3GPP TSG SA WG3 Security — S3#15 **S3-000623**

12-14 September, 2000

Washington D.C., USA

3G TR 33.8de V0.0.0 (2000-09)

Technical Specification

3rd Generation Partnership Project; Technical Specification Group SA3; Network domain security (Release 2000)

Keywords
Network Domain Security, Key management

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

An identified security weakness in 2G systems is the absence of security in SS7 networks. This was formerly perceived not to be a problem, since this network was the province of a small number of large institutions. This is no longer the case, and so there is now a need for security precautions. Another significant development has been the introduction of IP in the GPRS backbone network. The introduction of IP signifies not only a shift towards packet switching, which is a major change by its own accounts, but also a shift towards completely open and easily accessible protocols. The implication is that from a security point of view, a whole new set of threats and risks must be faced.

For 3G systems it is a clear goal to be able to protect the core network protocols, and by implication this means that security solutions must be found for both SS7 and IP based protocols.

Various protocols and interfaces are used for signalling in and between core networks. These include among the protocols MAP and GTP, among the interfaces A, Iu, and Iur, and possibly other protocols or interfaces that are new to R'00 or have yet to be identified. The security characteristics that have been identified as being in need of protection are confidentiality, integrity, and authentication. These will be ensured by standard procedures, based on cryptographic techniques.

1 Scope

The scope for this technical report is to define the requirements, functions and solutions for the network domain security of 3G mobile telecommunication systems. The TR focuses on new or modified functionality as compared to R99 and technical description of the features, functions and solutions of R00.

This interim/internal TR will act as a basis for the detailed Stage 2 specification work. Note that this is not a specification i.e. everything in this document may be changed at any time.

This TR is based on the contributions presented and approved at the SA3 meetings for the WI Network Domain Security and the WI Network Domain Security Key Managagement; see [3]-[n].

[We may also want to include the scope/results from the Denial-of-Service risk analysis WI that Motorola have proposed.]

According to the WI "Network Domain Security" the key objectives and the corresponding time plan are:

- Key objectives:
  - Developing security solutions for GTP
  - Developing security solutions for MAP-over-IP
  - Consider to develop security solutions for CAP
  - Consider to develop security for the interfaces where keys are transported (Iu, Iur and possibly A). The main motivation for developing security solutions for these interfaces will be to protect the transport of the keys CK/IK and Kc respectively. However, for the UMTS interfaces Iu and Iur one may want to develop a more complete solution. [have I got this right?]

One should note that the primary goal is to protect the control plane of the network domain. The user plane may also be covered where feasible. It is anticipated that user plane protection will not always be possible, and that for a number of cases it will not be necessary or indeed wanted.

- Timeplan (WI Network Domain Security):
  - August 2000, S3#14 Requirements capture (CAP, MAP-over-IP) / Feature specification of GTP security
  - September 2000, S3#15 Specification of security features for CAP, MAP-over-IP / Approval of GTP CRs
  - September 2000, SA#9 SA approval of GTP CRs to TS 33.102
  - November 2000, N4#5 N4 approval of GTP CRs
  - November 2000, S3#16 Feasibility study of proposals from this TR / Requirements capture for Iu/Iur interfaces
  - December 2000, CN#10 CN approval of GTP CRs
  -

  [… more to follow…]

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.
- For a non-specific reference, the latest version applies.

2.1 Normative references

[1] 3G TS 33.102: "3rd Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Architecture".

[2] 3G TR 23.821: "3rd Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; Architecture principles for Release 2000".

2.2 Informative references

[3] S3-000412 (Motorola): “A method to retain the Ipsec full security service in the three layer network domain security architecture”

[4] S3-000421 (Motorola): “Protect GTP signalling messages by IPsec”

[5] S3-000434 (Ericsson): “Principles for Core Network Security ” *(revised version)*

[6] S3-000444 (Siemens): “Core network security protocols ”

[7] S3-000511 (Telenor): “WI description: Network Domain Security ”

[Further references to be included, in particular NDS_KM references must be included]

2.3 RFC references

[8] RFC 1750 Randomness Recommendations for Security [is this the newest?]

[9] RFC 2401 Security Architecture for the Internet Protocol

[10] RFC 2402 IP Authentication Header

[11] RFC 2403 The Use of HMAC-MD5-96 within ESP and AH

[12] RFC 2404 The Use of HMAC-SHA-1-96 within ESP and AH

[13] RFC 2405 The ESP DES-CBC Cipher Algorithm With Explicit IV

[14] RFC 2406 IP Encapsulating Security Payload (ESP)

[15] RFC 2407 The Internet IP Security Domain of Interpretation for ISAKMP

[16] RFC 2408 Internet Security Association and Key Management Protocol (ISAKMP)

[17] RFC 2409 The Internet Key Exchange (IKE)

[18] RFC 2410 The NULL Encryption Algorithm and Its Use With IPsec

[19] RFC 2411 IP Security Document Roadmap

[20] RFC 2451 The ESP CBC-Mode Cipher Algorithms

[21] RFC 2521 ICMP Security Failures Messages

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply.

