Reducing False Synchronizations in 3G-WLAN Integrated Networks

Christoforos Ntantogian, Christos Xenakis, and Ioannis Stavrakakis, *Fellow, IEEE*

***Abstract*—Authentication in 3G encompasses a mechanism, which ensures that the authentication vectors (AVs) are used only once. To achieve this, the employed mechanism maintains counters at both sides (mobile station and network) and verifies that the provided AVs are among the last α generated. However, there are many cases in which the mobile station receives AVs that have not been previously used, but the employed mechanism rejects them as outdated. This phenomenon, called false synchronization, causes signaling overhead and delays, and increases the cost of the network use. False synchronizations are more frequent in 3G-WLAN integrated networks. The frequency of false synchronizations decreases with α, while at the same time the risk of a replay attack increases. This paper aims at analytically determining an appropriate value of α, which balances effectively in 3G-WLANs the tradeoff between the rate of false synchronizations and exposure to adversaries exploiting compromised AVs. This is done by determining a threshold value of α beyond which the further reduction in false synchronizations is marginal, while the potential for a replay attack is constantly increasing and substantial. To this end, an analytical model based on a four dimensional Markov chain is developed whose accuracy is verified through simulations.**

***Index Terms*—Authentication, 3G-WLAN, beyond 3G networks, sequence numbers, false synchronizations, Markov chain.**

I. Introduction

ONE of the most important steps for the provision of secure networked services is the authentication procedure between a mobile station (MS) and a network. In third generation (3G) mobile networks, authentication is executed between a MS, the access network (AN) that provides wireless connection to the MS, and the home network (HN) where MS is subscribed. The MS initiates the authentication procedure by sending an authentication request to the AN. If the latter has authentication credentials available for the specific MS, then the AN authenticates MS. Otherwise, the AN executes an Authentication Data Request (ADR) procedure: that is, the AN conveys the MS's identity to its HN which then generates a batch of L authentication credentials for the specific MS, called authentication vectors (AVs), and sends them to the AN. The AN provides to the MS one of them, which is used only in the specific (one) authentication and is deleted afterwards, and stores the remaining $L-1$ AVs to serve future authentication requests by the MS. The reason the HN generates a batch of L AVs, instead of only one as required for an authentication request, is to reduce the authentication delay in future authentication requests by avoiding executing an ADR and engaging the typically remote HN each time access to the AN is needed.

The authentication procedure of 3G networks encompasses a security mechanism, which ensures that each AV is used only once (freshness property). This protects both MS and AN from what is known as replay attacks: that is, an attempt by an adversary to use a compromised previously used AV in order to authenticate itself either as a valid MS or as a valid AN. To help detect such attacks, the MS and the AN try to keep track of previously used AVs in a way that creates some new issues that are identified and addressed in this paper.

In order to defeat replay attacks, the HN maintains a counter SQN_{HN} for each MS, which is increased by one each time a new AV is generated. The current value of SQN_{HN} is included in each generated AV by setting a parameter of the latter named SQN_{AV} equal to SQN_{HN} (i.e., SQN_{AV} = SQN_{HN}). In each authentication request, the AN always provides to MS the AV with the smallest value of SQN_{AV}. Therefore, MS should always receive an AV with SQN_{AV} value greater than the previously received. To be able to check against this, MS stores the greatest SQN_{AV} value of the received AVs in a counter called SQN_{MS}. The MS accepts an AV received from an AN if its SQN_{AV} value is greater than the one stored in MS (i.e., $SQN_{AV} > SQN_{MS}$) and subsequently stores the received SQN_{AV} value by setting SQN_{MS} equal to SQN_{AV} (i.e., $SQN_{MS} = SQN_{AV}$). Otherwise, it rejects the received AV, since it considers that it has been reused in the past [3]. This rejection provision ensures the freshness of the used credentials and constitutes a security mechanism.

However, the aforementioned security mechanism does not take into account the fact that the MS may change ANs. Suppose that after having been served by an AN 1, the MS moves to an AN 2 that belong to a different domain and attempts authentication to the newly visited network. The two different ANs possess different batches of AVs, generated by the MS's HN. Assume now that AVs residing at AN 2 are fresher than those residing at AN 1 (i.e., if the SQN_{AV} values of AVs of the AN 1 are $d+1, d+2, \ldots, d+L$, then the related values of the AN 2 are $d+(k\times L)+1, d+(k\times L)+2, \ldots, d+(k\times L)+L$, for some $d \in N$ and $k, L \in N^*$). In this case, the MS and the AN 2 are authenticated, mutually, and the latter continues to grant networked services to the former, uninterruptedly.

Manuscript received October 15, 2010; revised May 3, 2011; accepted July 16, 2011. The associate editor coordinating the review of this paper and approving it for publication was B. Liang.

C. Ntantogian and C. Xenakis are with the Department of Digital Systems, University of Piraeus, Karaoli and Dimitriou 80, PC 18534, Piraeus, Greece (e-mail: {dadoyan, xenakis}@unipi.gr).

I. Stavrakakis is with the Department of Informatics and Telecommunications, University of Athens, Panepistimioupolis, Ilisia, PC 15784, Athens, Greece (email:ioannis@di.uoa.gr).

Suppose now that the MS returns to the AN 1 after one or more authentications in the AN 2 in which case a new authentication is required. If AVs are available in the AN 1 from a previously used batch, then the MS will reject the received AV. This is because the received SQN_{AV} is smaller than the stored one $(d+1 < d+2 < \cdots < d+L < d+(k \times L)+1 < d+(k \times L)+2 < \cdots < d+(k \times L)+L)$, meaning that the involved AV has been used in the past, which is false. This phenomenon is called false synchronization [14], [15] and during this, AN deletes all the stored AVs for the specific MS and executes ADR to obtain a new batch of L fresh AVs from HN. Because of this unnecessary execution of ADR, false synchronizations impose signaling overhead that: (i) increases significantly the authentication latency, especially in cases that the MS is located far away from its HN [8]; (ii) increases the call blocking rate in cases that the MS has active real-time sessions (e.g., VoIP, videoconference) [6]; and (iii) overcharges the MS in cases that the AN and HN are located in different countries [9]. Therefore, false synchronizations: (a) deteriorate the overall network performance, (b) lower the quality of service offered to MS, and (c) increase the cost of the network use.

The frequency of false synchronizations in 3G networks, mainly, depends on the mobility of MS. The higher the mobility of an MS, the greater the probability of changing ANs, which causes false synchronizations. False synchronizations become more frequent in Beyond 3G (B3G) networks, which are characterized by the integration of different AN technologies, such as WLAN and WiMAX in 3G mobile networks [1],[10]. In B3G networks, the false synchronization problem can be intensified further since a unique geographical area is covered by more than one AN technologies, and MS may switch from one AN to another, based on various quality and service parameters such as data rate, available services, billing, signal strength, etc, without changing geographical area. In this paper, we focus on 3G-WLAN integrated networks [1] as a case study of B3G, since it is one of the most prominent and well studied network paradigms.

