Authentication Vector Management for UMTS

Lin-Yi Wu and Yi-Bing Lin, *Fellow, IEEE*

***Abstract*— In *Universal Mobile Telecommunication System* (UMTS), the security function provides mutual authenticity and key agreement between the core network and the *Mobile Station* (MS). Specifically, the *Serving GPRS Support Node* (SGSN) in the core network obtains an array of *Authentication Vectors* (AVs) from the *Home Subscriber Server/Authentication Center* (HSS/AuC), and consumes one AV for each mutual authentication. After the departure of the MS, the SGSN may keep the unused AVs for a time interval called the *Reservation Timeout* (RT) period. If the MS returns within the RT period, the SGSN uses the stored AVs for mutual authentication instead of obtaining new AVs from the HSS/AuC. Note that a long RT period results in fewer accesses to the HSS/AuC at the cost of extra AV storage in the SGSN. In this paper, we propose an analytic model to investigate the impact of the RT period on the system performance. Our study provides the guidelines for the mobile operators to select an appropriate RT period.**

***Index Terms*— 3G, authentication vector, security function, UMTS core network.**

I. Introduction

UNIVERSAL MOBILE TELECOMMUNICATION SYSTEM (UMTS) [1] supports multimedia applications with quality of services. The UMTS network includes three parts: *Mobile Station* (MS; Fig. 1 (1)) is the equipment through which a user accesses UMTS services. Core Network (Fig. 1 (2)) provides mobility management, session management and transport for IP-based services. *UMTS Terrestrial Radio Access Network* (UTRAN; Fig. 1 (3)) provides wireless connectivity between the MS and the core network. UTRAN consists of *Radio Network Controllers* (RNCs; Fig. 1 (8)) and Node Bs (Fig. 1 (9)). In the core network, *Serving GPRS Support Node* (SGSN; Fig. 1 (4)) and *Gateway GPRS Support Node* (GGSN; Fig. 1 (5)) provide mobility and session services to mobile users. One SGSN connects to several RNCs, and one RNC connects to one or more Node Bs. The coverage of the Node Bs connected to the same SGSN is called an SGSN area. In Fig. 1, SGSN area 1 (Fig. 1 (10)) corresponds to SGSN1, and SGSN area 2 (Fig. 1 (11)) corresponds to SGSN2. The

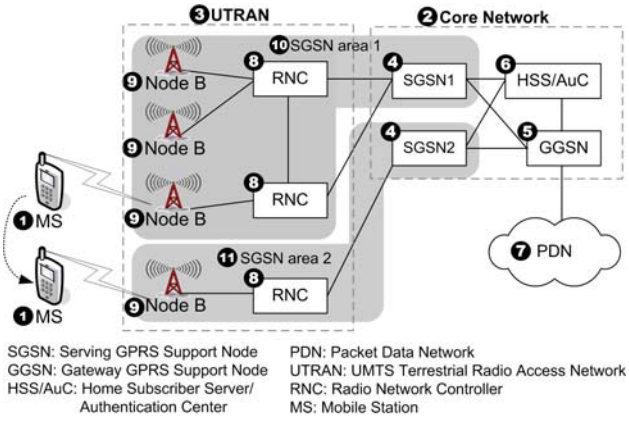


Fig. 1. The UMTS architecture.

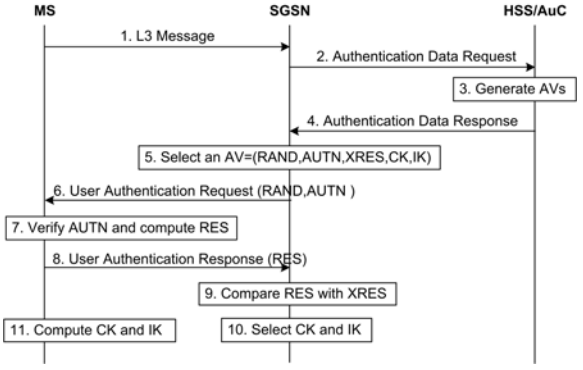


Fig. 2. Message flow for UMTS security.

GGSN connects to the external *Packet Data Network* (PDN; Fig. 1 (7)) by an IP-based interface. Both SGSN and GGSN communicate with the *Home Subscriber Server/Authentication Center* (HSS/AuC; Fig. 1 (6)) for mobility and session management. The HSS/AuC is the master database containing all user-related subscription and location information.

When an MS sends initial L3 messages (e.g., location update request, connection management service request, paging response, etc.) to the SGSN, the security function is activated to provide mutual authentication between the HSS/AuC and the MS [2]. Mutual authentication is achieved by showing the knowledge of a preshared secret key that is only available in the MS and the AuC. Specifically, an *Authentication Vector* (AV) is used for one authentication and key agreement between the SGSN and the MS. Fig. 2 illustrates the message flow for UMTS security.

Step 1. The MS sends an L3 message (e.g., registration) to the SGSN.

Steps 2-4. Upon receipt of the L3 message, if the SGSN possesses the AVs of the MS, Steps 2-4 are skipped.