Confidentiality: The property that information is not made available or disclosed to unauthorised individuals, entities or processes.

Data integrity: The property that data has not been altered in an unauthorised manner.

Data origin authentication: The corroboration that the source of data received is as claimed.

Entity authentication: The provision of assurance of the claimed identity of an entity.

Key freshness: A key is fresh if it can be guaranteed to be new, as opposed to an old key being reused through actions of either an adversary or authorised party.

3.2 Symbols

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

Ga	Charging data collection interface between a CDR transmitting unit (e.g., an SGSN or a GGSN) and a CDR receiving functionality (a CGF).
Gb	Interface between an SGSN and a BSS.
Gc	Interface between a GGSN and an HLR.
Gd	Interface between a SMS-GMSC and an SGSN, and between a SMS-IWMSC and an SGSN.
Gf	Interface between an SGSN and an EIR.
Gi	Reference point between GPRS and an external packet data network.
Gn	Interface between two GSNs within the same PLMN.
Gp	Interface between two GSNs in different PLMNs. The Gp interface allows support of GPRS network services across areas served by the co-operating GPRS PLMNs.
Gr	Interface between an SGSN and an HLR.
Gs	Interface between an SGSN and an MSC/VLR.
Iu	Interface between the RNS and the core network. It is also considered as a reference point.
Iur	Interface between RNSs in the access network.

3.3 Abbreviations

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

AAA	Authentication Authorisation Accounting
AH	Authentication Header
AKA	Authentication and key agreement
CS	Circuit Switched
ESP	Encapsulating Security Payload
GGSN	Gateway GPRS Support Node
HLR	Home Location Register
IKE	Internet Key Exchange
MAC	Message Authentication Code
ME	Mobile Equipment
MS	Mobile Station
MSC	Mobile Services Switching Centre
PS	Packet Switched
RNS	Radio Network Subsystem
SGSN	Serving GPRS Support Node
UE	User Equipment

UICC	UMTS IC Card
USIM	User Services Identity Module
VLR	Visitor Location Register

4 Security threats and requirements

4.1 Threats to the network domain

[A large portion of this text is just lifted from 21.133 (section 6.2 and 8.1) – it ought to be refined and adapted in later versions of this TR]

A number of threats to the network domain exist. A broad classification into active and passive attacks can be made and table 1 gives a non-exhaustive overview of these. Furthermore the network domain can effectively be divided into a CS/SS7 and an IP part since both the threats and requirements may depend on the network layer transportation method.

Table 1: Classification of threats

	Active attack		Passive attack	
	IP	**CS/SS7**	**IP**	**CS/SS7**
1) Unauthorised access to data	X	X	X	X
2) Threats to integrity	X	X	-	-
3) Denial of service	X	X	-	-
4) Repudiation	X	X	X	X
5) Unauthorised access to services	X	X	-	-

4.1.1 Unauthorised access to data

[This section is lifted directly from 21.133]

T5a **Eavesdropping user traffic:** Intruders may eavesdrop user traffic on any system interface, whether wired or wireless.

T5b **Eavesdropping signalling or control data:** Intruders may eavesdrop signalling data or control data on any system interface, whether wired or wireless. This may be used to access security management data which may be useful in conducting other attacks on the system.

T5c **Masquerading as an intended recipient of data:** Intruders may masquerade as a network element in order to intercept user traffic, signalling data or control data on any system interface, whether wired or wireless.

T5d **Passive traffic analysis:** Intruders may observe the time, rate, length, sources or destinations of messages on any system interface, whether wired or wireless, to obtain access to information.

T5e **Unauthorised access to data stored by system entities:** Intruders may obtain access to data stored by system entities. Access to system entities may be obtained either locally or remotely, and may involve breaching physical or logical controls.

T5f **Compromise of location information:** Legitimate user of a 3G service may receive unintended information about other users locations through (analysis of) the normal signalling or voice prompts received at call set up.

4.1.2 Threats to integrity

[This section is lifted directly from 21.133]

T6a **Manipulation of user traffic:** Intruders may modify, insert, replay or delete user traffic on any system interface, whether wired or wireless. This includes both accidental and deliberate manipulation.

T6b **Manipulation of signalling or control data:** Intruders may modify, insert, replay or delete signalling or control data on any system interface, whether wired or wireless. This includes both accidental and deliberate manipulation.

T6c **Manipulation by masquerading as a communications participant:** Intruders may masquerade as a network element to modify, insert, replay or delete user traffic, signalling data or control data on any system interface, whether wired or wireless.

T6d **Manipulation of applications and/or data downloaded to the terminal or USIM:** Intruders may modify, insert, replay or delete applications and/or data which is downloaded to the terminal or USIM. This includes both accidental and deliberate manipulation.

T6e **Manipulation of the terminal or USIM behaviour by masquerading as the originator of applications and/or data:** Intruders may masquerade as the originator of malicious applications and/or data downloaded to the terminal or USIM.

T6f **Manipulation of data stored by system entities:** Intruders may modify, insert or delete data stored by system entities. Access to system entities may be obtained either locally or remotely, and may involve breaching physical or logical controls.

