} \times \int_{\tau_u=T-\tau}^{\infty}(\tau+\tau_u)\lambda e^{-\lambda\tau_u} d\tau_u d\tau\right\} = \left(\frac{1}{\gamma}+\frac{1}{\lambda}\right)(1-e^{-\frac{\lambda}{\gamma}}) \tag{16}$$

From (13), (14), (15), and (16), (12) is re-written as

$$p = \frac{\gamma}{\lambda+\gamma} \tag{17}$$

B. Exponential Delayed Threshold

If T has the exponential distribution with the mean $1/\gamma$, then

$$E[T \ \& \ T<\tau_u] = \int_{T=0}^{\infty} T\gamma e^{-\gamma T}\int_{\tau_u=T}^{\infty}\lambda e^{-\lambda\tau_u} d\tau_u dT = \frac{\gamma}{(\lambda+\gamma)^2} \tag{18}$$

$$E\left[\begin{array}{c} T \ \& \\ \tau<T<\tau+\tau_u \end{array}\right] = \sum_{N=1}^{\infty}\left\{\int_{T=0}^{\infty} T\gamma e^{-\gamma T}\int_{\tau=0}^{T}\left[\frac{(\lambda\tau)^{N-1}}{(N-1)!}\right] \times \lambda e^{-\lambda\tau}\int_{\tau_u=T-\tau}^{\infty}\lambda e^{-\lambda\tau_u} d\tau_u d\tau dT\right\} = \frac{1}{\gamma} - \frac{\gamma}{(\lambda+\gamma)^2} \tag{19}$$

$$E[\tau_u \ \& \ T<\tau_u] = \int_{T=0}^{\infty}\gamma e^{-\gamma T}\int_{\tau_u=T}^{\infty}\tau_u\lambda e^{-\lambda\tau_u} d\tau_u dT = \frac{\gamma}{(\lambda+\gamma)^2} + \frac{\gamma}{\lambda(\lambda+\gamma)} \tag{20}$$

and

$$E\left[\begin{array}{c} \tau+\tau_u \ \& \\ \tau<T<\tau+\tau_u \end{array}\right] = \sum_{N=1}^{\infty}\left\{\int_{T=0}^{\infty}\gamma e^{-\gamma T}\int_{\tau=0}^{T}\left[\frac{(\lambda\tau)^{N-1}}{(N-1)!}\right] \times\lambda e^{-\lambda\tau}\int_{\tau_u=T-\tau}^{\infty}(\tau+\tau_u)\lambda e^{-\lambda\tau_u} d\tau_u d\tau dT\right\} = \frac{1}{\gamma}+\frac{1}{\lambda}-\frac{\gamma}{(\lambda+\gamma)^2}-\frac{\gamma}{\lambda(\lambda+\gamma)} \tag{21}$$

From (18), (19), (20), and (21), (12) is re-written as

$$p = \frac{\gamma}{\lambda+\gamma} \tag{22}$$

Both (17) and (22) indicate that p for fixed delayed threshold is the same as that for exponential delayed threshold.

REFERENCES

[1] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Presence Service; Architecture and functional description," Technical Specification 3G TS 23.141 version 8.1.0 (2008-06), 2008.

[2] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2," Technical Specification 3G TS 23.228 version 8.5.0 (2008-06), 2008.

[3] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Presence service using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3," Technical Specification 3G TS 24.141 version 8.1.0 (2008-06), 2008.

[4] 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," Technical Specification 3G TS 24.229 version 8.4.1 (2008-06), 2008.

[5] P. G. Escalle, V. C. Giner, J. M. and Oltra, "Reducing location update and paging costs in a PCS network," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 200-209, Jan. 2002.
[6] F. P. Kelly, *Reversibility and Stochastic Networks*. John Wiley & Sons, Inc., 1979.
[7] Y.-B. Lin, W.-R. Lai, and J.-J. Chen, "Effects of cache mechanism on wireless data access," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1247-1258, Nov. 2003.
[8] Y.-B. Lin and A.-C. Pang *Wireless and Mobile All-IP Networks*. John Wiley & Sons, Inc., 2005.
[9] A. Niemi, "Session Initiation Protocol (SIP) Extension for Event State Publication," RFC 3903, IETF, Oct. 2004.
[10] A.-C. Pang, C.-H. Liu, S.P. Liu, and H.-N. Hung, "A study on SIP session timer for wireless VoIP," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, vol. 4, pp. 2306-2311, Mar. 2005.
[11] A. B. Roach, "Session Initiation Protocol (SIP)-Specific Event Notification," RFC 3265, IETF, June 2002.
[12] J. Rosenberg, "A Presence Event Package for the Session Initiation Protocol (SIP)," RFC 3856, IETF, Aug. 2004.
[13] S. M. Ross, *Stochastic Processes*. John Wiley & Sons, Inc., 1996.
[14] S.-I. Sou and Y.-B. Lin, "Modeling mobility database failure restoration using checkpoint schemes," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 313-319, Jan. 2007.
[15] T.-P. Wang and K. Chiu, "An efficient scheme for supporting personal mobility in SIP-based VoIP services," *IEICE Trans. Commun.* (Special Section Mobile Multimedia Commun.), vol. E89-B, no. 10, pp. 2706-2714, Oct. 2006.
[16] S.-R. Yang, "Dynamic power saving mechanism for 3G UMTS system," *ACM/Springer Mobile Networks Applications (MONET)*, vol. 12, no. 1, pp. 5-14, Jan. 2007.
[17] S.-R. Yang, P. Lin, and P.-T. Huang, "Modeling power saving for GAN and UMTS interworking," accepted for publication in *IEEE Trans. Wireless Commun.*

Whai-En Chen (M'99) received a Bachelor of Science degree in Electric Engineering from Tam Kang University in 1997, and received a Ph.D. in Computer Science from National Tsing Hua University in 2002. He began serving as a Research Assistant Professor in National Chiao Tung University from 2002 to 2007. Since August 2007, Dr. Chen served as an assistant professor in Institute of Computer Science and Information Engineering (CSIE)and the chief in network division of computer and IT center in National Ilan University. Since September 2008, he served as the chair of Institute of CSIE. His research interests include 3G IP Multimedia Subsystem (IMS), SIP-based VoIP services, IPv6 translation mechanisms, IEEE 802.16 WiMAX MAC.

Yi-Bing Lin (M'96-SM'96-F'03) is Dean and Chair professor of College of Computer Science, National Chiao Tung University (NCTU). He is a senior technical editor of IEEE NETWORK. He serves on the editorial boards of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and textscIEEE Transactions on Vehicular Technology. He is General or Program Chair for prestigious conferences including ACM MobiCom 2002. He is Guest Editor for several first-class journals including IEEE TRANSACTIONS ON COMPUTERS. Lin is the authors of the books *Wireless and Mobile Network Architecture* (Wiley, 2001), *Wireless and Mobile All-IP Networks* (John Wiley, 2005), and *Charging for Mobile All-IP Telecommunications* (Wiley, 2008). Lin is listed in ISI-HighlyCited.Com among top 1% most-cited Computer Science researchers. He received numerous research awards including 2005 NSC Distinguished Researcher and 2006 Academic Award of Ministry of Education. Lin is an ACM Fellow, an AAAS Fellow, and an IET Fellow.

Ren-Huang Liou received the B.S. degree in computer science from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 2007. He is currently working toward the M.S. degree at NCTU. His current research interests include design and analysis of personal communications services networks, mobile computing and performance modeling.