Man-in-the-Middle in Tunnelled Authentication Protocols

N. Asokan, Valtteri Niemi, Kaisa Nyberg

Nokia Research Center, Finland

{n.asokan,valtteri.niemi,kaisa.nyberg}@nokia.com

Extended Abstract [1]

Abstract

Deploying a new security protocol is expensive. This encourages system designers to look for ways of re-using existing infrastructure. When security protocols and components are re-used, it is critical to re-examine the security of the resulting system as a whole. For example, it has become a common paradigm to run a legacy client authentication protocol within a server-authenticated tunnel. The commonest example of such composition is the use of HTTP authentication inside a TLS tunnel.

In this paper, we describe a man-in-the-middle attack on such protocol composition. The vulnerability arises if the legacy client authentication protocol is used both in tunnelled and untunnelled forms. We propose to solve the discovered problem by using a cryptographic binding between the client authentication protocol and the protection protocol.

1 Introduction

When new services and applications are developed, their security requirements and trust assumptions may necessitate designing and deploying new security protocols. Deploying a new security protocol is expensive. This encourages system designers to look for ways of re-using existing infrastructure. One form of reuse is to build on existing protocol components or frameworks so that the need to deploy new software is reduced. But the most difficult aspect of deployment is the cost of provisioning initial security associations, especially to end user devices. Consequently, there is considerable pressure to reuse security protocols and security context databases beyond their originally intended purposes and environments.

Re-use is generally considered a good thing. However, when security protocols and components are reused, it is critical to re-examine the security of the resulting system as a whole. This is particularly important when different components of the composition have different endpoints, either in different layers in the same entity, or different entities altogether. The fact that there are no easy-to-use tools or methodologies to verify the correctness of security protocols makes protocol re-use a risky task.

Faced with this difficulty, designers who needed to re-use legacy client authentication protocols and/or security associations have been using an obvious approach to shoring up security. They define an authentication protocol as a combination of two protocols. The legacy client authentication mechanism is run inside a server-authenticated tunnel. The most common example of this type of construction is the combination of a server-authenticated TLS tunnel [9] and the HTTP Digest Authentication mechanism [11]. Shortcomings of the legacy client authentication protocol in the new environment can apparently be offset by the tunnelling. For example, HTTP Digest Authentication may be based on a simple user-name/password; but the TLS tunnel will protect the password from any eavesdroppers. At the same time, the legacy authentication method may continue to be used legacy environments. Therefore, at first glance, secure tunnelling appears to be a reasonable approach. This is why, as described in Section 2, this type of protocol composition has been widely used when there is a need for reuse of legacy authentication mechanisms.

Unfortunately, when such a legacy client authentication protocol needs to be run both in the legacy environment as well as in a tunnelled environment, a man-in-the-middle attack becomes possible. In this paper, we describe the attack, and discuss how it can be avoided by cryptographically binding the inner and outer protocols. If the inner client authentication protocol produces a suitable key, this binding *does not require any changes* to the inner protocol. This is a very important requirement because the whole point of using legacy protocols is that they are already widely deployed.

---

[1] An earlier, longer version of this work appeared as a research report [6]

In section 2 some background information is provided by highlighting the common tunnelling approach and using PEAP as an example to describe how this is instantiated in practice. Then the man-in-the-middle attack is described in section 3. Ways of removing the vulnerability are presented in section 4. The use of weak password-based protocols is discussed in section 5. The conclusions are summarized in section 6. Finally, a brief status update is provided in section 7.