4.1.3 Denial of service attacks

[This section is lifted directly from 21.133]

T7a **Physical intervention:** Intruders may prevent user or signalling traffic from being transmitted on any system interface, whether wired or wireless, by physical means. An example of physical intervention on a wired interface is wire cutting. An example of physical intervention on a wireless interface is jamming. Physical intervention involving interrupting power supplies to transmission equipment may be conducted on both wired and wireless interfaces. Physical intervention may also be conducted by delaying transmissions on a wired or wireless interface.

T7b **Protocol intervention:** Intruders may prevent user or signalling traffic from being transmitted on any system interface, whether wired or wireless, by inducing protocol failures. These protocol failures may themselves be induced by physical means.

T7c **Denial of service by masquerading as a communications participant:** Intruders may deny service to a legitimate user by preventing user traffic, signalling data or control data from being transmitted by masquerading as a network element to intercept and block user traffic, signalling data or control data.

T7d **Abuse of emergency services:** Intruders may prevent access to services by other users and cause serious disruption to emergency services facilities by abusing the ability to make USIM-less calls to emergency services from 3G terminals. If such USIM-less calls are permitted then the provider may have no way of preventing the intruder from accessing the service.

4.1.4 Repudiation

[This section is lifted directly from 21.133]

T8a **Repudiation of charge:** A user could deny having incurred charges, perhaps through denying attempts to access a service or denying that the service was actually provided.

T8b **Repudiation of user traffic origin:** A user could deny that he sent user traffic

T8c **Repudiation of user traffic delivery:** A user could deny that he received user traffic

4.1.5 Unauthorised access to services

[This section is lifted directly from 21.133]

T9a **Masquerading as a user:** Intruders may impersonate a user to utilise services authorised for that user. The intruder may have received assistance from other entities such as the serving network, the home environment or even the user himself.

T9b **Masquerading as a serving network:** Intruders may impersonate a serving network, or part of an serving network's infrastructure, perhaps with the intention of using an authorised user's access attempts to gain access to services himself.

T9c **Masquerading as a home environment:** Intruders may impersonate a home environment perhaps with the intention of obtaining information which enables him to masquerade as a user.

T9d **Misuse of user privileges:** Users may abuse their privileges to gain unauthorised access to services or to simply intensively use their subscriptions without any intent to pay.

T9e **Misuse of serving network privileges:** Serving networks may abuse their privileges to gain unauthorised access to services. The serving network could e.g. misuse authentication data for a user to allow an accomplice to masquerade as that user or just falsify charging records to gain extra revenues from the home environment.

4.2 Security requirements

4.2.1 Requirements on system integrity

[This section is lifted directly from 21.133. The underlying assumption here is that system/control plane integrity is the most important part of NDS. We may want to elaborate on this assumption]

R3a It shall be possible to protect against unauthorised modification of user traffic. (T2a, T6a,c, T7b,c)

Note: It is assumed that user traffic contains sufficient redundancy such that a stream cipher provides a basic level of data integrity protection on the radio interfaces and that, if that is not sufficient and additional measures are required, the application should be aware and measures should be implemented at the application layer.

R3b It shall be possible to protect against unauthorised modification of certain signalling data and control data, particularly on radio interfaces. (T2b, T3b,c, T6b,c, T7a,b,c)

R3c It shall be possible to protect against unauthorised modification of user-related data downloaded to or stored in the terminal or in the USIM. (T6d,e, T6c, T10f,i)

R3d It shall be possible to protect against unauthorised modification of user-related data which is stored or processed by a provider. (T6c,f)

R3e It shall be possible to ensure that the origin and integrity of applications and/or data downloaded to the terminal and/or the UICC can be checked. It may also be necessary to ensure the confidentiality of downloaded applications and/or data. (T6c,d,e,f, T10e,f,i)

R3f It shall be possible to ensure the origin, integrity and freshness of authentication data, particularly of the cipher key on the radio interface. (T1a,b, T2b, T5c, T6c)

R3g It shall be possible to secure infrastructure between operators. (T5a,b,c, T6a,b,c, T7a,b,c, T9b,c)

5 Security architecture for the network domain

5.1 Protocols and interfaces covered by the network domain

[The following is a first attempt to give an overview of the protocols and interfaces that we consider to be part of NDS. The table below is not complete, but could serve as a starting point for analysis of NDS protocols/interface related security issues. Abis is deliberately not included. Iur should be considered to belong to the access network, but is included since CK/IK apparently is transported over that interface]

