3GPP TR 33.801 V1.0.0 (2005-11)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Service and System Aspects;
Access Security Review
(Release 7)

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organizational Partners and shall not be implemented.
This Specification is provided for future development work within 3GPP only. The Organizational Partners accept no liability for any use of this Specification.
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Select keywords from list provided in specs database.

Keywords

<keyword[, keyword]>

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

SA3 has agreed on short-term solutions to mitigate the worst effects of discovered A5/2 vulnerabilities. SA3 has also agreed that long-term security enhancements are needed to protect GERAN Access Network in the future. A deeper study of GERAN security weaknesses, in particular security dependencies between various uses of the GSM security context, and consideration of potential future attack scenarios is needed, to decide on suitable long-term enhancements of security for GERAN Access Network and other access types relying on GSM security context. Similar considerations for UMTS access security are also taken into account in order to evaluate and re-assess UMTS security features for future attack scenarios, not originally considered [21].

1 Scope

The present document comprises an analysis of the potential vulnerabilities and vulnerabilities coming from the use of GSM security context and threats coming from the re-use (transfer) of a security context between GSM and UMTS (and other access networks) in the absence of particular security features such as strong encryption algorithms, network authentication, key separation, etc. Interaction between GERAN/UTRAN/other access during hand-over is also to be considered.

Details in the scope of the study are

- Re-assess the security objectives for "GERAN security", i.e. consider whether the security objectives that were deemed required at the time "GERAN security" was designed is still sufficient or not.
- Perform a threat, risk and vulnerability analysis of GERAN and UTRAN access security. This should in no way be limited by e.g. A5/2 cipher vulnerabilities, but should rather look at (at least) those issues raised in [15].
- The study should take into account known potential threats and vulnerabilities, and should also try to identify new ones with a clean-sheet approach.

- Provide a survey of long-term countermeasures to limit these threats and risks. A possible countermeasure should not (at this point) be ruled out just because it would imply major changes, such as e.g. phasing out legacy SIMs. (In other words, no part of the GSM/UMTS “security context” is by default left out from the study).

- Study the feasibility of the introduction of these countermeasures in the time-frame of 3GPP Release 7. This includes not only cost of implementation, but also migration and backwards compatibility issues.

- Suggest a set of feasible to implement, long-term security enhancements to GERAN (and possibly UTRAN, WLAN interworking) that reduces relevant risks/realistic threats.

The following issues are explicitly left outside of the scope:

- DoS attacks of pure radio interference or “jamming” nature.

- Cryptographic analysis of individual algorithms.

- While it is possible for operators to define their own AKA algorithms, possible desire to maintain the secrecy of these algorithms shall not be considered as part of the scope. Indeed the study shall be based on the assumption that all algorithms (including COMP and GEA variants) are publicly known.

- MAP signalling is not in scope, since it is only used between network core nodes.

2 References

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

[1] Elad Barkan, Eli Biham and Nathan Keller, "Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication", Proceedings of Crypto 2003, Springer LNCS 2729.

[2] Vodafone, "Cipher key separation for A/Gb security enhancements", S3-030463, S3#29, 15 – 18 July 2003, San Francisco, USA.

[3] Ericsson, "Enhanced Security for A/Gb", S3-030361, S3#29, 15 – 18 July 2003, San Francisco, USA.

[4] A. Maximov, T. Johansson and S. Babbage, "An improved correlation attack on A5/1", Proceeding of SAC 2004.

[5] Ericsson, "On the introduction and use of UMTS AKA in GSM", S3-040534, S3#34, 6 - 9 July 2004, Acapulco, Mexico.

[6] Vodafone, "Analysis of the authenticated GSM cipher command mechanism", S3-040262, S3#33, 10-14 May 2004, Beijing, China.

[7] Vodafone, "Evaluations of mechanisms to protect against Barkan-Biham-Keller attack", S3-040263, S3#33, 10-14 May 2004, Beijing, China.

[8] Ericsson, "Comparison of Suggested A5/2 Attack Countermeasures", S3-040341, S3#33, 10-14 May 2004, Beijing, China.

[9] Qualcomm Europe, "An observation about Special RAND in GSM", S3-040572, S3#34, 6 - 9 July 2004, Acapulco, Mexico.

[10] Ericsson, "Enhancements to GSM/UMTS AKA", S3-030542, S3#30, 6 – 10 October 2003, Povoa de Varzim, Portugal.

[11] Lucent, "Eavesdropping without breaking the GSM encryption algorithm", S3-040360, S3#33, 10-14 May 2004, Beijing, China.

[12] C. Brookson, "Authentication: A mechanism for preventing man-in-the-middle attacks", S3-040036, S3#32, 9 - 13 Feb 2004, Edinburgh, Scotland, UK.

[13] Ericsson, "Generation of IOV-UI/IOV-I values during PS Handover", GP-041987, GERAN#21, 23 – 27 Aug 2004, Montreal, Canada.

[14] Nokia, "Handling of ciphering during PS Handover", GP-042046, GERAN#21, 23 – 27 Aug 2004, Montreal, Canada.

[15] Ericsson, "Future of GERAN Security ", S3-040789, S3#35, 5 - 8 October 2004, St. Paul's Bay, Malta.

[16] Project RC5, http://www.distributed.net/rc5/

[17] Electronic Frontier Foundation, "Cracking DES", O'Reilly.

[18] E. Skoudis and L. Zeltser, "Malware: Fighting malicious code", Prentice Hall, 2003.

[19] 3GPP TS 33.102 V6.2.0, "Technical Specification Group Services and System Aspects; 3G Security; Security architecture (Release 6)".

[20] U.Meyer and S. Wetzel, "On the impact of GSM encryption and man-in-the-middle attacks on the security of interoperating GSM/UMTS networks", 2004.

[21] 3GPP TS 21.133 v4.2.0; "Technical Specification Group Services and System Aspects; 3G Security; Security Threats and Requirements (Release 4)"

[22] ECRYPT Network of Excellence, "ECRYPT Yearly Report on Algorithms and Key sizes (2004)", Rev 1.1

[23] Nokia, "Future of GERAN Security ", S3-040288, SA3#33, 10-14 May 2004, Beijing, China

[24] Orange and Nokia, "Introducing the special RAND mechanism as a principle for GSM/GPRS", S3-040529, SA3#34, 6-9 July, Acapulco, Mexico

[25] Orange and Vodafone, "Further development of the Special RAND mechanism", S3-030588, SA3#30, 7-10 October, 2003

[26] David Wagner's home page, http://www.isaac.cs.berkeley.edu/isaac/gsm.html, last visited 2005-05-12

[27] Torvinen, et.al, "draft-torvinen-http-digest-aka-v2-02.txt", IETF-draft, work in progress.

[28] BT, "Protecting GSM/GPRS networks from attacks from compromised WLAN networks when interworking"

[29] 3GPP TS 43.318 "Generic Access to the A/Gb interface; Stage 2"

[30] Ericsson, "Enhancements to GSM/UMTS AKA", S3-030542, SA3#30, 6-10 October, 2003

[31] 3GPP TS 55.205 v6.1.0, "Specification of the GSM-MILENAGE Algorithms: An example algorithm set for the GSM Authentication and Key Generation functions A3 and A8"

[32] 3GPP TS 55.226 v1.0.0, "Specification of the A5/4 encryption algorithms for GSM and ECSD, and the GEA4 encryption algorithm for GPRS; Document 1: A5/4 and GEA4 specification"

[33] Elad Barkan and Eli Biham, "Conditional-Estimators and Correlation Attacks on A5/1", Proceedings of Selected Areas in Cryptography, 2005

3 Definitions, symbols and abbreviations

3.1 Definitions

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

Access Gateway: A node on the border between the access network and the Core Network with the property that it is trusted to terminate the access security. E.g. for UMTS the AG is the RNC, for GSM it is the BTS, etc.

Editor's note: should we be even more general and open up for terminating GSM security in another place? Seems a quite tough change.

GERAN security: Any and all security and privacy aspects related to the protection of GSM or UMTS security context and Layer 2 traffic (user plane CS/PS, and signaling) carried in GERAN A/Gb between the TE and the CN, and any other security function (e.g. application layer) building on these.

GSM security context: As defined in [19].

One-way function: a function easy to compute but infeasible to invert.

Terminal equipment: This consists of the subscriber's 2G MS/3G UE together with SIM/UICC, as well as any other device that can access and use GSM/UMTS security context information such as RES/SRES, Kc/CK/IK, etc. Thus, a laptop with a SIM card reader is considered TE, even if no mobile phone is used.

UMTS security context: As defined in [19].

UTRAN security: As above, but for UTRAN/GERAN Iu-interface.

3.2 Abbreviations

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

AG	Access Gateway
AN	Access Network
CN	Core Network
DoS	Denial of Service
PDG	Packet Data Gateway
SSO	Single Sign-on
TE	Terminal Equipment

4 Network model

Since the study covers GERAN, UTRAN and any other access technology, interworking with GSM/UMTS security context, the following abstract network model is used.

Editor's note: insert a figure showing a TEs with "SIM", and three access networks; GERAN, UTRAN, and "other access", together with the corresponding AGs

5 Threat and Risk Analysis Methodology

Every threat/risk analysis starts with *a system description*. In this case the system is rather well known, so the main purpose is to define which parts of the GSM/GPRS/UMTS networks that are part of the study. Also, *assumptions* may need to be made about changes that may have been introduced in the time frame of Release 7, e.g. the disabling of A5/2 in Rel-6 terminals.

Next, *assets* that are desirable to protect are defined, and classified according to sensitivity/value.

The consequences of a weak algorithm attack are highly dependent on the resources of the attacker. For instance, casual eavesdropping requires modest resources, whereas the active attacks (false base stations etc) require substantially larger resources. Therefore, *attackers/threat agents* that are considered relevant are defined, and classified into categories.

