

Mobility Management and Signaling Traffic Analysis for Multi-Tier Wireless Mobile Networks

Yuguang Fang, *Senior Member, IEEE*, and Yi-Bing Lin

Abstract—A multi-tier wireless system integrates the high-tier wireless systems and the low-tier wireless systems into a single system to provide the advantages of both tiers. Such a system is expected to provide better service (more service availability and more cost effectiveness to the users) at the expense of the extra tier-switching management. Examples of multi-tier systems include the interconnection network of a satellite communication network and a terrestrial cellular network, the integration of low-power systems such as wireless personal area networks and cellular systems, and the integration of higher mobility of cellular systems with low-mobility wireless local area networks. In such a system, mobility management is critical, because only when the location of a mobile user is known or tracked can the service quality for the user be guaranteed. In this paper, we study a few mobility management schemes for multi-tier wireless mobile networks under more realistic assumptions. Performance evaluation in terms of registration signaling traffic is carried out, and easy-to-use analytical results are obtained. It is also clear that for service availability, the two-tier system is better than the single one-tier system; we therefore study the probability that a call is forced to terminate in the single low-tier system because the low tier becomes unavailable during the call (such a call can be continued in the two-tier system).

Index Terms—Mobility management, multi-tier, registration, service availability, wireless network.

I. INTRODUCTION

SEVERAL *personal communications service* (PCS) technologies have been developed to provide wireless service in different environments [26], [29], [31]. A multi-tier wireless system integrates the high-tier wireless systems and the low-tier wireless systems into one single system to provide the advantages of both tiers. Such a system is expected to provide better service (greater availability and lower cost for users) at the expense of the extra tier-switching management. An example is the integration of cellular networks and satellite networks. Cellular networks have enjoyed great success in recent years and penetrate most major markets providing communication services to

anyone, anywhere, at anytime, and in any form. On the other hand, the satellite communication system has larger coverage flexibility and complements the PCS systems, in particular, it can solve the hot-spots problem that often occurs in cellular systems [1], [27]. The satellite communication systems were designed to compete with the terrestrial cellular systems; hence, most satellite communications research and development focused on how to design satellite networks (Satnet) to compete with terrestrial cellular networks (Cellnet). In this competitive mode, Satnet service providers have to build their own infrastructure, gateways or base stations, and communication switches have to be built. This is not a cost-effective approach. Although the terrestrial cellular networks already had the infrastructure in place and only need some modifications for their equipment to interwork with Satnet, it will be much more cost-effective for both industries to work together. Two-tier wireless systems have already been seen in the integration of PCS networks and the AMPS systems (the macro cellular and microcellular infrastructure); the advantage of such integration is witnessed in the handling mobile users moving at different speeds and in providing better quality of service [5]. More recently, integration of various wireless systems and overlaying networks has received tremendous attention in both academia and industry; the multi-tier architectures becoming more important [3], [15], [18], [25], [26]. The basic idea for such multi-tier integration is based on the following observations.

- 1) Because of the resource (bandwidth, power etc.) limitation on satellites, direct services to a huge number of customers with different requirements on mobility and QoS may be difficult to maintain. Besides, the satellite communications services may be more expensive, thus, Cellnet can help for indirect service distribution with a lower cost.
- 2) Cellnet may not be able to cover all service areas. Satnet is the natural choice for this support. Satnet obviously could handle the overflow traffic from Cellnet, so that service holes in Cellnet coverage can be filled.
- 3) Communication aircrafts can be used to provide some communication service over a certain region of interest (such as for business); such a system can also be integrated with Cellnet for better coverage and capacity increase. The Halo system [2] belongs to this category.
- 4) Sometimes, some areas are congested in service traffic for a short time (such as during rush hours at major highways). Adding more permanent resources along the congested areas may not be a good idea, rather it is better to use dynamic traffic relief. This can be done flexibly with a high-tier system such as Satnet. A wild idea is to use the communication blimps to fly over the congested areas for such service traffic relief.

Manuscript received December 27, 2001; revised August 26, 2003 and May 12, 2004. The work of Y. Fang was supported in part by the National Science Foundation Faculty Early Career Development Award under Grant ANI-0093241 and the Office of Naval Research Young Investigator Award under Grant N000140210464. The work of Lin was sponsored in part by NSC Program for Promoting Academic Excellence of Universities under Contract 93-2752-E-009-005-PAE, National Science Council under Contract NSC Performance Evaluation for B3G Network, FarEastone, IIS/Academia Sinica, CCL/ITRI and NCTU Joint Research Center, Chair professorship of Providence University and National Telecommunication Development Program (NTP).

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: fang@ece.ufl.edu).

Y.-B. Lin is with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsingchu, Taiwan, R.O.C. (e-mail: liny@csie.nctu.edu.tw).

Digital Object Identifier 10.1109/TVT.2005.851326

- 5) To fulfill the vision of real globalization, various communications systems will most likely be integrated, microcellular systems (such as PCS systems) can form the first tier; macrocellular DAMPS systems, communication blimps, or Halo systems can form the second tier; the mobile satellite systems (LEOS) can form the third tier; and the geostationary satellites can form the fourth tier. Thus, such multi-tier systems can provide better service to mobile users and realize the goal PCS systems originally envisioned.
- 5) The convergence of high-speed wireless local area networks and anywhere and anytime wireless cellular systems [3], [18] poses another challenging application of the multi-tier architecture.

In this paper, for ease of presentation, we concentrate on the two-tier system, which is to integrate the high-tier and low-tier systems into a single system seamlessly to provide better service to mobile customers. The results can be easily generalized for the multi-tier systems.

To make this integration of high-tier and low-tier systems work, both systems have to know where a mobile user is to effectively deliver services to the user. This problem is solved in the single-tier wireless cellular networks such as PCS networks by maintaining location databases (home location register and visiting location register) and by carrying out mobility management schemes. For the integrated multi-tier systems, not only the signaling message formats are different, but also the mobility management may be different. For example, in wireless cellular systems, the location information is stored in the location registers and updated through registration, whereas in wireless LANs, the location information is via association/disassociation with access points or distribution systems (DS) [17], [18]. Even for a single-tier system, signaling traffic may be a major concern and multi-tier-like architecture may be used to reduce the registration traffic [32]. Thus, when we have integrated/overlayed multiple wireless systems, we have to design the efficient mobility management scheme to have up-to-date location information; in particular, the multiple wireless systems must have the same correct view about users' location information. One simple solution is to send location updates to all systems whenever necessary; however, this may lead to too much registration-signaling traffic, which is definitely undesirable. Thus, more efficient mobility management schemes for the multi-tier integrated systems must be considered.

In [23], we have proposed a few mobility management schemes, *single-tier registration* and *multi-tier registration*, and carry out performance comparison under restrictive exponential assumption in specific scenarios. Moreover, a few open problems are posed without solutions. In particular, we suggested that the problem of relaxing the exponential assumptions for two scenarios in its analysis be solved by simulation. However, sufficient evidence from theory and experimental data from various (wired or wireless) networks shows that many time variables are no longer exponentially distributed. To extend the analysis in [23] to a more general scenario is not a trivial task. Two issues must be resolved: what distribution model should be used and how can we obtain tractable analytical results? In this paper, we

generalize the work in [23], propose to using a more general distribution model to characterize all time variables and obtaining easy-to-compute analytical results for the average number of registration-signaling messages used in each mobility management scheme. As a by-product, we have successfully solved the aforementioned open problem analytically instead of using simulation. More importantly, we develop a unifying analytical approach to the analysis for the mobility management schemes under more realistic assumptions.

This paper is organized as follows. In Section II, we present the mobility management schemes (the descriptions of two registration schemes). In Section III, we develop a new analytical modeling approach to analyzing the mobility management schemes. Service availability in multi-tier systems is investigated in Section IV. We conclude in Section V.

