Using the EAP Framework for Fast Media Independent Handover Authentication

Antonio Izquierdo, Nada Golmie, Katrin Hoeper and Lidong Chen
National Institute of Standards and Technology,
100 Bureau drive, stop 8920,
Gaithersburg, MD 20899-8920, USA.
E-mail: antonio.manzanares@nist.gov, nada.golmie@nist.gov, katrin.hoeper@nist.gov, lily.chen@nist.gov

Abstract—In this paper we compare different authentication techniques that may be used in order to reduce the time and resources required to perform a handover: namely, re-authentication and pre-authentication. We provide a detailed analysis of each mechanism, and highlight key performance trade-offs for select scenarios and parameters of interest.

Index Terms—Handover, Authentication, EAP, Performance, Pre-authentication, Re-authentication

I. INTRODUCTION
Key characteristics of today’s networking environment are the availability of different types of networks and the ability to roam across these networks on an as needed basis. As network heterogeneity has become more present, user mobility has also increased, making roaming from one network to another a common practice. However, in many cases roaming often implies a temporary service disruption, which is unacceptable for many user applications that have stringent quality of service requirements, such as voice and video. Thus, a number of tools and mechanisms have to be provided in order to minimize any service disruption. At a minimum, these mechanisms should include mobility protocols to manage the network transition, tools for continuously monitoring the network status, and mechanisms for fast authentication in the new network. Efficient mobility mechanisms provide the means to reduce the operations to be performed during the network entry. Some of these protocols include Mobile IPv6 [1] and Fast Mobile IPv6 [2] that improve the performance of the layer-3 handover. Similarly, there are mechanisms that aim to improve the layer-2 handover, such as the handover optimizations introduced in 802.16e [3] and the fast Base Station (BS) transition in 802.11r [4].

Although authentication may be considered as an intrinsic part of network entry, it is always treated separately due to its many security implications. Furthermore, and as we shall see in the later sections of this article, authentication involves complex operations requiring considerable amount of time and computational resources. Thus, reducing the authentication time is critical to the overall mobility performance and to achieving seamless handovers.

In this article we focus on studying efficient mechanisms for fast authentication in heterogeneous networks. We analyze re-authentication [5] and pre-authentication [6] in the context of a media independent framework such as the Extensible Authentication Protocol (EAP) [7] and compare these schemes to full authentication as performed at connection set-up time. For each technique we provide a detailed analysis and the means to estimate the authentication latency in terms of key network parameters, such as the propagation delay between nodes, the number of messages exchanged, and the cryptographic processing time required.

The remainder of this paper is organized as follows. Section II presents an overview of authentication in heterogeneous networks. In section III we describe and analyze each authentication scheme considered. In section IV we introduce the simulation environment used to obtain the results presented in section V. Finally, section VI has our concluding remarks.

II. AUTHENTICATION IN HETEROGENEOUS NETWORKS
In this section we provide a brief overview of how authentication is achieved in heterogeneous networks. Our goal is to provide insights on current practices and point out similarities found in various authentication techniques used in heterogeneous networks.

Currently, each network technology defines its own authentication mechanism that will allow the Point of Attachment (PoA) to validate the device’s credentials, and grant or deny access to the network. Considering authentication mechanisms defined in IEEE 802.11r, IEEE 802.16e, and Universal Mobile Telecommunications System (UMTS) [8], it is easy to see that, although the basic steps are comparable, all three technologies use different messages, cryptographic operations, and key derivation schemes.

However, in spite of these differences, there are a number of similarities when it comes to achieving fast authentication. The basic idea is to reuse the cryptographic information negotiated in previous authentications, so that most of the exchanged messages can be omitted during a fast handover. For example, in IEEE 802.16e the Pairwise Master Key (PKM) and the Traffic Encryption Key (TEK) can be reused, and in IEEE 802.11r during a fast Basic Service Set (BSS) transition the first level Pairwise Master Key (PMK-R0) is reused.

Extending these principles to heterogeneous networks implies reusing cryptographic material previously negotiated for
a different network. This can only be achieved using a common authentication architecture such as EAP, that provides a technology-independent framework for running authentication protocols. This framework consists of a set of messages and formats to transport the authentication messages, which are defined by the specific authentication algorithm used, called 'EAP method'. These methods define the messages to be sent by each peer and the way in which, upon successful authentication, a cryptographic key is agreed.

