A Quantitative Study of Authentication and QoS in Wireless IP Networks

Wei Liang Wenye Wang
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695-7911

Abstract— **With the increasing demand for secure and high-quality communications in public access wireless IP networks, it is very important to have an in-depth understanding of the relationship between the security and quality of service (QoS). In wireless networks, authentication can provide secure communications by preventing unauthorized usage and negotiating the credentials for data transmission. Nevertheless, it induces heavy overhead and delay to data transmission, further deteriorating overall system performance. Therefore, we analyze the impact of authentication on the security and QoS quantitatively in this paper. First, we introduce a system model based on a challenge/response authentication, which is widely used in various mobile environments. Then, a concept of security level is proposed to describe the protection of communications, which is classified with regard to the nature of security, i.e., information secrecy, data integrity, and resource availability. By taking traffic and mobility patterns into account, our approach establishes a direct and quantitative connection between the security and QoS through the authentication. Finally, the numerical results are provided to demonstrate the impact of security levels, mobility and traffic patterns on overall system performance in terms of authentication cost, delay, and call dropping probability.**

Key Words: *Wireless IP networks, challenge/response authentication, security association, performance analysis.*

I. Introduction

The tremendous advance of wireless communication technologies has facilitated the ubiquitous Internet service, whereas inducing more challenges to security due to open medium [1]. In order to provide security services in wireless IP networks, *authentication* is used as an initial process to authorize a mobile user (MU) for communication through secret credentials [2]. In an authentication process, an MU is required to submit secret materials such as certificates and challenge/response values for verification with a security association (SA), which is a relationship that affords security services with parameters such as session keys between the MU and its authenticator etc [3–6]. With authentication process, the network resource can be maintained by authenticating legitimate users. The information secrecy and data integrity can also be guaranteed by using the negotiated secret credentials for encryption and message authentication.

Meanwhile, authentication also affects the quality of service (QoS) greatly. When public/private-key based authentication mechanism is applied, the computation complexity of encrypting/decrypting data consumes more time and power [7]. As for secret key based challenge/response authentication mechanism, due to lack of end-to-end SA, the credentials of the MU are encrypted and transmitted for remote verification hop-by-hop among authentication servers [8–10]. The transmission and encryption/decryption of credentials affect many QoS parameters such as authentication cost in terms of signaling and encryption/decryption cost and authentication delay, which further affect other parameters such as call dropping probability and throughput. Therefore, in some scenario, a tradeoff between security service and system performance should be considered because different users have different preference between security and performance.

Moreover, the arrival rate of authentication requests is tightly related with the mobility and traffic patterns of the MU, which may cause great impact on QoS parameters such as aggregated authentication cost in different scenarios, because the cost needs to be calculated by adding up the costs in all of the authentication requests. Therefore, the impact of authentication on QoS parameters are far more sophisticated with different mobility and traffic patterns in different scenarios.

Since the authentication affects both of security and QoS, many authentication schemes are proposed, focusing on the security and efficiency [2, 5, 8–15]. However, none of them provide quantitative analysis of security and system performance, simultaneously, and nor do they show the connection between security and system performance. Furthermore, mobility and traffic patterns are not considered, which are important features in wireless networks. Therefore, new authentication solutions may not be adapted to mobile environments with the concerns of security, performance, mobility and traffic patterns.

In this paper, we investigate the performance of authentication in wireless IP networks. First, a system model is proposed to analyze the challenge/response based authentication, which is highly consistent with various wireless IP networks such as Mobile IP and wireless local area network (WLAN). This consistency guarantees that our analytical results are applicable in real mobile environments. A concept of security level, which is classified with the nature of security, i.e., information secrecy, data integrity, and resource availability, is introduced to indicate the protection of communications. Moreover, we analyze authentication cost, delay and call dropping probability at different security levels in combination with *mobility and traffic patterns*, which builds a solid ground for understanding the impact of authentication on security and QoS with the concern of adaptation to various mobile environments.

The rest of our paper is organized as follows. In Section II, we discuss the effect of authentication on security and QoS based on challenge/response authentication. In Section III, we describe a system model and define metrics used for performance evaluation in the paper. We analyze these metrics at different security levels based on the mobility and traffic patterns in Section IV. Then, we provide the numerical results of our analysis on authentication cost, delay and call dropping probability in Section V. Finally, we draw a conclusion in Section VI.

II. Effect of Authentication on Security and QoS

In this section, we introduce the challenge/response authentication, which is widely used in wireless networks, and describe the effects of authentication on security and QoS with challenge/response authentication.

A. Overview of Challenge/Response Authentication

The authentication in wireless networks is a process to identify MUs and to negotiate SAs for communications. An SA is a trust relationship with many parameters, such as keys and algorithms, for secure service with cryptographic techniques [16].

In a challenge/response-based authentication, a user is identified with shared SA by an authentication server that sends a *challenge value*, a random number, to the user for encryption, and verifies the returned value, called *response value*, with decryption. In a foreign network, a visiting mobile user (MU) sends an authentication request to an access point (AP), which is a function unit to transmitting data. The AP relays the request to a local authentication server (LAS), which only takes charge of authentication for visiting MUs from foreign networks. If the LAS has no information to verify the MU, it contacts the HAS of the MU through an authentication architecture. An HAS is an authentication server to identify the MUs who subscribe the service in its network. And, an authentication architecture is composed of many authentication servers that share SAs with the LAS and home authentication server (HAS). If the request is an inter-domain authentication request, the HAS sends a registration request to the MU's home agent (HA), which is a router in the home network that maintains the current location of the MU, to update the MU's location.

Throughout this paper, we assume that an MU is roaming in a foreign network domain. Then, the challenge/response authentication for an MU in a foreign network domain can be categorized into three types: intra-domain handoff authentication; session authentication; and inter-domain handoff authentication.

Intra-domain handoff authentication: When an MU crosses the boundary of subnets in the foreign network domain with an on-going service, an intra-domain handoff authentication is initiated. Since there is an on-going communication session between the MU and an AP, one session SA exists between the MU and the LAS in the visiting network domain. Therefore, it is unnecessary to contact the HAS of the MU for authentication. In the case shown in Fig. 1.A, the LAS who receives the authentication request from an MU sends a challenge value to the MU. The MU encrypts the challenge value using shared SA with the LAS and replies the response value to the LAS. After decrypting the replied value and comparing it with the original challenge value, the LAS can authenticate the MU.

Session authentication: When an MU starts a communication session in a subnet of a foreign network, a session authentication is initiated. At this time, session SA does not exist between the MU and the AP, and it is necessary to contact the HAS of the MU for authentication. In the case shown in Fig. 1.B, when the LAS receives the authentication request, it sends a challenge value to the MU. The MU encrypts the challenge value with the SA shared with the HAS, and replies the response value to the LAS. The LAS must relay the challenge and response values to the HAS of the MU for verification due to lack of SA to decrypt the response value. After authentication at the HAS, the secret credentials such as keys to protect the communication may be generated and sent to the LAS.

Inter-domain handoff authentication: When an MU is crossing the boundaries of different foreign network domains with an on-going service, an inter-domain handoff authentication occurs. Due to lack of SA between the MU and the LAS, the signaling diagram shown in Fig. 1.C is similar with that in the case of session authentication, except that the MU needs registration to its home agent (HA) through the HAS because we assume that the MU needs registration only if it is crossing the boundaries of different network domains.

B. Effect of Authentication on Security

Security services are to provide information secrecy, data integrity, and resource availability for users. Information secrecy means to prevent the improper disclosure of information in the communication, while data integrity is to prevent improper modification of data and resource availability is considered to preventing improper denial of services [16].

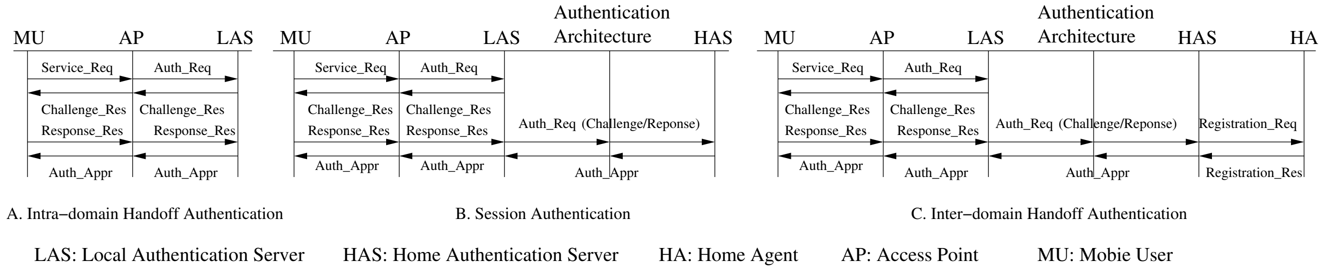
In order to provide security services in wireless networks, the challenge/response based authentication adopts several techniques. First, it requires the MU to share an SA with its HAS. The SA is unique and secret to other users. Therefore, the identification of the MU is unique, which can prevent unauthorized MUs from accessing the network resource. Second, session keys may be generated and sent to communication partners during authentication, which are used to encrypt the data of communication and provide message authentication code for data integrity check. Therefore, the authentication mechanism becomes a key role to protect the information secrecy and data integrity.

C. Effect of Authentication on QoS Metrics

Besides the effect on the security, authentication also affects the QoS metrics, such as authentication delay, cost, call dropping probability and throughput of communications.

