3GPP TSG SA WG3 Security — S3#19 **S3-010326**

4-6 July, 2001 **PA4**

Newbury, Great Britain

Source:	**Ericsson, Nokia, and Nortel**
Title:	**Security mode setup for the IMS registration**
Document for:	**Discussion / Decision**
Agenda Item:	**tbd**

Contents

1. Introduction

The so-called *security mode setup* procedure is generally required in all secure protocols to decide when and how the security services start. In the IMS system, authentication of users is performed during the registration of the SIP subscriber to the S-CSCF. Subsequent signalling communications in this session will be integrity protected after this authentication, based on keys derived during the authentication process e.g. through the use of AKA. Suggested message flows and formats exist for the authentication process [2] as well as for the integrity protection of the SIP signalling traffic [3].

However, it still remains to be decided for the IMS how the involved parties decide what kind of protocols, algorithms, and parameters shall be used to protect the communications, and exactly when does the protection start. According to earlier discussions it will not be possible to agree on a single algorithm that will be solely used, as cryptographic algorithms may be broken. The protocols that will be used in UMTS must be able to deal with these situations, and it must be possible for equipment that supports better algorithms to choose them over the broken ones.

The purpose of this contribution is to propose a particular way of performing the security setup mode process. In particular, we propose the following:

- The security setup mode shall be integrated as a part of the registration and authentication message flow
- The security setup mode shall be integrated as a part of SIP
- A minimal approach with the least number of added roundtrips will be used.

2. Previous Work

In the Madrid meeting, Ericsson's contribution [4] showed the basic two approaches to implementing the security setup mode:

- Following the approach taken in Release 1999 [5] as an independent phase.
- Integrating the Release 1999 procedures in a more "compressed" way to the existing SIP message flow.

In the Phoenix meeting, Nokia's contribution [6] discussed security problems related to registering to the P-CSCF before integrity protected was turned on. This contribution proposed that integrity protection be started already at the second (and last) registration message sent from the client to the P-CSCF.

3. Requirements

In this section we discuss the main requirements for the security setup mode. The requirements have been classified to the following groups:

- Functional
- Security
- Characteristics

3.1. Functional Requirements

It SHALL be possible to choose different integrity protection algorithms. This is the main requirement.

It SHALL be possible for the terminal to present its supported algorithms in priority order.

It SHALL be possible for the proxy to pick any algorithm from the terminal's list (or to reject the request) and to inform the terminal of the chosen algorithm.

It MAY be possible to choose different algorithm parameters, encryption algorithms, and integrity/encryption protection protocols. It would be beneficial for future expansion to be able to do these, but it isn't strictly necessary right now.

3.2. Security Requirements

It SHALL be possible to protect the negotiated parameters in a cryptographic manner against attackers who do not have the secret authentication keys and who can't forge any of the offered integrity protection algorithms in real-time.

When the SIP REGISTER messages are used in the authentication and security setup mode process, it SHALL also be possible to protect all other information in the SIP REGISTER messages and not just the above parameters.

3.3. Characteristics Requirements

The security setup mode SHALL NOT add new roundtrips in the SIP communications to the terminal or to the home infrastructure.

The security setup mode SHALL NOT add substantial amount of contents in the SIP messages.

4. Proposal – First Hop Signalling

This proposal starts from the assumption that we will use Ericsson's "compressed" approach [4] and start the integrity protection as proposed by Nokia in [6].

4.1. Normal Case

In this section we will show the proposed message flow for the registration. In this message flow we include authentication and integrity protection in order to provide a complete example. For authentication, we use the EAP-based SIP headers with AKA as described in [2], and use IPSec ESP for integrity[1]. In order to highlight the consequences of this contribution we show the standard SIP parts in **black**, the authentication extensions in **blue**, integrity protection in **green**, and the security mode setup parts in **red**.

First, the user will send a SIP Register request to the registrar i.e. the S-CSCF to start both the authentication and the security mode setup:

REGISTER sip:… **SIP/2.0**
Authorization: **eap** *base64_eap_identity_response*
Security-setup: esp hmac_sha1, esp hmac_md5
…

In this example the terminal is indicating that it will be able to handle IPsec ESP NULL protection, either with the HMAC_SHA1 or the HMAC_MD5 algorithms. The network will process both the authentication and the security mode setup requests, and come up with a response:

SIP/2.0 401 Unauthorized
WWW-Authenticate: **eap** *base64_eap_aka_challenge_request*
Security-setup: esp hmac_md5
…

Here the network has decided that it selects the HMAC_MD5 as the protection algorithm. In our example we use the AKA authentication scheme [7] under EAP. This means that after the first message from the network side back to the terminal, the terminal will be able to conclude the authentication of the network, produce an answer, as well as generate integrity protection keys (IK). We propose that the integrity protection be turned on exactly at this point that the terminal learns the IK.