After this, a t*rust model* reasonable for use in Release 7 time frame is defined. It is clear that with public WLAN access, re-use of (U)SIM authentication, etc, the trust model may be quite different than that used in the original design of GSM.

Then, *threats* to these assets are identified after which it is then investigated which threats that seem practically possible to be realized. During this investigation there is only a coarse-grained attempt to qualify how probable a particular threat is.

As the next step, the *risks* posed to the assets are evaluated. The risk is measured as a function of the probability that the threat is instantiated and the cost of damages, specifically, the risk is the expected damage:

RISK = f(PROBABILITY_OF_THREAT_REALIZED, DAMAGE),

for some function f.

The PROBABILITY_OF_THREAT_REALIZED, in turn, is the total probability over all attacks leading to the threat's realization. A scale of 1-5 is used for *seriousness* is shown in Table 1, 5 being the most serious. Similarly probabilities are estimated on a 1-5 scale, 5 being the most probable, as shown in Table 2.

Table 1. Threat seriousness levels

1	**Minimal** Example: Threats that would not imply anything for user privacy, QoS, or charging, e.g. being able to occasionally increase a phone's transmit power.
2	**Small** Example: Threats that, if realized, only causes very small annoyance for a single user during a short period of time.
3	**Medium** Example: Local threats, e.g., DoS targeted at a small set of BTSes under a single BSC. Could occasionally lead to single instances of incorrect charging data.
4	**High** Example: Something that, if realized, would be mentioned in IT/telecom media.
5	**Very high** Example: Something that would make front-page news, seriously damaging the trust in mobile networks, either from users' or operators' point of view, e.g. complete loss of privacy and/or robust charging.

Table 2. Attack probability levels

1	**Negligible** Example: Attack successfull with a probability comparable to guessing/breaking an (at least) 80-bit key, or requiring resources equivalent to breaking such a key. Alternatively, requiring full control of some critical function e.g. the AuC, from the "outside". (Note: 80-bit is marginal if one considers attacks by "national agencies", see [22].)
2	**Unlikely** Example: Organized crime with considerable resources would only occasionally be able to mount a successful attack.
3	**Medium** Example: Organizations, capable of erecting rogue network GERAN/UTRAN equipment, e.g. base-stations, are likely to be able to succeed.
4	**High** Example: Qualified/resourceful individuals or small groups, e.g. capable of manipulating consumer products on a limited scale, could succeed.
5	**Almost certain** Example: The attack is performed realized by single, averagely skilled "hackers" with standard PC/phone resources, possibly using "attacking tools" developed by someone else, found on the Internet. Cryptographic complexity: approximately 40-bits

Note that there is no correlation between the seriousness and the probability levels in the tables above, i.e. a threat's seriousness can in principle be completely independent of the probability that it is realized.

6 System Description

6.1 Assumptions

It is assumed that the following holds for 3GPP Release 6 with some enhancements (e.g., A5/2 has been removed):

1. A5/2 has been disabled from Rel-6+ MEs, A5/3 is supported by Rel-6+ MEs, and the available security measures are enabled by the operator (e.g., f5 is *not* implemented as the zero-function).
2. Any possible new access technology that uses GSM/UMTS security context information has a well-defined AG.
3. All used security algorithms are known, and can be analyzed by the public.
4. The Lawful Intercept systems are working properly and cannot be used by attackers to circumvent protection.
5. The AuC is securely protected and cannot be used by attackers to obtain security context data (e.g. fresh AV) or data needed to generate the appropriate security contexts (e.g. K, Ki).
6. It is assumed that an attacker does not have physical access to the (U)SIM of the victim.
7. It is assumed that node and protocol implementations are robust and are able to fail safely when faced with malformed messages etc. This is an important issue, but is out of scope for the study.
8. It is assumed that all protocols are designed to fail safely. For example, sending correctly formatted messages to a node cannot cause infinite loops, dead locks etc.
9. It is assumed that no nodes in the access network or UEs are infected by malicious software.
10. 3GPP TS 43.020 states that: "No information elements for which protection is needed must be sent before the ciphering and deciphering processes are operating." Hence it is assumed that encryption is on except for broadcast messages, the authentication procedure and initial identification of UE.

6.2 System

The system under study consists of

- the AN user and control plane traffic (GERAN/UTRAN/WLAN,…), from TE to the AG and CN,
- the security processing in AuC, MSC/VLR, SGSN, BSS, AG, PDG and TE,
- any application service relying on GSM/UMTS security context, e.g. a server using “SIM”-based SSO, GAA/GBA, etc,
- the (U)SIM and its communication with external entities such as TE

6.3 Assets

User data: user payload (CS or PS) in the AN.

Security context data:

- the subscriber key, K/Ki,
- replay counter, key sequence number (where applicable),
- SA data (Kc, CK, IK, etc),
- user identity, IMSI/TMSI,

Control signalling: signalling in the AN/CN:

- radio resource management (including cipher mode command, etc)
- mobility/hand-off signalling (including AKA procedure, triplet/quintuplet transport etc),

- call set-up signalling

Security signalling: higher layer signalling, directly related to security context:

- IPsec tunnel establishment for integrated WLAN or other interworking access,
- GAA/GBA related signalling

6.4 Actors and Threat Agents

The following are the main actors, which to varying extent (see next section) are trusted by each other. They are derived from clause 5.2 of [21].

Subscriber: A person or other entity which has an association with a home environment. A subscriber is responsible for the payment of charges to that home environment (which may be before or after service delivery, i.e. pre-pay or subscription).

Home network operator: the actor that has overall responsibility for the provision of a service or set of services to a subscriber based on the existing association between them.

Visited operator: The operator of the serving network, assuming that the serving network is different from the home network.

The (untrusted) threat agents are classified as follows:

Insider: Dishonest person working for any party with legitimate access to the assets.

Outsider: Any of the following.

- **Pedestrian hacker**: a single (or a small group of) individuals which are assumed to be able to launch passive attacks on the radio interface and with computing power equivalent of a small number of workstations/PCs connected to the Internet. This type of attacker can however be assumed to be able to transmit in unlicensed spectrum using off-the-shelf equipment such as WLAN cards.
- **Organized crime**: “cyber terrorists” or resourceful organization, powerful enough to put up false 2G/3G base stations, large computing power, etc. Such an organization could potentially bribe an **insider**, but in that case the attack is considered to be mounted by the **insider**. Note that an operator performing an attack against another operator is considered as an **Organized crime** unit.
- **Agency**: an extremely resourceful organization, e.g. a national agency. (For this study, attackers are not considered to be agencies)).

6.5 Trust Model

Editor’s note: in the time-frame of Release 7, how much do operator trust subscriber, does Home Network operator trust visited, WLAN access operator, UMA access, etc.

TBD.

7 Security Objectives

In the following, the most common security objectives that are generally considered during a threat analysis are briefly surveyed, some of which were considered irrelevant or found to be met implicitly when designing GSM security. Therefore, there are objectives in the list which were not explicitly considered during the design of GSM/UMTS security, but that may now show up as new requirements.

7.1 Confidentiality

Due to the nature of wireless communication, the need for confidentiality protection of the access has never been questioned, and it is obvious that future mobile networks must strive to meet this security objective.

7.2 Integrity

From 3G systems and onwards it has been deemed necessary to provide integrity for signalling but not for user data. However, there is clearly a big difference in difficulty between injecting/forging traffic on WLAN access respectively GSM access. Although there are currently no problems identified with the robustness of charging, there may be advanced charging scenarios that calls for a re-assessment of this requirement.

7.3 Authenticity

Subscriber authentication has from the start been an obvious security objective, if for no other reason to ensure robust charging. 3G and newer accesses have identified the need for strong authentication (mutual authentication with replay protection). Note that unless integrity protection is used, the reliability of authentication depends on the frequency of re-authentication. In fact, in scenarios where pair wise shared keys are used, authenticity and integrity are in direct correspondence; you cannot have one without the other.

7.4 Non-repudiation

The term non-repudiation in this TR is used to describe the property that there will be cryptographic evidence of a transaction between two parties, which can be used to verify that the transaction took place. This cryptographic evidence need not involve public key signatures. For example, two operators, A and B, can sign a contract with their subscribers that states that if a subscriber S belonging to operator A is roaming in operator B's network, the authentication procedure between a node of operator B and Subscriber S shall result in a value X which could only be produced using the knowledge of subscriber S's key Ki, and that value X serves as evidence that subscriber S did indeed authenticate within operator B's network. In both GERAN and UMTS access, the visited network is trusted to authenticate the user and to produce correct charging data. This means that a user who complains about a phone bill currently has no "cryptographic" evidence speaking either for or against him. Nor does the visited network have any "proof" of the user's presence in the network towards the HN. The trust model is differently for IMS access, where the home network always performs the authentication. Changes in the trust model might be considered e.g. when serving the objective of supplying non-repudation of authentication.

7.5 Availability

While attacks of "radio jamming" nature are not considered, other aspects are of importance for the study. For example, the subscriber should not loose network attachment, except if radio contact is lost/degraded. If the TE is handed over to another base station, there should indeed be a base station there to continue the service, etc. Furthermore, attacks that are non-persistant to their nature are not to be considered, i.e., the study only covers attacks against the availability that are still in effect after the attack has ceased.

7.6 Privacy

On the "access level", the important aspects of privacy are: protection of subscriber identity and location, protection against unsolicited paging, and privacy of user traffic, the latter being considered a confidentiality issue for the purpose of this study. The use of temporary identities in GSM gives some anonymity, since the real identity is only used when the MS gets connected to the network. Strong confidentiality protection (Section 6.1) provides privacy to the subscriber's communications.

