3GPP TR 33.821 V9.0.0 (2009-06)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Rationale and track of security decisions in Long Term Evolved (LTE) RAN / 3GPP System Architecture Evolution (SAE) (Release 9)

Keywords
LTE, SAE, Security

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

This document collects the identified threats[1] and proposed countermeasures, and includes the design choices and rationale for why proposed security mechanisms are accepted or rejected to record the history of the final security solution.

1 Scope

The scope of this 3GPP Technical Report is rationale and track of security decisions in Long Term Evolved (LTE) RAN and 3GPP System Architecture Evolution (SAE) for release 8.

Disclaimer: This TR reflects the discussions held in 3GPP SA3 while 3GPP SA3 was working towards TS 33.401. This TR is useful to better understand the basis on which decisions in TS 33.401 were taken, and the alternatives that were discussed towards the decision. Some of the text in this TR reflected 3GPP SA3's decision. However 3GPP's position on EPS Security Architecture is reflected in the normative text in TS 33.401.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non specific.
- For a specific reference, subsequent revisions do not apply.

[1] The possible attackers/intruders are hackers, operator's own personnel, third parties having access to the system, competing operators, competing vendors, criminals, ordinary subscribers (deliberately or non-deliberately), spies, etc. Motivations of attackers/intruders are espionage, violating operator's business or reputation, getting information about operators' system, business or services, just for fun, financial benefit, by mistake, to cover illegal actions, vandalism, to avoid charging, etc.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document in the same Release as the present document.

[1] "Revised Draft report of 3GPP TSG RAN WG3 meeting #50 & joint RAN WG2/RAN WG3/SA WG3 LTE meeting", S3-060119, 3GPP TSG SA WG3 (Security) meeting #42, Bangalore, India, 6 - 9 Feb 2006.

[2] "LS on the status of the study on LTE/SAE security", 3GPP TSG RAN WG3 Meeting #51, R3-060289, Denver, Colorado, USA, 13 - 17 February 2006.

[3] "Security Vulnerabilities in the E-RRC Control Plane", 3GPP TSG-RAN WG2/RAN WG3/SA WG3 joint meeting, R3-060032, 9-13 January 2006

[4] M. Zhang: "Security Analysis and Enhancements of 3GPP Authentication and Key Agreement Protocol", IEEE Transactions on Wireless Communications, Vol. 4, No. 2, March 2005.

[5] draft-haddad-alien-problem-statement-00, January 2007: "Anonymous Layers Identifiers for Mobile and Multi-homed Nodes: Problem Statement".

[6] EFF, "Cracking DES", O'Reilly, 1998.

[7] M. Wiener, "Efficient DES Key Search", originally presented at Crypto 93 rumpsession, reprinted in W. Stallings (ed), Practical Cryptography for Data Internetworks.

[8] S. Kumar, C. Paar, J. Pelzl, G. Pfeiffer, A. Rupp, and M. Schimmler, "How to Break DES for € 8,980", SHARCHS 2006 workshop, http://www.ruhr-uni-bochum.de/itsc/tanja/SHARCS/start06.html

[9] I. Devlin and A. Purvis, "Assessing the Security of Key Length", SASC 2007 workshop.

[10] IETF RFC 4270: "Attacks on Cryptographic Hashes in Internet Protocols"

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

aGW	Access Gateway
AS	Access Stratum
(D)DoS	(Distributed) Denial of Service
eNB	Evolved Node-B
LTE	Long Term Evolution
MAC	Media Access Control
MME	Mobility Management Entity
NAS	Non Access Stratum
PDCP	Packet Data Convergence Protocol
RAN	Radio Access Network
RB	Radio Bearer
RRC	Radio Resource Control
SA	Security Association
SAE	System Architecture Evolution
SMC	Security Mode Command
UE	User Equipment

4 Layered Security Approach in LTE

The general direction in the LTE security has been to separate the security between AS (RRC security in eNB) and NAS signalling, as well as to terminate the user plane security above eNB. The requirement is also that the radio link and the core network must have cryptographically separate keys.

The result is that LTE system has two layers of protection instead of one layer perimeter security like in UTRAN. First layer is the Evolved UTRAN (E-UTRAN) network (RRC security and User plane protection) and second layer is the Evolved Packet Core (EPC) network (NAS signalling security).

The design target has been to minimize the effects of the compromised E-UTRAN security layer (1st) to the EPC security layer (2nd). This principle improves the overall system security and allows placement of eNBs into more vulnerable locations without high risks for the operators. It also makes the overall system security evaluation and analysis easier in case of multiple access technologies connected to the EPC. However, care must be taken when designing the interface between these two security layers, namely the S1-C and S1-U interfaces.

In case attacker is able to compromise the first security layer, the second layer is not compromised. However, it is important to evaluate how the compromise of the first layer affects the whole SAE/LTE system security. The goal is to make this effect low and local so that the risk of compromised first layer is as low as possible. As a result, the use case of a home eNB (identified scenario in LTE) becomes more realistic as well.

The S1 interface (consists of S1-C and S1-U), is the point where the two security layers interact (see Figure 1). Careful design must be applied for this interface to disallow high security risks because of possibly partially compromised first security layer. Thus, particularly the messages from eNBs towards the EPC network elements should be properly analyzed from security perspective. The threat to think about is to see what an attacker can do if she/he can send whatever S1-C/S1-U messages on behalf of a legit eNB.

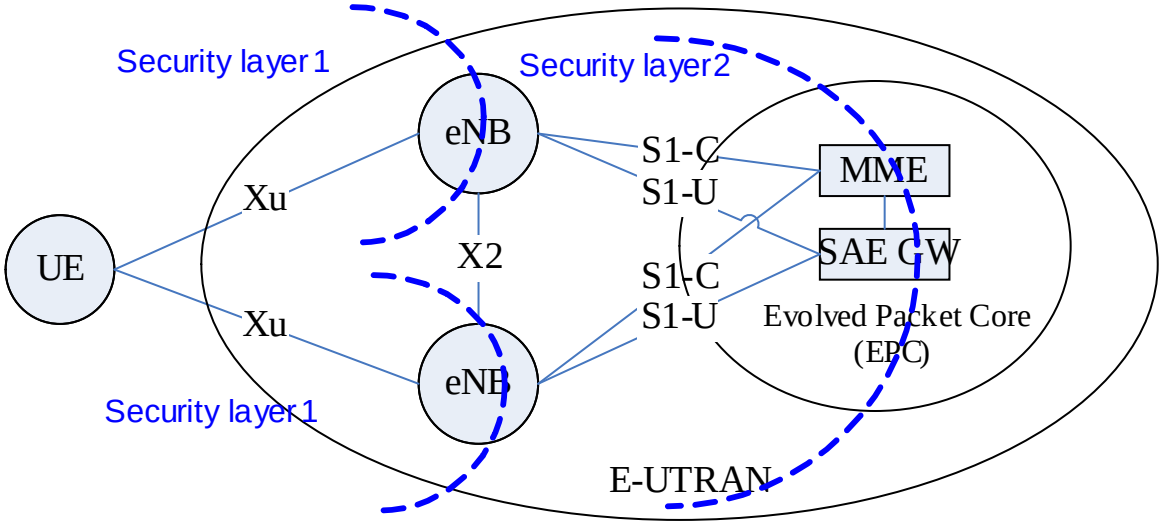


Figure 1 First and second security layers in LTE

4.1 S1-C interface security

For the interface between eNBs and MMEs (S1-C), NDS/IP or similar solution is used. SA is needed because the MME will provide confidential information like RRC keys and user profiles for the eNBs. This SA is independent of the first layer from security perspective.

Security analysis should be made for the messages originating from eNBs towards MMEs.

Security analysis should also be made for the key management inside the eNBs to minimize risk of compromised keys.

4.2 S1-U interface security

Security analysis should be made for the messages originating from eNBs towards SAE GW, if any.

4.3 Example case: Direct Path Switch Message Security

There is a proposal to use direct path switch message from target eNB to the SAE GW for improving the handover performance (see contributions to RAN WG3). It is assumed path switch messages need to be integrity protected .

In case of eNB compromise, attacker can send false path switch messages towards SAE GWs.With the S1 flex interface, the number of eNBs one SAE gateway can control can be quite high. On the other hand one eNB can have connections to many SAE gateways based on the nature of the interface. Thus the impact of this attack can be high,

making the risk of the attack high as well. Also, the attack is easy to launch as it requires only one message to the SAE gateway per UE. Note that the attacker can also blindly generate the messages towards multiple SAE gateways.

In case the attacker resides on the S1-U interface and there is no NDS/IP or physical security on S1-U, the result can be that the SAE gateway and MME loose synchronization of the UE's location (assumed that the S1-C interface is secure). In best case the MME may resynchronize the SAE gateway and the UE's route is correct again.

In a more severe cases, the eNB is compromised (compare to the eNB in vulnerable locations), meaning that the attacker resides in the eNB, and can send arbitrary messages towards both the MMEs and the SAE gateways. In this case it is hard for the MMEs and SAE gateways to detect if the messages are sent because UE has moved or because the eNB is compromised.

Solution 1: Use NDS/IP between eNBs and SAE gateways for the path switch message only. This may be hard to achieve in case the path switch message is considered to be in-band signalling. Also, managing a separate SA for the path switch message only may not be cost efficient. This solution, however, does not protect MMEs and SAE gateways against the compromised eNBs.

Editor's Note: This solution does not involve UE as compared to solution 2, thus this solution is only partially solving the problem.

Solution 2:.To mitigate the threat of forged path switch messages from a malicious or compromised eNB (i.e. eNBs which do not serve the UE), the target eNB could use UE related keying material to create authentication information for the path switch message (e.g. integrity protect the path switch message). Alternatively MME can provide UE specific one-time integrity key for the serving eNB, which can be used to sign the path switch message sent for the SAE GW. In either case the SAE GW can then verify that the target eNB sending the path switch message is actually serving the UE (e.g. has keying material related to UE). However, this requires that the SAE GW is able to verify the message authentication parameters. In case MME provides one-time authentication information, both SAE GW and MME can share a long term key which is used to create the token or verify the signature of the path switch message. For example MME can derive a UE specific key from the long term key using UE identity and a nonce (or sequence number) and provide the key and nonce (or sequence number) for the serving eNB. eNB then signs the path switch message with this key and includes the UE identity information and the nonce (or sequence number). SAE GW gets the message and derives UE specific key based on the message and/or UE context information, e.g. UE identity information and nonce (or sequence number). As a result the threat of forged path switch messages towards the SAE GWs is mitigated as only eNBs having the UE related authentication keying material can send the message to the SAE GW, even if the security of layer 1 (see Figure 1) is compromised. This solution is specific for direct messages between eNBs and SAE GWs. For each authentication token or message a fresh nonce or sequence number is needed to achieve replay protection.

4.4 Conclusion

Using security layer 2 keys for protecting messages affecting UEs between eNBs and the EPC is considered to be an implicit follow-up security requirement for the LTE system. It also makes the risk of compromised eNBs lower and localized. This means that eNB placement into vulnerable locations is more practical deployment scenario.

However, since the User Plane ciphering termination is performed in eNB, rather than in EPC, security between the eNB and the EPC should also cover the path switch message. Thus, NDS/IP on the backhaul link and secure eNB implementations are considered to be secure enough solutions for the path switch message protection. The complexity of solution 2 is not justified compared to the threat.

The path-switch message requires integrity protection to counter attacks where an attacker forges path-switches on behalf of eNBs. Since the move of user plane encryption from the EPC to the eNB, this may imply (see discussion in section 8) that S1-U needs to be encrypted (to prevent backhaul link threats). In this case, adding integrity protection on some (or all) messages between eNB and SAE GW adds close to no overhead in terms of establishing the integrity key (but of course adds overhead in terms of bulk processing).

Editor's Note: Any dependence in security between application layer and bearer layer (air interface user plane) will cause more complexity in the system than what gains. An example of such a dependency would be the deactivation of user plane ciphering when the user applies Application layer ciphering (or vice versa).

Editor's Note: The gain in the processing power and the storage of using shorter keys and less secure algorithms is a tradeoff that we do not believe in. (S3-070031)

5 Threats

5.1 Threats to UE

5.1.1 IMSI catching attack

5.1.1.1 Threats

International Mobile Subscriber Identity (IMSI) consists of Mobile Country Code (MCC), a Mobile Network Code (MNC), and a Mobile Subscriber Identification Number (MSIN). The total maximum length of IMSI is 15 digits, where MCC is 3 digits and MNC 2 or 3 digits depending on the area.

From subscriber's privacy point of view, the MSIN (also IMEI) identifies the subscriber and thus should be confidentiality protected. However, the subscriber's credentials can not be fetched before the subscriber has been properly identified. With the UTRAN AKA authentication method the network can not be authenticated before the user provides identification to the network. This is a reason why in UTRAN the UE can not deny plain text IMSI queries from the network.

- Attackers can utilize this hole by collecting IMSIs in an area or place (e.g. in airport). History information of seen IMSIs (or IMEIs) in some areas or places is considered to be confidential as whenever the IMSI is mapped to a user identity, the user's movements and presence can be tracked automatically (back in time or in the future).
- The IMSI provides a globally unique user identifier that even provides further information like home network and home country. For internet service usage, global unique identifiers are seen critical by the European Union [1], since unique user identifier allows the matching of user preferences and profiles from different sources.
- Service identifiers are today sent in clear, since the user identity can be revealed by the IMSI already.
- IMSI catching in the mobile environment might be considered a quite expensive exercise, but in the near future with the network convergence and smart cards, that can be connected directly to the PC e.g. via USB stick, this picture changes. If the device access the network via fixed network or WLAN, then the IMSI needed as a baseline for future service usage might be send in clear to an unauthorized requesting entity e.g. a specially configured WLAN access point.
- In roaming scenarios, the roaming partner receives the IMPI from the user. It is unclear, if all the roaming partners can be considered as trusted in this sense also in the future.

5.1.1.2 Countermeasures

To mitigate this threat, UE must be able to reject plain text IMSI queries coming from an untrustworthy source. This way the UE has control over when to send the plain text IMSI to an unauthenticated network or source. Public key cryptography or symmetric keys may be used to hide the IMSI.

A mechanism similar to TMSI mechanism in UMTS may be used. User permanent identity is rarely used. Temporary identity is often used to identify the user. Temporary identity is allocated by network. The procedure of allocating temporary identity should be provided confidential protection.

(The following paragraphs are from S3-060646)

Securing the IMSI so that an attacker can not get it over the air interfaces is important and provides improved security over UTRAN. This can be achieved if the UE has a key, which it can use to encrypt the IMSI (or MSIN part of it) before sending it to the network. This can also be achieved if the UE has a pseudonym that is assured to be understandable by the home network at least (corresponding IMSI can be identified).

One natural way to incorporate IMSI protection is to extend the identity request and response messages (see TS 33.102) for LTE to include an option to support one or more IMSI protection mechanisms. UE can for example provide the identity in an encrypted form that is then denoted in the identity response message. Alternatively the network can also denote in the identity request message that it supports encrypted IMSIs in a backwards compatible way and thus the

support of IMSI protection can be implemented into existing UTRAN networks as well for terminals that support IMSI protection.

Editor's Note: More studies are needed on how to support emergency calls.

Solution –A) Public Key Based Approaches

In a public key based approach the UE uses a public key to encrypt the MSIN part of the IMSI and provides it to the network. Either visited or home network then decrypts the MSIN part of the IMSI and uses the plain text IMSI to get proper authentication vectors from corresponding HLR.

Traditionally the public key comes in a form of a certificate, as has been the case with the 3GPP General Bootstrapping Architecture (GBA). The public key certificate needs to be provisioned for the UE before it can use it and reject the plain text IMSI queries. UE must also be able to authenticate the public key certificate. This has been considered a problem earlier, but once solved within the scope of the 3GPP GBA the same approach can be used for IMSI protection as well, which makes the IMSI protection solution a nice side product of the 3GPP GBA standardization.

Here the user identity request and response message procedure could be extended to support visited/home network certificate provisioning for the UE during the initial attachment. UE could then verify the received certificate based on the same principles as in GBA.

Alternatively the AKA quintet could be marked to be "used in home network only" based on a small extension into the AMF field in the quintet (one bit standardized to denote "used in home network only"). This would in effect allow the UE to authenticate the home network from any other networks, which is not possible with AKA today as it does not bind the quintets to the access network identities. Using the secure NAS signalling path to the home network UE could allow it to update any public keys or certificates for it if needed. This update for the AKA protocol would not need any changes in the USIM, but the UE could check whether the AMF field contains "home network bit" set to true or false.

Extending the AMF field would also allow other potential applications to be built on top of the secure connection with the home network and is independent of the LTE. The requirement is of course that the HLR does not provide quintets with the "home network only" AMF bit set to true for any visited networks. Thus, this extension is also transparent for the access networks, but requires a small HLR internal update. However, the protocols between HLR and other network elements should not need to be changed (provided that the HRL can distinguish home network elements from visited network elements based on the SAs).

Alternatively the UE could also use GBA Subscriber Certificates for IMSI encryption.

Solution-B) Pseudonyms Based Approach

In a secure pseudonym based approach (G. Ateniese, A. Herzberg, H. Krawczyk, and G. Tsudik, On Traveling Incognito, in journal of Computer Networks (31) 8 (1999) pp. 871-884, April 23, 1999), the HLR is modified in such a way that the quintet itself includes a pseudonym inside the AUTN parameter which the UE is able to authenticate based on its long term key in the USIM and shared with the HLR. When UE next time needs to provide identity and it does not share a P-TMSI with the network, it can use one of the pseudonyms from the previously authenticated AUTN parameters.

One way to create the pseudonym in the home network (HLR) is to have a secret key K with key identifier KID in HLR, which uses it to create the pseudonyms by encrypting the IMSI and some random variable with the key. This pseudonym is then put with the key identifier into the AUTN parameter.

The key K does not have to be transferred outside the HLR. Once HLR gets authentication request based on a pseudonym, it decrypts the pseudonym with the secret K associated with the key id KID and identify the full IMSI.

The pseudonym creation and the pseudonym based IMSI identification are HLR internal procedures. The exact algorithms used to create the pseudonyms do not affect the network and UE implementations as the pseudonym can be considered a bit string, possibly with a variable length.

To support this, the USIM in the UEs must be upgraded to support secure pseudonyms as the AUTN parameter in AKA is extended. Thus, we believe that this alternative is not feasible in practice unless there are other more important reasons to upgrade the AKA protocol, USIM and the HLR implementations.

Editor's Note: If users are moved between HLRs, pseudonyms based approach may have problems.

5.1.2 Threat of UE tracking

5.1.2.1 Threats

A) Tracking User temporary ID

Even though it is not yet settled how temporary RAN identifiers are going to be used in LTE, it is close to certain that some thing along the lines of U-RNTI used in UMTS will be present. Depending on the security mechanisms applied to the assignment of these identifiers, it may be possible to track users.

There are two main threats to consider:

1) The attacker is able to track (and record actions taken by) a UE as it moves between eNBs, but cannot immediately determine the user ID from the temporary ID(s). At a later stage the UE may reveal information (e.g., it connects to a web-service owned by the attacker where the user is required to give his name). When this happens the attacker can correlate the temporary ID with the user's name, and will be able to deduce that the user performed the actions previously recorded.

2) The temporary ID is assigned in such a way that the attacker immediately can correlate the temporary ID to the user's ID. For example, the user reveals his IMSI during the attachment procedure, and gets the temporary identity assigned in the clear. UMTS has the possibility to re-assign the temporary ID after confidentiality protection is activated, which counters this threat.

Editor's Note: There is other info other than ID which may give possibility of tracking.

B) User tracking due to Linkability of IMSI/TMSI and RNTI

A disadvantage of the 2G/3G temporary user identity confidentiality scheme is that a false network/eNB can always claim to have lost the TMSI and can ask the UE to reveal the IMSI upon registration. This will allow an attacker to record the usage of all (temporary) identifiers at the air-interface and then backwardly trace the UE behaviour when he succeeds in getting the IMSI correlated to the current TMSI. This attack may be difficult to prevent (See Section 3.1 IMSI catching) (only the successfulness to re-construct a UE's behaviour backwards in time can be limited. Essential to this is that the RNTI shall be unlinkable to the TMSI for an outsider.

In state LTE_IDLE and LTE_ACTIVE there exists a security association between the UE and MME, which can be used for protecting TMSI reallocations. But in LTE_IDLE the eNB does not possess a security association with the UE. The TMSI needs to be disclosed every time the UE has to contact MME from state LTE_IDLE (RNTI or similar identifier cannot be used to identify the requesting user to the MME).

This means that a passive attacker may be able to link the user's behaviour between different active sessions when TMSI is kept fixed, following an unexpected IMSI-TMSI disclosure by the network. The active attacker does not need an accidental IMSI-TMSI disclosure but can remount his attacks again during each next idle period.

C) User tracking due to IP-address linkability towards TMSI/IMSI/RNTI

The SAE gateway stores a UE context, e.g. parameters of the basic IP bearer service, keeps network internal routing information. The MME can store the UE context for long to allow for (re-)registration with temporary identity (user identity confidentiality). Within LTE the user gets an IP-address from the moment the registration (and authentication) has been successfully performed.

TR 25.813 V101 of table 10.1 currently describes within a NOTE that the protocol stack layer in which the ciphering takes place is FFS.

Assumed that user plane ciphering would be done at IP level than the initial assigned IP-address (allocated by confidentiality protected NAS signalling (requires SAE gateway/MME cooperation)) would be disclosed when starting data transfers.

Editor's Note: It needs to be checked whether IP-addresses will be sent in clear text or not.

When the IP-address would be kept static for a long time, it could allow the passive attacker to correlate reallocated TMSI with these static IP-addresses, and this would weaken the TMSI re-allocation scheme.

AS the User plane ciphering is being performed below/integrated to the PDCP layer, cf. section 4.3, there is no need to require frequent IP-address allocation as the IP-packets are tunnelled and encrypted within 'PDCP-ciphering'. This

also means that IP-address privacy mechanisms need not be used (e.g. MAC addresses in IPv6). However the identifier that is being used within 'PDCP' should then be re-assigned at least as frequently as the TMSI re-allocation.

NOTE: With user plane ciphering not activated, the passive attacker is not only able to observe the IP address of a user but might also be able to observe application layer identifiers, and as such be able to bypass TMSI-IMSI secure reallocation mechanisms.

D) Tracking based on new and old RNTI mapping

SA3 was notified in S3-060341 that C-RNTI will be used to identify a UE:

- The C-RNTI provides a unique UE identification at the cell level.

- It is assumed that this identity is used for scheduling unless the cost would turn out to be too high and the introduction of a separate MAC-Id is required.

RAN2 has agreed that C-RNTI is pre-allocated in the target eNB and transferred to the UE in Handover Command (see R2-061714). This means that a passive attacker can link new and old C-RNTIs together unless the allocation of C-RNTI itself is confidentiality protected.

E) Tracking based on handover signalling messages

Serving eNB commands UE to a target eNB with Handover Command message. UE sends Handover Confirm message to the target eNB. A passive attacker can map these messages together and conclude that a UE has changed eNB. This is just an example of what information an attacker can deduce from the RRC messages, which are not confidentiality protected. Note that identifying messages based on small differences in the message lengths is not obvious or most probably not even possible as the packets are sent in full frames etc.

F) Tracking based on cell level measurement reports

UE sends cell level measurement reports to the eNB within the RRC protocol. A passive attacker listening to the measurement reports from UEs can follow UE's movements based on the reports and track the position of the UEs more accurately than the information of current cell location. Note also that the location/position based services may be based on the cell level measurement reports.

G) Tracking based on packet sequence numbers

If the user plane (RLC, PDCP) or control plane (RRC, NAS signalling) packet sequence numbers are continuous it is easy for a passive attacker (listening) to follow UEs with high possibility based on the packets only (i.e. following the sequence number sequences).

A passive attacker can listen to user and control plane (AS and NAS) packets and track the UE based on the continuity of the packet sequence numbers between handovers or idle-to-active mode transitions.

H) Tracking based on UE's static IEEE MAC (Medium Access Control) address

If the UE is able to have access WLAN, the attacker may be able to track the UE based on its static IEEE MAC address, e.g. the attacker can record the MAC address at a certain hotspot and know a certain UE appears in the range. The attacker can know the victim's habit if he can match the MAC address with some high layer identities of the UE. Furthermore the attacker can track the UE from location to location. The detail of this kind of threat can be referred to [5].

5.1.2.2 Countermeasures

A countermeasure against these attacks is to confidentiality protect the assignment procedure of the temporary identities. Note that to fully counter the threat, it may also be necessary to confidentiality protect the measurement reports from the UE to the NW, since otherwise an attacker can predict that the UE is about to handover to a new cell, and then follow the UE to the new cell.

There are other ways than ciphering all NAS signalling messages. Several alternative solutions are listed below:

Editor's Note: The solutions 1 and 2 below are only a secondary choice under the assumption that there is no NAS confidentiality protection when sending temporary NAS identities. Similar countermeasures may be also used for RNTI.

Solution 1:

Before generating a new temporary identity, network should share keys with user. These pre-shared keys are used to deduce a key which is used to cipher the user temporary identity. Some fresh parameters should be included in procedure of deducing key to ensure the freshness of key. In message of allocating user temporary identity, a ciphered temporary identity is sent. Fresh parameters are also sent. UE uses pre-shared keys and fresh parameters to deduce the key and use it to get the user temporary identity.

Solution 2:

Before generating a new temporary identity, network should share keys with user. A new temporary identity is deduced by using pre-shared keys both in network and UE. Some fresh parameters should be included in procedure of deducing key to ensure the freshness of user temporary identity. In message of allocating user temporary identity, only fresh parameters are sent.

Editor's Note: Since the length of TMSI is short, there may be hash collision. The value of fresh parameter needs to be clarified.

1) Countermeasures against unintentional disclosure of IMSI by UE and MME

Requirement-1: The TMSI on initial and re-allocation by the MME shall be transferred via NAS signalling (confidentiality and integrity protected) towards the UE.

Requirement-2: The MME shall store the TMSI sufficiently long after user de-registration (transition to LTE_DETACHED) or Tracking Area-update time-out, in order for the user to be able to register again with TMSI.

Editor's Note: The time for the MME to keep the TMSI value is implementation dependent. It needs to be clarified what is sufficiently long.

Requirement-3: The UE shall give priority to use the last received TMSI over IMSI/IMEI when identification towards MME is needed.

2) Countermeasures against tracking a user between different LTE_ACTIVE and LTE_IDLE sessions

In order to prevent that a currently valid RNTI (which may be allocated insecurely) cannot be linked to the future TMSI i.e. via TMSI disclosures via MM-signalling in LTE_IDLE (e.g. periodic TA update) after the transition from LTE_ACTIVE to LTE_IDLE, it is necessary to perform TMSI reallocation after having activated NAS ciphering by the core network.

Requirement-4: The TMSI shall be re-allocated after each transition to LTE_ACTIVE transition when having activated NAS-security (and shall be transported confidentiality protected to the UE).

Editor's Note: It needs to be investigated whether it is sufficient to reallocate the TMSI on each cell change (rather than change to LTE_ACTIVE) in order to reduce NAS-signalling overhead. It needs also be studied how frequent these transitions can be.

In this case the RNTI can only be linked with the clear text TMSI used within the MM-procedure that initiated the previous state transition to LTE_ACTIVE. This prevents backwards traceability as the attacker cannot ask the IMSI related to the old TMSI anymore.

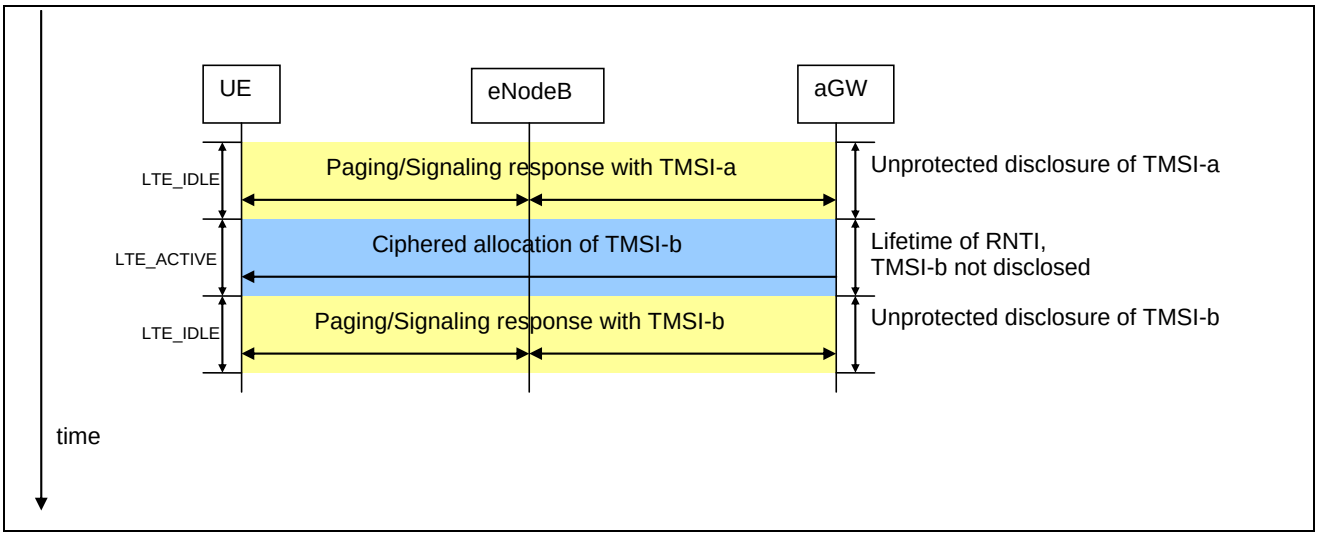


Figure 2 IMSI re-allocation in Time

The requirement 4 will result in isolation of the effects of user traceability against the passive attacker on accidental IMSI disclosure (e.g. TMSI mismatch)[2].

Restriction: The active attacker however can successfully retry after the user enters LTE_IDLEs state, after the first MM-signalling (e.g. Tracking Area Update) that needs to be identified by a TMSI and ask the user to identify himself with IMSI. This will allow the attacker tracing the user's behaviour during the next LTE_ACTIVE period assuming the RNTI allocation is not secure. The attacker will not be able to trace the user behaviour passively after that period without remounting the active attack.

Another countermeasure is to disallow the IP-address visibility. But if IP-addresses are exchanged in clear text then the reallocation of the IP-addresses shall be of a comparable frequency as the TMSI-reallocation.

Editor's Note: Frequent IP-address changes may have undesirable affect on the layers above IP.

3) Countermeasures against user tracking via RNTI during LTE_ACTIVE

A secure RNTI reallocation mechanism might further help in limiting the traceability of a particular user. It needs to be investigated whether the complexity that comes with it, warrants an increase in ID-confidentiality. An active attacker can use the LTE_IDLE state for his attacks. A passive attacker needs to take advantage of accidental IMSI disclosure. Under these circumstances it may be acceptable that the RNTI is transported and allocated without requiring confidentiality protection.

There exist several secure RNTI re-allocation solutions, with different complexity. It is thereby assumed that the assignment of an initial RNTI (could also be an initial MAC-ID) is being performed by the eNB before it is possible to confidentiality protect the transport of the RNTI to the UE. Following two alternative countermeasures therefore are intended for the secure reallocation of the RNTI.

a) Use of RRC encryption: In that case the RNTI could be re-allocated after activation of air-interface security and transported confidentiality protected to the UE. (This concerns both the state transitions from LTE_IDLE and LTE_DETACHED to LTE_ACTIVE).

b) Use of a derivation function at both the UE and the eNB to derive a secret subsequent RNTI that can be used without having to transfer the new RNTI-value. A potential problem with this is that collisions have to be avoided when generating the new value as the RNTI[3] has a limited length. This can be prevented by using a RAND that is chosen by the eNB, potentially going through some iteration by re-choosing RAND at eNB, in order to generate an unused RNTI value. Such a derivation function may be: new RNTI = HASH (old RNTI, RRC integrity key, RAND) and needs to be implemented on the ME en eNB.

---

[2] It's assumed that an attacker (excluding compromised eNB's) is not able to ask MME for the IMSI related to a traced TMSI within LTE_ACTIVE as MM-signalling shall be integrity protected on NAS level. Similarly the UE should not answer a paging request with IMSI or TMSI while in state LTE_ACTIVE . The newly assigned TMSI is therefore protected from disclosure via an active attack during the LTE_ACTIVE session. It's assumed that the protected MM-signalling during LTE_IDLE is routed towards NAS via the eNB on the basis of an internally linked RNTI-TMSI table (S1-interface).

[3] This also assumes that the RNTI is not structured.

Editor's Note: These solutions would potentially help to defend against the threat where a person is first passively identified and located, and then his position is tracked via used radio identifiers.

4) Countermeasures against UE tracking based on the sequence numbers

It is proposed to have RRC ciphering, similar to the UTRAN:

a) RRC ciphering prevents attackers from mapping RRC messages together during handovers (like "Handover Command" with "Handover Confirm")

b) With RRC ciphering new C-RNTI, which is transferred in the Handover Command message can not be linked to the old/current C-RNTI

c) With RRC ciphering an attacker can not track the UE based on the cell level measurement reports

5) Countermeasures against UE tracking based on packet sequence numbers

The packet sequence numbers must NOT be continuous over the air between handovers and possibly also between idle-to-active mode transitions:

The user and control plane packet sequence number sequences must not be continuous over handovers and idle-to-active mode transitions in the over-the-air signalling. The sequence number must be continuous for the ciphering function during a key lifetime. Thus, one possible solution is to use a random offset to make the user and control (AS and NAS) plane sequence numbers discontinuing in the over-the-air signalling. These random offsets are selected by the eNBs and carried along with the new C-RNTI to the UE via source eNB during the agreed handover procedure.

There are comments on this countermeasure as:

Attacker may not trace the user successfully by listening packet sequence.

If it is decided that sequence number should be discontinuous, above solution is just an alternative solution. There may be other solutions to mitigate the risk.

If random offset solution is used, it can only be concluded that C-RNTI and random offset should be confidentially protected. Both C-RNTI and random offset are not long in size. Some solutions can be used to provide confidential protection to C-RNTI and random offset rather than ciphering all RRC signalling.

The result from 3.2.2.4 and 3.2.2.5 is that a passive attacker can not track/follow the user based neither on control nor user plane packets.

6) Countermeasures against UE tracking based on static IEEE MAC (Medium Access Control) address

A possible countermeasure is to use fresh MAC address when accessing WLAN. UE could choose a new MAC address randomly or the network could choose for the UE (similarly to C-RNTI allocation in LTE). In case UE chooses a random MAC address, the collisions should be handled properly (e.g. a recovery mechanism). The countermeasures should be provided/designed by IEEE. However, IEEE has already indicated in S3-060762 that they may not provide any such countermeasures.

5.1.3 Forced handover

5.1.3.1 Threats within LTE

Threat 1:

In this threat we assume that the attacker is in possession of the currently used RRC keys because UE has previously been connected to the compromised eNB and the RRC keys have NOT changed since then.

The compromised eNB sends a false handover command message on behalf of its currently serving eNB to UE commanding UE to hand over to

a) The compromised eNB, which then drops the connection to UE.

b) Another eNB within the same SAE/LTE access network that is not prepared to handle UE, which will again make the UEs connection drop.

In both cases UE is denied service.

Threat 2:

A compromised eNB sends a powerful signal so that all UEs in its vicinity are handed over to the compromised eNB. Once the HO is complete, the compromised eNB drops the connection. As a consequence all UEs in the vicinity of the compromised eNB are denied service.

5.1.3.2 Countermeasures

For threat 1:

The attacker is only able to address UE when connected to another eNB if he knows the RNTI assigned to UE. If the RNTI is assigned with NAS involvement, an attacker in possession of the RRC keys does not have access to the assigned RNTI unless he can guess it from time-relations or because there is a limited range of RNTIs. It is important to note that the RNTI assignment is not decided upon. However, it may be of interest to introduce an RNTI assignment in two steps such that an initial temporary RNTI is assigned without NAS involvement and then a more permanent RNTI is assigned with NAS involvement after the NAS security is established.

Even if the attacker is in possession of the RNTI and the currently serving eNB drops the connection to UE, UE will try to establish a new connection with the best available eNB. In case the same RRC keys are used after the establishment of the new connection the attacker may be able to repeat the same attack several times. In case new RRC keys are used on a non-compromised eNB after the establishment of the new connection, the attacker cannot mount the attack again.

Furthermore, the above attack requires the attacker to send an individual false handover command message to each victim UE. As opposed to this a jamming of the corresponding radio frequencies of the currently serving eNB would affect all UEs in its vicinity at once.

The attacker can indeed extend the scope of his attack beyond a compromised eNB under his control, but the extension is fairly limited as the users must have been attached to the compromised eNB at one time. NAS involvement in the RNTI assignment would help to mitigate Threat 1, but may not completely prevent it.

For threat 2:

Threat 2 has a similar effect as Threat 1 as UEs are denied service. However, possible victims of the attacks previously described are only UEs that were at some point connected to the compromised eNB and the attacker has to explicitly address each victim UE. As opposed to this all UEs that are currently in the vicinity of the compromised eNB are possible victims of threat 6 and all of them can be denied access at once.

Threat 2 is one example for a threat that cannot be mitigated by the use of separate keys, but seems to be easier to mount and more effective than Threat 1. Furthermore, the use of separate keys seems much more complex than the use of common keys. As threat 6 shows, the security gain seems to be quite limited, which speaks in favour of using common keys.

5.1.4 Forced handover to legacy RAT[4]

5.1.4.1 Threats

An attacker may force an LTE UE that also supports legacy RAT to perform a handover to a legacy RAT with weaker security. The problem can be described as follows, cited from [3]:

"An attacker with the ability to generate RRC signalling—that is, any of the forms of compromise listed above—can initiate a reconfiguration procedure with the UE, directing it to a cell or network chosen by the attacker. This could function as a denial of service (if the target network cannot or will not offer the UE service) or to allow a chosen network to "capture" UEs.

An attacker who already had full control of one system (perhaps due to weaker security on another RAT) could direct other systems' UEs to "their" network as a prelude to more serious security attacks using the deeply compromised system. Used in this way, the ability to force a handover serves to expand any form of attack to UEs on otherwise

[4] This section is from S3-060200.

secure systems, meaning that a single poorly secured network (in any RAT that interoperates with the E-UTRAN) becomes a point of vulnerability not only for itself but for all other networks in its coverage area."

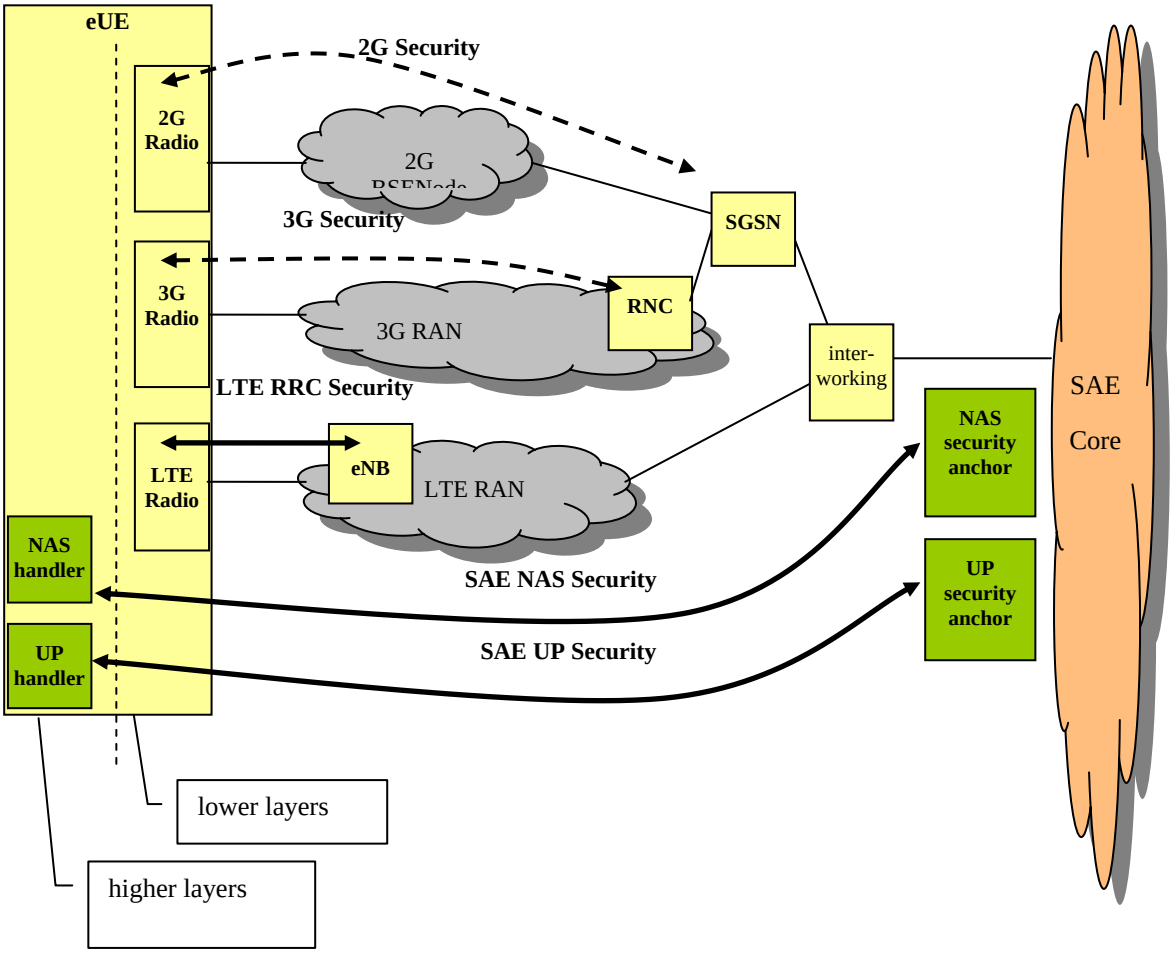


Figure 3 Forced handover to legacy RAT

5.1.4.2 Countermeasures

Two of the three security associations agreed for LTE/SAE are independent of the radio layer: the Non-Access-Stratum signalling and the User Plane security (NAS, UP). If usage of (NAS, UP) security is not confined to LTE-RAT access only, a handover attack will lose much of its attractiveness to an attacker. Even after compromising the radio layer security, an attacker can not send or eavesdrop UP traffic and NAS signalling, because they are protected by an additional security layer.

So LTE/SAE UEs will benefit from security enhancements, independent of the RAT they use to connect to the 3GPP system. Legacy 2G/3G UEs are not aware of the new NAS and UP security associations and continue to rely on their bearer-specific security only.

In order to counter the forced handover attack in the described way, an architectural decision must be made that allows a UE to utilize (NAS, UP) security over legacy RATs. This means that the NE that terminate the respective security associations must be above the interworking point with legacy RAT. Figure 3does not assign these security anchors and the interworking point to the LTE RAN or to the SAE CN, because discussion on their assignment is still ongoing in SA2.

Editor's Note: This is only one of the possible countermeasures. The architecture of SAE/LTE isn't known well. This countermeasure may affect mobility between different RAT. More explicit description of the threat is helpful. There already exists solution to enhance legacy RAT security.

5.1.5 Threats of unprotected bootstrap and multicast signalling in LTE

5.1.5.1 Threats

In UTRAN there is no protection of information received from the network before the security mode command, i.e. the bootstrap signalling is unprotected. Similarly, information which is sent from the network in a point-to-multipoint fashion, e.g. information triggering hand-over to other eNB while the UE is idle or information such as the GROUP_RELEASE command from the RNC lacks protection.

Protection of such signalling seems to require either:

- Public keys associated with RAN nodes and use of signatures,
- Source origin authentication schemes such as TESLA, [1], or,
- Other forms of "tailor made" symmetric key based solutions for specific problems, e.g. [2].

The threats associated with not using such protective measures seem mainly to be of DoS aspects, i.e. the UE will be fooled into camping on a false eNB, or, the UE would be detached from the network, etc. However, at the same time, the effects of these DoS attacks are more persistent than "radio jamming" attacks, as the UE will e.g. loose paging until the user/UE actively triggers an outgoing call. Thus, this sort of DoS attack is not completely persistent, neither is it exactly non-persistent.

Editor's Note: The former policy is only against persistent DoS, not non-persistent DoS. This is semi-persistent DoS. All the possibilities to identify threats related to broadcast and multicast should be identified in order to affect the design of the system. "GROUP_RELEASE" command is from the RNC and may be not needed in LTE/SAE.

5.1.6 Threat related to broadcast of system information

5.1.6.1 Threats

Broadcasts of SYSTEM INFORMATION are not protected in UMTS. If an attacker can imitate the network behavior and broadcast a set of SI, i.e. master information block, scheduling blocks and system information blocks having the same value tag and identities as in the current network, he can manage to introduce wrong SI parameters / predefined configurations to the UE.

The attacker should use false base station or produce standalone broadcast signaling, masking under the cell-ID of a real neighboring cell or different cell-ID, and transmitting with higher power than the real cells or in vicinity of the UE.

There are following basic ways, how a potential attacker can introduce wrong System Information (SI) (predefined configurations, other parameters such as measurement configurations, constants, counters, etc.)

- After the UE switch on / enter new PLMN using broadcasting wrong SI under the correct value tag. In LTE switching UE off/on might happen not very frequently, but vulnerable places such as airports should be considered.
- By paging all UEs using unprotected powerful paging message (similar to PAGING TYPE 1 in UMTS) by indicating that the SI is changed. Once all UEs have read the SI the attacker can change back the value tag to the value tag of the real neighboring cells and this time introduce wrong SI.
- Introduce some wrong parameters on the SI, which UE reads every time when entering new cell (scheduling blocks, SFN, value tag, PLMN Id, measurement Cell IDs, cell access restriction parameters); the effect of this might be the same as camping on a false BS, and may result in detach of UE from the network.

For example in UMTS network predefined configurations are broadcast in SIB type 16, which has multiple occurrences for each predefined configuration. Different parts of the system may provide the UE with one or more predefined UTRAN configurations, comprising radio bearer, transport channel and physical channel parameters. The UE should

store all relevant IEs included in the SIB. The availability of predefined configurations is communicated to the network during the call establishment, and thus, if available the network relies on this information instead of transmitting the complete configuration to the UE.

Thus a potential attacker can send some wrong PhyCH ot TrCH parameters such as spreading factor (SF) or Transport Format Combination Set (TFCS) , which will be written under the same value tag and identity as in current network. At call establishment the configuration stored in the UE will be different from the configuration that the network supposes. The UE will apply the wrong configuration and the communication will be spoiled somehow or the UE will be detached from the network until the next switch off / switch on or entering of new scope area (next re-read of system information). The following scheduled broadcasts of system information by the network will be ignored by the UE due to the fact that the value tag and identities are the same.

Similar threats can be expected to other information that the UE uses based on the system information, e.g. measurement configurations. But this is less critical due to it is valid only for one cell.

5.1.6.2 Countermeasures

Editor's Note: It should be checked how long lasting the attack is.

The solution against the described threat can be based on the methods:

- Source origin authentication using Signatures / PKI;
- Source origin authentication schemes such as TESLA.

Taking into account the complexity and issues related to implementation of both schemes, the following countermeasures which can mitigate this particular threat should be considered. These protective measures can not help against first introduction of wrong system information by attacker, but at least can identify that the system information is wrong, so UE can take actions to reconnect to the network.

In LTE security association between the UE and the network is maintained when the UE is in idle state. RRC security context is established and started whenever the RRC connection is established.

Solution 1

This is reasonable to consider using integrity protected RRC message to verify the correctness of SI received by the UE while attaching to a new cell. In order to save critical uplink resources this might be not standalone message but an ordinary RRC message sent for connection set-up.

Editor's Note: It should be checked what are the ways to re-select a cell.

This message in general can include the following information:

- Information for connection establishment
- Information which SI parameters are needed to be verified
- MAC-I, which is built on the following
    - RRC Message, including SI parameters needed to be verified
    - Checksum value calculated on the SI needed to be verified (agreed SI parameters; constants, timer values). The checksum is calculated in UE and it should be a unique identifier of the set of parameters which it was calculated on.
    - Other parameters to build MAC-I (Integrity Key, COUNT-I, Fresh, etc.)

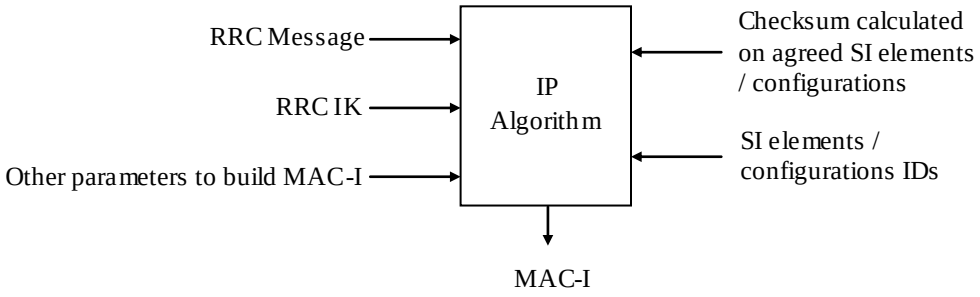


Figure 4. MAC-I for check message

Thus, nothing new is sent to the network apart from the information on which pre-configuration IDs / system parameters were used to build MAC-I. But this information needs to be anyway signaled to the network in order to allow it to rely on the fact that the UE has received this information. The set of standard SI parameters needed to be verified can be specified to eliminate the need to communicate it to the network.

In order the network to be able to calculate correctness of MAC it is necessary that the UE before transmitting the message including the MAC-I is informed about which SI parameters / pre-configurations the network knows (this can be received directly with the System Information), and only calculates the check code basing on this information.

The drawback associated with this method is that in case if MAC is wrong, the network can not distinguish whether the UE is not authenticated or whether the SI is wrong. But in normal case this fact should never happen for the UE, otherwise this will be an error case for which UE actions should be defined, e.g. it can try to resend the message pretending that there is no SI / predefined configuration available in UE. In this case UE can erase the SI and re-read it or network can send it in unicast mode in integrity protected way.

There is no need to make such verification every time UE wants to connect to the network. This can be made only at the first time when UE accesses the network after new reading / change of system information / enter new cell.

For example at first step the e-Node B would indicate the set of known configurations, then at second step the UE calculates the MAC-I adding information on its SI parameters / configurations and sends this information in the first integrity protected UL RRC message.

Whether the check is immediately done in the e-Node B, or whether the check is done in the AGw, with the e-Node B sending the necessary information to the AGW should be FFS. But e-Node B checking is preferred since the AGw does not need to be aware of the system information which is broadcast on a cell.

The additional complications associated with this solution in normal case are minor and are as follows:

1) Connections after new reading / change of system information / enter new cell:

   Additional calculation operations in UE and e-NB to build MAC are needed; Some small information added to existing signalling messages might be needed to agree on SI parameters / configurations to build / verify MAC

2) Subsequent connections: none

Solution 2

Suppose that attacker managed to introduce some wrong access parameters to the UE (e.g. RACH parameters), then even initial access can fail. Thus, in case that RRC message has failed for no clear reason (no reply from the network) the UE can erase all or some relevant IE and re-read it from the network.

Subsequently, if RRC connection can not be established again in this cell, the UE can trigger itself to search for another available cell with lower transmission power level (Ignore current cell-ID). The drawback in this case is that false-BS can mask under the correct cell-ID and no other cells with acceptable power level can be found. But, in LTE network (e.g. in a city), presumably consisting of different pico-, micro- and macro-cells, and even different RATs, this problem will have less impact. (Anyway the attacker can substitute all the cell-ID in this area but it would require more efforts from his side).

The additional complications associated with this solution in normal case: None; it should be triggered upon an error case, when the UE can not make initial connection.

Thus, both solutions can be complementary to each other.

In case that UE is not authenticated to the network, i.e. after switch on, the both described solutions can be applied. In this case the verification of SI can be made in the same way after the integrity protection start.

Alternative solution in case when UE is not authenticated to the network is using checksums for SI verification. UE can calculate checksum (using appropriate CRC or hash function) on received SI parameters / configurations every time when new SI is read. In the first RRC Connection Request message the UE can send this checksum, so the network can check whether the SI is correct. Also the indication on which SI parameters / configurations the checksum was calculated should be sent. In case that the UE has wrong SI, the network can instruct the UE to erase it and to re-read later or send the complete SI set which it wants to use.

To prevent against further introduce of wrong SI by the attacker, the UE should potentially be able to store more than one SI / configurations set under same identities (namely, received from the network and received from the attacker) during some small time interval. Also it can compare its current "wrong" checksum with the checksums calculated on received new SI to prevent against repeated introductions of wrong information. If the UE has received two new configurations during this small time interval, it should store both and try to connect using the first one. In case that the connection using the first predefined configuration is failed the UE can use another predefined configuration.

The complications associated with this method compared to today's UMTS are additional processing operations needed to calculate checksums and additional overhead signaling for sending current checksum to the network in the first RRC message after re-read / change of system information. The other described protecting measures should be activated only when it is identified that wrong system information has been introduced.

Editor's Note: The countermeasure is not bullet proof as the attacker can choose the checksum.

5.2 Threats to eNB and last-mile transport links

It's assumed that the LTE/SAE system will consist of smaller, lower cost radio site equipment, which will be deployed in increasingly vulnerable locations, and that less trusted types of transmission links will be used to interconnect that equipment to the "core network". This chapter covers the threats that may realize due to the

1) Small and low cost eNBs
2) Vulnerable eNB sites (e.g. public indoor site)
3) Less trusted transmission to/from eNB site (e.g. regular office Ethernet cables) (= last-mile)

This review is based on the SA3 assumption that evolved system will consist of 1), 2) and 3). In the following subsections the threats are listed, the possible countermeasures are described and the decisions are tracked.

5.2.1 User Plane packet injection attacks

5.2.1.1 Threats

A) The attacker injects packets in the eNB, which means that the physical security of the eNB has been compromised. The compromised eNB can inject upstream user plane packets to the core network and downstream user plane packets to the UE. Here, the assumption is that the SAE gateway and UE are not compromised.

B) The attacker injects user plane packets on the last-mile, while eNB, UE and SAE gateway are not compromised. DoS attack is also possible. Attacker may send broadcast packets to the access link and try to congest access network as much as possible.

C) Abuse of outsourced network access transit capacity, i.e. insider attack by access network operator employees is also possible. The result is that the access network operator reports more packets than in reality UEs have sent.

D) The attacker succesfully injects packets over the air on behalf of legally attached UEs. Attacker may also modify the packets of attached UEs. The result of this attack is service theft.

5.2.1.2 Countermeasures

The best countermeasure to A) is that the U-plane is integrity protected between UE and the SAE gateway. Using only confidentiality protection for the packets provides much higher security than no confidentiality protection, but still the packet modification attack is possible. However, when only confidentiality protection is used between UE and eNB, and

between eNB and SAE gateway, packet injection attack is mitigated when using appropriate mode of cipher, , i.e., cipher block chaining (CBC)..