To reduce false synchronizations, 3GPP proposes a variation of the freshness mechanism that allows a MS to accept a provided SQN_{AV} of an AV, *if it is among the last "α" generated sequence numbers SQN_{AV}*. This means that during authentication, MS accepts the provided SQN_{AV} (and the related AV): (i) if the provided SQN_{AV} is greater than the stored in the mobile device (i.e., $SQN_{AV} > SQN_{MS}$) or (ii) if it is at most α values smaller than SQN_{MS} (i.e., $SQN_{MS} - SQN_{AV} \leq \alpha$). Otherwise, (i.e., $SQN_{MS} > SQN_{AV}$ or $SQN_{MS} - SQN_{AV} > \alpha$), the MS rejects the received AV. The value of α is fixed and is referred to as the *offset* in this paper. Using this revised freshness mechanism, false synchronizations are reduced, since the MS accepts not only the *"absolute"* fresh AVs, but also the *"recently"* generated ones. The higher the value of α, the lower the frequency of false synchronizations. On the other hand, the offset α creates negative side effects from the security viewpoint. Specifically, if an adversary compromises an AV, the offset α creates a time gap in which the adversary may exploit the compromised AV for malicious purposes [6],[11],[12],[13]. Therefore, the value of α is of paramount importance since it regulates a tradeoff between security and performance. A relatively high value of α reduces false synchronizations and the associated burden, but at the same time increases the risk of a replay attack. On the other hand, small values of α reduce this risk, but increases false synchronizations.

This paper aims at investigating analytically the impact of the value of the offset α in 3G-WLAN integrated networks, and help draw guidelines on the selection of the appropriate value of the offset, so that the tradeoff between security and performance is effectively dealt with. To this end, an analytical model based on an identified Markov structure is developed and is employed in the derivation of key performance metrics and overall study of the problem. More specifically, we derive the probability and the mean number of false synchronizations in 3G-WLAN integrated networks and investigate how they are affected by various network and MS parameters, as well as the offset α. Moreover, we provide insights into the combined effects of the two AN technologies (i.e., UMTS, WLAN) on false synchronization. The carried out simulations confirm the accuracy of the developed analytical methodology.

The rest of the paper is organized as follows. Section II outlines the 3G-WLAN network architecture and includes the related work. In Section III, the analytical model that is based on a four dimensional Markov chain is developed and analyzed. Section IV elaborates on the numerical results of the model and finally, section V contains the conclusions.

II. 3G-WLAN Network Architecture and Related Work

The 3G-WLAN integrated network architecture consists of four individual parts [1], [7] (see Figure 1): (i) the MS, (ii) the UMTS radio AN (UTRAN), (iii) the WLAN, and (iv) the 3G core network. The MS comprises the user's device (e.g., laptop, smart phone) and the universal subscriber identity module (USIM), which contains the user's subscriber information. UTRAN consists of Nodes B that provide wireless connections to MS, and radio network controllers (RNCs) that provide radio channel management services. One or more Nodes B connect to an RNC, while one or more RNCs connect to a serving gateway support node (SGSN), which is located in the 3G-WLAN core network. WLAN includes wireless access points that provide Wi-Fi access and act like authentication, authorization, accounting (AAA) clients, forwarding security related messages to an AAA server. Finally, the 3G core network includes SGSN, the AAA server and an authentication centre (AuC). SGSN provides mobility and session management services in UMTS, while the AAA server provides authentication services in WLAN. Both SGSN and the AAA server are connected to AuC, which contains the authentication credentials of MS.

The security architecture of 3G-WLAN integrated networks [2] defines that in case that a MS wants to gain access to UMTS, it should execute the UMTS authentication and key agreement (UMTS-AKA) protocol. On the other hand, if it wants to have access to WLAN, it should carry out the authentication protocol EAP-AKA [4]. As mentioned in 3GPP specifications (see section 5.1.2 of [3]), an authentication request in UMTS is triggered after: (i) the initial registration of MS in a serving network, or (ii) a service request, a location update procedure, an attachment/detachment with a network request, or a connection re-establishment request in case that the maximum number of local authentications using the derived session keys has been reached. On the other hand, an authentication request in WLAN is exercised whenever MS initiates a new connection with an access point, considering that EAP-AKA fast re-authentication is not possible [4]. Moreover, a handover occurs when MS switches its AN from UMTS to WLAN and vice versa followed by an authentication request in WLAN and UMTS respectively.

A few recent studies aim at improving the authentication procedure in 3G mobile networks. Zhang and Fujise [9] proposed a proactive fetching mechanism of AVs in order to reduce the delay of AV distribution from AuC to SGSN, when the former is located far away (in terms of number of hops) from the latter. Wu and Lin [5] focus on determining an optimal time period that an SGSN should store AVs. Moreover, Lin and Chen [8] proposed a novel algorithm that determines an optimal value of L, reducing the authentication signaling cost. Based on this, Al-Saraireh and Yousef [16] proposed an algorithm to improve the estimation of the optimal value of L.

Apart from those, some other studies cope with false synchronizations in 3G networks. More specifically, [14], [15] identify that false synchronizations with negative side effects may occur in 3G, without however, mentioning any possible solution. Zhang and Fang [6] alleviate false synchronizations in 3G by proposing a new authentication scheme that does not require synchronization between MS and the network, but imposes several modifications in the existing 3G infrastructure. Our work differs from the previous in the sense that we first pinpoint that false synchronizations may occur more frequently in B3G integrated networks, as well as we develop an analytical model to balance effectively the tradeoff between the number of false synchronizations and exposure to adversaries exploiting compromised AVs, without introducing extensive modifications in the existing network infrastructure.

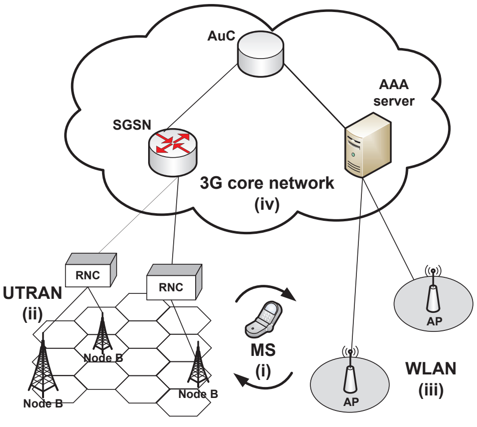


Fig. 1. 3G-WLAN network architecture

III. Analytical Study

In this section an analytical model based on a Markov chain is developed that leads to the derivation of the probability of false synchronizations P_{sync} and the mean number of false synchronizations $E[X_\tau]$ in 3G-WLAN integrated networks. X_τ is a random variable that denotes the total number of false synchronizations during a specific time period τ, in which we perform our experiments to gain insights and derive performance bounds for the observed system model.

A. The Embedded Markov Chain

To facilitate the system modeling, it is assumed that the residence time of a MS in UMTS and WLAN are exponentially distributed, and the authentication request rate forms a Poisson process (i.e., inter-authentication times are exponential distributions) [5]. Although long tail distributions (e.g., pareto) may provide better approximations of the MS residence time and authentication request rate, exponential distributions can be effectively used for mean value analysis, which does capture the trend of a system performance [17]. In general, exponential distributions strike a good balance between approximation of a real network and modeling simplification. The following definitions and notations will be used in the analysis that follows:

- The residence time of a MS in UMTS and WLAN is assumed to follow an exponential distribution with mean $1/\mu_u$ and $1/\mu_w$, respectively.
- The authentication request rate of a MS in UMTS and WLAN is assumed to be a Poisson process with rate λ_u and λ_w, respectively.
- The number of the generated AVs L_u and L_w in UMTS and WLAN, respectively, is assumed to be equal to L (i.e., $L_u = L_w = L$).
- The last SQN_{AV} that a MS has received from UMTS (i.e., SGSN) and WLAN (i.e., the AAA server), are denoted by SQN_{UMTS} and SQN_{WLAN}, respectively.