2 IETF Drafts on Tunnelled Authentication Protocols
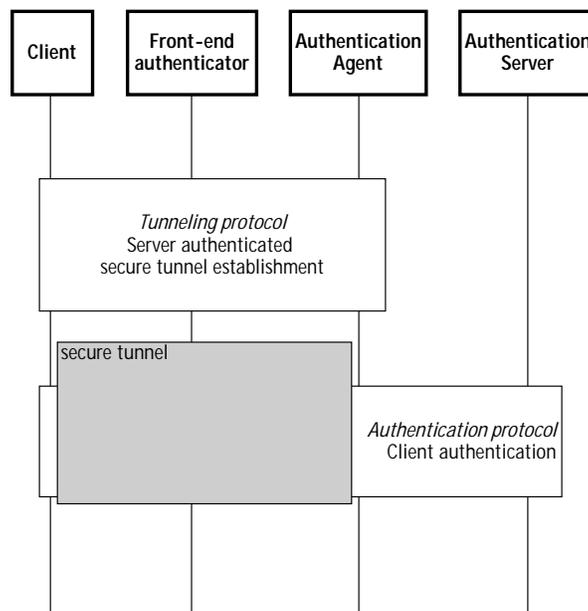
2.1 The general model

Figure 1: General model

A standard framework for client authentication protocols is the extensible authentication protocol (EAP) described in RFC2284 [16]. By using EAP in a system, it is possible to enable the system to support a number of legacy authentication schemes, including smart cards, Kerberos, public key mechanisms, One Time Passwords, cellular authentication mechanisms like GSM [14] or UMTS AKA (Universal Mobile Telecommunication System, Authentication and Key Agreement protocol) [4], and many others. In this section, we discuss tunnelled authentication protocols that use EAP as the inner client authentication protocol. But the discussion is applicable for any tunnelled authentication protocol.

EAP is a general authentication protocol run between a client and a server, possibly via a *front-end authenticator*. It does not require the front-end authenticator to authenticate the client itself. Instead, it allows the front-end authenticator to proxy authentication messages to an *authentication agent*. The authentication agent may use yet another server to help authenticate the client. We call this third server the *home authentication server*. Use of EAP allows new authentication methods to be developed without requiring deployment of new code on the front-end authenticator. The front-end authenticator acts as a "pass-through", and need not understand specific EAP methods.

Recently new protocols have been proposed in the Internet Engineering Task Force (IETF) for running EAP inside a server-authenticated tunnel. Examples of such protocols are PIC [20], PEAP [3], EAP-TTLS [12], POTLS [18], and most recently SLA [13], which led to the mechanisms for supporting legacy authentication methods in IKEv2 [10]. Some of these protocols, like PEAP, are motivated by a wish to correct perceived weaknesses of EAP, such as the lack of user identity protection and lack of a standardized mechanism for key exchange [3]. Others like PIC, POTLS and SLA are motivated by re-using legacy authentication databases for new applications. All of these new protocols are constructed in the same basic manner. First, a server authenticated tunnel is set up using a suitable protocol like TLS. Then the client authentication protocol is run inside this tunnel. This general model is illustrated in Figure 1. Other forms of tunnelled authentication protocols, such as HTTP Digest authentication inside a TLS tunnel, also conform to this general model.

2.2   Protected EAP

Protected EAP (PEAP) [3] is an example of the type of composition described in Figure 1: it wraps the EAP protocol messages within TLS [9]. It claims to provide user anonymity and built-in support for key exchange.

The relationship between the EAP peer (client), front-end authenticator, known as the "network access server" (NAS) in PEAP, and an authentication agent, known as the "back-end authentication server" in PEAP, is depicted in Figure 2. As described in the figure, the EAP conversation "passes through" the NAS on its way between the client and the back-end authentication server. While the authentication conversation is between the EAP client and the back-end authentication server, the NAS and back-end authentication server need to have established trust for the conversation to proceed.
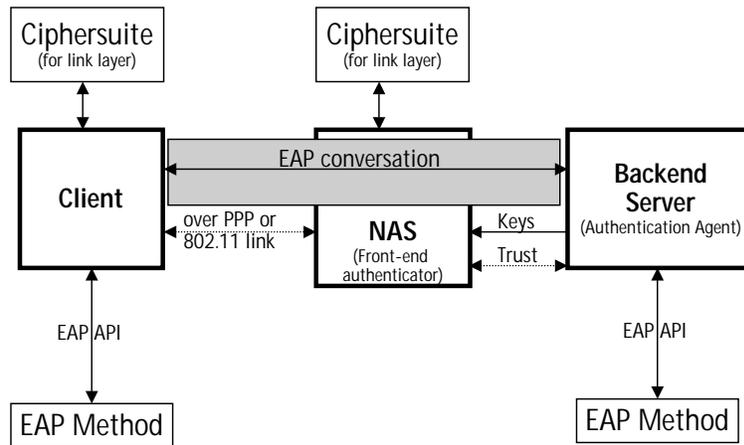

Figure 2: Relationship between EAP client, back-end authentication server and NAS in PEAP [3]

The client and the back-end server first set up a TLS channel over EAP. The client authentication protocol messages between the client and the back-end server are encrypted and integrity protected within this TLS channel. The NAS does not have knowledge of the TLS master secret derived between the client and the back-end authentication server, and cannot decrypt the PEAP conversation. The back-end server derives master session keys from the TLS master secret via a one-way function and conveys them to the NAS which can then use these session keys to protect subsequent link-layer communication between it and the client. The session key is transported from the server to the NAS using e.g. Radius or DIAMETER attributes.