It should be noted that the packet/byte counters, if any, in SAE gateway must be incremented only for valid packets (i.e. not for packets that result bogus after decryption). Also, duplicate packet detection has to be considered if counting packets/bytes so that the attacker can't send duplicate packets and affect the accounting for the users.

Another good countermeasure is to introduce counter check procedure in UMTS to LTE/SAE. Counter check procedure should be performed periodically between UE and network. Periodical authentication can also be performed in counter check procedure. There are several ways to implement counter check procedure in LTE/SAE. UE and aGW store some values of counters. These values can reflect the amount of data sent in uplink and downlink direction. UE and aGW periodically perform counter check procedure to check that these values are identical. If these values are not identical, aGW may release the connection. (This paragraph is from S3-060212)

Editor's Note: This countermeasure is only useful when there is no integrity protection. There may be different network nodes needed to store and check the counter. Complexity of counter management and the flexibility of this countermeasure need FFS. The threats mitigated by this countermeasure aren't clear and need full study of the contributor. There may be new threats brought by the countermeasure.

Signalling procedure for periodic local authentication (From S3a071021)

The following procedure is used optionally by the eNB to periodically perform a local authentication. At the same time, the amount of data sent during the AS connection is periodically checked by the eNB and the UE for both up and down streams. If UE receives the Counter Check request, it shall respond with Counter Check Response message.

The eNB is monitoring the COUNT values associated to each radio bearer. The procedure is triggered whenever any of these values reaches a critical checking value. The granularity of these checking values and the values themselves are defined by the visited network. All messages in the procedure are integrity protected.

Editor's Note: It is FFS whether the counters in SGW could be used instead or in addition of counters in eNB.

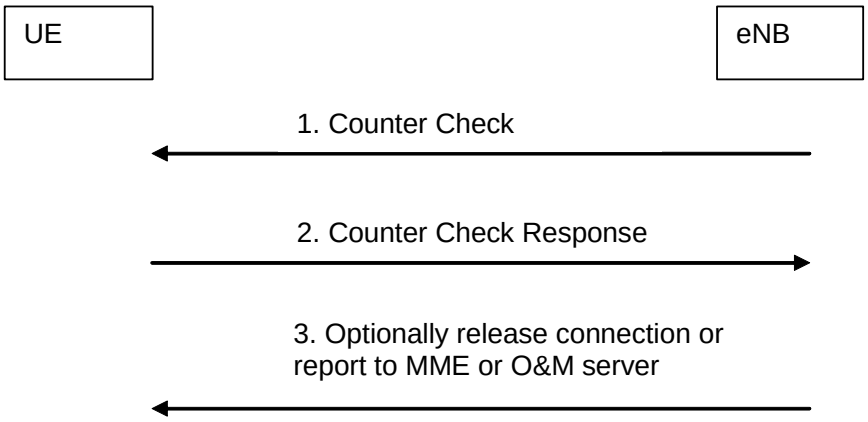


Figure 5 eNB periodic local authentication procedure

1. When a checking value is reached (e.g. the value in some fixed bit position in the hyperframe number is changed), a Counter Check message is sent by the eNB. The Counter Check message contains the most significant parts of the COUNT values (which reflect amount of data sent and received) from each active radio bearer.

2. The UE compares the COUNT values received in the Counter Check message with the values of its radio bearers. Different UE COUNT values are included within the Counter Check Response message.

3. If the eNB receives a counter check response message that does not contain any COUNT values, the procedure ends. If the eNB receives a counter check response that contains one or several COUNT values, the eNB may release the connection or report the difference of the COUNT values for the serving MME or O&M server for further traffic analysis for e.g. detecting the attacker.

5.2.2 User plane packet modification attacks

5.2.2.1 Threats

Here we assume that the user plane traffic is at least encrypted between UE and eNB and between eNB andSAE gateway. Thus, a result of packet modification attack would for example be that UEs would experience lower quality or denial of service.

A) The attacker modifies or drops user plane packets in the eNB or in the last-mile, in such a way that the UE must re-transmit etc. This affects the service quality that the UE (subscriber) is seeing. In addition educated modifications changing traffic content or affecting charging may be possible.

B) The attacker carries out attack A) by eNBhijacking for example the switches/routers on the SAE access network.

C) The attacker modifies or drops user plane packets in the eNB or in the last-mile, in such a way that the UE must re-transmit etc. This affects the service quality that the UE (subscriber) is seeing.

D) The attacker carries out attack B) by eNBor hijacking for example the switches/routers on the SAE access network.

5.2.2.2 Countermeasures

The countermeasure for threat A) and B)is to use user plane integrity protection between UE and SAE gateway. Using only U-plane ciphering between UE and eNB is not enough for mitigating packet modification attacks but provides higher security than no confidentiality protection. Only integrity protection can provide full mitigation for packet modification attacks.

The countermeasure for threat C) and D) is FFS.

5.2.3 User plane packet eavesdropping

5.2.3.1 Threats

The attacker may be eavesdropping at any interface between the UE and the SAE gateway or in a compromised eNB. The threats of this are:

A) Steal confidentiality of data transmitted in the packet payload (content confidentiality)

B) Steal confidentiality of context information such as identities, routing information and communication behaviour.

5.2.3.2 Countermeasures

User-plane confidentiality protection can be used to mitigate threats of type A).

For B, it can be said: The lower the layer at which confidentiality protection is applied the more information is protected. In particular, if confidentiality protection is applied below the IP layer then IP addresses and routing information are protected. For identities used below the IP layer, we need information from RAN2 on UE-ID.

Another advantage of performing confidentiality protection below the IP layer is the expected reduced overhead for security association establishment. So as not to destroy the effect of compression located in the PDCP layer, the encryption layer should be below the compression layer.

5.2.4 Physical attack threat on eNB

5.2.4.1 Threats

A) Breaking the eNB to get the keys and unencrypted data is theoretically possible, i.e. there may be some points in the eNB where the unencrypted data is exposed between two encrypted data pipes. The attacker may dig out the eNB-MME/SAE gateway shared secret or a long term certificate from the eNB and tries to add another eNB (in the same or another network).

By physically breaking into eNB the attacker will be able to circumvent RRC Integrity Protection and:

- Launch RRC DOS attack against UEs (idle and/or active). For example, the attacker can force an active UE into the idle mode and in the absence of UP confidentiality protection direct UP packets to the fraudulent UE (theft of service);
- Get a hold of UE's identities, thus compromising linkability and anonymity of the UE's;

B) The attacker steals an existing and deployed eNB to sell or deploy for own use..

C) The attacker gains access to the OM&A security context at the eNB. This security context might be used by the attacker to reconfigure the attacked eNB, or can be used to attack other eNBs.

5.2.4.2 Countermeasures

For threat A) this kind of attacks can be protected with physical security measures such as alarm systems to protect unauthorized opening of the eNB, putting keys into a hard to break chips etc. The countermeasure is to use identification and separate private keys between MME/SAE gateway and each eNB. eNB can have a secure module to store long term keys, which are used to bootstrap SA between eNB and MME. The identity of a eNB could be stored in a trusted physical module (TPM) and/or a possibly non-removable smartcard. Then the MMEs and SAE gateways compare the ID of the eNBs against a list of valid and revoked IDs. Depending on the cost this solution can be implemented.

Use physical security. Solution as for A, i.e., using not reset TPM, could help identifying the eNB if it is connected to an operator

For threat B) Use physical security for eNB implementation (i.e. burn identification information into the eNB during manufacturing phase). The ID is in tamper resistance chip (e.g. smartcard) and can not be changed without breaking the chip. The secret key (used in asymmetric cryptography) can not be read from the chip. MME is able to detect if there are two eNBs using same keys. When using eNB identification, it necessitates that MME's of different operators cooperate in detecting eNB's with the same identity.

Security context at the eNB (i.e., RRC keys, S1-C/U keys, eNB OM&A keys, etc.) can be also protected by the means of ensuring Platform Security and/or Physical Security of the eNB.

5.2.5 (D)DoS attacks against eNB from the network

5.2.5.1 Threats

A) A network node from the network, which is overtaken by an attacker, launches a logical (D)DoS attack against the eNB(s) by sending selected packets towards the eNB(s).

If IP multicast is used to send traffic to the eNBs, the effect of the attack is increased. For example, if IP multicast is used to deliver paging messages to all eNBs in the tracking area, all these eNBs (and their paging channels on the air interface) will be affected. (S3-070091)

5.2.5.2 Countermeasures

eNBs should not reserve any resources based on signalling without proper authentication. This would mean that the eNBs do not trust other eNBs without proper authentication.

Proper authentication in an IP multicast setting requires the use of public key cryptography signatures or a fully meshed symmetric key distribution if Data Origin Authentication is desired, or key hierarchies similar to MBMS, if only group authentication is required.

5.2.6 (D)DoS attacks against eNB from UEs

5.2.6.1 Threats

a) The attacker impersonating a UE sends selected packets against the eNBs to deny eNB services from others.

b) The attacker could launch a logical (D)DoS attack towards the eNBs from the RAN side.

c) The attacker could send random radio signals that impede the physical layer communication (radio jamming)

5.2.6.2 Countermeasures

The countermeasure is to integrity protect signalling after successful authentication. Before the UE is successfully authenticated, protocols should be used that are not highly vulnerable to (D)DoS attacks (for example cookies to avoid blind DoS attacks).

Editor's Note: The countermeasures for detection and report against jamming attacks need to be further detail.

Threat B) can be mitigated with mutual authentication between UE and eNB based on eNB-specific session keys. There are two possible solutions after that:

- Session keys are bound to the eNB identity and the master key for deriving eNB specific session keys are stored only in the UE and the MME. Attackers cannot leverage compromise of one eNB to compromise other eNBs. eNBs do not contain long term UE session keys (eNB keys with the MME are there) and they can not derive or create keys for other eNBs. Using the UE-eNB session keys provides protection against logical DoS attacks based on mobility signalling between eNBs. Context transfers and/or handoff commands can be authenticated and thus resource depletion attacks are mitigated. Attackers can't hijack UE's application level protected sessions with a hijacked eNB. Attackers can't hijack UE-MME session or initial access authentication key material with a hijacked eNB. Based on the eNB specific session keys attackers can't hijack sessions with other eNB with a hijacked eNB. Because of the separate UE session keys with every eNB, an attacker can not hijack UE sessions moving out of the hijacked eNB.

- After mutual authentication, rate limitation can be used to limit the amount of resources one UE can consume.

Radio jamming (threat C) attacks can be made with special hardware and countermeasures for these are not feasible to implement. However, jamming attacks may be detected and reported.

5.2.7 RLF recovery[5]

5.2.7.1 Description

Figure 6 shows the signalling flow for the Radio Link Failure (RLF) recovery RRC procedure. For this procedure to be successful, the eNB must have been prepared with a security context for the UE and a token which is computed as the MAC-I of the source C-RNTI, source PCI and the target cell ID using the keys and integrity algorithm in the source cell. The token is 16 bits long.

[5] This section is from S3-081017.

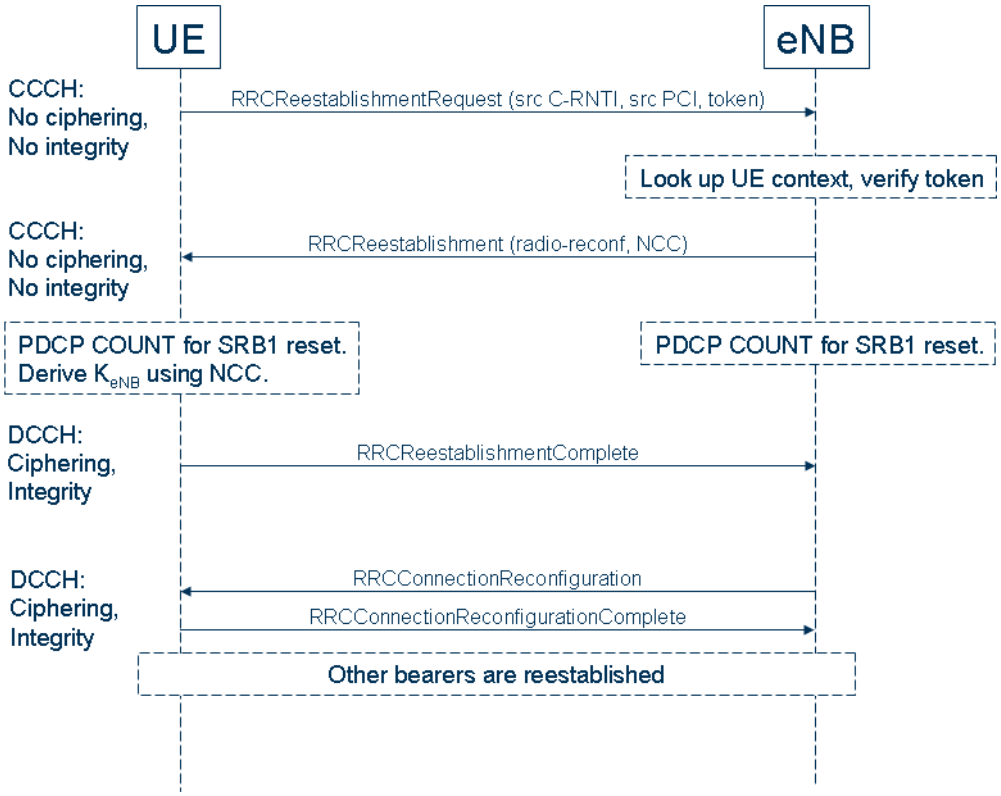


Figure 6 RLF recovery RRC procedure.

The UE contacts the eNB on the unprotected common control channel and provides the source C-RNTI, the source PCI and the token. The eNB verifies that it has the UE context corresponding to these parameters. If that is the case the eNB responds to the UE with the NCC value necessary for the K_{eNB} derivations and reconfiguration data to establish the protected dedicated control channel on SRB1. The UE configures the DCCH and responds with an integrity protected complete-message on SRB1. All other radio bearers remain suspended.

At this point the eNB has assurance that the eNB is the correct UE, since the UE both has proven ownership of the token and of the K_{eNB}.

Next the eNB run the RRC reconfiguration procedure and sends reconfiguration messages for the remaining radio bearers to the UE over the integrity protected and ciphered DCCH. Since the UE is in RRC_CONNECTED state (otherwise no RRC connection re-establishment would have been attempted), there is at least one radio bearer that needs to be re-established (e.g., the default EPS bearer), and hence the RRC reconfiguration procedure would always be run.

At this point the UE has assurance that the eNB is in possession of the K_{eNB} and is hence implicitly authenticated.

5.2.7.2 Threats

The two messages which lack integrity protection are the RRCReestablishmentRequest from the UE to the eNB and the RRCReestablishment from the eNB to the UE.

Message injection

If an attacker injects a faked RRCReestablishmentRequest on behalf of a UE, it is possible to know the source C-RNTI and the source PCI for a particular UE, but since the attacker does not have access to the K_{eNB} of the UE it will be a low probability that the attacker is able to compute the correct token. A mismatch in the token implies that the eNB will not be able to retrieve any UE context, and will hence not change state.

If an attacker awaits a RRCReestablishmentRequest from a UE and then injects a RRCReestablishment as a response, the LS reply from RAN2 (R2-084906) informs us that:

The reception of the message including false parameters might cause the UE to transmit some physical layer feedback signals on wrong radio resources, hence causing potential interference to the UL signals of other UEs. However, this would only last for a very short time, only until the UE receives the subsequent RRC Connection Reconfiguration message, which is integrity protected by the PDCP layer.

Hence the result of an injected RRCReestablishment is a short disturbance of the recovering UE's service, and a possible short disturbance of other UEs in the same cell.

Message modification

Modification of a UE's RRCReestablishmentRequest results in that the eNB will not be able to identify the correct UE context. The result is that the UE will get a RRCReestablishmentReject and will go via RRC_IDLE. If the UE has data to send it will come back via a normal RRCConnectionEstablishment procedure. If there is downlink data for the UE the UE will be paged, and then come back via the normal RRCConnectionEstablishment procedure.

In any case, the result of the attack is a short glitch.

If an attacker awaits a RRCReestablishmentRequest from a UE and then modifies the RRCReestablishment response from the eNB the result is the same with an injected RRCReestablishment in the same situation, i.e., less severe than a short radio-jamming attack.

Pre/replay attacks

An attacker may record a RRCReestablishmentRequest message from a UE and (continuously) disturb the retransmissions of the message. The attacker could then replay the message in a different cell (possibly in a different eNB). This attack is countered by including the identity of the target cell in the token derivation.

If an attacker records a RRCReestablishmentRequest message from a UE and replays it in the same cell, the attacker is simply functioning as a repeater, and this can hardly be considered an attack.

Replaying the same RRCReestablishmentRequest in the same cell after it has been used by the correct UE has the effect that the eNB will reply with a RRCReestablishmentReject message on the CCCH due to not being prepared for RLF recovery for that UE any longer.

Deletion attacks

An attacker may disturb the transmission of messages to/from the UE and hence they would be "deleted" from the message flow, but this is a classical radio jamming attack (or regular disturbance in the radio channel), and security protection is not generally applied.

5.2.7.3 Conclusion

The token based approach for authentication of the UE relies on similar principles as the NAS-token used in IRAT IDLE mode mobility from E-UTRAN to GERAN/UTRAN. The size of the token is 16 bits which was indicated as acceptable both for the NAS-token and explicitly for this token (see the LS from SA3 in S3-080226). The possible synchronization problem that exists with the NAS-token is not present in this case as the token is not based on a sequence number.

The procedure is sufficiently robust against deletion attacks.

5.3 Threats to MME/SAE gateway

5.3.1 (D)DoS attacks against MME through from RAN side

5.3.1.1 Threat

A) The attacker launches a logical (D)DoS attack against the MME utilizing signalling that comes from RAN side, for example initial access authentication

5.3.1.2 Countermeasures

The countermeasure is to integrity protect signalling after successful authentication. Integrity protection should be bound to authentication and there should be rate limitation in case of certain UE behaviour. Before the user is successfully authenticated, protocols should be used that are not highly vulnerable to (D)DoS attacks. Another countermeasure is to use cookies.

5.4 Threats related to mobility management

This chapter describes threats that are relevant to mobility management functionality. The term mobility management in this context covers both the protocol and the architecture used for UE handovers between different access networks. These handovers can be both inter-technology as well as inter-domain. The threats that we discern originate from the Technical Specification 3GPP TS 21.133 V4.1.0 (2001-12) on Security Threats and Requirements (Release 4).

5.4.1 Unauthorised access to control plane data

Mobility management traffic could disclose sensitive data related to users or network providers. An example is network provider resource utilisation data. Inference of information through observing mobility traffic can lead to a violation of confidentiality.

5.4.1.1 Threats

a) Eavesdropping to mobility management control plane traffic when carried unprotected.

b) When encryption is used to protect control traffic, an encryption termination point can be:

- Compromised.
- rogue
- masqueraded

5.4.1.2 Countermeasure

For mitigating the eavesdropping threat a), encryption can be used to protect traffic. For protecting against compromising in case b), the entity holding the keys (i.e. the encryption termination point) must be physically protected, and access must be authenticated and authorized. To prevent rogue and masquerading nodes accessing the control plane data, authentication must be used and where possible monitoring should be used for detection of these situations.

5.4.2 Privacy

Observation and/or analysis of mobility management traffic could lead to privacy violations such as disclosure of user location.

5.4.2.1 Threats

The same set of attacks defined for the confidentiality threat can be applied to violate privacy; additionally we identify the following threats:

a) Derivation of privacy sensitive information by linking of clear-text identifiers.

b) Browsing of mobility related information could disclose privacy sensitive information.

5.4.2.2 Countermeasure

In general the confidentiality countermeasures can be used to mitigate the afore-mentioned threats. For threat a) and b) encryption and hashing can be applied.

5.4.3 Unauthorised manipulation of control plane data

Unauthorised manipulation of mobility management control violates control plane integrity.

5.4.3.1 Threats

a) Replay attacks

b) Manipulation of mobility management control data when carried unprotected.

c) When encryption is used to protect control traffic, an encryption termination point can be:

- Compromised.
- rogue
- masqueraded

5.4.3.2 Countermeasure

Encryption can be used to prevent unauthorised access to control plane data in general in case of threat b). Furthermore, signatures can be used to guarantee the integrity of the data. Time stamping and packet counters can be used to mitigate the risk of replay attacks in threat a). For additional counter measures regarding c) see the counter measures described at the Confidentiality threat.

5.4.4 Disturbing or misusing network services

Disturbing or misusing network services leading to denial of service or reduced availability. Note that this concerns authorised users as opposed to unauthorised users described in the next threat.

5.4.4.1 Threats

a) Redirection of other users traffic and control traffic (to attacker or black hole, to flood a victim third party)

b) Flooding the RAN

c) Replay attack

d) Flooding the core network

- From outside the network (e.g., Internet)
- From inside the network (e.g., replace a network element)
- Rogue network entity misusing its privileges

5.4.4.2 Countermeasure

Authentication, monitoring and logging are appropriate countermeasures for mitigating the afore-mentioned threats.

5.4.5 Unauthorised access to network services

By circumventing authorization procedures unauthorized access to network services can be obtained.

5.4.5.1 Threats

a) Intruders can access services by masquerading as users or network entities (impersonation).

b) Users or network entities can get unauthorised access to services by misusing their access rights.

5.4.5.2 Countermeasure

Intrusion detection and authentication methods are suitable countermeasures for these threats.

6 User Plane Security

6.1 Consequences of (not) applying user plane integrity protection

Issue-1: Adding MACs to each user plane packet reduces the available bandwidth.

While it could be supposed that LTE access should not have the bandwidth limitations of 2G/3G systems, it should still be a design goal to maximize the available air interface throughput and minimize delays. Applying integrity protection to short packets (e.g. VoIP), adds a non-negligible amount[6] of overhead.

As an example suppose a voice sample with length 40 bytes. It requires a 20 byte IP header, 8 bytes UDP header and a 12 byte RTP header to transport on an IP network. The IP/UDP/RTP header can be compressed (e.g. ROHC according to RFC3085). Applying HMAC-SHA-1 produces a 160-bit MAC value which could be truncated e.g. to 128-bit (16 Byte). Suppose that the header compression succeeds in a 40 to 5 byte compression leading to a packet of 45 byte. Then adding a MAC of 16 bytes adds an overhead of 16 byte to the 45 byte and thus increases the packet size by 35%. If we decrease the MAC-length then adding integrity protection codes will consume less bandwidth but at a lower security level. Adding an 8 byte MAC code to each IP-packet, which could be seen as a minimum from a security point of view, would still expand the packet size by 17,5%.

Editor's Note: The length of the MAC could be much shorter, e.g, 4 bytes.

The calculation above assumes that there is one IP packet per PDCP PDU[7]. Possibly several short IP packets could be put into one PDCP PDU. This would reduce the MAC-overhead, but increase the effect of a bit error.

Issue-2: Most IP packets are small

The contribution R2-061858 to the RAN2adhoc in June concludes that it is important for an LTE access network to provide for efficient transmission of large fractions of small packets. It's quoted from that contribution: ' Internet traffic analysis studies (e.g., [1], [2]) highlight an important aspect that should be considered within the RAN groups in the context of LTE: more than 50 percent of all IP packets in the Internet are small (roughly 40 bytes or less). To a large extent those are the TCP acknowledgements and TCP connection management messages (SYNs / FINs). Note that a TCP receiver typically acknowledges every other data packet. Thus at least one third of the packets of a TCP-based bulk data transfers are TCP acknowledgements.

When assuming for an SAE/LTE access network a larger share of VoIP traffic then an even larger percentage of IP packets will be small. And when also assuming a wide use of IP based header compression within an SAE/LTE access network then those small IP packets will result in even smaller PDCP PDUs (e.g., roughly 5 bytes in the case of a TCP acknowledgement). '

So we conclude here that adding integrity protection will cause a considerable overhead when performed at PDCP layer both for TCP and for VoIP traffic (cf. Issue-1).

Issue-3: Implications on conversational (real-time) voice.

Most audio/video encoding schemes will produce acceptable quality from the user point of view, even in the presence of bit errors. When applying integrity protection, a single bit error, either in the data portion of the packet or in the MAC portion, will cause a packet to be dropped. The effect may be non-acceptable voice-quality, dependent on the value of the BLER (Block Error Rate) that is expected to be higher at the cell-edges.

---

[6] Similar considerations (but less severe) apply when block cipher encryption is used as this may already cause packet expansion before even integrity protection is applied

[7] It's assumed that confidentiality & integrity protection is applied at the PDCP Layer.

Editor's Note: As far as SA4 is aware, only PS services apply to LTE. SA4 doesn't foresee any PS services in LTE requiring that packets containing residual bit errors be received by the application layer and its media decoders. This is because SA4 assumes that the underlying layers (RLC/PDCP/IP/UDP/RTP) will discard any packets with errors anyway (i.e. Unequal Error Detection isn't used). So the effect of packets drops due to failed integrity verification is expected to be null as seen by the media decoders. (From S3-060737)

Issue-4: Implication on streaming media.

In general on streaming media fewer problems are expected regarding quality when packets have to be thrown away at the receiver because of integrity check failures. This is due to the fact that packet buffering applies at the receiver and missing packets could be retrieved by the application before play-out (retransmission requests). Whether this can be done without noticeable effect on the application depends on the buffer size and the round-trip-delay.

Issue-5: Effects on information retrieval services (Bursty in nature).

The TCP layer provides the reliability for many upper layer applications/protocols (e.g. http), and thus ensures that missing packets are re-fetched. PDCP packet drops due to failed integrity protection would be corrected. However, using TCP results in the use of many short packets (issue-2).

Issue-6: Integrity services may be provided already at the upper layers.

Applications that require high security will use application layer security mechanisms (e.g. TLS) and these services mostly run on top of TCP (issue-5). However, SA3 decided that the security features of LTE should be developed as an independent toolbox without taking into account application layer security services.

Issue-7: The benefits for an attacker replaying/modifying encrypted packets are practically not so clear (no integrity protection)[8]

It is well-known that encryption alone does not provide integrity protection features, but practically encryption alone may already increase the complexity to mount a successful attack.

Considering the effects of packet modifications, it may not be so difficult for an attacker to meaningfully modify packets in the presence of encryption. Especially in the case of a stream cipher if the attacker knows e.g. the IP address of the target and the position of the IP address in the bit stream, the attacker can change it to any other IP address without having to break the stream cipher. This could be used in a redirection attack. Encryption of the UP traffic on one hand makes it more difficult for an attacker to determine the location of the IP header(s) within a PCDP PDU. In addition, in order to modify the destination address of an IP packet that is encrypted with a stream cipher, the attacker has to know the original destination address. A prudent security design would include user plane integrity protection in order to future-proof the system.

Packet substitution or packet insertion of formerly sent (encrypted) packets will fail due to unmatched sequence numbering (SN)[9] of the payload as this SN is used within the key stream generation (cf. UMTS).

Issue-8: The benefits for an attacker replaying/modifying unencrypted packets (no integrity protection)

As there is no packet authentication for user plane data in this case, this allows packet modifications (e.g. redirection attacks) and replays. When we assume that Network Domain Security is applied on the S1-U[10] reference point in order to counteract S1 reference point threats, then the attacker needs to be active on the air-interface. In this case there is a benefit to apply user plane integrity protection. Dependent on the type of application this may reduce the perceived quality and available throughput (see issue-1/2)

Issue-9: Adding user plane integrity protection adds complexity/cost

Adding user plane integrity protection is not more costly from a performance point of view than ciphering alone. Assume that UIA1 and UEA1 can be reused then applying both ciphering and integrity protection seems to require twice as much cryptographic performance as for a UMTS UE. Keyed hashing can be done very fast. But for short packets integrity protection adds considerable overhead (cf. Issue-1/2). From an algorithm implementation point of view most implementation may be shared with the ciphering algorithm (e.g. UIA2, UEA2), but this is not the case

---

[8] It's assumed here that the encryption layer is at PCDP i.e. below the IP layer such that it is hard for an attacker to perform meaningful and sustainable packet (including IP header e.g. for redirection attacks) modifications.

[9] If encryption is applied in the way it is in 3G

[10] S1-User plane (between eNB and SAE gateway)

generally. When we suppose that user plane ciphering is based on a stream cipher then most of the complexity, i.e. sequence number handling, is already there. Note that as described below, secure activation of integrity for user plane needs to be ensured.

6.2 Track of Decision

The countermeasure “confidentiality protection” is required. Because of the advantages mentioned in the previous subsection (editor: user plane packet eavesdropping) confidentiality protection shall be performed at or below the PDCP layer (for PDCP, cf. TR 25.813). (from user plane packet eavesdropping conclusion)

It was decided at SA3#44, based on S3-060490, that confidentiality (and, if, required, integrity) protection shall be performed at or below the PDCP layer.

The work assumption is no integrity protection for user plane (from S3-060670).

It was agreed that PDCP for user plane ciphering will be moved from UPE to eNB in SA2-RAN2-RAN3 joint meeting in Feb 2007.

7 Control Plane Security

7.1 MAC, RLC and RRC layer security

In SAE/LTE the number of different MAC entities is reduced compared to UTRAN (e.g. MAC-d not needed in the absence of dedicated transport channels). The following analysis is under the assumption that there is no confidentiality or integrity protection at MAC layer.

In downlink (DL), anyone can receive the DL L1 control channel and find the DL time-frequency resource of a certain C-RNTI. Since TB is not encrypted, anyone can read TB to find the MAC C-PDU and D-PDU. Since C-PDU is not encrypted, anyone can read C-PDU. C-PDU in DL has the information only on ARQ, HARQ ((Hybrid) Automatic Request), not UE-specific information. Since D-PDU header is not encrypted, anyone can read sequence number, LCID (logical channel ID added by MAC), etc. If the plain text sequence number is NOT continuous in the handover, basically nothing can be followed. The payload is either RRC message or data from PDCP. The user data from PDCP is encrypted and resistant to confidentiality attack.

In uplink (UL) anyone can receive the UL L1 control channel and find the UL time-frequency resource of a certain C-RNTI. Because UL TB can be also received/demodulated/decoded in sub-frames, anyone can know the proper sequence number to be sent in the next sub-frame. But reusing that C-RNTI just collides with the transmission from the correct UE that has the C-RNTI. So, the only way to reuse C-RNTI is by requesting the capacity first by using the UL buffer status report MAC C-PDU. This C-PDU also does not have any UE ID inside, and the UE is supposed to be identified in L1. Thus, if UE is not properly checked in L1 (or if nothing is added in MAC), anyone can send UL buffer status report MAC C-PDU by reusing C-RNTI of other UE.

Buffer status reports from UEs to the eNBs are not protected and may be used by an attacker to make the eNB believe that other UEs don't have anything to transmit and get more resources as a result. The attack may also result in faster initial access times for the attacker (for example a burst of packets). However, the real UE is also sending buffer status reports, although not during DRX period if there is no urgent uplink data in the UE. As a result there may be a conflict in the eNB if an attacker is also sending reports on behalf of other UEs. It may be difficult to launch this attack as the UEs must follow the allocation tables. If they do not follow the allocation tables, the eNB can not decode the packets (noise).

If buffer status reports are sent in a random access channel (RACH) the attack is easier. RACH is used, when the UE is attaching to the eNBs as well. Sending false attach requests may be possible. The attack is comparable to Denial-of-Service attack as the attacker needs to send attach requests fast enough to consume the resources in the RACH.

A smart attacker can affect packet scheduling, load balancing, and admission control with false buffer status reports, but analyzing the real threat from these is not possible without proper knowledge of the algorithms (e.g. if the packet scheduling algorithm uses only the latest buffer status report or multiple reports to make decisions). Possible case might be that an attacker attacks only few other UEs and thus makes the attack more difficult to trace or notice,

The MAC header contains no sensitive data being mainly related to framing and segmentation. Apart form the MAC header, the only unprotected part of MAC is the peer-to-peer signalling, which is related to outer ARQ, retransmission window handling, and buffer status reporting. There is no confidential information in these messages.

Message insertion, deletion or modifications are not useful to the attacker, because the only result is the deterioration of the service, which can be achieved by simpler means (e.g. a simple analog interference transmitter, radio jammer). The only exception here might be the unprotected buffer status report, which may be easier and more effective for an attacker to use than radio jamming.

7.1.1 Conclusions

If there is already confidentiality and integrity protection at layers above MAC, there is no need for confidentiality or integrity protection at MAC layer. The worst thing that can happen caused by attack against the MAC layer is deterioration of the QoS, which can be achieved by simpler means like with a radio jammer.

- MAC layer does not need integrity protection or ciphering as attacks on MAC layer are comparable to radio jamming attacks. An attacker can not map MAC level messages together during handovers.

RRC ciphering prevents multiple UE tracking threats.

- RRC must be ciphered to prevent UE tracking based on cell level measurement reports, handover message mapping, or cell level identity chaining when ciphering key is available. If seen necessary higher layers messages transferred with RRC messages do not have to ciphered, if they are protected in the higher layers.

Tracking of UE based on packet sequence numbers is a threat especially in the LTE

Editor's Note: There is some concern on the cost of implementing RRC ciphering. If there is a low cost solution as a countermeasure to the threat above, SA3 is open to considering that solution.

7.2 SAE/LTE AKA

7.2.1 Requirements on SAE/LTE AKA

The possible options for SAE/LTE AKA have been discussed as:

- Use of "native" UMTS AKA.
- Use of EAP AKA.

UMTS AKA is considered to be a trusted protocol for authentication. The signalling sequence in high-level for authentication and key agreement in UMTS can be reused in SAE/LTE as well.

7.2.1.1 General

R0: The SAE CN and LTE AN SHALL allow for keys of size 128 or 256 bits. (From S3-060632)

- The MME shall be able to derive (key derivation function) keys of 256-bit length for CP, UP based on the information received in the Authentication vector (and potentially other information).
- The signalling protocols between the key derivation function in the MME and the key usage functions (i.e. the encryption and integrity protection functions) shall be able to transport keys of 256-bit and 128-bit length.
- The MME and eNB shall include Encryption and Integrity protection functions that are able to handle a key size of 128-bit. In case a 256-bit key is received then it needs to be truncated before interfacing with the security functions.
- The MME and eNB may include Encryption and Integrity protection functions that are able to handle a key size of 256-bit.
- Secure algorithm negotiation shall distinguish algorithms using 128 and 256-bit keys.

Rationale: the 128 bit level is needed for compatibility reasons (assuming that UMTS UEA/UIA security algorithms are to be possible to re-use also in LTE). While there currently is no need to go beyond 128-bit keys, even in 10-20 years perspective [1], and while the only threat to 128 bit keys appear to be quantum computers, it seems wise to guard SAE/LTE investments well beyond the 20 year time frame, hence the 256 bit level. The penalty to support also 256 bit keys seems very small, considering the amount of "future proofness" it provides.

7.2.1.2 Non-3GPP access

R1: AKA for non-3GPP access SHALL use USIM based EAP AKA.

Editor's Note: This requirement has to be confirmed when the other aspects are ready.

Rationale: Considering backwards compatibility with 3GPP I-WLAN and that EAP AKA is currently the only generic way to allow USIM-based access to non-3GPP networks, it appears the only feasible solution. Also, this appears to already be the working assumption in other 3GPP WGs.

7.2.1.3 LTE access

R2: 2G SIM Access to LTE SHALL NOT be granted.

Rationale: 2G security is not sufficient. When a UE has authenticated in GSM and later performs a handover to UMTS, the 64-bit ciphering key Kc is converted to the two 128-bit ciphering and integrity keys CK and IK. This operation does not add any entropy to CK or IK. Moreover, if an attacker breaks the encryption in GSM and gets hold of Kc, will be able to also decrypt the traffic even if the UE moves into LTE if direct hand-overs were allowed. If there is any time to phase out 2G SIMs, making LTE future proof, it is now. It seems likely that USIM can provide necessary security level, see also the next requirement. This requirement implies that an LTE UE that has previously only established GSM security shall be re-authenticated, establishing LTE security context, before granting LTE access.

- Security drawbacks of 2G SIM 1: Small key size

  The new GSM-Milenage algorithms can produce 128 bits keys, similar to USIM application. However, considering that GSM-Milenage is not (widely) deployed and since the only imaginable reason for supporting 2G SIMs in LTE would be to limit the need to physically distribute new UICCs, it is clear that allowing 2G SIM access to LTE will in practice imply a 64-bit security level. Put differently, if distribution of new 2G SIMs can be assumed, then one may as well assume distribution of USIMs.

  In 1998, special-purpose hardware machine was available that would retrieve 56-bit (DES) keys in about a day, [6]. The machine, a special-purpose ASIC design, was built at a cost of about US $250,000. The machine's cost/performance agreed well with predictions based on Moore's law and hardware proposals done already in 1993, [7]. It might seem that the 64-bit level of GSM would still remain secure, considering the cost/effort to break such keys. However, development has continued.

  In 2006, a similar (but FPGA-based) machine was presented that could be built at a cost of under € 9,000, and the machine would find 56-bit keys in a matter of a few days, [8]. Quite recently, an enhanced machine, specially targeted at dedicated stream ciphers (of which the A5/1, and UEA2 algorithms are examples) was presented, [9]. Using additional speed-up possible due to the nature of dedicated stream ciphers, it can be predicted that the effective security of LTE algorithms in general is even less.

  Assuming continued development, it can be predicted that breaking 64-bit keys will be "common place" in at most 10 years, but probably much earlier, to attackers with quite moderate resources.

  The conclusion is that 2G SIM key sizes will not remain secure for the economic life-time of SAE/LTE. Moreover, if 2G SIM support is kept in SAE/LTE, it can be envisioned that the practical problems of phasing out 2G SIMs is just pushed ahead, and when the "next-generation" systems are to be designed, the problem of the "SIM-legacy" will still exist.

- Security drawbacks of 2G SIM 2: Lack of Mutual Authentication

  As is well known, 2G SIM application does not support (home) network authentication. In SA3 it has been discussed whether in LTE, it should be possible to authenticate even the visited network. In any case, use of 2G SIM is clearly a major security risk.

- Security drawbacks of 2G SIM 3: Lack of replay protection

There is no guaranty of random freshness in GSM AKA. Related to the issue of network authentication is the issue of replay protection. This is one of critical aspects that makes the side-effects of the attacks on A5/2 so serious as it can spread also to other algorithms. Again, a significant risk is taken by using 2G SIMs.

Editor's Note: The security SA could be set up shortly after the authentication.

However, a hand-over from GSM BSS connected to a R99+ VLR/SGSN may very well be acceptable and is not excluded.

R3: LTE AKA SHALL be based on USIM and (possible) extensions to UMTS AKA. In particular, R99 USIM shall be sufficient for access to LTE

Rationale: This has already been agreed, and is in a sense therefore a superfluous requirement. Nevertheless, it is re-stated for self-containment. Note that the set of possible extensions include, but are not limited to EAP AKA and GBA. While security context transfer of UMTS security context to LTE is likely to provide sufficient security level (key size etc) at hand-over, it cannot be excluded that LTE security context will be a proper super-set of UMTS security context, this is FFS. As a derived requirement we get:

R4: LTE AKA SHALL produce keys forming a basis for UP/CP protection (ciphering, integrity).

Note: Other keys may also need to be produced, this is FFS.

R5: The LTE AKA keys of R4 SHALL be dependent on the algorithm with which they are used.

Rationale: Such "key separation" is being discussed as a countermeasure to GSM weaknesses discovered the last few years. While UMTS (and thus the re-use of UEA/UIA algorithms in LTE) is still believed to be secure, it seems prudent to introduce these mechanisms in LTE from day one.

7.2.1.4 3GPP non-LTE access

R6: SAE key management SHALL be able to produce keys (CK, IK, Kc) from the LTE AKA keys, compatible with GSM and Rel-6 access networks. Knowledge of these keys (only) SHALL not expose the LTE keys.

Rationale: Interoperability and security. It is FFS whether the same key-conversion functions as used in UMTS to GSM handovers suffice. The intention is that it must be possible to derive keys for UMTS and GSM from the keys resulting from AKA in LTE, so that the UE does not have to perform an authentication when doing hand over to UMTS or GSM to get CK/IK or Kc respectively

7.2.1.5 UE Attach in LTE

R7: As part of the initial attach request from the UE, it must be possible to signal ME security capabilities to the MME, i.e. the ME supported LTE key derivation algorithms.

Motivation: It is clear that the VPLMN must have this information, no later than at the time the security is to be activated ("cipher mode command"). While it may be possible to postpone this information until after AKA, we note that there are situations where AKA will not be needed (e.g. an already known UE) and it is therefore natural to signal this information at the same "place" of the procedure, independently of whether AKA is run or not.

7.2.2 Comparison of UMTS AKA or EAP AKA

7.2.2.1 High level items

H-1) 'Interworking with release 6 3GPP systems (i.e. 3GPP-PS core, 3GPP-IP access and IMS) shall be supported'[11] (TS 23.882 section 5 high level principles).

TSs 23.234 and 33.234 specify methods for interworking between 3GPP and WLAN (I-WLAN): Direct IP access and 3GPP IP access. WLAN and WiMAX are non-3GPP access systems. It is our understanding that any non-3GPP access system may be connected to the SAE core via one of two variants of the S2 interface: either using an IPsec tunnel

[11] Quoted text from SA2 TR: It's assumed that also Rel-7 interworking is meant by SA2.

between a UE and a PDG, as in 3GPP IP access, or without such a tunnel and/or PDG, (e.g. when the EPC is connected to non-3GPP access systems that are considered trusted by home 3GPP operators, or in the case of Direct IP access)[12].

Authentication for Direct IP access and 3GPP IP access is performed via EAP-AKA or EAP-SIM and creates a security context that is NOT shared for access via the PS domain of UMTS. TS 33.234 never meant to cover the so-called scenarios 4 and 5 i.e. session continuity and handover were outside the scope of Rel-6 and Rel-7. TS 33.234 will not be further developed to include scenarios 4 and 5, rather this is to be covered by SAE. When a user moves from 3GPP IP access to UMTS PS domain then a new authentication shall be performed in Rel-6 and 7. The pair CK, IK is available in UE and AAA server, and with some enhancements the keys and other context information could be reused in handover as the AAA server will remain the same and fast re-authentication would be used.

A major point to consider in this respect is the fact that the use of EAP would probably necessitate the involvement of the home AAA server in all handovers (See also P-3) if no anchor-MME concept would be used. This will affect all intra/inter-system 3GPP handovers and adds delays to it.

Assuming that handovers between 3GPP access systems (LTE-LTE, LTE-UTRAN) will be more frequent than between 3GPP and non-3GPP access systems, then maximal alignment of the authentication procedures/protocols of LTE and UMTS can result in optimal handovers between LTE and UMTS. With that respect inclusion of UMTS AKA NAS in LTE would be preferred over EAP AKA.

H-2) Verification of authentication in Home or Visited Network.

Home control of authentication is a basic characteristic of the EAP-architecture i.e. the AAA-server always resides in the Home Network. This protects the authentication vector from being stolen or spoiled and it has the advantage that authorization by HN is always timely. Note that the same authorization can be achieved by using UMTS AKA is a real-time fashion. On the other hand the connection set up time will be larger than in architectures where authentication verification in the VN is being performed as when each full (or fast re-) authentication has to go back to the Home Network. The more transit networks (and transport-proxies) there are between HN and VN, the more time will be added for subsequent (full and fast re-)authentications. In order to enable fast initial and resumed service access (with may require authentication), authentication verification in the VN is advantageous.

H-3) Transport protocols for authentication parameters in the Core network.

In IMS, as well as in GBA, authentication vectors can be transported over DIAMETER (i.e. Cx and Zh reference points). This would allow moving AAA-server functions to the visited network (for EAP-AKA use), but anonymity features seem to require Home control and solutions to discover the AAA-server would need further study[13]. The aGW would then need to implement DIAMETER in stead of MAP. The decision how much signalling protocol reuse there will be, is however to be taken independently by CT4. The aGW could have the choice between implementing MAP protocols (SendAuthenticationInfo) or DIAMETER. If EAP would be chosen for LTE access then DIAMETER would be the protocol for choice at the core network. Note that also the transport of non-security parameters needs to be considered in this decision in particular the transport of mobility management information would also create DIAMETER impacts.

So Issue H-3 is not seen a relevant issue in the comparison of EAP-AKA and UMTS AKA.

H-4) 'Access to Evolved 3GPP System shall be possible via existing Rel 99 USIM. Evolved 3GPP System shall also permit access to inbound roamers from mobile networks with Rel 5 HSS'; (TS 23.882 section 5 high level principles).

When using Rel-99 USIM together with EAP-AKA, then EAP needs to be implemented on the MT. EAP-AKA may optionally also be terminated on the UICC (Cf. TS 33.234 Rel-6).

Issue H-4 is not seen as a distinguishing issue in the comparison.

H-5) 'The authentication framework should be independent from the specific access network technology'; (TS 23.882 section 5 high level principles).

This issue could be interpreted in various ways. It's assumed that the choice of USIMs (and hence AKA) already fulfils this requirement on high level. If we interpret the high level requirement more in the sense of 'Extensibility to support future authentication methods' then the evaluation is different and looks more in favour of EAP support as it is designed

---

12 This understanding may need to be clarified with SA2

13 Usually, the address of the AAA server is derived from the NAI. Problems may also arise with states held in the AAA server when the user moves between networks and the AAA server is allocated dynamically.

to be extensible. Suppose that a new EAP method would be introduced then there need to be no changes on the aGW, but only in the Home Network and the UEs. A home operator could introduce new EAP methods without impacts in the Visited Network. Although this seems a tempting advantage, the threat or disadvantage might be an upcoming proliferation of authentication methods that will be introduced in the UEs and possibly in the HN. AKA is a well-established and secure authentication method, while other methods might be less-secure. On the other hand 3GPP SA3's choice of USIM (and hence AKA) has limited the (EAP-) authentication methods that need to be supported for LTE access to one currently. So there is currently no hard requirement for extensibility. Please also note that SAE will have to support both, UMTS AKA and EAP-AKA anyhow, as both UTRAN and I-WLAN will be accepted access systems in SAE, and the former requires UMTS AKA, while the latter requires EAP-AKA. This has been the situation since Release 6.

7.2.2.2 Particular EAP features

P-1) Privacy/Anonymity Features

RFC4187 includes an optional pseudonym management which is comparable with the TMSI mechanism. The TMSI mechanism is serving network controlled where the EAP mechanisms are home controlled.

Conclusion: The privacy and anonymity features of both seem to have equal strength.

P-2) Fast Re-authentication functionality

Fast re-authentication in EAP-AKA allows the UE to present a re-authentication identity which avoids to run a full AKA and hence to use a quintet. For each fast re-authentication the Home AAA-server needs to be contacted. The use of integrity protection on NAS-signalling (between UE and aGW) may also be considered as fast re-authentication as it also ensures the continued presence of the ME. Fast re-authentication achieves a change of session keys without consuming new authentication vectors. On the other hand, it would be more important to ensure the continued presence of the UICC in order to counter the rogue shell threat. This can be achieved only through a modification of the USIM e.g. EAP termination in UICC (but this would be against a backward compatibility requirement cfr. H-4) or through a new full authentication with new authentication vectors. At any time the aGW can decide to perform a full authentication for the user if the session keys need to be renewed.

When requiring more full-authentications the usage/generation of more quintets should not be considered the problem (i.e. the theoretical number of AVs that can be generated from a USIM, is practically never reached during the lifetime of a UICC). With respect to the performance requirement on the HSS/AuC, there is an increase but this is smoothened over time, while pre-computation (of AV-batches) is possible. The EAP-AKA concept requires a (new) AAA-function in the HN which implements the extra required signalling performance for fast re-authentications.

Conclusion: It is difficult to say which of the two concepts requires less authentication performance on AAA/HSS or HSS-only.

P-3) Fast re-authentication (latencies) versus security context transfers.

Using EAP, each change of authenticator (MME) should require a new authentication since MSKs should not be shared among authenticators [draft-ietf-eap-keying-14.txt].

[draft-ietf-eap-keying-14.txt]: 'EAP keying material and parameters provided to a lower layer MUST NOT be transported to another entity. For example, EAP keying material and parameters passed down to the EAP peer lower layer MUST NOT leave the peer; EAP keying material and parameters passed down or transported to the EAP authenticator lower layer MUST NOT leave the authenticator."

This makes handovers and idle mobility as in UMTS where the same CK, IK are forwarded from RNC to RNC, contradictory to the EAP keying framework. A handover between MMEs always has to involve the AAA server for security.

The IETF has recognized the above disadvantage of using EAP in wireless environments and has started some work cf. [draft-nakhjiri-aaa-hokey-ps-03.txt]. This work is in an early stage. The mentioned draft contains the problem statement.

Conclusion: "EAP-AKA show clear performance disadvantages in handovers between 3G access system: in LTE-LTE handovers, no security context transfer from MME to MME, in LTE-UTRAN handovers, no security context transfer from MME to SGSN, so in LTE-UTRAN handovers, new run of full UMTS AKA over UTRAN would be required."

P-4) Amount of messages for EAP-AKA versus NAS based UMTS AKA

Security capability negotiation and cipher mode setting is not part of EAP [draft-ietf-eap-keying-14.txt] Section 3 secure association protocol], so it still needs to be added after the EAP exchange.

Figure 7 shows the EAP-AKA exchanges on full-authentication. When compared to TS 33.102 flows (Figure 8) there is seems to be no difference in the amount of authentication messages on the air interface. However, if EAP-AKA is used, the security-related procedures as a whole may require more messages (including both core network and air interface messages). In EAP-AKA there are two round-trips across the core network, as opposed to one round-trip with MAP (UMTS AKA NAS).

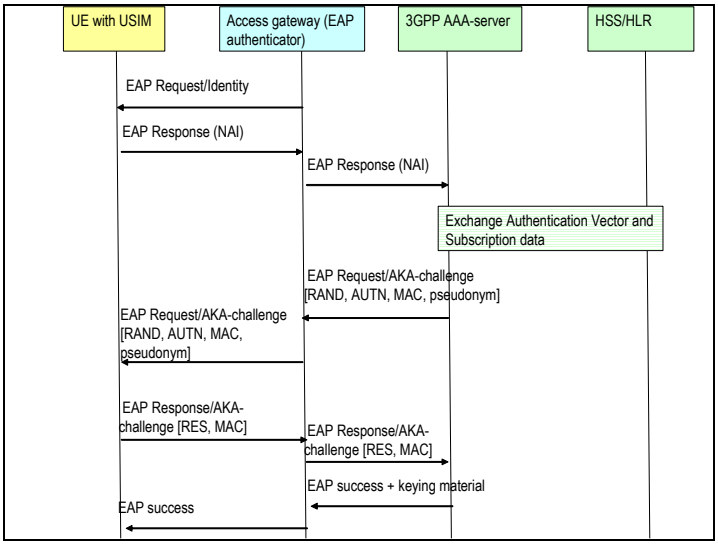


Figure 7 EAP-AKA full authentication

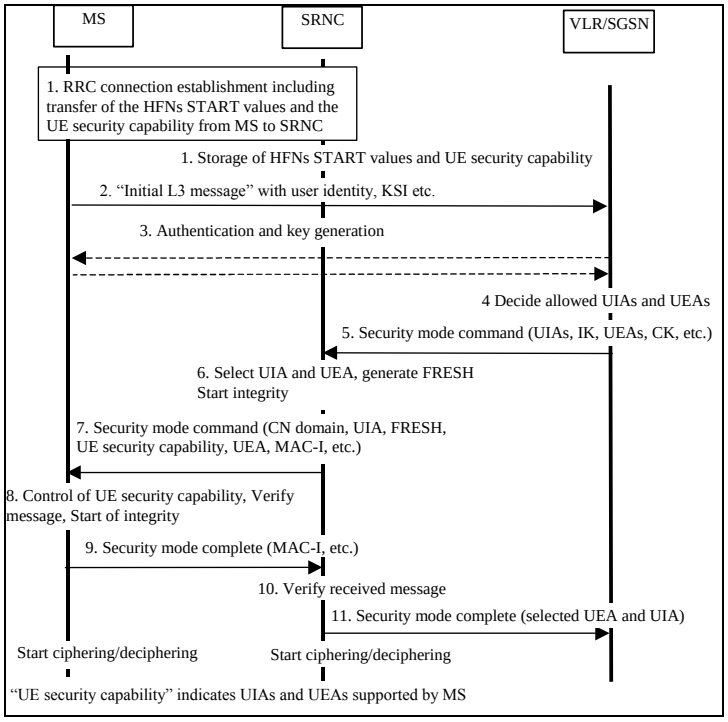


Figure 8 TS 33.102 procedures with AKA NAS in the third step

The following is observed:

1) The EAP Request/Identity may have to be repeated (according to EAP-AKA RFC).

2) In UMTS there is no equivalent to the EAP success message.

3) The security-related information, now carried in the initial L3-message need to be put into a separate signalling message or in EAP (ffs).

4) The security mode command procedure would have to be performed after the EAP procedure.

Conclusion: EAP-AKA is likely to require more messages across LTE.

7.2.2.3 Detailed impacts (outstanding standardization work)

I-1: Location of the EAP authenticator function

The network node that performs the authenticator function has access to all EAP exported keys i.e. MSK, TEK and IV. Therefore EAP needs to be terminated above eNB i.e. in the MME. The main reason is that the MME needs to derive keys for NAS and user plane.

I-2: Necessity of further key derivations for use in LTE (RRC and PDCP protocols).

Neither the EAP-AKA RFC, nor the EAP-framework provides a key derivation for the use of keys in the specific LTE protocols. This key derivation needs to be defined for both EAP AKA and NAS UMTS AKA. This can be documented in 3GPP specifications.

Figure 9 illustrates an example key derivation for NAS UMTS AKA.

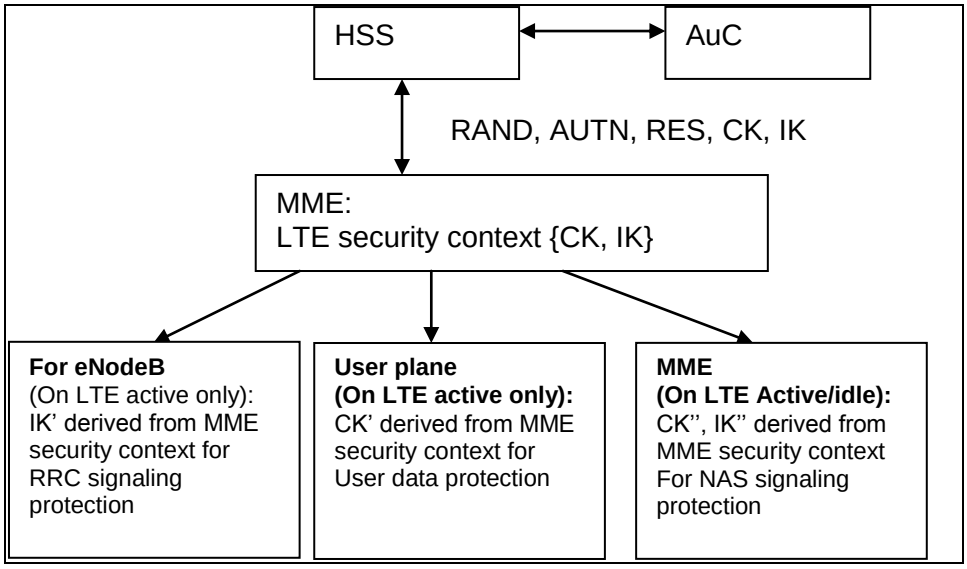


Figure 9 Example Security contexts/key derivation in LTE based on NAS UMTS AKA

In this key derivation solution for LTE access, the CK, IK would NOT be used directly to protect a particular protocol. In idle state only the LTE security context needs to be transferred when the UE performs a tracking area update to a different aGW. On User plane establishment and RRC state change to ACTIVE, keys can be derived within the MME and then distributed to the concerned entities and the RRC keys could be distributed to the eNB on SAE Bearer activation[14].