Consider the following quantities or processes embedded at the time instances k at which a network handover or authentication request (referred to as an H-A event) occurs:

- Let $N_k \in \{0,1\}$ denote whether the MS resides in UMTS ($N_k = 0$) or in WLAN ($N_k = 1$) following the k^{th} H-A event.
- Let $D_k \in (-\infty, +\infty)$, denote the difference between SQN_{UMTS} and SQN_{WLAN} (i.e., $D_k = SQN_{UMTS} - SQN_{WLAN}$), following the k^{th} H-A event.
- Let $U_k \in [0, L)$ and $W_k \in [0, L)$ denote the number of AVs stored in SGSN and the AAA server, respectively, following the k^{th} H-A event. All operations in the set $\{0, 1, 2.., L-1\}$ defined by the variables U_k and W_k use modular arithmetic with modulo L. To simplify notations, the symbol *"modL"* is omitted.

Table I summarizes the system and modeling parameters introduced above. In view of the aforementioned system modeling assumptions, it is not hard to show that the four dimensional process $E_k = \{N_k, D_k, U_k, W_k\}$ embedded at time instances k at which a network handover or authentication request occurs, is a discrete-time Markov chain.

It should be noted that the instances k at which an H-A event occurs are generated according to the following dynamics of handover and authentication requests:

1) Given that a MS is in UMTS (WLAN) in the current embedded instant, the next embedded time instant will be generated by an authentication request with probability $p_1 = \frac{\lambda_u}{\mu_u+\lambda_u}$ (with probability $p_2 = \frac{\lambda_w}{\mu_w+\lambda_w}$).
2) Given that the MS is in UMTS in the current embedded instant, the next embedded time instant will be generated by a handover from UMTS to WLAN (and will also trigger an authentication in WLAN) with probability $p_3 = \frac{\mu_u}{\mu_u+\lambda_u}$). Similarly, given that a MS is in WLAN in the current embedded instant, the next embedded time instant will be generated by a handover from WLAN to UMTS (and will also trigger an authentication in UMTS) with probability $p_4 = \frac{\mu_w}{\mu_w+\lambda_w}$).

TABLE I
ANALYTICAL MODEL PARAMETERS

Notation	Description
N_k	Denotes if MS resides in UMTS or WLAN
D_k	Difference between SQN_{UMTS} and SQN_{WLAN}
U_k	Number of AVs stored in SGSN
W_k	Number of AVs stored in the AAA server
SQN_{UMTS}	The last SQN_{AV} that MS has received from UMTS
SQN_{WLAN}	The last SQN_{AV} that MS has received from WLAN
$1/\mu_u$	Mean residence time in UMTS
$1/\mu_w$	Mean residence time in WLAN
λ_u	Authentication rate in UMTS
λ_w	Authentication rate in WLAN
L	Number of AVs that AuC generates in a batch
α	Offset value
X_τ	Number of false synchronizations during time τ
P_{sync}	Probability of false synchronization
$E[X_\tau]$	Mean number of false synchronizations during time τ

Next, we elaborate on the conditions that trigger a false synchronization. Assume that the Markov chain is in some state (δ, i, j, z) with $N_k = \delta, D_k = i, U_k = j$ and $W_k = z$. A false synchronization occurs, if the following conditions are satisfied:

1) A handover from UMTS to WLAN takes place, i.e., $\delta : 0 \rightarrow 1$.
2) WLAN (i.e., the AAA server) has at least one AV to provide to the MS, i.e., $W_k = z > 0$.
3) WLAN (i.e., AAA server) provides to the MS an AV that includes a SQN_{AV}, which is at least $\alpha + 1$ values smaller than the last SQN_{AV} that UMTS provided to the MS (i.e., $D_k = i > \alpha$, which means $D_k = \alpha + 1, \alpha + 2, \alpha + 3, \dots$).

The above are defined as *false synchronization conditions from UMTS to WLAN*. Note that the second condition guarantees that the AAA server has at least one "old" AV to provide to MS. On the contrary, if the second condition is not satisfied (i.e., $W_k = z = 0$), then a false synchronization does not occur when the MS moves to WLAN, since the AAA server performs ADR and fetches a batch of fresh AVs from AuC.

When a false synchronization occurs upon the MS handover from UMTS to WLAN (referred to as false synchronization in WLAN), we argue that the Markov chain jumps from the state $(0, i, j, z)$ with $i > \alpha$ and $z > 0$ to state $(1, -(j+1), j, L-1)$ since:

- The MS is located in WLAN after the handover and thus, $N_k = 1$.
- The false synchronization initiates ADR in WLAN, and thus, the AVs that resides in WLAN is fresher than this of UMTS, while it was the opposite just before the execution of ADR. Therefore, if the remaining j AVs in UMTS contain $SQN_{AV} = d+1, d+2, \dots, d+j$, then the newly generated AVs of WLAN contain $SQN_{AV} = d + j + 1, d + j + 2, \dots, d + j + L$, where $d \in N$. As a result, the difference between the last SQN_{AV} that UMTS has provided (i.e., $SQN_{UMTS} = d$) and the last SQN_{AV} that the MS received from WLAN in this handover (i.e., $SQN_{WLAN} = d + j + 1$) is $D_k = -(j + 1)$.
- The value of U_k is the same, $U_k = j$.
- From the newly generated AVs in WLAN the performed authentication (due to the handover) consumes one and thus, $W_k = L - 1$.

Summarizing the above we have:

Proposition 1: Upon a handover of MS from UMTS to WLAN, the Markov chain jumps from the state $(0, i, j, z)$ to $(1, -(j + 1), j, L - 1)$, if $i > \alpha$. In addition, if $z > 0$ is satisfied, then the handover from UMTS to WLAN triggers a false synchronization in WLAN.

Similarly, we can infer that a false synchronization occurs upon a MS handover from WLAN to UMTS (referred to as false synchronization in UMTS) when the following conditions (referred to as *false synchronization conditions from WLAN to UMTS*) are satisfied:

1) A handover from WLAN to UMTS takes place, i.e., $\delta : 1 \rightarrow 0$.
2) $U_k = j > 0$.
3) $D_k = i < -\alpha$, (i.e., $D_k = -(\alpha + 1), -(\alpha + 2), -(\alpha + 3), \dots$).

Finally, similarly to proposition 1, we have:

Proposition 2: Upon a handover of MS from WLAN to UMTS, the Markov chain jumps from the state $(1, i, j, z)$ to $(0, z + 1, L - 1, z)$, if $i < -\alpha$. Additionally, if $j > 0$, then the handover from WLAN to UMTS triggers a false synchronization in UMTS.