The authentication delay is defined as the time from when the MU sends out the authentication request to when the MU receives the authentication reply. During this authentication delay, no data for on-going service can be transmitted, which may interrupt the connections. Therefore, the call dropping probability is increased with the increase of authentication delay.

The authentication cost is defined as the signaling cost and encryption/decryption load of data. In a challenge/response authentication, the challenge/response values need to be transmitted back to the HAS of the MU through the authentication architecture for verification when the LAS has no SA shared with the roaming MU. The total number of signaling messages for authenticating the MU can be large if the distance between the MU and its HAS is long. Furthermore, the signaling messages need to be encrypted and decrypted hop-by-hop for protection due to lack of direct SA between the LAS and the HAS. These multiple



A. Intra–domain Handoff Authentication B. Session Authentication C. Inter–domain Handoff Authentication

LAS: Local Authentication Server HAS: Home Authentication Server HA: Home Agent AP: Access Point MU: Mobie User

Fig. 1. Challenge/Response Authentication in Public Wireless Access Networks.

operations increase the processing load of the networks. Moreover, the mobility and traffic patterns of MUs make the authentication happen frequently in different scenarios when an MU starts a communication session or crosses boundaries of subnets with an on-going service.

The throughput of the data communication is defined as the the effective data transmitted in a unit time. It can be greatly decreased due to authentication because of several reasons. First, the authentication delay causes a temporary pause for data transmission, which decreases the throughput. Second, the key size and algorithms negotiated in authentication to encrypt and decrypt the data affect the processing time, and the attachment of message authentication code for data integrity check will affect the payload of messages.

In summary, the authentication in wireless networks has great effects on both security and QoS metrics such as authentication delay, cost, and throughput. In order to improve the security and performance of wireless networks, it is necessary to analyze the authentication effects on both security and QoS metrics by taking into account the mobility and traffic patterns of the MU. To this end, we propose a system model with assumptions and definitions of performance metrics in next section.

III. System Model and Metrics

In this section, we introduce a system model to analyze the impact of challenge/response authentication in wireless networks. We consider the security and QoS metrics as system performance, in which the security is defined with regard to security levels, and the QoS metrics are evaluated with authentication cost, delay and call dropping probability.

A. System Model

We consider a generic system model for wireless networks from two aspects. One aspect is to describe the authentication interaction between autonomous wireless networks; the other is to illustrate the authentication within a wireless network.

The system model to describe the authentication interaction between inter-connected wireless networks is shown in Fig. 2. In this model, there are a number of n autonomous wireless networks. Each network domain has an LAS and an HAS, which are central authentication servers in a network domain. The trust relationships between these LASs and HASs are maintained through an authentication architecture [14]. The functions of the LAS, HAS, and authentication architecture are introduced in II-A. We assume that the LAS and HAS are integrated together, and the authentication architecture shares an SA with the LAS/HAS of a network domain.

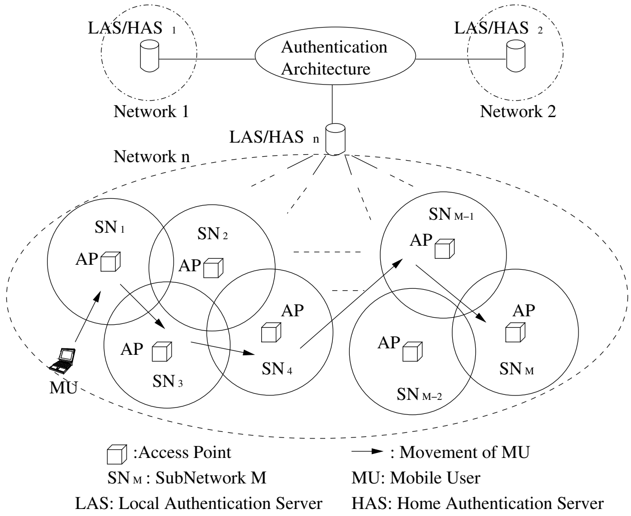


Fig. 2. System Model of Authentication between Wireless Network Domains.

The system model of authentication within a wireless network domain is demonstrated in Fig. 2. We assume that a network domain is composed of M subnets of equal size, and each subnet is controlled by an AP. An LAS/HAS shares SAs with M APs and controls the authentication requests from them.

The generic system model in our paper is consistent with many practical wireless networks such as the authentication, authorization, and accounting (AAA) architecture in Mobile IP networks and wireless local area networks (WLAN) [14]. It provides an authentication environment for MUs roaming among wireless networks. Based on this system model, in order to evaluate the performance of authentication, we need to describe specific conditions such as authentication mechanism, mobility and traffic patterns clearly .

Scenario: Assume that the challenge/response authentication is implemented on the generic system model with signaling diagrams shown in Fig. 1. Since *our initial assumption is that an MU is roaming in foreign network domains*, the intra-domain handoff authentication, session authentication, and inter-domain handoff authentication in foreign networks are illustrated in Fig. 1.A, 1.B, and 1.C, respectively.

Mobility pattern: The mobility pattern of an MU in our paper is represented with the residence time of the MU in one subnet, denoted as T_r. We assume that T_r is a random variable and the probability density function (PDF) of T_r, denoted as $f_{T_r}(t)$, is Gamma distribution with mean $1/\mu_r$ and variance V [17]. Then, the Laplace transform of $f_{T_r}(t)$, $F_r(s)$, is:

$$F_r(s) = (\frac{\mu_r \gamma}{s + \mu_r \gamma})^{\gamma}, \quad \text{where} \quad \gamma = \frac{1}{V\mu_r^2}. \tag{1}$$

Furthermore, if the number of subnets passed by an MU is assumed to be uniformly distributed between $[1, M]$, the PDF of the residence time in a network domain, denoted as $f_{T_M}(t)$, can be expressed with a Laplace transform $F_M(s)$ as follows [17]:

$$F_M(s) = \frac{1}{M}(\frac{\mu_r \gamma}{s + \mu_r \gamma})^{\gamma} \frac{1 - (\frac{\mu_r \gamma}{s+\mu_r \gamma})^{\gamma M}}{1 - (\frac{\mu_r \gamma}{s+\mu_r \gamma})^{\gamma}}. \tag{2}$$

Then, the mean value of residence time in this network domain, denoted as $\overline{T}_M$, can be expressed as:

$$\overline{T}_M = -\frac{\partial F_M(s)}{\partial s}|_{s=0} = \frac{M+1}{2\mu_r}. \tag{3}$$

Traffic pattern: In this paper, we consider the call arrival rate and call duration time of the MU as the traffic patterns of the MU. We assume that the call arrival rate of the MU, which includes the incoming calls and outcoming calls, is Poisson process with average rate λ_u, and a call duration time, denoted as T_D, has an exponential distribution with mean value $1/\eta$. Then, the PDFs of the call inter-arrival time and call duration time, denoted as $f_{T_A}(t)$ and $f_{T_D}(t)$, respectively, become:

$$f_{T_A}(t) = \lambda_u e^{-\lambda_u t}, \quad and \quad f_{T_D}(t) = \eta e^{-\eta t}. \tag{4}$$

Based on these assumptions on the mobility and traffic patterns of the MU, we evaluate the security and QoS metrics of authentication when the MU is roaming in our generic system model. The security and QoS metrics needed for evaluation are defined in next section.

B. Performance Metrics

We categorize the performance metrics into security and QoS parameters. The security parameter is represented by security levels, at which different levels of protection are provided. Meanwhile, we consider authentication cost, delay and call dropping probability as the system performance for evaluation.

B.1 Security Levels

There are much quantitative analysis of QoS in networks [18, 19], whereas less analysis of security exists. This gap between the QoS and security analysis demands quantization of security for the engineering research. Therefore, the concept of *security level* becomes widely used for security evaluation [20, 21]. The classification of security levels in these papers is either based on the information sensitivity, or based on the key length. In the classification with the information sensitivity, if a group of users are allowed to access most sensitive data, and the data in this group is prohibited to expose to other groups, thus, the security level of this group is the highest. In the classification with key length, if an encryption/decryption process is using a key length longer than other processes, this process has higher security level. As we can see, however, all of them do not consider the nature of security, i.e., data integrity, secrecy, and availability. Therefore, we argue that the nature of security should become the standard to classify the security levels.

In our paper, the *security level* is to indicate the level of protection provided by the authentication for quantitative analysis of security. The classification of security levels is shown in Table I according to the security functions described in Section II-B, i.e., protection for integrity, secrecy and resource availability. Because of different actions in challenge/response authentication, the protection of data integrity, secrecy, and availability may be different at different security levels.

- *Security Level 1*: Any MUs can send data through an AP without authentication. When an AP receives an authentication request, it checks the resources for this request. In intra-domain and session authentication shown in Fig. 3, if the resources for this service is available, the resource approval is replied to the MU to authorize the service. The signaling diagram for inter-domain handoff authentication is very similar with Fig. 1.C. The difference is that the LAS needs to send registration message to the HA and the HAS of the MU through the authentication architecture, instead of replying a challenge value to the MU. After registration, the service is authorized to the MU. At security level 1, the integrity, secrecy, and resource availability cannot be protected without cryptographic techniques.