ESP(
REGISTER sip:… **SIP/2.0**
Authorization: **eap** *base64_eap_aka_challenge_response*
…
)

Note that this message contained no parts relating to the security mode setup (however, note the remark in the following paragraph); the security is now on from the terminal's perspective. The network – if it is legitimate – will be able to check the integrity protection. There is no need for the selected algorithm to be echoed back to the network, as the network must in any case remember the selected algorithm, and use that. (Note that if the used integrity protection mechanism indicates the used algorithm in the SIP packet this information may not be relied upon. For ESP, this information isn't in the packet but there could be other protocols that have it.)

The network still has to respond and indicate that it accepted the user's authentication result. At the same time we also need to verify that the network got the list of offered protection mechanisms from the terminal unchanged. Otherwise someone may have been able to delete a particular algorithm. Another way to verify the same information would be to include it into the previous REGISTER message which would imply that it is the P-CSCF (instead of the UE) who checks that the terminal capabilities sent in unprotected and protected messages are identical:
:

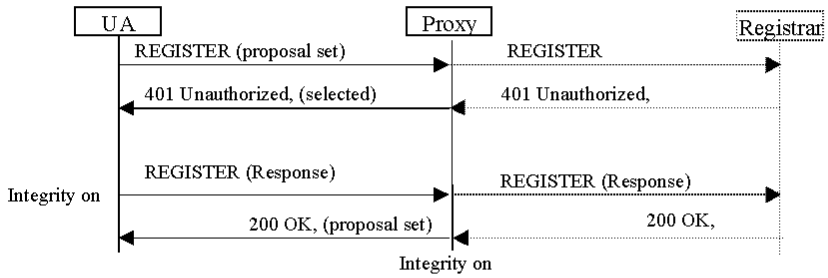
ESP(
SIP/2.0 200 OK
WWW-Authenticate: **eap** *base64_eap_aka_success*

[1] This is an example integrity protection mechanism. SA3 has not selected any mechanism for this purpose yet. Note that this choice of example does not indicate any stand-point of the contributing companies on this open issue.

Integrity-setup: esp hmac_sha1, esp hmac_md5

…

)



4.2. Failure Cases

The following failure situations need to be discussed:

- The network does not accept any of the proposals in the set offered by the UE.
- The authentication of the subscriber fails in the network due an incorrect RES.
- The UE reports a synchronization error.
- The authentication of the network fails in the UE.

The last three issues are not directly related to the security setup mode functionality as such, but still have to be discussed here since their handling may affect when and how the integrity protection can be turned on.

4.2.1. Unacceptable Proposal Set

In this situation, the network either dislikes or does not implement any of the given protocols or algorithms in the proposal set. The P-CSCF should respond as follows. The response to the first REGISTER message should indicate a failure. This can be done with a 403 Forbidden error code.

4.2.2. Subscriber Authentication Failure

In this situation, the UE sends a RES that is incorrect. This implies also that integrity check at the P-CSCF will fail since the IK and RES are based on the same process. There is nothing that the UE can do differently in this situation. Furthermore, the P-CSCF must be prepared to receive bogus messages, so the P-CSCF has really no information to determine if the received incorrect message came from the right UE (with some problems in the SIM) or from an attacker trying to confuse the registration procedure of legitimate UEs. For instance, it might be that an attacker sends a response with an incorrect RES and integrity check will fail, and a moment later the real UE sends the correct RES with proper integrity check. The first failure may not prevent the successful processing of the second message.

In order to handle this situation, the P-CSCF must implement a timer for the authentication process. When a message is received that passes the integrity checks and has the right RES, it is immediately processed. However, if during the registration timer the P-CSCF receives packets that do not pass the check, it throws them away. At the end of the registration timer, it reports an authentication failure back to the home network. It may use the messages thrown away as a clue to why the authentication failed. For instance, if no messages were received at all, the UE probably simply got disconnected. If messages with incorrect integrity check and RES were received, then it is likely due to subscriber authentication failure (but could also be an attack from outside).

4.2.3. Synchronization Failure

In this situation, the UE observes that the AUTN sent by the network contains an out-of-range sequence number. The UE could still compute an IK based on the sequence number sent by the network.

Here two alternative paths of action could be taken. One, the UE could send an unprotected message to the P-CSCF reporting the error. Second, the UE could derive an IK with the incorrect SQN (and an associated RAND) and send a protected message.