B. Markov Chain Truncation

In order to derive the probability of false synchronizations, the steady state probabilities of the Markov chain E_k should be derived first. To overcome the complexity introduced by the infinite state space of Markov chain E_k as $D_k \in (-\infty, +\infty)$, we truncate its state space properly so that the number of the possible states becomes finite, without compromising the accuracy of the resulting solution. The following proposition ensures that the original and the truncated Markov chains induce identical false synchronization events.

Proposition 3: Assume that $\tilde{E}_k$ is a truncated Markov chain of E_k, where $\tilde{E}_k = \{\tilde{N}_k, \tilde{D}_k, \tilde{U}_k, \tilde{W}_k$: $\tilde{N}_k \in \{0,1\}, \tilde{D}_k \in [-(\alpha+1), (\alpha+1)], \tilde{U}_k \in [0,L), \tilde{W}_k \in [0,L)\}$. Then, the evolution of E_k and $\tilde{E}_k$ is identical in the subspace that determines a false synchronization in UMTS or a false synchronization in WLAN and, thus, the two Markov chains induce the same number of false synchronizations.

Proof:

Notice that the values of D_k greater than $\alpha+1$ (i.e., $D_k = \alpha+2, \alpha+3, \dots$) or lower than $-(\alpha+1)$ (i.e., $D_k = -(\alpha+2), -(\alpha+3), \dots$) do not affect the evolution of the Markov chain in the subspace that determines the occurrence of a false synchronization in UMTS or a false synchronization in WLAN. Therefore, D_k parameter can be bounded between $-(\alpha+1) \leq D_k \leq \alpha+1$, as elaborated below.

If $-(\alpha+1) \leq D_k \leq \alpha+1$, it is evident that the two Markov chains $\tilde{E}_k$ and E_k are identical, $\tilde{E}_k \equiv E_k$. Now we investigate the evolution of the two Markov chains for $D_k > \alpha$. Assume that both chains are in the state $(0, \alpha+1, j, z)$ and an authentication in UMTS occurs, then E_k jumps to the state $(0, \alpha+2, j-1, z)$, since UMTS consumes one of the stored AVs reducing in this way U_k (i.e., $U_k = j-1$), and increasing D_k (i.e., $D_k = \alpha+2$). On the other hand, $\tilde{E}_k$ jumps to the state $(0, \alpha+1, j-1, z)$. After $n-1$ successive authentications in UMTS, E_k and $\tilde{E}_k$ are in states $(0, \alpha+n+1, j-n, z)$ and $(0, \alpha+1, j-n, z)$, respectively. We observe that for authentications in UMTS with $D_k > \alpha$, we have $\tilde{N}_k = N_k, \tilde{U}_k = U_k, \tilde{W}_k = W_k$, while $\tilde{D}_k$ remains equal to $\alpha+1$ and D_k increases (i.e., $D_k = \alpha+1, \alpha+2, \alpha+3, \dots$).

Assume now that a handover from UMTS to WLAN occurs. If $z > 0$, then in both chains this handover triggers a false synchronization in WLAN. More specifically, E_k jumps from the state $(0, \alpha+n+1, j-n, z)$ to state $(1, -(j+1), j-n, L-1)$, since $D_k = \alpha+n+1 > \alpha$ (see Proposition 1) and $\tilde{E}_k$ jumps from the state $(0, \alpha+1, j-n, z)$ to the state $(1, -(j+1), j-n, L-1)$, since $\tilde{D}_k = \alpha+1 > \alpha$. It is observed that $\tilde{D}_k = D_k = -(j+1)$ and thus, the two Markov chains are located in the same state from which their evolution is identical, since now $\tilde{D}_k = D_k < \alpha$. From this analysis, it can be deduced that whenever a false synchronization in WLAN occurs in E_k, the same false synchronization in WLAN also occurs in $\tilde{E}_k$. Similarly, we can prove that whenever a false synchronization in UMTS occurs in E_k, the same also happens in $\tilde{E}_k$. Thus, we can conclude that the evolution of E_k and $\tilde{E}_k$ is similar in the subspace that determines a false synchronization in UMTS or in WLAN.

End of proof.

It is evident that the truncated Markov chain $\tilde{E}_k$=$\{\tilde{N}_k, \tilde{D}_k, \tilde{U}_k, \tilde{W}_k\}$ is ergodic and converges to a steady state. Let $\pi\{N_k, D_k, U_k, W_k\}$ be the steady state probabilities of $\tilde{E}_k$.

C. Derivation of P_{sync} and E[X_t]

Having obtained the steady state probabilities of the truncated Markov chain, the probability of false synchronization P_{sync} can be calculated as follows. Let $P_{sync,W}$ be the probability of a false synchronization in WLAN, and $P_{sync,U}$ the probability of a false synchronization in UMTS. Considering the false synchronization conditions from UMTS to WLAN (see section III.A), $P_{sync,W}$ can be derived as $P_{sync,W} = Pr[D_k = \alpha+1$ and the WLAN has at least one AV stored for MS]×Pr[MS handover from UMTS to WLAN], which is equivalent to:

$$P_{sync,W} = \sum_{z>0,j=0}^{j=L-1} \pi\{N_k = 0, D_k = \alpha+1, U_k = j, W_k = z\} \times \frac{\mu_u}{\mu_u + \lambda_u} \tag{1}$$

Similarly, $P_{sync,U}$ can be derived as $P_{sync,U} = Pr[D_k = -(\alpha+1)$ and UMTS has at least one AV stored for MS] $\times Pr$[MS handover from WLAN to UMTS], which is equivalent to:

$$P_{sync,U} = \sum_{j>0,z=0}^{z=L-1} \pi\{N_k = 1, D_k = -(\alpha+1), U_k = j, W_k = z\} \times \frac{\mu_w}{\mu_w + \lambda_w} \tag{2}$$

It is evident that:

$$P_{sync} = P_{sync,U} + P_{sync,W} \tag{3}$$

In order to calculate the mean number of false synchronizations $E[X_\tau]$, we assume first that A_τ is the number of authentication transitions during τ, while MS resides in UMTS or WLAN. A_τ satisfies:

$$\begin{aligned} A_\tau &= \tau \frac{\mu_w}{\mu_u + \mu_w} \times \lambda_u + \tau \frac{\mu_u}{\mu_u + \mu_w} \times \lambda_w \\ &= \tau \left(\frac{\mu_w}{\mu_u + \mu_w} \times \lambda_u + \frac{\mu_u}{\mu_u + \mu_w} \times \lambda_w \right) \end{aligned} \tag{4}$$

Now, let H_τ be the number of handover transitions during τ from UMTS to WLAN and vice versa. H_τ is equal to:

$$H_\tau = 2 \frac{\tau}{\frac{1}{\mu_u} + \frac{1}{\mu_w}} = 2\tau \frac{\mu_u \mu_w}{\mu_u + \mu_w} \tag{5}$$

Let n be the number of total transitions in the Markov chain during τ. It is evident that:

$$n = A_\tau + H_\tau \tag{6}$$

Based on the definition of the frequency probability, we can derive the mean number of false synchronization E[X_τ] in a time period τ, which is sufficiently large (i.e., the Markov chain has reached its steady state) as:

$$E[X_\tau] = n \cdot P_{sync} \tag{7}$$

IV. Numerical Results

In this section we present and analyze the numerical results of the analytical model. To validate the accuracy of the analytical model, a discrete event-driven simulator written in C/C++ was developed. The statistical results collected from the simulation system after attained the equilibrium state, were averaged to eliminate the randomness effect. The