II. REGISTRATION SCHEMES

In this section, we present the mobility management schemes for multi-tier wireless systems. As we mentioned before, we focus on the two-tier systems, in particular, the integration of high-tier and low-tier systems such as the possible integration of Satnet and Cellnet. We assume that the reader is familiar with the concepts of mobility management (such as home location register and visitor location register) used in PCS networks. For more details, the reader is referred to [7] [16], [26]. In a multi-tier system, the individual tiers have their respective *visitor location registers* (VLRs) for their own tier coverage. These tiers share a single multi-tier *home location register* (HLR) or MHLR. Each tier provides services to the same area with its own coverage requirements; all tiers may overlay each other. Mobile users can be served by a tier at any given time and can be switched to another tier for their service. For example, in the integrated Satnet–Cellnet system, a mobile user can be served by the Cellnet. When the user roams out of the coverage area of Cellnet, the Satnet can provide the user with a continuous service. In such a system, a user's location information should be updated frequently. Registration (or location update) is the process for a mobile user to update the current location information. In a multi-tier system, there are two methods by which users can update (register) their location information. The first method is to require mobile users to register only at one tier at any given time; this is called the *single registration* (SR) method. In the second method, which is called the *multiple registration* (MR) method, mobile users are allowed to register at multiple tiers concurrently at any time. These two registration protocols for the multi-tier PCS networks were proposed in [24], [28], and the performance of these protocols has been analyzed in [23]. Because multiple registration induces too much signaling traffic, it is not practical; we will not study it further. In this paper, we focus on the SR methods. There are two alternatives in SR methods. (For the ease of our discussion, we consider a two-tier system. Our results can be easily generalized for multi-tier systems with more than two tiers.)

In the SR methods, the *mobile phones* or *mobile stations* (MSs) in the two-tier system register themselves in the following way:

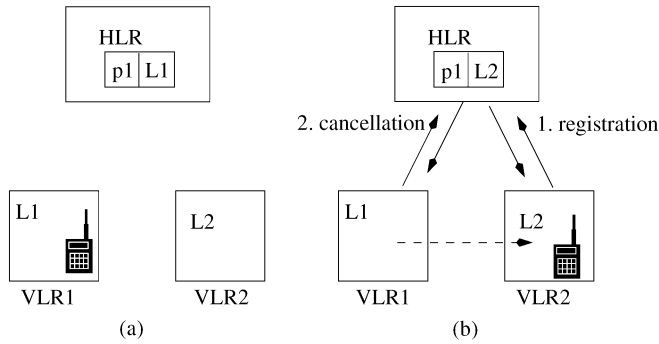


Fig. 1. The L system registration procedure. (a) Before the movement. (b) After the movement.

Case 1. When the MS is turned on, it first checks whether the low-tier system is available. If it is available, the MS sends a registration message to the MHLR via the low-tier VLR. The MS will then monitor and receive services in the low-tier system. If the low-tier system is not available, then the high-tier system is checked. If the high-tier system is available, then the strategy described in *case 2* is exercised. If neither low-tier nor high-tier is available, then the MS keeps monitoring both systems and repeats the process.

Case 2. If the high-tier system is available (and the low-tier system is not available), then the MS sends a registration message to the MHLR via the high-tier VLR and receives service through the high-tier system. At the same time, the MS keeps monitoring the low-tier system. Whenever the low-tier system is available, the MS will immediately switch to the low-tier system by exercising the procedure in case 1. If the high-tier system is not available, the MS keeps monitoring the two tiers by repeating case 1.

In [23], a performance model was proposed to compare the registration traffic generated by the single low-tier system (L system) and the two-tier systems. In the L system, users are only registered in the single (e.g., low-tier) system. If the single system is not available (either out of range or because of congestion), the user will not receive any service, and no registration messages are sent when the user crosses the single-tier boundary. While in the two-tier system, if the low-tier system is available, the MS will receive services from the low-tier system. Otherwise, the MS will utilize the high-tier system. The user continues to receive service through the high-tier system, until the low-tier is available.

In the L system, we assume that the IS-41 registration protocol [6] is used. Consider Fig. 1. Each MS p resides at a registration area L_1 , and the HLR record for p indicates that the current registration area (RA) is L_1 [see Fig. 1(a)]. When p moves to L_2 [see Fig. 1(b)], it sends a registration message to the HLR via the new VLR. The HLR updates the location information and acknowledges this action by sending a cancellation message to the old VLR. The HLR sends a cancellation message to the old VLR to delete the temporary record for p at the old VLR. An acknowledgment message is sent from the old

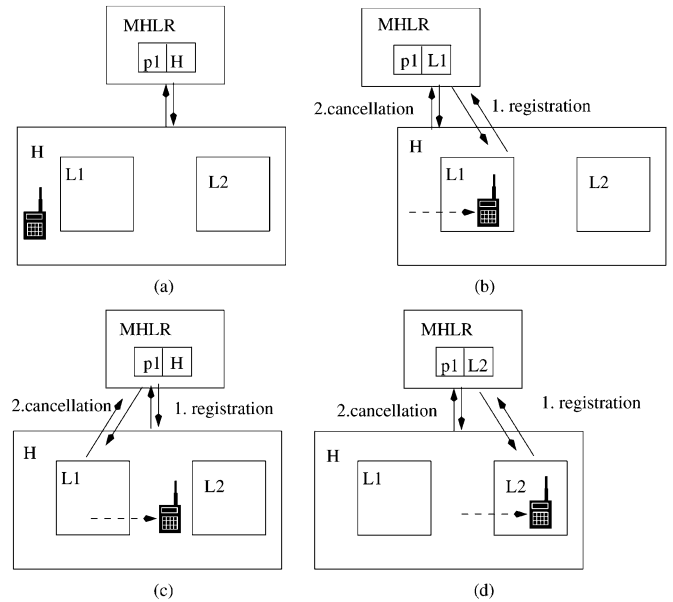


Fig. 2. The SR1 protocol. (a) Step 1. (b) Step 2. (c) Step 3. (d) Step 4.

VLR to the HLR. Hence, four signaling messages are needed. In this procedure, the authentication messages are omitted to simplify our discussion. Our results can be easily generalized when authentication is considered.

In single registration methods, there are two alternatives.

SR1. In this alternative, there is no distinction between the low-tier and high-tier RAs, and the switching between tiers is exactly the same as in the IS-41 registration process. Consider Fig. 2. The MS p is registered in one tier, say, the high-tier RA H , and the location is recorded in MHLR with the format (p, H) [see Fig. 2(a)]. When the MS moves to the low-tier RA L_1 , the VLR of L_1 sends a message to the MHLR to update the location record from (p, H) to (p, L_1) . The MHLR then sends a message to L_1 for acknowledgment and another message to H for cancellation. After H gets the message, it will send another acknowledgment to MHLR [see Fig. 2(b)]. A similar registration procedure is carried out when p moves from the low tier to the high tier [see Fig. 2(c)].

SR2. In this alternative every MHLR record consists of two location fields (one for the high tier and one for the low tier) and one bit to indicate which tier is available. The record format can be $(p, H; H_1, L_1)$ for the case when the MS p is in the high tier with RA H_1 and with the previous low-tier RA L_1 as shown in Fig. 3(a). When the MS p moves to the low-tier RA L_1 , the L_1 sends a message to MHLR to update the record and change the tier bit to L , i.e., the record is changed to $(p, L; H_1, L_1)$ as shown in Fig. 3(b). After this the MHLR sends an acknowledgment to L_1 . No cancellation/acknowledgment is necessary because the low-tier RA is the same as the previously visited one. In this case, only two signaling messages are needed. However, if the MS moves from the

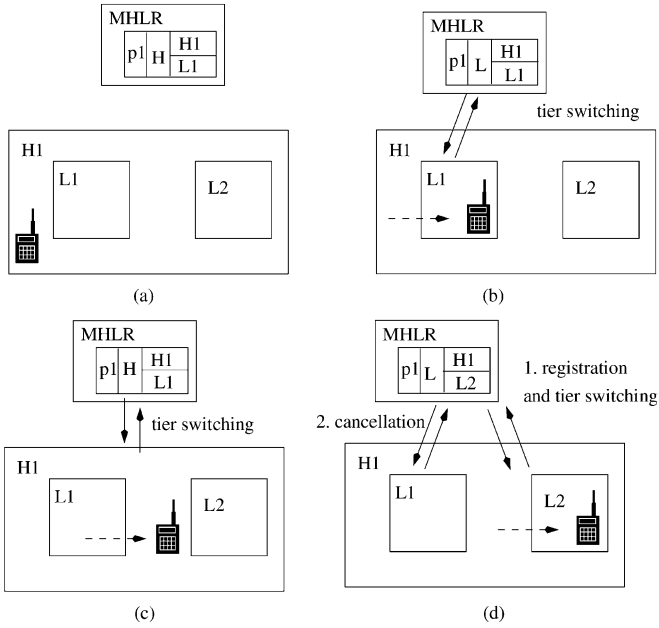


Fig. 3. The SR2 protocol. (a) Step 1. (b) Step 2. (c) Step 3. (d) Step 4.

high tier to a different RA, say, L_2 [see Figs. 3(c) and (d)], then the record in MHLR will be updated from $(p, H; H_1, L_1)$ to $(p, L; H_1, L_2)$, two signaling messages are exchanged between MHLR and L_2 , and two more messages are needed between MHLR and L_1 for cancellation/acknowledgment.