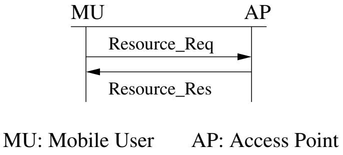


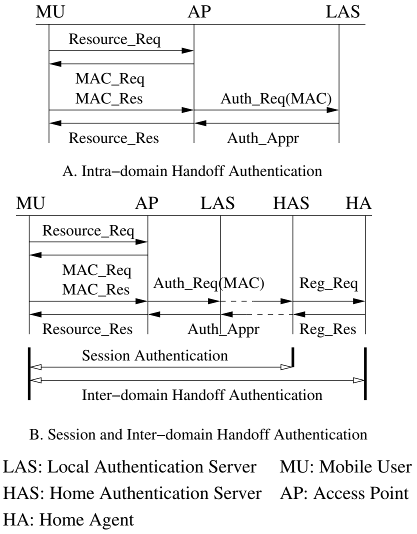
Fig. 3. Intra-domain Handoff and Session Authentication at Security Level 1.

- *Security Level 2*: Authentication is implemented with Media access control (MAC) address and no keys are generated for the subsequent communication. In this case, when an AP receives an authentication request, it requests the MAC address of the MU and relays the MAC address to LAS or HAS for verification, as shown in Fig. 4. For intra-domain handoff authentication, the LAS has the session SA of the MU, thus, the MU can be verified at LAS. For inter-domain and session authentication, the MU needs to be authenticated at HAS because there is no SA between the MU and the LAS. In particular, registration is required during inter-domain authentication. At security level 2, there is no protection available for data integrity and secrecy without keys distributed to the MU. But the network resource is slightly protected by identifying the MAC address although the MAC address can be forged easily.

- *Security Level 3*: Authentication is implemented with shared SA, and no keys are generated for the MU's communication. In this case, an SA between the MU and its HAS is used for inter-domain handoff and session authentication as shown in Fig. 5. The signaling process at security level 3 is almost the same as that at security level 2. The difference is that a pair of challenge/response values is used to authenticate the MU instead of the MAC address. The MU that receives a challenge value from the LAS encrypts the challenge value with corresponding SA. In the intra-domain handoff authentication, the corresponding SA is shared with the LAS during communication session. Then, the LAS verifies the challenge value replied from the MU by decrypting the response value with the SA. However, in inter-domain handoff and session authentication, the corresponding SA is shared with the HAS because there is no SA between the

TABLE I
SECURITY LEVEL CLASSIFICATION

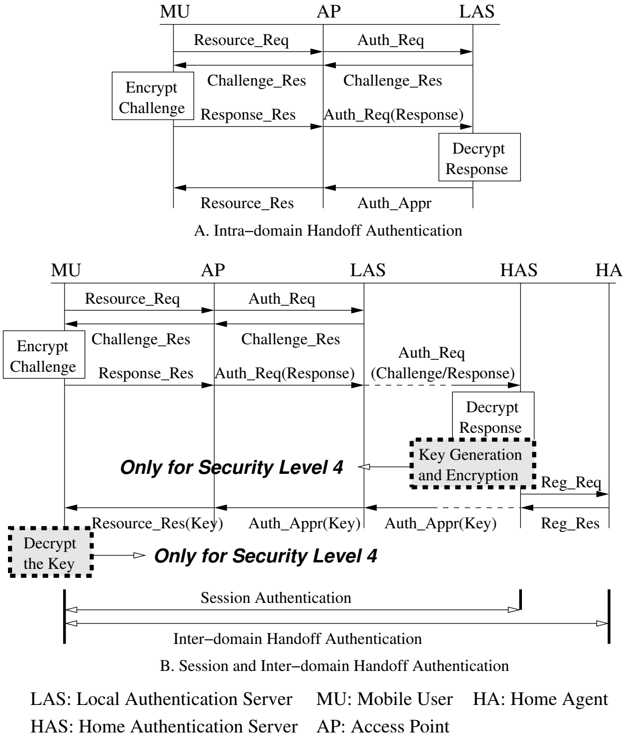
Security Level i	Security Service			
	Integrity	Secrecy	Confidentiality	Availability Protection
1	No	No	No	No
2	No	No	Low	Low
3	No	No	Medium	Medium
4	Yes	Yes	High	High



A. Intra−domain Handoff Authentication

B. Session and Inter−domain Handoff Authentication

Fig. 4. Signaling Diagram at Security Level 2.



A. Intra−domain Handoff Authentication

B. Session and Inter−domain Handoff Authentication

Fig. 5. Signaling Diagram at Security Levels 3 and 4.

MU and the LAS now. After verifying the MU at the HAS, the authentication approval is sent back to the LAS. Specially, the registration is required during inter-domain authentication. At security level 3, the network resources can be protected by only allowing the access of legitimate users. However, since no key is distributed for data transmission, the integrity and secrecy are not guaranteed. Furthermore, the network resource may be compromised due to lack of data integrity and secrecy.

- *Security Level 4*: Authentication is implemented with shared SA, and keys are generated for data communication. The signaling diagram at security level 4 is shown in Fig. 5, which is similar with that at security level 3. The difference is that keys are generated and transmitted to communication partners such as the MU, home and foreign agents. The keys are decrypted at the end users, and will be used to provide encryption and message authentication code to protect the communication. Therefore, the integrity of data can be guaranteed by message integrity check, and the secrecy is protected with data encryption. The network resource is also protected with strict identification.

From Table I and description above, we can see that the higher the security level, the better security services the authentication provides. However, the higher security levels are achieved by applying more complicated cryptographic techniques in the authentication process. The extra actions induce overhead that affects the QoS metrics, such as authentication cost, delay and call dropping probability during authentication.

B.2 Average Authentication Cost

In this context, we define *authentication cost* as the sum of signaling load and processing load for cryptographic techniques during one authentication operation. And, the *average authentication cost*, $C(i)$, is defined as the sum of the authentication cost over a number of authentication requests in a unit time at security level i, which can be written as:

$$C(i) = \sum_{\beta=1}^{3} \lambda_\beta [C_\beta^{(s)}(i) + C_\beta^{(p)}(i)], \qquad (5)$$

where β is the index of authentication type. $\beta = 1$ represents an intra-domain handoff authentication, $\beta = 2$ means a session authentication, and $\beta = 3$ is an inter-domain handoff authentication. We denote $C_\beta^{(s)}(i)$ and $C_\beta^{(p)}(i)$ as the signaling load and processing load of cryptographic techniques, respectively, of an authentication with type β at security level i. The arrival rate of requests for the authentication type β is defined as λ_β, which is related with the mobility and traffic patterns of MUs.

B.3 Average Authentication Delay

We define *authentication delay* as the time from when the MU sends out an authentication request to when the MU receives the authentication reply. The *average authentication delay*, $T(i)$, is defined as the sum of an authentication delay over a number of authentication requests in a unit time at security level i. Then, $T(i)$ can be written as:

$$T(i) = \sum_{\beta=1}^{3} \lambda_\beta T_\beta(i), \tag{6}$$

where $T_\beta(i)$ is the authentication delay per operation at security level i for authentication type β, and λ_β is the arrival rate of authentication requests with type β.

B.4 Average Call Dropping Probability during Authentication

In our paper, we consider a call is dropped due to either extended authentication delay, or an authentication failure. When an extended waiting time for authentication is induced and greater than a threshold time, the connection will be broken [22]. On the other hand, even though the authentication delay is small and the MU is a valid user, an authentication failure may happen regardless of security level because of damaged credentials caused by transmission error, packet drop at queues, attack of intruders and software application failure.

Therefore, we define the *call dropping probability* as the probability that the service of an MU is dropped during one authentication operation because of either extended authentication delay, or an authentication failure. When an MU roams among subnets in a network domain, the *average call dropping probability*, $P(i)$, is defined as the ratio of the sum of the call dropping probability per authentication in a unit time over the number of authentication requests sent by the MU within unit time at security level i. Let $P(i)$ denote the average call dropping probability at security level i, $P(i)$ can be written as:

$$\begin{aligned} P(i) &= \frac{\sum_{\beta=1}^{3} \lambda_\beta [P_\beta(i) + P_e]}{\sum_{\beta=1}^{3} \lambda_\beta}, \\ and \quad P_\beta(i) &= P_{T_\beta(i)}(T_\beta(i) > T_{th}), \end{aligned} \tag{7}$$

where T_{th} is a threshold time, $P_{T_\beta(i)}(T_\beta(i) > T_{th})$ is the probability that an authentication delay is greater than T_{th} in authentication type β. P_e is the probability that one authentication fails due to unknown effects. Since the factors that affect P_e include many unknown factors and there is no evidence on the pattern of attacks currently, we will use a mean value from experiments to represent P_e in the numeric results of our paper [23].

In summary, in order to evaluate $C(i)$, $T(i)$ and $P(i)$ in (5)$\sim$ (7), we need to analyze λ_β, $C_\beta^{(s)}(i)$, $C_\beta^{(p)}(i)$, $T_\beta(i)$, and $P_\beta(i)$. Next, we derive these parameters based on the system model shown in Fig. 2, assumptions described in Section III-A, and the definitions of the performance metrics in Section III-B.

IV. Performance Analysis

In this section, we analyze the impact of authentication on security and QoS metrics in terms of authentication cost, delay and call dropping probability from two key aspects. First is to observe the relationship between the security level and the QoS metrics for each authentication. Second is to obtain the interrelationship between the average QoS metrics during authentication and traffic load among networks by evaluating the total arrival rates of authentication requests.

A. *Performance Analysis per Authentication*

At different security levels, the authentication has different effects on the cost, delay and call dropping probability.

A.1 Authentication Cost per Operation

The authentication cost, $C_\beta(i)$, $(\beta = 1, 2, 3 \ and \ i = 1, 2, 3, 4)$, is composed of $C_\beta^{(s)}(i)$ and $C_\beta^{(p)}(i)$, which depend on the authentication type β and security level i. For convenient analysis, we define a set of cost parameters in Table II.