TABLE II
NUMERICAL RESULTS OF THE ANALYTICAL AND SIMULATION MODELS ($\alpha = 10, L = 5, \lambda_w = \lambda_u, \tau = 150h$)

λ_u	P_{sync}			$E[X_\tau]$ unit:μ_u		
unit:μ_u	Analytical	Simulation	Error	Analytical	Simulation	Error
1	0.00540.	0.00546	1.098%	5897.94	5898.00.	1.049%
5	0.03150.	0.03161	0.347%	68254.61	68291.67	0.054%
10	0.03450.	0.03460.	0.289%	121442.30.	121473.67	0.025%
25	0.02456	0.02459	0.122%	185666.80.	185843.00.	0.094%
45	0.01657	0.01659	0.120%	214798.30.	214958.67	0.074%
85	0.00991	0.00991	0.000%	235472.07.	235546.67	0.031%

simulations were carried out for a time period $\tau = 150$ simulation hours, which was sufficient for the Markov chain to reach its steady state. It was observed that the maximum related error between the analytical and simulation results was 1%, which verifies the accuracy of the analytical model. Table II displays indicative numerical results of the analytical and simulation models, as well as the related error for $\alpha = 10, L = 5, \lambda_u = \lambda_w$. For different input values, similar results were observed which are not presented.

The simulation model includes six different events: 1) handover_to_UMTS, 2) handover_to_WLAN, 3) authentication_in_UMTS, 4) authentication_in_WLAN, 5) ADR_in_UMTS, and 6) ADR_in_WLAN (see Figure 2). As long as MS is located in UMTS, authentication_in_UMTS and, when required, ADR_in_UMTS events are performed (i.e., steps (4)-(7)). At steps (8)-(9), the simulator checks if the next event is an authentication_in_UMTS or handover_to_WLAN. In the former case, steps (4)-(7) are repeated. In the latter case, the simulator verifies that the event list does not include any authentication_in_UMTS event (i.e., step (10)). This is essential in order to ensure that an authentication_in_UMTS will not be performed, while MS is located in WLAN. After the handover_to_WLAN is performed, the simulator generates a handover_to_UMTS which is inserted in the event list (i.e., step (11)) to ensure that MS will also perform a handover from WLAN to UMTS. After this, the simulator checks whether a false synchronization was caused by the handover from UTMS to WLAN (i.e., steps (12), (13)). The analogous procedure is followed if MS is located in WLAN (i.e., steps (14)-(20)).

The mobility model of MS can be viewed as a simple two state continuous time Markov chain that handoffs between UMTS and WLAN. It is assumed that MS does not have multiple connections in UMTS and WLAN at the same time. As long as MS remains in the same network, a false synchronization cannot occur. On the other hand, as MS moves from one AN to another (i.e., MS performs handovers), then false synchronizations occur. However, the number of false synchronizations is determined not only by the mobility of MS, but also by the values of the authentication request rate, the number of authentication vectors L and the value of offset α, as analyzed in the carried out experiments.

Overall, we have performed two sets of experiments. The objective of the first one is to determine an appropriate value of α that considerably reduces false synchronizations in 3G-WLAN, without however undermining the level of security. The second experiment aims at providing insights into the combined effects of the two AN technologies (i.e., UMTS and WLAN) on false synchronizations; and how the multiple parameters of the considered system model jointly affect its qualitative behavior.

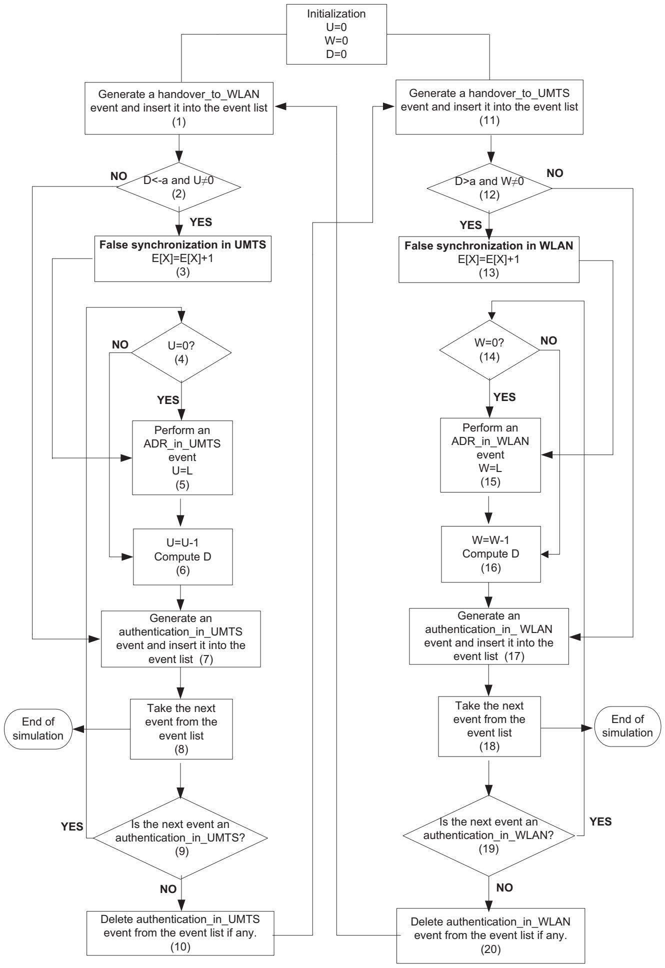
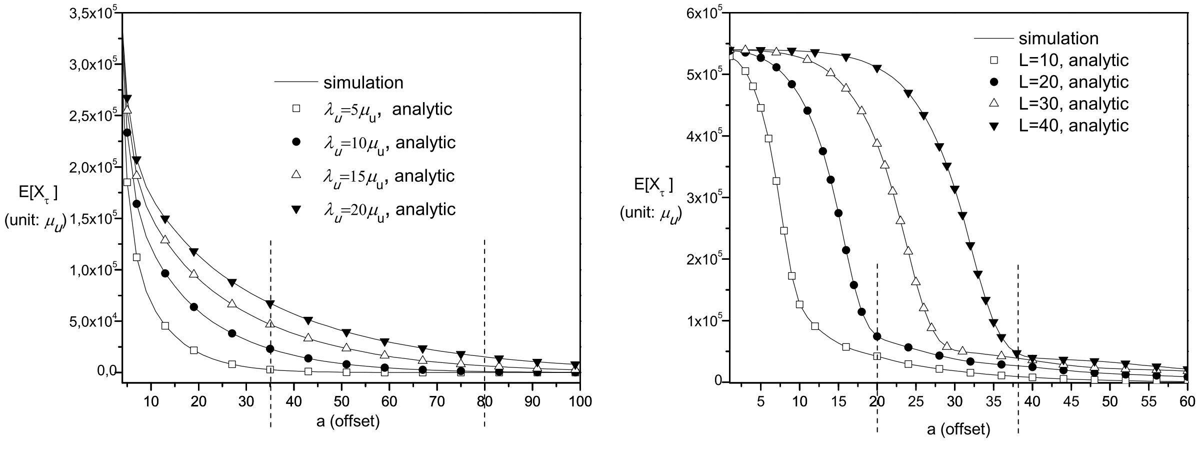