8 Known Vulnerabilities

8.1 Cryptographic Algorithm Vulnerabilities

8.1.1 Weak ciphering algorithms

Any access or application security solution which uses GSM/UMTS security context and a weak ciphering algorithm potentially jeopardizes confidentiality and possibly also a spreads the vulnerability to other accesses re-using the particular compromised security context. Even if A5/2 has been removed in this time-frame, it can not be ruled out that some other algorithm (A5/1 and GEA1 being the most likely victims), are also broken. Indeed, recent attacks on A5/1 ([1, 33] and [4]: about 20 seconds of known plaintext, and a ten minutes computational effort) raises the question how long A5/1 can be trusted. The main thing that protects GEA1 is probably the fact that the algorithm is still not known to the general public.

8.1.2 Key size

The 64-bit key size of GSM's A5/1-3 is marginal. The RC5-64 project [16], retrieved 64-bit keys by brute force in about 3 years using distributed Internet computing, which today (assuming Moore's law) could be done in less than a year. Even the pedestrian hacker type attacker could possibly launch such an attack by "stealing" CPU cycles from a large number of users by a large scale "malware" attack (in fact, several Internet "Worms" have been designed for this purpose [18]). Organized crime or other resourceful attackers can build special purpose hardware that would retrieve such keys in a matter of hours, extrapolating from [17]. A general recommendation for secure key size for the foreseeable future would be around 100 bits [22], and 128 bits may be a practical choice.

8.1.3 Weak AKA algorithms

Weak (GSM or UMTS) AKA algorithms could make authentication responses or cipher keys predictable or even reveal parts of the subscriber key. Though the AKA algorithms are not standardized, effects of weak AKA algorithms needs consideration.

Editor's note: Consider adding weak integrity algorithms.

8.2 Cryptographic Protocol Vulnerabilities

8.2.1 Lack of network authentication and Authentication Vector replay protection

GSM AKA has no network authentication and this is part of the reason that cipher weaknesses may spread outside of GSM access. UMTS AKA does provide replay protection, yet as noted in [20], it is possible to set up false GSM base stations (Not UMTS base stations), which makes it possible for an attacker to request a RAND and AUTN from the network on behalf of the victim. Then the attacker makes the victim attach to its false GSM base station, and provides the earlier retrieved RAND and AUTN. Since the base station is of GSM type, the attacker can force the victim into using a weak cipher (assuming one exists). This can be seen as a "pre-play" attack. Though the RAND/AUTN has not been seen before by the TE/USIM, it is still *not* the RAND/AUTN that the TE would have received, had it been communication with a trusted serving network. To guarantee freshness in this sense, the protocol would require the use of time-stamps and clock synchronization, or preferably, exchanging RAND values both TE-to-network, and network-to-TE. Using timestamp techniques implies that the authentication has to be perfomed between the TE and the home network each time (since it is then not possible to store batches of AVs in the serving network). This gives rise to a considerable increases in the load of the networks.

It should be considered very carefully whether it is really necessary to authenticate particular visited networks (or even base stations), or only to ensure (as today) that the visited network is authorized to serve. The former feature may be useful to counter certain attacks, but may raise a number of practical configuration issues, as the user then has to be aware which network that is desirable to attach to and reject the connection if the network is different. Furthermore, the impact of what part of the network is to be authenticated varies, e.g., it is completely impractical for the TE to authenticate single base stations.

8.2.2 Lack of integrity protection

As mentioned, without integrity (as in GSM), strength of authentication is weakened since it is theoretically possible to hijack the session after authentication has taken place. The introduction of integrity protection alleviates the need for frequent authentications. The encryption only provides limited protection, in particular the encryption may be switched off from time to time, which leads a situation where the strength of authentication depends on confidentiality, an undesirable dependence.

It is also noted that the termination point of integrity may be of importance. For instance, in WLAN access, even if stronger TKIP or AES based WLAN mechanisms are used, the L2 integrity terminates in the access point, which typically in a potentially “hostile”, public environment. Therefore, WLAN L2 integrity (alone) does not provide sufficient means for e.g. robust charging.

8.2.3 Key (in)separation

Key separation refers to the property that the same key can never be used for the different purposes, e.g. both for integrity and confidentiality, or even for confidentiality using two different algorithms. Key separation requires either guaranteed replay protection, or, an algorithm/access type specific conversion of the key using a one-way function. Without this, a weak GSM encryption algorithm will threaten the confidentiality of other GSM algorithms and there can also be inter-access security dependencies, e.g. GPRS or WLAN implications on GSM. Due to the use of exactly the same procedures and authentication functions in GSM/GPRS, it is of special importance to consider interaction between these two systems.

8.2.3.1 UMTS/GSM Hand-over and algorithm similarities

When performing handover from UMTS to GSM, the TS [19] specifies a key-conversion function. Specifically, the 128-bit CK/IK are turned into 64-bit GSM Kc by XOR:ing the four “halves”. This means that if Kc is used with a weak GSM encryption algorithm, 64 bits of information about the (CK,IK) pair leaks. This is not directly devastating, but shows that the choice of key derivation function is not an arbitrary one. One could consider a worse potential problem as follows.

Suppose that the 128-bit A5/4 algorithm has been introduced in GSM and that the following (quite natural) key conversion function would have been specified:

$$Kc = CK \text{ XOR } IK. \qquad \text{(Eq. 1)}$$

Now, GSM A5/4 is essentially identical to UMTS UEA1. This means that in the (unlikely) case that UEA1 is broken, so is A5/4 and vice versa. Now, per se this means that UMTS confidentiality is essentially lost. However, thanks to the integrity protection of UMTS it is (unlike the GSM case) still not possible to hijack the session since integrity is based on IK which cannot be deduced from CK. However, suppose an active attacker fools the TE into making a hand-over to GSM. The ciphering there will take place using Kc as in (Eq. 1) above, which can similarly be retrieved by breaking A5/4. But, this now means that the attacker knows CK and (CK XOR IK), from which also IK can be deduced. By handing over back to UMTS, the attacker can now also hijack the UMTS session.

Admittedly, this is a contrived example, but it again shows that key derivation/conversion functions cannot be arbitrarily chosen. The specification [20] lists the cases that could occur when a UE is authenticated over GERAN/UTRAN (including all combinations of GSM/UMTS capable MSCs and SIM/USIM) and then is moved over the other type of access network. The conclusion is (under the assumption that the GSM encryption can be broken) that GSM subscribers who perform hand-over from UTRAN to GERAN reveals information on both IK and CK for all UTRAN communication, both for keys used before the handover and keys used when moved back to UTRAN. UMTS subscribers will have 64 bits of the key-material revealed (if handover to a UMTS capable MSC) or both IK and CK if the handover is to a GSM only capable MSC.

Also, one could argue that since UEA1 is somewhat similar to UIA1, the fact that UEA1 is broken could have implications also for UIA1 and vice versa. It may thus be desirable to consider the addition of new UMTS UEA/UIA algorithms that are more “independent” of each other and of UEA1/UIA1.

NOTE: The discussion documents on MITM in GSM-UMTS Handovers (S3-050101) advices following configurations to ensure key freshness in UTRAN i.e. perform a re-authentication when roaming in IDLE mode from a 2G type access to a 3G type access for a 3G subscriber.

- Networks should be configured to ensure that there is a UMTS re-authentication after a change from a 2G MSC to a 3G MSC in IDLE mode when a GSM security context was transferred.
- Networks should be configured to ensure that all 2G-MSC/VLRs, which support handover to UTRAN, also support and use UMTS AKA.
- Networks should be configured to ensure that there is a UMTS re-authentication after a change from a 2G SGSN to a 3G SGSN in IDLE mode when a GSM security context was transferred.
- PS handover is currently being developed in 3GPP - see TS 43.129. If inter-system RAT handover (GERAN A/Gb to UTRAN) is supported, then networks should be configured to ensure that all 2G-SGSN, which support handover to UTRAN, also support and use UMTS AKA.

8.2.4 Two-time Pads

GSM/UMTS (for good reasons) relies on stream ciphers. These are vulnerable to replay attacks, causing so-called two-time-pads. One could imagine an attack as follows. An A5/1 (say) session is recorded. Later, the victim is (somehow) fooled into sending a known message (e.g. responding to an email, a form of "phishing" attack) using the same replayed RAND. This enables the attacker (using a false base station) to decrypt the recorded traffic. Even if *this* attack is not considered realistic, it shows an "unsoundness" in allowing replay that could potentially be exploited also in other ways.

Network authentication (7.2.1) is typically a pre-requisite to obtain authentication vector replay protection.

In [13,14] issues were raised concerning potential loss of security in connection to PS handover. The security of the GPRS ciphering depends on the uniqueness of a 32-bit IOV value; in case of collision (which may occur in such hand over situations) a two-time-pad is generated, revealing *at least* the XOR of the corresponding plaintexts. With the coming 128-bit GEA4 algorithm, the overall security will potentially depend on accidental collisions between 32-bit values. The collisions of IOV values are not under an attacker's control, and taking advantage of these collisions can be classified as a passive attack. Section 9.2 discusses the security problems of IOV-collision from a practical point of view.

8.3 Repudiation scenarios

There is a trend towards decreased trust in the visited network. For example, in IMS, authentication is done in the home. Consider the following "repudiation" scenario, which might be a WLAN access scenario. A somewhat dishonest visited network, X, claims that that home network Y's subscriber, S, is roaming in X. Y will happily (?) provide authentication vectors but will really not have any chance to determine if S is really in X's network. Later S might claim he never was. It is impossible to (robustly) decide if S was in X's network (or if X is lying in an attempt to get compensation) with current AKA mechanisms. However, it would be very easy to solve this cryptographically by introducing non-repudiation mechanisms.

Editor's Note: Solutions on how to achieve non-repudiation using (as-)symmetric key techniques need to be detailed further..

8.4 Potential Vulnerabilities with Suggested Enhancements

A number of suggested countermeasures to the A5/2 attack have been proposed, some which if implemented, *may* have security issues that needs consideration.