The call delivery procedure for SR (both SR1 and SR2) is exactly the same as the procedure for the single-tier system. That is, the network delivers the call based on the tier and RA indicated in MHLR. From the discussion in this section, we have the following conclusions.

- 1) To switch between tiers, extra signaling overhead is required in the multi-tier system (compared with the L system).
- 2) Compared with SR2, more messages are sent in SR1 (see the scenario in Figs. 2 and 3).
- 3) The advantage of SR1 over SR2 is that the standard IS-41 HLR can be used as the MHLR for SR1, whereas MHLR with extra record fields are required in SR2.

We observe that SR1 and SR2 both induce more registration traffic than the registration scheme (IS-41) in the L system; this is the tradeoff for more service coverage and better quality of service (in terms of the call-dropping rate). Thus, the problem we face now is how much more signaling traffic the new schemes will incur to gain the advantage of multi-tier architecture. A partial answer to this problem in [23] is given based on some simplified model. In this paper, we will provide detailed analysis for the registration overhead and the service availability in Sections III and IV.

III. ANALYTICAL MODELING FOR THE REGISTRATION SCHEMES

In this section, we calculate the signaling traffic for the three registration schemes L, SR1, and SR2. The performance anal-

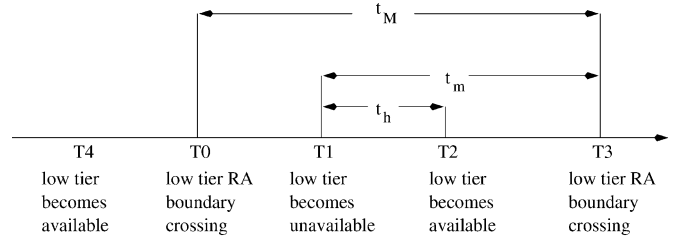


Fig. 4. The time diagram.

ysis has been carried out in [23] under the conditions in which some of the time variables are exponentially distributed. It was also pointed out that the possible extension of the work was to relax the exponential assumptions in that paper by the simulation study. In this paper, we attempt to generalize the results analytically without using the exponential assumptions. We obtain a general result, which also solves the problems for various situations in [23].

Fig. 4 shows the time diagram for the MS. Let T_0 denote the last low-tier RA boundary crossing before the low-tier is unavailable, T_1 denotes the instant that the current low-tier becomes unavailable, T_2 and T_4 are the time instants that the low-tier becomes available, and T_3 denotes the first low-tier RA boundary crossing when the low tier becomes available again. Thus, without a high-tier system, a mobile will not be served by the lower tier system because of the lack of coverage in the time between T_1 and T_2 when the mobile roams in the RA in the low-tier system. As indicated in the figure, t_h is the time that the MS spends in the high-tier system while the low-tier system is not available. Period t_M denotes the interboundary crossing time, i.e., the low-tier RA residence time, and t_m is the residual life of t_M .

Based on these notations, we could study the signaling traffic for each L, SR1, and SR2 system. As we notice, the interval $[T_4, T_1]$ is the period that the low-tier system is available, and the interval $[T_1, T_2]$ is the period that the low-tier system is not available and that the high-tier system is providing possible service. Thus, the interval $[T_4, T_2]$ will form the fundamental regenerative period (in an alternating renewal process) [30]. This fact allows us to concentrate on the signaling traffic analysis in one fundamental regenerative period. Our purpose is to compare the signaling traffic for SR1 and SR2 systems with the L system, since the signaling traffic incurred during $[T_4, T_1]$ for the three schemes are the same, we can focus on the signaling traffic in the interval in $[T_1, T_2]$. The following derivations can be found [23]. For the reader's benefit, we reiterate the derivations as follows.

The L System. If a user does not cross any low-tier RA boundary during $[T_1, T_2]$ (i.e., $T_2 < T_3$), the MS does not take any action at time T_2 . On the other hand, if the user crosses the low-tier RA boundary during $[T_1, T_2]$ (i.e., $T_2 > T_3$), then the MS detects the RA visited at T_2 is different from the RA visited at T_1 . Based on IS-41 protocol, four messages are required in the registration/de registration operations at T_2 . The average

number n_l of the signaling messages incurred during $[T_1, T_2]$ is

$$\begin{aligned} n_l &= 0 \times \Pr(T_3 > T_2) \\ &\quad + 4[1 - \Pr(T_3 > T_2)] \\ &= 4 - 4\Pr(T_3 > T_2). \end{aligned} \quad (1)$$

The SR1 System. From the operation of SR1 system, eight signaling messages are required for tier switching whether the MS crosses the low-tier RA boundaries during $[T_1, T_2]$ or not, four at T_1 and another four at T_2 . Thus, the average number n_{sr1} of signaling messages incurred during $[T_1, T_2]$ is

$$n_{sr1} = 8. \quad (2)$$

The SR2 System. If a user does not cross any low-tier RA boundary during $[T_1, T_2]$, then four signaling messages are incurred during $[T_1, T_2]$. On the other hand, if the user crosses the low-tier RA boundary during $[T_1, T_2]$ (i.e., $T_2 > T_3$), then six signaling messages will be exchanged. Therefore, the average number n_{sr2} of signaling messages incurred during $[T_1, T_2]$ is

$$\begin{aligned} n_{sr2} &= 4\Pr(T_3 > T_2) \\ &\quad + 6[1 - \Pr(T_3 > T_2)] \\ &= 6 - 2\Pr(T_3 > T_2). \end{aligned} \quad (3)$$

We observe that the performance of these three systems is boiled down to finding the probability $\Pr(T_3 > T_2)$. This can be calculated from the following result.

Lemma: Let X and Y be nonnegative random variables. Let $f_X(t)$ and $f_Y(t)$ denote probability density functions with Laplace transforms $f_X^*(s)$ and $f_Y^*(s)$, respectively. Then, we have

$$\begin{aligned} \Pr(X > Y) &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_Y^*(s)}{s} f_X^*(-s) ds \\ &= - \sum_{p \in \sigma_X} \text{Res}_{s=p} \frac{f_Y^*(s)}{s} f_X^*(-s) \\ &= 1 - \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_X^*(s)}{s} f_Y^*(-s) ds \\ &= 1 + \sum_{p \in \sigma_Y} \text{Res}_{s=p} \frac{f_X^*(s)}{s} f_Y^*(-s) \end{aligned}$$

where $\sigma > 0$ is a sufficiently small positive number, σ_X and σ_Y are the sets of poles of $f_X^*(-s)$ ($f_Y^*(-s)$) in the right half of the complex plane, respectively, and $\text{Res}_{s=p}$ denotes the residue at the pole $s = p$.

Proof:

$$\begin{aligned} \Pr(X > Y) &= \int_0^\infty f_Y(y) \int_y^\infty f_X(x) dx dy \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \int_0^\infty f_Y^*(s) e^{sy} \left[\int_y^\infty f_X(x) dx \right] dy ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} f_Y^*(s) \int_0^\infty \left[\int_y^\infty f_X(x) dx \right] e^{sy} dy ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} f_Y^*(s) \frac{-1 + f_X^*(-s)}{s} ds \\ &= \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} \frac{f_Y^*(s)}{s} f_X^*(-s) ds \\ &= - \sum_{p \in \sigma_X} \text{Res}_{s=p} \frac{f_Y^*(s)}{s} f_X^*(-s) \end{aligned}$$

where we have used the Residue Theorem [22]. Similarly, we can obtain the other equality. This completes the proof.