Initialization
U=0
W=0
D=0
Generate a handover_to_WLAN event and insert it into the event list (1)
Generate a handover_to_UMTS event and insert it into the event list (11)
NO
D<-a and U≠0 (2)
YES
D>a and W≠0 (12)
NO
YES
False synchronization in UMTS E[X]=E[X]+1 (3)
False synchronization in WLAN E[X]=E[X]+1 (13)
NO
U=0? (4)
YES
W=0? (14)
NO
YES
Perform an ADR_in_UMTS event U=L (5)
Perform an ADR_in_WLAN event W=L (15)
U=U-1 Compute D (6)
W=W-1 Compute D (16)
Generate an authentication_in_UMTS event and insert it into the event list (7)
Generate an authentication_in_ WLAN event and insert it into the event list (17)
End of simulation
Take the next event from the event list (8)
Take the next event from the event list (18)
End of simulation
YES
Is the next event an authentication_in_UMTS? (9)
Is the next event an authentication_in_WLAN? (19)
YES
NO
NO
Delete authentication_in_UMTS event from the event list if any. (10)
Delete authentication_in_WLAN event from the event list if any. (20)

Fig. 2. Simulation flowchart

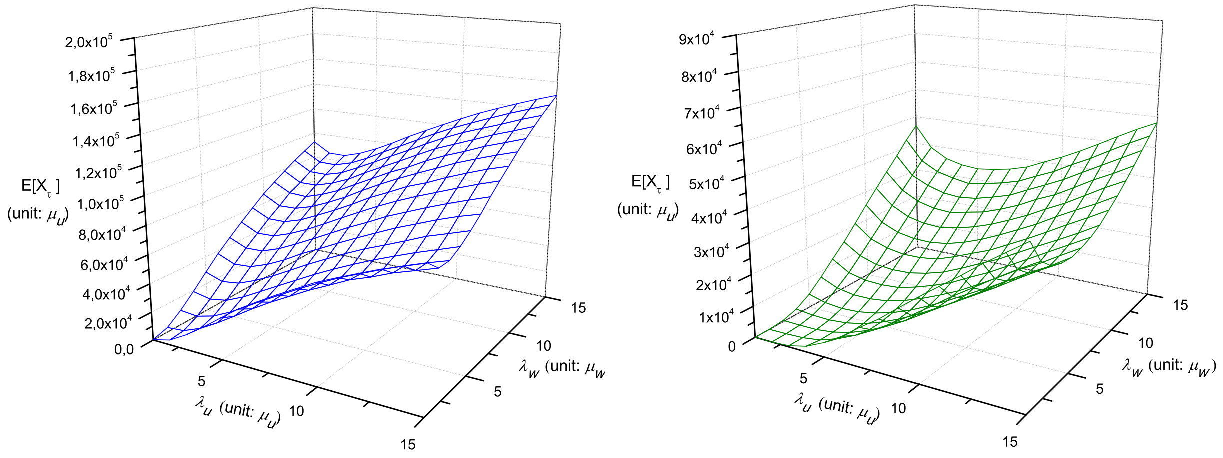
A. First Experiment

In this experiment, it is assumed that $\lambda_w = \mu_u = \mu_w$. The system showed the same qualitative behavior for different relations between λ_w, μ_u, μ_w, which are not presented to



(a) (b)

Fig. 3. Mean number of false synchronizations $E[X_\tau]$ as a function of offset α with parameter: (a) the rate of authentication requests in UMTS λ_u $(L = 5)$; and (b) the number of generated AVs L $(\lambda_u = 5\lambda_w)$



(a) (b)

Fig. 4. Mean number of false synchronizations as a function of λ_u and λ_w: (a) for $\alpha = 3L$ $(\alpha = 15$ and $L = 5)$; and (b) for $\alpha = 6L$ $(a = 30$ and $L = 5)$

conserve space. Figure 3(a) plots the mean number of false synchronizations E$[X_\tau]$, as a function of offset α and for various values of the authentication rate λ_u. The number L of AVs generated in a batch by AuC, each time the ADR procedure is executed, is equal to 5 (i.e., $L = 5$), as recommended by 3GPP [3]. Figure 3(a) shows that the mean number of false synchronizations is a decreasing function of α. However, it is observed that for high values of α, the mean number of false synchronizations becomes almost a constant function of α. By observing the values of the mean E$[X_\tau]$, we notice that as the offset α starts to increase from 0 to $\alpha_{optimum}$ the decrease of two successive values of E$[X_\tau]$ is very high (i.e., $\gg 5\%$) (i.e., exponential decrease). We also notice that by further increasing α beyond $\alpha_{optimum}$ the decrease of the $E[X_\tau]$ start to become quite small and remaining about 5%. Based on this, we define as $\alpha_{optimum}$ the smallest value of α in which the relative decrease between two successive values of the mean E$[X_\tau]$ is equal or less than 5%. More formally, let $(\alpha_1, \alpha_2, \ldots \alpha_k)$, with $\alpha_1 < \alpha_2 < \cdots < \alpha_k$ be a subset of successive values of α that satisfy $\frac{\{E[X]\}_{\alpha_{n+1}} - \{E[X]\}_{\alpha_n}}{\{E[X]\}_{\alpha_n}} \leq 5\%$. Then, $\alpha_{optimum} = min(\alpha_1, \alpha_2, \ldots \alpha_k) = \alpha_1$. Although this definition is arbitrary, it was derived after a rigorous analysis of the numerical results and it is quite accurate in determining the onset of the previously described phenomenon.

The value of $\alpha_{optimum}$ has the following important property. If $\alpha < \alpha_{optimum}$, then by increasing α, false synchronizations are significantly reduced. On the other hand, if $\alpha > \alpha_{optimum}$, then increasing α imposes a negligible reduction in false synchronizations. Based on the definition of $\alpha_{optimum}$, it can be deduced that $\alpha_{optimum}$ depends on the rate of authentication requests and the residence time

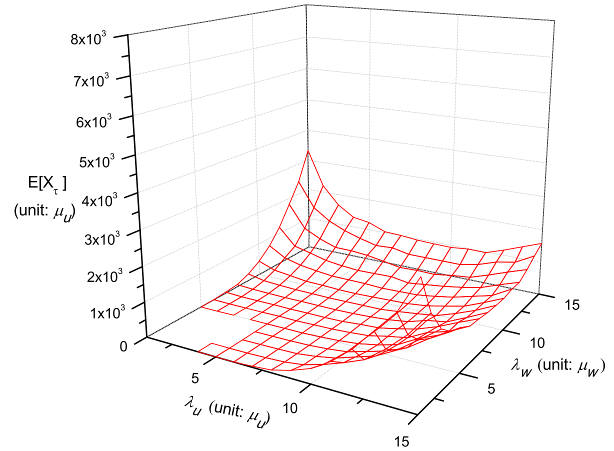


Fig. 5. Mean number of false synchronizations as a function of λ_u and λ_w for $\alpha = 18L$ ($\alpha = 90$ and $L = 5$)

of MS in each network. For example, if $\lambda_u = 5\mu_u$, then $\alpha_{optimum} = 35$, while if $\lambda_u = 20\mu_u$, then $\alpha_{optimum} = 80$ (see Figure 3(a)). To investigate further $\alpha_{optimum}$, Figure 3(b) plots the mean number of false synchronizations, as a function of offset α with parameter the number L of generated AVs (it is assumed that $\lambda_u = 5\lambda_w$). It can be realized that $\alpha_{optimum}$ not only depends on the rate of authentication requests and the residence time of MS in each network, but also from the different values of L. For example, if $L = 10$, then $\alpha_{optimum} = 20$, while if $L = 40$, then $\alpha_{optimum} = 38$.