Manuscript received May 11, 2006; revised January 1, 2007; accepted January 10, 2007. The associate editor coordinating the review of this paper and approving it for publication was X. Zhang. This work was sponsored in part by the NSC Excellence project NSC 95-2752-E-009-005-PAE, NSC 95-2218-E-009-201-MY3, NSC 94-2219-E-009-001, NSC 94-2219-E-009-024, NTP SIP-based B3G project under grant number NSC 95-2219-E-009-010, NTP IMS Integration Project under grant number NSC 95-2219-E-009-019, Intel, Chung Hwa Telecom, IIS/Academia Sinica, ITRI/NCTU Joint Research Center, and MoE ATU.

L.-Y. Wu is with MediaTek Inc., No. 1, Dusing Rd. 1, HsinChu Science-Based Industrial Park, HsinChu, Taiwan 300, R.O.C. (e-mail: lywyu@csie.nctu.edu.tw).

Y.-B. Lin is with the Department of Computer Science and Information Engineering, National Chiao Tung University, 1001 Ta Hseuh Rd., Hsinchu 30030, Taiwan (e-mail: liny@csie.nctu.edu.tw). He is also with the Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan.

Digital Object Identifier 10.1109/TWC.2007.060245.

Otherwise, the SGSN sends an Authentication Data Request message to the HSS/AuC to retrieve the AVs. The HSS/AuC generates an array of AVs, and returns them to the SGSN through the Authentication Data Response message. Steps 2-4 are called the *Authentication Data Request & Response* (ADR) operation. An AV contains 5 elements: a random challenge RAND, a network authentication token AUTN, an expected user response XRES, a cipher key CK, and an integrity key IK.

Steps 5 and 6. The SGSN selects the next unused AV in the AV array to perform the security function. Then the SGSN sends a User Authentication Request with parameters RAND and AUTN to the MS.

Steps 7 and 8. The MS authenticates the SGSN by verifying AUTN. After successfully authenticating the network, the MS computes the user response RES by RAND and preshared key, and then sends RES to the SGSN through User Authentication Response. Steps 6-8 are called the *User Authentication Request & Response* (UAR) operation.

Step 9. The SGSN compares RES with XRES. If they are matched, the MS is successfully authenticated.

Steps 10 and 11.The MS computes CK and IK by RAND and preshared Key. On the other hand, the SGSN retrieves CK and IK from the selected AV. CK and IK are used in both sides for encryption/decryption and integrity check.

Since the cost for accessing HSS/AuC is expensive (especially when SGSN and HSS/AuC are located in different countries), the issues of the AV management have been drawn considerable attention. In [8], long delay of AV distribution from HSS/AuC to SGSN was discussed. The authors proposed a new approach, in which the SGSN requests HSS/AuC for new AVs before all AVs kept in the SGSN are consumed. In [3], an analytic model was presented to investigate the appropriate number of AVs distributed in one access to HSS/AuC. Based on the analytic results, an automatic selection mechanism was proposed to reduce the network signaling cost.

In this paper, we will investigate on another significant factor in managing AVs. When an MS moves from one SGSN area to another, the old SGSN may store the unused AVs for an interval called the *reservation timeout* (RT) period. If the MS returns to the SGSN area within the RT period, the SGSN will utilize these stored AVs for authentication instead of obtaining new AVs from the HSS/AuC. Therefore, the signaling traffic for accessing the HSS/AuC is reduced. Note that a long RT period results in fewer accesses to the HSS/AuC at the cost of more AV storage required at an SGSN. This paper investigates the effect of the RT period on the system performance. In Section II, the AV usage mechanism is described. Section III presents an analytic model for measuring system performance. Numerical examples are given in Section IV. Finally, Section V concludes this paper.

II. The AV Usage Mechanism

Consider an SGSN area L_0. When an MS resides at L_0, the authentication activities are shown in Fig. 3. In this figure, the MS enters L_0 at time τ_1 (Fig. 3 (1)), and sends a registration request to L_0 at time $\tau_{1,1,1}$. This registration request activates a UAR for mutual authentication between the MS and SGSN L_0 (Fig. 3 (2)). Since L_0 does not have authentication information of the MS at its first visit, SGSN L_0 obtains an array of K AVs from the HSS/AuC through an ADR (Fig. 3 (3)), and utilizes the first AV for the UAR request. Subsequently, more UARs may be issued by the MS, and SGSN L_0 utilizes the next unused AV in the array to perform the following UARs (Fig. 3 (4), (5), and (6)). After $\tau_{1,1,k}$, all AVs have been consumed for UARs (Fig. 3 (6)). Therefore, when a UAR arrives at $\tau_{1,2,1}$ (Fig. 3 (7)), SGSN L_0 issues the second ADR (Fig. 3 (8)) to obtain the next AV array from the HSS/AuC and uses the first AV in the array to perform the UAR. At time τ_1^* (Fig. 3 (9)), the MS leaves L_0 with R_1 unused AVs. Then SGSN L_0 starts the RT timer of length T (Fig. 3 (10)), and keeps the R_1 unused AVs in its storage during the RT period. If the MS returns to L_0 before the RT timer expires (Fig. 3 (11)), SGSN L_0 will utilize these stored AVs for the next R_1 UARs (Fig. 3 (12), (13), and (14)). After the R_1 AVs are consumed, a UAR occurs at time $\tau_{2,1,1}$ (Fig. 3 (15)), and SGSN L_0 issues an ADR to obtain a new AV array (Fig. 3 (16)). At time τ_2^*, the MS leaves L_0 again (Fig. 3 (17)). Let the residence time of the i-th visit to L_0 be t_i. During t_i, N_i ADRs are executed. When the MS leaves L_0, there are R_i unused AVs, which will be subsequently used at the (i+1)-th visit of the MS. Note that if the RT timer expires before the MS returns to L_0 (Fig. 3 (18)), then these unused AVs are discarded; that is, R_i = 0 for the (i+1)-th visit.