Next, we apply this Lemma to find the probability $\Pr(T_3 > T_2)$. Assume that the t_h , t_M , and t_m have the probability density functions $f_h(t)$, $f_M(t)$, and $f_m(t)$ with the Laplace transforms $f_h^*(s)$, $f_M^*(s)$, and $f_m^*(s)$, respectively. Let $1/\lambda_M$, $1/\lambda_m$, and $1/\lambda_h$ denote the expected values of t_M , t_m , and t_h , respectively. Obviously, we have $t_M = T_3 - T_0$, $t_m = T_3 - T_1$, and $t_h = T_2 - T_1$. Thus, the event $(T_3 > T_2)$ is the same as the event $(t_m > t_h)$. Moreover, because t_m is the residual life of the RA residence time t_M , from the Residual Life Theorem [20], we have

$$f_m^*(s) = \frac{\lambda_M(1 - f_M^*(s))}{s}.$$

Applying the lemma with some mathematical manipulations, we obtain

Theorem 1: If the Laplace transform of either $f_M(t)$ or $f_h(t)$ has isolated poles (which is the case when it is a rational function), the probability $\Pr(T_3 > T_2)$ can be easily computed by the following:

$$\begin{aligned} \Pr(T_3 > T_2) &= \Pr(t_m > t_h) \\ &= \lambda_M \sum_{p \in \sigma_M} \text{Res}_{s=p} \frac{f_h^*(s)}{s^2} (1 - f_M^*(-s)) \\ &= -\lambda_M \sum_{p \in \sigma_M} \text{Res}_{s=p} \frac{f_h^*(s)}{s^2} f_M^*(-s) \\ &= 1 + \lambda_M \sum_{p \in \sigma_h} \text{Res}_{s=p} \frac{1 - f_M^*(s)}{s^2} f_h^*(-s) \end{aligned} \quad (4)$$

where σ_M is the set of poles of $f_M^*(-s)$ in the right half of the complex plane. If t_M is exponentially distributed, then we have

$$\Pr(T_3 > T_2) = f_h^*(\lambda_M).$$

If t_h is exponentially distributed, then we have

$$\Pr(T_3 > T_2) = 1 - \frac{\lambda_M}{\lambda_h} (1 - f_M^*(\lambda_h)).$$

Proof: We only need to prove the last two statements or just the second statement. If t_M is exponentially distributed with parameter λ_M , then we have

$$f_M^*(s) = \frac{\lambda_M}{s + \lambda_M}$$

taking it into Theorem 1, we obtain

$$\Pr(T_3 > T_2) = -\text{Res}_{s=\lambda_M} \frac{f_h^*(s)}{s} \cdot \frac{\lambda_M}{-s + \lambda_M} = f_h^*(\lambda_M).$$

This completes the proof.

Notice that in this result, the computation of the probability $\Pr(T_3 > T_2)$ can be easily carried out by using fractional expansion as long as all Laplace transforms are rational functions. It has been known that the set of probability distributions with rational Laplace transform is dense in the set of all probability distributions for nonnegative random variables, which implies that any nonnegative random variable can be sufficiently characterized by a probability distribution model with rational Laplace transform [11], [19], [20]. Thus, Theorem 1 is general enough and can be applied to all scenarios. From our previous discussion, as long as we know the calculation of this probability, we can evaluate the registration-signaling traffic.

However, the set of probability distribution functions with rational Laplace transform may still be too large. In what follows, we present a few classes of probability distributions of great interest and the analytical results we obtained from Theorem 1. The Erlang distribution with parameters (m, η) has been used in many applications; its probability density function and Laplace transform are given as

$$f(t) = \frac{\eta^m t^{m-1}}{(m-1)!} e^{-(m\eta)t}, \quad f^*(s) = \left(\frac{m\eta}{s + m\eta} \right)^m \quad (5)$$

where $1/\eta$ is the expectation. Here, we only concentrate on the modeling of RA residence time t_M , whose probability density function is $f_M(t)$. Similar results with respect to t_h can be easily obtained. If t_M is Erlang distributed with parameters (m, λ_M) , we have

$$f_M^*(s) = \left(\frac{m\lambda_M}{s + m\lambda_M} \right)^m.$$

Taking this into Theorem 1, we obtain [we will use $g^{(i)}(s)$ to denote the i th derivative of the function $g(s)$]:

$$\begin{aligned} \Pr(T_3 > T_2) &= -\lambda_M \sum_{s=\sigma_M} \text{Res}_{s=\lambda_M} \frac{f_h^*(s)}{s^2} \left(\frac{m\lambda_M}{-s + m\lambda_M} \right)^m \\ &= (-1)^{m+1} \lambda_M (m\lambda_M)^m \text{Res}_{s=m\lambda_M} \frac{f_h^*(s)}{s^2} \cdot \frac{1}{(s - m\lambda_M)^m} \\ &= \frac{(-1)^{m+1} \lambda_M (m\lambda_M)^m}{(m-1)!} \frac{d^{m-1}}{ds^{m-1}} (s^{-2} f_h^*(s)) \Big|_{s=m\lambda_M} \\ &= \frac{(-1)^{m+1} \lambda_M (m\lambda_M)^m}{(m-1)!} \end{aligned}$$

$$\begin{aligned} &\times \sum_{k=0}^{m-1} \binom{m-1}{k} f_h^{*(m-1-k)}(m\lambda_M) (s^{-2})^{(k)} \Big|_{s=m\lambda_M} \\ &= \frac{(-1)^{m+1} \lambda_M (m\lambda_M)^m}{(m-1)!} \\ &\times \sum_{k=0}^{m-1} \binom{m-1}{k} f_h^{*(m-1-k)}(m\lambda_M) \cdot \frac{(-1)^k (k+1)!}{(m\lambda_M)^{k+2}} \\ &= \frac{1}{m} \lambda_M \sum_{k=0}^{m-1} \frac{(k+1)(-m\lambda_M)^{m-1-k}}{(m-1-k)!} f_h^{*(m-1-k)}(m\lambda_M) \\ &= \frac{1}{m} \sum_{i=0}^{m-1} \frac{(m-i)(-m\lambda_M)^i}{i!} f_h^{*(i)}(m\lambda_M). \quad (6) \end{aligned}$$

When $m = 1$, we obtain the result obtained in Theorem 1 (which was given also in [23]). When $m = 2$, we obtain

$$\Pr(T_3 > T_2) = f_h^*(2\lambda_M) - \lambda_M f_h^{*(1)}(2\lambda_M). \quad (7)$$

The hyper-Erlang distribution [11], [19] has been shown to have very general approximation capability to the probability distribution of any nonnegative random variable. It has the following density function and Laplace transform:

$$\begin{aligned} f_{he}(t) &= \sum_{i=1}^M \alpha_i \frac{(m_i \eta_i)^{m_i} t^{m_i-1}}{(m_i-1)!} e^{-m_i \eta_i t} (t \geq 0), \\ f_{he}^*(s) &= \sum_{i=1}^M \alpha_i \left(\frac{m_i \eta_i}{s + m_i \eta_i} \right)^{m_i} \quad (8) \end{aligned}$$

where

$$\alpha_i \geq 0, \quad \sum_{i=1}^M \alpha_i = 1.$$

If t_M is hyper-Erlang distributed as in (8), where $\lambda_M = [\sum_{i=1}^M \alpha_i / \eta_i]^{-1}$, then we have the following result:

$$\begin{aligned} \Pr(T_3 > T_2) &= \sum_{k=1}^M \frac{\alpha_k}{m_k} \\ &\times \sum_{i=0}^{m_k-1} \frac{(m_k-i)(-m_k \eta_k)^i}{i!} f_h^{*(i)}(m_k \eta_k). \quad (9) \end{aligned}$$

Thus, if we find a hyper-Erlang distribution fit from the field data for the low-tier RA residence time, we can easily apply our theoretical results to evaluate the signaling traffic for the proposed registration methods in the paper. In fact, as long as all time variables are distributed whose probability density functions have rational Laplace transforms, our result will apply. Thus, if t_M or t_h has either Coxian or more generally phase-type distribution [20], [21], we can use Theorem 1 to find the probability above and carry out the signaling traffic analysis.

Finally, equipped with the calculation of $\Pr(T_3 > T_2)$ under realistic assumption (without the exponential assumption

random variables t_M or t_h or t_m), we are in the position to calculate the registration-signaling traffic, in particular, the extra traffic incurred from SR1 and SR2 in comparison with that incurred from registration in the L system. Notice that $T_1 - T_4$ is the period for the low-tier service to be available. Let $1/\lambda_l = E[T_1 - T_4]$ and λ be the low-tier RA crossing rate when the low-tier is available. Because every RA boundary crossing during the available period of low-tier system service will result in four message exchanges (including registration, deregistration, and acknowledgments), from Little's law the average number of signaling messages incurred during $[T_4, T_1]$ will be

$$N = 4 \times (\lambda E(T_1 - T_4)) = \frac{4\lambda}{\lambda_l}. \quad (10)$$

Let the percentage of the extra traffic created by SR1 and SR2 over the L system be δ_1 and δ_2 , respectively. Then, we have (11), shown at the bottom of the page. Similarly, we obtain

$$\delta_2 = \frac{1 + \Pr(T_3 > T_2)}{2[1 - \Pr(T_3 > T_2) + \lambda/\lambda_l]} = \frac{\delta_1}{2}. \quad (12)$$

Therefore, we have

Theorem 2: Let the percentage of the extra traffic incurred in SR1 and SR2 over the L system be δ_1 and δ_2 , respectively. Then,

$$\delta_1 = 2\delta_2 = \frac{1 + \Pr(T_3 > T_2)}{1 - \Pr(T_3 > T_2) + \lambda/\lambda_l}$$

where $\Pr(T_3 > T_2)$ can be calculated from Theorem 1.