The EAP framework also defines several entities that play different roles during the authentication process: the supplicant, the Network Access Server (NAS), the authenticator and the server. The supplicant is the device that needs to be authenticated. The NAS is the device that provides network access to the supplicant. The authenticator receives the authentication requests forwarded by the NAS, and starts the authentication exchange with the supplicant. After the authentication process has started, the authenticator plays the role of a cache for the server. The server is the entity that validates the supplicant’s credentials and grants or revokes the network access. The relationship between these entities is purely hierarchical, as an EAP server may have several related authenticators, which in turn have several NASs forwarding messages and providing access services to many supplicants.

In wireless networks, the NAS represents the PoA, which is the base station in IEEE 802.16 or the Access Point (AP) in IEEE 802.11, and the supplicant is the mobile node (MN). The physical location or the type of device of the authenticator and the server are not specified as they are usually located in the backbone network.

A base topology including all the entities described above is shown in Figure 1a. When considering handovers, this topology is easily extended to Figure 1b, where both the current and target networks share the same server in order to allow for cryptographic information reuse.

III. EAP-BASED FAST AUTHENTICATION

We consider two EAP-based fast authentication schemes: namely, re-authentication and pre-authentication, and compare them to full authentication. All three schemes are described in details and their performance characterized using the authentication signaling and the EAP method latency metrics. The authentication signaling latency, \( l_{\text{Sign}} \), is defined as the time elapsed between the sending of the first authentication message until the reception of the ACK for the last authentication message during the network entry. On the other hand, the EAP method latency, \( l_{EAP} \), is the time elapsed between the sending of the EAP Start message until the reception of either the EAP SUCCESS or EAP FAILURE messages.

A. Full authentication

When using full authentication, the mobile device carries out a full EAP exchange with the corresponding EAP server in the backbone network. This exchange takes place using specific layer-2 authentication messages that transport the EAP information between the mobile node and the PoA, which in turn transmits these messages over the wired network to the EAP authenticator. This transmission is usually achieved by encapsulating the EAP contents in other protocols, such as RADIUS.

The distinctive characteristic of this authentication scheme is that there is no optimization performed over the specification of the EAP method. As we shall see in section V, this leads to large handover delays, due to the number of messages exchanged between the peer and the server.

The EAP message flow of a full authentication is shown in Fig. 2a. It involves the exchange of the EAP Start messages between the mobile node and the target PoA, the execution of the full EAP method exchange between the mobile node and the EAP server, and the distribution of the cryptographic key from the EAP server to the EAP authenticator and the PoA. This key distribution involves a Request / Response exchange between the EAP entities in the hierarchy, and ensures that every node gets and derives a set of keys, as defined by the EAP method in use. This whole exchange is equivalent in terms of delay to a Request / Response exchange between the target PoA and the EAP server. Each of these phases may involve processing delays, especially when the nodes have to perform cryptographic operations. This cryptographic delay is also considered for our characterization of the authentication latencies.

### TABLE I

<table>
<thead>
<tr>
<th>Notation Used in the Expressions of the Security Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_{\text{EAP}} )</td>
</tr>
<tr>
<td>( m_{\text{Auth}} )</td>
</tr>
<tr>
<td>( c )</td>
</tr>
<tr>
<td>( d_{MP_C} )</td>
</tr>
<tr>
<td>( d_{MP_T} )</td>
</tr>
<tr>
<td>( d_{PA_T} )</td>
</tr>
<tr>
<td>( d_{AC\text{-}AT} )</td>
</tr>
<tr>
<td>( d_{PS_T} )</td>
</tr>
<tr>
<td>( d_{PC\text{-}P_T} )</td>
</tr>
</tbody>
</table>
B. Re-authentication

Re-authentication is an optimization over full authentication, based on the reuse of information already exchanged between the EAP peer and server when performing a subsequent authentication with the same EAP server.

There are several proposals being considered for how to perform re-authentication, such as [9], [10], and [5]. In this study, we use the re-authentication scheme proposed in the IETF HOKEY working group in [5]. This scheme presents a minimum set of modifications to the original EAP framework, while providing a complete key sharing and derivation model that allows security services to be maintained during a handover. Although we have considered other re-authentication schemes in the course of our work such [10], they are omitted here due to a lack of space.