Table 11: Overview of the interfaces and protocols in the network domain

Interface	Protocols etc
A	Interface between MSC and BSC. Protocols over A include: - MTP, SCCP, BSSAP (which includes DTAP and BSSMAP)
Ga	Charging data collection interface between a CDR transmitting unit (e.g., an SGSN or a GGSN) and a CDR receiving functionality (a CGF).
Gb	Interface between an SGSN and a BSS. Notice that user plane and associated control plane data is protected by encryption between SGSN and the MS. Security parameters are not transported over Gb. Protocols over Gb include: - For the user plane: Frame Relay, BSSGP, LLC, SNDCP and IP - For the control plane: Frame Relay, BSSGP, LLC, GMM/SM
Gc	Interface between a GGSN and an HLR. This interface is optional, and the GGSN can route the required signalling towards HLR via a SGSN. Protocols over Gc: - MAP/SS7 - In the future: MAP/IP ??
Gd	Interface between a SMS-GMSC and an SGSN, and between a SMS-IWMSC and an SGSN. Protocols over Gd: - SMS-TL carried by MAP/SS7 - In the future: SMS-TL carried by MAP/IP ??
Gf	Interface between an SGSN and an EIR. Protocols over Gf: - MAP/SS7 - In the future: MAP/IP ??
Gi	Reference point between GPRS and an external packet data network. Protocols over Gi: - User plane IP - AAA signalling (typically RADIUS)
Gn	Interface between two GSNs within the same PLMN. Protocols over Gn: - Lower layer IP - GTP-C - GTP-U
Gp	Interface between two GSNs in different PLMNs. The Gp interface allows support of GPRS network services across areas served by the co-operating GPRS PLMNs. Protocols over Gn: - Lower layer IP - GTP-C - GTP-U
Gr	Interface between an SGSN and an HLR. Protocols over Gc: - MAP/SS7

	- In the future: MAP/IP ??
Gs	Interface between an SGSN and an MSC/VLR Protocols over Gc: - MAP/SS7 - In the future: MAP/IP ??
Iu	Interface between the RNS and the core network. It is also considered as a reference point. Protocols over Iu: - User plane PS: ATM, AAL5, UDP/IP, GTP-U - Control plane PS: ATM, AAL5, SSCOP, SSCF-NNI, MTB3b, SCCP, RANAP, GMM/SM/SMS - Control plane PS (alternatively): ATM, AAL5, UDP/IP, SCTP, ITUN, SCCP, RANAP, GMM/SM/SMS - User plane CS: ATM, AAL2, … - Control plane CS: ATM, AAL5, SSCOP, SSCF-NNI, MTB3b, RANAP, GMM/SM/SMS
Iur	Interface between RNSs in the access network Protocols over Iur: - Control plane: same as for Iu-PS upto SCCP. Above SCCP one will find RNSAP - User plane: ATM, AAL2, …

5.2 Principles for Network Domain Security

[This section is based rather directly on the revised version of Erissons paper S3-00434 "Principles for Core Network Security".]

5.1.1 Introduction

The scope of this section is to outline the basic principles on which a security architecture for network domain should be based.

With the introduction of IP based transport to most, if not all, interfaces of the 3GPP specified network reference model follows new vulnerabilities of the network as well as new potential threats directed towards the network from outside. Instead of building, and managing, their own “private” transport networks, operators have a possibility to rent the transport capacity required between any two nodes of the reference model from virtually any ISP. Similarly also inter-network communications should not be considered unlikely to exploit the already existing transport network commonly known as Internet.

The most obvious security issue with such a view is that virtually any network connection could, in some sense, be considered “publicly” accessible and thus possible to exploit not only with the purpose of eavesdropping and fraud, but also with the purpose to attack the very business or reputation of the operator by means of e.g. hi-jacking, halting or in other ways disturbing the packet flow over such a connection.

The following sections discuss a basic architecture designed to support protected inter-network communications considering a scenario like the one described above. The very same principles might be applied, though, also for connections between e.g. two geographically separated sites within the same network.

5.1.2 The Security Architecture

When IP based transport is introduced to the cellular networks one could say that the network concept shifts from being a telecom centric one to a more datacom centric concept. The telecom network concept comprises fairly “private” networks initially built and maintained by a limited number of national telcos with their own infrastructure. In such a network based on timeslots or “fixed” communication lines (such as e.g. leased lines or AMT VPCs) it is relatively easy to employ strict access control.

The datacom centric concept, on the other hand, is to a much bigger extent built on available, often fairly “public”, transport. In such a network there is no way to distinguish one packet from another unless special care has been taken during the design of the network. Furthermore it was once designed to autonomously find a viable communication path

between two points of the network, which makes it much more difficult to force the packet flow through special point where e.g. access policy can be enforced.

When designing the security architecture for this “new” type of cellular network, it is wise to base it on the already existing knowledge from the datacom industry of today. The cellular network can from a security point of view be considered analogue to a corporate intranet. The entrance point of today’s corporate networks typically consist of an “air-gap” architecture with logically two firewalls creating a so called demilitarized zone, DMZ, between them. In this zone are resources required to be accessible from outside placed. In the corporate datacom world such resources typically includes e.g. a web-server and a mail-server.

We envision a similar entrance point also in the security architecture of the new IP based cellular networks. In this zone, which in the 3GPP architecture is denoted Extranet, resources like KACs, NATs, DNSes, I-CSCFs and other types of proxies should be placed. In order to get a simpler architecture we propose to introduce a new entity, the Security Gateway, SGW.