III. An Analytic Model

This section investigates the effect of the RT period T on the performance of the AV management. The following parameters and assumptions are made.

- The UAR arrivals are Poisson processes with rate λ .
- The SGSN residence time is exponentially distributed with rate μ (this exponential assumption will be relaxed in the simulation experiments).

Three output measures are evaluated in our study.

- α: the probability that the MS re-enters L_0 within the RT period T
- β: the expected AV storage consumed when $T > 0$, which is normalized by the expected AV storage consumed when $T = 0$
- δ: the number of ADRs performed in one visit to SGSN L_0 as comparing with that when $K = 1$. Let $E[N|K]$ be the expected number of ADRs performed in one visit to L_0, where K AVs are obtained in one ADR. Then

$$\delta = \frac{E[N|K]}{E[N|K=1]} \tag{1}$$

In the following sub-sections, we derive the above output measures.

A. Derivation of Probability α

We utilize a two-dimensional random walk to model the MS movement. Fig. 4 shows the layout of the SGSN areas, where

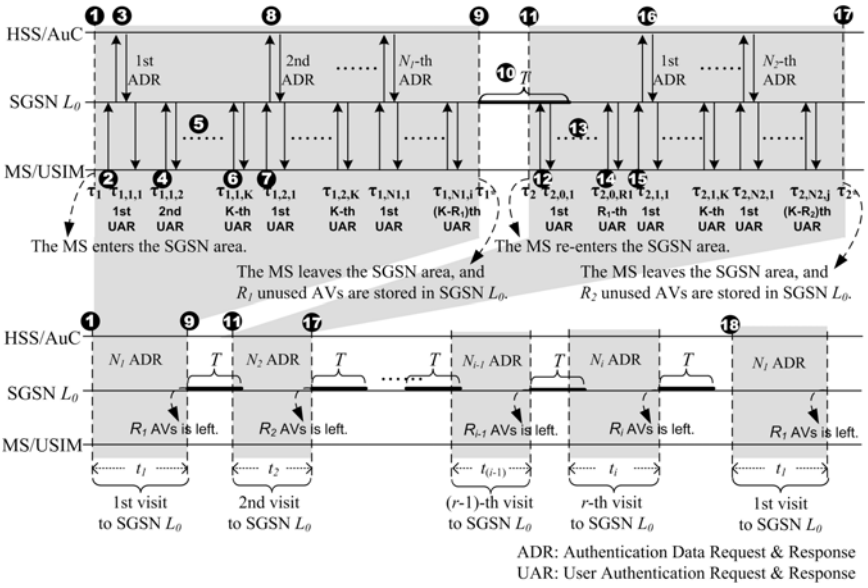


Fig. 3. MS authentication activities at an SGSN.

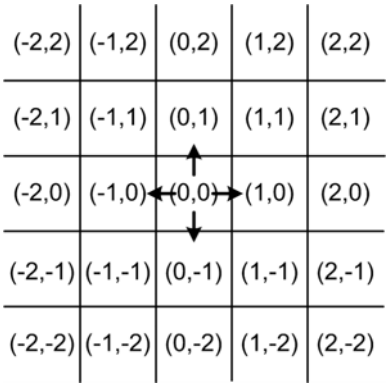


Fig. 4. Two-dimensional SGSN layout for the random walk model.

a coordinate (x, y) specifies the location of an SGSN area. An MS resides in an SGSN area for a period, then moves to one of its four neighbors with the same probability 1/4. Let L_j = (x_j, y_j) presents the coordinate of the SGSN where the MS resides after j movements. Initially, the MS resides at L_0 = (0,0).

Let P_j be the probability that the MS returns to L_0 at the j-th movement. That is, P_j = Pr[$L_j = L_0$]. Following the two-dimensional random walk model, $P_j = 0$ if j is odd. Consider the even movements where $j = 2n$. Assume that there are a movements to the east and the same number of movements to the west. Similarly, there are b movements to the north and the same number of movements to the south. Then $L_{2n} = L_0$ if $2a + 2b = 2n$. For $a \geq 0$ and $b \geq 0$, we have

$$P_{2n} = \left(\frac{1}{4}\right)^{2n} \left(\begin{array}{c} 2n \\ n \end{array} \right)^2 \tag{2}$$

Let Q_{2n} be the probability that the MS first returns to L_0 at the $2n$-th movement. In other words, Q_{2n} = Pr[$L_{2n} = L_0$, $L_{2l} \neq L_0$ for $0 < l < n$]. It is obvious that for $n = 1$, $Q_2 = P_2$.