A suitable key derivation could be build using TS 33.220 Rel-6 key derivation functions (Annex B) or other EAP-based examples.

Figure 10 illustrates an example key derivation in LTE based on EAP-AKA.

[14] Co-location of MME and SAE gateway (or their eventual split and resulting flows between them) are under discussion within SA2

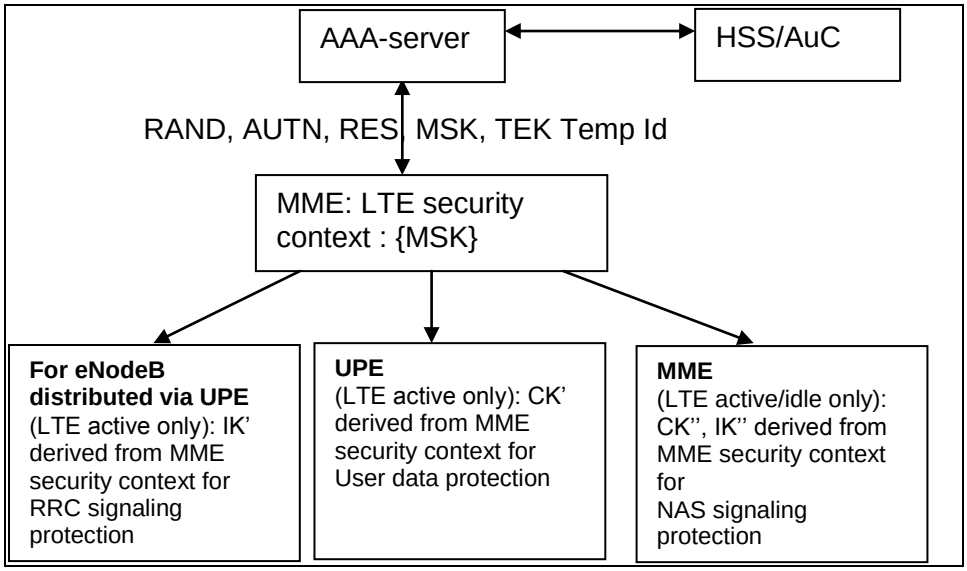


Figure 10 Example Security contexts/key derivation in LTE based on EAP-AKA

As noted before, using a security context forwarding concept (as known from GSM/UMTS) between LTE/SAE nodes will imply a deviation from the EAP keying framework. This issue needs further study.

I-3) EAP needs to be carried over LTE-access

An equivalent to EAPOL is needed and needs to be documented in IETF or 3GPP. Documentation in 3GPP may be preferable.

Editor's Note: RFC3748 has requirement on lower layer and should be included into this study. Known EAP security issues must also be addressed (for example "EAP Success/Failure" message insecurity).

7.2.2.4 Analysis overview

	UMTS AKA NAS	**EAP-AKA**
High level Items		
H-1: Interworking with release 6 3GPP systems ….	+ (more handovers to UMTS/GSM expected)	
H-2: Verification of authentication in HN or VN	+ (less authentication delays for VN concept).	Ensures more home control.
H-3: Transport protocols for authentication parameters …	*Issue found not relevant*	*Issue found not relevant*
H-4: Rel-99 USIM support	=	=
H-5: Authentication framework independence …	Needs to be supported for UTRAN access anyhow.	There seems to be no hard requirement for extensibility according to EAP. Needs to be supported for I-WLAN access anyhow however not in an authenticator role.
Particular EAP features		
P-1: Privacy/Anonymity	=	=
P-2: Fast re-authentication functionality	=	=

P-3: Fast re-authentication versus security context transfers	+ Allows handovers between 3G access systems without involving home network and without new authentication	- IDLE mode mobility and security context transfer against EAP framework.
P-4: Amount of messages Editor's Note: The possibility of message piggyback for EAP AKA isn't taken into account.		- Inherently more messages
Detailed impacts/outstanding standardization work	**UMTS AKA**	**EAP**
I-1: Location of authenticator	No such concept	EAP authenticator in MME
I-2: Necessity of further key derivations	Needed for both UMTS AKA and EAP-AKA	Needed for both UMTS AKA and EAP-AKA
I-3: EAP over LTE	Can be reused (TS 24.008)	Extra work is needed in 3GPP or IETF

7.2.3 RAND and 256-bit keys in E-UTRAN[15]

This section analyses potential threats from these parameter choices, the potential threat being that "full" 256-bit security is not achieved due to the restricted length of RAND.

Three possible weaknesses are analysed due to using 128-bit RANDs together with 256-bit EPS keys, but as the analysis below will show, an attacker seems to be unalble to exploit these, such that there is no justification for the the need of 256-bit RANDs in EPS given that this would cause lots of protocol impacts on the EPS.

It is throughout assumed that the USIM key, K, is 256 bits and that it is required that the AKA-resulting key, Kasme shall have a strength equivalent to 256-bits. It is also assumed that key derivation algorithms in the UICC and the ME are secure, without any exploitable weaknesses.

Some attack/compromise which occurs adversely/by accident with time/success ratio lower than 2^256 shall be considered as "attack" or "weakness.

Key Space Limitation

When RAND is 128-bits, there will only be 2^128 possible Kasme:s generated for any specific subscriber. But under the above assumption, although this set is sparse among all 256-bit keys, this set should have no exploitable structure if the key derivation function is (as assumed) secure (pseudo-random). That is, an attacker will still have to perform a workload of 256-bits to find out the set of possible Kasme's. Therefore, this property is not considered as a serious problem.

Off-line Attacks

It needs to investigate whether an off-line pre-computation attack might be possible if RAND is shorter than 256 bits. The possibility of such attacks are generally dependent on the number of keys in use (i.e. the number of subscribers of the system), but let us first leave this issue aside and look only at the theoretical possibilities.

In this scenario, the attacker (in advance) generates a table of T random, independent (K, RAND) pairs. The attacker also computes the corresponding (RES, Kasme) and sorts/indexes the table by (RAND || RES)-values. This has complexity O(T * log T). This phase can be done once and for all in advance.

---

[15] This section is from S3a090927.

Later, in the "attack phase", the attacker observes M runs of AKA (for randomly chosen subscribers) and records the (RAND, RES)-values. The attacker also records some small amount of ciphertext/MAC values from each "session" (see below). Each observed run of AKA is performed based on some (for the attacker "implict", yet unknown) key K'.

For each (RAND, RES) obtained in the attack phase, the attacker searches for a corresponding (RAND, RES)-value in the table. If a match is found, with some probability it has been derived from the same K as that in the table.

The attacker has T (RAND, K)-values in the table and M observed "implicit" (RAND, K')-values. If M*T > 2^(256+128) = 2^384, with high probability there is a (perfect) match. Choosing T = M = 2^(384/2) = 2^192 is an "optimum".

In the attack phase, for each of the M values, the attacker performs a search in the table, having complexity O(log T). But even if (RAND, RES) matches, it could be a "false" match as, depending on the size of RES, denoted k, the key K may not be uniquely determined.

It is expected that each (RAND, RES)-pair occurs T/2^(128+k) times in the table. Hence, the table is expected to hold equally many K-values, out of which we expect one key to be correct (for the above choice of T and M). For each candidate K-value in the table, the attacker can determine if it is correct or not by using the corresponding derived Kasme to check if ciphertext/MAC values agree or not. This overhead for each candidate is negligible compared to the other computations. (We assume that it is quite likely that there are sufficiently many "known-in-advance" signalling messages whose MAC tags can be pre-generated, given the key.)

Thus, the complexity (neglecting small overhead for each false candidate) is

$$O(T + T * \log T + M * \log T + T/2^{(128+k)}).$$

For the choice of parameters above, the dominating terms are T * log T and M * log T, which are each about 2^200.

This attack is thus non-trivial (better then 2^256), but still seems infeasible/unattractive due to:

- The system would need to have on the order of M subscribers for the attack to be possible. With parameters as above, in a system with, say, less then 2^64 subscribers it is unlikely that the pre-generated table includes even a single valid subscriber key. Therefore the attack is purely hypothetical.
- The required storage (proportional to 2^192).
- The fact that only a "random" subscriber can be attacked.
- The fact that the attacker would seem to have "global access" to AKA runs in order to argue that the subscriber-keys, K, are random and independent.

The main point above is that it is not possible to target a single user, so the incentive to perform the attack seems very small.

Key Collisions

The probability that any two RANDs (including the first two) collide is 2^(-128). The expected number of collisions after 2^t runs of AKA is about 2^(2t-128). If/when a collision occurs, the following can/will happen:

- An attacker who was present at the first use of the RAND can predict RES. However, he will not know the keys and can thus still not hijack the connection.
- Since Kasme will be the same, so will all other keys (assuming the cell ID etc are the same for UP). If the same ciphering algorithms are used, a "two-time-pad" will occur, revealing the XOR of the plaintexts to a passive eavesdropper.

Note that the attacker cannot force RANDs to be the same due to the network authentication based on AUTN. Thus, this security issue is purely "accidental" and outside the control of any attacker. It is questionable if this can even be considered as an attack.

The only relevant security threat identified due to 128-bit RANDs is that of accidental collision, in turn leading to two-time pad. Since this threat is not forcible by an attacker, it is concluded that the length of the RAND does not impact the usefulness of 256-bit keys in EPS.

7.2.4 Migration path to enable 256-EPS AKA[16]

The availability of 256-bit keys in order to ensure entropy of 256 bits in EPS implies changes to EPS AKA, i.e. UMTS AKA. The 256-UMTS AKA would have the following features concerning long terms keys shared between UE and HSS.

The key K used in new version of UMTS AKA should be 256 bits to enable the availability of keys with 256-bit entropy for EPS security. The design of UMTS AKA algorithm should be modified in order to enable the use of key K of 256 bits, this new version of UMTS AKA is referred as 256-UMTS AKA in this section.

To have 256-bit keys for EPS, the key K_{ASME} should be 256 bits. K_{ASME} is derived from the 256-bit key Ks, Ks results from the concatenation of CK and IK. With key K of 256 bits and changes to MILENAGE design, it should be possible to have Ks (CK, IK) with entropy of 256 bits. Consequently, K_{ASME} could have 256-bit entropy with keys CK and IK which are still 128 bits. An alternative could be possible to ask for new design of MILENAGE allowing to have CK and IK keys with length of 256 bits. But this alternative would require heavy changes to MILENAGE design and would have strong impacts on the UICC-ME interface. New length of CK and IK would modify the AUTHENTICATE command (output data of the AUTHENTICATE command in the 3G security context would be changed, computation of GBA bootstrapped key Ks in the GBA Bootstrap security context would be also modified). This alternative is considered as not relevant for 256-EPS AKA migration path.

Changes to UMTS AKA algorithm (e.g. MILENAGE algorithm) have impacts on the User Equipment.

The impacts on the UICC:

- New version of UMTS AKA algorithm

  The USIM should have a new version of UMTS AKA algorithm, e.g. new version of MILENAGE.

  New UICC, with 256-MILENAGE algorithm-based USIM, should be issued when the operators would like to have support for 256-bit keys with entropy of 256 bits in EPS.

- Storage of the key K

  Key K should be 256 bits instead of 128 bits. This change is minor for the USIM.

- AUTHENTICATE command

  There is no impact on the input and output data of the AUTHENTICATE command. CK and IK length remains the same.

  In case that acceptability of defining a new security context for the AUTHENTICATE command in order to provide higher level of security by computing and storing K_{ASME} in the UICC (confer Gemalto contribution S3a070739) then the use of 256-UMTS AKA algorithm would not change the input and output data of the AUTHENTICATE command with this new security context ("EPS Security context").

  There would be changes to 3GPP TS 33.102 and TS 31.102 specifications in order to describe the procedure of the AUTHENTICATE command in case of 256-UMTS AKA algorithm to deal with key K of 256 bits. Those changes are not significant.

- UICC-ME interface

  There is no impact on the UICC-ME interface since the input and output data of the AUTHENTICATE command are not modified by 256-UMTS AKA.

  There is no issue of backward compatibility with Rel-99 USIMs since the input and output data of the AUTHENTICATE command are the same for UMTS AKA and 256-UMTS AKA.

  The support of 256-UMTS AKA algorithm is optional. The decision to issue new UICCs containing USIM with 256-UMTS AKA algorithm will depend on the home operator who issues the UICCs.

Impacts on the ME

---

16 This section is from S3a070938.

There is no impact on the ME due to the new version of UMTS AKA to support 256-bit keys in EPS. The impacts of the support for 256-bit keys to protect UP, NAS and AS are described in S3a070922.

Impacts on the network

The AuC should contain the 256-UMTS AKA algorithm to perform user authentication with 256-UMTS AKA-based USIMs present on the field.

The use of USIM with 256-UMTS AKA will be possible in UTRAN networks since the length of CK and IK remains equal to 128 bits. New algorithm for 256-UMTS AKA with new length of key K (256 bits) impacts the USIM and the AuC only. There is no issue of interworking between E-UTRAN and UTRAN.

Other impacts in the network related to K_{ASME} and keys used for UP, NAS and AS protection, are described in S3a070922.

Summary of changes in the UE due to 256-UMTS AKA algorithm:

	Impacts on the UE
256-UMTS AKA algorithm	Change: new version of UMTS AKA algorithm in the USIM and in the AuC
Storage of key K	Change: storage of 256 bits instead of 128 bits for key K in the USIM and in the AuC
Length of CK and IK	No change for CK and IK length
AUTHENTICATE command	No change for output and input data of AUTHENTICATE command The description of the AUTHENTICATE procedure should be modified in TS 33.102 and TS 31.102 in order to take into account key K of 256 bits.
UICC-ME interface	No change
ME	No change due to 256-UMTS AKA algorithm

7.2.4.1 Track of decision

It was the decided that the introduction of a 256-bit key K should no be persued before the introduction of 256-bit encryption and integrity keys for EPS/E-UTRAN.

256-bit encryption and integrity keys for EPS/E-UTRAN will not be introduced in Rel-8.

7.3 Security set-up procedure

7.3.1 Security Mode Command

7.3.1.1 Separated Mode

In this so-called “separated mode”, SMC in LTE/SAE would be used in the way similar to UMTS. SMC would be performed between initial L3 message and its response message. In this case, SMC would be used to start/modify security configurations between UE and network. SMC would be integrity protected. Ciphering may also be performed after SMC. In addition, integrity check of signallings between UE and network could be used to authenticate UE when there is no authentication procedure between initial L3 message and its response message.

7.3.1.2 Combined Mode

Figure 11 is cited from section 7.14 of TR 23.882 v1.6.1:

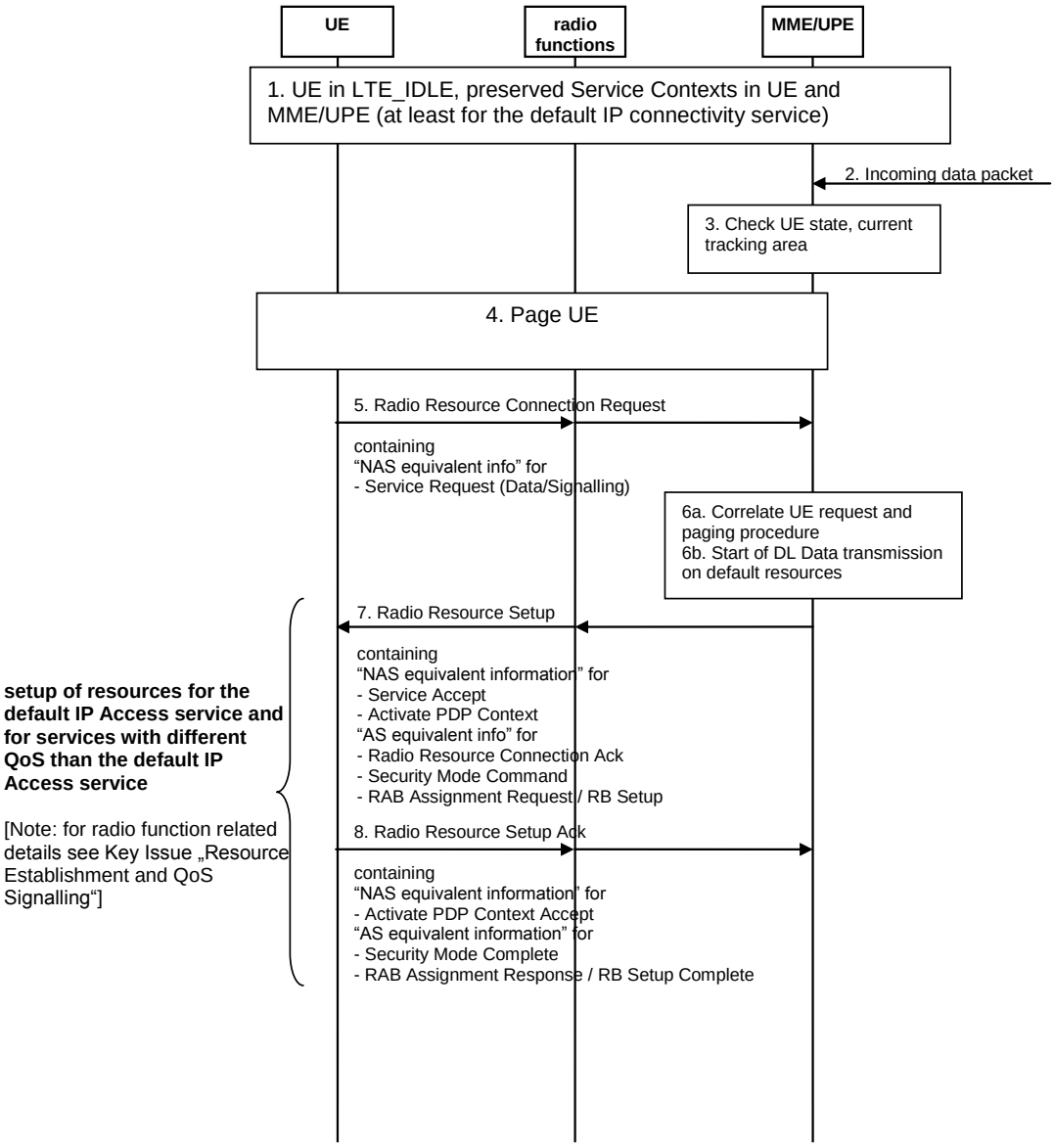


Figure 11 Combined Mode

A so-called "combined mode" concept for SMCs on idle to active transitions could be derived from the above figure. In this mode, SMC is combined with initial L3 message, its response message and acknowledge to response message.

The advantage of "combined mode is that the latency could be reduced since number of round trips would be decreased. Low latency is an important goal of LTE/SAE.

However, combined mode would meet with some security issues if used during initial attachment:

1) If combined mode was used on initial attachment, TMSI would be sent in its response message. But at this moment, because cipher algorithm negotiation has not been finished, response message could not be ciphered even if NAS keys are shared between UE and MME already. So TMSI could not be protected if TMSI is sent with response message. In this case, some solutions are needed to cipher TMSI.

2) If combined mode was used on intial attachment, some signalling procedures may be performed in network side before network send response message of initial L3 message to UE. If these signalling procedures would require integrity protection, a problem would arise as, UE would not be able to start integrity protection before the acknowledge message of the initial L3 message. In addition the network could not verify that the authenticate UE is in possession of the right integrity keys until the network received the acknowledge message from UE. Attackers may explore this weak point to perform DoS attack.

The above considerations show, that before the allocation of a new TMSI, the security mode command round trip has to be completed and, more generally, that the security mode command message cannot be combined with messages that require encryption. In particular the combined mode cannot be applied in case of initial attachment or re-attachment (Cf. also next remark.) However, it is ffs if SMC on idle to active transitions can be combined with the radio-resource-setup roundtrip.

In addition the security mode command procedure adopted by SA3 should take the following consideration into account. The Security Mode Procedure in UMTS was copied from GSM where they had performed a careful study about when to initiate ciphering. The conclusion was that ciphered messages should be sent only after confirmation was received from the other side that ciphering had been started. Otherwise, lots of ciphered messages could be sent from one side before an error was noticed. If we want to change this sequence of events, we need to analyse that nothing bad can happen.

7.3.2 Alternative not using Security Mode Command (SMC)

Editor's Note: Need evaluate further how this alternative works to avoid error situations.

7.3.2.1 Validity of the security association

RAN2 at its recent meeting confirmed their general understanding of the following:

A default bearer is established at the time of Attach. The relevant security context has to be established and started prior to the establishment of the default bearer. The security association is then retained at the aGW and the UE for the NAS signalling and for the user data until the UE detaches from the network.

The RRC security context is established and started whenever the RRC connection is established. The security association is retained and transferred from one eNB to another during handover until the RRC connection is released.

7.3.2.2 Start of Encryption and the Encryption of NAS message contents

In UMTS, encryption is started or reconfigured using an explicit Security Mode Command. The use of explicit signalling messages adds to the delay in the establishment of the procedures. This is perhaps more relevant when there is a reconfiguration.

Since LTE encryption of the NAS messages in the MME is independent of the encryption of the user plane, it brings new flexibility. In UMTS, RLC performs the encryption and RLC treats signalling identically to UP. In LTE MME does the encryption on its own and it only needs to encrypt NAS messages - so the encryption functionality can be tailored for this purpose.

Each LTE NAS message can be ciphered individually in the MME. It is possible to have non-ciphered header information elements in each NAS message and a ciphered remainder of the NAS message. This unciphered header information elements can be used to carry information such as Start ciphering, Security algorithm being used etc. So, for example, the new MME in its first NAS message could indicate to start ciphering and the security algorithm being used for this NAS message. This not only eliminates the need for an explicit security mode command procedure but also gets around the problem of MME relocation involving a change in ciphering algorithm. It also avoids the complexity of having to handle the Security mode command with a future activation time because the security is started on receipt of this message itself.

Similar principle could also be used for RRC messages. This also provides RRC with the independence in terms of security configuration and can avoid interactions between NAS and RRC to control the start of RRC encryption, algorithms etc.

7.3.3 Establishment of a security context

This section lists the potential SAE/LTE procedures that involve an establishment of a security context. For each procedure it is noted whether UE-eNB and/or UE to MME security context needs to be established and also discuss the importance of the procedure in SAE/LTE

1) Attachment

   - Both UE-eNB and UE-MME security needs to be established

- Mandatory

2) Idle to Active

- UE-eNB security must be established
- UE-MME may need to be established

  This is not expected to be frequent as there will be little NAS traffic, but it is not clear that it will never need to happen, e.g. errors cases when mobile went into Idle during a Security mode procedure
- Mandatory

3) eNode B handover without MME change

- Only UE-eNB security must be established
- Mandatory

4) eNode B handover with MME change

- UE-eNB security must be established

  This could be done using a key derived from the key at the old eNode B or with a key derived from K_{ASME} at the MME. The latter provides the stronger security as it provides cryptographic separation from the previous eNode Bs. The solution chosen would be a balance between the security gained against any complexity added.
- UE-MME security must be established

  This could be done by the passing of the NAS security context or may require a new context to be established if the MMEs support different security algorithms
- Mandatory

5) Idle Mobility

- UE-MME security must be established

  This could be done by the passing of the NAS security context or may require a new context to be established if the MMEs support different security algorithms
- Mandatory

6) Fresh AKA run

- Both UE-eNB and UE-MME security needs to be established
- Mandatory

7) Changing security context (key plus optionally algorithm) in Active without changing K_ASME

Editor's note: it's FFS whether this case is needed.

- UE-eNB security may need to be established

  This procedure would ensure that there is cryptographic separation from the previous eNode Bs. If a UE has been in Active for a long time and/or has been connected with many different eNode Bs, then there is an increased risk of compromised keys. The frequency of the used of this procedure would depend on Operator policy of the perceived risks of a UE in Active over a long time.
- UE-MME security may need to be established

There is not as strong a requirement for this case as the above as the amount of NAS traffic is small and the risk of compromise of NAS keys should be much lower.

- Network initiated refresh of UE-eNB security is strong requirement with support of the other cases not so strong.

8) Refreshing the keys at an eNode B

- Only UE-eNB security must be established

  This case needs to be mandatory if there is any risk of reaching the limit of ciphering under one key. Otherwise this procedure would be an Operator to set a limit on the amount of material protected a key.

- Mandatory if the ciphering limit could be reached.

7.4 Key handling

7.4.1 UMTS AKA

UMTS AKA is able to agree one pair of CK and IK. But there are more than one security associations in LTE/SAE. UE and network deduce the keys for AS, NAS and user data protection. The keys are delivered to the corresponding entities in the network side.

Editor's Note: it was decided not to use the following solution. Document TD S3-070095 contained the reasons why this was not chosen.

Two possible solutions as listed following can be used to generate and deliver the keys for UP security, NAS security and AS security to corresponding network entities.

1) More key generation functions are implemented in UE and HSS. UE and HSS are able to use these key generation functions to generate keys for NAS security, AS security and UP security. These keys are encapsulated into Authentication Vector. HSS sends these keys to MME. MME delivers these keys to corresponding entities which perform security operation.

2) UE and HSS deduce CK and IK as root keys. MME gets root keys from HSS. Based on the root keys, MME and UE deduce the keys for NAS security, keys for AS security and keys for UP security. MME delivers these keys to corresponding network entities which perform security operation.

Editor's Note: The better expression other than "root key" is needed.

Editor's Note: The backward compatibility on key generation and delivery should be guaranteed.

7.4.2 Serving Network Authentication for LTE[17]

7.4.2.1 Introduction

According to 3GTS 33.102, UMTS provides network authentication in the following sense:

Network authentication: the property that the user corroborates that he is connected to a serving network that is authorised by the user's HE to provide him services; this includes the guarantee that this authorisation is recent.

This means that the UMTS user obtains some guarantees about the authorization of the serving network, but he does not authenticate the serving network, i.e. he cannot corroborate its identity. UMTS has the further property that session keys are not bound to particular serving networks: an authentication vector may be used by a VLR or SGSN in any serving network. It may also be forwarded between serving networks. This is as in GSM, but is different from EAP: in the EAP framework, keys must not be shared among authenticators.

---

[17] This section is from S3-060716.

It is the purpose of this section to clarify the question whether in SAE session keys should be bound to serving networks identities and whether serving network authentication should be provided in SAE.

The scope of the discussion in this section is meant to apply to 3GPP networks. Abbreviation: we use SN for serving network in the sequel.

7.4.2.2 Threats

The following discussion applies to UMTS and, if a similar authentication approach as in UMTS was adopted, also to SAE.

When there is no SN authentication a user has no assurance to which SN he is connected. This may matter to the user because he may have preferences in network selection due to differences in e.g. security levels, tariffs. There is an attack if someone actively deceives the user about the SN identity. This requires the attacker to use a false base station and to broadcast a false SN identity on the radio interface or in the network information transferred during registration. For the attack to be successful, this SN identity has to have higher priority on the list of operators in the UE than all the other networks whose signals can be received by the UE, cf. next section

As usual in false base station attacks, there are two modes of operation in which the attack can be conducted:

a) the false base station may act as a relay towards a target network forwarding all traffic transparently, either by connecting to the target network through the Iub interface (requiring the target networks consent), or by acting as a UE towards the target network. Once, ciphering has been switched on the false base station sees only encrypted traffic.

b) the false base station may act as a UE towards the target network and assume the roles of NodeB and RNC towards the attacked UE. In particular, the false base station would terminate security in both directions.

Discussion:

None of the variants of the attacks would allow call theft, but eavesdropping and potential financial gain for the operator need to be discussed a little more:

Mode of operation a) is technically always possible, but of limited gain. A target operator could try to attract more users for financial gain, but it is very unlikely that this could go on at a commercially significant scale without being detected. The risk for the target operator is high: loss of roaming contracts. Therefore, the UMTS business model suggests that this attack is of no practical relevance in UMTS. But business models may change in SAE, e.g. because smaller operators having no long-term agreements with a home operator could be dynamically authorised to provide service to the home operators users, cf. the current discussion in SA1 on network composition as documented in TR 22.980.

Furthermore, an attacker may use mode of operation a) without the involvement of the target operator. Then there is no financial gain for the attacker, but there could be a potential motif for the attacker if the target network did not employ encryption. Then eavesdropping would become possible. An attacker would be likely to target specific users. For the attack to be successful, he would have to follow the victim around and wait until the victim makes a call while having the false base station up and running. Furthermore, the attacker would have to eavesdrop on the unencrypted radio interface to which the false base station is connected. (Remember the conditions of mode of operation a) .) The impact of the attack can be mitigated by the fact that a user may be warned by the ciphering indicator on his UE.

Mode of operation b) becomes possible only when an attacker can steal authentication vectors from a compromised network. Then it would not even be necessary for the attacker operating the false base station to connect to a real network, provided he could fake all the expected responses to the victim UE. Everything said above about the conditions for successfully operating a false base station holds also here, and the consequences of b) would again be eavesdropping and financial gain. Furthermore, the use of authentication vectors from a compromised or malicious network on a larger scale is quite unlikely to go undetected in the long run. The long term business relations with other operators embodied in roaming agreements are vital for operators in the UMTS business model. But again, we are not sure whether this still holds in SAE.

In addition, in UMTS AKA the use of a stolen authentication vector is limited to one instance of service provision, and it becomes unusable when its sequence number is too old to be accepted by the USIM. This considerably limits the scope of the eavesdropping attack, and stops the attack completely as soon no fresh authentication vectors are stolen. (Actually, this is the reason why sequence numbers were introduced. Other types of false base station attacks could have been countered by signalling integrity alone.)

Similar threats were described in the paper [4].

Summing up: there is a theoretical threat which, in UMTS, seems of little practical relevance. This may, or may not, change in an environment in which SAE may operate in the future. In how far serving network authentication could help here, is considered in the next section.

7.4.2.3 Countermeasures

Mechanism to provide serving network authentication

This section looks at the mechanism which could be used in SAE to provide serving network authentication in case it was decided to have this feature in SAE.

One apparently straightforward way of providing serving network authentication would be the following:

SN authentication could be achieved in SAE/LTE if SN-specific confidentiality and integrity keys CK', IK' were derived from CK, IK and the SN identity. The UE could perform this key derivation using the SN identity as received over the radio interface. The MME would obtain CK', IK' from the home environment and would not see CK, IK. This would then achieve SN authentication in the following sense: If the keys do not match on the UE and SN side, which would be the case if the SN identity was faked towards the UE, communication is not possible.

A prerequisite for this approach is, of course, that the SN identities (e.g. MCC+MNC) seen by the UE and by the home environment are the same. It should be noted here that an operator may use the same identity for GSM, UMTS and LTE. The SN identity must be able to be transported to the home environment in the correct protocol layer. This was a problem with earlier versions of MAP, but should not be a problem in SAE any more. Furthermore, it is a prerequisite that the home environment can authenticate the SN. (This cannot be taken for granted as we know from the discussions about MAP security.) The UICC and the Authentication Centre could operate as in 3G Release 99 if the derivation in the UE was performed by the ME, and the derivation in the home environment was performed by a key derivation server in front of the Authentication Centre. This key derivation server could be part of a AAA server, but it could also be included in the Authentication Centre or HSS in other ways. In particular, load balancing by pre-computation of authentication vectors in the Authentication Centre would still be possible. The UMTS AKA protocol would not change, only an additional key derivation step would be introduced after the completion of the authentication and key agreement protocol and before the use of confidentiality and integrity keys.

The use of different index values for different service domains in the array scheme for UMTS AKA sequence number management, as in TS 33.102, would still be possible, and would be independent of SN authentication. (Cf. Annex C.3.4 of TS 33.102: "Authentication vectors distributed to different service domains shall have different index values (i.e. separate ranges of index values are reserved for PS and CS operation)."

The MME would then further derive the keys required for the security associations on the links MME-UE, and eNB-UE from CK', IK', and distribute these derived keys to SAE gateway and eNB. The UE would derive these keys in the same way. It must be, of course, ensured that the user and the network side derive the same keys unambiguously. In particular, the UE must know whether such key derivation is required (i.e. in LTE) or not (in UTRAN).

Limitations of countermeasure: issues with serving network authorization

Mutual authentication is a feature, which is frequently demanded, with little regard to whether the result of the authentication can be useful to the verifier of the authenticated identity. To explain: it is all very well that the user, or his equipment, may be given the possibility to cryptographically verify the identity of a network he is connected to. But what does the user, or the UE, then do with the verified identity? How can the user, or the UE, decide whether this particular network is a network the user wants to connect to? In other words, the question is how the user can decide which network is authorised by him to serve him. (He already knows from UMTS AKA that the network is authorised by his home operator to serve him.)

UE and (human) user are condisered separately:

Authorisation of the SN by the UE:

In UMTS, PLMN selection is either manual or automatic. A UE can automatically perform network authorisation by checking the authenticated SN id against one of the lists used for PLMN selection. The USIM carries two ordered lists: one user controlled list and one operator controlled list of PLMNs. In automatic mode the UE first tries to connect to HPLMN, then in priority order to one of the PLMNs in the user controlled list, then to a PLMN in the operator controlled list and finally to the other available PLMNs in order of quality of radio reception. In particular a UE that receives its HPLMN with sufficient quality will always camp on one of its HPLMN's base stations.

The value, which serving network authentication adds here, is that the attacker can no longer broadcast the network identity of an SN high on the UE's priority list. His false base station has rather to blind out signals from SNs with higher priority on the list (assuming that the attacker could not steal authentication vectors from a high priority network). This makes the attacker's job technically more difficult, but not impossible, and may pose no restriction at all in certain roaming situations.

In addition to today's automatic network selection procedures in UMTS, one could think of plausibility checks of SN identity against other data reliably available to the UE, i.e. the identities of SNs of neighbouring cells, or geographical information. One could, e.g. think of comparing Mobile Country Codes in the SN identity against GPS information available in the UE to detect a mismatch. But the practicality of this kind of plausibility check would require much more study, and it would probably offer only limited protection. Authorisation of the SN by the user:

In manual network selection mode the user is presented a list of available networks ordered according to the same priorities as in automated mode. The user then selects the PLMN he wants to connect to from this list. With SN authentication, the user could be sure to be presented the correct identities. More generally, with SN authentication the user could always see a verified identity of the current SN on his display. But also this property is of limited value:

First of all, it has been a good principle in UMTS and GSM, not to encumber the user with security decisions. (Please remember the discussion in SA3 in the context of rejection of non-ciphered calls, where it was argued by operators that this feature was undesirable from a customer service point of view.)

Furthermore, according to TS 22.101, Annex A, the serving network can send Network Identity and Timezone (NITZ) during the registration, and then NITZ would be displayed to the user, and not PLMN names stored in the ME. This is so in order to ensure the most up-to-date information on the serving networks. In roaming situations SN identities and SN names may be often quite meaningless to the user as they may have never heard of them. The display of the country name is currently optional. It may certainly help to make the display of an authenticated country name mandatory, but the user may still easily overlook it. Hence authorisation of the SN by the user looking at the operator name on the UE's display has practical limitations.

Finally, it is seen as the prime interest of a user (and his home operator) that the user is able to obtain service anywhere at any time. A user may have a list of preferred operators, but a user cannot distinguish whether his preferred operator is not able to provide service e.g. due to a lack of coverage or overload, or whether its base station is blinded out by the false base station of an attacker. So, if service by the preferred operator is not available the user faces the choice to not obtain service or connect to another serving network. (Remember that the 3G AKA protocol guarantees that a user cannot be connected to just any serving network, but only to serving networks authorised by the user's home environment). The obvious choice from a marketing point of view can only be to permit connection to a non-preferred serving network. Serving network authentication then does not help here if the attacker's false base station can blind out the preferred SN, and if the (true) identity of an SN, from which the attacker may have stolen authentication vectors, does not alarm the user.

Summing up: while section 2 showed that serving network impersonation attacks are theoretically possible, but of little practical relevance in UMTS, this section showed that serving network authentication would have only limited value to counter these attacks for want of practical serving network authorisation. This is the main reason why serving network authentication was not introduced in UMTS although it would have been technically possible.

Scope of countermeasures: mobility aspects

Everything which was said in this section relates to authentication and key agreement. The described attacks assume that the victim user wants to register with a particular network and has to perform authentication. But in UMTS, it is possible to hand over to another SN without authentication. Rather, the session key CK, IK are transferred to the target SN.

It should be discussed in SA3 whether there is a risk in this way of doing handover, which would warrant a handling different from UMTS. It's stressed here that this is a different discussion. The use of SN-specific session keys and SN authentication in LTE may still be compatible with the forwarding of such keys to different SNs in handovers. But other alternatives are also conceivable which have a less severe impact on handover performance than re-authentication during handover. This is ffs.

Unused authentication vectors should probably not be forwarded from one SN to another. This is ffs.

7.4.2.4 Conclusions

This section shows that there are certain theoretical attacks in UMTS and SAE (if a similar authentication approach as in UMTS was adopted for SAE) which exploit the fact that authentication vectors can be used in any serving network,

and that UMTS does not provide serving network authentication. It was also shown that the attacks are of little practical relevance in UMTS, for which one of the reasons is the UMTS business model, and that serving network authentication would only provide limited protection against the residual risk due to practical difficulties with serving network authorisation. This trade-off led to the decision for UMTS not to introduce serving network authentication.

But in SAE, business models, trust relationships and roaming agreements may change. Hence, it may be worth looking into the possibility to provide serving network authentication in SAE/LTE. It was also shown here that, if desired, this could be done with little effort. Handover aspects should be considered separately, and would not necessarily be affected by a decision in favour of serving network authentication.

7.4.3 Key derivation

7.4.3.1 Key generation during initial access

Figure 12 shows an overview of SAE initial access authentication signalling and where the different keys are proposed to be generated. KDF (Key Derive Function) is used to derive different keys. After successful authentication, MME and UE will use KDF to generate keys based on CK, IK and RAND agreed during AKA authentication procedure.

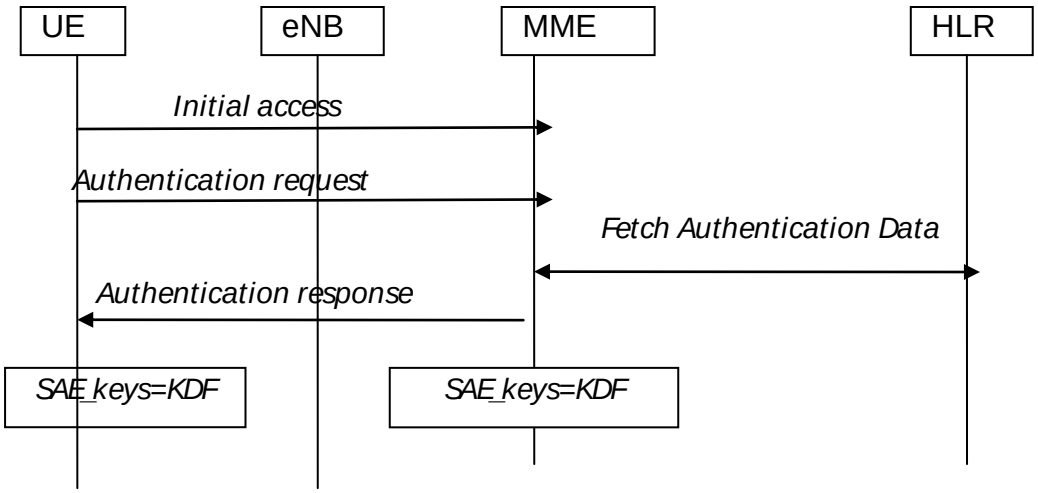


Figure 12 key generation for initial access

Many different KDF functions would be applicable for the purpose. Only three examples of them are listed:

Alternative-1: Specified KDF function in TS 33.220 Annex B.

SAE_keys = KDF (Ks, "static string", RAND, IMPI, SAE_Ids)

Where:

Ks is generated by concatenating CK and IK. IMPI could be obtained from the IMSI as specified in TS 23.003. SAE_ids could be e.g MME_id, eNB id and SAE-GW_id or MME's, eNB's and SAE GW's names.

SAE_keys will express then MME_key, UP_key, RRC_key.

"static string" could be "LTE_CK" and "LTE_IK" to generate CKs and IKs.

Editor's Note: It is for further study whether the identities of the nodes would be used.

Alternative-2: Uses Milenage f3 and f4 to derive keys

SAE_keys(K)=KDF(K, RAND XOR SAE_ids)

Where: SAE_ids could be e.g. MME_id, eNB id and SAE-.GW_id or MME's, eNB's and SAE-GW's names.

When K=CK, SAE_keys(CK)= f3(K, RAND XOR SAE_ids)

When K=IK, SAE_keys(IK)= f4(K, RAND XOR SAE_ids)

Alternative-3: (From S3-060692)

In this alternative, the LTE/SAE system uses the UMTS AVs and derives the other keys as follows:

CKNAS || IKNAS || CKAS || IKAS || CKUP = prf+ (Identity of UE || IK || CK)

The keys are derived in the MME and in the UE, after successful AKA procedure.

7.4.3.2 Key distribution during handover in inter-RAT

Editor's Note: This section was harmonized with Section 7.4.4 and the updated information was included in Section 9 and 10 of TS 33.abc

Continuous ciphered mode should be maintained during inter-RAT handover from E-UTRAN to UTRAN if ciphering has been activated and ongoing in E-UTRAN. The topic is current under discussion in RAN2. Thus, the distribution of security data (unused authentication vectors and/or current security context data, e.g. used CK, IK etc.) between SGSNs and MME should be discussed in SA3 as well. This section proposes to transfer security context in similar way as used between GERAN and UTRAN.

The following cases are distinguished related to the key conversion and key transfer of inter-RAT handover.

Case 1 Handover from LTE to 3G/2G:

LTE to 3G: MME convert Kasme with an one-way function to Ck,Ik. Then MME sends the Ck,Ik pair to 3G release8 SGSN as it's Ck,Ik pair.

LTE to 2G: MME convert Kasme with a one-way function to Kc. Then MME sends Kc to 2G SGSN.

Case 2, Inter –RAT Handover : 3G/2G -> LTE

Authentication vectors could be transferred from 2G/3G SGSN to MME. After MME received security context, e.g. CK, IK, it should be able to derive SAE_keys=KDF. KDF could be one of two KDFs described in 7.4.3.1.

Note: In addition to above cases, security context transfer from one MME to another MME in a PLMN might happen as well, however it is considered as rare case. Of course, security data could be distributed. In this case only SAE_key for NAS signalling needs to be updated with a new MME_id.

Editor's Note: AV forwarding is FFS.

Editor's Note: it's FFS there is difference when the SGSN is Rel8 or pre-Rel8. The prerefence is the differenct is transparent to UE.

7.4.4 Key management aspects for LTE/UMTS interworking[18]

Editor's Note: This section was harmonized with Section 7.4.3.2 and the updated information was included in Section 9 and 10 of TS 33.abc

- that LTE MME shall implement strong backwards key separation towards legacy systems,
- that Rel8 SGSNs shall implement strong backwards key separation towards LTE,
- that possibilities for forward security between Rel8 and LTE are to be further studied.
- LTE MME shall implement strong key conversion function when transferring key contexts to UTRAN, i.e. $K'_{ASME} = f(K_{ASME})$ and that Ck, Ik is derived from K'_{ASME}.

---

18 This section is from S3-060704.

7.4.5 Void

7.4.6 Key identities in LTE/SAE

Key Set Identifier (KSI) is used to identify the AKA quintet in UMTS networks (see TS 33.102). In UMTS the purpose of the Key Set Identifier is to make it possible for the network to identify the cipher key CK and integrity key IK which are stored in the mobile station without invoking the authentication procedure. This is used to allow re-use of the cipher key CK and integrity key IK during subsequent connection set-ups.

In UMTS security context is a state that is established between a user and a serving network domain as a result of the execution of UMTS AKA. At both ends "UMTS security context data" is stored, that consists at least of the UMTS cipher/integrity keys CK and IK and the key set identifier KSI. One is still in a UMTS security context, if the keys CK/IK are converted into Kc to work with a GSM BSS (see TS 33.102).

The amount of data that is protected by different key pairs can be different. It is easy to understand that the amount of data that is ciphered by $K_{eNB\text{-}UP\text{-}enc}$ can be more than the amount of data that is protected with $K_{eNB\text{-}RRC\,enc}$/$K_{eNB\text{-}RRC\,int}$ or K_{NASenc}/K_{NASint}. A mechanism that allows updating these key pairs separately would address the difference in amount of data that is protected with these keys. Such a mechanism could make LTE/SAE more flexible and reduce the unnecessary changes of security configurations, however there is a cost associated with it. Along these lines two options are shown below on how to identify the keys if they need to be identified separately or not.

- **K_{ASME} key identity only (security context identification)**

  In this case, K_{ASME} and hence the security context in the network and the UE will have an identity called KSI_{ASME}. If any of the derived keys needs to be invalidated, KSI_{ASME} is set to "111" and authentication shall be performed to update all the keys (i.e. create a new valid security context). A new KSI_{ASME} is then stored in MME and UE and delivered by the network for the UE.

- **Separate key identity for K_{ASME}, RRC, NAS and UP keys**

  In this case the identities are called KSI_{ASME}, KSI_{NAS}, KSI_{RRC} and KSI_{UP} respectively. If one of these keys (or key pairs) is invalid, the corresponding identity would be set to "111" to inform MME. For example, if key pair $K_{eNB\text{-}RRC\,enc}$/$K_{eNB\text{-}RRC\,int}$ is invalid, KSI_{RRC} would be set to "111".

  Similarly as in case of KSI_{ASME} identity only if K_{ASME} is invalid, KSI_{ASME} would be set to "111" and authentication procedure shall be performed to update all the keys. How to use the different key identities is out of the scope of this section..

Key hierarchy working assumption in Section 7.4.7 assumes that all LTE keys are derived based on a K_{ASME}. The key hierarchy does not allow, as is, explicit key updates, but RRC and UP keys are derived based on the K_{eNB} and certain dynamic parameters (like C-RNTI), which result as fresh RRC and UP keys in the eNB between inter-eNB handovers and state transitions. The K_{eNB} is not stored in eNB while UE is in idle mode. It seems enough to identify the security context with KSI_{ASME} that includes the K_{ASME} as all keys are derived from K_{ASME}. Thus, the original purpose of identifying CK and IK in UMTS with the KSI is similarly fulfilled by identifying K_{ASME} with KSI_{ASME}.

If RRC/UP keys are corrupted (e.g. ciphering/integrity fails continuously, keys are missing in UE/eNB, C-RNTI contained bit errors, etc.) UE will have to restart radio level attachment procedure (e.g. similar radio level procedure to idle-to-active mode transition or initial attachment). In case K_{ASME} is invalid KSI with value "111" is sent to the network, which then can initiate (re-)authentication procedure to get a new K_{ASME} based on a successful UMTS AKA authentication.

Editor's Note: Key change on the fly may require that a separate K_{eNB} identifier from the K_{ASME} identifier is needed but this is for further study.

Editors Note: It is for further study if a separate K_{eNB} identifier is needed for the purpose of separate lifetime handling of the K_{eNB} from the K_{ASME} lifetime handling.

A detailed solution in case of multiple KSI-identifiers is described below:

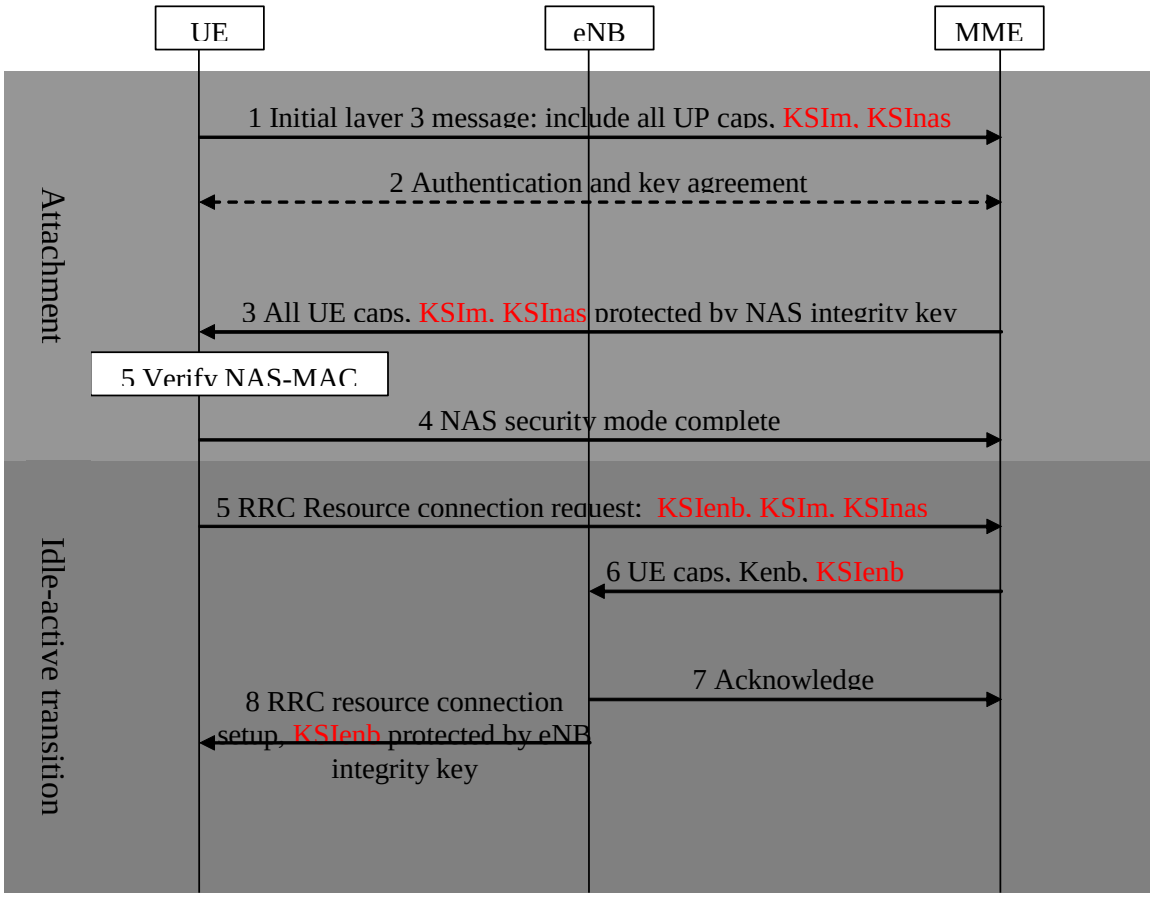


Figure 13 multiple KSI-identifiers

1. KSIm and KSInas were sent to MME. If there is no available master key in UE, UE would set KSIm to "111". If there is no available nas key in UE, UE would set KSInas to "111".

2. The KSIm "111" would trigger AKA procedure. If only KSInas was set to "111", AKA procedure would not be performed. After AKA procedure, K_{ASME} is shared between UE and MME.

3. If KSInas in step2 was set to "111", new keys for NAS signalling are derived. New KSInas was also generated. MME would send KSIm and KSInas to UE with integrity protection. If new NAS keys are generated, new KSInas would be send. The message should be integrity protected with new NAS key.

4. UE verify NAS-MAC. If NAS key are generated, UE would derive new NAS keys. UE send NAS security mode complete to MME.

5. In idle-active transition, KSIenb was send to MME. If keys are unavailable, related key identity would be set to "111". KSInas and KSIm may also be sent to MME in this message.

6. If KSIenb was "111", new keys for eNB would be generated based on master key. New KSIenb would be sent to eNB.

7. eNB send acknowledge to MME.

8. eNB send KSIenb to UE. If new keys for eNB were generated in step8, new KSIenb would be send.

7.4.7 Hierarchy of user-related keys in SAE/LTE[19]

7.4.7.1 General

This section deals with the **establishment** of user-related keys in SAE/LTE. User-related keys are keys shared between the UE and a network entity. It's considered the establishment of keys shared with entities in or at the border to the SAE core network and keys shared with entities in the LTE access network. There isn't consideration the establishment of any keys shared with entities inside non-LTE access networks here as these are dealt with in the standards relating to these other access networks. In order to have a common name for a key management entities at the border to the SAE core network we **define**:

*An **Access Security Management Entity** (ASME) is an entity which receives the top-level keys in an access network from the HSS.*

For LTE access networks, the role of the ASME is assumed by the MME. This is the only case, for which detailed information is available at present and which we consider in this section. Another example for an ASME may be an AAA server or a gateway residing in the home or visited network and serving a non-3GPP access network, e.g. a WiMAX network. (In the work on EAP-VKH [S3-060662], it has been proposed to also consider AAA servers in the visited domain. This is ffs.)

If the access network is UTRAN the key hierarchy proposed here does not apply, as it shall be possible for a legacy UMTS UE to attach to any UTRAN even if the UTRAN is connected to an MME. (With this statement, we want to allow for the possibility that an MME may also have the functionality of a 3G-SGSN to which a UTRAN is attached. This isn't required, though. This would be similar to the situation in UMTS, where a GERAN may be attached to a 3G-SGSN.)

New user-related keys will be established as a result of a new run of the user authentication and key agreement protocol. In particular, an AKA shall be run at initial attachment. But not all keys in the SAE hierarchy will necessarily be established at the same time. E.g. RRC keys may need to be established only when switching to active mode.

It's assumed, in accordance with the decisions of 3GPP SA3 that AKA is used for user authentication. Our considerations do not depend on the decision between UMTS AKA and EAP AKA, which is still pending at the time of writing this section.

This section does not deal with conditions for when to run AKA. Such conditions (e.g. operator defined conditions, conditions depending on active to idle transitions, conditions depending on timers, e.g. for connections of long duration) will have to be decided upon separately.

Key lengths are not considered in the present version of the section, but fit with the scope and could be added later.

Key derivation functions are not considered in the present version of the section, but fit with the scope and could be added later.

The focus on user-related keys implies that network-domain security in SAE is outside the scope of this section.

This section does not consider key handling on mobility events within an access network or between different access networks. This key handling will be addressed in separate sections. Key handling at mobility events may consist in a mere transfer of an already established key, or in a further key derivation from an already established key, or in a new run of the authentication protocol.

It makes sense to consider key establishment separately from key handling at mobility events because

- it helps the analysis and presentation of key-related issues by breaking the problem down into smaller problems;
- it allows to take into account the potentially different trade offs between risk of key compromise and complexity or performance for key establishment and handling of already established keys.