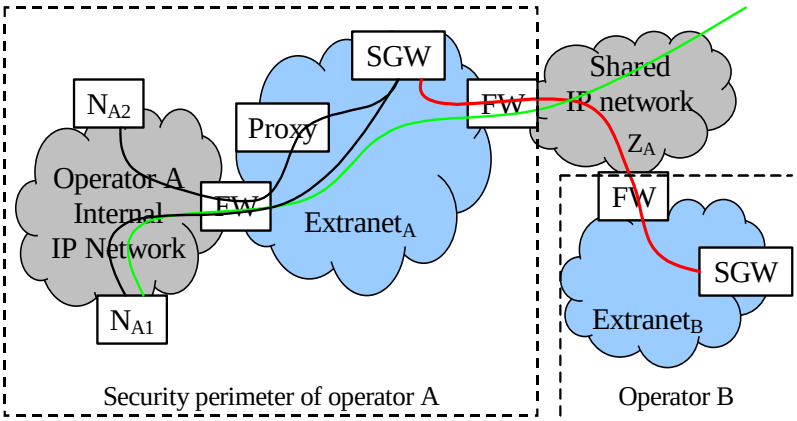

Fig. 1 An example security architecture

This SGW should be seen upon as a new node placed on the border of the network with the task to enforce the security policy regarding in- and outbound communications of the network. Typical responsibility/functionality of the SGW would thus include:

- Negotiation, establishment and maintenance of the security relationships (Security Associations, SAs) with other networks, represented by their respective SGW, by usage of IKE.
- Establishment and maintenance of the secured IP tunnels that realizes the negotiated SAs between two networks. IPsec is to be used for this purpose.
- Establishment and maintenance of other more optimal paths for certain packets flows as allowed by the security policy of the operator in question, e.g. packets from a subscriber desiring a raw, “open” and direct connection to Internet could be tunnelled directly between the two firewalls as indicated by the green line in the figure above.

Note that the SGW, as opposed to the previously introduced KAC (see e.g. T-Doc S3-000432), is able to maintain several different SAs where each is tailored for a certain type of packet flow, i.e. MAP could use a separate SA in respect to GTP-C.

Since the SGW internally handles all data in clear, i.e. unless the traffic is truly end-to-end encrypted, this would be a natural point to also locate functionality necessary to support eventual Lawful Interception requirements.

5.1.3 Reflections

The security architecture outlined above, including the introduction of an Extranet together with the proposed new SGW entity, comes with several important benefits, such as:

- The security strategy regarding the intra-network connections remains clearly separated from the inter-network security strategy. This allows an operator to independently choose his/her own security strategy for the internal network, while still maintaining inter-operability with e.g. roaming partners by adopting the proposed architecture for the inter-network communications.
- Many “original” network elements/nodes can be leveraged from the processing burden and complex functionality imposed by many security functions and procedures, such as encryption, decryption, authentication etc.
- Due to the network-to-network approach of the architecture, as opposed to a generic node-to-node approach, the total number of required keys to manage decrease significantly, which allows for an operator to start off with a simple pre-shared keys strategy and wait with the deployment of PKI till a later stage.
- The point for key management as well as policy enforcement in this architecture is centralized, i.e. in the SGW(s), which makes operation and maintenance easier to handle.
- The proposed security architecture can be seen as a natural migration from the architecture presented in the key management solution for MAP as proposed by Ericsson (see T-Doc S3-000432), which ensures the ability to still employ node-to-node security in cases where this would be preferred.

5.3 Lawful interception

[I don't know 33.106/33.107 very well and don't really know if it's appropriate to separate LI in a separate section, but obviously we must take LI into account at some stage. Suggestions on how to deal with LI is welcomed]

6 Key management and distribution in the network domain

[~~This section is included as a placeholder for input from WI NDS_KM~~Text to be included from documents identified below]

6.1 Working assumptions

1. A two-tiered key management architecture should be adopted in the first phase. Migration to a PKI-based flat key management will be considered for later phases.

2. IP-based communications secured using IPsec should be used for layer 1 and layer 2 which implies that all NEs and KACs support an IP stack.

3. IKE shall be used as the basis for key management (but some open issues need to be resolved - see below).

4. For communications secured using IPsec, the IETF IPsec security association will be adapted/profiled for 3GPP. For communications secured at the application layer, 3GPP will define new security associations (i.e. create a new DOI for ISAKMP). A first attempt at specifying a security association for MAPsec is given in S3-000433.

6.2 Open issues

1. Establishment of layer 3 SAs between KACs

There are two options: a) IKE is used; b) IKE is used between KACs to establish layer 1 SAs for IPsec between the KACs and another protocol is then used between the KACs to establish layer 3 SAs. Both options are described in S3-000445. The first option is favoured in S3-000432 (see start of section 4.3.1).

2. Establishment of layer 2 SAs between KACs and NEs

There are three options: a) IKE is used which implies that all NEs must support IKE; b) another protocol is used which implies a large specification effort; c) manual establishment which may be acceptable in the initial phase but which will require a proprietary anti-replay mechanism to be used. All options are described in S3-000445. The first option is favoured in S3-000432 (see end of section 4.3.4).

3. Distribution of layer 3 SAs to NEs

There are two options: a) a "push" approach using LDAP; b) a "pull" approach using SNMP. A preference for the first option is indicated in S3-000432. S3-000445 does not describe any options.