Suppose that the MS enters L_0 at the $2n$-th movement, and the prior visit to L_0 occurs at the $2m$-th movement, where $0 \leq m < n$. Then we have

$$\begin{aligned} P_{2n} &= Pr[L_{2n} = L_0] \\ &= \sum_{m=0}^{n-1} Pr[L_{2m} = L_0] \times Pr[L_{2n} = L_{2m} = L_0, \\ &\qquad L_{2l} \neq L_0 \ for \ m < l < n] \\ &= \sum_{m=0}^{n-1} P_{2m} \times Q_{2(n-m)} \end{aligned} \tag{3}$$

By rearranging (3), we have

$$Q_{2n} = P_{2n} - \sum_{m=1}^{n-1} P_{2m} Q_{2(n-m)} \tag{4}$$

Since the MS returns to L_0 at the $2n$-th movement, the MS moves across $2n - 1$ SGSN areas before it returns. For $j = 1, 2, ..., 2n - 1$, let the residence time at L_j be t_j^* with the density function $f(t_j^*) = \mu e^{-\mu t_j^*}$. Let t_r be the period between when the MS leaves L_0 and when it returns. That is $t_r = t_1^* + t_2^* + ... + t_{2n-1}^*$, where $L_{2n} = L_0$, and $L_{2l} \neq L_0$ for $0 < l < n$. Let $F(2n, t_r)$ be the cumulative distribution function that the MS returns to L_0 at $2n$-th movement at time t_r. It is clear that the $F(2n, t_r)$ is an Erlang distribution

$$F(2n, t_r) = 1 - \sum_{j=0}^{2n-2} \left[\frac{(\mu t_r)^j}{j!} \right] e^{-\mu t_r} \tag{5}$$

From (4) and (5), α is derived as follows.

$$\begin{aligned} \alpha &= \sum_{n=1}^{\infty} \{Pr[the \ MS \ first \ returns \ to \ L_0 \ at \ the \ 2n- \\ &\quad th \ movement] \times Pr[the \ MS \ moves \ 2n \ steps \\ &\quad within \ time \ T]\} \\ &= \sum_{n=1}^{\infty} Q_{2n} \times F(2n, T) \end{aligned}$$

$$= \sum_{n=1}^{\infty} Q_{2n} \times \left\{ 1 - \sum_{j=0}^{2n-2} \left[\frac{(\mu T)^j}{j!} \right] e^{-\mu T} \right\} \quad (6)$$

B. 3.2 Derivation for β

Consider the Markov chain illustrated in Fig. 5, where state S_k represents that there are k AVs stored in SGSN L_0. In this figure, the transition probability in a short observation interval Δs is considered. The descriptions of the transitions are given below.

- Transition 1: At state $S_k(k > 0)$, the MS resides in L_0 with probability $(1 - \mu\Delta s)$. A UAR arrives with probability $\lambda\Delta s$, which decrements the number of AVs by one. That is, the Markov chain moves from S_k to S_{k-1} with probability $(1 - \mu\Delta s)\lambda\Delta s$.
- Transition 2: At S_0, the MS resides at L_0 with probability $(1 - \mu\Delta s)$, and a UAR occurs with probability $\lambda\Delta s$. Since $k = 0$, SGSN L_0 issues an ADR to obtain K AVs from the HSS/AuC, and uses the first AV to perform the UAR. The remaining $K - 1$ AVs are stored in SGSN L_0. Therefore, the Markov chain moves from S_0 to S_{K-1} with probability $(1 - \mu\Delta s)\lambda\Delta s$.
- Transition 3: At state $S_k(k > 0)$, the MS leaves L_0 with probability $\mu\Delta s$, and it does not return to L_0 within T with probability $(1-\alpha)$. In this case, all AVs stored in SGSN L_0 are discarded. That is, the Markov chain moves from S_k to S_0 with probability $\mu\Delta s(1 - \alpha)$.
- Transition 4: At state $S_k(k > 0)$, the MS stays in L_0 with probability $(1 - \mu\Delta s)$, and the probability that no UAR occurs during Δs is $(1-\lambda\Delta s)$. In this case, the number of AVs stored in SGSN L_0 is not changed. Also, if the MS leaves L_0 with probability $\mu\Delta s$ and returns before RT expires with probability α, all unused AVs are still stored in SGSN L_0. Thus, the state remains in S_k with probability $(1 - \mu\Delta s)(1 - \lambda\Delta s) + \mu\Delta s\alpha$.
- Transition 5:At S_0, SGSN L_0 does not keep any AVs, and the state remains in S_0 with probability $1 - (1 - \mu\Delta s)\lambda\Delta s$.

Based on the above transitions, when $\Delta s \to 0$, the transition rate matrix H for the Markov chain can be expressed as

$$H = \begin{bmatrix} -\lambda & 0 & \dots & 0 & \lambda \\ \lambda + \mu - \mu\alpha & \mu\alpha - \mu - \lambda & & 0 & 0 \\ \mu(1-\alpha) & \lambda & & 0 & 0 \\ \mu(1-\alpha) & 0 & & 0 & 0 \\ \mu(1-\alpha) & 0 & & 0 & 0 \\ \mu(1-\alpha) & 0 & & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \mu(1-\alpha) & 0 & & \mu\alpha - \mu - \lambda & 0 \\ \mu(1-\alpha) & 0 & \dots & \lambda & \mu\alpha - \mu - \lambda \end{bmatrix}_{K \times K}$$