In a recent Internet draft, EAP SIM GMM authentication [8], an application of PEAP to GSM authentication is presented. The same approach can be used to combine PEAP with AKA which is the client authentication mechanism
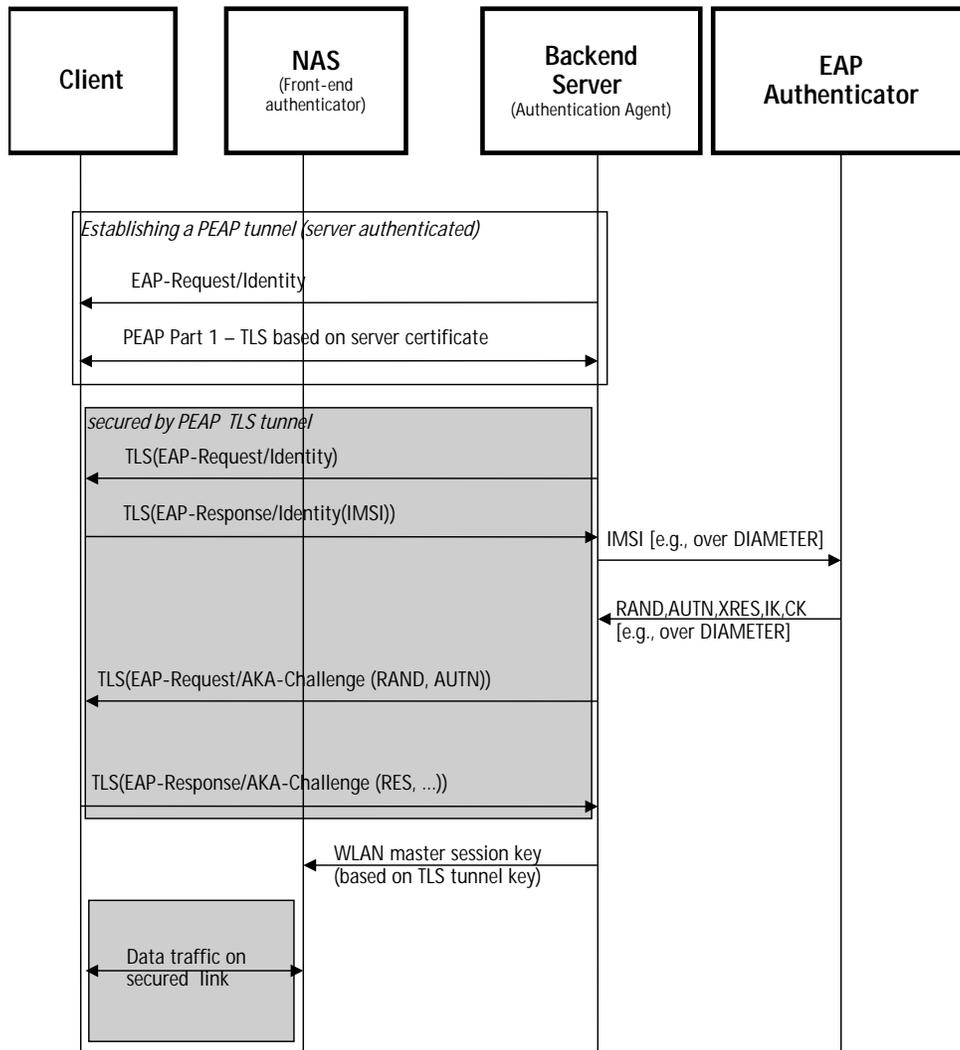
Figure 3: PEAP with EAP AKA: Example Message Flow

in the third generation UMTS networks. The message flow in PEAP with EAP AKA client authentication in the context of wireless local area network (WLAN) access authentication is depicted in Figure 3.