8.4.1 Identity module cloning

What would happen if two identical (same IMSI, Ki) were used in parallel? In current specifications, there is nothing that prevents this. For instance, would the first TE ,after being authenticated, be detached from the network when the second TE with the same IMSI and Ki authenticates?. In this document it is assumed that it is not possible to physically reverse engineer a SIM. However, as have been seen, COMP128-1 were weak and allowed retrieving Ki. Also, the process of population of Ki values into the AuC is a potential weak link in the security chain.

Currently only attacks against SIM have been identified, but it cannot be excluded that USIMs may also have weaknesses (although non have yet been identified).

8.4.2 Other potential vulnerabilities

A potential threat scenario is discussed in [11]. A simplified description follows. The attacker records the RAND used when an MS authenticates to the network, and then makes a call to the MS. During this call the attacker records the encrypted traffic going to the targeted MS. At this point the attacker knows both the plaintext and the ciphertext corresponding to the Kc the targeted MS derived for the particular RAND. Hence, the attacker can deduce the keystream the MS used during the call. Next, the attacker sets up a MITM between a network and the targeted MS, using the recorded RAND as challenge to the MS. The MS will derive the same Kc and will thus use the same keystream again. The attacker will now be able to inject arbitrary traffic to the MS and listen to any traffic to/from the MS.

As stated in [11], this attack can be mitigated by requiring network authentication.

9 Threat and Risk Analysis

9.1 Threat Analysis

For each of the assets, a threat analysis is performed against each of the security objectives relevant for that asset. For each threat, possible attacks are listed. Also the most important "sub-assets", comprising the "total asset", are identified. To simplify analysis, data modification attacks where an attacker uses a radio transmitter to change the content of messages mid-air are not considered, since these are seen as very difficult to mount and does not give the attacker any more power than if he control a relay node.

9.1.1 User payload

No sub-asset.

9.1.1.1 Threats to confidentiality/privacy

Threat: sensitive user conversation/packet data is revealed.

Attack(s):

- The ME is fooled to re-use a previously compromised key.
- The ME is/will be fooled to re-use the same key with an insecure algorithm (see Section 9.1.5).
- The key is disclosed by other means (see Section 9.1.5.1).
- The ME uses a stream cipher and re-uses a non-compromised key (and other data) that was earlier used to protect data known to the attacker (see, e.g., Section 8.4.2).
- The ME uses a stream cipher and later re-uses the same (non-compromised) key (and other data) to protect data known to the attacker.

- The ME is fooled into switching off ciphering (see Section 9.1.4.2).

Seriousness: 5 (A5/2 compromise made headlines, it will happen again if e.g. A5/1 is broken.)

Probability: 1 (UMTS security context) / 3 (GSM security context)

UP1: (Potential) vulnerability: A5/1 is theoretically broken, attacks close to being "practical", improvements cannot be excluded.

9.1.1.2 Threats to integrity/authenticity/non-repudiation

Threat: a subscriber generates traffic on behalf of another subscriber.

Attack(s): Cryptanalysis of AKA algorithm, enabling response to be predicted. The threat can also be realized by attacks on K/Ki, but these are treated in Section 9.1.5.1.

Seriousness: 5 (first attack would target single user, second would be general and very serious)

Probability: 1 (UMTS security context) / 3 (GSM security context) (security depends on strength of deployed AKA algorithms assuming the attacker does not have physical access to the SIM)

UP2: (Potential) vulnerability: Weak AKA algorithms (assuming the attacker does not have physical access to the SIM)

Threat: A subscriber's payload data is received incorrectly by a service (e.g. a credit card number sent over GPRS) or by another subscriber.

Attack(s): An attacker modifies user payload data blindly (or by knowing plaintext).

Seriousness: 4 (If it was possible to change the payload data in a controlled way, it would make the headlines of technical papers, but any sensitive application would be likely to use application layer security, such as TLS, to protect the data and the effects would not be that critical.

Probability: 1 (UMTS security context) / 3 (GSM security context) (Probability depends largely on the strength of the encryption. With strong encryption the probability is very low, but with a weaker encryption the only protection is the expense of mounting man in the middle attacks.)

UP3: (Potential) vulnerability: Lack of integrity protection.

9.1.1.3 Threats to availability

This is either a radio DoS attack (outside scope), or faked signalling (e.g. faked "detach", "hand-off", etc), which is handled below.

9.1.2 Call set-up signalling

Sub-assets: ME/NW control messages and "identifiers" (e.g. MSISDN).

9.1.2.1 Threats to data confidentiality/subscriber privacy

Threat: Someone can get information on who calls whom.

Attack(s): Attacker is able to eavesdrop on call setup traffic and retrieves the MSISDN of at least one of the two parties.

Seriousness: 4 (In most cases a user does not care too much whether this type of information leaks, but there may be privacy regulations that forces requirements on protection against this threat).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same reasoning as for attacks against confidentiality)

CS1: (Potential) vulnerability: Weak encryption algorithm.

9.1.2.2 Threats to integrity/authenticity/non-repudiation

Threat: Calls are redirected.

Attack(s): Attacker changes the destination MSISDN of the call in the signalling (requires MITM). The attacker could change the destination of the call to 911.

Seriousness: 5 (There have been headlines where VoIP operators have had problems with random redirections. If it is possible to redirect a single call it may not be too serious, but it is here assumed that it can be done generally. If the attacker is able to divert many calls to one destination, he can perform a DoS attack. The attacker can also divert calls to destinations that induce a high charging rate.)

Probability: 1 (UMTS security context) / 2 (GSM security context) (Probability depends largely on the strength of the encryption. With strong encryption the probability is very low, but with a weaker encryption the only protection is the expense of mounting man in the middle attacks. In UMTS this can be excluded, since the signalling traffic is integrity protected. The DoS attack is basically only effective if the attacker has means to perform a distributed attack.)

CS2: (Potential) vulnerability: Weak encryption algorithm / lack of integrity protection.

Threat: Calls are dropped.

Attack: Send faked "hang-up" or "call reject" signalling in the middle of a call.

Seriousness: 4 (In many respects it is very similar to the previous attack, but it completely lacks the DoS and charging aspects mentioned there).

Probability: 1 (UMTS security context) / 2 (GSM security context) (If encryption is secure then this is slightly more difficult to perform than the previous attack, because the attacker does now not have a particular message that he can change, but needs to create a message that decrypts to a "hang-up" or "call reject" message for an ongoing call. If on the other hand the encryption is weak, it may be easier, since the attacker can then inject the message during a silent period).

CS3: (Potential) vulnerability: Weak encryption algorithm / lack of integrity protection.

Threat: Calls are faked.

Attack(s): Send faked "call set-up" signalling. The real subscriber (the victim) must have authenticated prior to the attack. The attacker could set up numerous calls to 911.

Seriousness: 5 (This seems at least as serious as any of the two previous attacks, because now the attacker can also initiate calls that are being charged to a subscriber).

Probability: 1 (UMTS security context) / 2 (GSM security context) (By the same reasoning as for the two previous attacks).

CS4: (Potential) vulnerability: Weak encryption algorithm / lack of integrity protection.

Threat: A subscriber does not get charged for a call he/she did make.

Attack(s): A subscriber denies making a call he/she did make.

Seriousness: 3 (As long as the visited network provider is trustworthy and only a limited number of subscribers perform the attack, this is not big problem. A potentially more serious case would be if an attacker clones his own SIM. The cloned SIM is given to a collaborator who uses the SIM in a location different from the attacker (e.g. another country) for making one local call, and then switching power off. The attacker can now make a long distance call (close in time), and then provide the differences in location as evidence that the call could not have been made, and there is an error in the logs. Thus a hard to resolve non-repudiation scenario would occur.)

Probability: 5 (This is today possible and probably occurs).

CS5: (Potential) vulnerability: Lack of non-repudiation of generated charging records.

Threat: A subscriber gets charged for call-time he did not use.

Attack(s): A session is hijacked; making call longer than user think it is.

Seriousness: 4 (This is slightly less serious than the case when the attacker can make calls that he himself can make use of, but it is serious since there are phishing attacks (see the probability estimate below))

Probability: 1 (UMTS security context) / 2 (GSM security context) (Besides the possibility to disrupt the hang-up message of a call on the air link, the only way identified is to use a false base station. This requires that the other end of the call is not breaking the call, i.e., it is some form of automatic service rather than a human voice call. An example could be that the attacker first uses phishing to get a subscriber to call a high value service, then when the subscriber calls to the service the attacker does not terminate the call, and disturbs the hang-up message from the subscriber. This is a very complex attack).

CS6: (Potential) vulnerability: Protection terminates in a physically unprotected area (i.e., the base station), which makes MITM attacks possible (in GSM).

9.1.2.3 Threats to availability

This is either a radio DoS attack (outside scope), or faked signalling (e.g. faked "detach", "hand-off", etc), which is handled below.

9.1.3 Mobility signalling

Important sub-assets:

- Authentication signalling (e.g., AUTN, RAND and RES)
- Identification procedures
- (P)TMSI re-allocation signalling
- Location update (IMSI attach/detach)
- Access network discovery signalling.

9.1.3.1 Threats to confidentiality/ subscriber privacy

Threat: User/TE identity is revealed.

Attack(s): An attacker sends a faked identification request, to which the ME responds (requires a false base station). In GPRS the attacker performs a faked GPRS detach/attach.

Seriousness: 4 (While the IMSI may be less useful in itself than the MSISDN, it allows tracking attacks etc.)

Probability: 3 (UMTS security context) / 3 (GSM security context)

MS1: (Potential) vulnerability: Lack of integrity protection of network to ME signalling for this message.

Threat: A subscriber/TE is tracked.

Attack(s): An attacker listens to the attach signalling and records the IMSI/IMEI of a subscriber. The attacker can then follow the ME's subsequent updates of (P)TMSI. The attack requires that the confidentiality protection can be broken.

Attack(s): Subscriber is tracked by an attacker that can relate SQN (in AUTN) values to each other. The attacker succeeds in crypt-analysing the f5 or f5* functions and can read the SQN from the traffic.