7.4.7.2 Proposed hierarchy of user-related keys in SAE/LTE

Keys for all SAE access networks:

---

19 This section is from S3-070095.

Keys shared between UE and HSS:

- **K** is the permanent key stored on the USIM and in the Authentication Centre AuC
- **CK, IK** is the pair of keys derived in the AuC and on USIM during an AKA run. CK, IK shall be handled differently depending on whether they are used in an SAE context or a legacy context, as follows:
    - If the AKA is run over LTE or a non-3GPP SAE access network, CK, IK shall not leave the HSS.
    - If the AKA is run over a UTRAN access network, according to 3G TS 33.102, or a WLAN according to 3G TS 33.234, then CK, IK shall be transferred from the HSS to VLR, SGSN, or AAA server respectively.
      Note: whether this applies even to UTRAN attached to MME or a Release 8-SGSN is ffs. If it does not then the ME needs to be able to signal its capability to perform SAE key derivation.
    - CK, IK from an AKA run in one context (SAE or legacy) shall not be usable in key establishment procedures in the other context. The UE shall be able to check this condition.

Intermediate key shared between ME and ASME:

- **K_{ASME}** is a key derived by UE and in HSS from CK, IK during an AKA run. K_{ASME} shall depend on the type of the radio access technology. If the RAT is LTE type then K_{ASME} shall also depend on the PLMN identity (MCC + MNC). If the RAT is not LTE type then it is ffs what a PLMN identity known to UE and HSS could be. The identities become known to the UE during the attachment procedure. They are transferred from the ASME to the HSS as part of an SAE-specific authentication vector request. (Which protocol will be used in SAE for authentication vector requests, and how the above mentioned identities are carried in this protocol, is ffs.) The key K_{ASME} is transferred from HSS to ASME as part of an SAE-specific authentication vector response (remember that, for LTE, the MME is the ASME. Other cases are ffs).

Keys for LTE access networks:

Intermediate keys:

- **K_{eNB}** is a key derived by UE and MME from K_{ASME}. K_{eNB} may only be used for the derivation of keys for RRC traffic and the derivation of keys for UP traffic. K_{eNB} shall depend on the identity of the eNB requesting it from the MME.

Keys for NAS traffic:

- **K_{NASint}** is a key derived by UE and MME from K_{ASME}. It may only be used for the protection of NAS traffic with a particular integrity algorithm.
- **K_{NASenc}** is a key derived by UE and MME from K_{ASME}. It may only be used for the protection of NAS traffic with a particular encryption algorithm.

Keys for UP traffic:

- **K_{UPenc}** is a key, which may only be used for the protection of UP traffic with a particular encryption algorithm. This key is derived by UE and eNB from K_{eNB}, as well as an identifier for the encryption algorithm.

Keys for RRC traffic:

- **K_{RRCint}** is a key, which may only be used for the protection of RRC traffic with a particular integrity algorithm. K_{RRCint} is derived by UE and eNB from K_{eNB}, as well as an identifier for the integrity algorithm.
- **K_{RRCenc}** is a key, which may only be used for the protection of RRC traffic with a particular encryption algorithm. K_{RRCenc} is derived by UE and eNB from K_{eNB} as well as an identifier for the encryption algorithm (ffs).

Editor's note: It is ffs wether or not the same ciphering key can be used for the encryption of RRC and UP traffic in eNB.

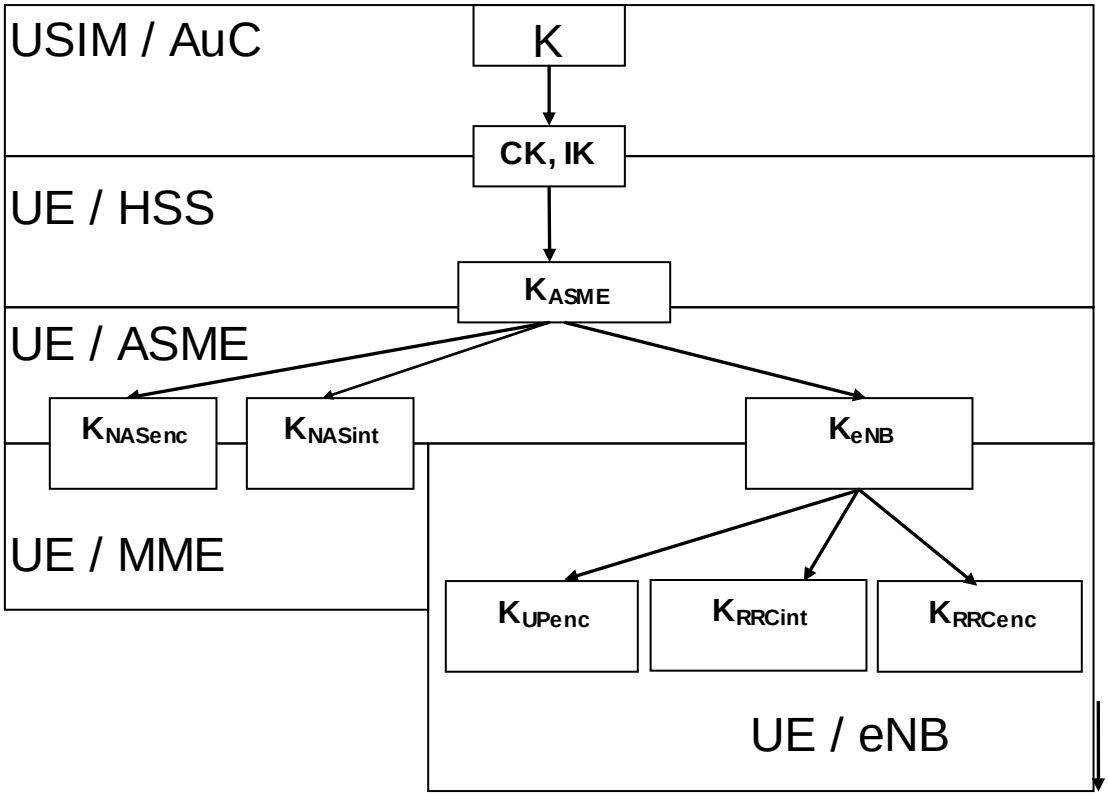


Figure 14: Overview on proposed key hierarchy

7.4.7.3 Justification of proposed key hierarchy

Editor's note: key hierarchy should not be used for UTRAN.

7.4.7.3.1 Binding of a context to a key:

As a guiding principle we propose to

- *bind context information to an established key in such a way that the compromise of the key cannot be exploited by an attacker in a key establishment procedure in a different context.*

The decision which context information to bind to the key depends on a trade-off between the reduction of risk achieved by the binding and drawbacks, if any, regarding e.g. complexity or performance, caused by the context binding.

The binding discussed in this section is meant to apply to key establishment. Whether this binding shall also imply that the use of this key is only allows in the context in which the key was established is a separate issue. When we propose here that the use of the key shall always be bound to the context of establishment we explicitly say so.

Example: a context to which a key is bound may be a PLMN identity. The binding may be achieved by deriving this key from a higher-order key using the PLMN identity as input. When this key is stolen from one PLMN then an attacker cannot use this key to impersonate another PLMN when the user tries to attach to this other PLMN. However, whether or not this key may be transferred to a different PLMN in handover is subject to a separate discussion. A different handling may be justified by a different trade-off between risk and performance.

7.4.7.3.2 Top-level key in the system

It follows from the SA3 decisions to use AKA for authentication and to allow Release 99 USIMs for access to SAE that the top-level user-related key in the SAE/LTE key hierarchy is the key K stored in USIM and Authentication Centre, as defined in 3G TS 33.102.

7.4.7.3.3 Binding CK, IK to SAE

In section 4 we proposed the following.

1. If the AKA is run over LTE or a non-3GPP SAE access network, CK, IK shall not leave the HSS.

2. If the AKA is run over a UTRAN access network, according to 3G TS 33.102, or a WLAN according to 3G TS 33.234, then CK, IK shall be transferred from the HSS to VLR, SGSN, or AAA server respectively.
   Note: whether this applies even to UTRAN attached to MME or a Release 8-SGSN is ffs. If it does not then the ME needs to be able to signal its capability to perform SAE key derivation.

3. CK, IK from an AKA run in one context (SAE or legacy) shall not be usable in key establishment procedures in the other context. The UE shall be able to check this condition.

The reasons for these requirements are explained in this section.

Requirement 2 above simply states that from Release 8 onward, authentication defined for pre-Release 8 access networks should continue to be done in the same way as before. This requirement is necessary because pre-Release 8 UEs shall be able to use these access networks even if the core elements to which they are attached are Release 8.

In order to explain the rationales for requirements 1 and 3 we need to provide some more background information.

The 3G TR 22.978 "All-IP Network (AIPN) feasibility study" explains the motivation and drivers for SAE as well as the expected changes in technology and business models. The expected changes in business models affect the trust models and threat analyses, which in turn provide the rationales for design decisions for a security architecture. These changes are therefore relevant in our context. In particular, 3G TR 22.978 states:

> *"..., an AIPN will need to follow architectural principles that facilitate operation of AIPN, access system and services by separate stakeholders."*

> *"With 3G and upcoming extensions of it, many new players will enter the scene. Small and very large AIPN operators and service providers will work together to offer the services the users expect in a competitive way. At the same time, the equipment of the end-users will become more complex and capable. ... In this environment, attacks may occur in many different places and in many different ways."*

> *"Transforming today's 3GPP system into an AIPN will introduce changes in the threat environment, introducing new threats but also changes in risk levels of already identified threats. Threats previously seen as having low risks may need to be reassessed leading to new security requirements and the need for new and/or improved security mechanisms. ..." One of the examples listed in this context is "System heterogeneity and multi-access (GSM, UMTS, WLAN, new accesses, etc)"*

3G TS 22.278 "Service requirements for evolution of the 3GPP system" goes one step further and derives one central requirement from the considerations in TR 22.978:

> *"Any possible lapse in security in one access technology shall not compromise security of other accesses."*

It's concluded from this that, in particular, a lapse in security in an LTE or any other SAE access technologies shall not compromise security of pre-SAE access technologies, and vice versa. In particular, the security lapse we discuss in this section is "stealing an authentication vector on SAE (or pre-SAE) networks and using it to impersonate a valid pre-SAE (or SAE) network.

Compromise of pre-SAE systems shall not affect SAE systems:

In 3GPP specifications before Release 8, 3G authentication vectors are handed out to various entities of 3G operators: VLRs, SGSNs and P-CSCFs in home and visited networks, S-CSCFs, I-WLAN AAA servers and BSFs in home networks. In HSPA, base stations will obtain CK, IK. With the decision of ETSI TISPAN to accept IMS AKA as their long term security solution, there is also the possibility for P-CSCFs serving fixed access networks to obtain 3G authentication vectors. As IMS AKA is access independent, P-CSCFs obtaining 3G authentication vectors may be, in principle, connected to any access network. If CK, IK were stolen from any of these entities they could be used in key establishment in an SAE network. In order to make use of stolen CK, IK an attacker would have to be able to set up a false eNB (at a bearable cost), attract the user to this eNB during the validity of the AV, and mount a network impersonation attack (for details of this discussion cf. S3-060716). The expected lower cost of LTE radio network equipment will make it easier to set up false eNBs, the expected larger amount of operators will make it more difficult to detect false eNBs and the lighter radio access network equipment will make it easier to set up false eNBs in the vicinity of victim users. Thus, without binding authentication vectors to their use within SAE, the effect of further key bindings within SAE could be easily defeated. E.g. it is argued in the next section that binding the PLMN identity to an LTE key in key establishment is useful. However, if keys are not

bound to SAE usage, CK, IK stolen from a UTRAN network could be used to impersonate an SAE network during user attachment.

This can be prevented if AKA authentication vectors given by the HSS to pre-SAE entities are verifiably different from those given to SAE entities and cannot be used in SAE systems.

Compromise of SAE systems shall not affect pre-SAE system:

The quotes from TR 22.978 above show that it is difficult to predict what business relations, and in particular trust relations, among operators we may assume for the lifetime of SAE systems. It seems likely that the current model of large operators with long-lived, stable business relations may not hold in the future. Some of the operators may be more trustworthy than others, and it may be difficult to assess their trustworthiness or rely on legal recourse when things go wrong. It therefore seems very advisable to design SAE in such a way that a security compromise in one SAE network affects the rest of the world as little as possible. Such a compromise should only minimally affect pre-SAE systems. In addition, other networks or RATs within SAE should not be affected either.

This can be prevented if CK, IK in AKA authentication vectors used for SAE never leave the HSS.

A possible mechanism to achieve a binding of AKA authentication vectors to SAE is the use of a bit in the AMF field.

It's assumed for this section, as for the next section, that a reasonable level of core network signalling security is provided such that e.g. HSS can authenticate the requesting PLMN or such that no AVs can be snooped in transit between home and visited network. Otherwise, it will be very difficult to guarantee good security with any architecture.

When trading off the expected security gain with the added complexity, we should also remember that, if we do not introduce key separation now, we will probably not be able to do it later for terminal backward compatibility reasons.

7.4.7.3.4 Binding top-level key for access network to PLMN and RAT

A detailed rationale why binding SAE keys to the PLMN identity during key establishment may be useful was given in S3-060716, which has become part of this TR.

The main reason given in S3-060716 was future-proofing SAE against network impersonation threats which were not practically relevant in UMTS, but may become relevant in SAE. The impersonation threat may be realised by stealing authentication vectors from one network, with possibly sloppy enforcement of security, and using them in another network. One should bear in mind that SAE/LTE is designed for use beyond 2015 and that the environment in which SAE/LTE will operate may be subject to drastic changes, including the business models and the assumptions on trust relations on which the UMTS security architecture was based. In particular, it is desirable for SAE that the dependency of the security in one network on the security in other networks shall be minimized.

If it is true that the security of one LTE network shall not depend on that of another LTE network, it is a fortiori true that it shall not depend on the security of a non-3GPP access network. Therefore the binding of the access network technology to the highest key available in an SAE access network is also advisable.

The binding of the identity of the ASME (MME in LTE) would ensure that the compromise of one ASME / MME under the control of an attacker does not affect other ASMEs / MMEs in the same access network. However, one may assume a uniform level of security for entities of the same type in one access network, and the consequences of a compromise of security would be felt only within one administrative domain, so the risk may be deemed lower. In addition, the ASME/MME identity may not be available to UE for key derivation as an operator may want to hide the MME identity towards the radio interface. It is therefore proposed not to include the ASME identity in the derivation of the top-level keys.

It is important to note that including the ASME identity in the top level key on key establishment does not imply that this top level key or derived keys cannot be transferred to other MMEs during mobility events.

7.4.7.3.5 Binding keys to traffic type in LTE

It is proposed that for NAS, UP and RRC traffic in LTE, specific keys are derived which may be used only with the specified traffic type. As the risk of compromise is different for the different traffic types it seems advantageous to limit the effect of a compromise to one traffic type. As separate keys are needed anyhow because the different traffic types terminate at different entities, the additional cost of binding the traffic type to the key seems low.

This binding was also proposed in S3-060648 and included in this TR

7.4.7.3.6 Binding keys to cryptographic algorithms in LTE

It is proposed that LTE keys may be used only with a particular cryptographic algorithm. The advantage of such a binding is that a compromised algorithm which allows retrieving the key would not affect traffic using stronger algorithms. This requirement is motivated by the experience with the very badly broken A5/2 algorithm in GSM. Similar attacks are believed not to be possible in UMTS because the cryptographic algorithms in UMTS are stronger and bidding down attacks are not possible due to signalling integrity protection. But, although no immediate risk is seen in LTE, it seems prudent to introduce this binding as it does not seem to cost much.

This key binding was also proposed in S3-060476 and included in this TR.

7.4.7.3.7 Binding keys to identities of eNBs in LTE

It is proposed to make RRC and UP keys dependent on the identities of the eNBs for which they are generated. This requirement does not preclude that these keys are transferred to and used by different network entities in handover.

The binding ensures that the compromise of one network entity would not affect other network entities of the same type in the same access network. But on the other hand, one may assume a uniform level of security in one access network, and the consequences of a compromise of security would be felt only within one administrative domain, so the risk may be deemed relatively low. It's proposed just the same to use this binding because (as already stated in Section 5.4) the moderate gain in security comes almost for free. This assumes that the relevant identities are easily available to entities deriving the keys.

This binding was also proposed in S3-060648 and S3-060692 and included in this TR

7.4.7.3.8 Binding keys to temporary identities of the UE

It is proposed to make LTE keys dependent on the temporary UE identities (i.e. C-RNTI for RRC). The binding ensures that the keys are renewed e.g. between multiple idle-to-active mode transitions under the same eNB. It's proposed to also consider whether the binding of S-TMSI to a further intermediate key derived from K_{ASME} could be beneficial to achieve key renewal at a higher level in the key hierarchy without a new AKA run. But this is not included here as it needs more discussion.

7.4.7.4 Storage of K_{ASME}[20]

All EPS keys are derived from key K_{ASME}; K_{ASME} is considered as master key. The EPS keys could be updated by reusing the existing master key K_{ASME} without new AKA procedure. K_{ASME} is a sensitive data and different attacks exist according to the type of storage of K_{ASME} in the UE.

Attacks description:

- K_{ASME} is stored in the ME

The input parameters used to derive EPS keys could be RAND, IMPI, Identities of eNB, MME, … All the derivation parameters (such as the identities of the network elements) are available on the ME and are not confidential.
An attacker accessing the ME can retrieve K_{ASME}, associated information such as the RAND and also IMSI or IMPI values. The knowledge of K_{ASME} and associated values allows the attacker to compute all EPS keys for any eNB, and MME entities.

Consequently, when K_{ASME} is stored on the ME, an attacker needs only one connection with the ME to allow a device not hosting the UICC to compute any set of EPS keys used by the ME during the availability of K_{ASME}. The attacker does not longer need a UICC to access the network during all the lifetime of K_{ASME} he maliciously got.

If K_{ASME} is stored in the ME there is no guaranty that the UICC is present during all the lifetime of K_{ASME} and thus the operator cannot have full assurance that the user equipment is not fraudulent.

- K_{ASME} is stored in the UICC

---

20 This section is from S3-080046.

An attacker accessing the ME can retrieve RAND and also IMSI or IMPI. He can also retrieve the current EPS keys used for communication. If the attacker wants to know another set of EPS keys associated to different elements (eNB, MME) then he needs to establish a new connection to the ME in order to make the ME send a command to the UICC asking for the derivation of new set of EPS keys. After the execution of the command, the ME gets a new set of EPS Keys which can be extracted by the attacker.

So, when the key K_{ASME} is stored on the UICC, an attacker who wants to discover a set of EPS keys from a device not hosting the UICC needs to establish one connection with the ME for each set of EPS keys. In this context, an attacker is able to use a set of keys only for a short period. Moreover, storing K_{ASME} in the UICC allows the operator to have full assurance that there is a UICC in the User Equipment when a new set of EPS keys is derived.

Editor's note: it's FFS if extra complexity will be caused by the solution of key derivation in UICC.

Conclusion: There isn't big enough security benefit to justify the added complexity of introducing this alternative solution.

K_{ASME} key lifetime:

- K_{ASME} is stored in the ME

K_{ASME} should be deleted when the ME is powered down or when the UICC is removed.

- K_{ASME} is stored in the UICC

There is no need to delete K_{ASME} when the ME is powered down or UICC is removed.

So, K_{ASME} key lifetime is longer when K_{ASME} is stored in the UICC, this leads to decrease the consumption of authentication vectors.

Alternative solution for EPS key hierarchy in case of K_{ASME} stored in the UICC

Editor's note: this alternative solution is not approved to be adopted in SAE/LTE and that further studies are needed.

The storage of K_{ASME} in the UICC implies the definition of an alternative solution to derive key hierarchy (K_{NASenc}, K_{NASint}, K_{eNB}) from K_{ASME}. This UICC-based key hierarchy requires the modification of the AKA authentication procedure in the USIM.

The storage of K_{ASME} in the UICC does not apply for UTRAN access network; this solution for UICC-based key hierarchy should not be used for UTRAN access network. The UICC shall be able to distinguish authentication requests for E-UTRAN access network requiring key hierarchy from authentication requests for UTRAN access network. Only EPS-capable USIM would be able to perform the alternative procedure to store K_{ASME} and derive EPS keys. The key hierarchy proposed in 7.4.7.2 should apply in case of non EPS-capable USIM.

A new security context of the AUTHENTICATE command, "EPS Security context", should be defined.

The AMF field would be used to distinguish authentication for E-UTRAN (EPS Security context) from authentication for UTRAN (3G security context). In case of AMF field indicating AKA for EPS (separation bit of AMF is set to 1) the ME would send AUTHENTICATE command with "EPS Security Context" and the required authentication data to perform EPS AKA. The Serving Network Identity is part of the input data of AUTHENTICATE command with "EPS security context".

The "EPS security context" of the AUTHENTICATE command should have two modes :

- "EPS authentication" mode to perform the authentication, compute K_{ASME} and send SRES to the ME. CK, IK are no longer sent to the ME. The USIM stores K_{ASME} and also associated KSI_{ASME}.
- "EPS key derivation" mode to derive K_{NASenc}, K_{NASint}, K_{eNB} on demand of the ME. The input data of this mode contains the parameters required to derive K_{NASenc}, K_{NASint}, K_{eNB} keys.

The term "EPS-capable USIM" is used to refer to new USIM application in the USIM implementing the AUTHENTICATE command with "EPS Security context".

The support of EPS-capable USIM in the UICC would be optional.

The creation of "EPS Security Context" for the AUTHENTICATE command impacts the UICC-ME interface.

Impacts on the ME

The ME should support the "EPS security context" for the AUTHENTICATE command.

When the EPS-capable ME receives authentication data from the network with Network Type equal to "E-UTRAN" then the ME should send to the UICC an AUTHENTICATE command with "EPS Security context - EPS authentication mode". Then the ME would receive RES from the UICC.

In order to retrieve K_{NASenc}, K_{NASint}, K_{eNB} the ME should send AUTHENTICATE command with retrieve "EPS Security context - EPS key derivation mode".

The support in the ME of AUTHENTICATE command with "EPS Security context" should be mandated. This would allow the home operator to issue when he wants new UICCs with EPS-aware USIM, independently of the type of EPS-aware ME.

Impacts on the network

There is no impact on the network. The AMF field in AUTN already provides indication on the type of authentication by means of the selection bit.

7.4.8 Use of AMF for SAE binding[21]

7.4.8.1 Background

In the SAE key hierarchy it is argued that binding authentication vectors to SAE use is crucial for enhanced security in SAE.

Briefly recap of the problem to be solved: authentication vectors used for SAE and pre-SAE systems shall be verifiably separated. In particular, authentication vectors delivered to pre-SAE network entities (e.g. SGSNs or RNCs in visited networks) could be stolen and then used to impersonate an SAE network. If this was possible it would defeat the purpose of key binding in SAE.

It's required a binding mechanism to meet the following two requirements:

1) the mechanism shall not require any changes to R99 USIMs;

2) it shall be possible to use the same USIM with SAE-capable MEs as well as with legacy MEs;

Requirement 1) is in accordance with SA3's decision to allow R99 USIMs for SAE access.

Requirement 2) means that a user can buy an LTE-capable ME at a certain point in time without having to change his USIM. A consequence of requirement 2) is that it is not possible for the HSS to generate special SAE authentication vectors for users, based on subscription information, as even SAE subscribers may at some point use their USIM in connection with legacy MEs that do not support a special SAE key derivation. Moreover, an ME cannot explicitly signal its capability to support SAE key derivation to a legacy SGSN. Therefore it must be assumed that the HSS has no information about the UE's capabilities to support SAE key derivation when the UE attaches to a legacy SGSN. As noted in the section on the SAE key hierarchy, it is ffs whether this is also true when the UE attaches to an MME or a Release 8-SGSN over UTRAN.

Here we present an effective solution for binding authentication vectors to SAE use. This solution is based on the AMF field of an authentication vector.

The proposed solution shows that it is possible to implement the binding of authentication vectors to SAE use without requiring any changes to R99 USIMs and while keeping AuC changes small.

The use of the mechanism considered here is independent of the decision whether UMTS 3G AKA or EAP-AKA will be used as it can be used with both variants of AKA.

---

21 This section is from S3-070096.

In this section, we show how a particular bit, which we call here the "Separation bit", in the AMF field of AKA could be used to indicate whether or not an authentication vector is usable for AKA in an SAE context. If the Separation bit is set to 1 the authentication vector is only usable for AKA in Release in an SAE context, if the bit is set to 0, the vector is usable in a non-SAE context only. For authentication vectors with the Separation bit set to 1, the secret keys CK and IK generated during AKA do never leave the HSS. The proposed procedure does not require any changes to current USIMs and keeps AuC changes small. Furthermore, we assume that the changes amount to a configuration of the AuC. Also it does not change the Release 99 specifications and higher versions of 3GTS 33.102.

The mechanism assumes that not all bits of the AMF are already in use for proprietary purposes. There is some evidence that, in fact, AMF is currently not used at all.

The 16 bit Authentication Management Field AMF (cf. TS 33.102, Section 6 and Annex F) is inserted into the authentication token AUTN in an authentication vector (AV) by the AuC during AKA in the clear (i.e. not blinded by the anonymity key AK). AMF is also included in the computation of the message authentication code MAC such that it cannot be changed during transfer to UE. There is currently no standardized interpretation for AMF. Examples for possible use cases are included in Annex F of TS 33.102.

7.4.8.2 SAE binding with AMF

HSS and ME follow the rules below for the new key separation:

Rules:

- the HSS must never issue an AV with the Separation bit set to 1 to a non-SAE network entity.
- The HSS performs further key derivation from CK, IK before sending an AV with Separation bit set to 1 to an SAE-MME (or any other SAE entity.).
- An ME attaching to LTE (or another SAE access network) must check during authentication that Separation bit is set to 1 and abort authentication if this is not the case.

Upon receipt of an authentication vector request from an MME, the HSS requests from the AuC one or more new authentication vectors AV* usable on SAE/LTE. AuC generates these authentication vectors AV* using the AMF with the Separation bit set to 1 and transfers them to HSS. For each received authentication vector AV*, the HSS derives a PLMN- and RAT-specific key K_{ASME} from the original CK, IK included in AV* and replaces these by K_{ASME} in AV*. How such a key could be transported from HSS to ASME is explained in S3-060632. HSS gets knowledge of the corresponding PLMN, RAT combination and the ASME (MME) identity as part of the SAE-specific authentication vector request received from MME as HSS needs this information anyway to be able to take authorization decisions of whether PLMN is allowed to serve a particular subscriber in combination with a particular RAT

HSS transfers the modified AV* to ASME (MME). ASME (MME) sends RAND, AUTN of the first AV* to UE. Upon receipt of a RAND, AUTN pair, the USIM checks whether MAC is correctly computed over AUTN. USIM is thus assured that AMF was not changed during transfer from AuC to USIM. This is all in accordance with Release 99 specifications. USIM does not interpret AMF. Instead ME interprets AMF and computes the key K_{ASME} from CK, IK if the corresponding Separation bit is set to 1. If network selection in SAE/LTE is performed as in UMTS then UE gets to know the RAT / PLMN combination it is currently attached to as part of the beacon information used on network selection (cf. TS 23.122) and can use this information upon key separation. It needs to be checked further how UE obtains the MME identity.

It is important to note that, due to the use of the Separation bit, the AuC cannot simply pre-compute authentication vectors for several sequence numbers any more. This is a result of the use of sequence numbers and the fact that the message authentication code MAC depends on AMF, and AMF depends on the context (SAE or not) which the AuC cannot predict. Nevertheless, it may still be useful to pre-compute authentication vectors for the same context as the previous one as the likelihood of a UE attaching to the same type of network may be reasonably high.

Editor's note: it's FFS if there is side effect caused by the proprietary use of the bit of AMF.

7.4.9 Key handling on active to idle and idle to active transitions in SAE[22]

7.4.9.1 General

In this section we propose a working assumption on how keys should be handled on active to idle and idle to active transitions within SAE/LTE. These state transitions are independent of mobility events such as handover or idle mode mobility.

As a general principle, on idle to active transitions, RRC protection keys and UP protection keys shall be generated as described in section 7.4.7 while keys for NAS protection as well as higher layer keys are assumed to be already available in the MME. These higher layer keys may have been established in the MME as a result of an AKA run, or as a result of a transfer from another MME during handover or idle mode mobility. On active to idle transitions, eNBs shall delete the keys they store after a predefined period such that state for idle mode UEs only has to be maintained in MME.

7.4.9.2 Idle to active transition

On idle to active transitions the MME shall generate and transfer the keys for RRC protection to eNB in the same way as during initial attachment (see the section on key hierarchy). In particular

- MME generates K_{eNB} and transfers it to eNB
- eNB subsequently derives , K_{RRCint} and K_{UPenc} from K_{eNB}
- eNB uses $K_{RRCenc}$for encryption K_{RRCint} for integrity protection of RRC trafficand K_{UPenc} for encryption of UP traffic.

In case UE is connected to the same eNB after idle to active transition as during any previous active phase since the last AKA run, the same K_{eNB} is transferred from MME to eNB as used during this previous active phase. In order to avoid the use of the same key stream with different instances of the same eNB, C-RNTI is included in the derivation of the keys K_{RRCenc} , K_{RRCint} and K_{UPenc}. [23]

EMM-IDLE to EMM-CONNECTED (From S3a070928)

> Editor's note: SA3 is aware of that the state names do not match the current naming in TS 23.401, and have to be updated

The following handling of keys is agreed by SA3.

A prerequisite for this K_eNB refresh procedure is that there exists a NAS security context which in particular contains replay protected uplink and downlink NAS SQN numbers. The K_eNB refresh procedure makes sure that the MME and the UE uses the same NAS SQN as freshness parameter in the K_eNB derivation.

The SQN of the NAS Service Request message sent from the UE to the MME is used as freshness parameter in the K_eNB derivation. The NAS Service Request is integrity protected. The current understanding is that there can be only one outstanding NAS Service Request message corresponding to the radio bearer establishment, so that both the UE and MME will know which NAS SQN to use (this needs to be confirmed by other groups). If this turns out to not be true, the MME needs to inform the UE about the NAS SQN of the NAS Service Request it used to derive the K_eNB. This information would be provided by the MME to the eNB on S1, and the eNB in turn informs the UE about it during the radio bearer establishment procedure.

7.4.9.3 Active to idle transition

On active to idle transitions we assume that eNB does no longer store state information about the corresponding UE. In particular eNB deletes the current keys from its memory.

In particular, on active to idle transitions:

---

[22] This section is from S3-070097.

[23] In order to avoid the use of the same key stream for NAS with different instances of the same MME (e.g. on subsequent attachment procedures) Start values similar to UMTS (TS 33.102) could be used. This is ffs.

- The eNB deletes K_{eNB}, $K_{RRC\,enc}$ and $K_{RRC\,int}$, and $K_{UP\,enc}$
- MME keeps K_{ASME} stored.

On active to idle transitions MME should be able to check whether a new authentication is required, e.g. because of prior inter-provider handover as described in the section on "Key handling on mobility events".

7.4.10 Key handling on mobility within an SAE/LTE network and between two different SAE/LTE networks[24]

Different alternatives for how keys could be handled upon mobility events within an SAE/LTE network and between two SAE/LTE networks. Key handling here includes key derivation and transfer. Mobility here refers to handover (mobility in active mode) as well as idle mode mobility. In our terminology, the user moves from a source network / entity to a target network / entity.

Former contributions to SA3 related to this topic proposed specialized solutions for some aspect of the key handling on mobility (as e.g. S3-060032, S3-050694, and S3-060236 focus on intra MME/UPE handover while e.g. S3-060704 focuses on key derivation and transfer on handover between LTE and UTRAN) , this sectionlook at the decisions SA3 has to take from a higher view point. The optimal choice will, however, depend on the actual handover procedures chosen by SA2. Currently, many alternative handover procedures are still under discussion (see TR 23.882, Section 7.15 for inter MME handover alternatives, see S2-063195 for rationale for Alternative 2, see TS 36.300, and TR 25.813 for intra MME handover).

When deciding key derivation functions and key management procedures the UE operations and UE – LTE RAN interface should be kept simple. It is preferred to maintain the same functionality regardless, which kind of handover type is in question, but this requirement needs to be traded off against other requirements.

7.4.11 K_eNB refresh at state transitions [25]

Editor's note: SA3 is aware of that the state names do not match the current naming in TS 23.401, and have to be updated.

The following handling of keys is agreed by SA3.

When the UE goes from EMM-DETACHED to EMM-CONNECTED, there are two cases to consider, either a complete NAS security context exists, or it does not.

If there is a NAS security context, the UE transmits a NAS Attach Request message. This message is integrity protected, and similarly to the EMM-IDLE to EMM-CONNECTED case, the NAS SQN of the Attach Request message is used to derive the K_eNB. Also here, it is the assumption (which needs to be verified with other groups) that there can only be one Attach Request message outstanding at any time. The RAN groups should be informed that this assumption has to be made to ensure that the UE and the MME use the same input for K_eNB refresh. Note that the same procedure for refresh of K_eNB can be used, regardless of if the UE is connecting to the same MME to which it was connected previously or to a different MME. Also note that in case UE connects to a different MME and this MME supports different NAS algorithms, the NAS keys have to be re-derived with the new algorithms as input. In addition, there is a need for the MME to send a NAS SMC to the UE to indicate the change of NAS algorithms and to take the re-derived NAS keys into use.

In the case that there is an AKA run (either because there is no NAS security context, or the network decides to run an AKA after the Attach Request but before the corresponding radio bearer establishment), the NAS (uplink and downlink) SQN:s are reset, and the start value of the uplink NAS SQN is used as input to the K_eNB derivation. Note that using the default value of the uplink NAS SQN in this case cannot lead to the same combination of K_ASME and NAS SQN being used twice. This is due to the fact that the first integrity protected NAS message UE sends to MME after AKA is the NAS SMC complete message. This message will include the default value of the NAS SQN that is used as input to the K_eNB derivation and the K_ASME is fresh. Following the AKA, a NAS SMC needs to be sent from the MME to the UE.

---

24 This section is from S3-070099.

25 This section is from S3a070928.

It is necessary that the uplink and downlink NAS SQN:s is only reset when a new AKA is run; in particular, the NAS SQN:s keeps running even if the NAS keys are re-derived from the same K_ASME (because of an MME relocation with NAS algorithm change).

If any of the NAS SQN:s are close to wrapping, a new AKA must be run to re-key the NAS keys.

It is necessary that the uplink NAS SQN in the used messages is integrity protected (this is however not an issue, since that is the de facto approach to achieve secure replay-protection).

7.4.12 Key handling on idle mode mobility[26]

7.4.12.1 Within one SAE/LTE network

Idle mode mobility within one SAE/LTE network leads to cell reselections in which a UE chooses new eNBs to camp on, and it leads to location updates in which MME changes are indicated to the HSS. A cell reselection does not lead to new keys being provided to the new eNB. However, a new MME selected upon idle mode mobility has to be provided with keys. A new MME can be provided with keys by one of the three alternatives described for inter MME handover. However, in case of idle mode mobility, new keys could also be provided to the MME by a new run of AKA. (On handover a new run of AKA would be too time consuming). In this case, a location update would always result in a new run of AKA during which MME obtains a new K_{ASME} from the HSS. However, a new run of AKA is not required in order to provide key separation between MMEs. For key separation between MMEs it would be sufficient to provide the new MME with a K_{ASME} which HSS derived from CK, IK with help of PLMN-ID, and RAT type as input. But this would not only require HSS to support some new form of fast re-authentication procedure but would also require the HSS to keep additional state about each UE, namely the CK, IK pair.

It's proposed that the specification shall allow a new run of AKA upon location update. But it is at the discretion of the operator to determine the frequency of AKA runs. If no AKA is run then the proposals are the same as for inter-MME handover.

IDLE mode mobility with MME relocation (From S3a070928)

> Editor's note: SA3 is aware of that the state names do not match the current naming in TS 23.401, and have to be updated

The following handling of keys is agreed by SA3

As can be seen from clause 5.3.3.1 of TS 23.401 (v1.3.0), unless the "active flag" is set in the TAU request message, the TAU procedure does not establish any radio bearers (nor any RRC state in the eNB). Because of this, there is no need to derive any K_eNB. When the UE transitions to EMM-CONNECTED, a K_eNB will be established, but that is via a different procedure.

If the "active flag" is set in the TAU request message, radio bearers will be established as part of the TAU procedure. In this case a K_eNB derivation is necessary, and the NAS SQN of the TAU request message sent from the UE to the MME is used as freshness parameter in the K_eNB derivation. The TAU request needs to be integrity protected.

In the case an AKA is run as part of the TAU procedure, the uplink and downlink NAS SQN:s are reset, and the starting value of the uplink NAS SQN is used as input the K_eNB derivations (at some later point). For the same reason as mentioned in Section 2.1, this will not cause the same K_eNB to be derived twice, since the K_ASME is different. After the AKA is run, a NAS SMC must be sent from the MME to the UE.

In the case an AKA is not run as part of the TAU procedure and source and target MME use different NAS algorithms, the target MME re-derives the NAS keys from K_ASME with the new algorithms as input and provides the new algorithm identifiers within an SMC, which may be combined with the TAU response.

> Editor's note: There may be an optimization possible, in that the NAS SMC is somehow sent together with the TAU response. This is up to CT1 and RAN3 groups to decide subject to SA3 agreeing that there are no unacceptable security issues. It is noted though, that the UE may need to send a NAS SMC confirm message back to the MME before confidentiality is started by the MME. The reason for this is that, depending on the contents of the TAU response, the MME may need different levels of assurance of that the TAU accept message gets ciphered by a correct algorithm.

---

26 This section is from S3-070099.

7.4.12.2 Between different SAE/LTE networks

Idle mode mobility between different SAE/LTE networks results in a MME change. It's advised requiring a new AKA run in this case in order to provide MME with new keys that depend on the identity of the new PLMN, and the RAT type,. In addition a key caching mechanism (ffs) could be used to avoid new AKA runs on frequent network changes.

7.4.12.3 Proposed procedure[27]

Upon idle mode mobility the old MME shall include the current values of the counters for NAS integrity and NAS encryption, as well as the old NAS keys and K_{ASME} in the MME context response message during tracking area updates. The procedure is illustrated in Figure 15:

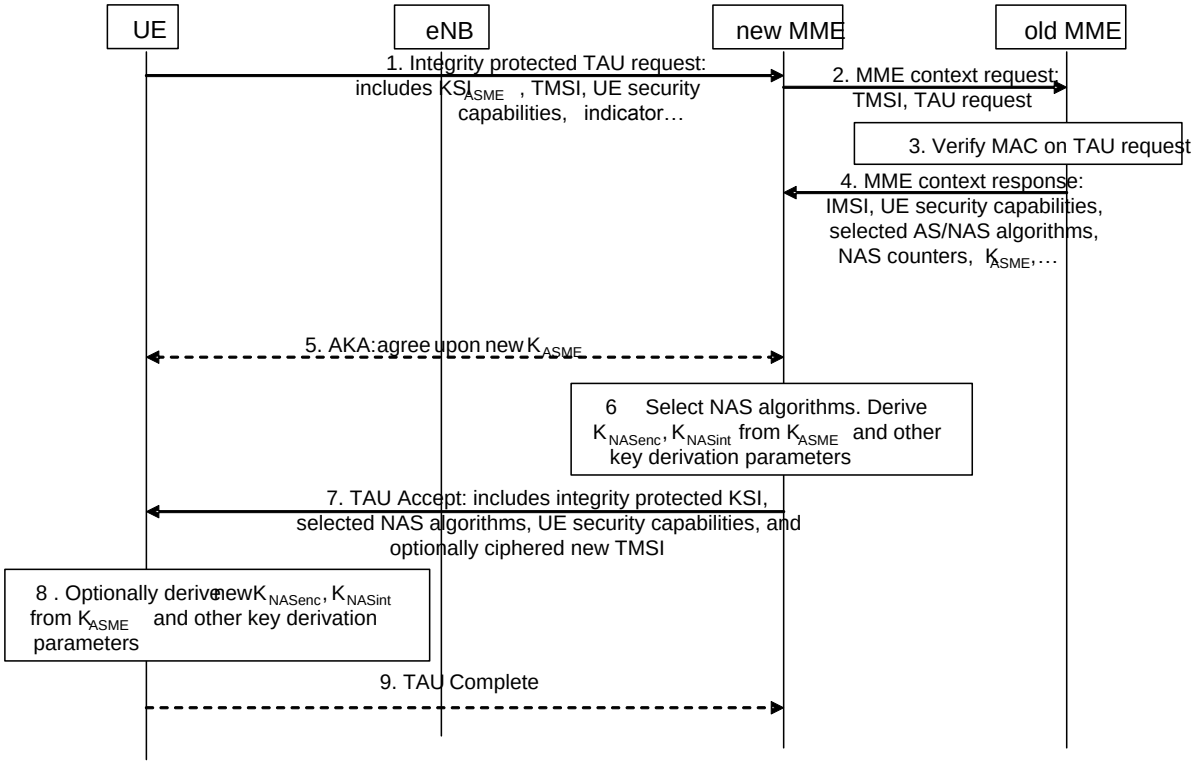


Figure 15 Key handling on idle mode mobility in E-UTRAN

1. UE sends a Tracking Area Update (TAU) request including UE's security capabilities to the new MME. UE includes KSI_{ASME} and integrity protects the TAU request.
2. New MME sends an MME context request including the TMSI and the TAU request to the old MME.
3. The old MME verifies the TAU request.
4. The old MME sends back an MME context response to the new MME including the TMSI, K_{ASME} as well as the current counter values for NAS to the new MME, the identifiers of the currently used NAS algorithms, and UE's security capabilities.
5. Optionally, the new MME initiates a new AKA authentication to get a fresh K_{ASME}.
6. The new MME select the NAS algorithms to use (according to its own, and UE's capabilities), and derives NAS keys (K_{NASenc}, K_{NASint}) from K_{ASME} using the identifiers of the NAS algorithms and other key derivation parameters as input parameters for the KDF.
7. The new MME includes the selected NAS algorithm identifiers and UE's security capabilities (including EPS/eUTRAN and UTRAN/GERAN if supported by UE), and optionally a ciphered (with the new NAS

[27] This section is from S3a071039.

ciphering key) new TMSI in the TAU accept message and integrity protects the message with the new NAS integrity key.

8. The UE optionally derives new NAS keys (K_{NASenc}, K_{NASint}) from K_{ASME}.. UE checks NAS-MAC and that the received UE security capabilities match with the sent ones.

9. Optionally UE sends integrity protected TAU Complete (see TS 23.401 v2.0.0 section 5.3.3.1).

Editor's note: it needs further study in the case that UEs need to change algorithms due to a handover (i.e. NAS Key handling when changing MME).

Editor's note: Deriving new NAS keys based on algorithms identifier as the only parameter is ffs (see S3-070533).

Editor's note: It is ffs if a separate NAS level SMC is used to change NAS algorithms on inter-MME handover (see S3-070533).

When UE is in idle mode, there is no RRC and UP security context, neither in the UE nor in the eNB. Thus, there is no need to derive RRC keys and UP key. (From S3a070917)

NOTE: there may be a case with active flag in the TAU, where the keys would be needed.

7.4.12.4 Key handling on idle mode mobility from UTRAN to E-UTRAN[28]

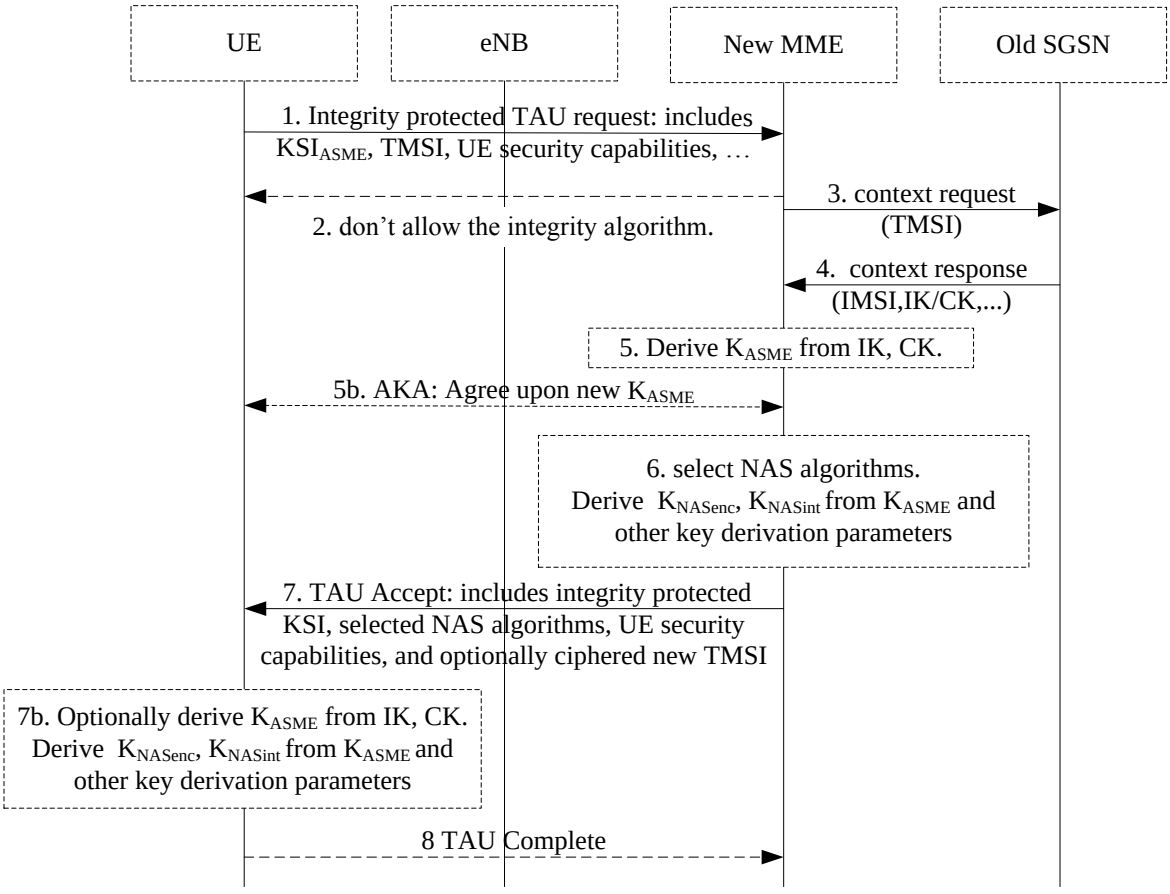


Figure 16 Key handling on idle mode mobility from UTRAN to E-UTRAN

1. UE sends a Tracking Area Update (TAU) request including UE's security capabilities and an integrity protection algorithm identifier that is used to protect the TAU request to the new MME. UE includes KSI_{ASME}

---

28 This section is from S31071040.

and integrity protects the TAU request if any keys available (e.g. cached EPS keys or keys derived from UTRAN/GERAN CK/IK).

2. Optionally, the new MME will ask the UE to resent the TAU request with another integrity protection algorithm if the new MME doesn't allow the integrity protection algorithm used in step 1.

3. New MME sends a context request including the TMSI to the old SGSN.

   Editor's note: It should be specified how the procedure changes the overall SMC, in that the TAU message can be combined with the SMC.

4. The old SGSN sends back a context response to the new MME including at least the IMSI, IK, CK to the new MME.

5. The new MME derives new K_{ASME} from IK, CK and uses the algorithm indicated in the TAU Request to derive K_{NASInt} and verify the integrity protection of the TAU Request. If the TAU Request verification fails,, the new MME initiates a new AKA authentication to get a fresh K_{ASME}.

6. The new MME selects the NAS algorithms to use (according to its own, and UE's capabilities), and derives ~~new K_{ASME} from IK~~, CK or uses the cached K_{ASME} if available and if UE used KSI_{ASME} corresponding to the cached keys. MME then derives new NAS keys (K_{NASenc}, K_{NASint}) from K_{ASME} using the identifiers of the selected NAS algorithms and other key derivation parameters as input parameters for the KDF.

7. The new MME includes the selected NAS algorithm identifiers and UE's security capabilities and KSI_{ASME} and optionally a ciphered (with the new NAS ciphering key) new TMSI in the TAU Accept message and integrity protects the message with the new NAS integrity key.

   The UE derives K_{ASME} from IK,CK and then derives new NAS keys (K_{NASenc}, K_{NASint}) from K_{ASME} if not already available. UE checks NAS-MAC and that the received UE security capabilities match with the sent ones.

8 Optionally UE sends integrity protected TAU Complete (see TS 23.401 v2.0.0 section 5.3.3.1).

7.4.12.5 Integrity protection of Attach and TAU message[29]

It is possible for an MME to receive both TAU and Attach Request message without integrity protection, e.g. Initial Attach and a TAU after being in GERAN/UTRAN. The likely MME's response to both messages is to successfully establish the security to authenticate the UE and then finish the procedures, e.g. in the case of Attach remove the current bearers. Currently all this is done without actually authenticating the actual message that was sent by the UE. This leads onto to some possible attacks and it is proposed that SA3 agree to provide integrity protection for the relevant portions of these messages.

Attack using Active Flag in TAU message

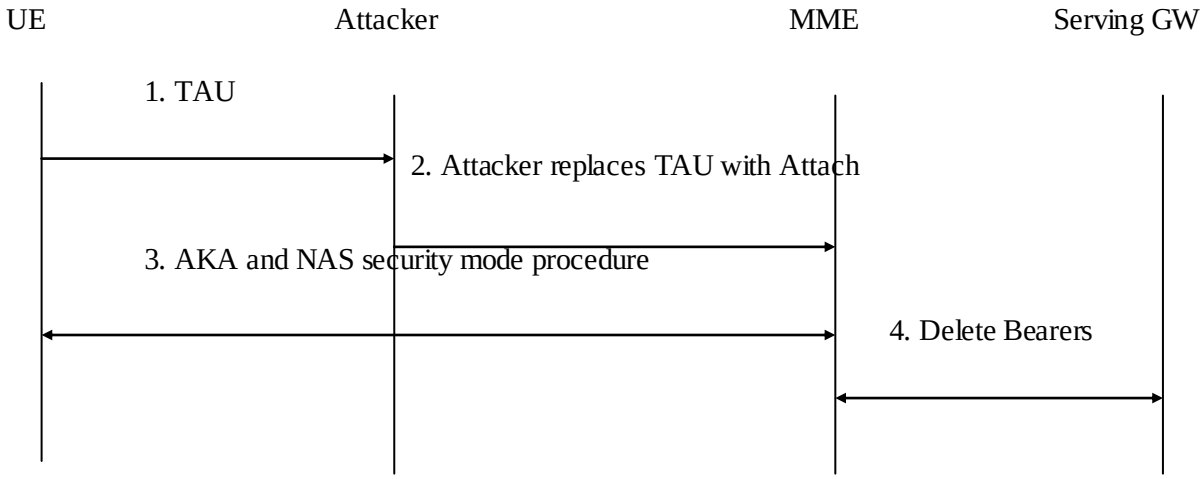
Suppose a UE send a TAU Request to perform Idle Mobility from UTRAN/GERAN to E-UTRAN. If there is no ISR, then it is not possible to integrity protect the content of this message. An attacker may change the content of this message to set the Active Flag to 1. The result of this attack is for the MME to effectively receive an unprotected Service Request message. In earlier meeting SA3 agreed that the Service Request message required at least 16 bits of MAC. To align with that decision, it is necessary to provide some integrity protection of the fact the UE sent a TAU message with the Active Flag set.

Attack by replacing a TAU message with an Attach Message

In this case an attacker substitutes an unprotected Attach for a TAU message. In response to this the network will probably perform an AKA followed by a NAS level security mode procedure. The UE will respond to these messages as it is a legitimate to expect this combination of message after a TAU. As part of the Attach procedure, the MME will delete old bearers (see step 6 of fig 5.3.2.1.1 in TS 23.401v800). The result of this is that despite the UE and MME successfully running AKA and NAS security mode command procedure, an attacker has forced the MME to delete the UE bearers. In the case of Service Request, SA3 have decided that a MAC of length greater than 16 bits is needed if the bearers are to be removed. The attack is described in the following figure (flow simplified to show only relevant steps).

---

29 This section is from S3-080178.



Another and possibly simpler way that an Attacker could launch this attack is to page the UE and replace the resulting Service Request with an unprotected Attach Request. The rest of the message flow proceeds as above. The advantage of launching the attack this way is that the attacker is not waiting for a UE to send a message it needs to replace.

Proposed solution

There are several ways that the data could be protected:

1. The Attach Request and TAU messages could be repeated along with one of the messages in the NAS level security mode procedure
2. The relevant information could be added to the NAS security mode complete message or the security mode complete message. Currently the only information that need integrity protection are what type of message was sent and if it was a TAU whether the Active flag was set or not.
3. The MAC on the security mode command is calculated assuming that the relevant bits of information have been included in the message.

Proposal 3 has the issue that different release UE and MMEs may have a different idea what needs to be included and hence there is always a risk of backwards compatibility. Proposal 1 works but it does require transferring more information than is necessary over the air. Proposal 2 has the advantage over proposal 1 in that it minimises the over the air information. Some possible text for inclusion in the TS to capture this is the following:

It is possible for the UE to send Attach Request and TAU message before integrity protection has been activated. In order to provide integrity protection for these messages the UE performs the following actions. If the UE receives a NAS level security mode command in response to an Attach Request before it receives an Attach Complete, then the UE shall include an indication in the NAS security mode complete message that its previous message was an Attach Request. Similarly if the UE receives a NAS level security mode command in response to a TAU message before it receives a TAU Accept message, then it shall include an indication in the NAS security mode complete message that its previous message was a TAU and repeat the Active flag from that message.

7.4.13 Key handling on active mode mobility[30]

7.4.13.1 Overview on alternatives for key handling on handover

Assume a target entity (eNB or MME) is to be provided with keys (for RRC, UP or NAS protection) during handover. Then we suggest further discussing the following general alternatives to provide the target entity with the corresponding keys.

---

30 This section is from S3-070099.

Alternative 1: Derivation of new target key for the target entity by the holder of the key one level up in the key hierarchy (parent key holder) from the key material (parent key) it holds. There are two subcases:

a) the parent key holder in the source network derives the key, which is then transferred to the target network;

b) the parent key holder in the source network transfers the parent key to the target network where key derivation takes place.

Alternative 2: Derivation of new target key for target entity by source entity from key material held by source entity (source key)

Alternative 3: Transfer source key used by source entity to target entity (possibly via another entity) and reuse it unchanged

It's assumed that key derivation is performed using a one-way function.

The three alternatives are illustrated in Figure 17.