The drawback with the first approach is that a distributed denial of service attack could be launched, by sending unauthenticated sequence number failure messages through a large number of P-CSCFs. These would forward the messages to the home network which eventually discovers that the messages were illegal by inspecting the AUTS. However, at this point the home network may already be congested.

The drawback with the second approach is that the UE uses an IK which it shouldn't be using. However, since the network was still authenticated, this can only be an old IK, not a new one. Furthermore, the IK is based on RAND which changes for each use of the sequence number. Finally, the IK is not in the message but rather used just as a key for an irreversible cryptographic function. Therefore, we feel that it is perhaps the most secure approach to use integrity protection even for this message.

This means that the UE sends an integrity protected SIP message that contains an EAP-USIM-Synchronization-Failure payload. The P-CSCF sees this message, verifies the integrity, and proceeds to ask for resynchronization from the home network.

4.2.4. Network Authentication Failure

In this situation, the UE observes that the AUTN sent by the network is incorrect. Since the network and the UE disagree on the shared secret, there is no possibility to use integrity protection successfully[2]. This means that P-CSCF will not be able to conclusively decide that a particular EAP-USIM-Network-Authentication-Reject in an incoming SIP message is legitimate. Therefore, it can not immediately act on it, as a legitimate message could arrive soon.

Instead, the P-CSCF throws away the message and waits for the registration timer to end. When it ends, the P-CSCF can use the messages thrown away as a possible explanation for the failure in the authentication.

5. Proposal - Network Signalling

It still remains to be discussed how the P-CSCF selects the appropriate protocols and algorithms. This is a typical three-party policy situation where two parties (the UE and the P-CSCF) are restricted by their local policies and their capabilities, and a third party (S-CSCF) is restricted by its policies. The SA3 should discuss whether the S-CSCF must be allowed to affect the decisions or if it is sufficient for just the UE and the P-CSCF make the decision together.

However, a preliminary network signalling approach is that the P-CSCF forwards the first REGISTER reguest to the home network, and as it receives the challenge back it also gets the following additional information inside SIP to enable it to do the policy decisions and integrity checks:

[2] If the underlying security schemes require this, integrity protection could still be used simply with IK set to 0 or some other incorrect value. The end result will be the same.

- Integrity protection policy information (acceptable algorithms) from the home network.
- IK for checking the next incoming REGISTER message i.e. the REGISTER message that includes the response to the challenge from the S-CSCF.

6. Conclusions

Preliminary security analysis results in the following notes:

- It is not possible for anybody to change the selected proposal sent by the P-CSCF, because if it was changed, then the integrity check of the second REGISTER message in the P-CSCF would fail, given that the P-CSCF remembers the selection.
- It is not possible for anybody to change the proposal set, given that this is copied back in the answer to the second REGISTER message, which is already under integrity protection. It would be possible to forge this, however, if it was possible to do a real-time attack on any of the integrity check algorithms proposed by the terminal. However, this does not seem a feasible danger and is not required in 3.2
- All parameters in the SIP REGISTER message are protected, given that the second REGISTER message is already under integrity protection.

The following further work needs to be started by SA3:

- Detailed network signalling diagrams between the P-CSCF and the home network, including the need for the home network to affect the integrity policy decision.
- Further work on the details of the failure cases.
- Analysis of requirements fullfillment.
- Characteristics analysis.
- Standardization of the necessary SIP extensions in IETF. Note that if an Informational RFC status is desired, then it is sufficient to document this but not necessary to have an acceptance of the SIP WG. It is only necessary to register the header names in IANA, and to ensure that this work does not collide with any other work in IETF.

In conclusion a relatively simple and straightforward yet efficient mechanism has been discussed in this paper to perform the so called security setup mode operation in SIP. Further work remains, but the parties behind this contribution propose that this approach be selected as the working assumption.

References

[1] "The Internet Key Exchange (IKE)", IETF, RFC 2409. November, 1998.
[2] "Proposal to Use a Generic Authentication for SIP", Ericsson, S3-010263. May, 2001.
[3] "Integrity Protection for SIP Signalling", Ericsson, S3-010199. May, 2001.
[4] "Security Mode Setup Example", Ericsson, S3z010056. April, 2001.
[5] "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Security Architecture (Release 1999)", 3GPP, TS 33.102 V3.8.0. March, 2001.
[6] "Integrity Protection of the IMS Registration", Nokia, S3-010220. May, 2001.
[7] "EAP AKA Authentication". Draft-ietf-arkko-eap-aka-00.txt, IETF. May, 2001.