The WLAN authentication server is in the role of the back-end server. The terminal has a smart-card containing a secret key. The same key is available in the subscriber database of HSS (Home Subscriber Server) located in the home network of the subscriber. The terminal sends a cellular identity like the International Mobile Subscriber Identity (IMSI) as part of EAP AKA. The WLAN server uses a protocol like DIAMETER to send the IMSI to the HSS. HSS returns an AKA authentication vector, which is a quintuplet containing a challenge (RAND), an authenticator for the challenge (AUTN), expected response (XRES), and session keys IK (for integrity) and CK (for confidentiality) to the WLAN server. The WLAN server forwards RAND and AUTN to the terminal via EAP AKA. The terminal can now verify AUTN to confirm that RAND comes from HSS, and compute its own response RES and the keys IK and CK. The terminal sends RES back to the WLAN server using EAP AKA. If RES and XRES are the same, WLAN server considers the terminal to be authenticated. The AKA protocol was defined as part of the third generation (3G) cellular standardization activities. More detailed descriptions of 3G protocols are available elsewhere, e.g., [15].

Again, only the TLS master secret is used to derive the session keys to be used to protect the WLAN link. The secret key material carried within the 3G AKA [15] authentication quintuplets is not used.

3   Man-in-the-Middle Attack

Why can the tunnelling approach go wrong? There are two reasons:

- The legacy client authentication protocol is used in other environments, e.g., plain EAP without any tunnelling, or without any EAP encapsulation at all, e.g., direct use of one-time passwords, or cellular authentication protocols.

- The client fails to verify the server certificate properly (even when the authentication protocol is used *within* a tunnel).

The active attack by a Man-in-the-Middle (MitM) proceeds as follows:

1. MitM either waits for a legitimate device to start an untunnelled legacy remote authentication protocol or actively fools the legitimate device into initiating an untunneled remote authentication protocol. It then captures the initial message sent by the legitimate client.

2. MitM initiates a tunnelled authentication protocol with an authentication agent

3. After the tunnel is set up between MitM and the authentication agent, the MitM starts forwarding legitimate client's authentication protocol messages through the tunnel.

4. MitM unwraps the legacy authentication protocol messages received through the tunnel from the authentication agent and forwards them to the legitimate client.

5. After the remote authentication ended successfully, MitM derives the session keys from the same keys it is using for the tunnel.

Some legacy authentication protocols do not support mutual authentication or session key agreement. When they are used in an open environment, they are obviously vulnerable to MitM attacks. However, tunneling a generic authentication protocol framework like EAP makes the MitM attack possble even when the actual authentication method used within the framework does support mutual authentication and session protection. This situation arises because the authentication agent cannot be certain that the client in the legacy authentication protocol and the client endpoint of the tunnel are the same entity.

Figure 4 shows how the MitM attack works in PEAP with EAP AKA as the example client authentication protocol in a WLAN access authentication setting. The victim terminal assumes that the MitM is a UMTS radio access network. Note that even though the UMTS AKA protocol itself [15] and its EAP encapsulation [4] provide for mutual authentication of the user terminal and the radio access network: AUTN parameter is used to authenticate the source of the RAND challenge. The attempt to tunnel EAP AKA through PEAP effectively defeats this protection. In other words, the attack on the composed protocol does not imply that the inner authentication protocol is insecure.