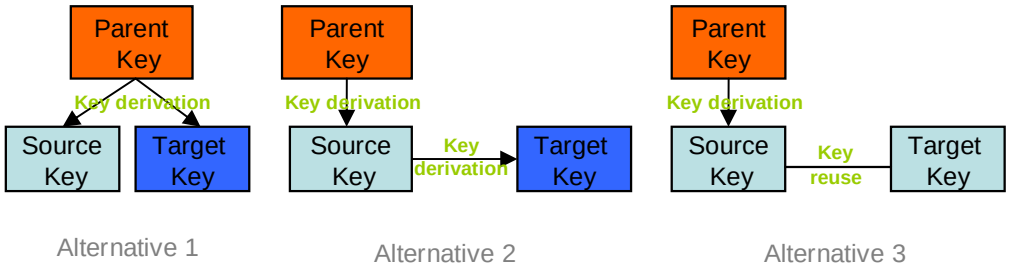


Figure 17 Alternatives 1 , 2, and 3 for key handling on handover between a source and a target entity

Alternative 1 guarantees that separate keys are used for the protection of traffic between UE and the source entity and between UE and the target entity. Therefore Alternative 1 is the preferred option bearing in mind only security. However, Alternative 1 requires the parent key holder (MME respectively HSS) to either be involved in each handover procedure or requires the parent key holder to predict the potential target entities and distribute encrypted keys for the potential target entities to the source entities before handover (for providing a target eNB with RRC keys this option was suggested in S3-060032). Involving the parent key holder does not currently seem feasible for all alternative handover procedures discussed by SA2 (see e.g intra MME handover described in TR 25.813 or TS 36.300). Predicting the target entities and distribute encrypted keys for them. This is out of the scope of this section.

Alternative 2 (which was also proposed in S3-060236 for RRC key handling on handover) provides backward security for the source keys: a target key compromised while used by the target entity cannot be used for impersonation of any source entity or for decrypting previously recorded traffic exchanged between the source entity and UE. Alternative 2 does not require the parent key holder to be involved in the handover procedure.

Finally, Alternative 3 does not protect the source entities from compromised target entities. RRC, UP or NAS keys compromised while used by the target entity can be used to impersonate any other entity of the same type or decrypt previously recorded encrypted traffic exchanged between source entity and UE. However, Alternative 3 adds the least overhead to handover procedures and seems to be acceptable as long as eNBs and MME can be assumed to be equally well protected (as e.g. in case of intra-PLMN handover).

As a general principle we suggest using Alternative 1 whenever the handover procedures selected by SA2 allow for an easy implementation of this alternative. Otherwise we suggest using Alternative 3 due to the additional complexity and the limited security gain of Alternative 2. In the following two sections we discuss the above alternatives in more detail for the different handover types and show how the decision on which solution to select depends on the way handover procedures will be implemented in SAE.

SA3 agreed the following requirements: (From S3-070475)

(1) If the sequence numbers for the UP or RRC ciphering/integrity protection are about to wrap around, it shall be possible to change the respective keys.

(2) If a UE has been in LTE_ACTIVE for a long period of time, it shall be possible to update the keys for UP and RRC ciphering/integrity protection, even though the sequence numbers are not close to wrapping.

(3) The operator shall be able to restrict the lifetime of K_{ASME} (independently of the key usage in LTE).

(4) If the UE has performed an inter-RAT handover from UTRAN/GERAN to LTE, it shall be possible to update all keys within seconds.

7.4.13.2 Key handling on handover within one SAE/LTE network[31][32]

7.4.13.2.1 The necessity of forward security for KeNB derivation

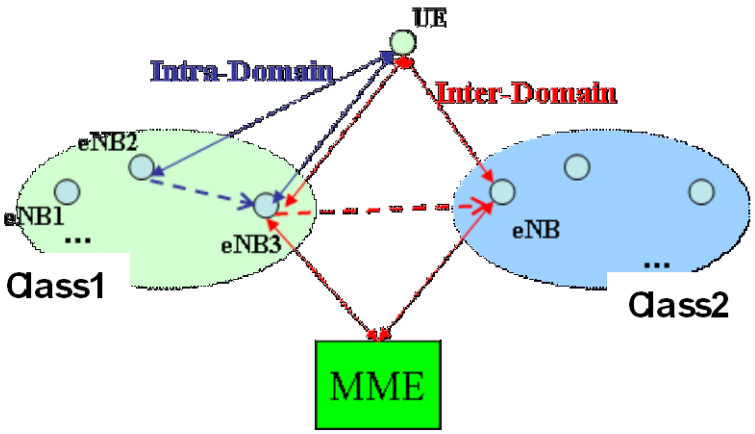


Figure 18 Security class

Figure 18 shows the example of different eNB classes where, for example, Class1 is eNBs considered secure and Class2 is eNBs considered not secure.

Different eNBs may have different security levels because of their deployment environments, physical protection mechanisms, and so on. It is easy to understand that an eNB deployed in hot-spot without any physical protection mechanisms has lower security level than an eNB deployed in an operator's building. So eNBs can be classified to several security classes according to the location or security configuration of eNB. The eNB with the same security level can be assigned to the same security class.

It is FFS how an operator can decide the handover type depending on deployment, network topology etc. This may not be up to SA3 to decide.

eNodeB cannot always be considered a secure entity. Under the new assumption, it is much more advantageous for the adversary to attack eNodeB. By gaining access to a single eNodeB, an adversary can get access to significantly more security resources than under the old assumption. It is imperative to protect K_{eNB} during a handover procedure by limiting the damage caused by a compromised eNodeB. Thus, forward and backward K_{eNB} secrecy are vital.

7.4.13.2.2 AS key Handling Properties[33]

7.4.13.2.2.1 Definitions

Forward security in the context of this Technical Report (and TS 33.401) is defined as the property that an eNB is unable to calculate AS keys that will be used between a UE and another eNB to which the UE is connected after a series of subsequent handovers. More specifically: n-hop forward security is the property that eNB X is unable to calculate AS keys that will be used between a UE and another eNB to which the UE is connected after n subsequent handovers of any type starting from eNB X.

---

[31] This section is from S3-080058.

[32] This section is updated by S3-080498.

[33] This section is from S3-090103.

A fresh AS key for an eNB is a key (generated in the MME) which can not be derived from a former AS key which was distributed to the eNB.

7.4.13.2.2.2 Rationales and decisions for forward security

Following decisions have been taken to achieve that n in "n-hop security" is as low as possible in various scenario's but at the same time fitting into the signaling flow design.

1) An MME will never send the same AS keys to more than one eNB.

2) An source eNB will never forward the same AS keys to more than one target eNB.

3) The MME will always generate fresh AS keys (cfr clause TS 33.401 Subclause 7.2.8) such that forward security can be applied i.e.

(a) On S1 handovers 1-hop security is achieved between eNBs.

(b) On X2 handovers 2-hop security is achieved between eNBs.

NOTE: In case of specific error conditions the 2-hop security property will not be achieved e.g. in case the PATH SWITCH REQUEST ACKNOWLEDGE was not received. However an attacker cannot exploit it because S1-interface is protected and target eNB is not compromised by assumption.

4) On an intra-eNBs handover (prior to X2 or S1 handover) a fresh key (when available) will be taken into use.

Each AS key has a key index in order to ensure that the UE can determine which key (and key derivations) has been used by the eNB or MME to provide forward security. For this purpose

a) The MME when generating fresh AS keys will never decrease the key index.

b) The source node will always include a key index when sending a key to a target node.

c) The eNB will send the key index to the UE in the various inter-LTE handover scenario's.

NOTE: The key index for forward security is differently from the key index KSI which is used to identity the AKA-run.

7.4.13.2.3 Key refresh on Intra eNB handover

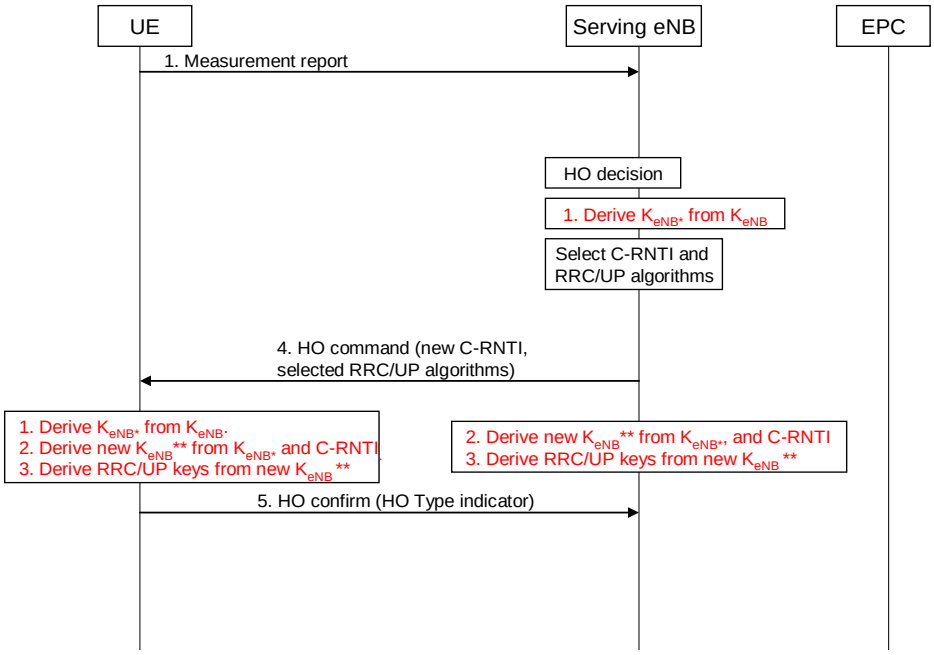


Figure 19 Key re-fresh in intra eNB handover

For intra-eNB handovers the C-RNTI binding is used for key refresh purposes. This improves the security whenever the MME is not involved in the key derivation procedure (e.g. intra-MME handover and path switching without Next-Hop-KeNB and inter-MME handover without key derivations and thus no indication for the UE of the MME involvement in the key derivation).

7.4.13.2.4 Key refresh on Inter eNB, intra MME handover

In this handover case the three alternatives for providing the target eNB with the keys for RRC protection are:

Alternative 1: MME derives a new K_{eNB} or a new K_{RRCenc}, K_{RRCint} pair from K_{ASME} and transfers it to the target eNB. If MME transfers K_{eNB} then the target eNB subsequently derives K_{RRCenc}, K_{RRCint} and K_{UPenc} from K_{eNB}

Alternative 2: eNB derives a temporary key K'_{eNB} from K_{eNB} , or K_{RRCenc} , or K_{RRCint} , and transfers it to the target eNB (directly or via MME). The target eNB subsequently derives K_{RRCenc}, $K_{RRC\,int}$ and K_{UPenc} from K'_{eNB} for RRC protection

Alternative 3: eNB (or MME) transfers K_{eNB} to the target eNB, target eNB derives K_{RRCenc}, K_{RRCint} and $K_{U\,enc}$ from K_{eNB} dependent on the encryption and integrity protection algorithms it is going to use. For this alternative it is crucial that the intermediate key K_{eNB} is used such that the target eNB can derive separate K_{RRCenc}, K_{RRCint} if it uses encryption and integrity protection algorithms different from the ones used by the source eNB.

RAN (see TS 36.300, TR 25.813) currently assumes that MME is not involved in intra MME handover procedures. Therefore Alternative 1 does not seem to be easily applicable during this type of handover. In order to circumvent this difficulty, it was suggested in S3-060032 that MME should provide an eNB with keys not only for itself but also for potential target eNBs. These keys would then be encrypted with the help of a keys shared between MME and the target eNBs.

According to 25.813, v 7.10, Section 9.1.5, on intra MME handover the source eNB sends a handover request to the target eNB. The target eNB replies with a handover response. The handover response includes information required by UE (e.g. the C-RNTI). The source eNB includes this information in the handover command it sends to UE.

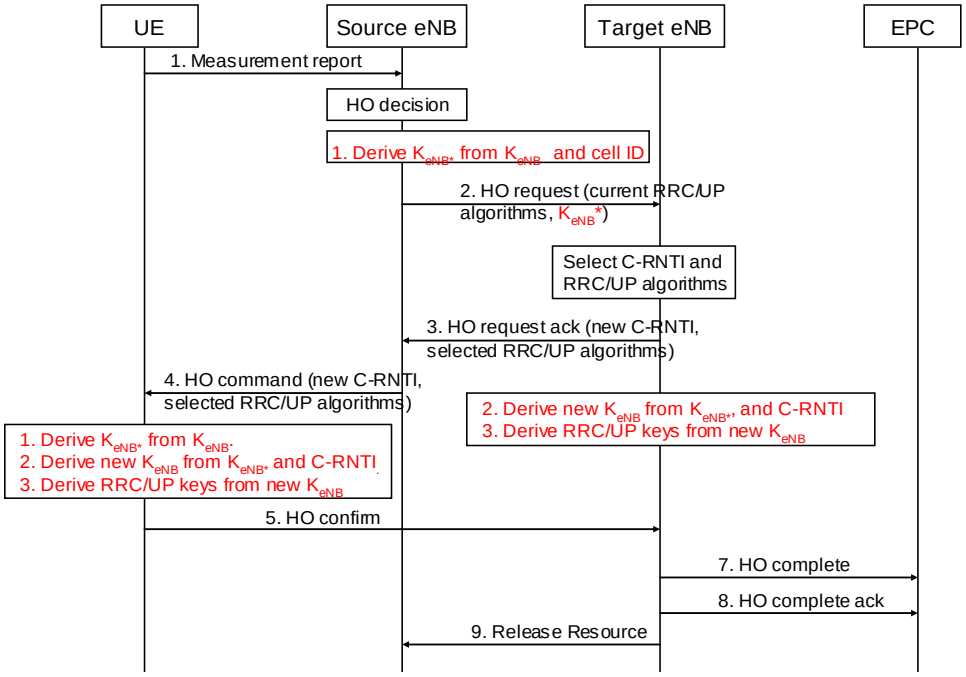


Figure 20: Key re-fresh and algorithms selection on intra MME handover

Figure 20 shows how K_{RRCenc}, K_{RRCint}, K_{UPenc} are refreshed on intra-MME handover.

1. UE measurement report
2. Source eNB calculates a one way hash over the current K_{eNB} and the physical cell ID of the target cell to get K_{eNB*} and transfers it to the target eNB in the handover request message including current RRC/UP algorithms
3. Target eNB sends handover response message to the source eNB, which includes the new C-RNTI, selected RRC/UP algorithms, and some other parameters (see 25.813, section 9.1.5). Target eNB derives a new K_{eNB} from C-RNTI and K_{eNB*} by $K_{eNB}\_new = KDF(K_{eNB*} \| \text{C-RNTI})$ and further derives K_{RRCenc}, K_{RRCint}, K_{UPenc} from the K_{eNB}_new.
4. Source eNB sends integrity protected and ciphered handover command message to the UE including C-RNTI and selected RRC/UP algorithms. In case the algorithms do not change they can be omitted.
5. UE derives the K_{eNB*}, new K_{eNB}, K_{RRCenc}, K_{RRCint}, and K_{UPenc} and sends handover confirm message to the target eNB integrity protected and ciphered with the new RRC keys.

Editor's note: Recovery from failed handover needs further study

Editor's note: it's FFS how to re-use the original keys before the handover attempt.

Editor's note: The possibility of the target eNode B's key being supplied by the MME is still open and ffs.

The proposed mechanism is described in Figure 21 and includes a Next-Hop-KeNB parameter from MME to the target eNB within the path switch acknowledgement message. Feeding both the serving eNB -related K_{eNB} and the K_{ASME} to the Next-Hop-KeNB derivation function in the MME results as cryptographically separate parameter for the target eNB compared to the parameter in the source eNB.

NOTE: Because the path switch message is transmitted after the radio link handover, it can only be used to provide keys for the next handover procedure and target eNB. Thus, perfect forward security happens only after 2 hops because the source eNB knows the target eNB keys (the fresh key derivation parameter, Next-Hop-KeNB, for target eNB is provided by the source eNB). In other words, the forward security step comes after two hops, as the source eNB does not have a way to know the keys that the target eNB uses to prepare handover to its own target eNBs (the fresh key derivation parameter, Next-Hop-KeNB, comes from the MME to the target eNB in the path switch acknowledgement message).

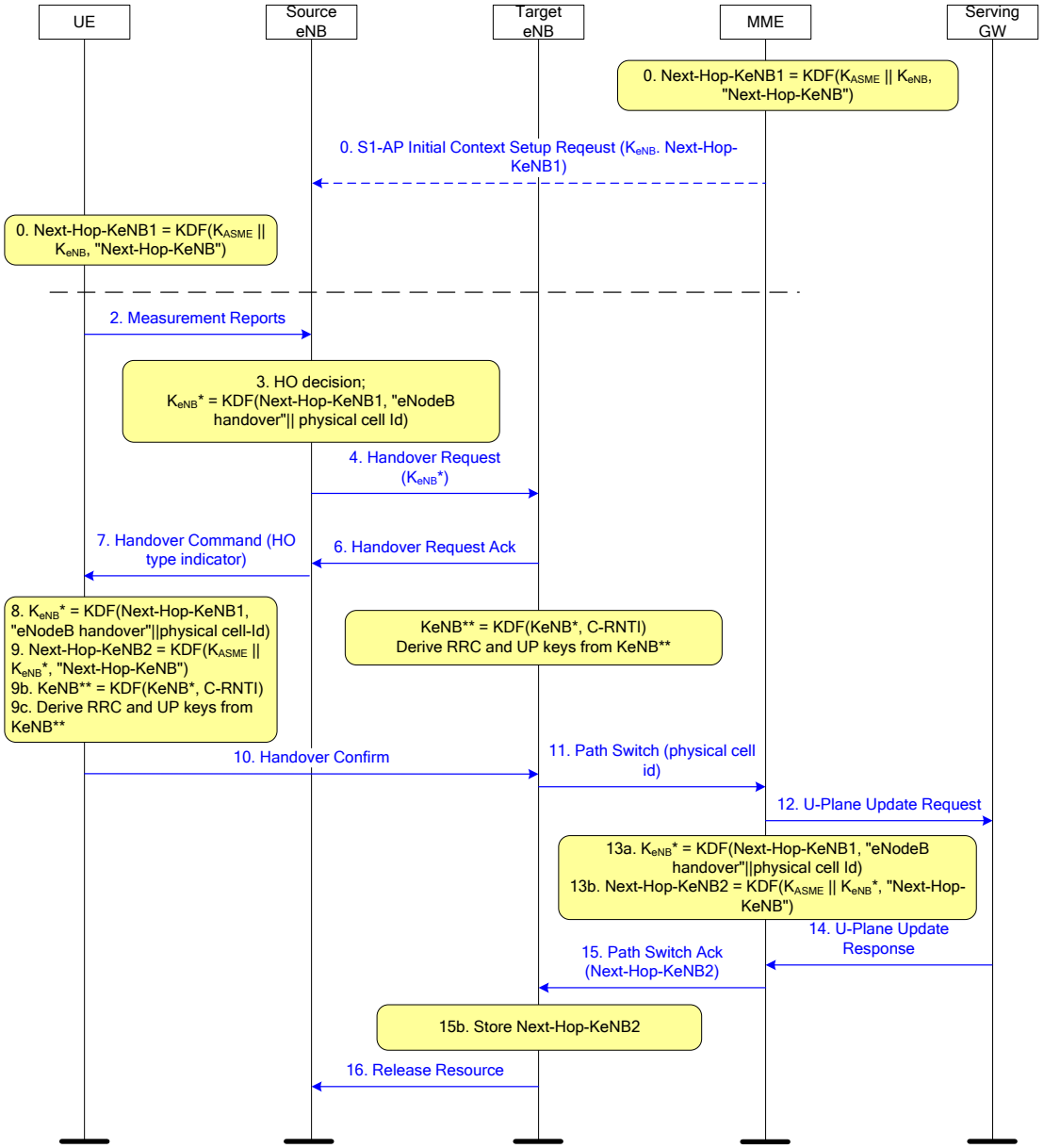


Figure 21 Perfect forward security for K_{eNB} after 2 hops (works also for inter-MME handovers)

The "K_{eNB}" is the key used between UE and source eNB, "Next-Hop-KeNB" is an intermediate parameter only used in K_{eNB}* derivations. The K_{eNB}* is the key used between UE and target eNB to derive KeNB** from target eNB C-RNTI and KeNB*. KeNB** is used to derive RRC and UP keys.

The target Cell Id is not be available for the UE in the HO Command message, but a physical cell Id is used instead. If physical cell Id is used then also MME needs to know the physical cell Id and this needs to be included into the path switch message.

Another alternative is to include target Cell Id into the HO Command messages for all inter-eNB handovers and do Cell Id binding only for inter-eNB handovers as from security perspective it is meaningful only for inter-eNB handovers. This would then work also as a HO type indicator for inter-eNB handovers, even tough it is not as efficient as a one bit indicator.

Note: need for synchronization procedure is FFS.

7.4.13.2.5 Key refresh on Inter MME handover

SA2 currently discusses whether or not MME relocations within one SAE/LTE network are necessary for certain handover types (see TR 23.882 and S2-063195). If MME relocations are implemented, keys have to be provided to the

target MME and to the target eNB. There are following alternatives for key handling on handover with MME relocation:

Alternative 1: HSS derives new K_{ASME} from CK, IK (with target MME-ID as well as the target PLMN-ID and the target RAT type as input) and transfers it to the target MME. The target MME derives K_{eNB} from K_{ASME} and transfers it to eNB. In addition, the target MME derives K_{NASenc} and K_{NASint} from K_{ASME}. In case the target MME transfers K_{eNB} eNB derives K_{RRCenc}, and K_{RRCint} and K_{UPenc} from K_{eNB} **(requires HSS to be involved in key derivation and transfer upon inter MME handover or requires HSS to predict potential MMEs to which UE may relocate and send several encrypted keys.)**

Alternative 2: Source MME derives a temporary key K'_{ASME} from K_{ASME} using the target MME's identity and the target PLMN-ID[34] as input. Target MME derives K_{NASenc} and K_{NASint} from K'_{ASME}

a) The target MME subsequently derives the key K_{eNB} from K'_{ASME} and transfers it to the eNB. , The eNB then derives K_{RRCenc}, and K_{RRCint} and K_{UPenc} from K_{eNB} **(requires MME to be involved in key transfer)**

b) K'_{eNB} is derived by the source eNB (with the target eNB-ID and the target PLMN-ID as input) and keys are transferred to the target eNB as in Alternative 2 described above **(allows for direct context transfers between eNBs)**

c) K_{eNB} are reused by target eNB as in Alternative 3 described for intra-MME handover. **(allows for direct context transfers between eNBs)**

Alternative 3: The source MME transfers K_{ASME} to the target MME. In addition, the target MME derives K_{NASenc} and K_{NASint} from K_{ASME}.

a) The target MME subsequently derives the keys K_{eNB} from the same K_{ASME} that was already used by source MME and transfers it to eNB, then eNB derives K_{RRCenc}, and K_{RRCint} and K_{UPenc} from K_{eNB} **(requires MME to be involved in key transfer)**

b) K_{eNB} is transferred from source eNB to target eNB as in Alternative 3 described for intra-MME handover **(allows for direct context transfers between eNBs)**

HSS involvement during handover procedures with MME relocation seems too time-consuming. In addition, HSS involvement would require HSS to keep additional state about each UE, namely the CK, IK pair from which K_{ASME} can be derived. Or else, the HSS would have to predict potential MMEs to which UE may relocate and send several keys K_{ASME} encrypted with keys shared between HSS and MME. But, apart from the complexity, this solution would require that core network security is realized in an end-to-end fashion between HSS and MME, which may not be assumed. Therefore, Alternative 1 in connection with HSS involvement upon handover seems infeasible.

In case Alternatives 2 or 3 are chosen by SA3 we propose to use Option a) if the handover procedures adopted by SA2 allow for it.

According to 23.882, v 1.18, Section 7.15 inter MME handover does either not occur at all (due to S1 flexible nature) or is executed with involvement of a target MME. We assume here that in the latter case, the handover command and handover confirm messages are exchanged between UE and the source eNB in the same way as on intra-MME handover such that inter and intra-MME handover are indistinguishable for the UE. It is ffs if this assumption holds.

On inter-MME handover as on intra-MME handover, the fresh K_{eNB*} is transferred to the target eNB. A new K_{eNB} is derived from the K_{eNB*} and C-RNTI, and K_{RRCenc}, K_{RRCint}, K_{UPenc} are refreshed with the help of this new K_{eNB}. The proposed procedure is detailed in Figure 22.

---

34 Note that according to TR 23.882, Section 7.20.2, MME-ID and eNB-ID are unique within a PLMN. Consequently on PLMN changes the PLMN-ID should be used as an additional input for key derivation. In order to support the same procedures in case of Inter-MME handover between PLMNs as within a PLMN, we suggest to use the PLMN-ID in any of the two handover cases.

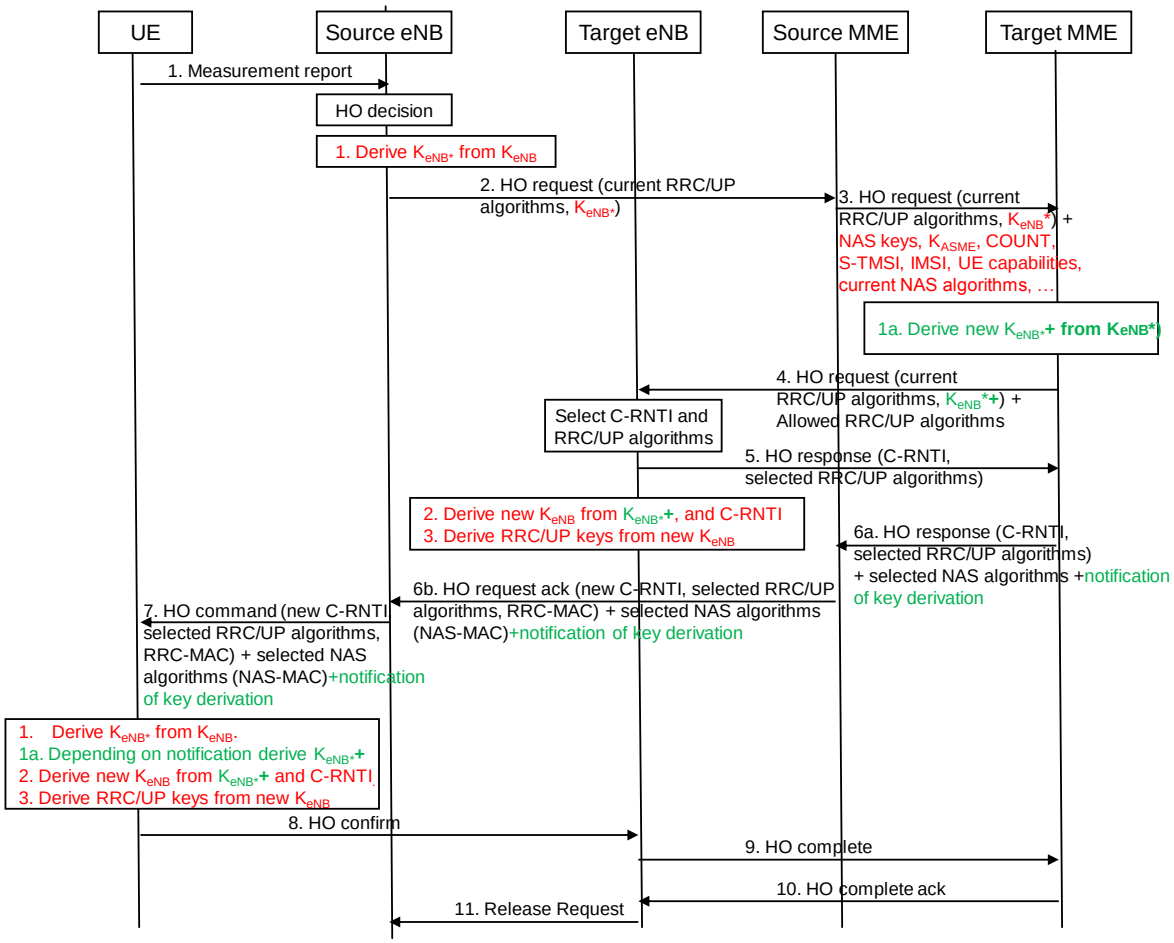


Figure 22 Forward secure key refresh on inter MME handover

1. UE measurement report

2. Source eNB calculates a one way hash over the current K_{eNB} and the physical cell ID of the target cell to get K_{eNB*} and transfers it to the source MME in the handover request message

3. Source MME transfers the K_{eNB*} and other related MME security context information, like NAS keys, COUNT values for NAS protection, S-TMSI, IMSI, and K_{ASME} to the target MME in the handover request message.

4. Target MME derives K_{eNB*+} from K_{eNB*} and K_{ASME}: K_{eNB*+} = KDF(KDF(K_{ASME} ‖"Handover String") || K_{eNB*})). KDF(K_{ASME} ‖"Handover String") is a key derived from K_{ASME}. "Handover String" is a constant.Target MME includes the K_{eNB*+} in the handover request sent to target eNB with allowed RRC/UP algorithms.

5. Target eNB selects the same RRC/UP algorithms if possible. Target eNB sends handover response message to the target MME, which includes the new C-RNTI, selected RRC/UP algorithms, and some other parameters (see 25.813, section 9.1.5). For the target eNB, K_{eNB*+} from K_{eNB*} look the same and are used identically. Target eNB derives a new K_{eNB} from C-RNTI and K_{eNB*+} by K_{eNB}_new = KDF(K_{eNB*+} ‖ C-RNTI) and further derives K_{RRCenc}, K_{RRCint}, K_{UPenc} from the K_{eNB}_new.

6. Target MME forwards the handover response with selected MME algorithms to source MME which sends it to source eNB including NAS-MAC. There needs to be a bit included to signal to the UE how the K_{eNB} is to be derived.

7. Source eNB sends the handover command message to the UE including NAS level message with the selected NAS algorithms and NAS-MAC. This AS level message is protected with the old RRC integrity and ciphering keys shared with the source eNB. The message also includes target eNB algorithms (for RRC and UP) if different than the source eNB algorithms. There needs to be a bit included to signal to the UE how the K_{eNB} is to be derived.

8. Based on the notification bit the UE knows whether to first calculate K_{eNB*+} to then derive K_{eNB}. UE derives the K_{eNB*}, potentially K_{eNB*+}, new K_{eNB}, K_{RRCenc}, K_{RRCint}, and K_{UPenc} and sends handover confirm message to the target eNB integrity protected and ciphered with the new RRC keys.

9. UE derives the K_{eNB*}, new K_{eNB}, K_{RRCenc}, K_{RRCint}, and K_{UPenc} and sends handover confirm message to the target eNB integrity protected and ciphered with the new RRC keys. Based on the notification bit the UE knows whether to first calculate K_{eNB*+} to then derive K_{eNB}.

Using a key derived from K_{ASME}, i.e. KDF(K_{ASME} ||"Handover Key"), for derivation of K_{eNB*+} allows delegation of forward providing forward security for K_{eNB} without potential of compromising any of the other keys.

NOTE: source and target MME might be identical.

Editorial note: the notification bit of this section and the handover type indicator need to be merged.

Editor's note: Deriving new NAS keys based on algorithms identifier as the only parameter is ffs (see S3-070533).

Editor's note: It is ffs if a separate NAS level SMC is used to change NAS algorithms on inter-MME handover (see S3-070533).

Considerations on C-RNTI and its randomness (S3-070511):

The solutions above derive the new key using the hash of the old eNB key and the C-RNTI value, KeNB_new = KDF(KeNB* || C-RNTI) in both inter and intra MME handovers. The goal of this transformation is to make the job of an attacker, who has an eNB key, more difficult because he would need to overhear all the messages that allocate C-RNTI in order to derive the current new eNB key.

Suppose the UE moves from eNB1 to eNB2. The attacker has the key at eBS1, but did not hear the C-RNTI allocation in the HO messages, but the attacker collects the rest of the conversation from eNB2. According to the S3-070306, the attacker should not be able to decrypt the rest of the conversation happening through eNB2. Unfortunately, we demonstrate that this is not the case – the attacker can, with a modest effort, get the new key, KeNB_new:

Knowing the eNB1 key, the attacker creates 2^{32} candidate eNB_new keys; one candidate key for each possible C-RNTI value. Using the candidate keys, the attacker tries to decrypt the conversation at eNB2. For all the candidates save one, the decrypted text would appear to be random. For the candidate key with the correct C-RNTI value and the correct eNB_new key value, the decrypted text would have recognizable and expected formats, like protocol headers, etc. Thus the attacker would be able to recognize the correct key to decrypt the rest of the conversation that went through eNB2.

The 2^{32} choices and verifications would not take much time even on a single modern PC. If the attacker does not know the C-RNTI for two intermediary eNB in the chain then the complexity is 2^{64}; for the case of three missing C-RNTI value, the complexity is 2^{96}. One needs four missing intermediary C-RNTI values to reach 2^{128} complexity.

S3-070511 recommended that instead of using a 32 bit C-RNTI value, the target eNB should generate a 128 bit random value and use that as the input to the key derivation. The target eNB should also send this value to the UE via the source eNB.

To save on signalling bandwidth at the handovers, SA3 #48 proposed to augment 32-bit C-RNTI (which has to be transported during the handoff anyway) with 96-bit random value, thas bringing randomness of the concatenated length of random string to 128.

It is ffs how to generate such random value.

Inter-eNB handover with MME relocation (From S3a070928)

Editor's note: SA3 is aware of that the state names do not match the current naming in TS 23.401, and have to be updated.

The following handling of keys is agreed by SA3.

At an inter-eNB hand over with MME relocation, the K_eNB is chained in the same way as if it was a regular intra MME eNB hand over. However, there is the possibility that the source MME and the target MME do not support the same set of NAS algorithms or have different priorities regarding the use of NAS algorithms. In this case, the target MME re-derives the NAS keys from K_ASME using the NAS algorithm identities as input to the NAS key derivation functions. All inputs, in particular the K_ASME, will be the same in the re-derivation except for the NAS algorithm

identity. It is essential that the NAS SQN is not reset unless the K_ASME changes. This prevents that, in the case a UE moves back and forth between two MMEs the same NAS keys will be re-derived time and time again resulting in key stream re-use. Since K_ASME only changes when a new AKA is run, it is a requirement that the NAS SQN is only reset when there is a new AKA run. In case the target MME decides to use NAS algorithms different from the ones used by the source MME, a NAS SMC must be sent from the MME to the UE.

Considerations on prepared handovers

It is a threat that multiple cells belonging to other eNBs are prepared to handover, and hence all have access to enough information to derive the K_{eNB} the UE will use after the handover (by simply cycling through all possible C-RNTIs).

This is countered by including the physical cell ID of the target cell in the derivation of K_{eNB}*, which makes the K_{eNB}* unique per target cell.

Editor's note: the use of physical cell ID need to be checked by RAN2.

7.4.13.3 Alternatives for key handling on handover between different SAE/LTE networks

The alternatives for key handling on handover between different SAE/LTE networks are the same as described in the case of inter MME handoverin the last section.

If Alternative 2 or 3 are chosen for this type of handover, the target operator should be able to initiate a new authentication as soon as possible after the handover. It is currently ffs whether or not authentication can take place during an ongoing connection. If this is not the case, the target operator should at least be able to initiate a new authentication as soon as UE transits from active to idle (see the section on "Key handling on active to idle and idle to active transitions".

7.4.13.4 Summary of evaluation of alternatives

Table 1 Key derivation alternatives compared for the different handover types

	Assumption	**Alternative-1**	**Alternative-2**	**Alternative_3**
Inter eNB, Intra MME handover	MME is not involved in intra-MME handover	preferred if generation of encrypted keys for multiple eNBs in MME is acceptable	OK (alternative 3 preferred)	OK
Inter-MME Handover (within same PLMN)	MME relocation	Unwanted due to creating HSS state.	OK (alternative 3 preferred)	OK
Inter-MME Handover (between PLMNs)	MME relocation	Unwanted due to creating HSS state.	OK	OK

7.4.13.5 Key handling on handover from UTRAN to E-UTRAN[35]

When UE handovers from UTRAN to EUTRAN, SGSN shall transfer CK || IK to MME in the relocation request message. MME shall derive K'_{ASME} from CK || IK with the help of a one-way key derivation function KDF:

KDF(CK || IK) = K'_{ASME} . MME shall derive the NAS keys and K_{eNB} from K'_{ASME}."

For $\mathbf{K_{eNB}}$ derivation, apparently there are two mechanisms, but neither of them can work here:

1, In MME and UE, use K_{ASME}, NAS SQN and other parameters. But when HO from UTRAN to EUTRAN, there is no valid NAS SQN

---

[35] This section is from S3-080067.

2, In target eNB and UE, use $\mathbf{K_{eNB}}$ * and C-RNTI. But when HO from UTRAN to EUTRAN, there is no $\mathbf{K_{eNB}}$ * in target eNB and UE

Furthermore, for the parameter used to derive $\mathbf{K_{eNB}}$, the following 3 conditions need to be considered.

i , If MME derives $\mathbf{K_{eNB}}$, it must happen between forward redirection request and HO request, since the 1st message sends the IK, CK and the second message sends out $\mathbf{K_{eNB}}$ to eNB.

ii , possible security threat here: when HO fails after the target eNB receives the $\mathbf{K_{eNB}}$, then later if the HO to another target eNB happens, the same $\mathbf{K_{eNB}}$ could possibly be used if the same key derivation parameter is used. If the first target eNB is compromised, the attacker can have current $\mathbf{K_{eNB}}$.

iii , UE and MME need a way to share this key derivation parameter.

According to the above analysis, we propose to use a random number generated in MME every time it receives the forward redirection request to derive $\mathbf{K_{eNB}}$.

In relocation response and UTRAN HO command this random number is transferred to UE to derive $\mathbf{K_{eNB}}$.

During UTRAN to EUTRAN HO, MME generates a random number and uses it with K'_{ASME} to derive $\mathbf{K_{eNB}}$. The random number is sent to UE during HO and UE uses it with K'_{ASME} to derive $\mathbf{K_{eNB}}$.

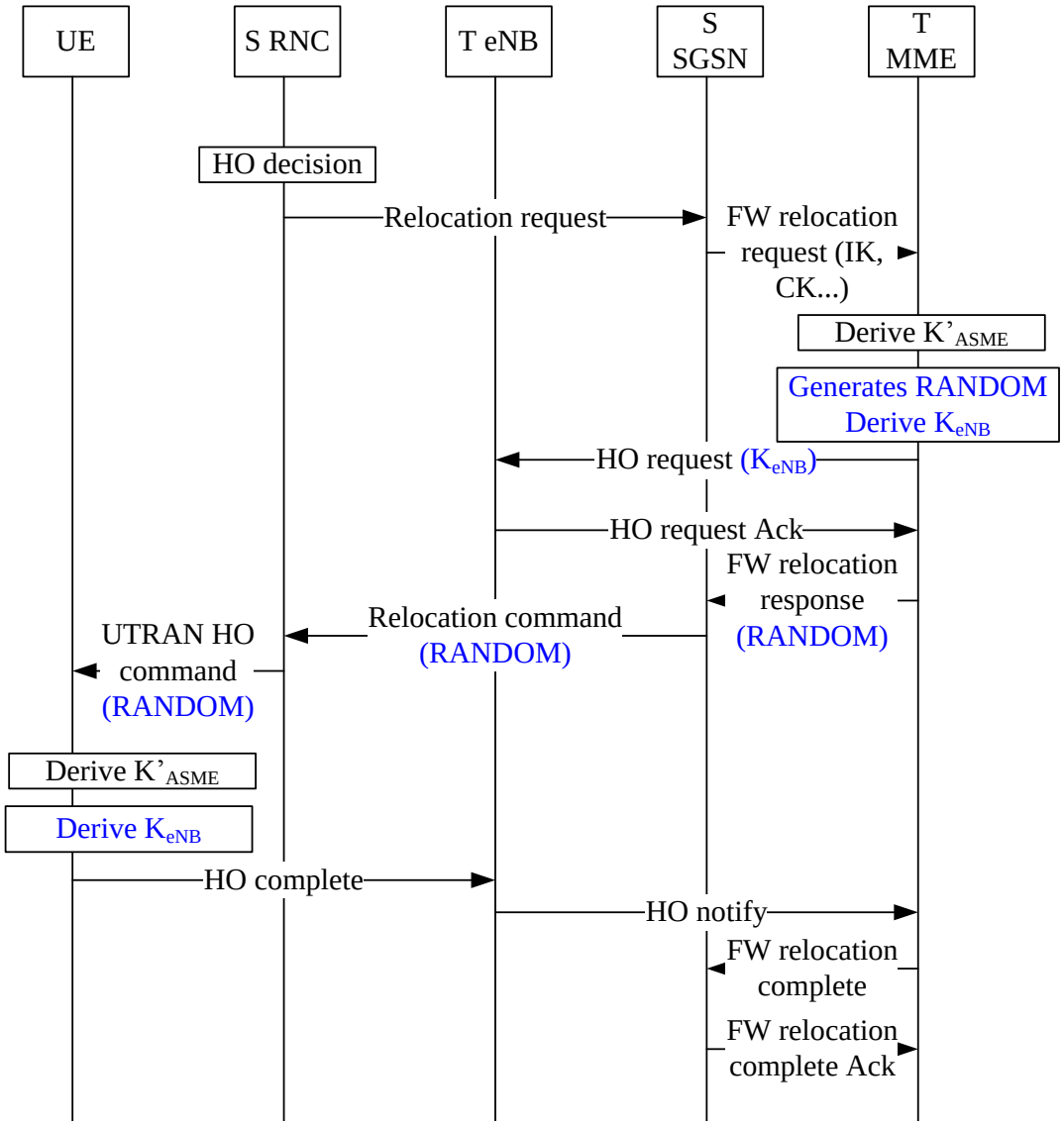


Figure 23 $\mathbf{K_{eNB}}$ Derivation during UTRAN to EUTRAN HO

1, Source RNC makes HO decision base on measurement report or other conditions.

2, Source RNC sends Relocation request to Source SGSN.

3, Source SGSN forwards Relocation request to Target MME. In this message, IK, CK are sent to the target MME.

4, Target MME derives K'_{ASME} using IK, CK.

5, Target MME generates a random number, and uses K'_{ASME} and this random number to derive K_{eNB}.

6, In HO request, target MME sends K_{eNB} to target eNB.

7, The Target eNB replies HO request Ack.

8, In Forward relocation response, relocation command, and HO from UTRAN command, the random number is transferred to UE. Note that HO from UTRAN command is protected by UTRAN security, so is the random number.

9, UE derives K'_{ASME} using IK, CK. UE derives K_{eNB} using K'_{ASME} and the random number.

10. UE sends HO complete to target eNB.

10, Target eNB sends HO notify message to Target MME.

11, Target MME forwards relocation complete to source SGSN.

12, Source SGSN replies relocation complete Ack.

Editor's Note: it needs to check if it's better to put difference at NAS level to cover also idle case.

Editor's Note: the solution can be based on the Cell-ID used instead of RANDOM.

Editor's Note: it should be studied whether the detailed signalling flows are appropriated at a stage 2 specification.

The following is another alternative:[36]

Editor's Note: the availability of eNodeB ID need to be verified.

[36] This section is from S3-080369.

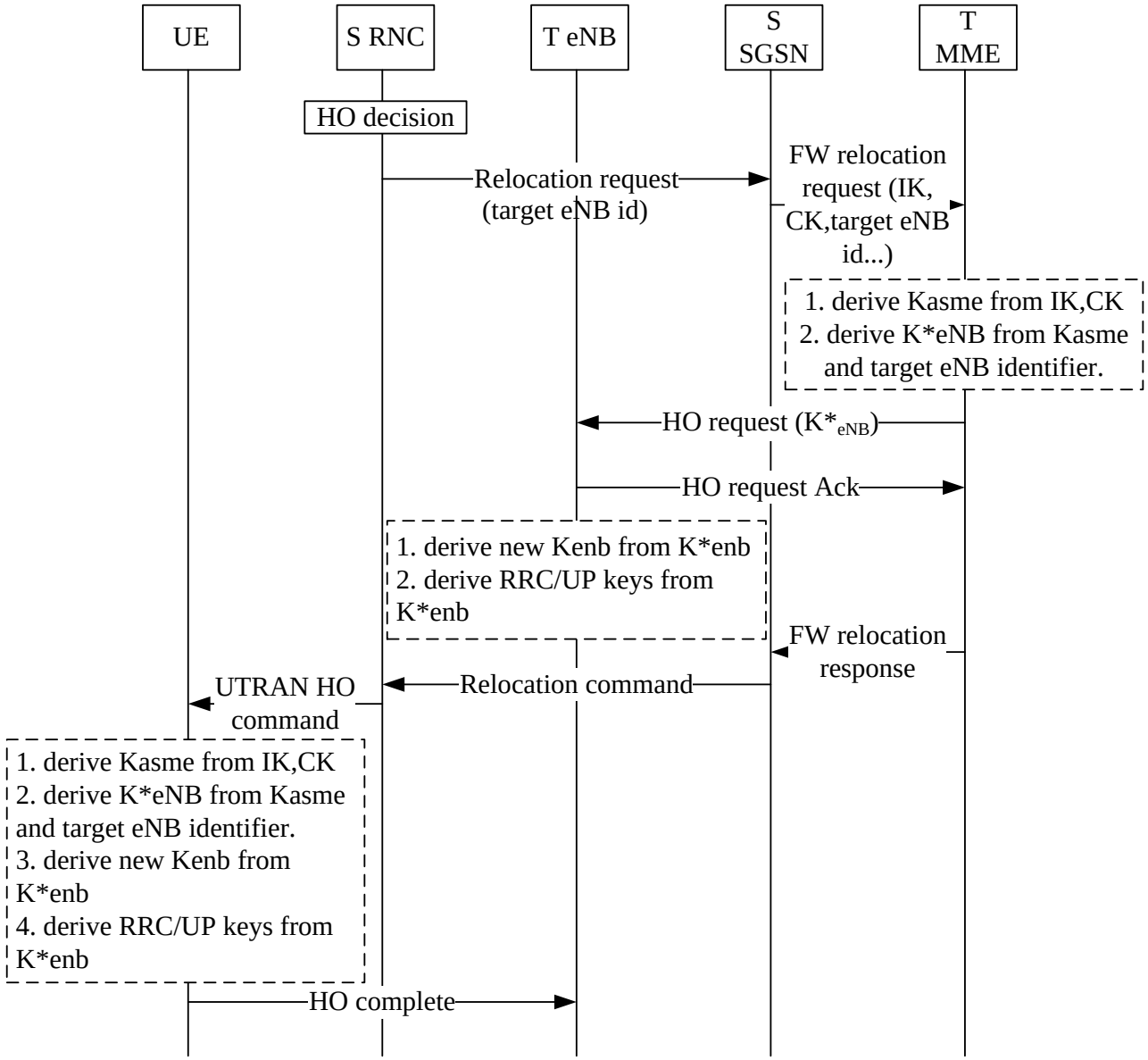


Figure 24 Kenb derivation during inter-RAT handover

1, Source RNC makes HO decision base on measurement report or other conditions.

2, Source RNC sends Relocation request to Source SGSN, which includes the target eNB identifier.

3, Source SGSN forwards Relocation request to Target MME. In this message, IK, CK and the target eNB identifier are sent to the target MME.

4, Target MME derives K'_{ASME} using IK, CK, then derives K^*_{eNB} from K'ASME and the target eNB identifier.

5, In HO request, target MME sends K^*_{eNB} to target eNB.

6, The Target eNB replies HO request Ack to the target MME. The target eNB id is contained in this message. The target eNB derives new Kenb from K*enb and other parameters in the same way as inter-eNB handover, and then derives RRC/UP keys from new Kenb.

7, The target MME replies Forward relocation response to the source SGSN, which replies relocation command to source RNC, which sends HO from UTRAN command to the UE. The target eNB id is transferred in these messages.

8, The UE derives RRC/UP keys from new kenb in the same ways as in the MME/eNB.

9. UE sends HO complete to target eNB.

7.4.14 Security algorithm negotiation and Security mode command in SAE/LTE networks[37]

7.4.14.1 General

In this section we propose different alternatives for how the security algorithms for RRC, UP and NAS protection could be agreed upon between UE and a serving network during network attachment and on idle to active transitions. In addition, it also includes security algorithm selection upon mobility events within LTE. Here the UE and the target network entities have to agree on security algorithms to use after a mobility event occurred.

7.4.14.2 Background: algorithm selection in UMTS

In UMTS (TS 33.102) the security mode setup procedure is used to negotiate an encryption algorithm and an integrity algorithm between RNC and ME upon network attachment. This works as follows:

- ME sends its UMTS security capabilities (i.e. the encryption and integrity protection algorithms it supports) and optionally its GSM encryption capabilities to the RNC during RRC connection setup. RNC stores the received capabilities of ME
- After successful authentication, MSC sends RNC a list of allowed encryption and integrity protection algorithms (along with CK and IK). RNC chooses an encryption mechanism and an integrity protection mechanism that are supported by RNC and UE and that are allowed by MSC.
- RNC acknowledges its choice to ME in the security mode command (SMC). The SMC replays the capabilities of ME and is integrity protected. The replay guarantees that ME notices it if an attacker manipulated ME's capabilities on transfer to RNC (protection against "bidding down"). The integrity protection of the selected algorithm guarantees that an attacker cannot manipulate RNC's choice of algorithm without ME detecting the manipulation.

7.4.14.3 Requirements for algorithm selection in SAE/LTE

Requirement 1: An active UE and a serving network shall agree upon algorithms for

- RRC encryption, RRC integrity protection (to be used between UE and eNB)
- UP encryption (to be used between UE and eNB)
- NAS encryption and NAS integrity protection (to be used between UE and MME)

Requirement 2: The serving network shall select the algorithms to use dependent on

- the capabilities of UE,
- the capabilities of the currently serving network entity
- restrictions set by the home network of the subscriber (ffs, cf TR 23.008)
- SN-wide policies on allowed security algorithms

(In this document, "capabilities" always refers to the supported encryption and integrity algorithms.)

Requirement 3: Each selected algorithm shall be acknowledged to UE in an integrity protected way such that UE is ensured that the algorithm selection was not manipulated ("bidding down protection of networks choice").

Requirement 4: The capabilities the UE sends to the network shall be repeated in an integrity protected message to UE such that "bidding down attacks" against UE's capabilities can be detected by UE.

---

37 This section is from S3-070100.

7.4.14.4 Alternatives for security mode command and algorithm selection in SAE/LTE

7.4.14.4.1 Security mode command and algorithm selection at initial attachment or in transitions to active mode

According to the requirements described in the last section we need to answer the following questions

- Which network entity or entities should select the security algorithms to use (cf. Requirement 2)?

The security algorithms could either be selected by the eNB, and the MME for RRC, UP and NAS protection respectively or all algorithms could be selected by MME. As opposed to the second case, in the first case MME does not have to be configured to know the capabilities of all eNBs under its control.

- Which network entity should acknowledge the selection of which algorithms to UE in an integrity protected way (cf. Requirement 3)?

It seems natural to assume that MME acknowledges the selection of NAS algorithm and eNB acknowledges the selection of RRC and UP algorithms or MME acknowledges all selected algorithms to UE. However, in case eNB selects RRC and UP algorithms and MME selects NAS algorithms, there are several options for the acknowledgement, see below.

- To which network entity should UE send its capabilities in which message, and which network entity should repeat the received capabilities to UE in an integrity protected message (cf. Requirement 4)?

In UMTS, UE sends its capabilities to RNC during RRC connection establishment and RNC selects the algorithms to use. MSC does not have to obtain the UE capabilities. However, in SAE/LTE MME has to obtain knowledge of (at least the NAS part) of UEs capabilities in order to be able to select the NAS protection algorithms. As a consequence it seems most natural to assume that UE sends its capabilities to MME in an initial layer 3 message.

- When should the selection and acknowledgement of algorithms take place?

NAS signaling protection has to be started after each attachment, e.g. for subsequent protection of a "re-attachment request" (cf. TR 23.882, Section 7.13). Therefore the algorithm selection by the network entity (or entities) and the acknowledgement of the selection to UE has to take place immediately after or during each network attachment. RRC signaling protection and UP encryption need to be started on each idle to active transition. Idle to active transitions e.g. occur due to mobile originated service requests or due to paging. The selection and acknowledgement of RRC signaling protection and UP encryption algorithms therefore has to take place upon idle to active transitions. This leads to the following alternatives:

Alternative 1: In this alternative it is the MME that selects all algorithms and acknowledges the selected algorithms to UE. The message flow for Alternative 1 is illustrated in Figure 25 and Figure 26and described in more detail bellow.

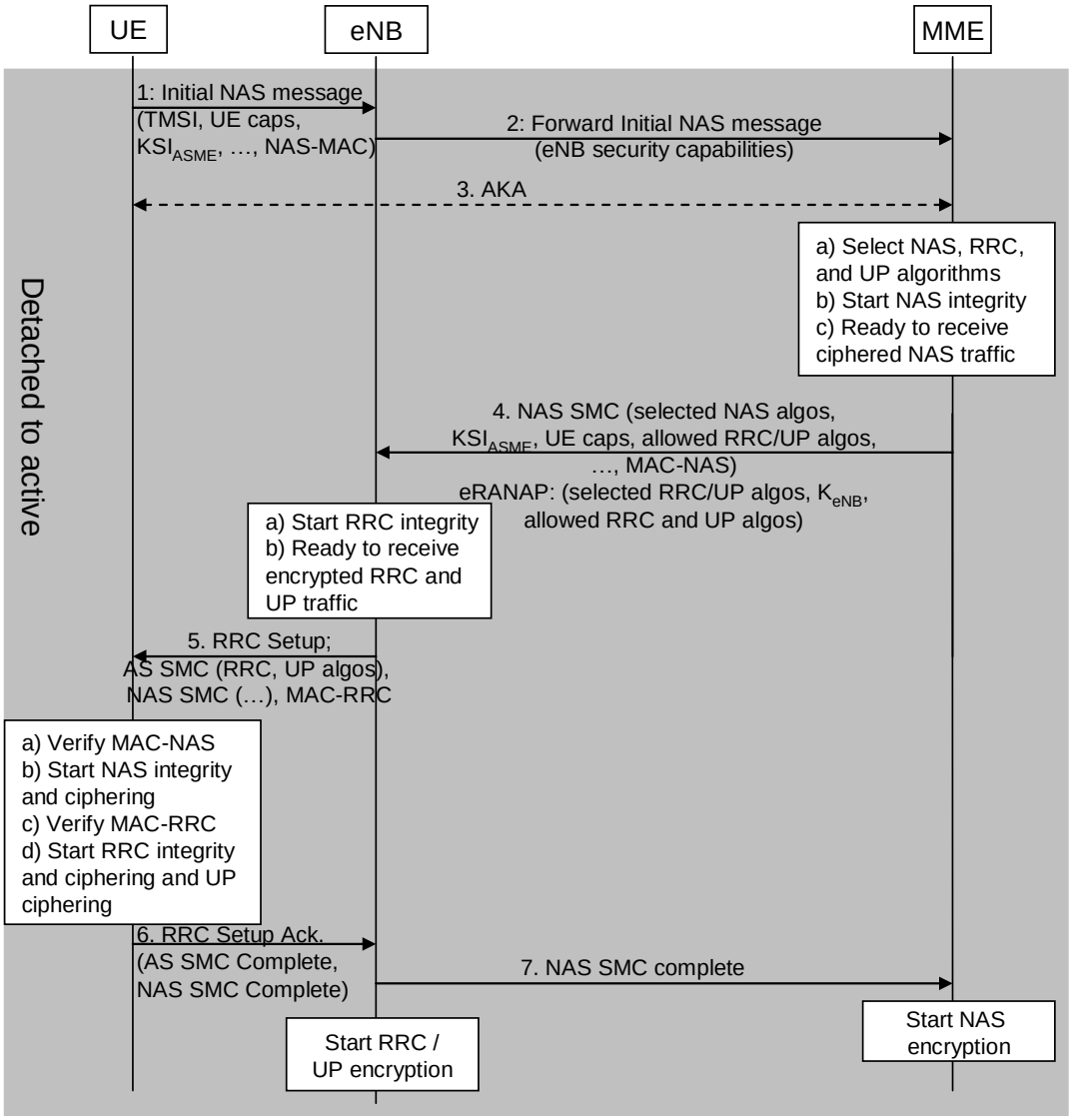


Figure 25 : Alternative 1, detached to active mode state transition, where MME selects all algorithms and MME protects the acknowledgement for NAS algorithms, eNB protects the acknowledgement for UP and RRC algorithms.