Seriousness: 5 (The attacker must set up receiver stations in the entire area where the tracking is to be performed (or follow the victim around with a receiver), and an attack would certainly make headlines.)

Probability: 1 (The identified attack is very complex and requires breaking the encryption).

MS2: (Potential) vulnerability: Weak encryption algorithm.

9.1.3.2 Threats to integrity/authenticity/non-repudiation

Threat: ME is forced to use a different network.

Attack(s): An attacker changes an "location update accept" message to an "location update reject" with a cause code of "PLMN not available", and hence forces ME to look for another one. This is a man-in-the-middle attack.

Seriousness: 3 (An attacker is able to stop a visited network operator from getting users to connect to the visited network).

Probability: 1 (UMTS security context) / 2 (GSM security context) (Same reasoning as for the man-in-the-middle attack on the signalling).

MS3: (Potential) vulnerability: Lack of integrity protection on signalling traffic.

Threat: The ME accepts faked authentication signalling messages.

Attack(s): (UMTS security context) The attacker manages to replay RAND and/or fake AUTN. In the latter case he has to break f1. Note that it is not necessary to break f5, since the f5 output is XORed with SQN. Thus implying that it is at least 50% chance that flipping the least significant bit of the SQN part of the AUTN produces an acceptable SQN.

Seriousness: 3 (There will potentially be two-time pads, and *if* one UMTS encryption/integrity algorithm is weak the weakness may spread to other UMTS algorithms).

Probability: 1 (There is currently no reason to believe that weak f1 implementations will be used).

MS4: (Potential) vulnerability: Weak integrity protection on signalling traffic.

Attack(s): (GSM security context) The attacker sends a replayed RAND to the ME.

Seriousness: 5 (Confidentiality will be lost, see reasoning for confidentiality)

Probability: 2 (Same reasoning as for the confidentiality, except that in this case it is an active attack, hence one degree less).

MS5: (Potential) vulnerability: Lack of replay protection on this message.

Threat: Successful impersonation of a subscriber.

Attack(s):Cryptanalysis of A3.

Seriousness: 5 (In principle there is still some protection against the attacker being able to make calls on the subscriber's behalf. But consider that A3 is often very similar to A8).

Probability: 3 (There are weak implementations of A3 in the market. If the security context is re-used in application independent way, and the attacker uses real SIM as an oracle it *will* be possible to impersonate a subscriber. This works if the application security is based only on that the authentication succeeded. Attacks of this type that uses the victim as an oracle are discussed in [27]).

MS6: (Potential) vulnerability: Weak AKA algorithms.

Attack(s): Cryptanalysis of f2.

Seriousness: 5 (In principle there is still some protection against the attacker being able to make calls on the subscriber's behalf. But consider that f2 is often similar to f3/f4).

Probability: 1 (There is currently no reason to believe that weak f2/f3/f4 implementations will be used. In addition, the RAND is authenticated. If the security context is re-used in application independent way, and the attacker uses real USIM as an oracle it *will* be possible to impersonate a subscriber. This only works under the conditions stated in the previous attack).

MS7: (Potential) vulnerability: Weak authentication response calculation algorithm.

9.1.3.3 Threats to availability

Threat: The MEs batteries are drained and the network signalling is increased.

Attack(s): A false base station broadcasts the location update timer, and it has a very low value causing the MEs to do the updates very often. The lower limit of the timer is six minutes.Seriousness: 3 (The attack will cause local annoyance for a limited set of users and time).

Probability: 3 (Requires false base station. The reason to perform the attack is a bit unclear, it could be one operator disturbing another operator's network. Note that there is no protection on these messages).

MS8: (Potential) vulnerability: Lack of integrity protection.

Threat: ME is disabled.

Attack(s): An attacker fakes a base station and changes an "location update accept" message to an "location update reject" with a cause code of "illegal equipment" to the ME. The attack is working as long as the SIM is not removed, or the ME is rebooted. Another attack is if the attacker is a MITM and changes the IMEI in the messages from the ME.Seriousness: 4 (Slightly worse than the attack where the ME is forced to chose another PLMN, because the ME is now completely disabled).

Probability: 2 (Same reasoning as for the man-in-the-middle attack on the call-setup signalling).

MS9: (Potential) vulnerability: Lack of integrity protection.

Threat: The ME fails to authenticate properly.

Attack(s): The attacker changes the RAND in the challenge, or changes the RES in the response from the ME, or he can just drop the messages. This requires a man-in-the-middle.

Seriousness: 3 (Local annoyance).

Probability: 1 (UMTS security context) / 2 (GSM security context) (Requires a false base station. The case with changing the RAND can be excluded in UMTS. There are much easier ways to accomplish a DoS, so it is questionable if this attack is attractive to perform).

MS10: (Potential) vulnerability: Lack of integrity protection on this message.

Threat: A ME is illegitimately detached from NW.

Attack(s): Fake "IMSI detach" command from the attacker to the NW that a certain ME requests detach (this requires that the attacker can circumvent the authentication or that the attacker is a MITM that uses the ME as an oracle).

Seriousness: 3 (Local annoyance. This is a persistent attack, but the attack will cease to take affect as soon as the next location area update is received by the network.).

Probability: 1 (UMTS security context) / 2 (GSM security context) (Requires a false base station. There are much easier ways to accomplish a DoS, so it is questionable if this attack is attractive to perform).

MS11: (Potential) vulnerability: Weak / lack of integrity protection on signalling traffic.

Threat: A ME is unable to establish IP connectivity to hosts due to lack of mappings in a NAT.

Attack(s): An attacker exhausts the state space of the NAT by initiating numerous connections.

Seriousness: 4 (Subscribers will not be able to create new connections.)

Probability: 4 (Although, there is no theoretical bound on the number of mappings, there are implementation decisions that has to be made to limit this number. Assuming a port based NAT the attacker can allocated 2^{16} mappings per IP address he has. Assuming the NAT has only one external IP address, only one attacker is sufficient to exhaust the mapping space (attacks of this type has been performed). For "IP-address to IP-address mapping NATs", an attacker will only get as many mappings as he has addresses, hence it is unlikely that these kind of NATs will be exhausted (on the other hand, it is questionable if this type of NATs are commonly used).

MS12: (Potential) vulnerability: Lack of limitation on the number of mappings allowed per user.

Threat: ME is tricked into camping on a false base station.

(See RS12)

9.1.4 Radio resource management signalling

Important sub-assets:

- ME capability ("Classmark") info,
- location/Cell-ID where ME is located,
- security setup signalling (e.g., cipher-mode command),
- radio measurement data,
- NW detach signalling,
- handover procedures.

9.1.4.1 Threats to confidentiality/subscriber privacy

Threat: Outsider can deduce information about a subscriber’s location.

Attack(s): Eavesdropper retrieves the Cell ID from the signalling from the UE to the NW. Note: seriousness depends on also compromising subscriber ID (see Section 9.1.3).

Seriousness: 1 (It is less serious than the IMSI/TMSI tracking attack, since the ID of the subscriber is not known by simply eavesdrop on the Cell ID of an unknown subscriber).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same as for breaking encryption for user payload).

RS1: (Potential) vulnerability: Weak encryption algorithm.

Threat: Outsider can deduce information about the ME capabilities.

Attack(s): The attacker listens to the attach signalling (or requests the Classmark information).

Seriousness: 1 (A subscriber would probably not care to much if the capabilities of his ME is known by someone else).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same as for breaking encryption of user payload)

RS2: (Potential) vulnerability: Weak encryption algorithm.

Threat: Outsider may trick ME into using no/wrong /weak encryption algorithm.

Attack(s): MITM fakes capabilities of the ME. E.g., the ME and NW are tricked into using GSM security even if both are capable of UMTS security. In GERAN access, a MITM changes the Classmark revision level (e.g., in Classmark 2 sent in CM Service Request message by the ME, unencrypted) from "R99+" to "GSM ph2"

Seriousness: 5 (User privacy is compromised)

Probability: 3 (Requires man-in-the-middle)

RS3: (Potential) vulnerability: Lack of integrity protection.

9.1.4.2 Threats to integrity/authenticity

Threat: A ME is illegitimately moved to another NW.

Attack(s): Forge radio measurement data signalling, causing handover to another NW.

Seriousness: 3 (Same as the attack when the "ME is forced to use a certain network").

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same as for breaking encryption for user payload, and requires man-in-the-middle).

RS4: (Potential) vulnerability: Weak / lack of integrity/encryption algorithms.

Threat: MEs are made to hand over to non-existing/faked base station.

Attack(s): Faked h/o signalling towards the ME (probably only applicable to GPRS).

Seriousness: 3 (Same as the attack when the "ME is forced to use a certain network").

Probability: 1 (Very complex attack, requires *at least* that the encryption is broken).

RS5: (Potential) vulnerability: Lack of integrity protection.

Threat: Network gets the incorrect information of the status of the radio link.

Attack(s): The attacker sends incorrect/faked measurements to the NW on behalf of a ME.

Seriousness: 3 (Same as the attack when ME is made to hand over to non-existing/faked base station).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Requires that the encryption is broken, and a ME/PC).

RS6: (Potential) vulnerability: Weak / lack of integrity/encryption protection.

Threat: The ME sends traffic outside of its allocated timeslots.

Attack(s): The attacker sends a message to the ME that instructs it to send traffic a little before the timeslot begins (this is used when the ME is at the border of the cell, to achieve correct synchronization).

Seriousness: 3 (Local DoS against a particular ME. Its uncertain if this attack is persistent or not).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same as for breaking encryption for user payload and requires man-in-the-middle).

RS7: (Potential) vulnerability: Weak / lack of integrity/encryption protection.

Threat: False base station.

Attack: Attacker disables (for example cutting antenna connection or cutting the power) a real base station, puts up a false base station, faking a base station (e.g. over non-authenticated micro wave link) towards the NW and fakes a NW towards the ME. The attacker can then easily tap into the information sent.