Surprisingly, this type of protocol composition is very popular. As mentioned before, the most widespread of these may be the use of HTTP authentication [11] through a TLS tunnel (i.e., to an `https` URL). Many web sites use this type of a scheme for controlling access to their resource;, for example, on-line access to bank accounts are usually controlled this way. Naturally, if the authentication method is usable without a TLS tunnel (e.g., to a plain `http` URL), a MitM attack is possible.
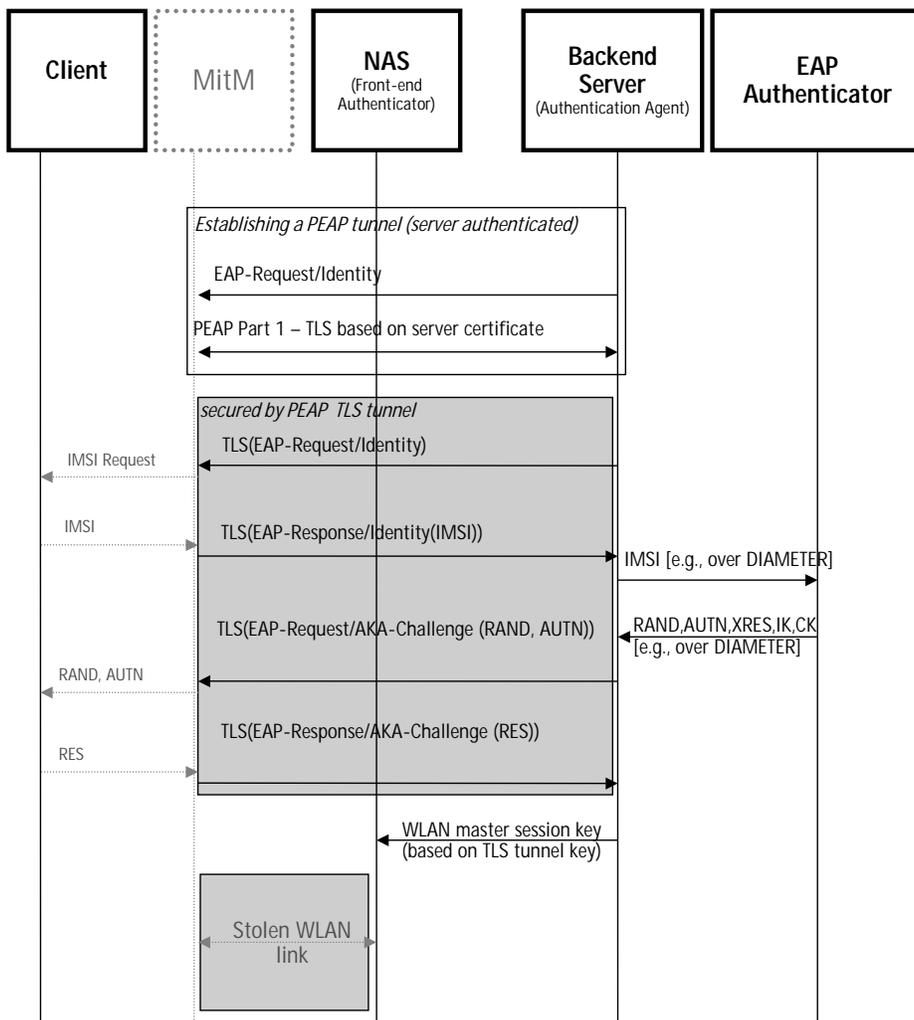
Figure 4: Man-in-the-Middle in PEAP, e.g., with EAP AKA

4 Avoiding the attack: Cryptographic Binding

The vulnerability discussed above is due to the fact that the entity that after the protocol gets hold of the client's copy of the session keys or credentials is not properly authenticated by the protocol. Two approaches to provide the necessary authentication are possible: by implicit or explicit authentication. The goal of implicit authentication is to ensure that the correct session keys or credentials are accessible only to the legitimate parties. In explicit authentication a separate authentication step is performed to verify that the master secrets from which the session keys are derived, are in possession of legitimate parties only. In both cases, the authentication is provided by a cryptographic binding between the inner and the outer protocols.

These solutions are applicable to mutual authentication protocols that are constructed as a combination of two authentication protocols. The outer protocol is assumed to support the construction of a secure tunnel based on server authentication. As we saw, this is the case with all the protocols cited in section 2.

A pre-requisite to cryptographic binding is that the inner authentication protocol must produce a key S in one of two ways:

1. S is a session key resulting from a run of the inner protocol ; i.e., it is an authentication *and* key agreement protocol; or

2. S is the client's long-term authentication key and is accessible for derivation of session keys directly, outside the authentication protocol.

A typical example of a protocol of the first type is EAP AKA. Figure 3 illustrates how EAP AKA can be used within the PEAP protocol. The AKA quintuplets contain 256 bits of secret session keys which would be readily available for session key derivation by the authentication agent. These AKA session keys are also available in the client's device. Similarly, most existing EAP types provide a method for session key derivation. It is likely that a common key agreement framework will be defined for EAP [2]. If such a framework is defined, all EAP types can result in session keys. For such protocols, no changes are required to provide the necessary cryptographic binding.