4. Achieving anti-replay protection at layer 3 for IPsec case

There are two options: a) a proprietary anti-replay mechanism is used; b) IKE based on the pre-shared secret established by layer 1/2 is used to dynamically negotiate SAs between NEs. Both options are described in S3-000445. S3-000432 does not describe any options.

6.3 Workplan

S3#15

- Resolve all the open issues listed above.

S3#16

- Write the stage 2 description of the IKE-based key management architecture.
- Select algorithms to be supported for MAPSec.
- Write MAPsec DOI for ISAKMP. Find out if it has to be an RFC.
- Agree on standard profiles of MAP-PP.
- Agree on use and format of SPI in the MAPSec component headers.
- Define database formats for SPD and SADB.
- Select algorithms to be supported for IPsec.
- Adapt IPsec DOI for ISAKMP. Find out if a new RFC is required.
- Select algorithms to be supported for CAP.
- Write CAPsec DOI for ISAKMP. Find out if it has to be an RFC.
- Agree on standard profiles of CAP-PP.

7 Security mechanisms

7.1 The use of IPsec

[This section is based on the S3-000421 input paper from Motorola]

IPSec mainly consists of IP Authentication Header (AH) and IP Encapsulating Security Payload (ESP). The Authentication Header provides data integrity, data origin authentication, and optional limited anti-replay services to IP.

Encapsulating Security Payload provides confidentiality, data origin authentication, anti-replay, data integrity, and limited traffic flow confidentiality.

The concept of Security Association (SA) is fundamental to IPSec. The SAs are unidirectional contracts between two communication entities. SAs determine the IPSec protocols used for securing the packets, the algorithms, the keys, and the duration for which the keys are valid. A “Security Association Database” (SADB) maintains the SAs.

Both AH and ESP make use of SAs. A key management protocol IKE is employed to establish and maintain SAs. For UMTS purposes we assume that SAs are established by Key Administration Centers (KACs) defined in TS 33.102, so that both SADB and SPD are established for UMTS use. The security services afforded in the SAs will be applied to UMTS protocols as needed.

For control plane signalling messages, the provision of the following security services is suggested:

- data integrity;
- data origin authentication;
- confidentiality; and
- anti-replay.

Use of ESP, AH, as well as any possible combination of them, should conform to standard IPSec to the extent possible.

Two types of SAs are defined in IPSec:

- **transport mode**
  In this mode only the higher layer protocols (transport and above) is protected by IPSec

- **tunnel mode**
  Tunnel mode provides for protection of the entire IP datagram, including the full original IP header information.

According to RFC 2401, “a host MUST support both transport and tunnel mode. A security gateway is required to support only tunnel mode. If it supports transport mode, that should be used only when the security gateway is acting as a host, e.g., for network management.”

For UMTS control plane messages, a host-to-host SA can be either transport mode or tunnel mode. However, whenever at least one end is a gateway, then it must be in tunnel mode. Furthermore, tunnel mode would provide source and destination address confidentiality. It must be noted that the use of tunnel mode is largely incompatible with the use of NATs. This can be a problem in Ipv4 networks since address space is scare and NATs is common way of solving the address space problem.

7.2 Router filtering and firewalls

ffs

8 Security for the Network Domain protocols and interfaces

8.1 Security for the MAP protocol

[This section is based on the Siemens input paper "Core network security protocols" (S3-000444). The Siemens paper also briefly mentions CAP, but this part has been left out.]

8.1.1 Introduction

Prior to Release99, an SS7 based protocol stack always carried the MAP protocol. Starting from Release00, transport of MAP in a UMTS core network may be either based on SS7 or on IP. Therefore, in the medium term, entities in the UMTS core network must be able to support the following protocol stacks:

- MAP over SS7 (for short: MAP/SS7)
- MAP over IP (for short: MAP/IP)
- Application protocols over IP transport with no equivalent protocol over SS7 transport. (for short: native IP-based protocols).

It may well happen that a single platform communicates over both SS7 and IP, e.g. the HSS. Examples for the nodes supporting only native IP-based protocols include the CSCF.

For native IP based protocols the decision on where in the protocol stack the security functionality should be applied is a fairly easy one as a network layer solution (IPsec) that will work for all IP based protocols is available.

When trying to protect protocols that may be carried by both SS7 and IP, it is no longer obvious that a network layer solution is the best choice.

In these cases, core network security can be provided either at the application layer or at a lower layer. If security is provided at the application layer we denote this by the term application layer security. If IPsec provides security at the (network) IP layer, we use the term IP security.

The following (necessarily symmetric) matrix shows which type of network entity needs to communicate with which other type in the medium term:

	MAP/SS7	MAP/IP	Native IP
MAP/SS7	Yes	Yes	No
MAP/IP	Yes	Yes	No
Native IP	No	No	Yes

8.1.2 Security for MAP

Although this section is discussing MAP, one should be aware that the same type of arguments applies to CAP.

It is assumed here that, for a long period after the introduction of IP as the transport for MAP, MAP/IP nodes (e.g. VLRs in network 1) need to be able to communicate with MAP/SS7 nodes (e.g. HLRs in network 2).