Another interesting outcome of Figure 3(b) is the relation between α and L. For relatively small values of α, compared to L (i.e., $\alpha << L$), the mean number of false synchronizations yields high values. However, as α increases, the mean $E[X_\tau]$ reduces exponentially. For instance, for $L = 40$, if $\alpha = 10$, the mean number of false synchronizations is $E[X_\tau] = 5 \cdot 10^5 \mu_u$; while if $\alpha = 50$, $E[X_\tau] = 0.5 \cdot 10^5 \mu_u$, which means almost 90% reduction in false synchronizations. This can be attributed to the fact that if the offset α is much smaller than L, then almost any execution of the ADR procedure in UMTS or WLAN (as a result of the MS handover to UMTS or WLAN, respectively) entails the occurrence of a false synchronization. Finally, Figure 3(b) illustrates that if L increases, then the mean number of false synchronizations also increases. This is because whenever the value of L is increased, the counter SQN_{HN} is swiftly increased in every execution of ADR, and thus, false synchronizations occur more frequently.

From the analysis of the numerical results of the first experiment, we draw the following conclusions regarding the selection of α values. Selecting values of α smaller than $\alpha_{optimum}$ (i.e., $\alpha < \alpha_{optimum}$) increases false synchronizations. On the other hand, selecting values of α greater than $\alpha_{optimum}$ (i.e., $\alpha > \alpha_{optimum}$) does not considerably reduce false synchronizations, but at the same time increases the time period that an adversary may exploit compromised AVs to perform various attacks in UMTS [6],[11],[12],[13]. The value of $\alpha_{optimum}$ is estimated as a function of the rate of authentication requests in each network (i.e., UMTS and WLAN), the mean residence time of MS in each of them, and the number of generated AVs. Since the first two parameters for each network (i.e., $\lambda_u, \lambda_w, \mu_u, \mu_w$) may change over time, the 3G-WLAN network should, dynamically, estimate the value of $\alpha_{optimum}$ for every MS as $\alpha_{optimum} = f(\lambda_u, \lambda_w, \mu_u, \mu_w, L)$ and the AuC set $\alpha = \alpha_{optimum}$. In this way, false synchronizations are significantly reduced, without however undermining the provided level of security. The mechanism that calculates $\alpha_{optimum}$ can be implemented in the AuC, which will be informed by SGSN and the AAA server regarding the values of λ_u, μ_u and λ_w, μ_w of each MS, respectively. The AuC may invoke this calculation mechanism each time it generates AVs for an MS (i.e., the ADR procedure is executed). Then, it may convey the value of $\alpha_{optimum}$ to MS, through SGSN or the AAA Server, using the AMF field of AVs. As mentioned in the 3GPP specifications [3], the AMF field can be used to inform MS for dynamically changing network parameters. Thus, no extensive modifications to the UMTS or WLAN infrastructure are required to support the dynamic selection of the value of α.

B. Second Experiment

In the second experiment it is assumed that $\mu_u = \mu_w$. For different relations between μ_u and μ_w, the system showed the same qualitative behavior and thus, they are not presented in this paper. Figure 4(a) plots the joint effects of λ_u and λ_w on the mean number of false synchronization E[X_τ] for $\alpha = 3L$. It is observed that as λ_u and λ_w increase, the mean number E[X_τ] also increases, as expected. However, for very high values of λ_w, we pinpoint that, initially, E[X_τ] is subtly decreased, as λ_u increases. Symmetrical results for E[X_τ] are observed for high values of λ_u, as λ_w increases. This behavior of E[X_τ] is more perceptible, as the value of offset α increases (i.e., for $\alpha = 6L$, see Figure 4(b)). More specifically, Figure 4(b) depicts that for high values of λ_w (i.e., $\lambda_w > 2.5\mu_w$), the mean number E[X_τ] is reduced, as λ_u increases from 0 to $8\mu_u$. Only if $\lambda_u > 8\mu_u$ and λ_u increases, the mean number E[X_τ] is increased. Symmetrical results are also observed for high values of λ_u, as λ_w increases. This initial decrease of the mean number E$[X_\tau]$ as λ_u increases is justified as follows. A MS that is located in UMTS, it executes UTMS-AKA increasing the difference D_k. When at some point in time MS handovers from UMTS to WLAN, the system has not attained the third false synchronization condition from UMTS to WLAN (i.e., $D_k > \alpha$ - see section III.A), due to the high value of α. In WLAN, MS now executes EAP-AKA and therefore, the difference D_k swiftly starts to reduce, because of the high value of λ_w. This means that the system moves away from the third false synchronization condition from UMTS to WLAN, decreasing the mean number of false synchronizations. Finally, in Figure 5, it is observed that by further increasing α (i.e., $\alpha = 18L$), the mean E$[X_\tau]$ exhibits the same qualitative behavior, as for $\alpha = 6L$. The only difference is that the values of the mean number of false synchronizations E$[X_\tau]$, are considerably reduced, because of the high value of α.

From the above discussion, we can draw the following results regarding the combined effects of λ_w and λ_u on the mean number of false synchronizations E$[X_\tau]$. For small values of α, compared to L, the mean number E$[X_\tau]$ increases, as λ_w or λ_u increases. On the other hand, for high values of α, compared to L, a counterintuitive outcome is observed: that is for high values of λ_w (or λ_u), the mean number E$[X_\tau]$ initially decreases, as λ_u (or λ_w) increases, creating a parabolic curve. The interpretation of this phenomenon lies in the fact that for high values of λ_u, WLAN counterbalances the effect of UMTS on the false synchronization condition as λ_w increases and thus, the mean number of false synchronizations is reduced. Similarly, for high values of λ_w, UMTS counterbalances the effect of WLAN as λ_u increases, reducing the mean number of false synchronizations.

V. Conclusions

This paper aims at analytically determining an appropriate value of the offset α, which balances effectively in 3G-WLAN integrated networks the tradeoff between the rate of false synchronizations and exposure to adversaries exploiting compromised AVs. This is done by determining a threshold value of the offset α beyond which the further reduction in false synchronizations is marginal, while the potential for a replay attack is constantly increasing. To this end, an analytical model based on an embedded four dimensional Markov chain was developed. The numerical results showed that for high values of α, the mean number of false synchronization becomes almost a constant function of α. Based on this, we defined as $\alpha_{optimum}$ the smallest value of α in which the relative decrease between two successive values of the mean number of false synchronization is equal or less than 5%. Selecting values of α smaller than $\alpha_{optimum}$ (i.e., $\alpha < \alpha_{optimum}$) increases false synchronizations. On the other hand, selecting values of α greater than $\alpha_{optimum}$ (i.e., $\alpha > \alpha_{optimum}$) does not considerably reduce false synchronizations, but at the same time increases the risk of a replay attack. The value of $\alpha_{optimum}$ is estimated as a function authentication requests rate in each network (i.e., UMTS, WLAN), the mean residence time of MS in each of them, and the number of generated AVs. Since these parameters may change over time, AuC should dynamically calculate the value of $\alpha_{optimum}$ each time it generates AVs for a specific MS. The calculated value will be conveyed to MS using a reserved field of AVs and thus, the proposed dynamic estimation of α does not require extensive modifications in the UMTS or WLAN infrastructure.