As shown in Fig. 2b, the EAP messages are reduced to the EAP Start, Identification (ID), and result (SUCCESS/FAILURE) messages exchanged between the EAP peer and the server. The result message also carries information about the new derived key and is propagated back to the peer through the authenticator and the PoA. A final message is exchanged between the old and the new authenticator in order to transfer the keys that the old authenticator got in the previous authentication. So, for re-authentication, the EAP latency becomes:

\[
    l_{EAP} = 2 \cdot d_{MP_T} + 2 \cdot (d_{MP_T} + d_{PS_T}) + 2 \cdot d_{PS_T} + c
\]

where \( c_{Reauth} \) is the total delay introduced by the cryptographic operations in this reduced EAP exchange. Since most of the cryptographic operations are eliminated, and the key derivation operations are simpler than computing new keys, we can say that \( c_{Reauth} < c \). Also, we note that in the case of re-authentication the EAP method exchange \((2 \cdot (d_{MP_T} + d_{PS_T}))\) consists of only two messages, instead of the variable number \( (m_{EAP}) \) that we saw in the full authentication case. This makes re-authentication EAP method-independent.

The resulting signaling latency is the same as the one computed for the full authentication case:

\[
    l_{Sign} = l_{EAP} + d_{MP_T} \cdot m_{Auth}
\]

Thus, reducing the number of messages exchanged and simplifying the cryptographic operations leads to a lower authentication signaling latency.

C. Pre-authentication

Pre-authentication is another optimization, which aims at reducing the resources and time needed for authentication during handovers. The main advantage of pre-authentication is that cryptographic material is not reused, hence it may be considered more secure. This also removes the constraint of having a common EAP server for both the current and target networks.

In order to make use of pre-authentication, the mobile device needs to perform a full EAP method exchange with the target network’s EAP server via the target PoA well before a handover is initiated. To reach the target PoA the mobile node may use a layer-2 or layer-3 tunnel through the current PoA, depending on the network topology, as highlighted in Fig. 2c.
The EAP latency is computed as:

\[ l_{EAP} = 2 \cdot (d_{MP} + d_{PC,P_T}) + \\
+ (d_{MP} + d_{PC,P_T} + d_{PS_T}) \cdot m_{EAP} + \\
+ 2 \cdot d_{PS_T} + c + c_{POA} \]

In this case, the EAP latency is likely to increase compared to that of a full authentication, due to the additional hops the messages have to go through, and the additional cryptographic operations performed between the PoAs if they establish a secure tunnel to protect the information forwarding, represented as \( c_{POA} \).

Once this EAP message exchange is done, the mobile node is authenticated in the target network, so during the handover the EAP phase is skipped. However, all the remaining steps in the network entry phase are performed in order to effectively bind the authentication credentials to the mobile device. Given that, the authentication signaling latency is expressed as:

\[ l_{Sign} = d_{MP} \cdot m_{Auth} \]

In this case, the authentication signaling latency during the network entry is significantly reduced. However, note that this improvement occurs only if the EAP method was successfully completed in advance, that is, before the mobile node lost the connection with the old network, or before the authentication with the new network starts, whichever happens first. If the pre-authentication phase is not completed before the handover starts, then the signaling latency of a full authentication applies.

Note that the notation used in the analysis is summarized in Table I.

IV. SIMULATION ENVIRONMENT

The simulation environment we use is based on the NS-2 [11] simulator, with extensions for the IEEE 802.16 model [12], and the authentication schemes described previously.

In addition to the performance metrics defined previously, namely, the EAP and the authentication signaling latencies, we define two new metrics as follows. Let the handover delay be the time elapsed between the decision to switch to a new network and the traffic redirection to the new interface. Similarly, we define the cryptographic processing delay as the time used by the mobile node to perform cryptographic operations related to the authentication.

The parameters used in the simulations are shown in Table II, and the network topology is based on the one shown in Fig. 1, with the following parameters:

- There is an 802.11 network and an 802.16 network, which share the EAP server.
- A common router between both networks and the backbone also hosts the EAP authenticator for both EAP NAS installed in both PoAs.
- The EAP server is connected to the EAP authenticator by a link with a 5 ms propagation delay.
- The link between the EAP authenticator and each of the PoAs have a 15 ms propagation delay.

Additionally, in order to provide the simulator with realistic information about the cryptographic performance, we used as an example a Palm Tungsten T3\(^1\) to obtain the time required to compute all the cryptographic operations. Although the performance results of the cryptographic processing are dependent on the choice of platform used, and may therefore vary from system to system, the relative gain or loss in performance between different authentication schemes remains valid for any platform choice.

Regarding the simulation parameters, a topic which deserves special attention is the EAP methods used, as their complexity, security level and number of messages exchanged vary greatly from one method to another.

EAP Methods

Our results have been obtained using two different EAP methods, with different levels of complexity: the EAP Generalized Pre-Shared Key (EAP-GPSK, [13]) and the EAP Tunneled TLS Authentication Protocol version 1 with MD5 challenge-response authentication (EAP-TTLSv1-MD5, [14]). EAP-GPSK is defined in an Internet draft, and aims at providing a lightweight EAP method with well-defined security requirements (such as mutual authentication), while keeping it as simple and efficient as possible. The resulting EAP method only requires four messages to be exchanged between the EAP server and the EAP peer.