Seriousness: 5 (Same as loss of confidentiality, and in addition other attacks can now easily be performed).

Probability: 1 (UMTS security context) / 4 (GSM security context) (Requires detailed knowledge about network equipment. This attack is particularly attractive at airports, GSM in-office etc, and hence the incitement to perform it raises the probability by one. Note that in UMTS the protection is not terminated in the NodeB, so it is not only to tap in to NodeB to RNC link).

RS8: (Potential) vulnerability: Encryption terminates in the base station.

Threat: Forcing NW into performing unnecessary MAP signalling.

Attack(s): A ME sends many attach request for random/selected IMSIs.

Seriousness: 2 (Small annoyance for the network, unless a distributed version of the attack is performed).

Probability: 3 (Even though the attack can be mounted by a technically skilled person, the gain of the attack is questionable, and hence this receives a lower probability).

RS9: (Potential) vulnerability: Lack of integrity protection on signalling traffic.

9.1.4.3 Threats to availability

Threat: The MEs are not able to use signalling towards the network.

Attack(s): An attacker sets up a false BSC (could be implemented in a false base station), that broadcasts a "barred access class" message (unencrypted), that disables signalling between the network and a set of MEs. MEs does not try to reconnect after this (except for emergency calls).

Seriousness: 5 (Local but persistent DoS attack).

Probability: 3 (Requires detailed knowledge about network equipment).

RS10: (Potential) vulnerability: Lack of integrity protection on signalling traffic (broadcast in this case).

Threat: One or more MEs are illegitimately detached from NW (or are never able to attach).

Attack(s): Fake "Group Release" command from the NW to one or more MEs .

Seriousness: 4 (May annoy all subscribers below an RNC).

Probability: 3 (Requires detailed knowledge about network equipment).

RS11: (Potential) vulnerability: Lack of integrity protection on signalling traffic (broadcast in this case).

Threat: ME is tricked into camping on a false base station.

Attack(s): The attacker sets up a false base station that has a better reception at the ME, e.g., transmits with a higher power than the real base stations in the area, making the ME select this base station. After this the ME will only get traffic from the false base station and will only be able to send traffic to the false base station.

Seriousness: 3 (Only MEs located in the same geographical area are affected).

Probability: 3 (Attacker needs to be able to put up network nodes. Note that this is also possible to do in UMTS).

RS12: (Potential) vulnerability: Lack of integrity protection on signalling traffic (broadcast in this case).

9.1.5 Security context data

Important sub-assets:

- Long-term subscriber key (Ki/K), IMSI, TMSI,
- session confidentiality/integrity key(s) (Kc, CK, IK, etc),
- replay information (SQN_MS),
- application identifier (information on in which application, if any, the security context is being used in).

9.1.5.1 Threats to confidentiality/subscriber privacy

Threat: Ki/K is disclosed:

Attack(s):

- Ki/K is disclosed by passive cryptanalysis of the AKA algorithm.
- Ki/K is disclosed by active cryptanalysis of the AKA algorithm.
- Ki/K is disclosed by injection (see threats to integrity/authenticity).
- Ki/K is leaked from manufacturer.
- Ki/K is leaked when installed in AuC.

Seriousness: 5 (If Ki/K is leaked there is nothing to bootstrap the security on).

Probability: 2 (The most probable attacks are that an insider is bribed by organized crime. Weak AKA algorithms, e.g., COMP128-1 are possible to cryptanalyze, and there are newer side-channel attacks, but these are all out of scope by the assumptions).

SD1: (Potential) vulnerability: unreliable insiders.

Threat: a session key (Kc, IK and CK) is disclosed.

Attack(s):

- A particular session key is disclosed by cryptanalysis of the encryption/integrity algorithm using it.
- A particular session key is disclosed by cryptanalysis of the AKA, A8, f3 or f4 algorithms.
- All Session keys are disclosed by successfully attacking Ki/K (see above threat).
- A known value is "injected"/replayed in the protocol (see threats to integrity).
- The key is disclosed by cryptanalysis of a hand-over key conversion function.
- An attacker cryptanalysis an application, where the GSM/UMTS security context is used in an application independent way.
- Key is exposed during access network transport.
- Key is disclosed by physical tampering of AG.

Seriousness: 5 (Confidentiality/integrity is lost).

Probability: 1 (UMTS security context) / 3 (GSM security context) (Same as breaking confidentiality for user payload).

SD2: (Potential) vulnerability: Weak AKA/encryption algorithms.

Threat: SQN_MS is forced out of synch.

Attack(s): Only identified attack is by manipulating AUTN (see mobility signalling).

Seriousness: 2 (Annoyance for single MEs).

Probability: 1 (There is currently no reason to believe that weak f1 implementations will be used).

SD3: (Potential) vulnerability: Weak AKA algorithms.

9.1.5.2 Threats to integrity/authenticity

Threat: Session key(s) are modified to a known value.

Attack(s): A known key is replayed. Only identified ways to achieve this is to replay a challenge as part of the mobility signalling, or use Wagner's et.al. attack to send an "equivalent" RAND.

Seriousness: 5 (Confidentiality can be broken. The attacker first have to record the session he wants to listen to. Next he initiates a new connection to the subscriber (where he knows the plaintext). When this is done he XORs the data from the two sessions together, and in this way he gets the XOR of the plaintexts of the two sessions. Since he knows the plaintext from the second session he can derive the plaintext from the first.)

Probability: (GSM security context): 3 (There is no replay protection, and requires man-in-the-middle).

Probability: (UMTS security context): 1 (There is no reason to believe there will be weak f1 implementations).

SD4: (Potential) vulnerability: Weak AKA algorithms / lack of integrity/replay protection.

9.1.5.3 Threats to availability

Only DoS aspects.

9.1.5.4 Threats to non-repudiation

All threats related to disclosure of keys open up repudiation scenarios involving other assets than security context data (see above); no other threat has been identified.

9.2 Risk Analysis

Editor's note: This section will assign "seriousness" and" probability" to the threat found above.

9.2.1 IOV-collisions

RISK:
The 'problem' of IOV-collision existed also in the past (without PS handover) when an MS performs an inter-SGSN RAU and the 'currently used keys' are transferred to the new SGSN. When this new SGSN, decides to continue to use these 'currently used keys' and by chance, generates the same IOV-I or IOV-UI as the previous SGSN (and these IOVs are 32 bit values, i.e. the probability should be ~ 1 : 4 000 000 000), then the generated cipher stream will be the same.

Solutions: Collisions of IOV can be prevented by various means of which re-authentication seems to be the most suitable solution.

A) GP-042046 [14]: The included countermeasure changes the OC definition.
Evaluation: This is not backward compatible with old MS, so ciphering with old MS will fail with New OC - SGSN.

B) GP-041987 [13]: The included countermeasure changes the IOV structure.
Evaluation: Now the new SGSN also needs to know the incremental part of the IOV (IOV-I and IOV-UI), which requires a new parameter at the Gn-interface.

C) A possible alternative (which does not require impacts on protocols, and neither has compatibility issues) is to require re-authentication at each inter-SGSN RAU. (NOTE: XID-reset is performed at each inter-SGSN RAU and at each GPRS Attach).

9.2.2 Risk assessment

The risks presented in Table 3 are computed as the product of the seriousness and probability for each attack described in Section 8. Some attacks are valid both for GSM and UMTS security context. Therefore both probabilities are given, separated by a slash (this of course then also holds for the risk). Attacks that are not applicable to a particular security context are marked with an 'x'.

Table 3. Summary of attacks

Attack	Seriousness	Probability (UMTS/GSM)	Risk (UMTS/GSM)
UP1:	5	1/3	5/15
UP2:	5	1/3	5/15
UP3:	4	1/3	4/12
CS1:	4	1/3	4/12
CS2:	5	1/2	5/10
CS3:	4	1/2	4/8

CS4:	5	1/2	5/10
CS5:	3	5/5	15/15
CS6:	4	2/2	8/8
MS1:	4	3/3	12/12
MS2:	5	1/x	5/x
MS3:	3	1/2	3/6
MS4:	3	1/x	3/x
MS5:	5	x/2	x/10
MS6:	5	x/3	x/15
MS7:	5	1/x	5/x
MS8:	3	3/3	9/9
MS9:	4	x/2	x/8
MS10:	3	1/2	3/6
MS11:	3	1/2	3/6
MS12:	4	4/4	16/16
RS1:	1	1/3	1/3
RS2:	1	1/3	1/3
RS3:	5	3/3	15/15
RS4:	3	1/3	3/9
RS5:	3	1/1	3/3
RS6:	3	1/3	3/9
RS7:	3	1/3	3/9
Rs8:	5	1/4	5/20
RS9:	2	3/3	6/6
RS10:	5	3/3	15/15
RS11:	4	3/x	12/x
RS12:	3	3/3	9/9
SD1:	5	2/2	10/10
SD2:	5	1/3	5/15
SD3:	2	1/x	2/x
SD4:	5	1/3	5/15

The attacks with the highest risk (above 15) are listed in Table 4 for GSM and UMTS security context separately.

Table 4. Vulnerabilities (and root causes) that pose the highest risks.