Let $\pi = (\pi_0, \pi_1, \pi_2, ..., \pi_{K-1})$ be the probability matrix where π_k is the probability that k AVs are stored in SGSN L_0 at the steady state. Since $\pi H = 0$, we have

$$\left.\begin{aligned} -\lambda\pi_0 + \lambda\pi_1 + \mu(1-\alpha)(\pi_1 + \pi_2 + \ldots + \pi_{K-1}) &= 0 \\ (\mu\alpha - \mu - \lambda)\pi_1 + \lambda\pi_2 &= 0 \\ (\mu\alpha - \mu - \lambda)\pi_2 + \lambda\pi_3 &= 0 \\ \vdots & \\ (\mu\alpha - \mu - \lambda)\pi_{K-2} + \lambda\pi_{K-1} &= 0 \\ (\mu\alpha - \mu - \lambda)\pi_{K-1} + \lambda\pi_0 &= 0 \end{aligned}\right\} \quad (7)$$

By rearranging (7), we have

$$\pi_k = A^{K-k}\pi_0 \quad for\ l \le k \le K-1, \quad where\ A = \frac{\lambda}{\mu + \lambda - \mu\alpha} \quad (8)$$

By solving (8) with $\sum_{i=0}^{K-1} \pi_i = 1$, we have

$$\left.\begin{aligned} \pi_0 &= \tfrac{1-A}{1-A^K} \\ \pi_k &= A^{K-k}\left(\tfrac{1-A}{1-A^K}\right) \quad for\ 1 \le k \le K-1 \end{aligned}\right\} \quad (9)$$

Let Ψ_T be the expected AV storage consumed in one visit to L_0 when $T \ge 0$. Then

$\Psi_T = E[the\ number\ of\ AVs\ stored\ in\ SGSN\ L_0\ at\ any\ time] \times E[the\ time\ period\ that\ SGSN\ L_0\ stores\ AVs]$

where

$$\begin{aligned} &E[the\ number\ of\ AVs\ stored\ in\ SGSN\ L_0\ at\ any\ time] \\ &= \sum_{k=0}^{K-1} k\pi_k \\ &= \left(\frac{1-A}{1-A^K}\right)\left[\frac{A^{K-1}-1}{(1-A^{-1})^2} + \frac{1-K}{1-A^{-1}}\right] \end{aligned} \quad (10)$$

As previously defined, t_r is the period between when the MS leaves L_0 and when it returns. We have

$$\begin{aligned} &E[the\ period\ that\ SGSN\ L_0\ stores\ AVs] \\ &= E[the\ period\ that\ MS\ resides\ in\ L_0] \\ &\quad + E[the\ period\ that\ SGSN\ L_0\ stores\ the\ unused\ AVs\ after\ the\ MS\ leaves\ L_0] \\ &= E[the\ period\ that\ the\ MS\ resides\ in\ L_0] + Pr[t_r > T] \times E[T|t_r > T] + Pr[t_r \le T] \times E[t_r|t_r \le T] \\ &= \frac{1}{\mu} + (1-\alpha)T \\ &\quad + \int_{t=0}^{T} t\left[\frac{d\left(\sum_{n=1}^{\infty} Q_{2n} \times F(2n,t)\right)}{dt}\right] dt \\ &= \frac{1}{\mu} + (1-\alpha)T + \sum_{n=1}^{\infty}\left\{\left[\frac{Q_{2n}(2n-1)}{\mu}\right] \times \left\{1 - e^{-\mu T}\left[1 + \sum_{r=0}^{2n-2}\frac{(\mu T)^{2n-r-1}}{(2n-r-1)!}\right]\right\}\right\} \end{aligned} \quad (11)$$

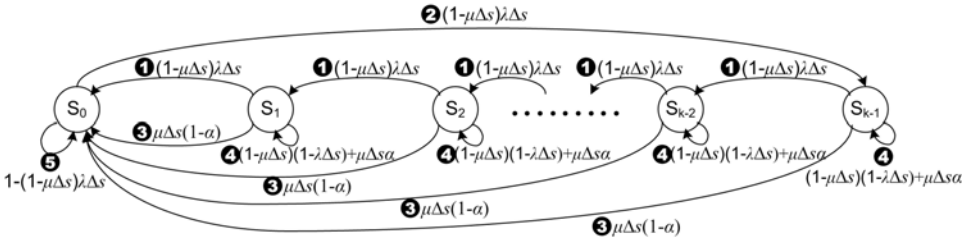


Fig. 5. The state transition diagram for the AVs size at SGSN L_0.

From (10) and (11),

$$\Psi_T = \left\{\left(\frac{1-A}{1-A^K}\right)\left[\frac{A^{K-1}-1}{(1-A^{-1})^2}+\frac{1-K}{1-A^{-1}}\right]\right\} \times\left\{\frac{1}{\mu}+(1-\alpha)T+\sum_{n=1}^{\infty}\left\{\left[\frac{Q_{2n}(2n-1)}{\mu}\right] \times\left\{1-e^{-\mu T}\left[1+\sum_{r=0}^{2n-2}\frac{(\mu T)^{2n-r-1}}{(2n-r-1)!}\right]\right\}\right\}\right\} \tag{12}$$

From (12), β is derived as

$$\beta=\frac{\Psi_T}{\Psi_0} \tag{13}$$

C. Derivation for δ

Suppose that K AVs are obtained in one ADR. The expected number $E[N|K]$ of ADRs performed in one visit to L_0 is derived as follows. For $i\geq 1$, let $\theta(N_i,R_{i-1},R_i,t_i)$ be the probability that

(i) at the i-th visit to L_0, R_{i-1} unused AVs are stored in SGSN L_0, where $R_0=0$ and $0\leq R_{i-1}<K$,
(ii) the residence time of the i-th visit to L_0 is t_i,
(iii) during t_i, N_i ADRs occur, and
(iv) there are R_i unused AVs when the MS leaves L_0, where $0\leq R_i<K$.