TABLE II
Authentication Cost Parameters

Symbol	Description
c_s	Transmission cost on one hop
c_p	Encryption/decryption cost on one hop
c_v	Verification cost at an authentication server
c_{us}	Encryption/decryption cost for a session key
c_g	Key generation cost
c_{ts}	Transmission cost for a session key to other communication identities
c_{rg}	Registration cost

Then, the transmission costs, $C_\beta^{(s)}(i)$, can be derived from the signaling diagrams in Figs. 3, 4, and 5, respectively, as follows:

$$C_\beta^{(s)}(i) = a_{\beta,i} c_s, \quad \forall \beta = 1, 2, 3 \text{ and } i = 1, 2, 3, 4, \tag{8}$$

where $a_{\beta,i}$ is an element of matrix A, indicating the number of hops by which the authentication process passes for authentication type β at security level i. For example, when $\beta = 3$ and $i = 4$, $a_{3,4} = 2(N_h + 3)$ denotes the number of hops that the authentication signalings pass when $\beta = 3$, $i = 4$, which can be obtained from Fig. 5.B. Thus, we obtain A as:

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & 8 & 8 \\ 2 & 2(N_h+1) & 2(N_h+2) & 2(N_h+2) \\ 2(N_h+1) & 2(N_h+2) & 2(N_h+3) & 2(N_h+3) \end{bmatrix}, \tag{9}$$

where β and i represent the row and column of A, respectively. N_h is the number of the hops between the MU and its HAS.

Similar with the analysis in (8), according to the signaling diagrams in Fig. 3, 4, and 5, $C_\beta^{(p)}(i)$ can be written as:

$$C_\beta^{(p)}(i) = \vec{\mathbf{b}}_{\beta,i} \cdot \vec{\mathbf{x}}_p, \quad \forall \beta = 1, 2, 3 \text{ and } i = 1, 2, 3, 4. \tag{10}$$

Here, $\vec{\mathbf{x}}_p$ is a vector defined as:

$$\vec{\mathbf{x}}_p^T = [c_p,\ c_v,\ c_{us},\ c_g,\ c_{ts},\ c_{rg}], \tag{11}$$

where all of the cost parameters are defined in Table II. And, $\vec{\mathbf{b}}_{\beta,i}$ are also vectors determined by:

$$\begin{aligned} &\vec{\mathbf{b}}_{1,1} = \vec{b}_{2,1} = [0,\ 0,\ 0,\ 0,\ 0,\ 0], \\ &\vec{\mathbf{b}}_{1,2} = [2,\ 1,\ 0,\ 0,\ 0,\ 0], \\ &\vec{\mathbf{b}}_{1,3} = \vec{b}_{1,4} = [4,\ 1,\ 1,\ 0,\ 0,\ 0], \\ &\vec{\mathbf{b}}_{2,2} = [2(N_h-1),\ 1,\ 0,\ 0,\ 0,\ 0], \\ &\vec{\mathbf{b}}_{2,3} = [2N_h,\ 1,\ 1,\ 0,\ 0,\ 0], \\ &\vec{\mathbf{b}}_{2,4} = [2N_h,\ 1,\ 2,\ 1,\ 1,\ 0], \\ &\vec{\mathbf{b}}_{3,1} = [0,\ 0,\ 0,\ 0,\ 0,\ 1], \\ &\vec{\mathbf{b}}_{3,2} = [2N_h,\ 1,\ 0,\ 0,\ 0,\ 1], \\ &\vec{\mathbf{b}}_{3,3} = [2(N_h+1),\ 1,\ 1,\ 0,\ 0,\ 1], \\ &\vec{\mathbf{b}}_{3,4} = [2(N_h+1),\ 1,\ 2,\ 1,\ 1,\ 1]. \end{aligned} \tag{12}$$

The coefficients in front of the cost variables in $\vec{x}_p$ denote the number of the costs we should consider during one authentication. In the case of $\beta = 3$ and $i = 4$, no encryption/decryption is applied between the MU and the AP during authentication. Therefore, the number of c_p is $2(N_h+1)$ since N_h is the number of hops between the MU and its HAS. At this time, one registration at the HA, one credential verification and key generation at the HAS, one decryption for the credentials at the HAS, one decryption of the key at the MU, and the credential transmission to the communication partner of the MU are needed. Therefore, the coefficients of c_v, c_{us}, c_g, c_{ts}, and c_{rg} are 1, 2, 1, 1, and 1, respectively. The derivation of other coefficients in different cases is the same as this analysis.

A.2 Delay per Authentication

To derive the delay for different types of authentications in different security levels, we use the same signaling diagrams shown in Figs. 3, 4, and 5. We also define a set of time parameters shown in Table III for convenient description.

TABLE III
AUTHENTICATION COST PARAMETERS

Symbol	Description
T_{pr}	Message propagation time on one hop
T_{tr}	Message transmission time on one hop
T_{ed}	Message encryption/decryption time on one hop
T_a	Authentication request service & waiting time at the AP
T_{sg}	Authentication request service & waiting time at the proxy authentication server
T_v	Authentication request service and waiting time at the HAS
T_{us}	Key encryption & decryption time
T_g	Key generation time at the HAS
T_{ts}	Transmission time for the session key to the other communication identities such as HA
T_{rg}	Registration request service and waiting time at the HA

Then, $T_\beta(i)$ can be expressed as:

$$T_\beta(i) = \vec{\mathbf{d}}_{\beta,i} \cdot \vec{\mathbf{x}}_t, \quad \forall \beta = 1, 2, 3 \text{ and } i = 1, 2, 3, 4. \tag{13}$$

Here, $\vec{\mathbf{x}}_t$ is a vector defined as:

$$\vec{\mathbf{x}}_t^T = [T_{pr} + T_{tr},\ T_{ed},\ T_a,\ T_{sq},\ T_v,\ T_{us},\ T_g,\ T_{ts},\ T_{rg}], \tag{14}$$

where all the time components are defined in Table III. And, $\vec{\mathbf{d}}_{\beta,i}$ are the vectors defined as follows:

$$\begin{aligned}
&\vec{\mathbf{d}}_{1,1} = [2,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{1,2} = [6,\ 2,\ 3,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{1,3} = d_{1,4} = [8,\ 4,\ 4,\ 0,\ 2,\ 1,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{2,1} = [2,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{2,2} = [2(N_h+1),\ 2(N_h-1),\ 3,\ 2(N_h-2),\ 1,\ 0,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{2,3} = [2(N_h+2),\ 2N_h,\ 4,\ 2(N_h-2),\ 1,\ 1,\ 0,\ 0,\ 0],\\
&\vec{\mathbf{d}}_{2,4} = [2(N_h+2),\ 2N_h,\ 4,\ 2(N_h-2),\ 1,\ 2,\ 1,\ 1,\ 0],\\
&\vec{\mathbf{d}}_{3,1} = [2(N_h+1),\ 0,\ 2,\ 2(N_h-1),\ 0,\ 0,\ 0,\ 0,\ 1],\\
&\vec{\mathbf{d}}_{3,2} = [2(N_h+2),\ 2N_h,\ 3,\ 2(N_h-2),\ 2,\ 0,\ 0,\ 0,\ 1],\\
&\vec{\mathbf{d}}_{3,3} = [2(N_h+3),\ 2(N_h+1),\ 4,\ 2(N_h-2),\ 2,\ 1,\ 0,\ 0,\ 1],\\
&\vec{\mathbf{d}}_{3,4} = [2(N_h+3),\ 2(N_h+1),\ 4,\ 2(N_h-2),\ 2,\ 2,\ 1,\ 1,\ 1].
\end{aligned} \tag{15}$$

The coefficients in front of the time variables in $\vec{x}_t$ denote the number of time variables we should consider in the authentication case. For example, in the case of $\beta = 3$ and $i = 4$ as shown in Fig 5, the number of hops that the signaling messages pass is $2(N_h + 3)$, and the number of hops that we should consider the encryption/decryption is $2(N_h + 1)$. Since the authentication process in the case of $\beta = 3$ and $i = 4$ needs to pass the AP four times, the intermediate authentication server $2(N_h - 2)$ times, the HAS twice, the coefficients of T_a, T_{sg} and T_v are 4, $2(N_h - 2)$, and 2, respectively. And because a registration at the HA, a key generation, transmission process, and twice key encryption/decryption at the HAS and the MU are needed, the coefficients for T_{rg}, T_g, T_{ts}, and T_{us} are 1, 1, 1, and 2, respectively. The other cases of authentication share the same analysis.

A.3 Call Dropping Probability

In Section III-B.4, we consider a call is dropped during authentication if the waiting time for authentication is greater than a threshold value T_{th}, or an authentication failure happens. In (7), we use a mean value from an experiment for P_e, due to the unknown distribution model. Therefore, to evaluate $P_\beta(i)$, $(\beta = 1, 2, 3\ and\ i = 1, 2, 3, 4)$, the authentication delay shown in (13) is critical.