For MAP/IP security, there are basically two options:

- Security on the MAP application layer
- IP security (network layer security)

If IP security is used, the need for application-to-network layer security gateways (ANLSG) arise when interworking between IP and SS7 transport becomes necessary. Such an application-to-network layer gateway would have to translate application layer MAP security (in the SS7 domain) into network layer security (in the IP domain). This is highly undesirable for several reasons:

- There is high additional complexity introduced by such a gateway
- To receive protected MAP messages and to transform them into IPsec secured messages, a SS7/IP gateway must be capable of terminating application layer (MAP) security on the SS7 side. Since MAP routing is based on the IMSI number and does not happen at the MAP-layer, an SS7 end-entity cannot directly address (and usually does

not even know) gateways at the network layer or other MAP entities. Therefore, it seems to be difficult to set up a MAP security association between a MAP end-entity and an ANLSG.

- The trust issues raised by this solution are difficult. The endpoints of the MAP communication would have to trust the ANLSG. But how can a MAP/SS7 node even know, which gateway the MAP messages pass? ANLSGs could even be located in intermediate networks, e.g. if the originating network has no direct link to the IP world. So ANLSGs were likely to influence and even restrict the worldwide PLMN topology, for guaranteeing a closed chain of trust between all communicating MAP entities.
- An ANLSG would seem to contradict the principle of a separation between transport stack and application.

This speaks in favour of providing security at the application layer also for MAP/IP.

On balance, this scenario isn't the only one. One may also require that all networks support both secure SS7 and IP based MAP version. This scenario does have some advantages like allowing for rapid transition toward IP-only networks, but it comes at the prohibitive cost of having two implementations for securing MAP.

A serious drawback of the application layer security approach is that every application protocols must be separately adapted to provide the desired security. In reality, this is not a serious problem as only MAP and CAP have been identified as targets for application layer security. Furthermore, the work to adapt MAP has already taken place.

Security for protocols that can be carried by both SS7 and IP should therefore be secured on the application layer. An additional advantage of this approach is that no additional specification and implementation effort is foreseen for MAP security when IP-based transport for MAP is introduced.

Consequently, security for MAP and possibly CAP shall be provided at the application layer.

[Since we (at least I) haven't seen the specification for the MAP/IP implementation, we should be prepared for at least some minor incompatibilities between MAP/SS7 and MAP/IP. If any of these are security related we must be sure to document them]

8.2 Security for the GTP protocol

[This section is almost entirely based on the S3-000421 paper by Motorola. Only minor modification have been made. Some of the contents from S3-000421 is moved to 7.1 "The use of Ipsec"]

8.2.1 Introduction

GPRS Tunnelling Protocol (GTP) is defined in 3G TS 29.060 v3.5. It includes both the GTP control plane signalling (GTP-C) and user plane data transfer (GTP-U) procedures. GTP is defined for Gn interface, i.e. the interface between GSNs within a PLMN, and for the Gp interface between GSNs in different PLMNs.

Some mobility management messages accommodated in GTP-C include sensitive information, for example, authentication vectors and MM context. Therefore, it is necessary to apply security protection to GTP signalling messages (GTP-C). GTP-U is the tunnelling part of GTP, and as such GTP-U will itself carry user plane IP in its payload. Special care is required when applying security protection to GTP-U in order not to unnecessary duplicate protection on the two IP layers.

GTP uses UDP/IP path to transfer GTP signalling messages as well as to tunnel user data packets. IPSec is a set of protocols that integrate security into IP and provide data source authentication, data integrity, confidentiality, and protection against replay attacks. Therefore, IPSec is a natural candidate to provide protection for GTP messages.

8.2.2 Use of IPSec to protect GTP signalling messages

It is possible to protect both GTP signalling messages (GTP-C) and user data packets (GTP-U) by IPSec. However, in most of the applications, the user data may be protected by higher layer security mechanisms. It is not efficient or may not be necessary to apply double protection to user data.

Therefore it's generally assumed that the protection provided by IPSec only apply to GTP control plane signalling messages (GTP-C). Nevertheless, protection of GTP-U traffic by IPSec remains as an option for the network operators.

8.2.3 No changes in GTP messages

IPSec is independent of any higher layer protocols. Therefore, it does not require any changes to the GTP messages. This differs from the approach taken to add security to MAP messages (in TS 29.002), where the security functions were applied at the application layer.

8.2.4 Error and failure handling

In RFC 2521, a set of ICMP messages are defined to deal with the errors and failures in using AH and ESP. The new ICMP messages defined include “bad SPI”, “authentication failure”, “decompression failure”, “decryption failure”, “need authentication”, and “need authorization”.

IPSec error status may be conveyed to the sender by means of a local network management function. This function is beyond the scope of GTP standardization.

8.2.5 IPSec SPD and its implication to GTP message protection

In the IPSec architecture, a look-up table, called Security Policy Database (SPD), is used to discriminate among traffic that is afforded IPSec protection and traffic that is allowed to bypass IPSec.