Since $N_i=0$ has no effect on $E[N_i]$, it suffices to consider $N_i>0$ in the derivation. In this case, $N_iK+R_{i-1}-R_i$ UARs are performed in the period t_i, and $N_iK+R_{i-1}-R_i>0$. Therefore

$$\theta(N_i,R_{i-1},R_i,t_i)=e^{-\lambda t}\left[\frac{(\lambda t_i)^{N_iK+R_{i-1}-R_i}}{(N_iK+R_{i-1}-R_i)!}\right]$$

Let $\varphi(N_i,R_{i-1})$ be the probability that when the MS enters L_0 at the i-th visit, R_{i-1} unused AVs are stored in SGSN L_0, and N_i ADRs are performed during the residence time of the i-th visit. For $N_i>0$, $\varphi(N_i,R_{i-1})$ is derived as

$$\begin{aligned}&\varphi(N_i,R_{i-1})\\&=\sum_{R_i=0}^{K-1}\left[\int_{t_i=0}^{\infty}\theta(N_i,R_{i-1},R_i,t_i)\times f(t_i)dt_i\right]\\&=\left(\frac{\lambda}{\lambda+\mu}\right)^{N_iK+R_{i-1}+1}\left[\left(\frac{\lambda+\mu}{\lambda}\right)^{K}-1\right]\end{aligned} \tag{14}$$

Let $\Gamma(N_i,T)$ be the probability that N_i ADRs are performed at the i-th visit to L_0, where $i\geq 1$ and $N_i>0$, and the length of the RT period is T. Consider the following two cases:

Case 1: The MS re-enters L_0 within T. Thus, SGSN L_0 still stores R_{i-1} unused AVs, where $0\leq R_{i-1}<K$. The probability of Case 1 is α.

Case 2: The MS re-enters L_0 after the RT timer expires, and $R_{i-1}=0$. The probability of Case 2 is $(1-\alpha)$.

Then we have

$$\Gamma(N_i,T)=\alpha\left[\sum_{k=0}^{K-1}\pi_k\varphi(N_i,k)\right]+(1-\alpha)\varphi(N_i,0) \tag{15}$$

The first term of the right hand side in (15) is derived as follows. From (9) and (14)

$$\begin{aligned}&\alpha\left[\sum_{k=0}^{K-1}\pi_k\varphi(N_i,k)\right]\\&=\alpha\left(\frac{1-A}{1-A^K}\right)\left(\frac{\lambda}{\lambda+\mu}\right)^{N_iK+1}\left[\left(\frac{\lambda+\mu}{\lambda}\right)^{K}-1\right]\\&\times\left\{1+A^K\left[\frac{\lambda}{A(\lambda+\mu)-\lambda}\right]\left[1-\left(\frac{\lambda}{A(\lambda+\mu)}\right)^{K-1}\right]\right\}\end{aligned} \tag{16}$$

The second term of the right hand side in (15) is derived as follows.

$$\begin{aligned}&(1-\alpha)\varphi(N_i,0)\\&=(1-\alpha)\left(\frac{\lambda}{\lambda+\mu}\right)^{N_iK+1}\left[\left(\frac{\lambda+\mu}{\lambda}\right)^{K}-1\right]\end{aligned} \tag{17}$$

Let $E[N|K]=\lim_{i\to\infty}E[N_i|K]$ be the expected number of ADRs performed during one visit to SGSN L_0 at the steady state (i.e., when $i\to\infty$). From (15), (16), and (17), $E[N|K]$ is expressed as:

$$\begin{aligned}&E[N|K]\\&=\sum_{N=1}^{\infty}N\times\Gamma(N,T)\\&=\left\{\frac{\alpha\left(\frac{\lambda}{\lambda+\mu}\right)\left(\frac{1-A}{1-A^K}\right)}{1-\left(\frac{\lambda}{\lambda+\mu}\right)^K}\right\}\left\{1+\left[\frac{\lambda A^K}{A(\lambda+\mu)-\lambda}\right]\right.\end{aligned}$$

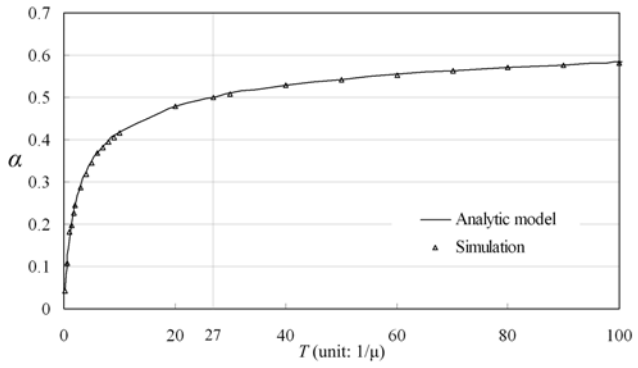


Fig. 6. Effects of T on α.