We notice that the signaling traffic for SR1 is always twice the signaling load induced by the SR2. Because of the simple relationship above, we only need to concentrate our comparison study for either δ_1 or δ_2 . We will study the effect of the parameters on the signaling traffic on δ_1 in this section. Because we have not made any assumption on some of the random variables except a weak assumption that the Laplace transforms of their probability density function are rational functions, we have successfully solved the open problem posed in [23].

In the rest of this section, we present a few examples showing how to use our results to calculate the extra signaling traffic due to single-tier registrations. We adopt the same scenario from [23]:

- 1) $\lambda_M = \alpha\lambda$, where $\alpha > 1$ is the scaling factor that reflects the mobility (the user speed). The larger the α , the larger the λ_M , the shorter the mobile stays in the low-tier RA; hence, the faster the mobile moves.
- 2) $\beta = (E[T_1 - T_4]/E[T_2 - T_4]) = \lambda_h/(\lambda_l + \lambda_h)$ is the proportion of the time when the low tier is available. Thus,

we have

$$\lambda_l = \frac{1 - \beta}{\beta} \lambda_h$$

In many applications [23], some time variables are characterized by the Gamma distribution [with parameters (γ, η)] which is given by

$$f(t) = \frac{\eta^\gamma t^{\gamma-1}}{\Gamma(\gamma)} e^{-\eta t}, \quad f^*(s) = \left(\frac{\gamma\eta}{s + \gamma\eta} \right)^\gamma \quad (13)$$

where $\Gamma(\gamma)$ is the Gamma function. We observe that if γ is an integer, then we obtain the Erlang distribution with parameters (γ, η) and with $\Gamma(\gamma) = (\gamma - 1)!$. When $\gamma = 1$, we obtain the exponential distribution with parameter η . Thus, Gamma distribution covers many interesting cases. For illustration purposes, we assume now that t_h is Gamma distributed with parameters (γ, λ_h) . If t_M is exponentially distributed with parameter $\lambda_M = \alpha\lambda$, then we have

$$\Pr(T_3 > T_2) = \left(\frac{\gamma\lambda_h}{\alpha\lambda + \gamma\lambda_h} \right)^\gamma = \left(\frac{\gamma(\lambda_h/\lambda)}{\alpha + \gamma(\lambda_h/\lambda)} \right)^\gamma. \quad (14)$$

When T_M is Erlang distributed with parameters $(2, \alpha\lambda)$, we have

$$\begin{aligned} P(T_3 > T_2) &= f_h^*(2\lambda_M) - \lambda_M f_h^{*(1)}(2\lambda_M) \\ &= \left(\frac{\gamma\lambda_h}{2\lambda_M + \gamma\lambda_h} \right)^\gamma + \lambda_M \frac{\gamma}{2\lambda_M + \gamma\lambda_h} \\ &\quad \cdot \left(\frac{\gamma\lambda_h}{2\lambda_M + \gamma\lambda_h} \right)^\gamma \\ &= \left[1 + \frac{\lambda_M \gamma}{2\lambda_M + \gamma\lambda_h} \right] \left(\frac{\gamma\lambda_h}{2\lambda_M + \gamma\lambda_h} \right)^\gamma \\ &= \left[1 + \frac{\alpha\gamma}{2\alpha + \gamma(\lambda_h/\lambda)} \right] \left(\frac{\gamma(\lambda_h/\lambda)}{2\alpha + \gamma(\lambda_h/\lambda)} \right)^\gamma \end{aligned} \quad (15)$$

Using λ to scale all parameters, let $\rho = \lambda_l/\lambda$, then $\lambda_h = \beta\rho/(1 - \beta)$. Thus, from (14) and (15), when T_M is exponentially distributed, we have

$$\delta_1 = \frac{1 + \left(\frac{\gamma\beta\rho}{\alpha(1-\beta) + \gamma\beta\rho} \right)^\gamma}{1 - \left(\frac{\gamma\beta\rho}{\alpha(1-\beta) + \gamma\beta\rho} \right)^\gamma + \frac{1}{\rho}} \quad (16)$$

$$\begin{aligned} \delta_1 &= \frac{\text{signaling traffic in } [T_4, T_2] \text{ for SR1} - \text{signaling traffic in } [T_4, T_2] \text{ for } L}{\text{signaling traffic in } [T_4, T_2] \text{ for } L} \\ &= \frac{(n_{sr1} + N) - (n_l + N)}{n_l + N} = \frac{n_{sr1} - n_l}{n_l + N} \\ &= \frac{4 + 4\Pr(T_3 > T_2)}{4 - 4\Pr(T_3 > T_2) + 4\lambda/\lambda_l} = \frac{1 + \Pr(T_3 > T_2)}{1 - \Pr(T_3 > T_2) + \lambda/\lambda_l}. \end{aligned} \quad (11)$$

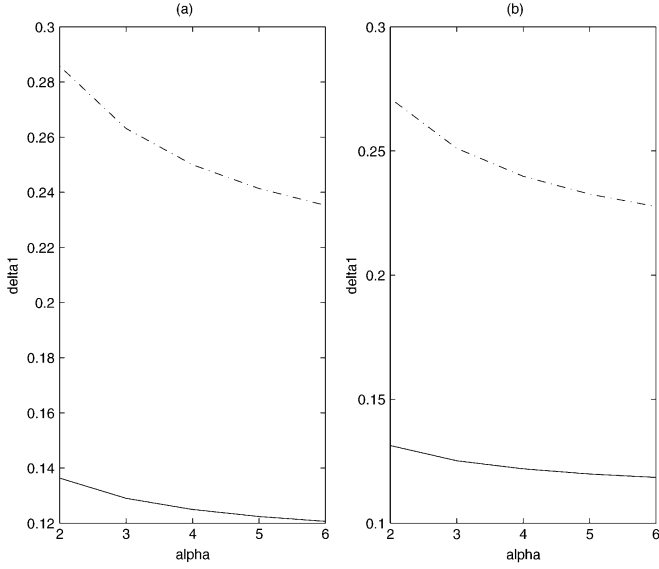


Fig. 5. The effects of ρ and α on the two-tier registration traffic: $\beta = 0.8$; solid line, when $\rho = 1/8$; dashed line, when $\rho = 1/4$. (a) $m = 1$. (b) $m = 2$.

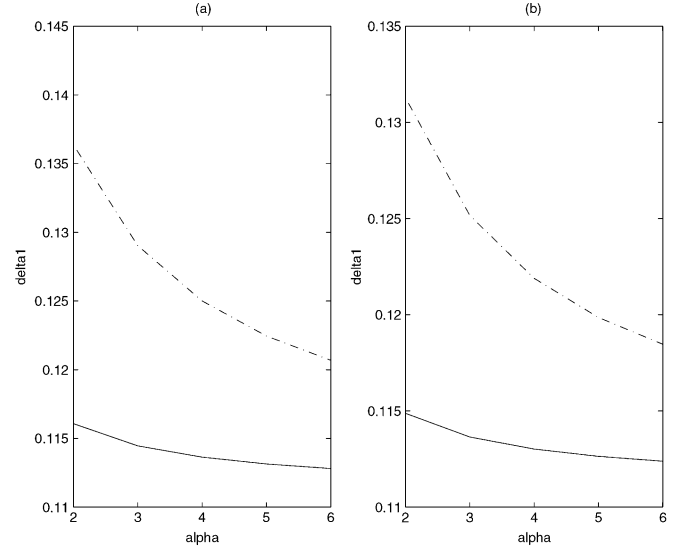


Fig. 7. The effects of β on the two-tier registration traffic: $\rho = 1/8, \gamma = 1$; solid line, when $\beta = 0.4$; dashed line, when $\beta = 0.8$. (a) $m = 1$. (b) $m = 2$.