Editor's Note: Initial radio layer messages not included.

Upon detached to active transitions:

1. UE sends a RRC connection request to eNB including initial NAS (L3) request to MME (see TR 23.882, Section 7.14). The initial L3 message includes all UEs security capabilities, UE identifier (IMSI or TMSI), and KSI$_{ASME}$. NAS-MAC is included if UE has valid NAS keys (i.e. KSI is also valid).

2. eNB forwards the NAS request to MME within eRANAP including eNB security algorithm capabilities

3. Optionally MME requests a new AKA authentication

4. MME selects RRC, and UP algorithms. MME sends a list of allowed RRC algorithms, allowed UP algorithms, KeNB, selected RRC and UP algorithms to the eNB as part of the UE's security context for the eNB via eRANAP.. MME additionally selects the NAS algorithms, starts NAS integrity and prepares to receive encrypted NAS traffic. MME sends the NAS SMC message to eNB.

   Editor notes: It is ffs whether allowed RRC algorithms, allowed UP algorithms shall be acknowledged to UE in NAS SMC, so that UE can verify that AS SMC includes RRC and UP algorithms from the allowed set (in idle-to-active state transitions and also during mobility).

5. eNB starts integrity protection, gets ready to receive encrypted RRC and UP traffic and sends the AS SMC to UE, including the NAS SMC message received from MME. The AS SMC message includes the selected RRC algorithms and the selected UP algorithm and is protected with AS integrity protection.

6. UE verifies MAC-NAS on the NAS SMC and starts NAS integrity and NAS encryption. UE verifies MAC-RRC on the AS SMC message and starts RRC integrity and RRC and UP encryption. UE sends AS SMC Complete and NAS SMC Complete message.

7. Upon receipt of the RRC setup acknowledge message, eNB starts RRC and UP ciphering and forwards NAS SMC complete message for the MME. Upon receipt of the NAS SMC complete message, MME starts NAS ciphering

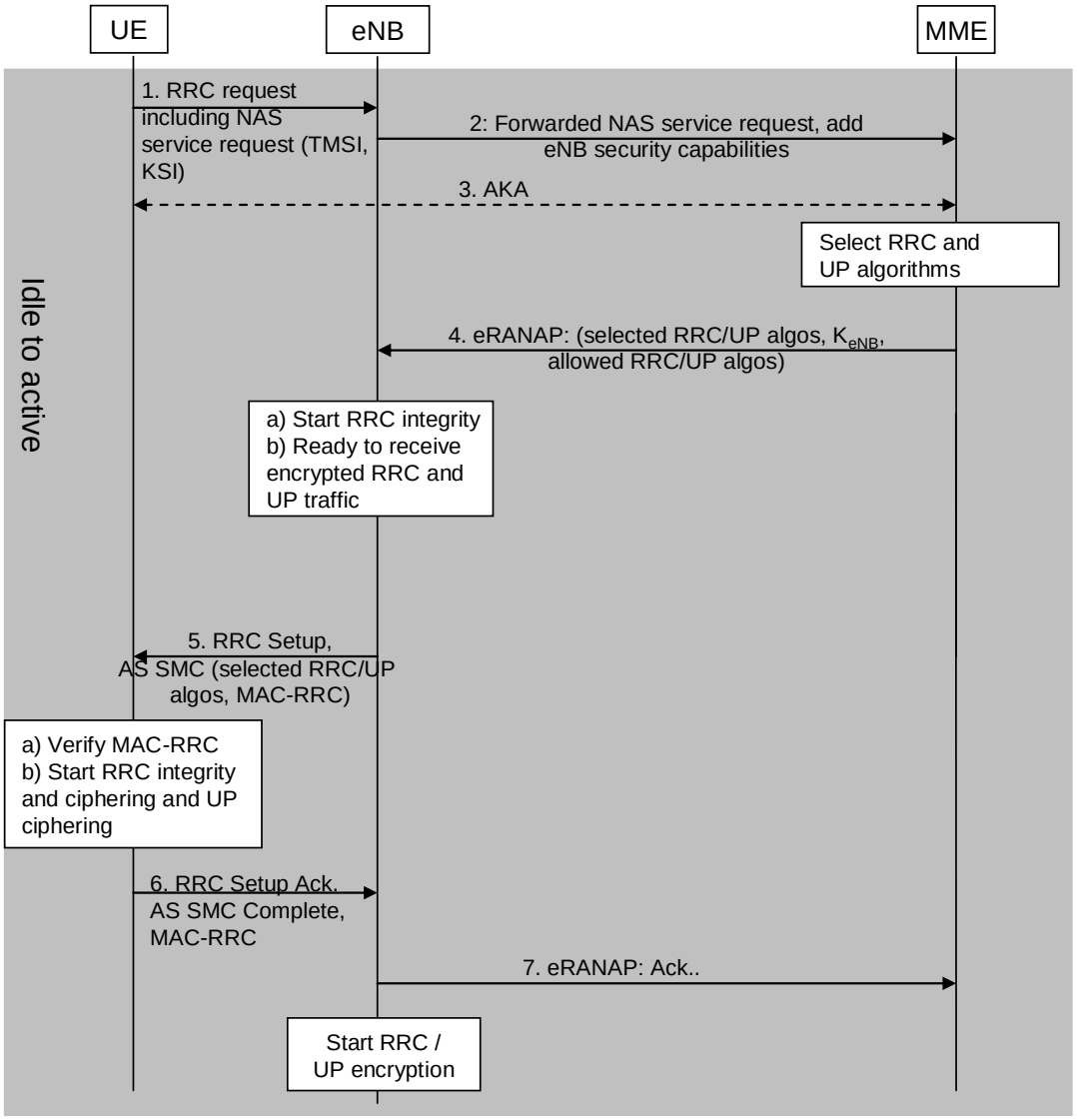


Figure 26: Alternative 1, idle to active mode state transition, where MME selects RRC and UP algorithms and eNB protects the acknowledgement for UP and RRC algorithms.

Editor's Note: Initial radio layer messages not included.

Upon idle to active transitions:

1. UE sends a RRC connection request to eNB including service request to MME. The service request message includes UE identifier (IMSI or TMSI), KSI_{ASME} and NAS-MAC.

Editor's note: It is ffs whether KSI_{ASME} is needed on idle-to-active state transition if the NAS level service request message has to be length optimized. In case KSI is not included error handling must be designed properly to address KSI (i.e. K_{ASME}) mismatch.

2. eNB forwards the request to MME within eRANAP including eNB security algorithm capabilities
3. Optionally MME requests a new AKA authentication
4. MME selects RRC, and UP algorithms. MME sends a list of allowed RRC algorithms, allowed UP algorithms, K_{eNB}, selected RRC and UP algorithms to the eNB as part of the UE's security context for the eNB via eRANAP.

   Editor's note: It is ffs whether NAS algorithms and keys can be changed during idle-to-active mode state transition using an additional NAS-SMC.

5. eNB starts integrity protection, gets ready to receive encrypted RRC and UP traffic and sends the AS SMC to UE. The AS SMC message includes the selected RRC algorithms and the selected UP algorithm and is protected with AS integrity protection.
6. UE verifies MAC-RRC on the AS SMC message and starts RRC integrity and RRC and UP encryption. UE sends AS SMC Complete.
7. Upon receipt of the RRC setup acknowledge message, eNB starts RRC and UP ciphering and sends an ack to MME.

Alternative 2: The NAS algorithm selection is the same as in Alternative 1, eNB selects RRC and UP algorithms itself such that MME does not have to know the capabilities all eNBs.

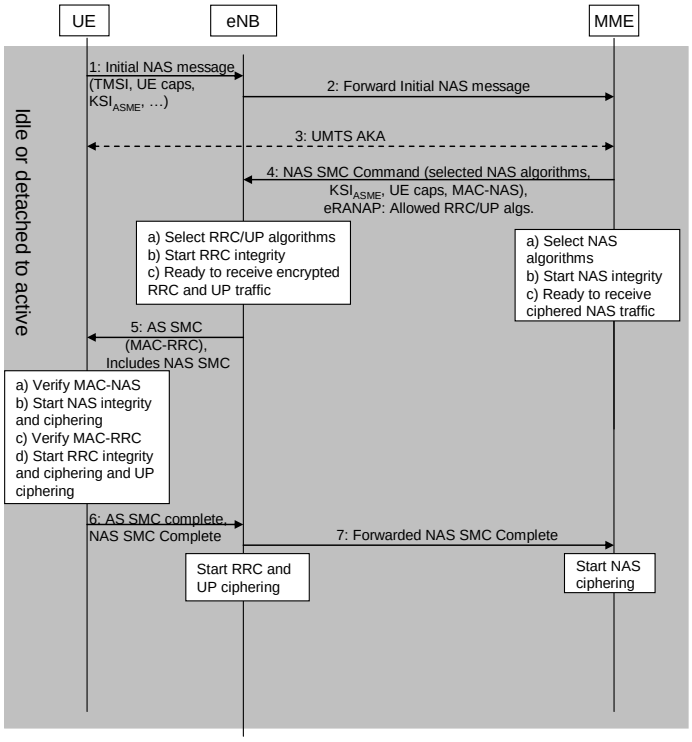


Figure 27: Alternative 2, MME and eNB select the corresponding algorithms and MME protects the acknowledgement for all algorithms

Editor's Note: Initial radio layer messages not included.

Upon idle or detached to active transitions:

1. UE sends a RRC connection request to eNB including initial NAS (L3) request to MME (see TR 23.882, Section 7.14). The initial layer 3 message includes all UEs security capabilities

2. eNB forwards the request to MME

3. Optionally MME requests a new authentication

4. MME sends a list of allowed RRC algorithms, allowed UP algorithms and the RRC and UP capabilities of UE to the designated eNB. In the detached to active case, MME additionally selects the NAS algorithms (in the idle to active case the NAS algorithms are already selected), starts NAS integrity and gets ready to receive encrypted NAS traffic.. MME sends the security mode command message to eNB.

5. eNB starts integrity protection, gets ready to receive encrypted RRC and UP traffic and sends the AS SMC to UE, including the NAS SMC message received from MME. The AS SMC message includes the selected RRC algorithms and the selected UP algorithm and is protected with AS integrity protection.

6. UE verifies MAC-NAS on the NAS SMC and starts NAS integrity and NAS encryption. UE verifies MAC-RRC on the AS SMC message and starts RRC integrity and RRC and UP encryption. UE sends AS SMC Complete and NAS SMC Complete message.

7. Upon receipt of the RRC setup acknowledge message, eNB starts RRC and UP ciphering and forwards SMC complete message for the MME. Upon receipt of the security mode complete message, MME starts NAS ciphering

Alternative 3: In Alternative 3, eNB, and MME each selects the corresponding algorithms and protects the acknowledgement messages sent to UE. The message flow for Alternative 3 is illustrated in Figure 28 and the message flow is detailed below.

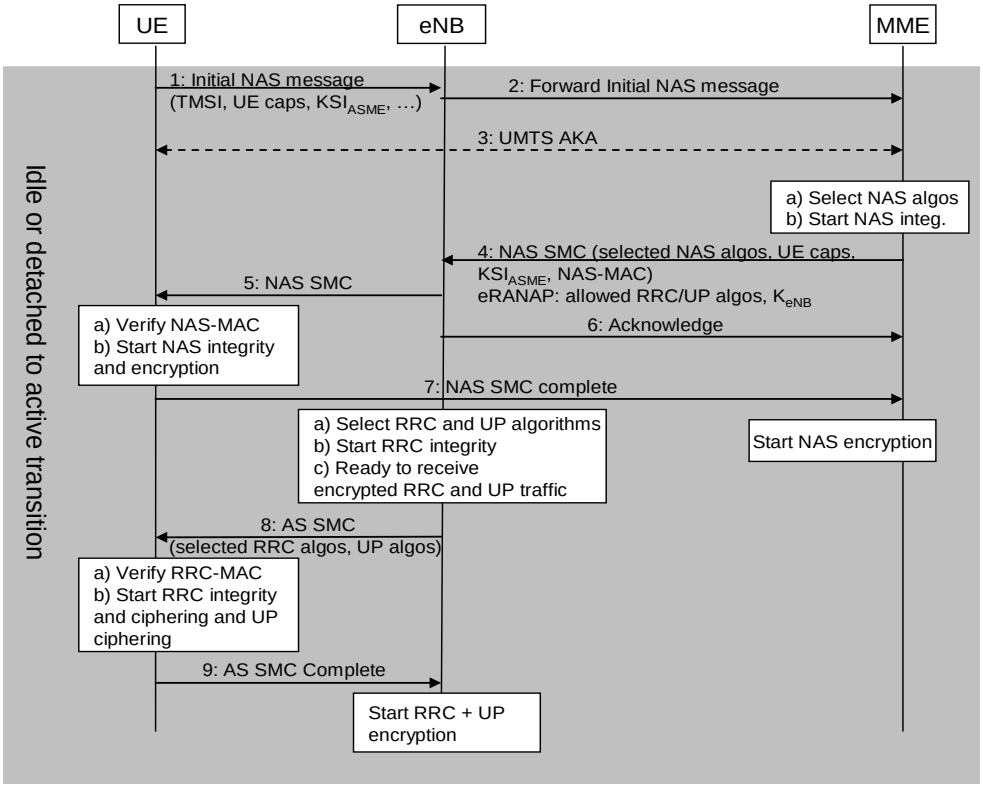


Figure 28: Alternative 3, MME and eNB select and MME protects the acknowledgement NAS, eNB protects the acknowledgement for RRC and UP algorithms

Editor's Note: Initial radio layer messages not included.

Editor's Note: It is ffs whether the order of the messages 6 + 7 and 8 + 9 should be reversed. In this case it would also be possible to combine the messages 4 and 8.

Editor's Note: if the eNB included its algorithm capabilities in message 2 then MME could select the eNB algorithms and send the selected algorithms to the eNB.

Upon idle to active and detached to active transitions:

1. UE sends a RRC connection request to eNB including initial NAS (L3) request to MME (see TR 23.882, Section 7.14). The initial layer 3 message includes all UEs security capabilities
2. eNB forwards the request to MME
3. Optionally MME requests a new authentication
4. Upon receipt of a service request, MME sends a list of allowed RRC algorithms and the allowed UP algorithms to the designated eNB , including the RRC and UP capabilities of UE. MME sends a NAS security mode command message including the selected NAS algorithms to UE. The message is integrity protected
5. eNB forwards the NAS SMC to UE
6. eNB acknowledges the eRANAP message.
7. UE sends NAS SMC Complete to MME. MME verifys the NAS SMC Complete message and starts NAS encryption.
8. eNB selects the RRC security algorithms and UP algorithm in compliance with its own capabilities, the allowed algorithms and UEs capabilities.[38]eNB starts RRC integrity and gets ready to receive encrypted RRC traffic. eNB acknowledges its choice of UP and RRC algorithms to UE in an integrity protected RRC security mode command included in the radio resource setup message
9. UE verifies the correctness of the integrity protection, starts RRC integrity and RRC and UP encryption and returns an AS SMC Complete message. After receipt of the correctly protected AS SMC Complete message, eNB starts ciphering of RRC and UP messages

Table 2. Comparison of the three alternatives

	Alternative 1	**Alternative 2**	**Alternative 3**
Selection of security algorithms	In MME	In MME and eNB	In MME
Protection of acknowledgement	NAS integrity protection	NAS integrity protection	NAS integrity protection (NAS SMC), RRC integrity protection (AS SMC)
Signaling Efficiency		Extra round trip between eNB and MME due to the eNB RRC, UP algorithm selection	More messages over the air interface due to the separated NAS SMC and RRC level procedures.

All alternatives have equal strength of security protection. It is a security decision, however, which entity is in control of selecting algorithms. Preference is give here to control by the MME, which favours alternative 1. From an efficiency point of view, alternative 1 is clearly much better than the optimised alternative 3.

[38] MME transfers the UE capabilities and not the intersection of the UE capabilities with the allowed algorithms such that eNB can transfer them upon handover.

7.4.14.4.2 Security mode command and algorithm selection on idle mode mobility

Idle mode mobility results in a location update procedure. The security algorithm selection upon location updates shall be performed in the same way as on initial attachment.

7.4.14.4.3 Security mode command and algorithms selection on handover

Upon handover the change of a network entity may require a change of the currently selected security algorithms if the source and the target entity do not have the same capabilities or belong to different administrative domains.

For intra-MME handover with eNB change, MME involvement is currently not foreseen (TR 25.813).

Handover without MME involvement:

In this case, the source eNB will transfer UE-context to the target eNB. This context shall include the UE algorithm capabilities, allowed RRC/UP algorithms for the UE, and the currently used security algorithms in the source eNB.

The target eNB selects the RRC and UP algorithms for use after handover and transfers it to the source eNB. If the currently used algorithms are supported by the target eNB the choice shall be the currently used security algorithms. In other cases target eNB selects an algorithm based on the UE capabilities and allowed algorithms set for the UE (provided originally by MME) and includes the selected algorithms in the integrity protected and ciphered Handover Command message to the UE (see Figure 15 on section "Key refresh on Intra-MME handover (S3-070306)").

The source eNB may check that the target eNB algorithm selection complies with the allowed algorithms for the UE.

The threat of a compromised eNB downgrading ciphered communications is mitigated by sending allowed eNB algorithms in the NAS SMC. In case non-NULL ciphering algorithm has been selected for AS in detached to active mode state transition, downgrading to NULL ciphering algorithm is not allowed during active or idle modes. UE may check the target eNB algorithms compliance with the allowed eNB algorithms received in the NAS level SMC.

Error case handling situations on handover without MME involvement are ffs.

Handover with MME involvement:

This case shall be handled in the same way as handover without MME involvement with the following exception:

The target MME shall select the NAS algorithms to use and the target eNB shall select the UP/RRC security algorithms based on the allowed RRC/UP algorithms informed by the MME and acknowledge its choice to MME. The MME then sends all the NAS/RRC/UP security algorithms choice in the handover command to UE via both the source MME and the source eNB or changes the algorithms after handover with the SMC procedure(s). See Figure 16 in section "Key refresh on Inter-MME handover".

Editor's note: NAS algorithm change during the active mode with a NAS level security mode command without a handover is ffs. As an example MME could be changed, but the eNB would be the same (i.e. no handover for UE).

(From S3-80054)

There is a threat of bidding down to the NULL algorithm at handover. Assuming that the target eNB knows the *correct* UE security capabilities, the *correct* MME_allowed_set and its own O&M_allowed_set, it is possible to select the algorithm with highest priority at a handover, and hence avoiding that a lower priority algorithm "sticks". In particular it would be possible to ensure that the NULL-algorithm does not stick.

Since any UE supports at least two encryption algorithms, it is not possible that the UE security capabilities only contain the NULL algorithm. Therefore, the source eNB cannot claim that the UE capabilities only contain the NULL algorithm.

However, the source eNB can still change the MME_allowed_set to contain only the NULL algorithm and the NULL algorithm would again be sticky.

As can be seen the failing-point is that the source eNB can manipulate the MME_allowed_set.

There are many possible options to prevent the manipulation of the MME_allowed_set, some of them are:

1. Accept the fact that a compromised eNB can make the NULL algorithm (and any other less preferable algorithm choice) stick even after handover to non-compromised eNBs.
2. Ensure that the UE reports the MME_allowed_set to the target eNB in the handover confirm message. If there is a mismatch with the MME_allowed_set received from the source eNB, algorithm re-negotiation can take place. Possibly also reporting of the misbehaving source eNB to the O&M system.
3. Ensure that the UE reports the selected algorithm to the MME in a NAS message after the handover is complete. If the UE is using the NULL algorithm, the MME would know that the eNB supports a better algorithm, and that the MME_allowed_set contains a better algorithm, so the MME could take action.
4. Have the MME send the MME_allowed_set to the UE at each TAU procedure.

Options 1, 2, and 3 have been discussed in previous meetings. Option 4 has not been discussed so far.

The option with least impact (but providing no protection) is option 1. Options 2 and 3 allow the network to detect the attack and react. Option 3 requires a new NAS procedure, whereas option 2 and 4 only piggy back the information on existing messages. Therefore, option 2 and 4 have the least impact and still provide protection.

There are two main differences between option 2 and option 4:

- Option 2 detects the attack immediately, whereas option 4 only detects the attack when the UE changes tracking area.
- Option 2 induces higher bandwidth consumption than option 4, since it sends the MME_allowed_set more frequently.

A severe problem with option 4 is that the UE cannot know if the O&M_allowed_set for the target eNB only contains the NULL algorithm. This means that the NULL algorithm may well be a valid choice for the target eNB, and hence option for does not solve the problem.

There is a threat of bidding down to a general algorithm. To prevent that the compromised source eNB bids down to any algorithm, the UE security capabilities must also be given to the target eNB in a secure fashion. This can be achieved in the same way as the MME_allowed_set is communicated from the UE to the network.

A related issue is how to determine which algorithms are preferable. Currently it is FFS is this is to be done by having the MME_allowed_set being an ordered list An ordered list cannot be stored as compactly as a non-ordered set (which can be represented by a bit-vector, c.f., supported algorithms in GERAN/UTRAN). In all of the above options it is better the shorter the MME_allowed_set is (since it is transferred between nodes). This has as a simple consequence that it is preferable to keep the O&M_allowed_set in priority order instead of the MME_allowed_set.

7.4.14.4.4 Algorithms selection on handover to and from 2G/3G

Editor's Note: The information contained in this Section was included in Section 9 and 10 of TS 33.abc.

Handover from LTE to 2G/3G:

UE capabilities send from UE to MME in the initial layer three messages shall include the GERAN and UTRAN UE capabilities. On handover to GERAN, MME shall include the UE capabilities in the handover request sent to SGSN. SGSN shall select the GERAN algorithm to use and indicate its choice in the handover command sent via MME to UE. On handover to UTRAN, MME shall include the UE capabilities in the handover request sent to RNC via SGSN. RNC shall select the UTRAN algorithm to use and indicate its choice in the handover command sent via SGSN and MME to UE.

Handover from 2G/3G to LTE:

An SGSN shall be able to ask UE for its NAS, UP, RRC security capabilities. On handover to LTE, SGSN shall include the NAS, UP, RRC security capabilities in the handover request send to MME. MME shall select the NAS algorithms to use and include the UE's UP and RRC security capabilities and its allowed RRC/UP security algorithms in the handover request send to eNB. eNB shall select the UP/RRC security algorithms and acknowledge its choice to MME. The MME then acknowledge all the NAS/RRC/UP security algorithms choice in the handover command sent to UE over SGSN.

7.4.15 Key-change-on-the-fly[39]

7.4.15.1 Serving network operator restricts the KASME lifetime

This corresponds with case three from S3-070475.

The goal of the K_{ASME} lifetime[40] restriction is to limit the effects of a K_{ASME} breach at the UE. This timer covers the IDLE as well as the ACTIVE periods. Periodic Tracking area updates will increment the NAS COUNT, so the NAS COUNT being at maximum value will trigger as well a fresh AKA run, possibly before K_{ASME} lifetime timer expiration. NAS COUNT length will determine which of the events takes place first.

On K_{ASME} lifetime timer expiration **the whole key hierarchy** SHALL be updated.

If other 'possibly frequent events' like service requests would trigger an AKA in the MME then the expiration of the K_{ASME} life time will not be a frequent event.

It may be possible that 7.4.15.6 (see below, "eNB sequence number wrap around") never occurs dependent on the settings of the timer i.e. if set to a 'low' value and dependent on the length of AS COUNT.

The usefulness of a separate K_{ASME} lifetime timer will depend on the NAS COUNT length.

The assumptions made above is that in EMM-DEREGISTERED the NAS security context (including K_{ASME} and NAS COUNT) is not available and that the K_{ASME} lifetime timer is started after taking a fresh key set into use. If the MME (and the UE) would keep the NAS security context in EMM-DEREGISTERD than the timer also need to cover this period.

7.4.15.2 Serving network operator restricts the ECM-CONNECTED lifetime

This corresponds with case two from S3-070475.

This ECM-CONNECTED timer would collect the sum of the periods that any eNB is in the possession of a key KeNB during the lifetime of the same K_{ASME}. The recorded connected time needs to be exchanged between MME's.

It may be possible that 7.4.15.6 (see below, "eNB sequence number wrap around") never occurs dependent on the settings of the 2.2 timer i.e. if set to a 'low' value and dependent on the length of HFN+SQN of AS.

It has no sense to set the ECM-CONNECTED timer is larger than the K_{ASME} lifetime timer.

Letting an eNB supervise the time of having the possession of a key, is not seen as a valuable alternative.

Therefore, for this use case **the whole key hierarchy shall** be updated if the ECM-CONNECTED time reaches a pre-determined value.

The added value of this timer seems dependent on the actual use of the eNB. Every time a UE goes from ECM-IDLE to ECM-CONNECTED, a fresh KeNB is generated (on the basis of a new NAS COUNT). In this case the K_{ASME} lifetime (or NAS COUNT) + the AS SQNs to restrict the lifetime of each individual KeNB should be sufficient. For cases where the UE says long in ECM-CONNECTED the timer could have more value.

7.4.15.3 NAS COUNT reaches maximum

This relates to previous section.

If the NAS COUNT reaches a maximum value, then before wrap-around the MME needs to trigger an AKA run to refresh K_{ASME} and subsequently the NAS keys. For this use case the whole key hierarchy SHALL be updated.

7.4.15.4 After Inter-RAT handover from UTRAN/GERAN to LTE

This corresponds with case four from S3-070475.

---

39 This section is from S31070929.

40 K_{ASME} may be time-restricted in addition to the restriction by the maximum NAS COUNT value

In order to protect EPS from a breach of the UTRAN/GERAN keys that were used for deriving EPS keys after the handover, **the whole key hierarchy** based on K_{ASME} **shall** be updated based on a new AKA run. The procedure is then as follows after the handover from the legacy 3GPP system.

1) Run AKA in the background in LTE
2) Take the new K_{ASME} into use resulting in new K_{eNB} and NAS-keys.

First the new keys SHALL be established in the eNB and in the UE. Secondly, the new keys shall be taken into use.

7.4.15.5 Intra LTE Inter-operator Handover

This event may trigger the need for AKA determined by the target MME, and thus a key change-on-the-fly due to following reasons:

a) the recorded K_{ASME} lifetime timer by the source MME transferred to the target MME is higher than the maximum value set by the target MME.
b) the recorded ECM-CONNECTED timer by the source MME and transferred to the target MME is higher than the maximum value set by the target MME.
c) the current NAS COUNT set by the source MME and transferred to the target MME is higher than the maximum value set by the target MME to trigger an AKA.

For this use case **the whole key hierarchy** shall be updated.

7.4.15.6 KeNB sequence numbers are about to wrap around.

This corresponds with case one from S3-070475.

This case would only happen if the UE stays long in active mode on the same eNB therefore this should not happen frequently, thus mainly for stationary[41] usage of the E-UTRAN.

This seems to be the only case where an independent AS key change (i.e. eNB and all derived keys) could have some advantages. The security difference between the case of involving the MME or not, is that for the first case the new eNB-keys will be cryptographically independent from the previous eNB-keys while when not involving the MME then the old and new key will not have that property. But as the eNB is the same, this does not make a big difference provided that the hash-function for key chaining and the KDF are not weak. Therefore we should rather look at potential complexity, error cases, and advantages of AS independent key change compared to the support of combined AS + NAS key change coordinated by the MME.

Before analysing the different possibilities we note that in practice it may be possible that 7.4.15.6 never occurs dependent on the timer settings of use 7.4.15.1 and 7.4.15.2 (and on AS COUNT length).

The first question that arises is whether the UE or the eNB should trigger that a wrap-around is approaching. According to TS 36.300 the handover decision is performed by the eNB, therefore it seems logical also to allocate the responsibility to supervise the AS COUNT value to the eNB and trigger an approaching wrap-around (this would enable the use of handover procedure i.e. intra-cell HO to fit in as a solution to support the AS key change).

For an AS key change without corresponding NAS key change the MME does not need to be involved and hence there is no need for S1 messages or NAS messages for informing the MME (which means that some messages in the preparation phase could be saved). It may be possible that at the same time an MME-timer (See 7.4.15.1 and 7.4.15.2) would expire, and hence this could trigger NAS + AS change as well. This case seems similar from eNB viewpoint to the case that the UE is performing an intra-MME handover at the moment that the MME wants to activate AS-keys based on a new K_{ASME}.

In case of realizing an AS + NAS combined key[42] change then the eNB needs to send an S1-message to the MME in order to trigger an AKA-run for a specific UE, such that new keys can be taken into account. AS activation time is supplied to the eNB and UE (cfr R3-072410: LS on active mode key change RAN#58).

---

41 E.g..for Home eNB usecases.

42 Alternatively it is also possible to generate a new KeNB in the MME based on NAS-COUNT but this requires S1 signaling. This provides (as well as a new Kasme) independent cryptographic eNB keys. It could save the need for an authentication run, which as well is a property of the independent AS-key change before AS COUNT-rollover.

7.4.16 Independence of keys at different eNodeBs

A discussion was conducted on whether a requirement should be imposed that the keys used by a UE at different eNodeBs (specifically the key K_{eNB}) be independent of each other or not in the case where the MME is not involved in deriving keys at handover.. Such a requirement would imply, for example, that when a handover is executed, the key K_{eNB} used by a particular UE at the target eNodeB after handover must be independent of the K_{eNB} used by the UE at the source eNodeB prior to the handover. (See S3-070252 for arguments that were advanced in favour of such a requirement.) It was concluded that such a requirement is not necessary, based on the following reasoning:

- It will be assumed that a source eNodeB in a handover securely deletes the K_{eNB} for the UE as soon as the handover of the UE to a target eNodeB is complete. Thus, any compromise of the source eNodeB *after the handover is completed* cannot reveal the key K_{eNB} that was used by the UE at the source eNodeB prior to the completion of the handover. Further, it is computationally infeasible for an attacker to obtain the K_{eNB} used by a UE at a particular eNodeB by cryptanalysis only without also physically compromising the eNodeB, since the keys that are actually used over the air are derived from K_{eNB} by a one-way function. Thus the only possible way for an attacker to obtain K_{eNB} for a particular UE is by physically compromising the eNodeB *while the UE is connected to that eNodeB.*

- If an attacker manages to compromise the UE's K_{eNB} at a particular eNodeB while the UE is still connected to that eNodeB, then UE's communications are compromised irrespective of whether the handover keying method ensures independence of keys at handover or not.

Hence there is no need to impose a requirement that the method of key derivation at handover must ensure independence of the new K_{eNB} after the handover from the old K_{eNB}.

Note that the above reasoning and decision do not imply that the keys K_{eNB} used by a UE at different eNodeBs need to be identical, only that it is not required that they be independent of each other. There may be other reasons for changing the K_{eNB} used by a UE at handover, for example to simplify the initialization of various counter settings at handover etc.

It was noted that measures need to be put in place to detect a physical compromise of the eNodeB, and react by taking the appropriate steps to limit the consequences of the compromise.

7.5 START value transfer

7.5.1 Why does START value have to transfer from UE to CN

In UMTS, START value is defined to record the amount of data that is protected by a set of cipher/integrity keys. When START value reaches maximum value, THRESHOLD, the cipher/integrity keys shall be updated.

When a radio connection is set up, this START value should be transferred from UE to RNC to initialize COUNT-C and COUNT-I, which is as an input parameter of f8 and f9 respectively in ME and RNC. Upon radio connection release and when a set of cipher/integrity keys is no longer used, the ME updates STARTCS and STARTPS in the USIM with the current values.

But in LTE/SAE system, Integrity/Encryption of AS signaling function is located in eNB, Encryption function of user data is located in UPE, and Integrity/Encryption function of NAS signaling was located in MME. Each node need a COUNT as an input parameter of Encryption/Integrity function (in this section, encryption and Integrity function is assumed to share one COUNT). And the three COUNT can be called COUNTas, COUNTnas and COUNTup.

So in order to initialize COUNT in these nodes, START value should be transfer from UE to eNB, MME and UPE respectively.

7.5.2 How does START transfer from UE to CN

In this subclause, two alternatives of solution to pass START value from UE to CN will be given:

- A).Alternative 1: transmit START value to CN using initial L3 message

Because START value is used to initialize COUNT, which is an input parameter of Encryption and Integrity function, START value should be transferred to CN as soon as possible. Initial L3 message is the first message from UE to CN, so it seems reasonable that transmit START in initial L3 message.

- B).Alternative 2: transmit START value to CN using RRC and RANAP message

When MME, UPE, and eNB use the same START to initialize their COUNT value, it is suggested to transfer of START through RRC and RANAP message directly. During Radio Resource Connection establishment message, START will be sent to eNB. eNB will save a copy of this START to initialize COUNT in the eNB, then the eNB forwards this START to CN in eRANAP message. MME save a copy of the START and then forwards the START to UPE.

7.5.3 How many START value should be used

There are two possibilities for the number of START value.

1) only one START value is defined in UE

In this solution, START can record the maximum of COUNT values of the three COUNTas, COUNTnas, COUNTup for all AS signalling protected using CKas/ IKas, NAS signalling protected using CKnas/ IKnas, and user data protected using CKupe. When a session is set up, the START is used to initialize the three COUNTas,COUNTnas, and COUNTup.

In this solution, START value can not record the exact amount of data being protected by each key pair (Key for AS protection/or Key for NAS protection /or Key for UPE protection). And when the START reaches its THRESHOLD, all these Keys should be updated.

2) Three START Values are defined in UE

In this section, the three START value can be called $START_{AS}$, $START_{NAS}$ and $START_{UP}$. And each can be used to record corresponding COUNT value. i.e., $START_{AS}$ record COUNTas value.

In this solution, each of the three START values can record the exact amount of data being protected by each key pair. And eNB/MME/UPE can initialize its COUNT according each START, and each key can be update independently.

7.6 Security algorithms

7.6.1 Choice of algorithms

It has already been agreed that UEA2 and UIA2 is to be supported. Due to lessons learnt it is also agreed to have a second “back up” option, or complement to UEA2/UIA2. Natural candidates are UEA1/UIA1. However, the general opinion in SA3 seems to favour an “AES based” alternative. Various arguments can be given in favour of both of these main “tracks”.

Security: The most important aspect is of course security. The complement to UEA2/UIA2 should be “sufficiently different” so that a compromise of UEA2/UIA2 should not automatically mean a compromise also of the other algorithm. UEA2 shares some components with AES (the S-box), so advances in AES analysis could affect also UEA2. On the other hand, UEA1 and AES are both block cipher based so a new, general attack method on block ciphers might affect both. Linear and differential cryptanalysis are examples of attacks which in the past affected several block ciphers. Though new “breakthroughs” of similar impact may not be too likely, they can on the other hand not be discarded as “impossible”.

Performance: it is felt that this does not speak significantly for or against either of the choices as many trade-offs regarding optimizations are available for both UEA1 and AES.

Implementation: It can be noted that for UEA1/UIA1, a great deal effort spent on getting good implementations of these for UMTS which would be cost effective to re-use. On the other hand, implementation aspects of AES were quite deeply investigated already during NIST’s AES effort which is available to the public. Terminals capable of both UMTS and LTE would benefit from having to only implement UEA1/UIA1 and UEA2/UIA2. Pure LTE terminals would not benefit in the same way.

Specification work: If an “AES based” option is chosen, more specification work is needed than if UEA1/UIA1 is chosen. A mode of operation for AES needs to be specified, but this is probably not a major issue. A (slightly) more involved issue is to specify how to use AES for integrity protection purpose (assuming SA3 does not opt for a dedicated MAC such as HMAC). One option would be to run AES in “f9” mode as done with Kasumi in UMTS. This arguably be considered “different” (from UIA2) unless significant difference to Kasumi in f9 mode is also wanted at the same time.

7.6.2 Terminal support

A question is whether both or only one algorithm should be mandatory to implement in the ME. To minimize risk of incompatibility between ME and network, it is felt that the usual approach to mandate the ME to support all algorithms seems wise.

7.6.3 Network support

There is today a large problem in some GSM networks that only support the broken A5/2 algorithm. It is therefore felt that SA3 should work towards mandating that all networks (at least the eNodeB:s) support both UEA2/UIA2 and the alternative algorithm (whichever is chosen). In the future, new algorithms may be added to the networks with optional support, but from day one, LTE networks should implement both of the initial two algorithms.

7.6.4 Algorithm input

7.6.4.1 Input parameters to RRC signalling ciphering algorithm

All input parameters to ciphering algorithm in UMTS are needed in SAE/LTE for the same use. So the input parameters to RRC signalling ciphering algorithm shall include ***COUNT$_{RRC}$***, DIRECTION, LENGTH, K_{RRCenc}, Bear ID.

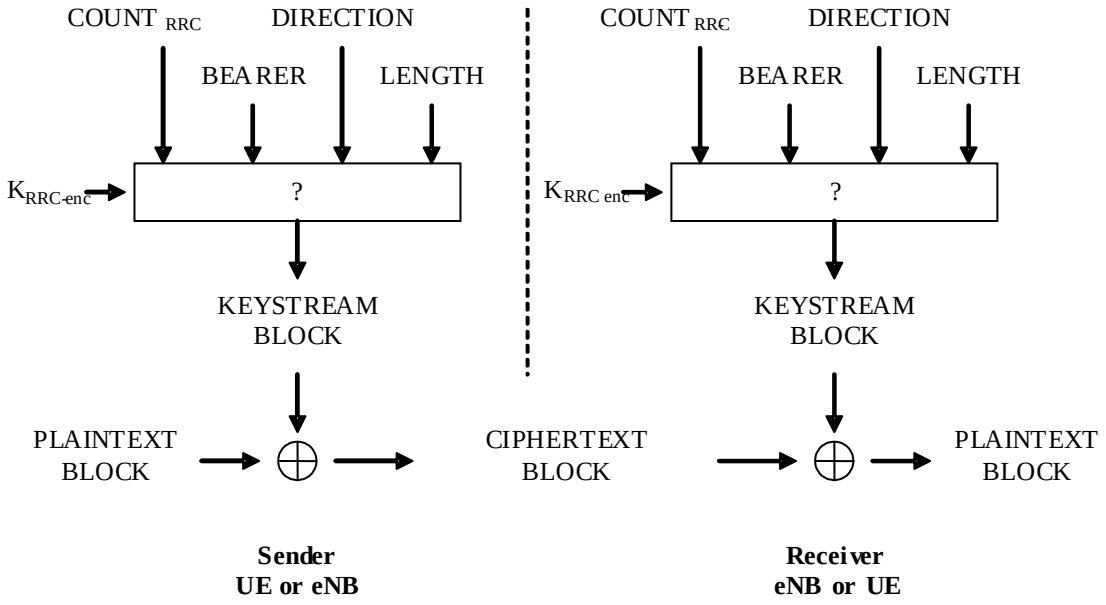


Figure 29 Input parameters to RRC signalling ciphering algorithm

COUNT$_{RRC}$

COUNT$_{RRC}$ is the ciphering sequence number. It can be composed of two parts: PDCP SN that is available in each PDCP PDU, and PDCP hyper frame number (PDCP HFN) which is incremented at each PDCP SN cycle.

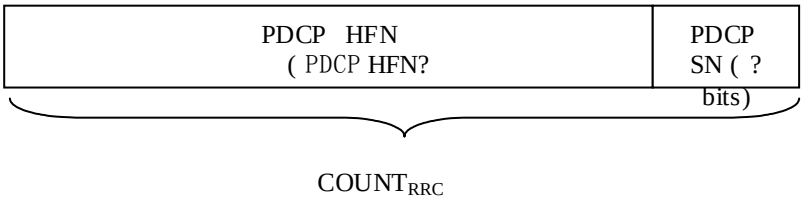


Figure 30 The structure of COUNT$_{RRC}$

Both the PDCP HFN and PDCP SN are initialised to 0.

Editor's note: Whether PDCP SN can be initialized to 0 needs to be confirmed with RAN2.

Editor's note: length of PDCP HFN and PDCP SN are ffs.

DIRECTION: See 6.5.4.4 of TS 33.102.

K_{RRCenc}: K_{RRCenc} is the key derived from K_{eNB}.

LENGTH: See 6.5.4.4 of TS 33.102.

BEARER: The radio bearer identifier BEARER is 5 bits long.

Editor's note: Whether the radio bearer identifier BEARER can be 5 bits long needs to be confirmed with RAN2.

There is one BEARER parameter per radio bearer associated with the same user. The radio bearer identifier is input to avoid that for different keystream an identical set of input parameter values is used.

7.6.4.2 Input parameters to NAS signalling ciphering algorithm

The input parameters to RRC signalling ciphering algorithm shall include ***$COUNT_{NAS}$***, DIRECTION, LENGTH, K_{NASenc}, Bear ID.

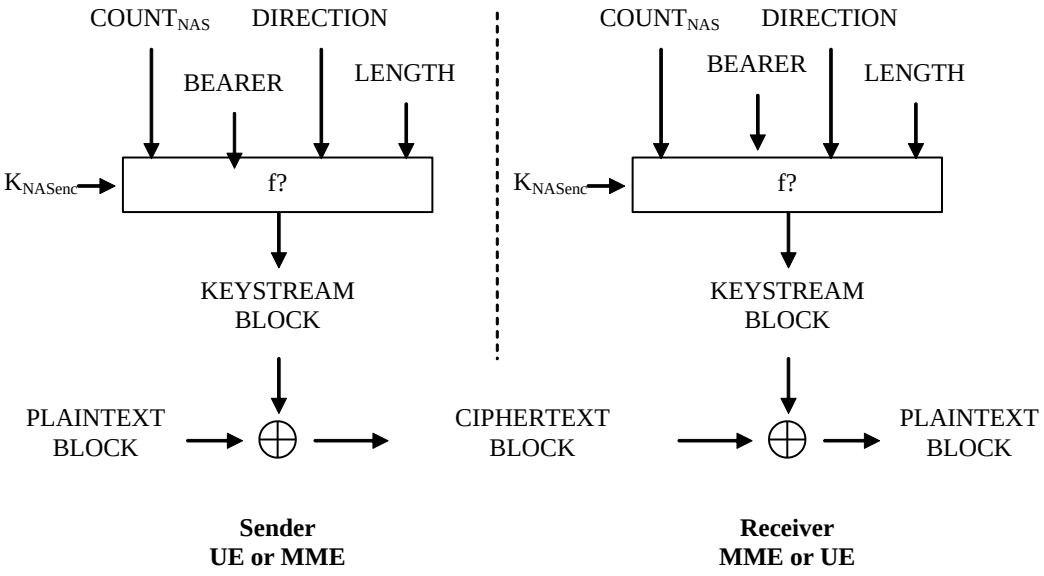


Figure 31 Input parameters to NAS signalling ciphering algorithm

$COUNT_{NAS}$

$COUNT_{NAS}$ is the NAS ciphering sequence number.

For NAS signalling there is one $COUNT_{NAS}$ value per up-link NAS signalling bearer and one $COUNT_{NAS}$ value per down-link NAS signalling bearer.

$COUNT_{NAS}$ is composed of two parts: a "short" sequence number and a "long" sequence number. The "short" sequence number forms the least significant bits of $COUNT_{NAS,}$ while the "long" sequence number forms the most significant bits of $COUNT_{NAS}$. The "short" sequence number is the NAS sequence number (NAS SN) that is available in each NAS PDU. The "long" sequence number is the NAS hyper frame number (NAS HFN) which is incremented at each NAS SN cycle.

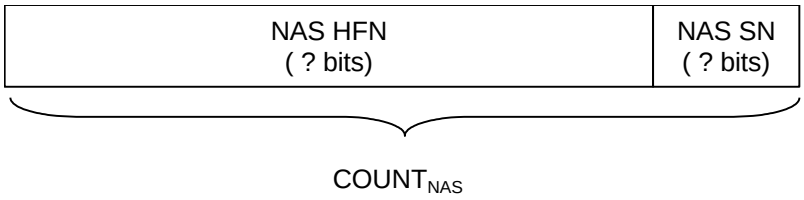


Figure 32 The structure of $COUNT_{NAS}$

Editor's note: length of NAS HFN and NAS SN are ffs.

DIRECTION: See section 7.6.4.1.

K_{NASenc}: K_{NASenc} is the key used to cipher NAS signaling, it is derived from K_{ASME} in MME and UE respectively.

LENGTH:See section 7.6.4.1.

BEARER: Since there is only one NAS signalling bear for one user, the BEARER can be set to a default value.

7.6.4.3 Input parameters to UP ciphering algorithm

As both RRC signalling and UP data are ciphered PDCP layer. The parameters to UP ciphering algorithm can be the same as that to RRC ciphering algorithm except that ciphering key shall be different.

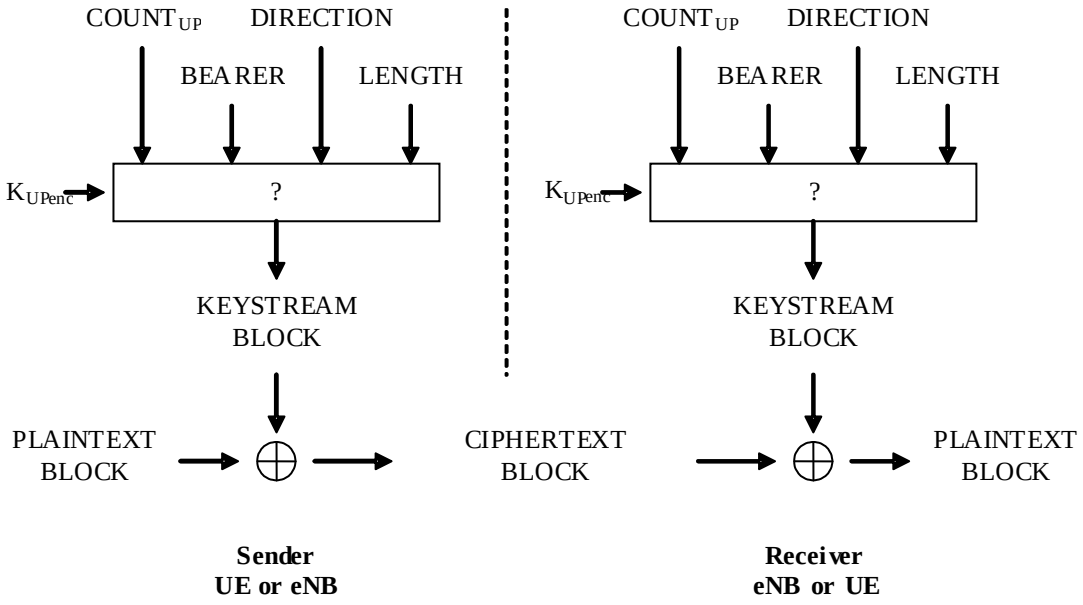


Figure 33 Input parameters to UP ciphering algorithm

$COUNT_{UP}$

$COUNT_{UP}$ is the data ciphering sequence number.

$COUNT_{UP}$ can be composed by two parts: PDCP SN that is available in each PDCP PDU, and PDCP hyper frame number (PDCP HFN) which is incremented at each PDCP SN cycle.

At begin of each session, $COUNT_{UP}$ is initialised to 0.

Editor's Note: The value to which the initialization is done needs to be confirmed with RAN2.

DIRECTION: See section 7.6.4.1.

K_{UPenc}: K_{UPenc} is the key used to cipher user data. It is derived from K_{eNB} in eNB and UE respectively.

LENGTH: See section 7.6.4.1.

7.6.4.4 Input parameters to RRC signalling Integrity algorithm

The input parameters to RRC signalling Integrity algorithm shall include ***COUNT***$_{RRC}$, DIRECTION, K_{RRCint}, BEARER,.

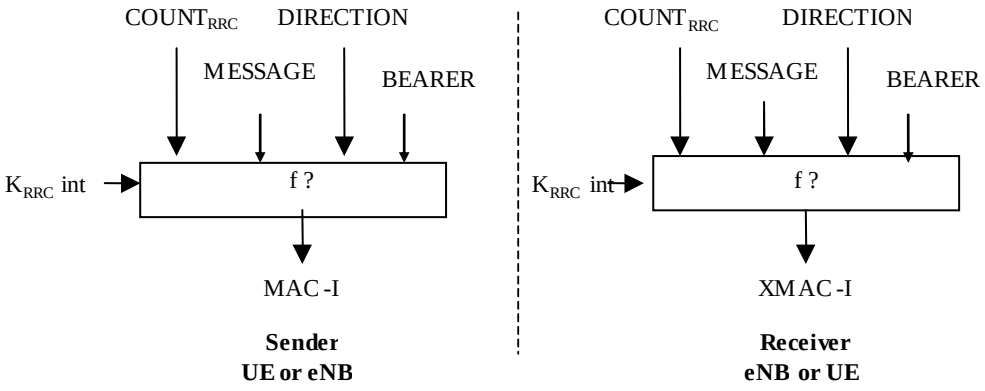


Figure 34 Input parameters to RRC signalling integrity algorithm

$COUNT_{RRC}$: See section 7.6.4.1.

DIRECTION: See section 7.6.4.1.

K_{RRCint}: K_{RRCint} is the integrity key derived from KeNB.

BEARER: See section 7.6.4.1.

7.6.4.5 Input parameters to NAS signalling Integrity algorithm

The input parameters to NAS signalling Integrity algorithm shall include ***COUNT***$_{NAS}$, DIRECTION, K_{NASint}.

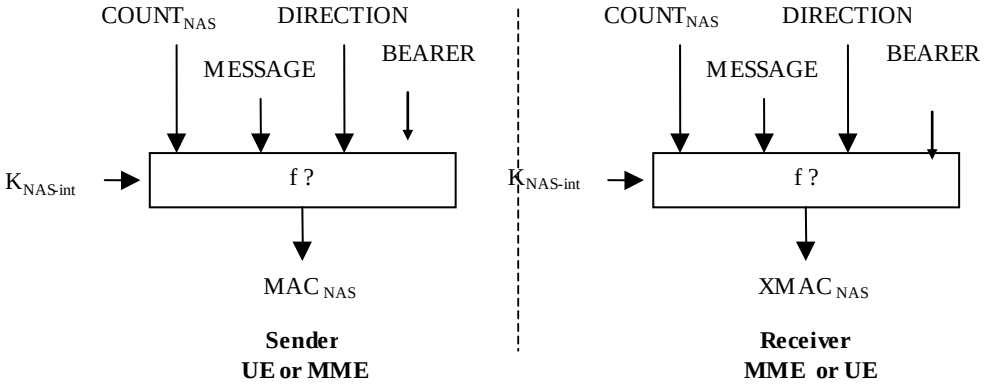


Figure 35 Input parameters to NAS signalling integrity algorithm

$COUNT_{NAS}$: See section 7.6.4.2.

DIRECTION: See section 7.6.4.1.

K_{NASint}: K_{NASint} is the key used to integrity protection to NAS signaling, it is derived from K_{ASME} in MME and UE respectively.

BEARER: Since there is only one NAS signalling bear for one user, the BEARER can be set to a default value.

7.6.5 Algorithm IDs in EPS[43]

The usage of algorithm IDs is:

1. Identify the algorithms, e.g. when UE performs HO between 2 eNBs or between 2 MMEs, or UE perform TAU between 2 MMEs, if old algorithm can be supported, then the same algorithm may be chosen.
2. As an input parameter to NAS keys if NAS algorithms change.
   But for AS keys, it will change anyway when HO, so maybe algorithm ID is no need as input parameter of new AS keys derivation.

Although the algorithms have not been defined yet, the identifiers are needed in some message and needed as input parameter to the key derivations.

Since the algorithms will be based on SNOW 3G and AES, 4 bits (same as for UMTS) should be enough for algorithm identifiers. We are trying to give the following values for the EPS algorithm identifiers.

These algorithm IDs indicates the key length is 128bits.

Encryption Algorithm IDs for EPS:

- EPS AS encryption algorithm:
  - "0000_2" EA EA0 Null algorithm
  - "0001_2" EA EA1 SNOW 3G
  - "0010_2" EA EA2 AES
- EPS NAS encryption algorithm
  - "0000_2" ENEA0 Null algorithm
  - "0001_2" ENEA1 SNOW 3G
  - "0010_2" ENEA2 AES
- EPS UP encryption algorithm
  - "0000_2" EUEA0 no encryption
  - "0001_2" EUEA1 SNOW 3G
  - "0010_2" EUEA2 AES

Integrity Algorithm IDs for EPS:

- EPS AS integrity algorithm
  - "0001_2" EAIA1 SNOW 3G
  - "0010_2" EAIA2 AES
- EPS NAS integrity algorithm
  - "0001_2" ENIA1 SNOW 3G
  - "0010_2" ENIA2 AES

Editor's Note: The modes of operation of Snow 3G and AES need FFS.

---

43 This section is from S3a070918.

7.6.6 KDF negotiation[44]

7.6.6.1 Overview on the use of KDF functions for EPS

KDF's are specified in TS 33.401 to support key derivations in HSS, MME and eNB.

NOTE: The use of KDF functions for EPS can be seen from Annex A of the TS 33.401. This section provides a categorization and overview of the use of the KDFs.

K_{ASME} is derived from CK, IK and PLMN ID in HSS. The keys derived from K_{ASME} in MME include NAS keys, K_{eNB} and NH for eNB, and NAS-token for IDLE mode change to an SGSN. During intersystem Handover keys are derived from and to K_{ASME} (mapping of security context or refreshing of security context) in MME. UP and RRC keys and key chaining during X2-Handover are derived in eNB.

There are three types of KDF usages:

(a) Derivation of intermediate keys (e.g. $K_{eNB,}$ NH, K_{ASME}) i.e. keys not used with protocols seen on the air interface.

(b) Derivation of keys used with a specific algorithm (e.g. K_{upenc} with EEA1) and used with protocols seen on the air interface.

(c) Derivation of one-time values like NAS-TOKEN.

The use of KDFs is there on one node extra (HSS) than implied by (b) i.e. eNB and MME. KDF usage is possibly performed in other functions of the node than the ciphering/integrity engines which are possibly implemented in hardware But when a KDF is realized outside these engines, the addition of a new KDF may be easier performed by software upgrades.