In (13), we only consider the time variables, T_{sq}, T_a, T_v, and T_{rg}, as the random variables because the variance of the other time variables are small. Thus, to find $P_\beta(i)$ becomes to find the PDFs of the different combinations of T_{sq}, T_a, T_v, and T_{rg} in $T_\beta(i)$. If we assume that:(1) $M/M/1$ queues are applied at APs, authentication servers, and HAs; (2) The PDFs of T_{sq}, T_a, T_v, and T_{rg} are independent identical distribution (iid), then the PDF of T_{sq}, T_a, T_v, and T_{rg}, i.e., $w(t)$, can be shown as [24]:

$$w(t) = (\mu_s - \lambda_s)e^{-(\mu_s - \lambda_s)t}, \tag{16}$$

where μ_s and λ_s are the service and arrival rates of authentication requests, respectively. Furthermore, the PDFs of the different combinations of T_{sq}, T_a, T_v, and T_{rg} in $T_\beta(i)$, i.e., $f_{\beta,i}(t)$, can be expressed in (17), on next page as the components of a matrix $f(t)$. In (17), β and i represent the row and column, respectively. $\Gamma(x) \triangleq \int_0^\infty s^{x-1}e^{-s}ds$, and $\xi = \mu_s - \lambda_s$. With these PDFs, $P_\beta(i)$ can be obtained in different cases.

To summarize, we have obtained authentication cost, delay, and call dropping probability for one authentication operation. However, in order to obtain the average authentication cost, delay, and call dropping probability defined in (5), (6), and (7), we still need to evaluate the arrival rates of different types of authentication requests, that is, λ_β, $(\beta = 1, 2, 3)$.

B. Arrival Rates of Authentication Requests

Since the authentication requests are categorized into three types: intra-domain handoff authentication, session authentication, and inter-domain handoff authentication, we analyze the arrival rates of different types of authentication, i.e., λ_β, $(\beta = 1, 2, 3)$, based on the mobility and traffic patterns of the MUs.

B.1 Arrival Rate of Intra-Domain Handoff Authentication, λ_1

The intra-domain handoff authentication happens when an MU crosses the boundaries of subnets inside a network domain with an on-going service. In order to calculate the arrival rate of intra-domain handoff authentication requests, we categorize the calls into four types that happen in four events:

- Y_1 is the event that an MU starts a connection before entering the network domain, enters the network domain with the

$$f(t) = \begin{bmatrix} \xi e^{-\xi t} & \frac{\xi(\xi t)^3 e^{-\xi t}}{\Gamma(4)} & \frac{\xi(\xi t)^5 e^{-\xi t}}{\Gamma(6)} & \frac{\xi(\xi t)^5 e^{-\xi t}}{\Gamma(6)} \\ \xi e^{-\xi t} & \frac{\xi(\xi t)^{2N_h-1} e^{-\xi t}}{\Gamma(2N_h)} & \frac{\xi(\xi t)^{2N_h} e^{-\xi t}}{\Gamma(2N_h+1)} & \frac{\xi(\xi t)^{2N_h} e^{-\xi t}}{\Gamma(2N_h+1)} \\ \frac{\xi(\xi t)^{2N_h} e^{-\xi t}}{\Gamma(2N_h+1)} & \frac{\xi(\xi t)^{2N_h+1} e^{-\xi t}}{\Gamma(2N_h+2)} & \frac{\xi(\xi t)^{2N_h+2} e^{-\xi t}}{\Gamma(2N_h+3)} & \frac{\xi(\xi t)^{2N_h+2} e^{-\xi t}}{\Gamma(2N_h+3)} \end{bmatrix}. \tag{17}$$

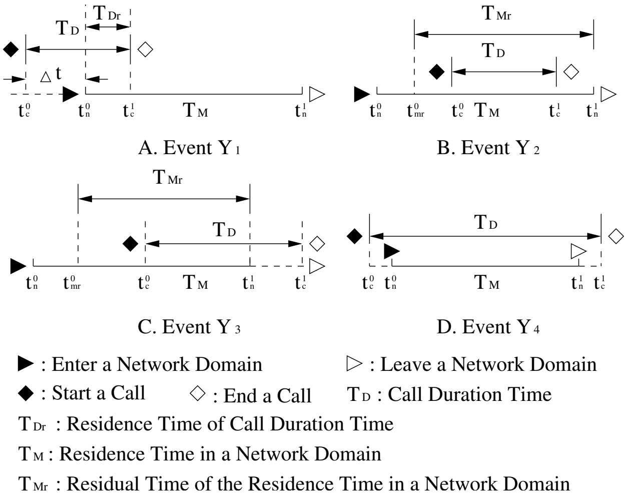


Fig. 6. Time Diagrams of Events.

on-going connection and this connection ends before the MU moves out of the network domain.

- Y_2 is the event that an MU starts a connection within current network domain and this connection ends before the MU moves out of the network domain.
- Y_3 is the event that an MU starts a connection within current network domain and this connection ends after the MU moves out of the network domain.
- Y_4 is the event that an MU starts a connection before entering the network domain, enters the network domain with the on-going connection, and the connection ends after moving out of the network domain.

Then, the arrival rate of intra-domain handoff authentication requests, λ_1, can be written as:

$$\begin{aligned} \lambda_1 = & \lambda_u P_{r1}(\lceil \overline{N}_{a1} \rceil - 1) + \lambda_u P_{r2}(\lceil \overline{N}_{a2} \rceil - 1) \\ & + \lambda_u P_{r3}(\lceil \overline{N}_{a3} \rceil - 1) + \lambda_u P_{r4}(\lceil \overline{N}_{a4} \rceil - 1), \end{aligned} \tag{18}$$

where λ_u is the call arrival rate defined in (4), P_{r1}, P_{r2}, P_{r3}, and P_{r4} are the probabilities that event Y_1, Y_2, Y_3, and Y_4 happen, respectively. $\overline{N}_{a1}$, $\overline{N}_{a2}$, $\overline{N}_{a3}$, and $\overline{N}_{a4}$ are the average numbers of subnets passed by an MU in current network domain in event Y_1, Y_2, Y_3, and Y_4, respectively. The time diagrams of these events, Y_1, Y_2, Y_3, and Y_4, are shown in Fig. 6, where t_c^0 and t_c^1 are the call beginning and ending time, respectively. t_n^0 and t_n^1 are the time when an MU enters and leaves the network domain, respectively. t_{mr}^0 is the beginning time of the residual time of the residence time in a network domain. Then, P_{r1}, P_{r2}, P_{r3}, and P_{r4} can be derived next.

According to the time diagram in Fig. 6.A, and denote $\Delta t = t_n^0 - t_c^0$, we have:

$$\begin{aligned} P_{r1} = & \textstyle\int_0^\infty P_r[I(t_c^0 + \Delta t, t_c^0) = 1] \cdot P_r(T_D > \Delta t) d(\Delta t) \\ & \cdot P_r(T_{Dr} \leq T_M), \end{aligned} \tag{19}$$

where $I(t_c^0 + \Delta t, t_c^0)$ is the number of calls that arrive in time interval $[t_c^0, t_c^0 + \Delta t)$. Since we assume that the call arrival rate is a Poisson process,

$$P_r[I(t_c^0 + \Delta t, t_c^0) = 1] = \lambda_u \Delta t e^{-\lambda_u \Delta t}, \tag{20}$$

where λ_u is the average arrival rate of the calls. In (19), T_D is the call duration time with PDF defined in (4), and T_M is the residence time of an MU in the network domain with Laplace transform of PDF in (2). Thus, we have:

$$P_r(T_D > \Delta t) = \int_{\Delta t}^\infty f_{T_D}(t) dt = e^{-\eta \Delta t}, \tag{21}$$

where $f_{T_D}(t)$ is defined in (4), $1/\eta$ is the average call holding time and $\Delta t = t_n^0 - t_c^0$.

Furthermore, T_{Dr} is the residual time of the call duration time with the same PDF as T_D defined in (4) due to the memoryless property of exponential distribution. Since we have the Laplace transform of the PDF of T_M defined in (2), $P_r(T_{Dr} \leq T_M)$ can be determined by:

$$P_r(T_{Dr} \leq T_M) = \int_0^\infty f_{X_1}(t) dt, \tag{22}$$

where $X_1 \triangleq T_M - T_{Dr}$, and $f_{X_1}(t)$ can be computed from:

$$f_{X_1}(t) = \mathscr{L}^{-1}\left\{\frac{(\eta + s)F_M(s)}{\eta}\right\}. \tag{23}$$

Here, $1/\eta$ is the average call holding time, $\eta/(\eta + s)$ is the Laplace transform of the PDF of T_{Dr}, and $F_M(s)$ is the Laplace transform of the PDF of T_M defined in (2).