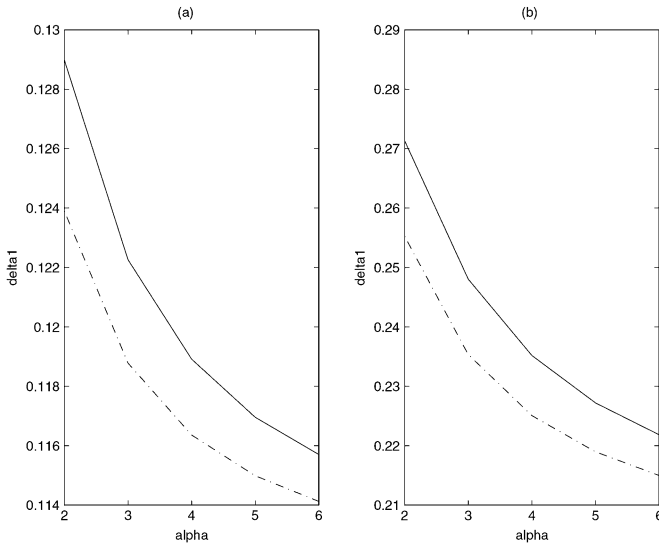


Fig. 6. The effects of α and m on the two-tier registration traffic: solid line, when $m = 1$; dashed line, when $m = 2$. (a) $\rho = 1/8$. (b) $\rho = 1/4$.

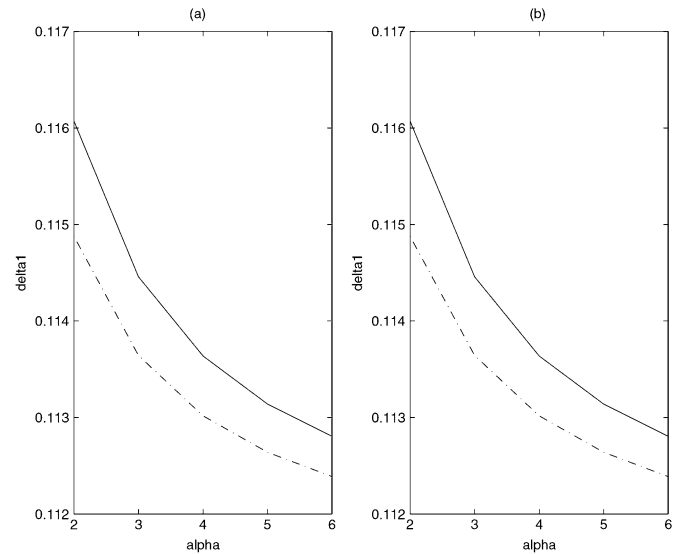


Fig. 8. The effects of m on the two-tier registration traffic for different β : $\gamma = 1.5$; solid line, when $m = 1$; dashed line, when $m = 2$. (a) $\beta = 0.4$. (b) $\beta = 0.8$.

When t_M is Erlang distributed with parameters $(2, \alpha\lambda)$, we have

$$\delta_1 = \frac{1 + \left[1 + \frac{\alpha\gamma(1-\beta)}{2\alpha(1-\beta) + \gamma\beta\rho} \right] \left(\frac{\gamma\beta\rho}{2\alpha(1-\beta) + \gamma\beta\rho} \right)^\gamma}{1 - \left[1 + \frac{\alpha\gamma(1-\beta)}{2\alpha(1-\beta) + \gamma\beta\rho} \right] \left(\frac{\gamma\beta\rho}{2\alpha(1-\beta) + \gamma\beta\rho} \right)^\gamma + \frac{1}{\rho}} \quad (17)$$

Figs. 5–8 show the effects of α on the signaling traffic δ_1 under various β, γ , and ρ values. These figures generalize the results in [23] that δ_1 decreases as α increases. A large α implies a small probability $\Pr[T_3 > T_2]$ (if the user moves fast, it is likely that the user will cross RA boundaries during $[T_1, T_2]$). Because δ_1 decreases as $\Pr[T_3 > T_2]$ decreases, we have the conclusion that δ_1 is a decreasing function of α .

Note that a large α factor is expected in a two-tier system where the low-tier system is allowed to serve high-speed users.

Thus, we observe that a two-tier system generates less registration traffic if the low tier is allowed to serve high-speed users.

Fig. 5 shows the effects of ρ on the two-tier registration traffic δ_1 , where $\beta = 0.8$ and $\gamma = 1$. It is shown that the registration traffic is sensitive to ρ , the larger the value ρ , the more traffic induced, which is consistent with our intuition that more low-tier crossings lead to more signaling traffic.

Fig. 6 shows how registration traffic is affected by the variance of the low-tier RA residence time t_M ($\gamma = 1.5$). We recall that the variance for t_M is $1/(m\lambda_M^2)$. We conclude that the variance of the low-tier RA residence time really affects the signaling traffic: the smaller the variance (i.e., the larger the value m), the lower the registration traffic.

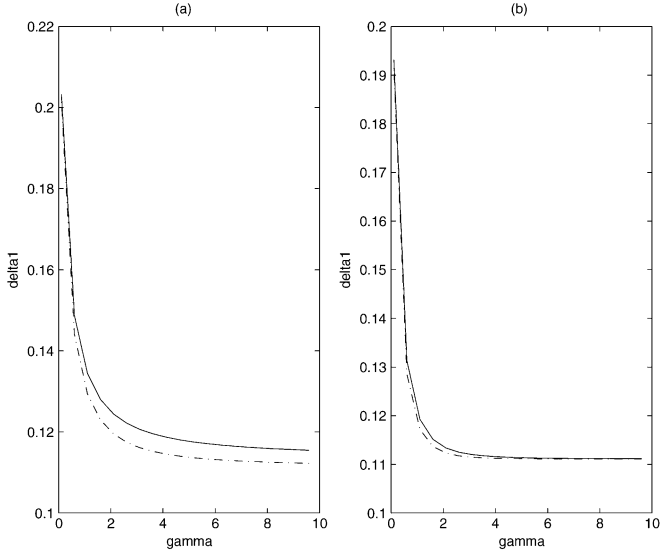


Fig. 9. Effects of γ on the two-tier registration traffic: solid line, when $m = 1$; dashed line, when $m = 2$. (a) $\alpha = 2$. (b) $\alpha = 6$.

Fig. 7 shows the effects of α, β , and m on the registration traffic, where $\rho = 1/8$. It shows that the registration traffic is very sensitive to β . Fig. 8 shows the effect of the variance of the low-tier RA residence time.

Fig. 9 shows the effect of the variance of t_h (i.e., γ). We observe that registration traffic is significantly affected when the variance is large (i.e., when γ is small). When α is larger (i.e., the speed factor is larger), the variance of the low-tier RA residence time (i.e., the value m) will have no significant effect on the signaling traffic.

From the demonstrative examples, it is clear that our general analytic formula provides guidelines to choose appropriate registration schemes based on the statistics collected from the practical systems.

IV. SERVICE AVAILABILITY OF LOW-TIER SYSTEMS

Because of the limited resources in the multi-tier systems, it may be possible that requested services cannot be provided. It is very important for system designers to know the service availability of various tiers so that resources can be provisioned to meet the network service requirement. For illustration purposes, we still concentrate on the two-tier systems.

In a two-tier system, the service availability is better than in the single-tier systems. For the single high-tier system, β portion of the phone connections are delivered at a more expensive cost than the two-tier system where calls are delivered through the relatively low-cost low-tier system. For the single low-tier system, $(1 - \beta)$ of the phone calls are lost compared with the two-tier system. Also, when the low tier becomes unavailable during a phone call, the user cannot complete the call in the single low-tier system, but the phone call can continue in the two-tier system. Although it is difficult to switch between these two systems (such as in a terrestrial cellular system and a satellite communications system), the customer can continue the conversation by redialing the call.

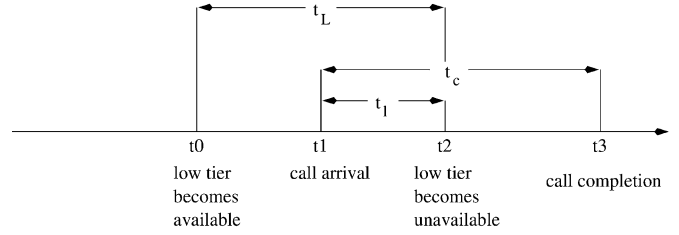


Fig. 10. Timing diagram for low-tier service availability.