GSM Security context		UMTS Security context	
20	RS8: Encryption terminates in a physically unprotected area, i.e. the base station..	20	
16	MS12: No protection against NAT mapping depletion	16	MS12: No protection against NAT mapping depletion
15	UP1: A5/1 is more or less broken UP2: A5/1 is more or less broken SD2: A5/1 is more or less broken	15	
15	CS5: Lack of non-repudiation of generated charging records.	15	CS5: Lack of non-repudiation of generated charging records.
15	MS6: Weak implementations of A3 are in use.	15	
15	RS3: MITM fakes capabilities of the ME. E.g., the ME and NW are tricked into using GSM security even if both are capable of UMTS security. This is possible due to lack of integrity protection on signalling in GSM.	15	RS3: MITM fakes capabilities of the ME. E.g., the ME and NW are tricked into using GSM security even if both are capable of UMTS security. This is possible due to lack of integrity protection on signalling in GSM.
15	RS10: An attacker sets up a false BSC. This could be implemented in a false base station and sends broadcasts a "barred access class". This is possible due to lack of integrity protection on broadcast signalling.	15	RS10: An attacker sets up a false BSC. This could be implemented in a false base station and sends broadcasts a "barred access class". This is possible due to lack of integrity protection on broadcast signalling.
15	SD4: A known key is replayed. Only identified ways to achieve this is to replay a challenge as part of the mobility signalling, or use Wagner's et.al. attack to send an "equivalent" RAND. This is due to lack of replay/integrity protection in GSM AKA.	15	

10 Overview of Possible Enhancements

Editor's note: This section will discuss protection mechanism to counter the identified risks. Note that the list given below is by no means exhaustive and an evaluation of which proposals are to be recommended are TBD.

Note that to circumvent the vulnerability RS8, it is probably necessary to move the termination point of the confidentiality protection further inside the network (as is done in UMTS). This would probably imply that too much equipment needs updating, to be economically feasible. Furthermore, to mitigate RS10 there is a need to integrity protect broadcast messages. This either requires a pre-shared key between the network and each user, or that the network signs broadcast messages (which leads to the need for certificate handling). Both alternatives have very high cost/complexity. It is suggested that the vulnerability RS10 is left without countermeasures. Since RS8 is the attack that ranked highest in Section 9, it will still be considered along side the other vulnerabilities, despite the anticipated high cost of any countermeasure. Note that if it is decided to move the endpoint of the encryption in GERAN further into the network, it would be easier to update/manage the encryption algorithms used, since they then would be located in a more central point in the network.

The following possible security enhancements for A/Gb mode are so far identified:

1) Protection against algorithm negotiation bidding down and disablement attacks.

2) Ensure that the use of 3G AKA (i.e. UMTS security context) is always possible when USIM is available so as to provide re-play protection (session key freshness)
3) Providing key separation.
4) Protection against access signalling modification (GSM security context).
5) Enlarging the GERAN A/Gb encryption algorithms key size e.g. 128 bit.

6) A/Gb tunneling within IPsec (aligned with TS 43.318 [29]).

7) Adding new encryption algorithms.

8) Removing insecure encryption algorithms

9) Extend the set of signalling messages that are integrity protected (UMTS security context).

10) Add integrity protection to user payload.

10.1 Description of possible identified countermeasures

This section describes the realisations of the identified countermeasures. The feasibility of implementing the countermeasures is then analysed in Section 11.

10.1.1 Protection against algorithm negotiation bidding down and disablement attacks

Authenticated ciphermode command (as described in [6]):

- Network returns the list of algorithms the ME sent previously as the ones the ME support. The message is integrity protected by some key from the AKA.

- The mobile compares the list received form the network to the list it sent.

- The start ciphering command is made mandatory.

Special RAND (as described in [24, 25]):

- The network signals preferred algorithms by setting certain bits in the RAND value.

- Special RAND solutions uses part of the RAND to signal in which context and which algorithms are allowed to be used with the resulting key. In [25] there is a discussion on the idea. These kinds of solutions decrease the maximum entropy of the RAND from 128 to 76 bits (in the case of [23]) and 84 bits (in case of [24]). Thus, they also decrease the theoretically effective key-space of Kc by the same amount. The decrease of entropy also means that off-line pre-computation attacks against Ki are reduced in complexity from 2^{128} to about 2^{104}. Still, this is more than enough to rule out the practical feasibility of such attacks. SAGE has estimated that the entropy of RAND could be reduced even to 64 bits without making practical, non-trivial attacks more likely to succeed.

- The collision attack on COMP128-1 [26] due to Wagner et. al. can be extended into a pre-image attack, that given a Special RAND x finds a non-special RAND y that results in the same RES. The pre-image attack can be expected to succeed after that the attacker has observed 2^{28} authentications using the same SIM. Even though this may seem as a very impractical attack, it indicates that for Special RAND solutions (such as [23] and [24]) to work, there is still a dependence on the choice of implementation for the A3/A8 algorithms.

10.1.2 Ensure use of 3G AKA is always possible when USIM is available so as to provide re-play protection (session key freshness)

All core networks should be upgraded to support 3G AKA which means that a USIM capable mobile with a USIM inserted could **mandate** *3G AKA over GERAN*. This effectively means that a UMTS security context will be available in the core network and the mobile, therefore ensuring an effective 128-bit key availability for GERAN. See Annex for Figure 18 of TS 33.102 [19].

10.1.3 Providing key separation

This feature ensures that a session key can only be used within a particular context [28]. The benefit is that a retrieved session key cannot be reused outside of the defined context (e.g. retrieved within GSM and reused within WLAN, retrieved by A5/1 vulnerability and reused within A5/3). This context can be for instance on access domain level or algorithm level.

Special RAND (see above):

Cryptographic post-AKA processing of Kc

- Cryptographic wise key separation (by session key post-AKA processing in core or access network and mobile, e.g., deriving Kc' = PRF(Kc, algorithm_ID) may require that 11.12 is taken into account. For more details see [30].

10.1.4 Protection against access unicast signalling modification.

According to good security practices, requires a key derivation solution to create an independent encryption and authentication key. The enhancement requires to take 11.12 into account.

This is a generalization of 10.1.1, where more messages than the ciper mode command are integrity protected. Could be implemented in different flavours e.g. with a small subset of commands or a wide list.

10.1.5 Enlarging the GERAN A/Gb encryption algorithms key size.

This in practice means starting to introduce A5/4, GEA4 and Milenage according to [31, 32]. Note that the encryption algorithm key size must be upgraded to 128 bits and simultaneously subscribers must start using USIMs (UMTS AKA) to get the 128-bit key, since the GSM AKA only gives a 64-bit key. Note that it is not enough to simply convert the 64-bit key from GSM AKA using the conversion function c3, since this will not give 128-bits of entropy for the key.

10.1.6 A/Gb tunneling within IPsec

When the work on Generic Access to A/Gb interface was initiated (TS 43.318) it was realized that:

- We could not necessarily rely on the local link security alone e.g. Bluetooth, WEP, WPA etc due to lack of control over the configuration of this security – The access point is in the customer premises and under their control.

- The use of the A/5 algorithms may have been confined to "GSM operators" and for use for GSM air interface only.

- The use of the existing mechanisms in a fixed environment may have increased the scope for exploitation of the identified vulnerabilities.

Instead of switching back to "GSM" air interface security on it own, when the handset reverts to the mobile network connection, it invokes IP transport with IPsec as an additional security mechanism.

This would be an option for operators to support for their own subscribers in their own network,(support their subscribers in visited networks should be a second priority and support for inbound roamers from other networks the third priority.)

10.1.7 Adding new algorithms

By this is meant development of new algorithms, A5/5, A5/6, GEA5 etc. This assumes that 10.1.5 has been implemented.

10.1.8 Disabling insecure algorithms

By this is meant disabling of, e.g., A5/1 and GEA1 from the networks.

10.1.9 Integrity protection of broadcast signalling

Integrity protection of all broadcast messages is not possible since for some there may not yet be a key in place to protect them. However, some messages could potentially be protected, but it may be required to develop a broadcast key management to support this.

10.1.10 Add integrity protection to user payload

This countermeasure consists of addition of integrity protection to the user payload between the UE and the base station.

11 Feasibility Study

This section evaluates the security enhancements identified in Section 10 from the viewpoint of their interdependencies in order to prepare for prioritization of security enhancements. The enhancements are denoted by the same numbers as in Section 10 in brackets. It is noted that these interdependencies are on a technical level and implementation cost has not been considered yet.

11.1 Protection against algorithm negotiation bidding down and disablement attacks.

This is the weakest spot in the whole GERAN 'A/Gb'. The enhancements (3) and (5) can be circumvented if there is no solution for (1).

- **The authenticated cipher mode command**
  Implications:
  - Need new ME/BTS signalling and processing (generation/verification of MAC, generation of integrity key by post processing of Kcetc.),
  -

  Remaining security issues:
  - See the considerations in Section 11.12.
  - To avoid bidding down in connection to handover, also certain handover signalling (i.e., change of algorithm) from the NW to the ME must be integrity protected. This will have performance impacts.
  - General signalling integrity issues.

  Advantages:

  Possible disadvantages:

Main vulnerabilities countered:

- SD2 (partially, only attacks related to bidding down are countered.)
- RS3

- **The special-RAND solution**

Implications:

- New processing ME and HLR/AuC.

Remaining security issues:

- See the considerations in Section 11.12.
- This special RAND solution can only disallow the use of particular algorithms in a particular access context, but does not provide a solution against False BS choosing the weakest of the permitted algorithms for the mobile. This is however not considered to be a problem, as long as the Home network is always informed about the capabilities of the VN and can send special-RANDs tailored to the VN capabilities. The down grade attack then consists of choosing between strong algorithms. Separate considerations apply when for unequal key-length algorithms (see (5)).
- Special RAND does not protect against the attack where a false base station can eavesdrop by forcing the TE into unciphered mode, even if the special RAND prohibits the use of A5/0 (= no encryption). This is because the false base station can simply omit the authentication procedure as well as the cipher mode command in order to force the use of an unciphered connection. Some further modification to the special RAND mechanism would be needed to protect against disablement attacks.
- General signalling integrity issues.

Advantages:

- The special-RAND solution can be introduced stepwise.
- Home operator can control the algorithms chosen.

Possible disadvantages:

- Reduced key entropy (see 10.1.1).
- Special-RAND does not protect against bidding down protection in a network which (allows) supports a mixture of different encryption-key-length algorithms e.g. GEA3 and GEA4 if both would be allowed. (see 11.5).