Second, P_{r2} can be derived from Fig. 6.B as:

$$\begin{aligned} P_{r2} & = P_r(T_D < T_{Mr}) \cdot P_r(t_{mr}^0 \leq t_c^0 < t_{mr}^0 + T_{Mr}) \\ & = \textstyle\int_0^\infty f_{X_2}(t) dt \cdot \int_0^\infty \lambda_u t e^{-\lambda_u t} f_{Mr}(t) dt, \end{aligned} \tag{24}$$

where $X_2 \triangleq T_{Mr} - T_D$, $f_{X_2}(t)$ and $f_{Mr}(t)$ are the PDFs of X_2 and T_{Mr}, respectively, which can be obtained by:

$$\begin{aligned} & f_{X_2}(t) = \mathscr{L}^{-1}\left\{F_{Mr}(s)\tfrac{\eta+s}{\eta}\right\}, \\ & f_{Mr}(t) = \mathscr{L}^{-1}\left\{F_{Mr}(s)\right\}, \end{aligned} \tag{25}$$

where $1/\eta$ is the average call holding time, and $F_{Mr}(s)$ is the Laplace transform of the PDF of T_{Mr}, the residual time of the residence time in a network domain. $F_{Mr}(s)$ is determined by:

$$F_{Mr}(s) = \frac{1 - F_M(s)}{s\overline{T}_M}, \tag{26}$$

where $\overline{T}_M$ is defined in (3), $F_M(s)$ is defined in (2), Moreover, we can obtain P_{r3} from Fig. 6.C:

$$\begin{aligned} P_{r3} &= P_r(T_D > T_{Mr}) \cdot P_r(t_{mr}^0 \le t_c^0 < t_{mr}^0 + T_{Mr}) \\ &= \textstyle\int_0^\infty f_{X_3}(t)dt \cdot \int_0^\infty \lambda_u t e^{-\lambda_u t} f_{Mr}(t)dt, \end{aligned} \quad (27)$$

where $X_3 \triangleq T_D - T_{Mr}$, $f_{Mr}(t)$ is the PDF of T_{Mr} defined in (25), $f_{X_3}(t)$ is the PDF of X_3, which can be obtained by:

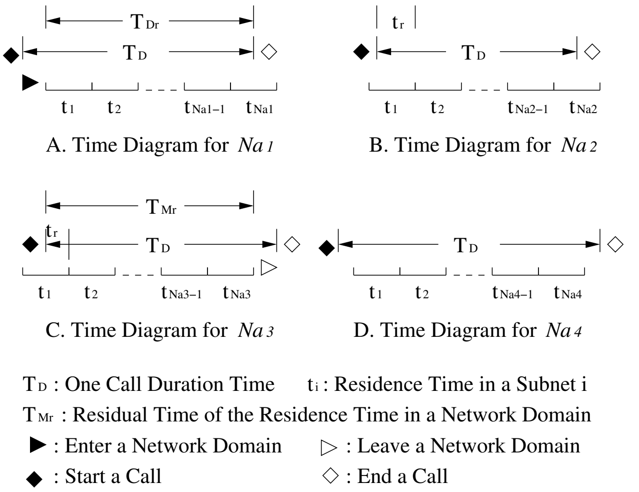
$$f_{X_3}(t) = \mathscr{L}^{-1}\left\{\frac{\eta}{(\eta+s)F_{Mr}(s)}\right\}, \quad (28)$$

where $F_{Mr}(s)$ is defined in (26), η is defined in (4).

Finally, P_{r4} can be determined from Fig. 6.D as follows:

$$\begin{aligned} P_{r4} = &\textstyle\int_0^\infty P_r[I(t_c^0 + \Delta t, t_c^0) = 1] \cdot P_r(T_D > \Delta t)d(\Delta t) \\ &\cdot P_r(T_{Dr} > T_M), \end{aligned} \quad (29)$$

where $P_r[I(t_c^0 + \Delta t, t_c^0) = 1]$ is shown in (20), $P_r(T_D > \Delta t)$ is defined in (21), and $P_r(T_{Dr} > T_M) = 1 - P_r(T_{Dr} \le T_M)$ with $P_r(T_{Dr} \le T_M)$ defined in (22).


T Dr
T D
t1
t2
tNa1−1
tNa1
tr
TD
tNa2−1
tNa2
T Mr
tNa3−1
tNa3
tNa4−1
tNa4
TD : One Call Duration Time
ti : Residence Time in a Subnet i
TMr : Residual Time of the Residence Time in a Network Domain
▶ : Enter a Network Domain
▷ : Leave a Network Domain
◆ : Start a Call
◇ : End a Call

A. Time Diagram for *Na 1*

B. Time Diagram for *Na 2*

C. Time Diagram for *Na 3*

D. Time Diagram for *Na 4*

Fig. 7. Time Diagram for Number of Subnets Passed by in One Call.

After we obtain P_{rj}, $(j = 1, 2, 3, 4)$, in order to evaluate λ_1, we need to evaluate $\overline{N}_{aj}$, $(j = 1, 2, 3, 4)$. The time diagrams to evaluate $\overline{N}_{aj}$ are shown in Fig. 7.

In order to evaluate $\overline{N}_{a1}$ and $\overline{N}_{a2}$, we consider a theorem in [25], which says that given call holding time and subnet residence time with Gamma distribution, the average number of subnets passed by an MU within a call, $\overline{K}$, is determined by:

$$\overline{K} = -\alpha \sum_{p\in\sigma_c} \text{Res}_{s=p} \frac{1 - f^*(s)}{1 - (1-p_f)f^*(s)} f_c^*(-s), \quad (30)$$

where $1/\alpha$ is the average residence time of an MU in a subnet, p_f is the probability that a handoff call is blocked, $f^*(s)$ is the Laplace transform of the PDF of the residence time of an MU in a subnet, $f_c^*(s)$ is the Laplace transform of the PDF of the call holding time of the MU, σ_c is the singular points of $f_c^*(-s)$, and $\text{Res}_{s=p}$ denotes the residue at singular point $s = p$.

In event Y_1 and Y_2, the call duration time in the network domain are T_{Dr} and T_D, respectively, which are exponential distribution, one special case of Gamma distributions. Therefore, $\overline{N}_{a1}$ and $\overline{N}_{a2}$ can be obtained with (30). By assuming that $p_f = 0$, we can carry out $\overline{N}_{a1}$ and $\overline{N}_{a2}$ as:

$$\overline{N}_{a1} = \overline{N}_{a2} = \frac{\mu_r}{\eta}, \quad (31)$$

where $1/\eta$ is the average call duration time of the MU and μ_r is the average residence time of the MU in a subnet in our paper.

On the other hand, note that T_{Mr} and T_M are not Gamma distributions, we cannot obtain $\overline{N}_{a3}$ and $\overline{N}_{a4}$ with (30). Thus, we derive $\overline{N}_{a3}$ and $\overline{N}_{a4}$ next.

From Fig. 7.C, the relationship between different time components can be written as follows:

$$T_{Mr} = t_r + \sum_{i=2}^{N_{a3}} t_i, \quad (32)$$

where T_{Mr} and t_r are the residual time of the residence time of an MU in a network domain and in a subnet, respectively. The Laplace transform of the PDF of t_r, $F_{tr}(s)$, is:

$$F_{tr}(s) = \mu_r \frac{1 - F_r(s)}{s}, \quad (33)$$

where $1/\mu_r$ is the average residence time of an MU in a subnet, $F_r(s)$ is defined in (1). In (32), t_i is the residence time of an MU in subnet i, which is assumed to be Gamma distribution with Laplace transform of PDF defined in (1), and N_{a3} is the random number of the subnets passed by an MU in current network domain in event Y_3.

Based on the relationship in (32), we can obtain:

$$F_{Mr}(s) = F_{t_r}(s) G_{N_{a3}-1}(z)|_{z=F_r(s)}, \quad (34)$$

where $F_{Mr}(s)$ is defined in (26), $F_{t_r}(s)$ is defined in (33), $G_{N_{a3}-1}(z)$ is the generating function of the PDF of $N_{a3} - 1$. Then, $\overline{N}_{a3}$ can be obtained by:

$$\begin{aligned} \overline{N}_{a3} &= \tfrac{\partial G_{N_{a3}-1}(z)}{\partial z}|_{z=1} + 1 \\ &= \tfrac{2M^2-M-1}{12\overline{T}_M \mu_r} + \tfrac{(M+1)}{4}\tfrac{(\gamma+1)}{\gamma} + 1, \end{aligned} \quad (35)$$

where $\overline{T}_M$ is defined in (3), M is the number of subnets in current network domain, $1/\mu_r$ is the average residence time that an MU stays in a subnet, and γ is defined in (1).

According to Fig.7.D, $\overline{N}_{a4}$ is equal to the average number of subnets that an MU passes in a foreign network domain. Recall that we assume that the number of subnets that the MU passes by in a network domain, N_{sn}, is uniformly distributed between 1 and M, i.e.,

$$P(N_{sn} = m) = \frac{1}{M}, \quad m = 1, 2, \cdots, M. \quad (36)$$

Here, M is the total number of subnets in current network domain. Thus, we have:

$$\overline{N}_{a4} = \overline{N}_{sn} = \sum_{j=1}^{M} \frac{j}{M} = \frac{M+1}{2}. \quad (37)$$

Now we have obtained all $\overline{N}_{aj}$ at event Y_j, $j = 1, 2, 3, 4$. Since we get P_{rj} in previous part of this subsection, we can evaluate λ_1 by substituting the values of P_{rj} and N_{aj}, $j = 1, 2, 3, 4$, into (18). Next, in order to obtain $C(i)$, $T(i)$, and $P(i)$ defined in (5), (6), (7), we need to evaluate λ_2 and λ_3.

B.2 Arrival Rate of Session Authentication, λ_2

After an MU has moved into a network domain, a session authentication is initiated whenever a call arrives. Therefore, the arrival rate of session authentication requests for one MU, e.g. λ_2, is equal to the call arrival rate of an MU,

$$\lambda_2 = \lambda_u, \tag{38}$$

where λ_u is assumed to be the call arrival rate in (4).

B.3 Arrival Rate of Inter-Domain Handoff Authentication, λ_3

The inter-domain handoff authentication requests happen when an MU enters the network domain with an on-going service. Therefore, the arrival rate of inter-domain handoff authentication requests, λ_3, can be obtained by:

$$\lambda_3 = \lambda_u(P_{r1} + P_{r4}), \tag{39}$$

where λ_u is the call arrival rate assumed in (4), P_{r1} and P_{r4} are the probabilities that events Y_1 and Y_4 occur. The events Y_1 and Y_4 are defined in IV-B.1. P_{r1} and P_{r4} are evaluated in (19) and (29), respectively.