Let θ denote the probability that the low tier becomes unavailable during a phone call connection (i.e., it is the probability that a call cannot be completed in the single low-tier system). The low-tier system is unavailable during a phone call if and only if the low-tier residence time is shorter than the call-holding time. We will derive the analytical formula based on the more realistic assumptions. Assume that the call originations/terminations are Poisson process, the call-holding time is generally distributed with rational Laplace transform, the low-tier availability time (residence time) has a general density function (with rational Laplace transform). Let $f_c^*(s)$ and $f_L^*(s)$ be the Laplace transforms for the call-holding times and the low-tier availability time, respectively. Let $1/\mu$ and $1/\eta$ denote the average call-holding time and the average low-tier availability time, respectively. Let t_0 and t_2 denote the beginning and ending of a typical low-tier availability time, respectively. Let t_1 and t_3 be the beginning and ending of a typical call-holding time, then $t_L = t_2 - t_0$ is the length of the availability time and t_c is the length of the call-holding time. Suppose that the call arrival t_1 is in between the low-tier availability period, then $t_l = t_2 - t_1$ is the period that the call is served by the low-tier system. Because t_l is the residual life of t_L , its Laplace transform $f_l^*(s)$ is given by

$$f_l^*(s) = \frac{\eta[1 - f_L^*(s)]}{s}. \quad (18)$$

Detailed timing diagram is shown in Fig. 10.

It is obvious that the low-tier system is not available for a call service if and only if the call-holding time is longer than the low-tier availability time, i.e., $\theta = \Pr(t_c > t_l)$. Using the similar technique, we could obtain the following result.

Theorem 3: If the call-holding times have rational Laplace transform, the low-tier service unavailability probability is given by

$$\begin{aligned} \theta &= - \sum_{p \in \sigma_c} s \stackrel{\text{Res}}{=} p \frac{f_l^*(s)}{s} f_c^*(-s) \\ &= - \sum_{p \in \sigma_c} s \stackrel{\text{Res}}{=} p \frac{\eta[1 - f_L^*(s)]}{s^2} f_c^*(-s) \\ &= 1 + \sum_{p \in \sigma_l} s \stackrel{\text{Res}}{=} p \frac{f_c^*(s)}{s} f_l^*(-s) \\ &= 1 + \eta \sum_{p \in \sigma_L} s \stackrel{\text{Res}}{=} p \frac{f_c^*(s)}{s^2} f_L^*(-s) \end{aligned} \quad (19)$$

where σ_c, σ_l , and σ_L are the set of poles of $f_c^*(-s), f_l^*(-s)$, and $f_L^*(-s)$ in the right half of the complex plane, respectively.

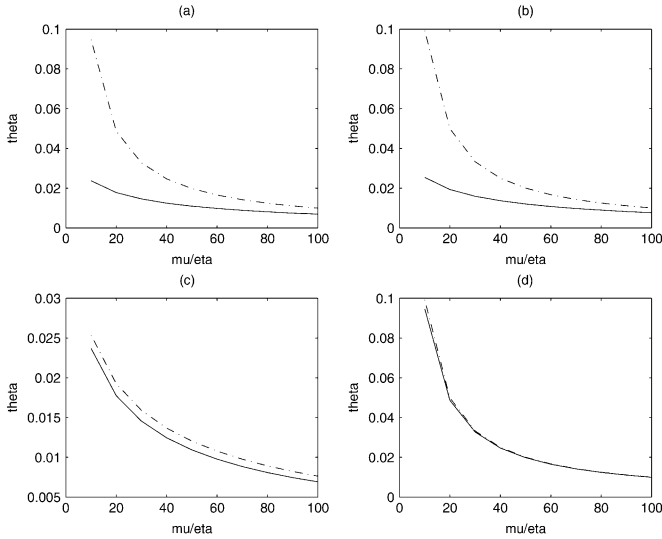


Fig. 11. Service availability. (a) and (b) Solid line is for $k = 0.01$ and dashed line is for $k = 10$. (c) and (d) Solid line is for $\gamma = 1$ and dashed line is for $\gamma = 2$.

If the call-holding times are Erlang distributed with parameters (k, μ) , then we have

$$\theta = \frac{(-1)^{k-1} \eta (k\mu)^k}{(k-1)!} \frac{d^{k-1}}{ds^{k-1}} \left[\frac{1 - f_L^*(s)}{s^2} \right] \Bigg|_{s=k\mu} \quad (20)$$

If the low-tier availability time t_L is Erlang distributed with parameters (γ, η) (here γ is a positive integer), then we have

$$\theta = 1 + \frac{(-1)^\gamma \eta (\gamma\eta)^\gamma}{(\gamma-1)!} \frac{d^{\gamma-1}}{ds^\gamma} \left[\frac{f_c^*(s)}{s^2} \right] \Bigg|_{s=\gamma\eta} \quad (21)$$

The second statement in the theorem was obtained in [23] by using a different approach. Obviously, whenever either t_c or t_L is distributed with a probability density function having a rational Laplace transform, the results in Theorem 3 can be easily used to derive easy-to-use analytical formula to compute service unavailability probability (hence, service availability probability).

In [23], the property of θ when the call-holding time is Erlang distributed has been investigated. Here we investigate how the variances of call-holding time and low-tier availability time affect the service availability probability. We assume that the call-holding time t_c is Gamma distributed with parameters (k, μ) (here k can be any positive real number) and that the low-tier availability time t_L is Erlang distributed with parameters (γ, η) . From Theorem 3, when $\gamma = 1$, we have

$$\theta = 1 - f_c^*(\eta) = 1 - \left(\frac{k\mu}{\eta + k\mu} \right)^k;$$

when $\gamma = 2$, we have

$$\begin{aligned} \theta &= 1 + \eta f_c^{*(1)}(2\eta) - f_c^*(2\eta) \\ &= 1 - \left(1 + \frac{k\eta}{2\eta + k\mu} \right) \left(\frac{k\mu}{2\eta + k\mu} \right)^k. \end{aligned}$$

Fig. 11 shows the results for service availability. We observe that the low-tier service unavailability probability θ decreases

as the call-to-mobility factor μ/η increases, which is consistent with our intuition. The variance of the call-holding time affects θ significantly, the lower the variance (i.e., the larger the value k), the more significant effect it has on θ . From Fig. 11(c) and (d), we observe that the variance of the low-tier service availability time affects θ only when the variance of call-holding time is large (i.e., k is small).

V. CONCLUSION

In this paper, we have studied the registration traffic analysis and service availability problem for multi-tier wireless networks.

The two-tier system generates more registration traffic than the single low-tier system. Under the range of the input parameters in our study, we showed that the two-tier system generates less than 15% extra registration traffic than the single low-tier system in most cases and generates less than 30% registration traffic in the worst cases. By relaxing the constraints in [23], the study in this paper shows results consistent with those in [23].

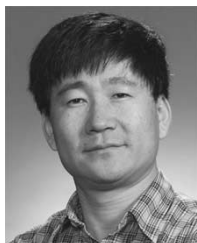
It is clear that, for service availability, the two-tier system is better than the single one-tier system. We also studied the probability that a call is forced terminated in the single low-tier system because the low tier becomes unavailable during the call (such a call can be continued in the two-tier system). Similar to the results in [23], under more general conditions, our study indicated that such a probability can be as large as 10% (5–10 times the engineered blocking probability of a typical PCS system). These forced terminated calls can be avoided in the two-tier system.

In summary, we generalize some of previously known results and obtain some analytical formulae that are easy to compute. The technique developed in this paper can be used to analyze more complicated tradeoff analysis for signaling and performance evaluation for mobility management in wireless mobile networks [9], [10].