In the second case, it is necessary to specify an additional key derivation application that has direct access to client's authentication key. This would be in addition to the existing authentication protocol and is to be judged separately for each case.

Let K be the ultimate session key. In PEAP and EAP-TTLS the session key K is the master key that finally becomes available to the local authentication agent (access server), which uses K to derive further session keys. Let T denote the master key that is used to derive the secret keys for the protection tunnel. For example, the TLS master key derived in the TLS handshake of PEAP is a typical example of T.

There are two ways of using S to achieve the necessary binding between S and K. In the first method the binding is established directly by taking S in addition to T as input to the session key computation. This provides implicit authentication of the client. The second method is to make use of a cryptographic check value to verify that the client who is in possession of T is also in possession of S. This second type of binding provides explicit authentication of the client.

Both methods rely on secure communication between the network entities. On the network side some entity, let us call it "binding agent", is responsible for collecting the secret key information S and T and creating the binding value. Typically the binding agent is either the local authentication agent or the home authentication server. If S is a long term authentication key of the client, then the binding agent is preferably co-located with client's home authentication server, to avoid transfer of S across the network.

In implicit binding, the binding agent and the client each compute its copy of the session key K from S and T using a pseudo random function suitable for key derivation. The binding agent distributes the session key to the network entities that need to use it for further communication with the client.

In explicit binding, the binding agent and the client each compute its copy of a verification value V from S and T using a cryptographic hash function or a message authentication function. The client and the binding agent transfer their verification values to some network entity responsible for comparing the two verification values. If they are equal, the client is granted access to the network service. The comparing entity can be the authentication agent or the home authentication server.

A weaker variation of explicit binding works as follows. In the computation of the verification value V, some tunnel-indication data D is used instead of T. D indicates to the server that the client is communicating with it through a server-authenticated tunnel. Binding D with S further assures the server that the end point of the tunnel is the same as the entity that knows S. For example, D can be an entire protocol message sent by the client, or just a text string unique to the application. This variation is useful in cases where access to T is not available because of implementation constraints (e.g., a black-box implementation of TLS will not allow the TLS master key to be accessible outside TLS usage). This approach is weaker than the normal explicit binding, described earlier, because it relies on the client carrying out server authentication properly during tunnel setup.

It is also possible to implement both implicit and explicit binding. In such a case the explicit binding verification acts as a key confirmation for the agreed session key.

5   Weak Password-based Protocols

Consider a client authentication protocol based on weak secrets like passwords. Suppose further that the protocol does not use any modern techniques to protect the weak secret against guessing attacks. Many of the tunnelling protocols described earlier, e.g., the PIC protocol, are particularly intended to address such weak authentication protocols, which are insecure without added protective tunnelling. One may ask what is the impact of the cryptographic binding described in section 4 in such a case. Does the binding possibly improve the overall security? It is clear that if the client is unable to perform server authentication for the outer tunnel protocol, then the binding does not improve the security: an attacker could masquerade as the server in protocol run, and use the information gained in a dictionary attack on the password.

However, all the examples we considered in section 2 do assume that the client can perform server authentication for the tunnel protocol. In this case, the attacker cannot pretend to be the server end of the tunnel towards the client. If the attacker is a passive eavesdropper, he cannot perform the dictionary attack on the binding because the binding includes the outer tunnel key which is assumed to have sufficient entropy. If the attacker pretends to be the client towards the server and if the server does the binding first (e.g., if the server sends the binding verification value to the client, or if it starts using the key K or S to encrypt or authenticate messages), then he can perform a dictionary attack. But this is easily prevented by requiring that the client does the binding first and demonstrate knowledge of S and T to the server. If this step fails, the server must terminate the session.

Thus, the cryptographic binding does not *reduce* the security of a weak password-based protocol. However, to be fully secure, weak password-based protocols used with server authenticated tunnels must satisfy two conditions:

A1 *Correct server authentication*: the client **must** perform server authentication correctly, and

A2 *No mixing of authentication modes*: if a client uses tunnelled authentication, it **must not** use the same authentication protocol outside secure tunnels.

Assumption A2 is unavoidable for weak password-based protocols. But it clearly diminishes the capability of the tunnelling protocol to leverage the advantages of already deployed legacy authentication protocols. Well-designed authentication protocols do not need assumption A2.