7.6.6.2 Effects on the security of overview of the use of KDF functions

7.6.6.2.1 Can a KDF be broken ?

There are two types of attacks on hash functions and the effects on KDFs according to [10] are:

- **Attacks against the "one-way" property**: A "first-preimage attack" allows an attacker who knows a desired hash value to find a message that results in that value in fewer than 2^L attempts. A "second-preimage attack" allows an attacker who has a desired message M1 to find another message M2 that has the same hash value in fewer than 2^L attempts.
- **Attacks against the "collision-free" property**: Attacks against that have to show that two messages M1 and M2 can be found to have the same value in fewer than 2^ (L/2) where L is the hash length.

Hash attacks concentrate mostly on collision free property attacks (e.g. MD5 and SHA-1). Most protocols that use hash algorithms do so in a way that makes them immune to harm from collision attacks. However, the KDFs used for EPS only require the one-way property. Attacks finding collisions as such pose no risk to the use of KDFs in EPS. So, the recent attacks on MD5 and SHA-1, even if they were extensible to SHA-256, which is used in EPS, would not constitute a breach of security of EPS.

In most of the cases where the KDF is used in EPS, both input and output values are secret to outsiders if we restrict the attacks to real outsiders i.e. the air interface attacker. For the K_{ASME} derivation one could consider the MME as the outsider too, and similarly for K_{eNB} derivation in the MME, one could consider the eNB as the outsider.

Considering the 3 types of KDF usage's (cfr clauses 7.6.6.1) in derivation of intermediate keys, derivation of keys used with a specific algorithm and derivation of one-time values, and considering further that the likelihood of a compromise of an MME has to be assumed to be much lower that that of a compromise of an eNB, the risk of finding CK, IK through a broken KDF used to derive K_{ASME} is reduced as a compromised eNB would have to successfully invert the KDF twice. Furthermore, EPS would be compromised only if big weaknesses with the KDFs are found i.e. a compromised eNB being able to reverse engineer in a practical way the input (KDF), the K_{ASME} could be reverse engineered by an eNB from $K_{eNB,}$ or the CK or IK could be found back by an MME from the K_{ASME}. But such way of

---

44 This section is from S3-081256.

going back from an arbitrary output to the input value of a hash algorithm would require a disastrous break of that algorithm, which is considered very unlikely.

7.6.6.2.2 What is the impact if the KDF is broken?

Breaking a KDF means that it would be possible to reconstitute its input from its output. Currently there is no such attack in sight for the KDF currently favored for EPS: SHA-256.

EPS KDFs are used in three different network nodes: HSS, MME, eNB.

eNB-KDF:

(1) Backward security in case of horizontal handovers[45] would be lost

(2) Key separation in case of preparing multiple eNBs would be lost

(3) Key separation of KeNB to RRCint, RRCenc and UPenc would be lost

Possible network based countermeasures in case of a break:

(1) can be solved by resorting to vertical[46] handovers whenever backward security is required

(2) can be resolved by grouping the eNBs into security domains and only allowing preparation of multiple eNBs of the same security domain,

(3) is irrelevant, as the same entity eNB knows KeNB and all derived keys.

MME-KDF:

(4) Separation of K_{eNB} and NH from K_{ASME} and between NH and K_{eNB} would be lost i.e. there would be no forward and backward security anymore.

The impact of this would be that there would be the potential for someone with access to keying material of one eNB to gain access to the keying material for all sessions of the UE ever attached to this eNB.

Possible network based countermeasures in case of a break:

- None known, note that the threat is the compromised eNB and not the outside attacker.

HSS-KDF:

(5) separation of CK, IK from K_{ASME} would be lost

(6) Separation of K_{ASME} between different visited networks would be lost

The impact of (5) would be low, as the long term secret K shared between AuC and USIM is not endangered.

Possible network based countermeasures in case of a break:

- note that the threat is the compromised MME and not the outside attacker.
- The impact of the (6) can be mitigated by enforcing a new AKA after handover between visited networks of different security domains.

7.6.6.3 Possible solutions for KDF negotiation and their requirements

For eNB-KDF, MME-KDF, HSS-KDF negotiation, the following two solutions may be used:

1. The KDF negotiation is not needed, i.e. one KDF is specified and used in all network entities. This solution is simple. This has been decided for Rel-8.

---

45 Handovers between eNBs without using new keying material from the MME

46 Handovers involving the MME

2. The KDF negotiation is needed. There are several KDF negotiation solutions possible:

   (A) One common KDF is negotiated by MME and UE;

   (B) eNB-KDF, MME-KDF, HSS-KDF are negotiated by MME and UE;

   (C) eNB, MME and HSS respectively negotiates an appropriate KDF with UE;

   (D) eNB-KDF and MME-KDF are negotiated by MME and UE, HSS-KDF is negotiated by HSS and UE.

   (E) KDF negotiation between UE and eNB & MME could be implicit

   In the future, UE and network devices may be upgraded, and they may support a newer and securer KDF algorithm. If the KDF negotiation is used, the upgraded network devices and UE can negotiate to use the securer KDF. In this way, the KDF can be flexibly updated.

The following clauses sketches the impact of the above KDF negotiation solutions to the network.

7.6.6.3.1 (A) One common KDF is negotiated by MME and UE

For this solution, UE, eNB and HSS need to inform their supported KDFs to MME.

The supporting UE's KDFs can be included in UE Security capabilities and informed to the MME by mean of a NAS message. The supporting eNB's KDFs could be included on S1 to the forwarded attach request or TAU request towards the MME. The supporting HSS's KDFs can be included in authentication data response and informed to the MME, but it will need extension to the authentication data response message. Alternatives are solutions with O&M pre-configurations e.g. on S1 interface but this might be unmanageable between HSS and MME.

After MME knows the UE's KDFs, eNB's KDFs and HSS's KDFs, MME selects one common KDF supported by UE, eNB, MME and HSS and informs the selected KDF to UE, eNB and HSS. Selected KDF can be informed to UE, eNB and HSS via NAS downlink message, S1 and S6a interface respectively. Here it need modify the S1 and S6a interface.

Only when UE, eNB, MME and HSS are upgraded synchronously, a newer and securer KDF that it supported by all of them can be negotiated by MME and UE. If any one of four entities is not upgraded to support a newer and securer KDF, this KDF can not be used.

7.6.6.3.2 (B) eNB-KDF, MME-KDF, HSS-KDF are negotiated by MME and UE

For this solution, UE, eNB and HSS will also inform their supported KDFs to MME. MME will select the appropriate eNB-KDF according to the KDFs supported by UE and eNB, MME will also select the MME-KDF and HSS-KDF according to the KDFs supported by UE and MME, KDFs supported by UE and HSS respectively. These selected KDFs may be different or same, i.e. the selection procedure of three KDFs is independent. And then, MME will inform the selected KDF to UE, eNB and HSS. This solution will modify or extend some interfaces' protocols or messages in the same way as the first solution.

When UE and eNB/MME/HSS are upgraded synchronously, a newer and securer eNB-KDF/MME-KDF/HSS-KDF will be negotiated by MME and UE.

7.6.6.3.3 (C) eNB, MME and HSS respectively negotiates an appropriate KDF with UE

For this solution, UE will inform it's supported KDFs to eNB, MME and HSS.

7.6.6.3.3.1 eNB-KDF and MME-KDF negotiation

The supporting UE's KDFs can be included in UE Security capabilities and informed MME by NAS message. MME will inform the UE's KDFs to eNB via S1 interface.

After eNB/MME knows the UE's KDFs, eNB/MME will select an eNB-KDF/MME-KDF and inform the KDF to UE. eNB-KDF/MME-KDF can be informed to UE via AS/NAS SMC procedure.

7.6.6.3.3.2 HSS-KDF negotiation Alternative 1

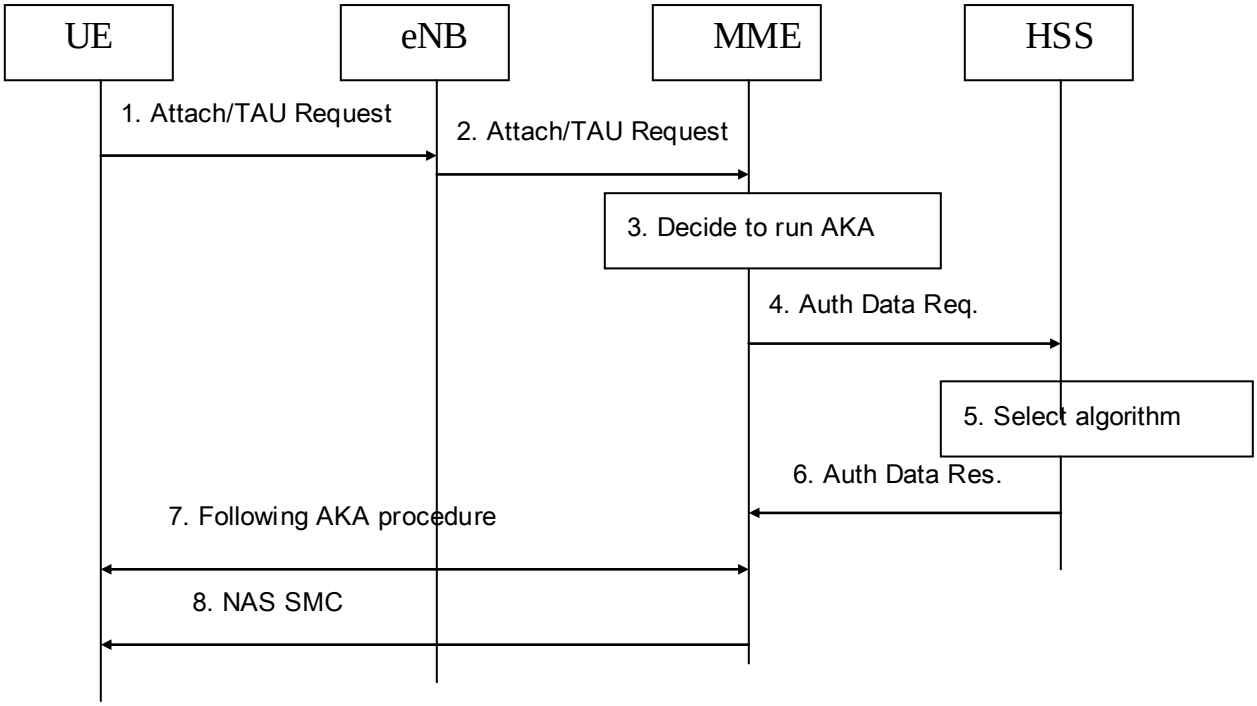


Figure 36 The KDF negotiation

1. UE sends Attach/TAU Request message to eNB, the UE network capability which included in Attach Request/TAU message will include the KDFs of K_{ASME} supported by UE.

2. MME receives the Attach/TAU forwarded by eNB.

3. MME decides to run an AKA.

4. MME sends Authentication Data Request message to HSS, the KDFs of K_{ASME} supported by UE is included in Authentication Data Request message.

5. HSS receives the Authentication Data Request and achieves the KDFs of K_{ASME} supported by UE, and then HSS selects one KDF used to derive K_{ASME} according to the local policy.

6. HSS sends Authentication Data Response message to MME, the selected KDF of K_{ASME} is also included in this message.

7. The following AKA procedure is performed successfully.

8. MME sends NAS Security Mode Command to UE. This message includes K_{ASME}, UE security capability, ENEA, ENIA, NAS-MAC, and the selected KDF of K_{ASME} by HSS is also included in this message.

9. When UE receives NAS Security Mode Command, the selected KDF of K_{ASME} by HSS is known by UE, i.e. the KDF of K_{ASME} between UE and HSS is negotiated successfully.

7.6.6.3.3.3 HSS-KDF negotiation Alternative 2

The supporting UE's KDFs can be informed to HSS by MME via Authentication Information Request.

After HSS knows the UE's KDFs, HSS will select a HSS-KDF and inform the KDF to UE. The KDF can be informed to UE via Authentication Data Response.

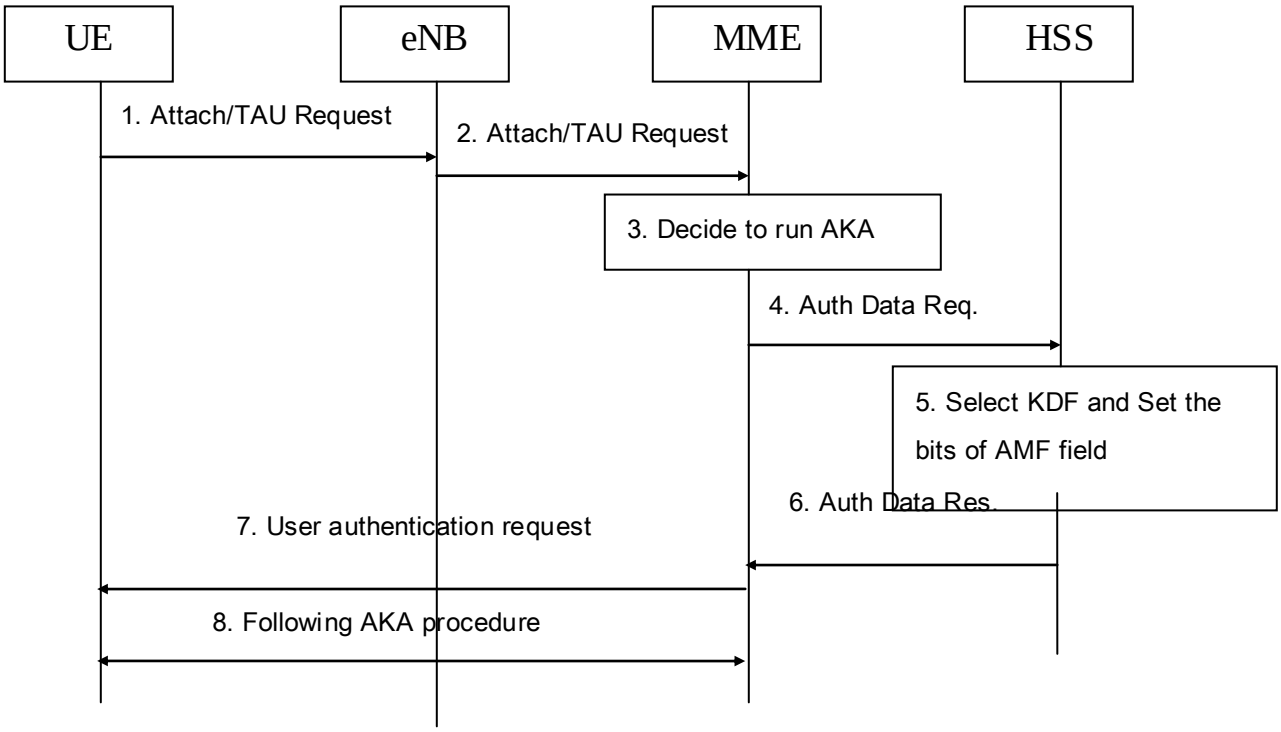


Figure 37 The HSS-KDF negotiation

1. UE sends Attach/TAU Request message to eNB, the UE network capability included in Attach Request/TAU message will include the HSS-KDFs supported by UE.

2. MME receives the Attach/TAU forwarded by eNB.

3. MME decides to run an AKA.

4. MME sends Authentication Data Request message to HSS, the HSS-KDFs supported by UE is included in Authentication Data Request message.

5. HSS receives the Authentication Data Request and achieves the HSS-KDFs supported by UE, and then HSS selects one KDF used to derive Kasme according to the local policy and sets certain bits in the AMF field of AUTN to indicate the selected KDF of K_{ASME}.

6. HSS sends Authentication Data Response message to MME.

7. UE receives User authentication request which included RAND and AUTN, checks the certain bits and knows the selected HSS-KDFs by HSS, i.e. the HSS-KDF between UE and HSS is negotiated successfully.

8. The following AKA procedure is performed successfully.

It is possible to reduce the impact for uplink message in S6a. The following solution is a possible way:

An IE which contained in Authentication Information Request may be extended to indicate the KDFs supported by UE, For example, four bits are extended in IE-Requesting Node Type which contained in Authentication Information Request and indicated the KDFs supported by UE.

7.6.6.3.4 (D) eNB-KDF and MME-KDF are negotiated by MME and UE, HSS-KDF is negotiated by HSS and UE

For this solution, eNB-KDF and MME-KDF negotiation are as same as eNB-KDF and MME-KDF negotiation in the clause 7.6.6.3.2, HSS-KDF negotiation is as same as HSS-KDF in the clause 7.6.6.3.3.

7.6.6.3.5 (E) KDF negotiation between UE and eNB & MME could be implicit

Currently we have only one KDF specified in EPS (For MME, UE, eNB, HSS), an explicit negotiation mechanism does not add anything now.

A straightforward realization of an explicit KDF negotiation is to include KDF-ids in the UE security capabilities and negotiate them also with the ciphering and integrity protection algorithms for E-UTRAN in the same way as the algorithm selection. In this case there are impacts e.g. for indication of selected KDF and configuration of KDF list at eNB. But as we currently have only one single KDF and that a disastrous compromise is very unlikely the effort can be avoided in Rel-8. One way of doing this is for the negotiation to be implicit (no impacts) too, i.e. with EEA1 and EEA2 always SHA-256 would be used. On introduction of a new air inter face algorithm e.g. EEA3 a new implicit KDF could be added or not. This would be based on the assumption that a KDF lifetime is longer than the algorithm lifetime, and the fact that MME's and K_{eNB} have to be prepared for such an upgrade anyhow.

7.6.6.4 Attacks on KDF negotiation solutions and requirements for secure solutions.

7.6.6.4.1 Requirements and resistance to bidding down attacks.

The requirements on a secure solution for KDF negotiation is not different from a secure algorithm negotiation solution i.e. care needs to be taken that a bidding down attack can be prevented. As there is already a solutions for secure algorithm negotiation in EPS, which extends between the MME/eNB and UE, the analysis and attack scenarios below focuses only on the KDF negotiation between the UE and the HSS.

The attack scenarios:

1. If the MME could be assumed to be not compromised then hop-by-hop integrity protection between UE and MME and between MME and HSS of the negotiation of the KDF used between UE and HSS would be sufficient. But then K_{ASME} would not be disclosed from the MME, and CK, IK could not be re-engineered from K_{ASME} even if the KDF was weak, as K_{ASME} was not known to the attacker.

2. If the possibility of a compromise of the MME had to be assumed then end-to-end integrity protection between UE and HSS of the KDF negotiation would be needed.

7.6.6.4.2 Resistance to bidding down attacks for HSS-KDF negotiation solutions

Solution in section 7.6.6.3.3.2 (Alternative 1) proposes to have KDF selection integrated in authentication signalling. The selected KDF-ID is inserted in the AV-response Signalling. For the AuthInfoRequest Signalling, UE KDF capabilities need to be sent towards the HSS.

A change to the MAP or DIAMETER protocol between MME and HSS would be required. In order to prevent bidding down (e.g. from the serving PLMN) on the HSS and UE KDF capabilities the HSS would need to integrity protect the HSS capabilities and the received KDF UE capabilities in the AV–response. For this purpose, an integrity algorithm shared between UE and HSS would be required. But there are problems with this approach: How can we ensure that this integrity algorithm is safe from being compromised in the future? Would we need to negotiate this integrity algorithm used for KDF negotiation protection as well? How would this negotiation be protected? Please note that breaking this integrity algorithm could be done offline as a compromised MME/attacker is able to pre-fetch EPS-AVs.

The solution in section 7.6.6.3.3.3 (alternative 2) proposes to use AMFs bit to indicate selected HSS-KDF, which would be transparent for the AV-response Signalling, but not for the AuthInfoRequest Signalling because the UE KDF capabilities needs to be sent towards the HSS. It will eat up more AMF bits if the number of KDFs increases. Sending KDF capabilities in the AV-Req could be omitted after initial negotiation but this creates a state with the selected KDF in the HSS after initial user registration.

Also, the HSS's choice may depend on the UE capabilities. So, in order to prevent bidding down (e.g. from the serving PLMN) on the UE KDF capabilities the HSS would need to integrity protect the received KDF UE capabilities in the AV–response. This eats up again AMF-bits or requires extra parameters during authentication, leading to additional requirements as with alternative 1 above. At any rate, the UE capabilities would have to be sent to the HSS, so a change to the MAP or DIAMETER protocol between MME and HSS would be required.

7.6.6.5 Summary and decision made for Rel-8

For Rel-8, meeting **SA3#52bis has decided not to include means to negotiate KDF's**. Rel-8 only provides one KDF function (as specified in Annex A of TS 33.401) which is used for various purposes and between different network nodes. KDF negotiation may however be introduced in a later release as soon as a second KDF needs to be introduced.

7.7 Rationale for approach to security handling in inter-RAT mobility procedures[47]

During SA3#52bis it was agreed the handling of the freshness parameters and the messages in the TAU procedures differed in the following cases:

A) Idle mode mobility from UTRAN to E-UTRAN using mapped context

B) Idle mode mobility from UTRAN to E-UTRAN using cached context

C) Handover from UTRAN to E-UTRAN using mapped context

and the following case

D) TAU after handover from UTRAN to E-UTRAN using cached context

does not exist.

7.7.1 Idle mode mobility from utran to e-utran using mapped context

Agreements at SA3#52bis on idle mode mobility from utran to e-utran using mapped context (cfr. clause 9.1.2 in TS 33.401, S3-081234, and S3-081217) are:

a) TAU request is not integrity-protected

b) $Nonce_{UE}$ is included in TAU request

c) MME includes $Nonce_{UE}$ and $Nonce_{MME}$ in SM Command sent after receiving TAU request and before sending TAU accept.

d) K_{ASME} is refreshed based on $Nonce_{UE}$ and $Nonce_{MME}$.

The justifications of these agreements are:

a) TAU request is not integrity-protected because any K_{ASME} the UE could compute from CK, IK used in UTRAN at the point in time of sending the TAU request would not be guaranteed to be fresh. This is so because the CK, IK could have been cached in the SGSN, and the UE could have switched back and forth between UTRAN and E-UTRAN, resulting in always the same mapped K_{ASME} if no freshness parameters were used. But such freshness parameters are not available yet when the UE sends the TAU request, so the MME cannot not know whether the TAU request is a replay from a previous TAU procedure run. But the protection with a NAS key derived from a mapped K_{ASME} not guaranteed to be fresh would give only a marginal security gain, therefore it is better to leave the TAU request unprotected.

This lack of protection of the TAU request carries the risk that an attacker registers a user for a tracking area, in which the user is actually not present. This makes the user unreachable. This threat is mitigated by the use of the Security Mode Command procedure, which according to S3-081217mandatorily follows the unprotected TAU request, and is executed before the TAU accept. In this SMC procedure, the fresh K_{ASME} derived using $Nonce_{UE}$ and $Nonce_{MME}$, is established between UE and MME. The UE proves to be present in the tracking are by correctly responding with a protected Security Mode Complete message protected with a NAS key derived from this fresh K_{ASME}.

b) $Nonce_{UE}$ is used as one input to compute a fresh K_{ASME} from CK, IK. Its inclusion guarantees freshness of K_{ASME} to the UE. There is no other source of freshness for the UE.

---

[47] This section is from S3-081387.

c) $Nonce_{MME}$ is used as the second input to compute a fresh K_{ASME} from CK, IK. Its inclusion guarantees freshness of K_{ASME} to the MME.

d) The rationale for refreshing K_{ASME} based on $Nonce_{UE}$ and $Nonce_{MME}$ is given above.

7.7.2 Idle mode mobility from utran to e-utran using cached context

Agreements at SA3#52bis on Idle mode mobility from utran to e-utran using cached context (cfr. clause 9.1.2 in TS 33.401 and S3-081234) are:

a) TAU request is integrity-protected and not ciphered. As the MME cannot know the difference between this TAU request and other TAU requests the TAU request shall never be ciphered. This is no security risk as the use of temporary identities is assumed.

b) $Nonce_{UE}$ is included in TAU request

c) No $Nonce_{MME}$ is used

d) K_{ASME} is not refreshed

The justifications of these agreements are:

a) TAU request is integrity-protected: this is possible due to the availability of cached context and desirable to avoid DoS attacks against the user

Bb) No use is made of $Nonce_{UE}$ when cached context is used as there is no need to refresh K_{ASME}.

The reason why $Nonce_{UE}$ is included in TAU request nevertheless is that the UE cannot know whether the MME still has the cached context. If this is the case the MME discards $Nonce_{UE}$. If this is not the case the MME proceeds with using mapped context, and the $Nonce_{UE}$ is needed.

In case the MME does not have cached context, it runs NAS SMC procedure as above.

Bc) No $Nonce_{MME}$ is used as there is no need to refresh K_{ASME}.

Bd) K_{ASME} need not be refreshed as the replay protection is provided by the use of NAS COUNT which is stored as part of the EPS security context together with K_{ASME}.

7.7.3 Handover from utran to e-utran using mapped context

Agreements at SA3#52bis on handover from utran to e-utran using mapped context (cfr. clause 9.2.2 in TS 33.401 and S3-081233) are:

a) The TAU request following the handover is integrity-protected and not ciphered, with a NAS key derived from a fresh K_{ASME}

b) No $Nonce_{UE}$ is included in the TAU request.

c) $Nonce_{MME}$ had been included in the E-UTRAN HO Command prior to the TAU request.

d) K_{ASME} is refreshed by deriving it from CK, IK and $Nonce_{MME}$.

The justifications of these agreements are:

Ca) The TAU request is integrity-protected: this is possible due to the fact that the $Nonce_{MME}$ is made available to the UE in the HO Command before it sends the TAU request. This is the difference to the idle mobility case, cf. Aa). The integrity protection is desirable for the same reason as in idle mode.

Cb) No $Nonce_{UE}$ is included in TAU request for the following reason: the UE receives $Nonce_{MME}$ generated by the MME over protected interfaces, either core network interfaces or the protected UTRAN or GERAN air interface. (There may be a risk, however, when the GERAN air interface is not ciphered. But then there a even bigger security risks, especially in the GERAN PS domain which is the domain under consideration here, when ciphering is off.) In this way, the UE knows that $Nonce_{MME}$ was generated by the genuine network, and not an attacker. As the UE trusts both, the UMTS core network and the EPS core network, the UE also trusts that

$Nonce_{MME}$ is indeed fresh and has no need to generate a freshness parameter of its own. It should be remarked, however, that the inclusion of a $Nonce_{UE}$ by the UE would be possible.

The difference to the idle mode case in 7.7.1 b) is that there, in principle, the $Nonce_{MME}$ could have been used in a previous run of the procedure and, hence, the messages the UE receives in the current run, namely SM Command and TAU Accept, which are protected with a key derived using this very $Nonce_{MME}$, could have been replayed from the previous run if no $Nonce_{UE}$ was used. In the handover case here, the $Nonce_{MME}$ is sent HO Command which is replay-protected by UTRAN procedures independently of any mapping of keys on the E-UTRAN side.

c) The rationale for including $Nonce_{MME}$ in the E-UTRAN HO Command is given above.

d) The need for refreshing K_{ASME} is the same as for the idle mode case (caching of CK, IK in SGSN, and UE switching back and forth between UTRAN and E-UTRAN). It is explained above that it is sufficient to derive the fresh K_{ASME} from CK, IK and $Nonce_{MME}$.

7.7.4 TAU after handover from UTRAN to E-UTRAN using cached context

This case does not exist in TS 33.401 for the following reasons:

- Negotiating between UE and MME during the handover procedure whether a cached context was available on both sides was considered to add too much complexity, cf. reply from SA3 in S3-081138 to LS from RAN2.
- Furthermore, the HO Complete message would have to be sent not ciphered in case such a negotiation would be performed. This would have been a deviation from current RAN procedures.
- Therefore, at AS level the mapped security context is established during handover.
- As a consequence, it was decided that the mapped context shall also be used at NAS level as
    - the key hierarchy is built on the assumption that AS keys in use and NAS keys in use are derived from the same K_{ASME}. A deviation from this assumption would again add complexity to the specification.
    - there would be little security gain if the NAS level used cached context, but the AS level still used mapped context (assuming that the cached context may provide a higher degree of security than the mapped context).
- This does not preclude a subsequent switch to the cached context, if available, by performing a key-change-on-the-fly.

7.8 Track of decision

7.8.1 MAC, RLC, and RRC layer security

See Annex A for decisions about "MAC, RLC and RRC layer security".

“It was decided that RRC is always integrity protected.” (from Annex A)

“It was decided that a separate key set for RRC protection is necessary if RRC is terminated is in Node-B in order to prevent the derivation of NAS and User Plane keys” (from Annex A)

Further on, no MAC layer integrity protection or ciphering as a working assumption (from S3-060565)

Further on, RRC ciphering is a working assumption (from R2-062718)

Editor's Note: There is some concern on the cost of implementing RRC ciphering. If there is a low cost solution as a countermeasure to the threat above, SA3 is open to considering that solution.

Further on, combined COUNT-I and COUNT-C considered having no security concerns for RRC (from S3-060833)

Further on, RRC Integrity and ciphering algorithm can only be changed in the case of the eNB handover (from S3-060833)

Further on, RRC Integrity and ciphering will be started only once during the attach procedure (i.e. after the AKA has been performed) and can not be de-activated later. The combination of assumptions 1 and 2 means that integrity and ciphering cannot be switched to a "dummy" algorithm except at handover; this restriction is acceptable to SA3 (from S3-060833)

Further on, RRC Integrity and ciphering will always be activated in one procedure. However, it should be noted that SA3 cannot offer a guarantee that integrity and ciphering will be activated at the same time within the procedure; integrity may start before ciphering, even though the two activations are triggered as a single procedure (from S3-060833)

7.8.2 LTE AKA requirements

Comparison between EAP-AKA + EAP-ER and UMTS AKA (conclusion from SA3#45).

Requirement-0: The SAE CN and LTE AN SHALL allow for keys of size 128 or 256 bits. (from S3-060632)

NOTE: A condition to protect AS and NAS by means of 256-bit keys with an entropy of 256 bits is to have permanent key K of length 256 bits. (From S3-080044)

Requirement-1: AKA for non-3GPP access SHALL use USIM based EAP AKA (from section 7.2.1).

Editor's Note: This requirement has to be confirmed when the other aspects are ready.

Requirement-2: 2G SIM Access to LTE SHALL NOT be granted (from section 7.2.1).

Editor's Note: The security SA could be set up shortly after the authentication.

Requirement-3: LTE AKA SHALL be based on USIM and (possible) extensions to UMTS AKA. In particular, R99 USIM shall be sufficient for access to LTE (from section 7.2.1).

Requirement-4: LTE AKA SHALL produce keys forming a basis for UP/CP protection (ciphering, integrity) (from section 7.2.1).

Requirement-5: The LTE AKA keys of R4 SHALL be dependent on the algorithm with which they are used (from section 7.2.1).

Reuse UTMS AKA in LTE/SAE authentication (S3-070085).

7.8.3 NAS level signalling security

"It was decided that a separate key set for RRC protection is necessary if RRC is terminated is in Node-B in order to prevent the derivation of NAS and User Plane keys" (from Annex A)

Refer to A5.1.1 of S3-060119 [1]:

- "Clear requirement that keys used in the CN (for user-plane ciphering) should NOT be provided to the Node-B"
- "NAS protected above Node-B"
- "SMC to manage user-plane and NAS security above Node-B"

A mechanism similar to TMSI mechanism should be used. Protection against active identity and location confidentiality attacks (e.g. IMSI catching) should not be a high priority requirement in the LTE/SAE security design. However, if an effective solution can be developed at a relatively low cost, then it should be introduced into the specifications (from IMSI catching attack threat conclusion).

LTE/SAE shall support the same level of User Identity Confidentiality as today's 3GPP system (e.g. Idle mode signalling and attach/re-attach with temporary user identities). NOTE: This is from Section 5 of TR 23.882 (Requirements on the architecture). (from Threat of UE tracking conclusion)

NAS signalling is used for local authentication, key agreement, algorithm negotiation, and negotiation of parameters that are needed for ciphering and integrity protection of NAS signalling and ciphering of U-plane. (from threat of forced handover conclusion)

Further on, combined COUNT-C and COUNT-I considered having no security concerns for NAS signalling (from S3-060833).

7.8.4 Key handling

Draft Report SA3#42: "So, at this stage there is no convincing argument that separate keys have significant benefit, but SA3 would like to reserve the right to continue study on it. It is understood that RAN still needs to go forward with the Handover, architecture and it was decided that RAN should be given the go ahead on common keys. "

"Opinion were requested so that a decision could be made whether to have serving network authentication or not. Delegates were asked to determine if this is a threat or not."

Editor's Note: The added value isn't clear.

SA3 agreed to use key hierarchy presented in 7.4.7 as a working assumption for LTE (S3-070095).

SA3 agreed to not bind the K_{eNB} or RRC/UP keys to the eNB identity, because:

- The K_{eNB} is renewed on each IDLE to ACTIVE transition, so in this case binding to a certain eNB ID does not give any extra security
- Re-binding the K_{eNB} to the target eNB identity at inter-eNB handover requires that the eNB identity is sent to the UE in an encrypted NAS message. This requires that a NAS message is introduced purely for this purpose, which was seen as too complex.
- The UE only knows the RAN as cells, not as eNBs. Introducing the eNB identity in the KDF requires that the network must expose its topology.

Editor's Note: it's FFS if cell ID and tracking area ID is needed to be bound.

SA3 agreed to not bind the K_{ASME} or K_{eNB} MME identity, because:

- The UE is agnostic of the MME it is connected to, and extra signalling would have to be introduced just to achieve this binding.
- Introducing the knowledge of the MME the UE is connected to exposes the network topology to the UE.

Editor's Note: Whether this key hierarchy should be introduced in UTRAN if ffs (relation to S3-070089).

SA3 agreed to bind authentication vectors to SAE usage with the AMF field (7.4.8) (S3-070096)).

Key handling in idle and active mode mobility presented in 7.4.9 was adopted as a working assumption (S3-070097).

Editor's Note: Usage of START value is ffs.

Key handling in mobility, presented in 7.4.10, was agreed as the baseline for further discussions (S3-070099).

7.8.5 Security procedures

Editor's Note: This section needs to be re-checked in case the PDCP is relocated to eNB.

- **the meaning of "transparently to the UE RRC and the eNB" in assumption 4 (in S3-060833) "Change of integrity and ciphering keys will be performed transparently to the UE RRC and the eNB at state transition from idle to active mode"**

At state transition from idle mode to active mode an RRC context will be established in the UE and the eNB respectively. At this occasion the UE RRC and the eNB will be provided with keys from higher layers and the MME respectively that are used for applying RRC integrity protection and ciphering. Whether at this procedure new keys or already applied keys are given to the UE RRC or the eNB does not affect the procedure, supposing that in each case a suitable START value is negotiated between the UE and the

NodeB which might be 0 in the case of new keys. In this case the change of keys from previously used keys can be seen as “transparent” to the eNB / RRC.

- **The assumption 4 (in S3-060833) refers only to RRC signalling.**

- **Would incrementing the RLC sequence number by an offset at handover, instead of resetting it to zero, be acceptable to RAN2?**

RAN2s intention is to remove the need to inform the target eNB about the last used SN in the source eNB since this prevents the transmission of messages from the source eNB to the UE after initiating the handover procedure towards the target eNB. Therefore applying an offset to the SNs after the handover compared to the SNs used before the handover is not desirable. RAN2 does not see any problem with restarting the RLC SN from an arbitrary value. However it is the RAN2 understanding that it is anyway easily detectable from the signalling of the target cell that a new UE has just arrived due to the fact that a new C-RNTI is used in the signalling of the target cell, and thus we do not see any gain from this proposal.

- **Would incrementing the PDCP sequence number by some offset at handover be acceptable to RAN2, if done by eNB and the UE?**

The PDCP SN is assigned in the PDCP entity in the UE and in the UPE in the network, and is supposed to be handled transparently by the UE lower layers and the eNB. Incrementing the PDCP SN in the UE and the eNB would imply a violation of this layering principle. Introducing a gap in the SNs in the UPE and the PDCP entity in the UE would not work due to the fact that in the DL all PDCP PDUs may not have been transmitted in the source eNB and will be forwarded to the target eNB which would imply that the PDCP SN would be consecutive. Furthermore due to consecutive handovers RAN2 is concerned about the fact that the PDCP SN would increase very quickly. For these reasons incrementing the PDCP sequence number by some offset at handover is not seen as a possible solution in RAN2.

7.8.6 Security Algorithms

SA3 has agreed the following decision regarding the security algorithms used in LTE.

Encryption algorithms that shall be supported are:

- NAS: UEA2 and AES (AES mode of operation is FFS)
- UP: UEA2 and AES (AES mode of operation is FFS)
- RRC: UEA2 and AES (AES mode of operation is FFS)

Integrity algorithms that shall be supported are:

- NAS: UIA2 and an integrity algorithm based on AES (design of AES based algorithm is FFS)
- RRC: UIA2 and an integrity algorithm based on AES (design of AES based algorithm is FFS)

The arguments for the choice of AES as core algorithm for the second algorithm (compared to UEA1/UIA1) were the following. Apart from these they were perceived as equally good choices:

- The eNB needs to support AES in any case because the eNB needs to support NDS/IP, which uses AES.
- The licensing conditions on the core of UEA1/UIA1 (Kasumi) do not make it free for use for other purposes than 3GPP access protection.
- Similarity with other non-3GPP accesses.

Both network and terminals shall support both algorithms from the start.

The input parameters to the encryption algorithms shall be the same, and shall be:

- NAS: FFS
- UP: FFS
- RRC: FFS

The input parameters to the integrity algorithms shall be the same, and shall be:

- NAS: FFS
- RRC: FFS

8 Network Domain Security

This chapter describes how Network Domain Security according to TS 33.210 could be used to counteract certain IP-based threats on the LTE reference points. Section 8.1 gives a general overview; section 8.2 clarifies which threats from section 3 until 5 can be counteracted and which not. Finally section 8.3 provides a summary of the required security of NDS/IP.

Editor's Note: If relevant threats are added to section 3 or 4 then this chapter may also need further changes.

8.1 Introduction

8.1.1 NDS/IP architecture applied to LTE

TS 33.210 defines a Za and a Zb-interface that is applied between NE's (Network Elements) and SEGs (Security Gateways) in order to protect the transfer of signalling data.

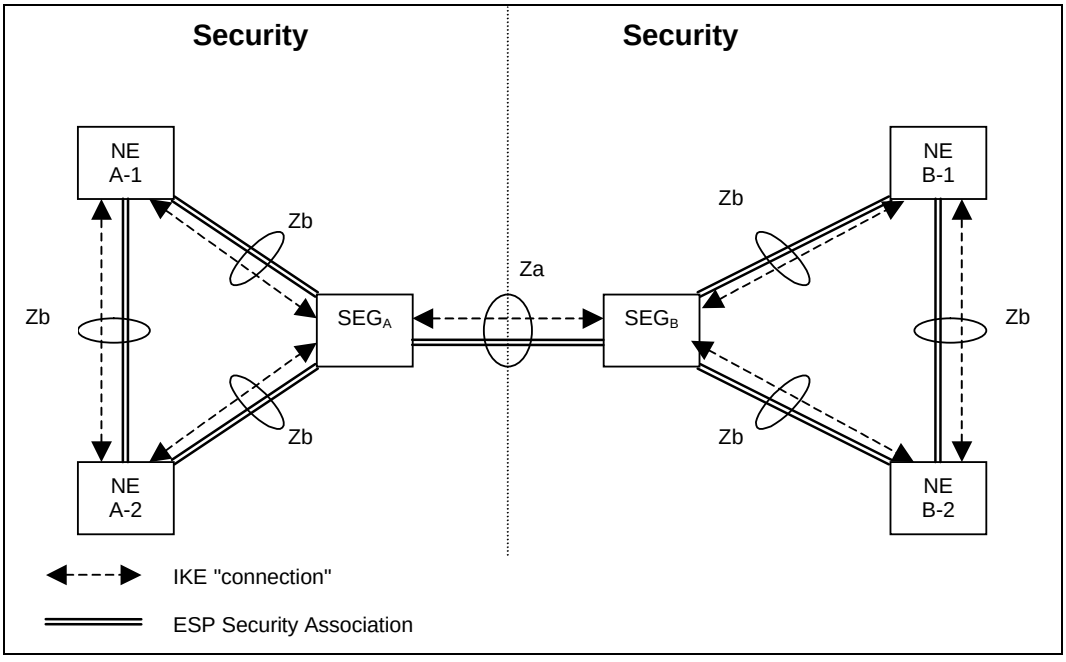


Figure 38 NDS architecture for IP-based protocols from TS 33.210

If we convert Figure 38 towards SAE/LTE entities then NE A-1 may be a core network node (i.e. an SAE GW or an MME) and NE B-1 may be the eNB. The core network node and the eNB may reside in different security domains (e.g. if they are connected over the Internet). The SEG may be integrated into the NE or may be a standalone device. If the link between the SEG and the NE can be trusted (e.g. the link between the core network node and the SEG resides in the same building of the operator) than no additional security (other than the physical measure) needs to be applied between

them (i.e. the Zb reference point security is optional). Alternatively, if the core network node and eNB reside in the same security domain, they may be mapped to NE A-1 and NE A-2 respectively and the optional Zb interface would be used between them.

If several nodes are placed within the same trusted environment, then it may be advantageous to concentrate the security processing in a stand-alone device i.e. a SEG at the border of the trusted domain. This may be the fact for the core network node but also for eNBs. In any case the number of (semi-static) security associations for NDS/IP on the S1-reference points between eNB and the core Network will largely be determined by the number of eNBs.

It is assumed that the S1-reference points between eNBs and the core network may go via the open internet or over equivalent solutions with similar low protection level (e.g. the operator leases an IP-line from a carrier that cannot guarantee the prevention of security threats on that leased line). While the core network node resides in a trusted location, this is not necessarily the case for the eNB. In this case, the physical links in the vicinity of the eNB may be vulnerable. Therefore, in the general case, IPsec functionality (according to TS 33.210), terminating either Za (i.e. SEG functionality) or Zb will have to be integrated in the eNB, to prevent breaches if there would be a separate SEG to eNB link. However we should not rule out the deployment option where the vicinity of eNBs is sufficiently trusted, but the backhaul link to the core network is not. In this case, it may be advantageous to use a SEG aggregating the traffic from several eNBs.

8.1.2 Key Management solutions for NDS/IP

In the distributed case signalling and packet forwarding exists between the eNBs. At the same time the transmission links between eNBs are considered to be insecure, meaning that the threat of packet injection, packet eavesdropping, and packet modifications exists on these links. Handovers can also happen between many different eNBs, depending on the network configuration and management.

There are various methods to provide key management for NDS/IP between eNBs:

1) NDS/IP could be used to secure connections between eNBs, based on pre-shared secrets. This would mean that Operations & Management is required to create the SAs between the required eNBs, or that the pre-shared secrets are transferred to the right eNBs by some other means.

2) eNB specific certificates could be used to bootstrap security associations between eNBs. This would mean that each eNB shall have its own certificate signed by a Certificate Authority (CA) and the corresponding root certificate from the CA for certificate validation. This would also probably mean that certificate revocation methods should be implemented or short enough certificate lifetimes should be used. The latter requires provisioning of new certificates, before the lifetime of the current ones is exceeded. Choosing the right lifetime becomes a trade-off issue between a fresh and possibly a disclosed certificate.

3) Centralized node(s) in the network could bootstrap eNB-eNB security associations automatically when needed. This would mean that the centralized node(s) know the topology of the eNBs (i.e. at least neighbour eNBs for each eNB).

Evaluation:

In cases 1 and 3, when adding a new eNB to the network, the existing neighbouring eNBs need to be updated to incorporate the security association or needed credentials with the new eNB. In case 2 the certificate management must be implemented and the certificates in the eNBs must be protected and provisioned.

8.1.3 Alternatives

An alternative for NDS/IP is to provide the keying material inside a subscriber context from the core network to the eNBs. The MME encrypts a subscriber specific signalling protection (symmetric) key for multiple eNBs at the same time and sends all these encrypted keys to the subscribers' current eNBs in the subscriber's context.

When secure signalling between eNBs is needed the source eNB uses the subscriber specific signalling protection key to protect the messages, finds the encrypted entry for the target eNB and sends it along with the messages to the target eNB. Target eNB then decrypts the key and the corresponding received messages. This way the source eNB can securely communicate with all eNBs that are included in the subscribers context received from the core network. This does not mandate neighbouring relationship between the eNBs.

In this case, there is no need to maintain security associations between eNBs, because the exchanged messages themselves include needed material for message decryption securely delivered to the corresponding eNB.

Editor's Note: There isn't preference to the above countermeasures.

Editor's Note: The text of this section needs further clarification how a subscriber context could be used as alternative to NDS/IP.

8.2 How particular threats can be counteracted.

In the distributed eNBeNB-architecture, signalling and packet forwarding exists between the eNBs. At the same time the transmission links between eNBs are considered to be insecure, meaning that the threat of packet injection, packet eavesdropping, and packet modifications exists on the links. Handovers can also happen between many different eNBs, depending on the network configuration and management.

In this section we analyse the IP-based threats, and evaluate whether and how NDS/IP provides a countermeasure. In this section we only consider outsider attacks between UE and the first uplink core network node (i.e. the MME and the SAE GW), and on the IP-based reference points between eNB.

NOTE: Only those threats from Section 3 and 4 were evaluated which were found relevant.
NOTE: The threats within this section are numbered as NDS-Threat-x in order to have a numbering independent from section 3 and 4. This will allow to renumbering of sections 3 and 4 with minimal impacts in this chapter.

8.2.1 Threats to User Data

NDS-Threat-1: Section 3.1 User Plane packet injection attacks (Threat-B): 'The attacker injects user plane packets on the last-mile, while eNB, UE and SAE GW are not compromised. DoS attack is also possible. Attacker may send broadcast packets to the access link and try to congest access network as much as possible.'

Evaluation:

If the interface between SAE GW and eNB is accessible for an attacker then an attacker could indeed inject packets via that interface towards the UE. The eNB would simply forward these packets towards the UE, irrespective of whether there would be a higher layer protection mechanism on the user plane data. In this way, an attacker could overload the air interface and deny service. Packet filtering methods must be used here. However the use of NDS could prevent that the eNB sends bogus packets further into the radio access network.

Protecting S1 user plane with integrity protection (between eNBs and SAE gateways) adds a requirement for the eNB to start processing each user plane packet going through the eNB, both uplink and downlink. This adds to the cost of the eNB as additional hardware crypto chip is required. Taking into account the high bandwidth of LTE, the crypto hardware must be powerful enough making it unsuitable to use the same hardware as is currently used.

Having S1 user plane integrity protection also increases the processing requirement of crypto hardware in the SAE gateway for all user plane packets that are integrity protected on the S1 interface, both up and downlink. This adds to the cost of the whole LTE system.

Adding integrity protection to the S1 user plane interface also increases the packet processing times on the system (first in eNBs and then in SAE gateways). Power consumption in the eNBs and SAE gateways also increases.

Having integrity protection between eNBs and SAE gateways in case of separated MME and SAE gateway increases the number of Security Associations on the LTE system, as each eNB must then have also an SA to the SAE gateways (or worse, to separate security GWs). This has an impact to the total system performance and management of the SAs.

However if NDS-threat-4 has to be counteracted by applying confidentiality protection on S1_U then the cost of adding integrity protection would be much lower when starting from null.

S1 user plane interface in case the attacker is flooding packets with very high speed and the receiving buffers in the eNB are overflowing. Attacker having access to the S1 links means that she/he may also try to congest the link regardless if there is integrity protection on S1-U or not. The result is service level degradation and possible packet drops. Integrity protection of S1 user plane packets does not solve these problems.

In the uplink, the effect of User Plane packet injection towards the SAE GW is similar as described for the downlink direction. NDS could not stop an attacker from bombarding the SAE GW with bogus packets. Packet filtering methods must be used here. However the use of NDS with integrity could prevent that the SAE GW sends bogus packets further into the core network. Note that User Plane packets (with no integrity activated on S1-U) are forwarded by the SAE GW and eNB only if the attacker could correctly guess the required headers. Since it is not possibly to reliably apply

replay protection without integrity protection guessing the required headers is not necessary for an attacker. The attacker may simply copy an old packet and may by changing a few bits in the higher parts of the packet easily damage transport layer and application layer messages. The result of this would from the (both on server side and UE side) IP stack and application point of view be more or less randomly looking errors that are very difficult to trace.

As a result the packet injection attack threat described in the security rationale document is high enough to justify S1 user plane packet integrity protection both for the uplink and downlink. If the headers can be guessed correctly by an attacker injected packets on the S1 interface could go through the eNB to the air interface for the downlink, or in the uplink the injected packets could pass the SAE Gateway..

NDS-Threat-2: Section 3.2 User Plane packet modification injection attacks between eNB and the UE: (Threat-A) 'The attacker modifies encrypted user plane packets, so as to deny service from the UE by modifying UE packets in such a way that the UE must re-transmit etc. In this way the attacker acts as man-in-the-middle between UE and UPE. This affects the service quality that the UE (subscriber) is seeing'.

Evaluation: Applying NDS between eNB and SAE GW does not seem to help against attack between eNB and UE.

NDS-Threat-3: Section 3.3 User plane packet eavesdropping between the eNB and the UE

Evaluation: Applying NDS between eNB and SAE GW does not seem to help against attacks between eNB and UE.

NDS-Threat-4: User plane packet eavesdropping between the eNB and the SAE GW (S1-U) or between two eNB's (X1-U)

Evaluation: Applying NDS with confidentiality activated does counteract this threat.

NDS-Threat-5: IPsec tunnels that provide confidentiality but not integrity may be put out of synch.

Evaluation: The default behaviour of ESP is to use extended sequence numbers. That works similar to encryption in UTRAN where there is a Hyper Frame Counter that is not sent over the link, but is increased every time there is a wrap around of the sequence number (which *is* sent over the link). An attacker can force the "Hyper Frame Counter" out of synch between the two peers by injecting a bogus IP packet (or replaying an old IP packet) with a low sequence number, when the real sequence numbers are high. ESP has a recovery mechanism for this, but that one is based on that there is integrity protection in place. If integrity protection is used, the described attack would not work, since the "Hyper Frame Counter" is covered by the MAC, and the re-synch mechanism can be used if something still goes wrong.

8.2.2 Threats to Signalling Data

NDS-Threat-5: Section 4.1 Dos Attacks from false MME against eNB

Evaluation: This concerns control plane traffic which is originated from a false MME towards genuine eNB. As control traffic we distinguish S1-signalling (Iu-like) between eNB and MME and NAS signalling between UE and MME. The vice-versa case is similar.

It's assumed that NAS signalling shall be integrity protected and may be confidentiality protected between the UE and the MME. Similar consideration as for NDS-Threat-1 applies i.e. the availability of higher layer protection mechanism can not prevent packet processing and forwarding at the eNB. IP packet authentication is needed to prevent that DoS attacks towards eNB's spread further towards the air interface.

However note that signalling on the S1-reference point will transfer RRC and PDCP User plane keys, so there is a requirement for confidentiality protection of the S1-signalling between MME and eNB

NDS-Threat-6: Dos Attacks from false eNB to eNB.

Evaluation: Similar as NDS-Threat-1: IP packet authentication is needed.

NDS-Threat-7: Attacks on the eNB-eNB interface.

Evaluation: Similar as NDS-Threat-1: IP packet authentication is needed to prevent spoofed handover commands. It is likely that sensitive information will be transferred on this interface which will require confidentiality protection (e.g. RRC or PDCP user plane keys in handover). For the protection of User data, the same rationales as on the S1_U references points applies i.e. the countermeasures against user plane packet injection attacks are not good enough reason compared to the loss in bandwidth and processing performance.

8.3 Summary

Reference point and data type / security requirement	Integrity/authentication	Confidentiality	Remarks
User Plane Data			
S1-User plane (SAE GW-eNB)	Yes	Yes	TS 33.210 only covers signalling data
eNB-eNB (X2-U)	Yes	Yes	TS 33.210 only covers signalling data
Signalling Plane Data			
S1-C transferring NAS signalling (MME and -eNB)	Yes	No	
S1-C (Iu-alike) between MME and –eNB.	Yes	Yes (transfer of sensitive information e.g. RRC and PDCP user plane keys)	
eNB-eNB (X2-C)	Yes	Yes if sensitive information is exchanged (RRC and PDCP user plane keys)	

8.4 Network Domain Security Evolution[48]

TS 33.210 provides an overview on how IPsec/IKE shall be used for protection of signalling protocols between two core nodes. Signalling traffic going outside or entering a security domain needs to pass a Security Gateway (SEG). Starting from Rel-4, IPsec tunnel mode was selected as the only IPsec mode. At SA3#48, a CR was approved allowing the use of IPsec in transport mode within a security domain, but at the same time not mandating the implementation.

As the amount of free IPv4 addresses is getting shorter and shorter, and may exhaust within a few years, the deployment of IPv6 capable nodes will increase which alleviates the need to use NATs (and smaller security domains). When transport mode can be used within a security domain, then it has an advantage over tunnel mode due to the smaller IPsec header overhead[49]. This overhead consideration is in particular interesting where IPsec needs to be used to protect **user data** of smaller packet size and without cross-border firewalling/inspection requirements.

Proposal-1: Mandate the support of IPsec Transport mode on particular interfaces that need to handle lots of data i.e. S1_U and X2-interface for LTE.

This could be performed by adding a separate chapter or an Annex to TS 33.210

With regard to the key management protocol IKEv1, only pre-shared keys support is needed in 3GPP Rel-7. Certificate based IKE authentication is included in TS 33.310 but only between Security Gateways and thus not for use on intra-security domain interfaces. The introduction of TS 33.310 contains following introduction text related to this:

"In the case of NDS/IP this Specification concentrates on authentication of Security Gateways (SEG), and the corresponding Za-interfaces. Authentication of elements in the intra-operator domain is considered an internal issue for operators. This is quite much in line with [1] which states that only Za is mandatory, and that the security domain operator can decide if the Zb-interface is deployed or not, as the Zb-interface is optional for implementation. ***However, NDS/AF can easily be adapted to intra-operator use since it is just a simplification of***

---

48 This section is from S3-070760.

49 Transport bandwidth will be a scarce resource for quite some time, as long as there will be eNBs which are not connected by fiber.

the inter-operator case when all NDS/IP NEs and the PKI infrastructure belong to the same operator. Validity of certificates may be restricted to the operator's domain."

In the light of more dynamically changing networks configurations as we expect for the interconnection of E-UTRAN with the EPC, the use of IKE certificates with automatic enrolment seems advantages also for intra-domain usage.

However, if transport mode is used on the S1 interface it implicates that both MME and S-GW have to implement IPsec. Especially for large networks having the S-GW, and to a lesser extent the MME, doing IPsec will not scale and the nodes would have to do tasks that just as well could be done by a SEG. SA3 has agreed that IPsec Transport mode for S1 is optional for implementation. However, IPsec tunnel mode, used in combination with a SEG, does not have these limitations.

Proposal-2: Mandate the support of IKE certificates with automatic enrolment on these E-UTRAN and EPC nodes that need to handle lots of interconnections i.e. S1 and X2 interfaces for E-UTRAN.

The requirement could be added to TS 33.abc or to RAN Specification while TS 33.210 does not refer to certificate support and TS 33.310 does not list specific interfaces.