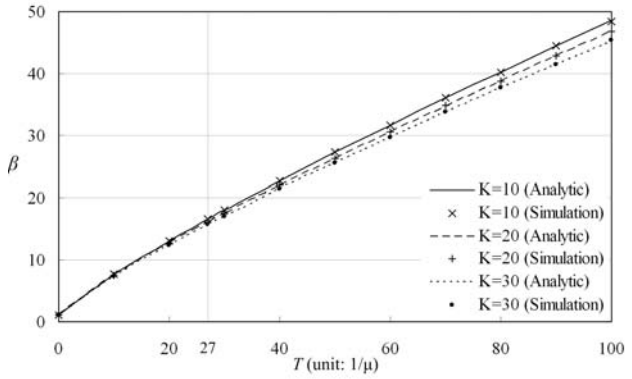


Fig. 7. Effects of T and K on β ($\lambda = 20\mu$).

$$\times \left\{1 - \left[\frac{\lambda}{A(\lambda+\mu)}\right]^{K-1}\right\}\Bigg\} + \frac{1}{\left[1 - \left(\frac{\lambda}{\lambda+\mu}\right)^{K}\right]^{2}} \times \left\{(1-\alpha)\left(\frac{\lambda}{\lambda+\mu}\right)\left[1 - \left(\frac{\lambda}{\lambda+\mu}\right)^{K}\right]\right\} \tag{18}$$

Finally, δ is derived from (1) and (18).

IV. Numerical Examples

Based on the analytic models, we use numerical examples to investigate how the RT period T affects the performance of AV management. These numerical examples also validate the simulation model against the analytic analysis in Section III. As shown in Figs. 6-8, the discrepancies between the analytic and simulation models are less than 1%. Therefore the analytic and simulation models are consistent.

Based on (6), Fig. 6 plots the probability α of re-entrance to L_0 against the RT period T. The figure indicates that for $T < 27/\mu$, α significantly increases as T increases. For $T \geq 27/\mu$, the impact of T on α becomes less significant. We note that the α curve is determined by the probabilities of the movement directions. In the two-dimensional random walk, if the routing probabilities of the movement directions are not the same, then it is very likely that the MS will never return to L_0. In the real world, the MS movement may exhibit locality, and the MS eventually moves back to L_0.

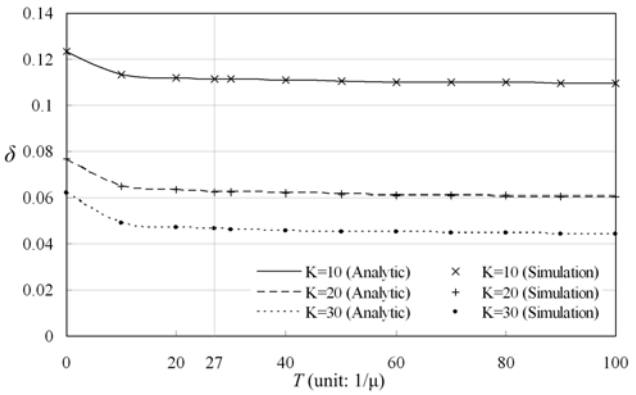
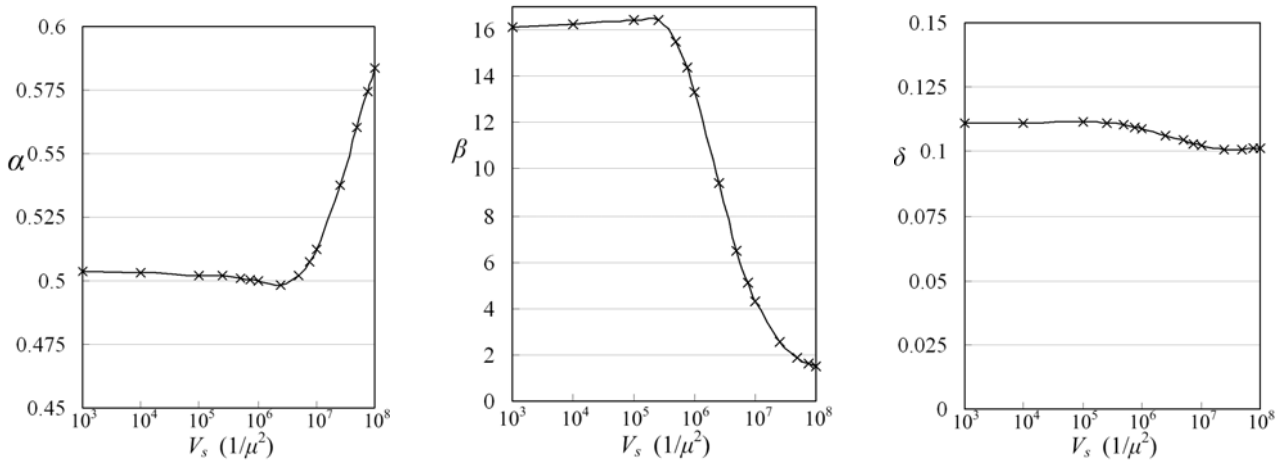


Fig. 8. Effects of T and K on δ ($\lambda = 20\mu$)

According to (12) and (13), Fig. 7 plots the normalized AV storage β against T and K. The figure indicates that β is an almost linearly increasing function of T. When $T = 27/\mu$, SGSN L_0 consumes 17 times as much AV storage as that when $T = 0$.