If assumption A2 can be made in general, cryptographic binding is not necessary, but the legacy advantages are lost. On the other hand, cryptographic binding is not applicable to the legacy authentication protocols that do not yield a key suitable for cryptographic binding.

Generic tunnelling protocols, by definition, should be able to work with all types of authentication protocols while making as few assumptions about them as possible. Therefore a generic tunnelling protocol must not take A2 as a blanket assumption to be imposed on all authentication protocols. Instead it allow the possibility of using the type of cryptographic binding described in section 4 where appropriate. This way, the tunnelling protocol can be secure, generic (supporting both weak and strong authentication protocols), and non-invasive (avoiding unnecessary restrictions on strong authentication protocols).

6   Conclusion

In this paper we have shown that when a client authentication protocol is tunnelled within another protocol, it is necessary for each endpoint to demonstrate that it has participated in both protocols within the authentication exchange. If this is not demonstrated then the tunnelled authentication protocol is vulnerable to a Man-in-the-Middle attack.

We have also shown that the required demonstration can be provided in an implicit or explicit way in a form of a cryptographic binding between the tunnel protocol and the authentication protocol. In our proposals the binding facility is implemented in the outer tunnel protocol. It requires the authentication protocol to provide some secret key values for the use of the binding. This approach is preferred since it requires minimal or no changes to the EAP protocols. It allows for flexible and secure usage of an authentication protocol in multiple authentication environments without the authentication protocol being aware of the specific environment.

The cryptographic binding proposed in this paper does not diminish the security of the tunnelled protocols in any case. If the inner authentication protocol is a weak authentication protocol based on a weak client secret, the tunnel must be constructed based on server authentication, and the client should not use the same secret in different environments. Otherwise, the protocol is vulnerable to dictionary attacks, with or without cryptographic binding. Strong authentication methods are not vulnerable to dictionary attacks, and hence should not be restricted to tunnelled environments only. Hence the need for the type of binding described in this paper.

7 Current developments

We pointed out the MitM problem to the authors of some of the tunnelling proposals in early October, 2002. Since then a number of positive developments have taken place. The IETF EAP working group has recognized the problem of securely binding a sequence of authentication methods together. It is now marked as an open issue to be solved [7]. Another draft [19] describes the type of solutions prescribed in this paper.

Client authentication protocols may take additional safeguards to protect against MitM. For example, in a newer version of the EAP AKA specification [5], the authentication response must include RES as well as a message authentication code on RES. Thus the attack described in Figure 4 is no longer possible. However, the MitM can still pretend to be an EAP authenticator towards the victim. Therefore, if mixed mode usage is unavoidable, then ultimately the tunnelling protocols have to be fixed.

The surprising aspect was the sheer number of different protocol combinations where the same MitM problem has manifested. It is applicable to a number of protocols and drafts in progress, including PIC, IKEv2, XAUTH, PEAP, EAP TTLS and POTLS [1]. These are not academic toy protocols, but protocols with a good prospect of being deployed and used widely. For IKEv2 it concerns, in particular, the Secure Legacy Authentication (SLA) protocol, which is intended to be included in the main body of IKEv2. The conclusion from the discussion at the IPSEC list [21] was that an explicit cryptographic binding for SLA will be specified. The binding is mandated in case the legacy authentication method provides a shared secret. This change has already been included in a more recent version of the IKEv2 draft [10]. As we mentioned already, the problem is not limited to EAP only. For example, HTTP Digest AKA specification [17] effectively uses the HTTP authentication protocol [11] as a framework. The standard combination of HTTP authentication with TLS is thus rendered insecure when the underlying HTTP authentication is HTTP Digest AKA, despite the fact that HTTP Digest AKA is a secure protocol providing mutual authentication and strong session key agreement.

We therefore expect that in due course of time the problem will be solved in all the identified specifications. However, it requires significant analysis and development for each protocol. In each case one must decide at which protocol message the binding is implemented, which parties shall create the binding, and how to derive the shared secrets used for the binding. Different decisions result in different protocol properties.

A particular problem is created by the large variety of legacy protocols. Some of them provide strong mutual authentication, and others are based on a user selected password. It is not clear if it is useful or even possible to accommodate all different protocols within the same tunnelling framework. The MitM problem discussed in this paper were discovered when the protocols were used for purposes that the designers of the protocol did not explicitly foresee, but which were within the defined scope and satisfy the specified requirements.