EAP-TTLSv1-MD5 is more complex. First, the EAP server and the EAP peer establish a secure tunnel using a TLS-like message exchanges. Once the tunnel is ready, both entities perform an EAP authentication using a different method (in this case, MD5-Challenge, which only requires two messages to be sent). Besides the added complexity of the additional

<table>
<thead>
<tr>
<th>Handover optimization</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application traffic bursts</td>
<td>32 Kb every 50 ms</td>
</tr>
<tr>
<td>Application traffic type</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>Application traffic rate</td>
<td>640 Kb/s</td>
</tr>
<tr>
<td>Size of EAP IDs</td>
<td>64 bytes</td>
</tr>
<tr>
<td>Size of symmetric keys</td>
<td>128 bits</td>
</tr>
<tr>
<td>Size of Diffie-Hellman keys</td>
<td>1024 bits</td>
</tr>
<tr>
<td>Key lifetime</td>
<td>More than the simulation time</td>
</tr>
<tr>
<td>802.11 coverage area radius</td>
<td>50 m</td>
</tr>
<tr>
<td>802.16 coverage area radius</td>
<td>500 m</td>
</tr>
</tbody>
</table>

\(^1\)This device was used as an example platform, and its use in this research does not constitute an endorsement by National Institute of Standards and Technology.
(70% or more) over the full authentication case, regardless of the EAP method chosen. Second, we observe the same differences uncovered by the analysis in section III. Since for re-authentication, the EAP message exchange is significantly reduced, this leads to lower EAP and authentication signaling latencies. In the case of pre-authentication, although the EAP message exchange may take longer to complete, it is carried out before a handover is started. As a result, the authentication signaling latency during the handover is even lower than in the case of re-authentication.

Also, in Table IV we can see that when using pre-authentication the mobile node has to start the EAP message exchange more than 30 seconds prior to the start of a handover in order to complete it in time. In some situations, this may be hard to achieve since it is not always possible to decide which target network to use that much time in advance. On the other hand, by comparing the re-authentication data in Table III and Table IV we can see that, as we mentioned previously, when using re-authentication the latencies are not affected by the choice of the EAP method. This means that when using pre-authentication the mobile node has to decide which network it will move to well in advance, so it has enough time to perform the entire EAP message exchange, which can be time consuming.

B. Cryptographic Processing Delay

As cryptographic operations are usually computationally extensive, it is important to consider how these operations may be reduced when using either authentication scheme. The cryptographic delay is defined as the time spent by the mobile node to perform different cryptographic operations during the authentication phase. The values of this delay when using each of these schemes can be seen in Table V (for EAP-GPSK) and Table VI (for EAP-TTLSv1-MD5).

These results are consistent with the authentication schemes analysis presented in section III. Since pre-authentication does not shorten or eliminate any step in the authentication process, the cryptographic delay is the same as in the case of a full authentication. However, in the case of re-authentication, there are less operations carried out, and thus, the cryptographic processing delay is reduced. This may be an important consi-
C. Handover Delay

Our next step is the study of the impact of these authentication schemes on the total handover time. The handover delay represents the time elapsed between when a decision to handover is executed until the traffic is redirected to the new interface. In our case, the decision to perform a handover is made when a new link is detected and if the new link is better than the current link or if the current link is disconnected. The handover delay results are shown in Table VII.

We observe that re-authentication reduces the handover delay in a predictable way, regardless of the EAP method used, since the actual message exchange of the method is omitted. Also pre-authentication reduces the handover delay as long as the EAP message exchange is completed before the handover starts. We can see these two situations with EAP-GPSK, which completes before a handover, and EAP-TTLSv1-MD5, which does not complete before the handover starts.

VI. Conclusions

The emergence of mobile and heterogeneous network environments is making roaming between networks more common. However, according to the results obtained in our simulations, many applications and environments cannot afford the service disruptions that typically occur when a full non-optimized authentication process is required. Therefore, it is necessary to consider optimizations to network authentication in order to expedite the authentication and achieve a true seamless mobility experience. In this article, we compare the performance of re-authentication and indirect pre-authentication showing that both proposals have several advantages over full authentication.

Re-authentication reduces the number of messages exchanged and cryptographic operations, while pre-authentication may provide the best optimization of the authentication signaling delay during the handover. Both schemes can be used independently and at different times in the same topology depending on the needs and constraints of the scenario considered.

VI. References