Thus, we have obtained the arrival rates of authentication requests in the cases of intra-domain handoff authentication, session authentication, and inter-domain handoff authentication. Since two key aspects, i.e., the relationship between the security and system performance, and the relationship between the QoS metrics and traffic load, have been evaluated, the impact of authentication on security and the system performance can be observed clearly through $C(i)$, $T(i)$, and $P(i)$ in (5)$\sim$ (7).

V. Numerical Results

In this section, we evaluate the effects of mobility and traffic patterns on authentication cost, $C(i)$, delay, $T(i)$, and call dropping probability, $P(i)$, at different security levels.

A. Assumptions and Parameters

The numerical results are proposed with the assumptions introduced in Section III and IV-A.3. In Section III, we consider an MU roaming in a foreign network shown in Fig. 2. The residence time of the MU in a subnet of the network domain is assumed to be Gamma distribution with mean value $1/\mu_r$. The Call arrival rate of the MU is assumed to be Poisson process with exponentially distributed inter-arrival time with mean value $1/\lambda_u$, and the call duration time of the MU is assumed to be exponential distribution with mean value $1/\eta$.

In Section IV-A.3, we further assume that $M/M/1$ queues are used at APs, authentication servers such as LAS and HAS, and HAs with service rate μ_s and arrival rate of authentication requests λ_s. Let $\xi = \mu_s - \lambda_s$. According to (16), the service and waiting time at an AP, authentication server, and HA, e.g., T_a, T_{sq}, and T_v, become the random variables with identical exponential distribution with mean value $1/\xi$. The parameters to evaluate the authentication cost and delay are shown in Table IV.

There are many ways to decide the values for the authentication costs. For example, the authentication cost for signaling can be measured with the number of messages, and the authentication cost for encryption can be measured with the number of CPU cycles. However, the most important problem here is how to make them consistent, i.e. the values of the costs can be compared with each other in the same scale. To solve this problem, we assume that the encryption/decryption cost on one hop, c_p, and the key generation cost, c_g, are all equal to 1 because they are the lightest load compared to other costs and they have the similar operation in cryptography techniques [26, 27]. The values of other costs are determined by comparing to c_p and c_g with the time to finish the operation, i.e., we use the ratio of processing time to represent the authentication cost instead of the actual processing time. The reason is that the time needed to finish an operation represents the load of the server to complete it. However, we do not use the processing time to represent the cost directly because we do not want to confuse the authentication cost with the authentication delay and the authentication cost can be evaluated with many other ways.

TABLE IV

Parameters for Evaluation on QoS Metrics

Parameters for Authentication Cost					
c_s	c_p	c_v	c_g	c_{ts}	N_h
10	1	20	1	110	10
Parameters for Authentication Delay					
T_{th}	T_{pr}	T_{tr}	T_{ed}	T_g	M
3s	40 μs	20ms	2ms	2ms	120
Parameters for Random Variables					
λ_u	η	γ	μ_r	ξ	
0.1 min^{-1}	0.3 min^{-1}	225	1/15 min^{-1}	15 sec^{-1}	

Therefore, the values of the time variables become critical. When the maximum authentication message size is 4096 bytes [3], the transmission delay is about 20 milliseconds with the assumption of 2 Mbps link capacity [26]. The values of T_{ed} and T_g come from [26]. By assuming one network domain is about $100km^2$ with radius $6km$, the value of the propagation time, T_{pr}, can be determined and shown in Table IV.

B. Effects of Mobility Pattern at Different Security Levels

The effects of mobility pattern on the authentication cost, delay, and call dropping probability are shown in Fig. 8, 9, and 10. In these figures, we illustrate the relationships between the residence time of an MU in a subnet, authentication cost, delay, and call dropping probability, respectively.

In Fig. 8, authentication costs at different security levels decrease with the increase of the residence time of an MU in a subnet because the longer an MU stays in the subnets, the less the intra-domain handoff authentication requests. And, if the residence time of an MU approaches to infinity, the authentication cost will be stable on the session authentication cost because only session authentication exists in this case. Moreover, we can see that the security levels have different effects on the cost at the same residence time in a subnet. The higher the security level, the more the authentication cost because higher security levels impose more effort to provide secure services. For example, if we degrade the security level from 4 to 3, the authentication cost can be reduced up to 32%.

Fig. 9 illustrates the effect of residence time on the authentication delay. As we can see, the authentication delay decreases with the increase of the residence time of an MU in a subnet.

Similar with the authentication cost, this trend is due to the decrease in the intra-domain handoff authentication requests. And, the higher security level causes more authentication delay because of more operations needed for more secure services. The improvement of authentication delay by changing security level from 4 to 3 is around 0.1 seconds.

The effect of call dropping probability in authentication is shown in Fig. 10. The call dropping probability increases with the increase of the residence time of an MU in a subnet. When the residence time of an MU in a subnet increases, the arrival rate of intra-domain handoff authentication requests will decrease. Then, the session authentication requests become the major part of the authentication requests. Note that the call dropping probability for session authentication is far more than that in intra-domain handoff authentication due to the longer authentication delay caused by the remote authentication. The call dropping probability will approximate the call dropping probability in session authentication if the residence time of an MU goes to infinity. In other words, the upper bound of the call dropping probability can be achieved when the authentication requests are all *session* authentication requests. Similar with the cost and delay, call dropping probability is greatly affected by the securit level. When the security level increases from 3 to 4, call dropping probability increases about 45%.

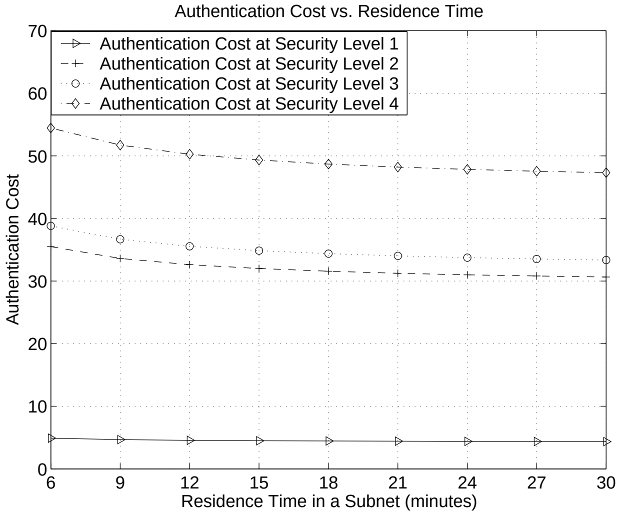


Fig. 8. Authentication Cost vs. Residence Time in a Subnet.

C. Effect of Traffic Load at Different Security Levels

The effects of traffic pattern on the authentication cost, delay, and call dropping probability at different security levels are demonstrated in Figs. 11, 12, and 13.

Figs. 11 and 12 show that the authentication cost and delay increase with the call arrival rate of an MU. As shown in (5) and (6), the authentication cost and delay are proportional to the call arrival rate λ_u since $\lambda_\beta,\ (\beta = 1, 2, 3)$ are proportional to λ_u. Moreover, a higher security level needs more cost and delay than a lower one. For example, if the security level increases from 1 to 2, the authentication will need about 7 times more cost and 29% more time than those at security level 1.

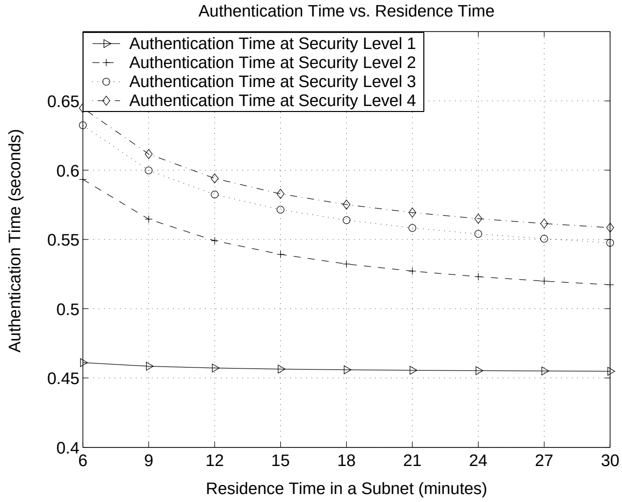


Fig. 9. Authentication Time vs. Residence Time in a Subnet.

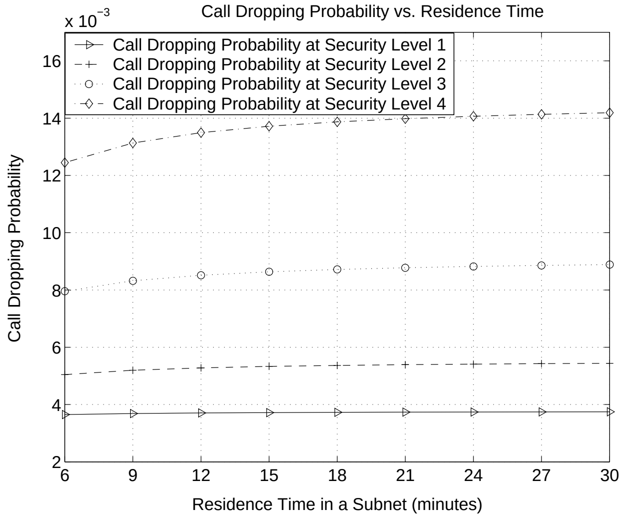


Fig. 10. Call Dropping Probability vs. Residence Time in a Subnet.