References

[1] Bernard Aboba. Review of man-in-the-middle problem statement draft. Message to IETF saag mailing list, January 2003. `http://jis.mit.edu/pipermail/saag/2003q1/000684.html`.

[2] Bernard Aboba and Dan Simon. The EAP Keying Problem, October 2002. IETF personal draft `draft-aboba-pppext-key-problem-03.txt`.

[3] H. Andersson, S. Josefsson, Glen Zorn, Dan Simon, and Ashwin Palekar. Protected EAP Protocol (PEAP), September 2002. IETF personal draft `draft-josefsson-pppext-eap-tls-eap-05.txt`.

[4] J. Arkko and H. Haverinen. EAP AKA Authentication, June 2002. IETF personal draft `draft-arkko-pppext-eap-aka-04.txt`.

[5] J. Arkko and H. Haverinen. EAP AKA Authentication, January 2003. IETF personal draft `draft-arkko-pppext-eap-aka-08.txt`.

[6] N. Asokan, Valtteri Niemi, and Kaisa Nyberg. Man-in-the-middle in tunneled authentication protocols. Technical Report 2002/163, IACR ePrint archive, October 2002. `http://eprint.iacr.org/2002/163/`.

[7] L. Blunk, J. Vollbrecht, and Bernard Aboba. Extensible Authentication Protocol (EAP), October 2002. IETF *pppext* working group draft `draft-ietf-pppext-rfc2284bis-07.txt`.

[8] Adrian Buckley, Prasanna Satarasinghe, Vladmir Alperovich, Jose Puthenkulam, Jesse Walker, and Victor Lortz. EAP SIM GMM Authentication, August 2002. IETF personal draft `draft-buckley-pppext-eap-sim-gmm-00.txt`.

[9] T. Dierks and C. Allen. The TLS Protocol Version 1.0, January 1999. IETF RFC 2246.

[10] Charlie Kaufman (Editor). Internet Key Exchange (IKEv2) Protocol, February 2003. IETF *ipsec* working group draft `draft-ietf-ipsec-ikev2-05.txt`.

[11] J. Franks et al. HTTP Authentication: Basic and Digest Access Authentication, June 1999. IETF RFC 2617.

[12] Paul Funk and Simon Blake-Wilson. EAP Tunneled TLS Authentication Protocol (EAP-TTLS), February 2002. IETF *pppext* working group draft `draft-ietf-pppext-eap-ttls-01.txt` (expired).

[13] Dan Harkins, Derrel Piper, and Paul Hoffman. Secure Legacy Authentication (SLA) for IKEv2, December 2002. IETF personal draft `draft-hoffman-sla-00.txt`.

[14] H. Haverinen and J. Salowey. EAP SIM Authentication, October 2002. IETF personal draft `draft-haverinen-pppext-eap-sim-06.txt`.

[15] Heikki Kaaranen, Siamäk Naghian, Lauri Laitinen, Ari Ahtiainen, and Valtteri Niemi. *UMTS Networks: Architecture, Mobility and Services*. John Wiley & Sons, 2001.

[16] J. Vollbrecht L. Blunk. PPP Extensible Authentication Protocol (EAP), March 1998. IETF RFC 2284.

[17] Aki Niemi, Jari Arkko, and Vesa Torvinen. Hypertext transfer protocol (http) digest authentication using authentication and key agreement (aka). IETF RFC 3310, September 2002.

[18] Yoshihiro Ohba, Shinichi Baba, and Subir Das. PANA over TLS (POTLS), September 2002. IETF personal draft `draft-ohba-pana-potls-00.txt`.

[19] Jose Puthenkulam, Victor Lortz, Ashwin Palekar, Dan Simon, and Bernard Aboba. The compound authentication binding problem, October 2002. IETF personal draft `draft-puthenkulam-eap-binding-01.txt`.

[20] Y. Sheffer, H. Krawczyk, and Bernard Aboba. PIC, A Pre-IKE Credential Provisioning Protocol, October 2002. IETF *ipsra* working group draft `draft-ietf-ipsra-pic-06.txt`.

[21] IETF IPsec working group. Secure legacy authentication for IKEv2. Discussion thread on the IPSec mailing list. `http://www.vpnc.org/ietf-ipsec/mail-archive/threads.html#02763`.