For any inbound or outbound datagram, three processing choices are possible:
- discard
- bypass,
- apply IPSec.

We recommend that GTP-U packets simply “bypass” the IPSec process and that GTP-C packets be afforded protection by the IPSec.

SPD specifies what security services are to be applied to an IP datagram based on a set of selectors, among which the most important ones are source IP address, destination IP address, source UDP port, and destination UDP port. The SPD must be consulted during the processing of all traffic (inbound and outbound), including non-IPSec traffic.

By using SPD, different security mechanisms can be applied to GTP-C messages and GTP-U messages, since they use different UDP ports according to the latest version of 29.060.

- **GTP-C**

  REQEST: The UDP Destination Port number for GTP-C is 2123. The UDP Source Port is a locally allocated port number at the sending GSN.

  RESPONSE: The UDP Destination Port value shall be the value of the UDP Source Port of the corresponding request message. The UDP Source Port shall be the value from the UDP Destination Port of the corresponding request message.

- **GTP-U**

  The UDP Destination Port number for GTP-U is 2152. The UDP Source Port is a locally allocated port number at the sending GSN

For Release99 and newer versions of GTP, this will allow us to apply IPSec mechanisms to only GTP-C messages. For pre-Release99 versions of GTP no port number distinction between GTP-C and GTP-U is made, and both GTP-C and GTP-U uses port number 3386. Notice that for pre-Release99, support for TCP is also defined for GTP-U over port 3386.

It may be possible to further classify GTP-C messages so that they can be protected by different security mechanisms, but it is a local matter for the application layer to signal the IPSec processing for selection of security mechanisms on a message-by-message basis.

8.2.6 IPSec protocols and applications to GTP messages

For GTP control plane signalling messages, the provision of the following security services is suggested:

- data integrity;
- data origin authentication;
- confidentiality; and
- anti-replay.

For GTP control plane messages, a host-to-host SA can be either transport mode or tunnel mode. However, whenever at least one end is a gateway, then it must be in tunnel mode. Furthermore, tunnel mode would provide source and destination address confidentiality.

Security services will probably also be needed to extend over the Gp interface, and thereby passing the Border Gateway (BG). This will imply mandatory support of tunnel mode.

8.3 Security for the interfaces Iu and Iur

[This will at least imply protection for parts of RANAP. We must also decide whether both CS and PS is to be targeted. That is, we don't really care about the CS/PS distinction for signalling protocols, but signalling protocols that control CS user plane tends to be SS7 based themselves. For "CS" it may mean protection of SS7 based RANAP protocol; this again suggest that RANAP may be easiest protected on the application layer since may have to deal with both SS7 and IP serving as network layers.

I really haven't looked at Iur at all, but I suspect that the same scenario as for RANAP also will surface here too.

Of course, we may decide to be radical and only provide protection say for the IP based versions of the Iu and Iur protocols. This is far from what we really want, but it **may** be the most realistic to achieve for now. I much rather see something done well for the part of the stack that is the most likely to persist than to end up with nothing for lack of focus.]

8.3 Security for the interfaces Gi and Gp

[This is were we may include recommendations for filtering and firewall configuration etc – much of the DoS protection stuff. Must be seen together with 7.2 "Router filtering and firewalls". Notice also that some of the GPRS documents at least mentions the use of FW. So 03.20, 09.60/29.060 and 09.61/29.061 should be consulted.]

ffs

8.4 Security for the CAP protocol

[Do we want to proceed with CAP security? In theory I'm for CAP security, but I don't know much about CAP except that its based on SCCP/TCAP just as MAP/SS7 is. It might be feasible to carry out the same kind of trick with CAP as has been done with MAP. But its going to take a real effort and this far I have not seen any contributions in this direction.]

ffs

8.5 Security for the A-interface

[What exactly do we want to secure? Currently I'm assuming that what we really want to protect is the transport of the key **Kc** and assoc.params. This can be achieved by protecting a few BSSMAP messages over the A-interface. BSSMAP is transported over SCCP (a GSM functionally modified version of SCCP) using both connectionless and connection oriented SCCP classes. Oversimplifying a bit, this clearly points to an application layer solution. Provided that only **Kc** and associated sec.params. would need to be protected, a preliminary analysis (OK, very preliminary) shows that only one information element (IE) needs to be protected. This then suggests that a minimal solution could be had that simply has a special encrypted version of this IE.

Of course, protecting **Kc** over the A-interface is nice and perhaps needed (but I notice that a good many MSCs and BSCs are co-located in the real world), it isn't sufficient as long as **Kc** also need to be transported to the BTS. And while protection of the A-interface might be feasible, the Abis-interface is not nearly as standardized and would therefore probably be hard to protect in a standardized way. (maybe somebody can prove me wrong here?)

Anyway, I haven't seen any contributions in this area and unless we receive any I suspect that nothing much will be done here.]

ffs

Annexes are only to be used where appropriate:

Annex <A> (normative): <Normative annex title>

Annex <X> (informative): Change history

It is usual to include an annex (usually the final annex of the document) for specifications under TSG change control which details the change history of the specification using a table as follows:

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New