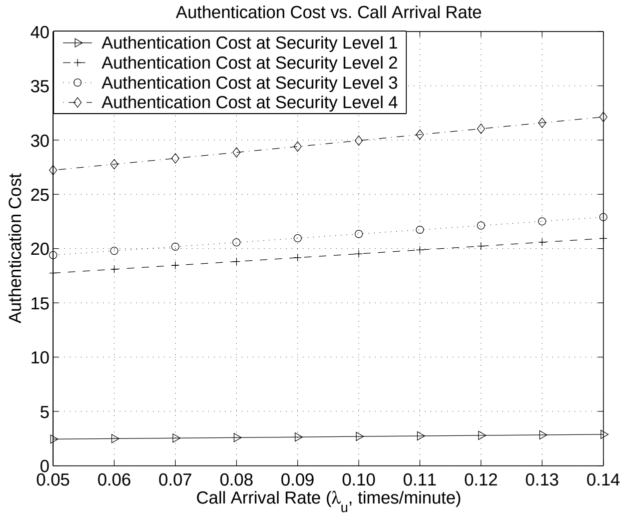


Fig. 11. Authentication Cost vs. Call Arrival Rate.

As for the call dropping probability at different call arrival rates, the call arrival rate of an MU does not affect the call dropping probability. As we can see in (7), $P(i)$ is average call dropping probability computed in the cases of intra-domain handoff authentication, session authentication, and inter-domain hand-

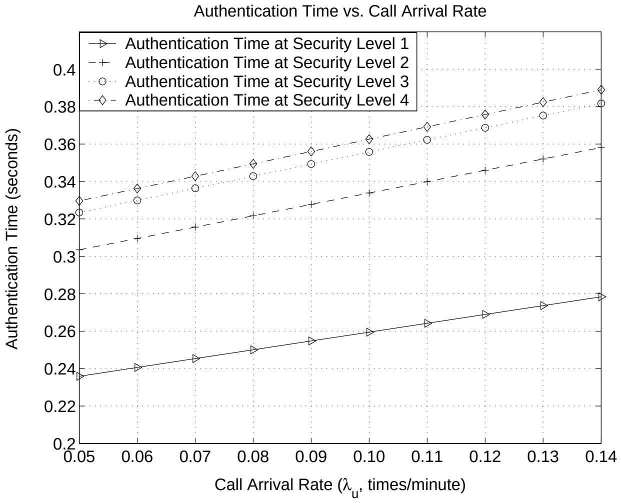


Fig. 12. Authentication Time vs. Call Arrival Rate.

off authentication. Since $\lambda_\beta, (\beta = 1, 2, 3)$ are all proportional to λ_u, λ_u disappears in $P(i)$'s definition equation (7). Therefore, once the PDF of the call duration time and the mobility patterns of the MU are known, i.e., η, μ_r, and γ are fixed, the call dropping probability of the MU can is a constant at different call arrival rates shown in Fig. 13. However, the call dropping probability is different at different security levels. As we can see in Fig. 13, the call dropping probability at security level 4 is about 56% more that that at security level 3.

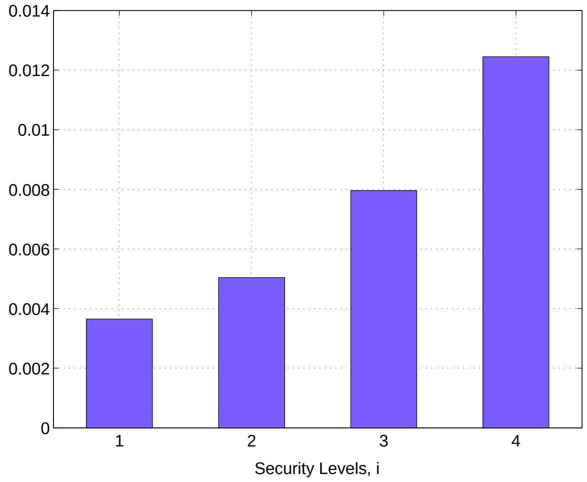


Fig. 13. Call Dropping Probability vs. Security Levels.

VI. Conclusion

In this paper, we investigated the impact of authentication on security and quality of service (QoS) in combination of mobility and traffic patterns, which is critical to deliver secure and efficient services in wireless IP networks. We analyzed the system performance with respect to authentication cost, delay, and call dropping probability at different security levels based on a system model with a challenge/response mechanism. To our best knowledge, our study is the first work on providing a quantitative connection between the security and quality of service, which is of extreme importance to the adaptation of new security solutions to various mobile environments. Therefore, this work provides an in-depth understanding of the authentication impact, and demonstrates a framework for the future design of efficient authentication schemes for wireless IP networks.

References

[1] A. Arumugam, A. Doufexi, A. Nix, and P. Fletcher, "An Investigation of the Coexistence of 802.11g WLAN and High Data Rate Bluetooth Enabled Consumer Electronic Devices in Indoor Home and Office Environments," *IEEE Transactions on Consumer Electronics*, vol. 49, pp. 587–596, August 2003.

[2] L. Salgarelli, M. Buddhikot, J. Garay, S. Patel, and S. Miller, "The Evolution of Wireless LANs and PANs - Efficient Authentication and Key Distribution in Wireless IP Networks," *IEEE Personal Communications on Wireless Communications*, vol. 10, pp. 52–61, December 2003.

[3] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko, "Diameter Base Protocol," *draft-ietf-aaa-diameter-17.txt*, December 2002.

[4] S. Jacobs, "Mobile IP Public Key Based Authentication," *draft-jacobs-mobileip-pki-auth-02.txt*, March 1999.

[5] C. Perkins and P. Calhoun, "Mobile IPv4 Challenge/Response Extensions," *RFC3012*, November 2000.

[6] *IEEE 802.11 Working Group. http://grouper.ieee.org/groups/802/11/index.html*.

[7] V. Gupta, S. Gupta, and S. Chang, "Performance Analysis of Elliptic Curve Cryptography for SSL," in *WiSe'02-ACM Workshop on Wireless Security*, September 2002.

[8] H. Kim and H. Afifi, "Improving Mobile Authentication with New AAA Protocols," in *IEEE International Conference on Communications*, vol. 1, pp. 497–501, 2003.

[9] W. Simpson, "PPP Challenge Handshake Authentication Protocol (CHAP)," *RFC1334*, August 1996.

[10] S. Shieh, F. Ho, and Y. Huang, "An Efficient Authentication Protocol for Mobile Networks," *Journal of Information Science and Engineering*, vol. 15, pp. 505–520, 1999.

[11] W. Liang and W. Wang, "A Cost-Aware Control Scheme for Efficient Authentication in Wireless Networks," in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'04.*, December 2004.

[12] B. Aboba and D. Simon, "PPP EAP TLS Authentication Protocol," *RFC2716*, October 1999.

[13] L. Dell'Uomo and E. Scarrone, "The Mobility Management and Authentication/Authorization Mechanisms in Mobile Networks beyond 3G," in *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 1, pp. c44–c48, September 2001.

[14] S. Glass, T. Hiller, S. Jacobs, and C. Perkins, "Mobile IP Authentication, Authorization and Accounting Requirements," *RFC2977*, October 2000.

[15] *http://standards.ieee.org/getieee802/download/802.1X-2001.pdf*.

[16] W. Stallings, "Network Security Essentials," *Applications and Standards*, 2000.

[17] W. Wang and I. Akyildiz, "Intersystem Location Update and Paging Schemes for Multitier Wireless Networks," in *Proc. of ACM/IEEE MobiCom'2000*, pp. 99–109, August 2000.

[18] S. Das, E. Lee, K. Basu, and S. Sen, "Performance optimization of VoIP calls over wireless links using H.323 protocol," *IEEE Transactions on Computers*, vol. 52, pp. 742–752, June 2003.

[19] Y. Xiao and J. Rosdahl, "Performance Analysis and Enhancement for the Current and Future IEEE 802.11 MAC Protocols," *ACM SIGMOBILE Mobile Computing and Communications Review (MCCR), special issue on Wireless Home Networks*, vol. 7, pp. 6–19, April 2003.

[20] E. Bertino, S. Jajodia, L. Mancini, and I. Ray, "Advanced Transaction Processing in Multilevel Secure File Stores," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, pp. 120–135, Feburary 1998.

[21] S. Sutikno and A. Surya, "An Architecture of F(2^{2^N}) Multiplier for Elliptic Curves Cryptosystem," in *Proceedings. ISCAS 2000 Geneva on Circuits and Systems*, vol. 1, pp. 279–282, May 2000.

[22] W. Wang and I. Akyildiz, "A New Signaling Protocol for Intersystem Roaming in Next-Generation Wireless Systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 2040–2052, October 2001.

[23] *http://www.paganini.net/ask/paper/node4.html*.

[24] D. Gross and C. Harris in *Fundamentals of Queueing Theory*, 1974.

[25] Y. Fang, I. Chlamtac, and Y. Lin;, "Channel Occupancy Times and Handoff Rate for Mobile Computing and PCS Networks," *IEEE Transactions on Computer*, vol. 47, pp. 679–692, June 1998.

[26] A. Hess and G. Schafer, "Performance Evaluation of AAA / Mobile IP Authentication," in *http://www-tkn.ee.tu-berlin.de/publications/papers/pgts2002.pdf*, 2002.

[27] P. Calhoun, T. Johansson, and C. E. Perkins, "draft-ietf-aaa-diameter-mobileip-13.txt," *IETF AAA Working Group*, October 2002.