Based on (1) and (18), Fig. 8 plots δ against T and K. We observe that δ decreases as T increases. For $T > 27/\mu$, the effect of T on δ is negligible. When $T \to \infty$, all AVs are utilized for the UARs, and $\delta = 1/K$. For the same T value, it is obvious that δ increases as K decreases. Consider the case $K = 30$ and $T = 27/\mu$, the ADR traffic decreases 24.9% as compared with $K = 30$ and $T = 0$. Fig. 7 and Fig. 8 indicate the relation between the storage usage and the ADR traffic, and provide the guidelines for the mobile operators to configure the RT timer. For example, if the operator sets $K = 10$ and wants to reduce 88.86% of the ADR traffic (as compared with when $K = 1$), the RT period $T = 27/\mu$ should be selected. In this case, the SGSN utilizes 17 times the AV storage as that when $T = 0$.

Fig. 9 shows the effects for the variance of the SGSN residence times. The Gamma distribution with mean $1/\mu$ and variance V_s is considered for SGSN residence times because it has been shown that the distribution of any positive random variable can be approximated by a mixture of Gamma distributions (see Lemma 3.9 in [4]). Following the past experience [5]–[7], we can measure the SGSN residence times in a real mobile network, and the measured data can be approximated by a Gamma distribution as the input to our simulation model. Fig. 9 shows the effect of variance V_s for the SGSN residence time distribution on the system performance. When $V_s < 2.5 \times 10^5/\mu^2$, the impact of V_s on α and δ is insignificant, and β increases as V_s increases. For $V_s > 2.5 \times 10^5/\mu^2$, as V_s increases, α significantly increases, β significantly decreases, and δ insignificantly decreases. This phenomenon is explained as follows. As V_s increases, more short and long SGSN residence times are observed, and the increase of the number of short SGSN residence times is more significant than that of long SGSN residence times. Since t_r is composed of SGSN residence times, the increase of short t_r is also more significant than that of long t_r. Short t_r results in large α value, and the SGSN consumes less AV storage after the MS leaves the SGSN area (i.e. small β value is expected). Moreover, as α increases, more stored AVs are used for UARs, and the number



(a) Effect of V_s on α. (b) Effect of V_s on β. (c) Effect of V_s on δ.

Fig. 9. Effects of V_s ($\lambda = 20\mu, T = 27/\mu$, and $K = 10$)

of ADRs decreases. Therefore, the AV usage mechanism has better performance when the variance of SGSN residence times becomes large.

V. Conclusions

In UMTS, when an MS leaves a SGSN area, the SGSN may keep the unused AVs for an interval called the RT period. If the MS returns to the SGSN area within the RT period, the SGSN uses these stored AVs for mutual authentication instead of obtaining new AVs from the HSS/AuC. This AV usage mechanism reduces the signaling traffic between the SGSN and the HSS/AuC. On the other hand, this mechanism results in extra AV storage at the SGSN. In this paper, we proposed an analytic model to investigate the impact of the RT period on the system performance. Three output measures are considered: the re-entrance probability, the extra AV storage, and the ADRs traffic between the SGSN and the HSS/AuC. The analytic results were validated against the simulation experiments. Our study provides the guidelines for the mobile operators to implement the AV usage mechanism.

References

[1] Y.-B. Lin, Y.-R. Haung, A.-C. Pang, and I. Chlamtac. "All-IP approach for UMTS third generation mobile networks," *IEEE Network*, vol. 16, no. 5, pp. 8-19, 2002.

[2] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; 3G Security; Security Architecture (Release 6), Tech. Spec. 3GPP TS 33.102 V6.0.0 (2003¡V09), 2003.

[3] Y.-B. Lin, and Y. K. Chen. "Reducing authentication signaling traffic in third generation mobile network," *IEEE Trans. Wireless Commun.*, vol 2, no. 3, pp. 493-501, 2003.

[4] F. P. Kelly. *Reversibility and Stochastic Networks.* New York: John Wiley & Sons, 1979.

[5] I. Chlamtac, Y. Fang, and H. Zeng. "Call blocking analysis for PCS networks under general cell residence time," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Sep. 1999, pp. 550-554.

[6] FarEasTone Telecom, Private communication, 2003.

[7] H. Zeng, and I. Chlamtac. "Handoff traffic distribution in cellular networks," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Sep. 1999, pp. 413-417.

[8] Y. Zhang, and M. Fujise. "An improvement for authentication protocol in third-generation wireless networks," *IEEE Trans. Wireless Commun.*, vol 5, no. 9, pp. 2348-2352, 2006.

Lin-Yi Wu received the B.S., M.S., and Ph.D. degrees in Computer Science from National Chiao Tung University, Taiwan, in 1999, 2001, and 2006 respectively. She is now a senior software engineer in MediaTek Inc., Taiwan. Her research interests include wireless metro area network, personal communications services, voice over IP, and network security.

Yi-Bing Lin (M'95-SM'95-F'03) is Chair Professor of Computer Science, National Chiao Tung University. His current research interests include wireless communications and mobile computing. Dr. Lin has published over 210 journal articles and more than 200 conference papers. Dr. Lin is the author of the book *Wireless and Mobile Network Architecture* (co-author with Imrich Chlamtac; published by John Wiley & Sons) and the book *Wireless and Mobile All-IP Networks* (co-author with Ai-Chun Pang; published by John Wiley & Sons). Dr. Lin is an IEEE Fellow, an ACM Fellow, an AAAS Fellow, and an IET(IEE) Fellow.