Acknowledgement

The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] 3GPP TS 23.234 (v9.0.0), “3GPP System to WLAN Interworking; System description”, Release 9, 2009.

[2] 3GPP TS 33.234 (v9.0.0), “3G security; WLAN interworking security; System description”, Release 9, 2009.

[3] 3GPP TS 33.102 (v9.1.0), “3G Security; Security architecture”, Release 9, 2009.

[4] J. Arkko, H. Haverinen, “EAP-AKA Authentication”, RFC 4187, Jan. 2006.

[5] L.-Y. Wu and Y.-B. Lin, “Authentication vector management for UMTS”, *IEEE Transactions on Wireless Communications*, vol. 6, no 11, pp. 4101-4107, Nov. 2007.

[6] M. Zhang, Y. Fang, “Security analysis and enhancements of 3GPP authentication and key agreement protocol”, *IEEE Transactions on Wireless Communication*, vol.4, no.2, pp. 734-742, Mar. 2005.

[7] C. Xenakis, C. Ntantogian, “Security architectures for B3G mobile networks”, *Telecommunication Systems*, Springer, vol.35, pp. 123-139, Aug. 2007.

[8] Y.-B. Lin, Y.-K. Chen, “Reducing authentication signalling traffic in third-generation mobile network”, *IEEE Transactions on Wireless Communications*, vol.2, no. 3, pp. 493-501, May 2003.

[9] Y. Zhang, M. Fujise, “An improvement for authentication protocol in third generation wireless networks”, *IEEE Transactions on Wireless Communications*, vol.5, no. 9, pp. 2348-2352, Sep. 2006.

[10] K. S. Munasinghe, A. Jamalipour, “Interworked WiMAX-3G cellular data networks: an architecture for mobility management and performance evaluation”, *IEEE Transactions on Wireless Communications*, vol.8, no 4, pp. 1847-1853, Apr. 2009.

[11] Z. Ahmadian, S. Salimi, A. Salahi, “Security enhancements against UMTS-GSM interworking attacks”, *Computer Networks*, Elsevier, vol. 54, no. 13, pp. 2256-2270, Sept. 2010.

[12] U. Meyer, S. Wetzel, “On the impact of GSM Encryption and Man-in-the-Middle Attacks on the Security of Interoperating GSM/UMTS Networks”, *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, pp. 2876 - 2883, Sept. 2004.

[13] U. Meyer, S. Wetzel, “A Man-in-the-Middle Attack on UMTS”, *Proceedings of the 3rd workshop on Wireless Security*, pp. 90-97, Oct. 2004.

[14] V. Niemi, K. Nyberg, “UMTS Security”, John Wiley & Sons, 2003.

[15] G. Koien., “An introduction to access security in UMTS”, *IEEE Wireless Communications*, vol. 11, no. 1, pp. 8-18, Feb. 2004.

[16] J. Al-Saraireh, S.Yousef, “Analytical model for authentication transmission overhead between entities in mobile networks”, *Computer Communications*, Elsevier, vol. 30, no. 8, pp. 1713-1720, Jun. 2007.

[17] Y. Zhang, S. Xiao, M. Zhou, M. Fujise, “Authentication traffics modeling and analysis in next generation wireless networks”, *Wireless Communications and Mobile Computing*, Wiley, vol. 8 no. 5, pp. 615-625, Jun. 2008.

Dr. Christoforos Ntantogian (Dadoyan) has received his B.Sc degree in Computer Science and Telecommunications in 2004 and his M.Sc degree in Computer Systems Technology in 2006 both from the Department of Informatics and Telecommunications of University of Athens. In 2009 he received his Ph.D. from the University of Athens (Department of Informatics and Telecommunications). Currently, he is a postdoctoral researcher in the University of Piraeus. Dr. Christoforos Ntantogian has participated in numerous projects realized in the context of EU Programs (e.g., CONTENT, ANA, CASCADAS) and his research interests are in the field of system and network security.

Prof. Christos Xenakis received his B.Sc degree in computer science in 1993 and his M.Sc degree in telecommunication and computer networks in 1996, both from the Department of Informatics and Telecommunications, University of Athens, Greece. In 2004 he received his Ph.D. from the University of Athens (Department of Informatics and Telecommunications). From 1998 2001 he was with a Greek telecoms system development firm, where he was involved in the design and development of advanced telecommunications subsystems. From 1996 2007 he was a member of the Communication Networks Laboratory of the University of Athens. Since 2007 he is a faculty member of the Department of Digital Systems of the University of Piraeus, Greece, where currently is an Assistant Professor and member of the System Security Laboratory. He has participated in numerous projects realized in the context of EU Programs (ACTS, ESPRIT, IST, AAL) as well as National Programs (Greek) and his research interests are in the field of systems, networks and applications security.

Prof. Ioannis Stavrakakis, IEEE Fellow: Diploma in Electrical Engineering, Aristotelian University of Thessaloniki, (Greece), 1983; Ph.D. in EE, University of Virginia (USA), 1988; Assist. Prof. in CSEE, University of Vermont (USA), 1988-1994; Assoc. Prof. of ECE, Northeastern University, Boston (USA), 1994-1999; Assoc. Prof. of Informatics and Telecommunications, University of Athens (Greece), 1999-2002 and Prof. since 2002. Teaching and research interests are focused on resource allocation protocols and traffic management for communication networks, with recent emphasis on: peer-to-peer, mobile, ad hoc, autonomic, delay tolerant and future Internet networking. His research has been published in over 180 scientific journals and conference proceedings and was funded by NSF, DARPA, GTE, BBN and Motorola (USA) as well as Greek and European Union (IST, FET, FIRE) Funding agencies. He has served repeatedly in NSF and EU-IST research proposal review panels and involved in the TPC and organization of numerous conferences sponsored by IEEE, ACM, ITC and IFIP societies, including: organizer of the 1999 IFIP WG6.3 workshop, the COST-NSF NeXtworking03, the Workshop on Autonomic Communications (WAC2005); co-organizer of the 1996 ITC Mini-Seminar, the IEEE Autonomic Opportunistic Communications (AOC'07&'08); technical program co-chair for the IFIP Networking'00, EWC04, IFIP WiOpt05, COST-NSF NeXtworking07; general co-Chair for Networking2002, IFIP MedHocNet07. He has served as the chairman of IFIP WG6.3 and elected officer for the IEEE Technical Committee on Computer Communications (TCCC). He is an associate editor for the ACM/Kluwer Wireless Networks and Computer Communications journals and has served in the editorial board of the IEEE/ACM transactions on Networking and the Computer Networks Journals. He is currently the head of the Communications and Signal Processing Division of his Department.