REFERENCES

- [1] I. F. Akyildiz and S. H. Jeong, "Satellite ATM networks: A survey," *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 30–43, Jul. 1997.
- [2] N. J. Colella, J. Martin, and I. F. Akyildiz, "The HALO network," *IEEE Commun. Mag.*, vol. 38, no. 6, pp. 142–148, Jun. 2000.
- [3] K. Balachandran, "Convergence of 3G and WLANs," in *Proc. IEEE ICC'2002*, New York: Panel and Business Applications Sessions, Apr. 28–May 2, 2002.
- [4] D. C. Cox, "Wireless personal communications: What is it?," *IEEE Pers. Commun. Mag.*, vol. 2, no. 2, pp. 20–35, Apr. 1995.
- [5] E. Ekici and C. Ersoy, "Multi-tier cellular network dimensioning," *ACM Wireless Netw.*, vol. 7, no. 4, pp. 401–411, Jul. 2001.
- [6] "Cellular Intersystem Operations, Revision C," Electronic Industries Alliance/Telecommunications Industry Association, Tech. Rep. IS-41, 1997.
- [7] "Mobile Application Part (MAP), Specification, Version 4.8.0," European Telecommunications Standards Institute/Technical Committee, Tech. Rep. Recommendation GSM 09.02, 1994.
- [8] D. E. Everitt, "Traffic engineering of the radio interface for cellular mobile networks," *Proc. IEEE*, vol. 82, no. 9, pp. 1371–1382, Sep. 1994.
- [9] Y. Fang, "Movement-based location management and tradeoff analysis for wireless mobile networks," *IEEE Trans. Comput.*, vol. 52, no. 6, pp. 791–803, Jun. 2003.
- [10] —, "General modeling and performance analysis for location management in wireless mobile networks," *IEEE Trans. Comput.*, vol. 51, no. 10, pp. 1169–1181, Oct. 2002.

- [11] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling for PCS networks," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1062–1072, Jul. 1999.
- [12] Y. Fang, I. Chlamtac, and Y. B. Lin, "Channel occupancy times and hand-off rate for mobile computing and PCS networks," *IEEE Trans. Comput.*, vol. 47, no. 6, pp. 679–692, Jun. 1998.
- [13] —, "Call performance of a PCS network," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 7, pp. 1568–1581, Oct. 1997.
- [14] —, "Modeling PCS networks under general call holding time and cell residence time distributions," *IEEE Trans. Netw.*, vol. 5, no. 6, pp. 893–906, Dec. 1997.
- [15] S. Helal, C. Lee, Y. Zhang, and G. G. Richard III, "An architecture for wireless LAN/WAN integration," in *Proc. IEEE WCNC'2000*, Chicago, IL, Sep. 2000.
- [16] J. S. M. Ho and I. F. Akyildiz, "Local anchor scheme for reducing signaling cost in personal communications networks," *IEEE/ACM Trans. Netw.*, vol. 4, no. 5, pp. 709–725, Oct. 1996.
- [17] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," Reference no. ISO/IEC 8802-1:1999(E), 1999. ANSI/IEEE Standard 802.11 Part 11.
- [18] P. Karlsson, "Integration of WLAN and cellular networks," in *First Mobile & Wireless Telecommun. Summit*, Thessaloniki, Greece, Jun. 17–19, 2002.
- [19] F. P. Kelly, *Reversibility and Stochastic Networks*. New York: Wiley, 1979.
- [20] L. Kleinrock, *Queueing Systems: Theory*. New York: Wiley, 1975, vol. I.
- [21] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Philadelphia, PA: SIAM, 1999.
- [22] W. R. LePage, *Complex Variables and the Laplace Transform for Engineers*. New York: Dover, 1980.
- [23] Y. B. Lin, "A comparison study of the two-tier and the single-tier personal communications services systems," *ACM/Baltzer Mobile Netw. Appl.*, vol. 1, no. 1, pp. 29–38, Aug. 1996.
- [24] Y. Lin, L. F. Chang, A. R. Noerpel, and K. Park, "Performance modeling of multi-tier PCS system," *Int. J. Wireless Info. Netw.*, vol. 3, no. 2, pp. 67–78, 1996.
- [25] Y. B. Lin and I. Chlamtac, "Heterogeneous personal communications services: Integration of PCS systems," *IEEE Commun. Mag.*, vol. 34, no. 9, pp. 106–113, Sep. 1996.
- [26] Y. B. Lin and I. Chlamtac, *Wireless and Mobile Network Architecture*. New York: Wiley, 2000.
- [27] M. J. Miller, B. Vucetic, and L. Berry, *Satellite Communications: Mobile and Fixed Services*, 2nd printing ed., Boston, MA: Kluwer, 1995.
- [28] K. Park and Y. Lin, "Registration methods for multi-tier personal communications services," *IEEE Trans. Veh. Technol.*, vol. 46, no. 3, pp. 597–602, Aug. 1997.
- [29] J. E. Rizzo and N. Sollenberger, "Multitier wireless access," *IEEE Pers. Commun. Mag.*, vol. 2, no. 3, pp. 18–31, Jun. 1995.
- [30] S. M. Ross, *Stochastic Processes*, 2nd ed., New York: Wiley, 1996.
- [31] J. E. Russell, "Emergence of a paradigm shift in the communications industry," *Int. J. Wireless Inf. Netw.*, Jul. 1994.
- [32] L. Tassiulas and F. Anjum, "A hierarchical multiresolution registration structure for mobility tracking," in *IEEE Int. Conf. Universal Pers. Commun. (ICUPC'96)*, Cambridge, MA, Sep. 1996.



Yuguang Fang (S'92–M'94–S'96–M'97–SM'99) received the Ph.D. degree in systems and control engineering from Case Western Reserve University in January 1994, and the Ph.D. degree in electrical engineering from Boston University in May 1997.

From June 1997 to July 1998, he was a Visiting Assistant Professor in the Department of Electrical Engineering at the University of Texas at Dallas. From July 1998 to May 2000, he was an Assistant Professor in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology. In

May 2000, he joined the Department of Electrical and Computer Engineering at the University of Florida where he got the early promotion with tenure in August 2003 and has been an Associate Professor since then. He has published more than 140 papers in refereed professional journals and conferences.

Dr. Fang received the National Science Foundation Faculty Early Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002. He is currently serving as an Editor for many journals including *IEEE TRANSACTIONS ON COMMUNICATIONS*, *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, *IEEE TRANSACTIONS ON MOBILE COMPUTING*, and *ACM WIRELESS NETWORKS*. He is also actively participating in conference organization such as the Program Vice-Chair for *IEEE INFOCOM'2005*, Program Co-Chair for the Global Internet and Next Generation Networks Symposium in *IEEE Globecom'2004* and the Program Vice Chair for 2000 *IEEE Wireless Communications and Networking Conference (WCNC'2000)*.



Yi-Bing Lin received the B.S.E.E. degree from National Cheng Kung University in 1983, and the Ph.D. degree in computer science from the University of Washington in 1990.

From 1990 to 1995, he was with the Applied Research Area at Bell Communications Research (Bellcore), Morristown, NJ. In 1995, he was appointed as a Professor of the Department of Computer Science and Information Engineering (CSIE), National Chiao Tung University (NCTU). In 1996, he was appointed as Deputy Director of Microelectronics and Information Systems Research Center, NCTU. During 1997–1999, he was elected as Chairman of CSIE, NCTU. He has been appointed as Chair Professor of NCTU since 2004. His current research interests include design and analysis of personal communications services network, mobile computing, distributed simulation, and performance modeling. He is an author of the book *Wireless and Mobile Network Architecture* (coauthor with Imrich Chlamtac; New York: Wiley & Sons, 2000). He has published more than 180 journal articles and more than 200 conference papers.

Dr. Lin is a Senior Technical Editor of *IEEE NETWORK*, an Editor of *IEEE TRANSACTION ON WIRELESS COMMUNICATIONS*, an Associate Editor of *IEEE TRANSACTION ON VEHICULAR TECHNOLOGY*, an Associate Editor of *IEEE COMMUNICATIONS SURVEY AND TUTORIALS*, an Editor of *IEEE WIRELESS COMMUNICATIONS MAGAZINE*, an Editor of *ACM/BALTZER WIRELESS NETWORKS*, Program Chair for the 8th Workshop on Distributed and Parallel Simulation, General Chair for the 9th Workshop on Distributed and Parallel Simulation. Program Chair for the 2nd International Mobile Computing Conference, Guest Editor for the *ACM/Baltzer MONET Special Issue on Personal Communications*, a Guest Editor for *IEEE Transactions on Computers Special Issue on Mobile Computing*, a Guest Editor for *IEEE Transactions on Computers special issue on Wireless Internet*, and a Guest Editor for *IEEE Communications Magazine special issue on Active, Programmable, and Mobile Code Networking*. He received 1998, 2000, and 2002 Outstanding Research Awards from National Science Council, ROC, and 1998 Outstanding Youth Electrical Engineer Award from CIEE, ROC. He also received the NCTU Outstanding Teaching Award in 2002. He is an Adjunct Research Fellow of Academia Sinica, and is Chair Professor of Providence University. He is an ACM Fellow.