Proposal-3: Extend TS 33.310 such that it explicitly covers the use of certificates within a security domain.

Another evolution is the use of IP multicast on particular reference points for user or signalling traffic. In particular the use of IP multicast on user data saves processing power in the source node. As described in the LS S3-070618 on "security for the eMBMS architecture" to RAN3, a particular usage may require the support of a recent IPsec RFC i.e. RFC4303 than currently required by TS 33.210.

Proposal-4: Extend TS 33.210 to include protection of multicast traffic for particular interfaces/usages.

Add a specific chapter(s) on the support of security solutions for protecting Multicast data. Add an Annex if specific interfaces shall follow these requirements.

8.5 IKE version in NDS/IP for EPS[50]

IKE (here used to denote both IKEv1 and IKEv2) is a key management protocol, which establishes an IKE Security Association (IKE-SA, or phase 1 SA) between two endpoints. Using this IKE-SA, IKE can be used to establish so called child SAs (or phase 2 SAs), which are used to protect the actual IP traffic between the nodes.

IKEv2 was developed because IKEv1 have drawbacks. The table below shows differences between the two protocols. It is obvious that this is not a comparison between two fairly competing proposals, but rather a list of some important improvements that IKEv2 provides to IKEv1. Basically the only thing that speaks for IKEv1 is that it is present in TS 33.210 as is shown in the first row.

Property	IKEv1	IKEv2	Comment
Used in TS 33.210	Yes	No	
Round trips to establish IKE-SA	6	4	IKEv2 can establish a child SA during this exchange as well.
Round trips to establish child SA	3	2	
SA selectors	Limited choice	Better	IKEv2 has better specification and flexibility for choosing traffic selectors.
SA management	Open only	Open/close	IKEv1 does not specify how to close SAs, whereas IKEv2 does.
Handling of QoS	Limited	Better	IKEv2 allows creation of different SAs between the same endpoints, which can be used for different QoS classes.

50 This section if from S3a070925.

Flexibility of SA handling	Less	Better	For example, the SA lifetimes (in IKEv1, lifetimes are negotiated at the beginning while in IKEv2, each party can choose its own SA lifetime independently of the other party).
Authentication flexibility	Less	Better	IKEv2 allows for EAP. If IPsec is also re-used for O&M protection, connections to different existing O&M systems can use the credentials they already support.
Support for multi homing	No	Yes	SCTP, which support multi homing, is used on S1 and X2 interfaces. This cannot be utilized in combination with IKEv1.
NAT traversal	No	Built in support	
Protocol complexity	Large	Less	IKEv1 has (righteously) been accused of being a complex protocol. IKEv2 has less types of phases and less messages in each phase.
Ease of implementation	Complex	Less complex	Due to the simpler protocol structure of IKEv2, it is easier to implement, which leads to a smaller probability of errors and less cost.

It is clear that IKEv2 is far better than IKEv1 on all accounts. What is further clear is that these features are useful for EPS.

The number of roundtrips to establish SA:s between nodes is far less for IKEv2, which reduces the time to set up security at installation and re-boots of eNBs.

IKEv1 can only establish SA:s, whereas IKEv2 gives the possibility to establish, close and manage their life-times individually. Further, IKEv2 have more flexible ways of dealing with SA selectors, which makes management and policy specifications easier. For nodes that are multi homed, IKEv2 allows selectors to work also in this case, which is of interest since SCTP is used on S1 and X2.

Since IKEv2 provides built in support for NAT traversal, more flexibility is added in deployment and network design.

IKEv2 allows authentication using EAP, which gives more room in the design of protection to home base stations, which may not use the same type of credentials as regular eNBs (considering that the trust model for home base stations is very different).

IPv6 enabled nodes must support IPsec according to RFC 4301, in which IKEv2 is specified as the default key management protocol.

A non-specification issue, but very important from implementation and deployment point of view, is that IKEv2 is a much simpler protocol than IKEv1. This leads to cheaper implementations with less probability for implementation errors and interop-problems.

On the theoretical side, IKEv2 is built on the SIGMA approach to Diffie-Hellman key agreement, which has a proof of security.

There is no doubt that IKEv2 is preferable to IKEv1 when compared. It is therefore proposed that IKEv2 shall be used as key management protocol for NDS/IP in EPS.

Editor's Note: This is agreed as a working assumption for X2 and S1 unless there is show stopper. Further analysis on the migration issues and impacts of 33.210 needs to be performed.

8.6 S1/X2 reference point security[51]

The backhaul transmission link carrying traffic and signalling to/from the eNB over the S1and X2 reference points may be vulnerable to external attack, particularly due to the fact that radio interface encryption terminates in the eNB. Therefore mechanisms shall be available to encrypt and authenticate the user traffic, signalling and management data carried over this link. In addition, the following requirements are identified:

---

[51] This section is from S3a070955.

- The mechanisms to secure the S1 and X2 interfaces shall be bandwidth efficient.
- It shall be possible to re-use the mechanisms to secure the S1 and X2 interfaces to secure backhaul link communications associated with other types of 3GPP and non 3GPP radio technologies that may be supported at the base station site.
- The mechanisms to secure backhaul link communications, particularly the key management part, shall be designed such that they can be easily extended or modified to support the specific requirements when base stations are installed in customer premises (cf. H(e)NB security study item).

8.7 S6a Reference Point Security[52]

Subscription and authentication data is transferred over S6a. That data is valuable for the operator, and the operator needs be able to trust the data. The data is also valuable from subscriber point of view, e.g. for privacy reason. In order to ensure that the data transported over S6a is trustworthy and kept out of reach from 3rd parties, the following security requirements are assumed:

1. The confidentiality of the S6a messages shall be ensured
2. the integrity and replay protection of the S6a messages shall be ensured
3. Mutual authentication of the communicating entities shall be ensured
4. If proxies are used on S6a, then the requirements 1-3 shall apply on each hop.

In further discussion, S6a security is considered using the concept of security domains as a starting point. The following scenarios need to be considered:

1. The MME and the HSS both reside in the same security domain
   In this scenario, it is the responsibility of the security domain operator to enforce a security policy that will ensure confidentiality, integrity and mutual authentication. This could be achieved for example by physical means, or by enforcing a suitable security protocol.

2. The MME and the pre-rel8 HLR reside in the same security domain
   This scenario is analogous to scenario 1, and the same security considerations apply.

3. The MME and the HSS reside in different security domains
   In this scenario, adequate explicit protection mechanisms need to be put in place to protect the traffic. Two options are possible. The first option is that the MME and the HSS have a (secured) direct connection between each other. The second option is that the MME in security domain A communicates (securely) with a proxy in security domain B. The proxy would then further communicate with the HSS in security domain B, with adequate protection in place between the proxy and the HSS. If there would be several proxies on the path, then adequate protection should be in place on each hop as assumed by requirement 4 above.

4. The MME and the pre-rel8 HLR reside in different security domains
   This scenario is analogous to scenario 3, and the same security considerations apply.

The protection mechanisms are already present in the 3GPP specifications.

8.8 Authentication Failure Reporting (AFR) functionality for EPS[53]

Section 6.3.6 of TS 33.102 defines functionality for reporting authentication failures back to the Home network.

1) There seems to be little home operator benefit for this feature.Technically the network "under attack" is the visited network. There is little the home network can do to prevent this type of 'attack'. If the amount of illegal

---

52 This section is from S3-070731.

53 This section is from S3a070953.

authentication attempts becomes unacceptable, tracking down of the bad guys will need to be done in the visited network. All necessary information (IMSI, LAC, etc.) is available in the visited country/network. It may of course be useful to inform the home network when his IMSI range is abused in an attack, but we do not immediately see why it would be useful that the home network is informed in **real time** about every single authentication failure.

2) The mechanism to report the failures back to the home network creates additional network signaling with marginal benefit.

3) The proposed actions if executed (see italic text below) as described by TS 33.102 section 6.3.6 for the home network may lead to a possible DoS-attack on the subscriber.

"The HE may decide to cancel the location of the user after receiving an authentication failure report and may store the received data so that further processing to detect possible fraud situations could be performed"

According to the analysis aboveIt is agreed by SA3 to leave out this non-essential functionality (i.e. with little benefit) in order to simplify the EPC signaling design (S6a- protocol)

8.9 EPS interworking with a pre-Rel-8 HSS/HLR[54]

8.9.1 Current approach to binding authentication vectors to E-UTRAN serving network identity

E-UTRAN authentication vectors shall be derived from UTRAN authentication vectors so that Rel-99 USIMs can be used. In particular, the {CK, IK} keys in a UTRAN authentication vector shall be converted into a corresponding K_ASME key for use in E-UTRAN.

The message flow between the MME and HSS is as follows:

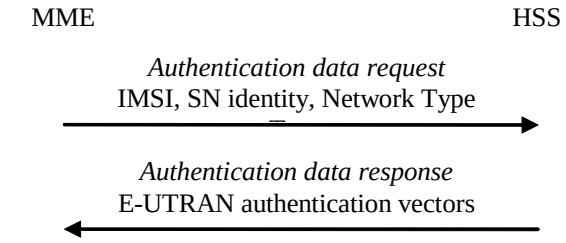


Figure 39 message flow between the MME and HSS

The method currently envisaged to provide the binding of authentication vectors to the E-UTRAN serving network identity makes use of 1 bit of the 16-bit Authentication Management Field (AMF) field in the authentication vector. This bit is termed a "separation bit" and is used as follows:

- The HSS shall never issue an authentication vector with the separation bit in the AMF set to 1 to a non-EPS network entity.
- For an EPS network entity, the HSS shall set the separation bit to 1 and generate an EPS serving network specific K_ASME from {CK, IK} using a key derivation function with the serving network identity as an input. If the separation bit is set to 1, then CK and IK shall not leave the HSS.
- An ME attaching to E-UTRAN (or another EPS access network) must check during authentication that the separation bit is set to 1 and abort authentication if this is not the case. If the separation bit is set to 1, then the EPS serving network identity is used as an input to the K_ASME derivation.

The binding of authentication vectors to the E-UTRAN serving network provides two security benefits:

- **E-UTRAN serving network authentication:** This allows the UE to be assured that it is connected to a specific E-UTRAN serving network. The binding of the authentication vector to a specific E-UTRAN serving network means that one serving network cannot masquerade as another.

---

54 This section is from S3a071031.

- **Cryptographic separation of E-UTRAN authentication vectors:** This prevents a non-E-UTRAN authentication vector from being used for E-UTRAN security. It helps avoid that security vulnerabilities in other applications of the authentication protocol (e.g. GERAN/UTRAN) leak into E-UTRAN.

The rationale for these enhancements is explained in section 7.3.2 of the SA3 TR 33.821 on EPS security.

8.9.2 Solutions for interworking with a pre-Rel-8 HSS/HLR

Six candidate solutions for interworking with a pre-Rel-8 HSS/HLR are described in the following sub-sections. It is proposed that 3GPP select and standardize one solution for all operators. In particular, it is assumed that multiple solutions do not need to co-exist. All solutions presented in this section can and shall co-exist with the target solution currently described in TS 33.abc v020.

8.9.2.1 Solution 1: K_ASME derivation and protocol conversion in HPLMN

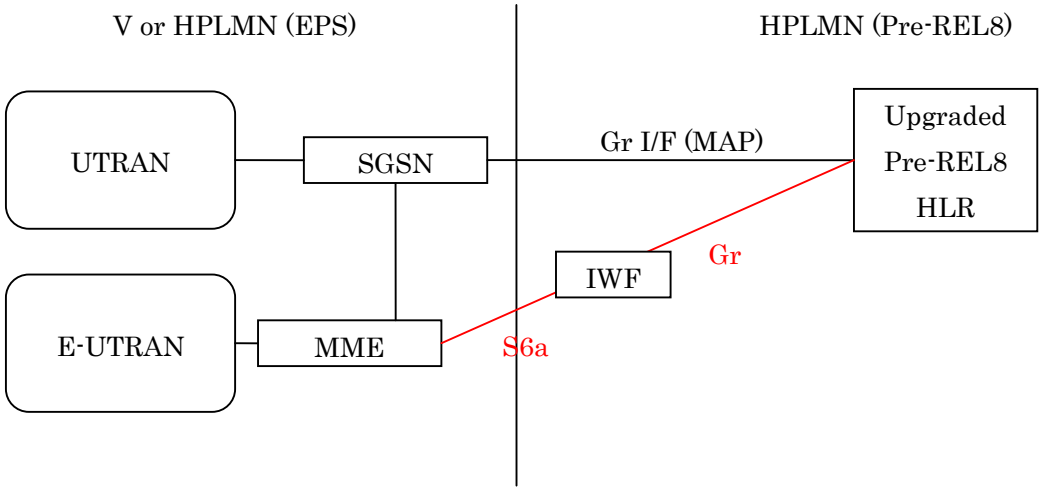


Figure 40 K_ASME derivation and protocol conversion in HPLMN

1. The AuC part of the pre-Rel-8 HLR is upgraded so that MAP authentication vector requests from nodes serving E-UTRAN can be identified. For such requests, the AuC sets the separation bit of the AMF to 1, otherwise it is set to 0.
2. An IWF in the HPLMN derives K_ASME using {CK, IK} and the serving network identity, and provides the necessary MAP-Diameter conversion of the authentication vector request/response.

Variants: The IWF may be split into separate boxes: one to perform K_ASME derivation, the other to perform protocol conversion. Key derivation may be performed either before or after protocol conversion.

Solution 1b: K_ASME derivation in HLR and protocol conversion in IWF in HPLMN

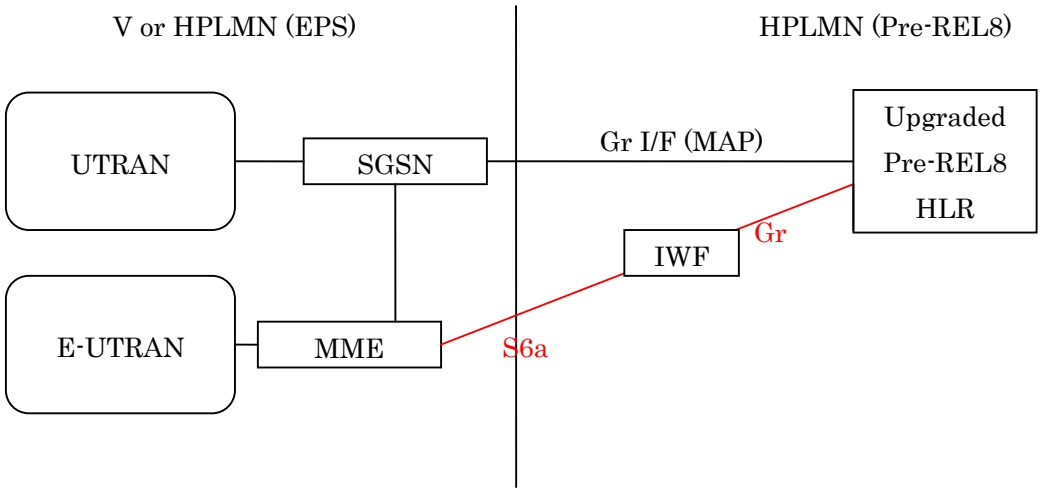


Figure 41 K_ASME derivation and protocol conversion in HPLMN

1. The AuC part of the pre-Rel-8 HLR is upgraded so that MAP authentication vector requests from nodes serving E-UTRAN can be identified. For such requests, the AuC sets the separation bit of the AMF to 1, otherwise it is set to 0.
2. The upgraded pre-Rel-8 HLR also derives K_ASME using {CK, IK} and the serving network identity. In this manner {CK, IK} does not leave the pre-Rel-8 HLR. The IWF provides the necessary MAP-Diameter conversion of the authentication vector request/response.

8.9.2.2 Solution 2: K_ASME derivation in HPLMN, protocol conversion in VPLMN

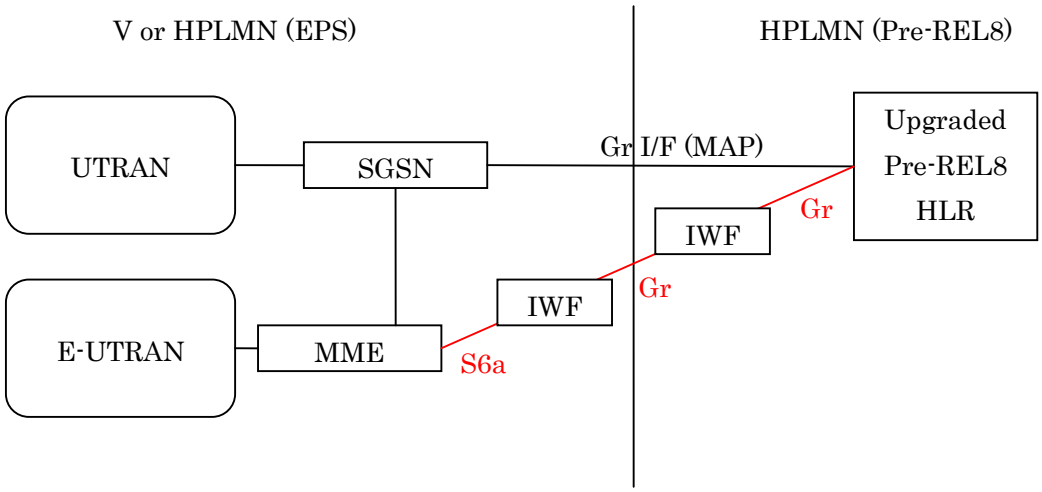


Figure 42 K_ASME derivation in HPLMN, protocol conversion in VPLMN

1. The AuC part of the pre-Rel-8 HLR is upgraded so that MAP authentication vector requests from nodes serving E-UTRAN can be identified. For such requests, the AuC sets the separation bit of the AMF to 1, otherwise it is set to 0.
2. An IWF in the HPLMN derives K_ASME using {CK, IK} and the serving network identity, which is determined from the source address of the authentication vector request. K_ASME is then carried in the {CK, IK} fields of the MAP authentication vector response.
3. An IWF in the VPLMN (or HPLMN when not roaming) provides the necessary MAP-Diameter conversion of the authentication vector request/response.

8.9.2.3 Solution 3: K_ASME derivation and protocol conversion in VPLMN (with dynamic setting of separation bit in HLR)

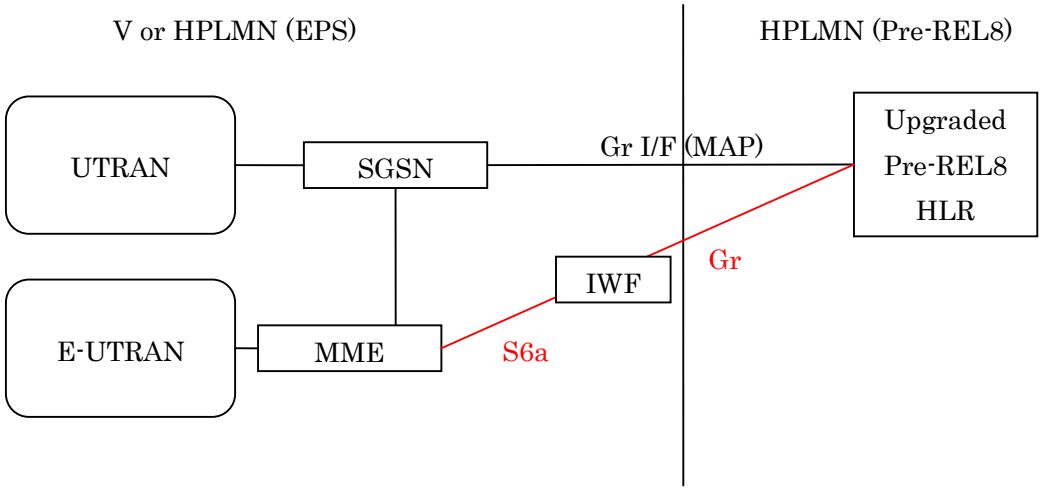


Figure 43 K_ASME derivation and protocol conversion in VPLMN (with dynamic setting of separation bit in HLR)

1) The AuC part of the pre-Rel-8 HLR is upgraded so that MAP authentication vector requests from nodes serving E-UTRAN can be identified. For such requests, the AuC sets the separation bit of the AMF to 1, otherwise it is set to 0.
2) The IWF in the visited network derives K_ASME using {CK, IK} and the serving network identity, and provides the necessary MAP-Diameter conversion of the authentication vector request/response. (Note that with this method there is little value in using the serving network identity as an input to the K_ASME derivation.)

Variants: The IWF may be split into separate boxes: one to perform K_ASME derivation, the other to perform protocol conversion. Key derivation may be performed either before or after protocol conversion.

8.9.2.4 Solution 4: K_ASME derivation and protocol conversion in VPLMN (with static setting of separation bit in HLR)

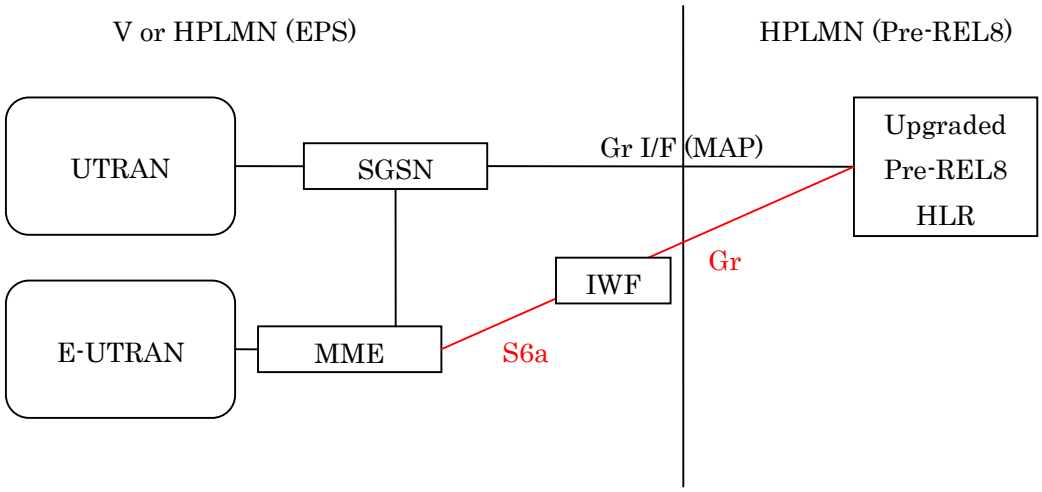


Figure 44 K_ASME derivation and protocol conversion in VPLMN (with static setting of separation bit in HLR)

1) The AuC part of the pre-Rel-8 HLR is upgraded so that the separation bit of the AMF is set to 1 for all authentication vector requests. Otherwise the HLR/HSS is unchanged and returns standard MAP authentication vector responses to the IWF.

2) The IWF in the visited network derives K_ASME using {CK, IK} and the serving network identity and provides the necessary MAP-Diameter conversion of the authentication vector request/response. (Note that with this method there is little value in using the serving network identity as an input to the K_ASME derivation.)

Variants: The IWF may be split into separate boxes: one to perform K_ASME derivation, the other to perform protocol conversion. Key derivation may be performed either before or after protocol conversion.

8.9.2.5 Solution 5: IWF in VPLMN with UMTS level security in EPS

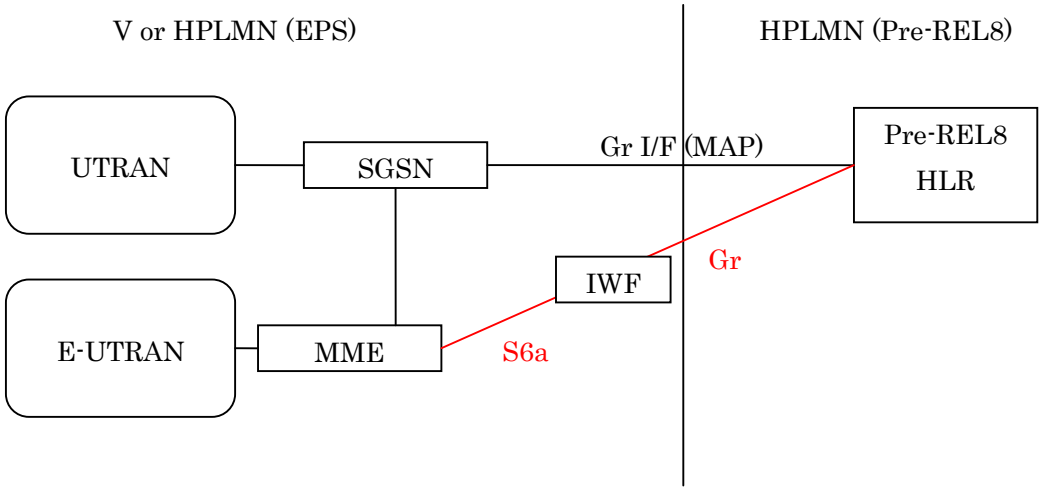


Figure 45 IWF in VPLMN with UMTS level security in EPS

1) The pre-Rel-8 HLR is not upgraded. Instead the security features of serving network authentication, and separation of E-UTRAN authentication vectors, are not provided in EPS networks. Consequently the separation bit is not set in the HSS/HLR and does not need to be interpreted by E-UTRAN capable ME.

2) The IWF in the visited network derives K_ASME using {CK, IK} and provides the necessary MAP-Diameter conversion of the authentication vector request/response. (Note that with this method there is little value in using the serving network identity as an input to the K_ASME derivation.)

Variants: The IWF may be split into separate boxes: one to perform K_ASME derivation, the other to perform protocol conversion. Key derivation may be performed either before or after protocol conversion.

8.9.2.6 Solution 6: Gradual upgrade of HLR using indicator on Rel-8 USIM

This solution allows an operator to start from solution 5 (UMTS level security and unmodified pre-Rel-8 HLR) and, at some later time after the start of EPS, upgrade his pre-Rel-8 HLR to a Rel-8 HSS. A Rel-8 USIM will contain an HLR indication bit (HI bit), which is set to 1 only when the HLR has been upgraded. However, the use of Rel-99 USIMs for E-UTRAN access is still possible. An ME attached to E-UTRAN will always check for this HI bit on the USIM.

User on pre-Rel-8 HLR:

If the HI bit is present on the USIM it is set to 0. An ME attached to E-UTRAN does not enforce cryptographic separation of E-UTRAN authentication vectors if the HI bit is absent (Rel-99 USIM) or set to 0, i.e. it also accepts authentication vectors with the separation bit in the AMF set to 0. The network entities behave according to solution 5. In case current USIM versions should not support adding the HI bit, new USIMs providing this support could be issued even while the user was still on a pre-Rel-8 HLR, with the HI bit set to 0.

User on upgraded pre-Rel-8 HLR:

If the user has an Rel-8 USIM the HI bit may now be set to 1. If it is set to 1 an ME attached to E-UTRAN enforces cryptographic separation of E-UTRAN authentication vectors, i.e. it does not accept authentication vectors with the separation bit in the AMF set to 0. The network entities now behave from a security point of view according to the current TS 33.abc.

To set the HI bit on the USIM, it would be advantageous to be able to add / toggle the HI bit on already deployed USIMs using Over-The-Air (OTA) protocols. However, this approach has the limitation that some already deployed USIMs may not have the correct permissions to allow the flag to be provisioned using OTA techniques. Also some operators may not have an OTA server. If the USIM cannot be upgraded over the air the user will not enjoy the EPS security feature “cryptographic separation of E-UTRAN authentication vectors” as long as no new USIM is issued to him. But the user will be able to communicate over E-UTRAN using all the other features.

Variant of solution 6: Instead of upgrading the pre-Rel-8 HLR to an HSS the operator could also choose to upgrade it to one of the solutions 1 through 4 first and then later upgrade it to an HSS.

8.9.3 Distinguishing E-UTRAN authentication vector requests from other types

Solutions 1, 1b, 2, 3 and 6 assume that the HLR can distinguish between authentication vector requests from E-UTRAN and other authentication vector requests. In solutions 1 and 1b, this could be done based on the source address of the IWF. However, for solutions 2, 3 and 6 an approach based on source address would not be practical. For those solutions a better approach would be to indicate "E-UTRAN" in the Requesting Node Type of the MAP authentication vector request. This would require a change to the Rel-8 MAP protocol.

A more general issue not related to interworking with pre-Rel-8 HLR is that a combined SGSN/MME can have both E-UTRAN and UTRAN attached, and, in this situation, it must be possible for the HSS to distinguish authentication vector requests from the same SGSN/MME relating to E-UTRAN from those relating to UTRAN. This could be done by using different source addresses, but a better solution would be to explicitly indicate the type of authentication vector needed in the Diameter-based authentication vector request.

How to ensure that K_ASME derivation is performed exactly once?
In some of the solutions there may be two or more IWFs in the path between the MME and the HLR/HSS. An HSS performs K_ASME derivation, and it must be ensured for solutions 3, 4 and 6 that the IWF in the visited network does not perform K_ASME derivation a second time. Hence, the IWF must know whether the authentication vector was sent by an HSS or a pre-Rel-8 HLR according to one of the solutions 3, 4 and 6. This is illustrated in the figure below.

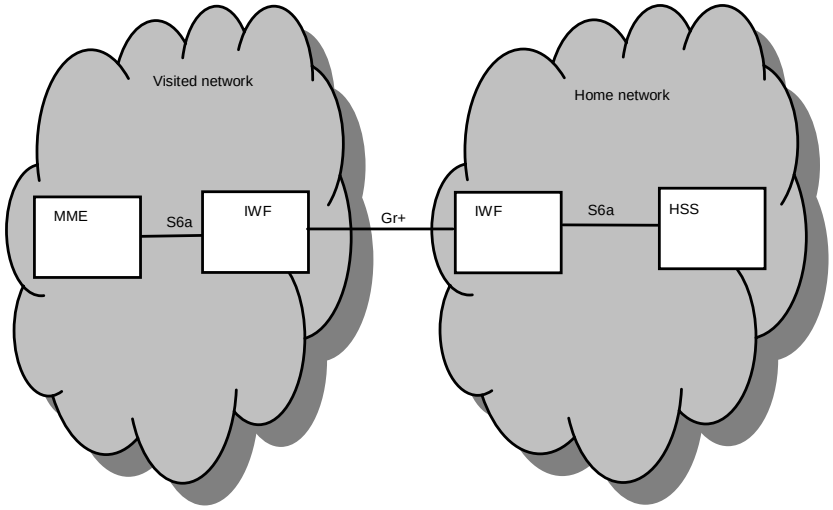


Figure 46 Two or more IWFs in the path between the MME and the HLR/HSS

Furthermore, it may be completely transparent to an MME whether protocol interworking is performed or not. Then the MME would need an indication whether the authentication vector was sent by an HSS or a pre-Rel-8 HLR even if only one IWF was present in the path. This is illustrated in the figure below.

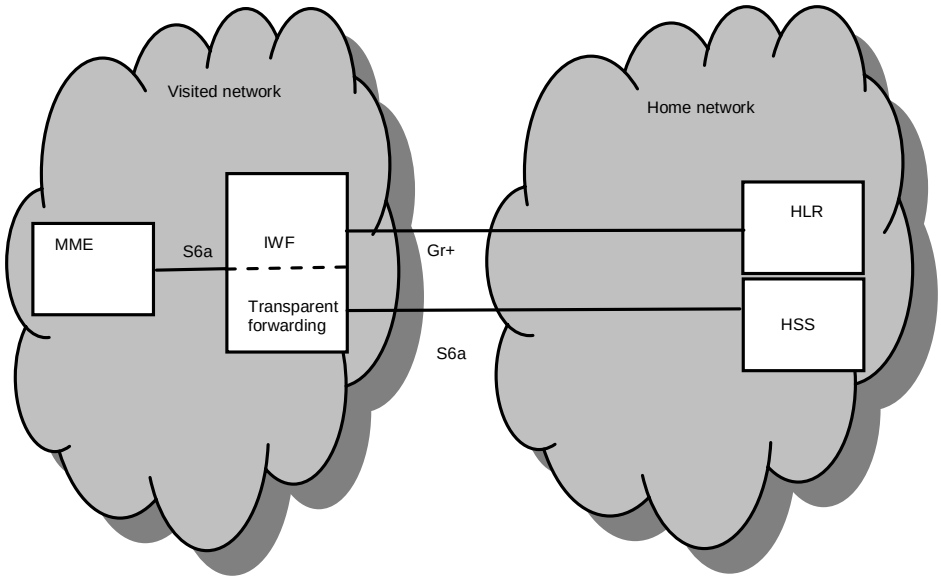


Figure 47 One IWF in the path between the MME and the HLR/HSS

A solution for the problem to ensure that K_ASME derivation is performed exactly once would be that entities which perform K_ASME derivation (HSS, IWF) indicate a corresponding capability in Gr+ and DIAMETER.

Also in order to avoid different type of IWFs (with and without K_ASME conversion) it could be useful to allocate K_ASME conversion functionality to the MME. (This would apply to solution 3, 4 and 6.)

8.9.4 Considerations on migration towards full security solution

The "full" security solution is one which provides both E-UTRAN serving network authentication and cryptographic separation of E-UTRAN authentication vectors, as for solution 1, and provides an S6a interface between MME and HSS.

For solutions 1, 1b, 2, 3 and 4, migration to the full security solution is possible for each HLR independently at any time. However, it is required that all pre-Rel-8 HLRs are upgraded as required by the respective solution before the start of EPS.

For solution 5, migration to the full security solution would be very difficult for the following reason: If, in a later release of EPS, cryptographic separation was to be introduced then this would require that post-Rel-8 MEs supporting this feature would have to reject authentication vectors with the separation bit in the AMF set to 0 when attached to E-UTRAN. Therefore, it would have to be ensured that, at the time when the first such upgraded ME was introduced into the system, all pre-Rel-8 HLR anywhere in the EPS would have been upgraded to Rel-8 HSS, otherwise access of these MEs to E-UTRAN might be impossible because pre-Rel-8 HLR may produce authentication vectors with the separation bit in the AMF set to 0. But if it is considered difficult to upgrade all pre-Rel-8 HLRs before the start of Rel-8 EPS for whatever reasons, it may be difficult for similar reasons to ensure this upgrade for all HLRs before the introduction of a later EPS release. Furthermore, Rel-8 MEs would be around for a long time, and hence an operator (visited or home) could never be sure that cryptographic separation was performed by the ME.

For solution 6, migration to the full security solution is possible for each HLR independently at any time. It is not required that all pre-Rel-8 HLRs are upgraded before the start of EPS. However, the EPS security enhancements will come into effect only when the USIM of a subscriber is also upgraded.

8.9.5 Evaluation of proposed solutions

The six solutions are compared in the table below.

	Solution 1: K_ASME derivation and protocol conversion in HPLMN	**Solution 1b: K_ASME derivation in HLR and protocol conversion in IWF in HPLMN**	**Solution 2: K_ASME derivation in HPLMN, protocol conversion in VPLMN**	**Solution 3: K_ASME derivation and protocol conversion in VPLMN (with dynamic setting of separation bit in HLR)**	**Solution 4: K_ASME derivation and protocol conversion in VPLMN (with static setting of separation bit in HLR)**	**Solution 5: UMTS security in E-UTRAN**	**Solution 6: Gradual upgrade of H (HI bit = 0)**
E-UTRAN serving network authentication	Yes	Yes	Yes	No	No	No	No
Cryptographic separation of E-UTRAN authentication vectors	Yes	Yes	Yes	Yes	No	No	No
Possibility for home network to upgrade security by upgrading to Rel-8 HLR	N/A	N/A	N/A	Yes	Yes	No	Yes
Impact on MAP	New AV requesting node type indication would be useful for Rel-8 nodes.	New AV requesting node type indication would be useful for Rel-8 nodes.	New AV requesting node type indication would be required for Rel-8 nodes.	New AV requesting and sending node type indication would be required for Rel-8 nodes.	New AV sending node type indication would be required for Rel-8 nodes.	No	New AV send node type indication wo be required fo Rel-8 nodes.
Impact on pre-Rel-8 HLR	Medium Dynamic setting of separation bit. Change of MAP.	Medium Dynamic setting of separation bit. K_ASME derivation. Change of MAP.	Medium Dynamic setting of separation bit. Change of MAP.	Medium Dynamic setting of separation bit. Change of MAP.	Small Static setting of separation bit.	No impact	No impact
Impact on other parts of home network	Medium IWF performs K_ASME derivation and protocol translation.	Low IWF performs protocol translation.	Low IWF performs K_ASME derivation.	No impact	No impact	No impact	No impact
Interoperability problems for UEs on visited EPS networks in the case that home network is not yet upgraded to support the interworking solution	No Visited network cannot contact home since home does not yet support Diameter-based roaming interface.	No Visited network cannot contact home since home does not yet support Diameter-based roaming interface.	Yes Visited network uses MAP-based roaming interface. Authentication vectors will be rejected by mobiles roaming in E-UTRAN.	Yes Visited network uses MAP-based roaming interface. Authentication vectors will be rejected by mobiles roaming in E-UTRAN.	Yes Visited network uses MAP-based roaming interface. Authentication vectors will be rejected by mobiles roaming in E-UTRAN.	No Visited network uses MAP-based roaming interface. Authentication vectors will be accepted by mobiles roaming in E-UTRAN.	No Visited netwo uses MAP-ba roaming interface. Authenticatio vectors will b accepted by mobiles roam in E-UTRAN.
Impact on visited network	No impact	No impact	Low IWF performs protocol translation.	Medium IWF performs K_ASME derivation and protocol	Medium IWF performs K_ASME derivation and protocol	Medium IWF performs K_ASME derivation and protocol	Medium IWF perform K_ASME derivation an protocol

				translation.	translation.	translation.	translation.

8.9.6 Conclusion

Solutions 1, 1b and 2 are preferred from a security point of view as the K_ASME derivation is done in HPLMN. Furthermore in solution1b the K_ASME derivation is done in HLR which is considered slightly more secure than 1 and 2; solution 1b also reduces the overall complexity of the IWF. Solutions 3 and 4 are less desirable from a security point of view, but they do at least allow the home operator to upgrade security later. Solution 5 is highly undesirable for SA3, since it would mean that EPS security enhancements that were previously agreed in SA3 would be completely abandoned. Furthermore, it would be very difficult with solution 5 to introduce these specific enhancements in later releases if they are not introduced in the first release of the EPS specifications.

One way of making a later introduction of EPS security enhancements possible is described in solution 6. Solution 6 has the advantage that operators may start into EPS without the need to upgrade the HLR, but has the disadvantage that the EPS security enhancements will come into effect only when the USIM of a subscriber is also upgraded. Note that for GSM and UMTS networks access with SIM was allowed, which is forbidden for EPS. Therefore upgrades from SIM to USIM to allow for E-UTRAN access is anticipated to be more frequent than was in the case of allowing UMTS-access.

If interworking with a pre-Rel-8 HSS/HLR whilst maintaining security is determined to be too difficult to achieve, then another option is to simply not allow interworking with a pre-Rel-8 HSS/HLR in EPS. This would of course be an acceptable solution from a security point of view.

9 Security Requirements for LTE eNBs

9.1 Terminology

This section defines the terminology as used in the subsequent section. So we spend some more time on discussing a suitable definition of *'secure environment'* in this contribution. As the LS from SA3#46bis S3-070283 to RAN2/3 describes, the term *'secure vault'* has been used by some companies in a similar context as *'secure environment'*. We prefer however to keep on working with *'secure environment'*. The word *'vault'* refers to a kind of secure storage function (i.e. a safe), while for LTE also some secured processing has to be performed inside. Also the term 'trusted environment' could be used e.g. see OMTP (http://www.omtp.org/docs/OMTP_Trusted_Environment_OMTP_TR0_v1.1.pdf) if it would be clear that no misunderstandings would arise from this and it would not violate any usage, disclosure and reproduction restrictions set by OMTP Ltd.

Proposed terminology:

Last-mile is the path or link from the eNB towards the physically secure core network (e.g. security gateway)

Physically secure means that attacker does not have physical access to the device/link

Physically insecure means that attacker can have access to the device/link

There seems to be different approaches to define a secure environment i.e. at very high level (see Wikipedia definition[55]) or at a very detailed level covering threats and protected assets (cfr OMTP). Going for the second approach will require a lot of SA3 effort, and is furthermore infeasible at this stage of standardization. The new proposal is a kind of middle way.

Proposed definition of secure environment:

---

55 In computing, a **secure environment** is any environment which implements the controlled storage and use of information. Often, secure environments employ cryptography as a means to protect information. Some secure environments employ cryptographic hashing, simply to verify that the information has not been altered since it was last modified.

A secure eNB environment is an environment which implements the controlled storage and processing of information. Dependent on the type of data that needs to be handled, this may or shall include protection against tampering of stored and processed information, and confidentiality protection. A secure environment does not prohibit multi-chip realizations, neither does it dictate certain architecture of communication busses and memory.

Editor's note: The definition of the term "*secure eNB environment*" needs further improvement.

9.2 eNB security requirements

NOTE 1: The eNB requirements are mostly identical with those presented at the SA3 adhoc although not all, i.e.

- Requirement 3 and 4 is the result of splitting and clarifying former requirement 3.
- a NOTE was added on Requirement 8

A) Requirements for eNB setup and configuration.

Setting up and configuring eNBs shall be authenticated and authorized so that attackers shall not be able to modify the eNB settings and software configurations via local or remote access.

1. Communication between the SAE core and the eNB shall be mutually authenticated.
2. Communication between the remote/local O&M systems and the eNB shall be mutually authenticated.
3. The eNB shall be able to ensure that software/data change attempts are authorized
4. The eNB shall use authorized data/software.
5. Sensitive parts of the boot-up process shall be executed with the help of the secure environment.
6. Confidentiality of software transfer towards the eNB shall be ensured.

B) Requirements for key management inside eNB

The SAE core network provides subscriber specific session keying material for the eNBs, which also hold long term keys used for authentication and security association setup purposes. Protecting all these keys is important.

7. Keys stored inside eNBs shall never leave a *secure environment* within the eNB eNB except when done in accordance with this or other 3GPP specifications.

C) Requirements for handling User plane data within the eNB

It is eNB's task to cipher and decipher user plane packets between the air interface and the *last-mile* link.

8. User plane data ciphering/deciphering shall take place inside the *secure environment* where the related keys are stored.
9. *The transport of user data over S1_U* shall be ciphered in case the *last-mile* link is *physically insecure*

NOTE 2: The use of ciphering on S1_U is an operator's decision. Various security configurations are possible for protection according to TS 33.210 (NDS/IP). In case the eNB has been placed in a physically secured environment then the 'secure environment' may include other nodes and links beside the eNB.

NOTE 3: SA3 aims for a single set of high level security requirements for all types of eNodeB (i.e. femto, pico and macro eNB). However, SA3 recognizes that different deployment environments dictate that different security solutions are needed to meet these requirements. SA3 has not yet agreed whether the requirements on the solutions for different deployment environments will be documented by 3GPP.

Annex A:
Decision made in RAN2/3-SA3 joint meeting in Jan 2006

A.1 RRC

Refer to A5.3 of S3-060119 (RAN2, RAN3 and SA3 joint meeting report from Sophia-Antipolis Jan 2006) [1]:

- “It was decided that RRC is always integrity protected.”
- "It was decided that a separate key set for RRC protection is necessary if RRC is terminated is in Node-B in order to prevent the derivation of NAS and User Plane keys. Keys per Node-B if RRC in Node-B TBD (TBD, SA3 to analyse if it is needed, answer by RAN Denver meetings latest ,else default in RAN group is no need)"
- "RRC protection resides in the node where RRC function terminates. i.e. if RRC is split in upper RRC and lower RRC then different security locations"
- "No identified show stopper in security vulnerability depending on the location for RRC => other criteria (cost complexity, performance, etc for overall RRC functions i.e. RB management, mobility, complexity/cost of security, etc) will be used for decision in RAN on RRC termination point(s). Conclusions will be provided to SA3 to continue joint work on security procedures"
- "RRC ciphering TDB (SA3)"
- "possibly user ID ciphering (scrambling) TBD (SA3 to investigate first)"
- "Allocation of IDs to be studied also (RAN2 will summarize information for SA3 and send it in an LS)"

Refer to chapter 1 of R3-060289 (LS from SA3#42 Bangalore on Feb 2006 to RAN2, RAN3, and SA2) [2]:

- “RRC ciphering and possibly user ID ciphering (scrambling). SA3 can’t decide now if RRC ciphering is needed without knowing the signalling messages and IDs used in RRC signalling. If there is need to protect the confidentiality of user IDs, there may be other ways than ciphering all RRC messages (potentially, by allocating IDs with a suitable scheme or only the identities themselves could be confidentially protected).”

As agreed in RAN plenary #31, the Evolved UTRAN functionality is distributed into eNBs.

A.2 MAC

Refer to A5.3 of S3-060119 (RAN2, RAN3 and SA3 joint meeting report from Sophia-Antipolis Jan 2006) [1]:

- “MAC security TBD (conclusion in April in SA3)”

Annex B:
Issues and Threats of emergency calls

B.1 General

The emergency call function is a highly important service that is required to work under almost all circumstances. This clause looks at DoS threats against the EC function and possibilities to protect against the threats and log attacks for post-fact analysis.

General DoS threats that are targeted at the GERAN/UTRAN/E-UTRAN and IMS subsystems are of course threats against the EC function, but this annex does not focus on the general case. Threats which are generally applicable are only mentioned in the context of EC.

Subclause B.2 looks at the threats from the perspective of the involved nodes.

Subclause B.3 looks at how network configuration and architecture can be used to limit the effects of DDoS attacks, in particular attacks from the Internet (compared to attacks originating from the access network).

Subclause B.4 looks at UE implementation aspects that could limit the possibilities for malware to be used for DDoS attacks from the UEs.

Other potential issues and threats that are not detailed further in this clause are:

- Mobility including I-RAT: For mobility within EPS and inter-RAT 33.401 has developed security solutions. This should be taken care of also for emergency calls.
- Man-in-the-middle by attacker acting as eNB or HeNB: In this case the user will be denied emergency services and the attacker could cause problem for the given subscriber or terminal if the network logs unwanted calls.
- False location: The UE could send false location information leading to major issues for emergency services this is already being observed for IP telephony services.

B.2 DoS threats against EC function

B.2.1 Threats against IMS nodes

IMS user agent (UE):

- According to clause 7.4 of TS 23.167 a UE shall not attempt to set up an anonymous emergency session to the same network again if it receives an error indication from the P-CSCF on the first try. This applies in the case the UE tries the emergency session setup without a prior emergency registration. The P-CSCF may send such an error based on local policy. This can be used by an attacker to send such an error to a UE, and the UE would not be able to establish the call.
    - **Spatial scope:** local to UE **Temporal scope:** persistent
    - **Protection:** Strict configuration of (emergency) PDN to ensure that IP layer attacks can not be achieved such as IP spoofing etc. Much harder to protect against attacks at the users premises in case the UE is not directly connected to EPS but, e.g., through a home GW (this might however be out of scope).
    - **Detection:** If the attack happens from the core network, normal logging procedures in different nodes can detect any attack attempts. If the attack happens in the users premises, it will not be easily detectable.
    - **Logging possibilities:** Possible in the core network.

P-CSCF:

- Overload P-CSCF with valid (emergency) registrations. This can be achieved by using dedicated emergency registrations or regular registrations. The attack can be launched as a DDoS from malicious software installed on legitimate UEs.

  - **Spatial scope:** local to serving NW **Temporal scope:** semi-persistent
  - **Protection:** Rate limiting on number of registrations allowed from a UE. Requiring dedicated emergency registration over emergency PDN only.
  - **Detection:** Counting number of registrations within a time period from a specific UE.
  - **Logging possibilities:** Log identities of registering UEs.

- Overload P-CSCF with emergency session requests. This can be achieved either when being registered as a regular emergency request, or be an unauthenticated emergency session request when UE is in limited service mode (if local regulations allow this). The attack can be launched as a DDoS from malicious software installed on legitimate UEs.
  - **Spatial scope:** local to serving NW **Temporal scope:** semi-persistent
  - **Protection:** Rate limiting on number of emergency session requests allowed from a UE (note that as no supplementary services are used, the current requirement is that a user will only have one active emergency session at time).
  - **Detection:** Counting number of (successful) emergency session requests within a time period from a specific UE.
  - **Logging possibilities:** Log identities and sessions of UEs.

S-CSCF:

- Overload S-CSCF with valid emergency registrations. The attack can be launched as a DDoS from malicious software installed on legitimate UEs.
  - **Spatial scope:** local to S-CSCF **Temporal scope:** semi-persistent
  - **Protection:** Rate limiting on number of registrations allowed from a UE. Should be coupled with rate limiting in P-CSCF as well.
  - **Detection:** See protection.
  - **Logging possibilities:** Log identities of registering UEs (not so helpful for anonymous registrations).

E-CSCF:

- Overload E-CSCF with emergency session requests. This can be achieved by using anonymous emergency session requests or regular ones. The attack can be launched as a DDoS from malicious software installed on legitimate UEs.
  - **Spatial scope:** local to serving NW **Temporal scope:** Semi-persistent
  - **Protection:** Rate limiting on number of emergency requests allowed from a UE (note that as no supplementary services are used, the current requirement is that a user will only have one active emergency session at time).
  - **Detection** Counting number of (successful) emergency requests within a time period from a specific UE.
  - **Logging possibilities:** Counting number of (successful) emergency requests within a time period from a specific UE.

LRF:

- Overload LRF with requests for UE locations. The LRF interface would probably only be accessible to a restricted set of trusted nodes (e.g., E-CSCF, PSAP), so this does not seem like a dangerous threat.
  - **Spatial scope:** Global **Temporal scope:** non-persistent
  - **Protection:** Restrict access to LRF to a limited set of trusted nodes.
  - **Detection:** Overload.
  - **Logging possibilities:** Log sources of requests.

B.2.2 Threats against EPS nodes

eNB:

- Crude radio jamming.
  - **Spatial scope:** local to eNB **Temporal scope:** non-persistent
  - **Protection:** Not possible

- **Detection:** Severe radio disturbance.
- **Logging possibilities:** Logging of time of attack.

MME:

- Overload MME with emergency bearer establishment requests.
  - **Spatial scope:** local to MME **Temporal scope:** semi-persistent
  - **Protection:** Rate limiting on number of emergency APNs per UE. For UEs which cannot be authenticated they could lie about their ID and make multiple requests. However, if the UE ID in the set up signaling is not possible to change via software, only physically hacked UEs can be used in a DDoS, i.e., malware is not sufficient. See UE implementation considerations below.
  - **Detection:** Overload.
  - **Logging possibilities:** Logging of times and sources of requests. In case of unauthenticated requests, request source logging is of limited value.

- Crude overload of MME with any type of NAS requests (e.g., Attach requests, bogus NAS messages).
  - **Spatial scope:** local to MME **Temporal scope:** semi-persistent.
  - **Protection:** Rate limiting/filtering of NAS messages from one UE. An attacker rapidly changing the UE ID and trying to overload the MME with NAS messages will probably first overload the eNB (needs to use one RRC connection per new UE ID, since the MME could filter out NAS messages with different UE IDs on the same S1 UE-connection). Again, a carefully implemented UE with limited access for applications to the radio APIs would limit the threat of a malware attack.
  - **Detection:** Overload.
  - **Logging possibilities:** Logging of times and sources of requests. In case of unauthenticated requests, request source logging is of limited value.

S-GW:

- Attacker injects bogus traffic on the Uu or S1-U interfaces.
  - **Spatial scope:** local to eNB/S-GW **Temporal scope:** non-persistent.
  - **Protection:** In case of existence of emergency calls, the eNB can make sure to carefully schedule traffic and not grant more traffic than it can handle for the emergency calls (the data rate required for emergency calls is not great). The same form of rate limiting can be performed by the S-GW if the attack is coming from the S1-U interface.
  - **Detection:** Overload.
  - **Logging possibilities:** Logging of times and sources of requests. In case of unauthenticated requests, request source logging is of limited value.

B.3 Protection via network configuration

None of the network nodes is immediately accessible from the Internet, save for the PDN gateway. This implies that DoS threats can be assumed to come from the access network. To protect the part of the EC function residing in the PDN gateway, the PDN gateway should be implemented in such a way that sufficient resources to handle the EC function are set aside. The PDN gateway is assumed to be connected to the PSAPs over a dedicated, trusted network.

In general, all nodes in the network must be provisioned to be able to cater for emergency call sessions from all connected UEs. Since there will be no more than one EC call session per UE, any additional EC call session set ups can be rejected by the network.

B.4 UE implementation considerations

The possibility to launch a DDoS attack by installing malicious software on the UEs is much dependent on the access to lower layer functions the UE makes accessible to applications.

To be able to setup a PS emergency call the UE must establish an emergency bearer with the network. If the UE does not export functions for emergency bearer establishment in the APIs visible to general applications, the risk of DDoS is severely reduced.

In general the more restricted the APIs to the lower layer radio functions are, the more risk is reduced.

If the UE implementation does not allow for setting the UE identity used in authentication etc from general purpose applications or only from a limited set of trusted applications, then masquerading during a DDoS will require a hardware modification of the UE. This seriously limits the effects that a malware DDoS can achieve.

Annex C:
Change History

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2006-02					Creation of document based on S3-060191.	0.0.0	0.0.1
2006-04					Including S3-060200, S3-060212, etc.	0.0.1	0.0.2
2006-07					Contributions in SA3 #44 added.	0.0.2	0.0.3
2006-11					Contributions in SA3 #45 added.	0.0.3	0.0.4
2006-12					Editorial changes (TR format).	0.0.4	0.0.5
2007-02					Contributions in SA3 #46 added.	0.0.5	0.1.0
2007-04					Contributions in SA3 SAE ah hoc added.	0.1.0	0.2.0
2007-05					Contributions in SA3 #47 added.	0.2.0	0.3.0
2007-07					Including S3-070511, S3-070518, S3-070523, S3-070529, S3-070555, S3-070574, S3-070617	0.3.0	0.4.0
2007-10					Including S3-070731, agreed parts of S3-070760, S3-070744, S3-070879, S3-070475	0.4.0	0.5.0
2007-12					Including S3a071021, S3a071031, S3a070927, S3a070938, S3a070953, S3a070917, S3a070928, S3a071039, S3a071040, S3a070929, S3a070925, S3a070955, S3a070918.	0.5.0	0.6.0
2008-02					Including S3-080044, S3-080046, S3-080058, S4-080054, S3-080178, S3-080067, S3-080060,S3-080123.	0.6.0	0.7.0
2008-04					Including S3-080369, S3-080381, S3-080498, S3-080320	0.7.0	0.8.0
2008-09					Including S3-081017.	0.8.0	0.9.0
2008-11					Including S3-081256, S3-081387 and S3-081396.	0.9.0	0.10.0
2008-12					MCC clean up for presentation to SA. Version 1.0.1 replaces previous erroneous version 1.0.0.	0.10.0	1.0.1
2009-01					Including S3-090103,	1.0.1	1.1.0
2009-03	SA-43	SP-0090132	--	--	Presentation to SA for approval	1.1.0	2.0.0
2009-03					SA approval	2.0.0	8.0.0
2009-06	SA-44	SP-090279	001	3	Emergency call aspects	8.0.0	9.0.0