Main vulnerabilities countered:

- SD2 (partially, only attacks related to bidding down are countered.).
- RS3 (partially, bidding down to a lower security algorithm possible, but the ME could be implemented so that it does not accept "no encryption" if some bits are set in the special RAND).

11.2 Ensure use of 3G AKA is always possible when USIM is available so as to provide re-play protection (session key freshness)

Implications:

- The features relates to the discussion around *MITM in GSM-UMTS Handovers (S3-050101)*[1], but goes one step further in that this effectively forbids Rel 98- VLR/SGSN in the core network, such that upgraded mobiles shall enforce 3G AKA.

[1] - Networks should be configured to ensure that all 2G-MSC/VLRs, which support handover to UTRAN, also support and use UMTS AKA.

USIMs needs to be rolled out to SIM subscribers.

Remaining security issues:

- Instead of a configuration advice as for *MITM in GSM-UMTS Handovers*, the upgrade requirements need a solution against new 'access' network security feature downgrade attacks (see 11.12). Providing a GSMA agreed cut-off date[2] (which is timely introduced before upgraded mobiles 'mandating 3G AKA over GERAN' become available for use), provides a possible means to ensure secure implementation (see 11.12).
- This only gives authentication of the RAND, there are still possibilities for bidding down attacks on algorithm negotiations.
- General signalling integrity issues.

Advantages:

- This would provide an integrity key (IK) which can be used for other countermeasures, e.g., Authenticated Cipher mode Command.
- Ensuring 3G AKA (2) ensures time-constrained session key separation such that a successful weak algorithm attack has only a limited time-window in the sense that the cipher key retrieved by the attacker cannot be used indefinitely as 3G authentication vectors cannot be re-used and become invalid when the next authentication run occurs. Could be enhanced with re-authentication when a UE moves between networks (network access type wise key separation). But re-authentication after handover is technically difficult to achieve and, although permitted by current specifications, has never been tested.

Possible disadvantages:

- Unlike "soft upgrades" to MEs that are relatively easy to roll out due to subscribers wanting "latest" terminal, this requires convincing subscribers of the benefit of switching to USIM.

Main vulnerabilities countered:

- UP2
- SD2 (Attacks where the AKA is attacked are countered, but it does not help against keys being available during network internal transport.)
- MS6
- SD4

11.3 Providing key separation

Special RAND

Implications: see 11.1*Remaining security issues:*

- PS handover is currently being developed in 3GPP - see TS 43.129. If inter-system RAT handover (GERAN A/Gb to UTRAN) is supported, then networks should be configured to ensure that all 2G-SGSN, which support handover to UTRAN, also support and use UMTS AKA.

[2] Flag based solutions (cf S3-030542, S3-040262 Annex) on the UE to allow '*new mobiles to roam in old networks*', are considered complex and not practical.

- It needs further study to find out whether RAND is not overloaded with all sorts of flags and the random part becomes too short, if all the many different key separation cases are to be supported.

- General signalling integrity issues.

- See also 11.12.

Advantages:

- Due to the home control and access transparency properties of this method, the special-RAND structure could provide access domain level key separation by a specific flag.

Possible disadvantages: see 11.1

Main vulnerabilities countered:

- SD2 (Only attacks where the user is tricked into using the same key with one strong and one weak algorithm, and the attacker breaks the weaker one and hence is able to read cipher text produced also by the stronger one are countered)

- RS3 (It is possible to stop the user from using a weak algorithm, but it is not possible to prevent the attacker from refraining to send a start ciphermode command from a faked base station)

Cryptographic post-AKA processing of Kc

Implications:

- Update of ME and BTS/SGSN. In the case that the RAND is also included in the computation, the MSCis required. The reason for including the RAND is to avoid pre-computation attacks on the Kc'.

Remaining security issues:

- See the considerations in Section 11.12.
- Bidding down attacks on algorithm negotiation still possible and home operator has no control of which algorithms are possible to choose (c.f., Special RAND).
- General signalling integrity issues.

Advantages:

- Some types of attacks against weak GSM AKA algorithms *may* be more difficult by this. For example, attacks where Kc (retrieved by, e.g., breaking A5/1) is related to the corresponding RAND, could be more difficult, since Kc is not used directly.

Possible disadvantages:

Main vulnerabilities countered:

- UP2 (The threat that the attacker by snooping RANDs, breaking the encryption algorithm and use the Kcs in combination with the RANDs in an attack on a weak A3, can retrieve the Ki/K is probably eliminated)

- SD2 (The attack where the attacker injects a known RAND, and forces the user to use a weaker algorithm in that case is countered.)

- MS6 (The threat that the attacker by snooping RANDs breaking the encryption algorithm and use the Kcs in combination with the RANDs in an attack on a weak A3, can retrieve the Ki/K is probably eliminated)

11.4 Protection against access unicast signalling modification.

Implications:

- Has bigger impacts and higher realization complexity on GERAN than the other features. It has the same implications as Authenticated Ciphermode command, but affects more signalling. In this case it is also a stronger need to introduce negotiations of protection algorithms. Effected nodes include MSC, BSS, SGSN and ME.
- Requires introduction in all networks to be effective, no gradual or partial introduction seems satisfactory, see 11.12.

Remaining security issues:

- In theory this provides the possibility to protect any sensitive signalling. Potentially all attacks against integrity and authenticity could be solved. In practice there could be obstacles, e.g, messages could be sent before an integrity key has been established, or messages that are broadcast can be difficult to fit into the mechanism.

Advantages:

Possible disadvantages:

Main vulnerabilities countered:

- SD2 (Authenticated ciphermode command is a special case)
- SD4 (The attack where the attacker injects a known RAND, and forces the user to use a weaker algorithm in that case is countered.)
- RS3

11.5 Enlarging the GERAN A/Gb encryption algorithms key size.

Implications:

- Implementation of new algorithms in AuC, SIM and ME and BTS/SGSN. There are also modifications to the signalling between the nodes (HLR, ME, BTS/SGSN/MSC), and new key conversion functions to/from UMTS.

Remaining security issues:

- There is still no network authentication and it is still possible to do algorithm negotiation bidding down attacks. Should be combined with (8) to counter bidding down.
- Even when the AKA provides a 128-bit key, it is possible that an ME attaches to a Rel 98- network that only supports a weak 64-bit encryption algorithm, e.g., A5/1. If the conversion function for the encryption key is not carefully done, this potentially leaks 64 bits of the original key. See further 11.12.
- Most non-confidentiality issues remain.

Advantages:

Possible disadvantages:

- Has less value without requiring (1).

Main vulnerabilities countered:

- UP1
- SD2 (partially)

Can benefit from (2) and requires that a mobile that supports this GERAN enhancement, shall support USIM interface which is fulfilled already in Rel-5.

11.6 A/Gb tunneling within IPsec

Implications: Nodes that are affected are MS and potentially base station, SGSN, MSC and GGSN. New nodes, GANC and AAA server, needs to be implemented in the system.

It is recognised that the following issues should be resolved:

- The use of the IP Stack for transporting GSM protocols when the connection is not via ADSL or Bluetooth WiFi (impact on BS handover etc)
- Whether the same security association is used or a new one negotiated
- If and how certificates are provisioned on the GANC and how revocation is handled.

Remaining security issues:

Advantages:

- Support for the Generic Access to A/Gb security solution may already be provided in future handset designs and in an operator's core network.

Possible disadvantages:

- Overhead in terms of bandwidth and processing.

Main vulnerabilities countered:

- UP1
- UP2
- SD2
- SD4
- MS6
- RS3
- R10
- RS8

.

11.7 Develop new encryption algorithms

Implications:

- Same as 11.5 and in addition tyo this cryptographical development work needed (though the forthcoming UEA2 UMTS algorithm could be considered a complimentary candidate for A5/5).

Remaining security issues:

- There is still no network authentication and it is still possible to do algorithm negotiation bidding down attacks. Should be combined with (8) to counter bidding down.
- Most non-confidentiality issues remain.
- Also see Section 11.12.

Advantages:

Possible disadvantages:

Main vulnerabilities countered:

- UP1 (Does not prevent bidding down attacks etc.)
- SD2 (Same as for UP1 above)

11.8 Disabling insecure GERAN encryption algorithms

Implications:

- Upgrades of ME, base stations, MSC and SGSNs required.

Remaining security issues:

- Most non-confidentiality related issues remain.
- Also see Section 11.12.

Advantages:

Possible disadvantages:

Main vulnerabilities countered:

- UP1 (Most attacks are made impossible, though security provided by 64-bit keyed A5/3 is still considered marginal.)
- SD2 (All attacks except the ones where the key is retrieved during network transport and application layer cryptanalysis are countered.)

11.9 Integrity protection of broadcast signalling

Implications:

- Special key management solution needs to be developed. Potentially a new node is needed for handling the key management.
- Upgrades of at least ME, BTS, and SGSNs required.

Remaining security issues:

- Most confidentiality related issues remain.
- Key separation issues remain.
- Also see Section 11.12.

Advantages:

Possible disadvantages:

Main vulnerabilities countered:

- RS10

Editor's note: This could also be beneficial for UTRAN.

11.10 Add integrity protection to user payload

Implications:

- Assuming that the integrity protection terminates at the same point as the ciphering, the same nodes are affected as in 11.4.

Remaining security issues:

- Most non-confidentiality related issues remain.
- Also see Section 11.12.

Advantages:

- Robust protection

Possible disadvantages:

- Increase of bandwidth usage.
- Effects of transmission bit errors needs to be studied.

Main vulnerabilities countered:

- UP2 (The attacker must now in addition to breaking Kc break the integrity protection).

11.11 Summary of technical Feature dependencies

- A solution for (1), (2), (3), (4), (5), (6), (7) and (8) may all require to take 11.12 into consideration.
- Enhancements (3) (for algorithm separation), and (5) may be circumvented if (1) is not provided.

- Feature (2) is useful for (5).
- Enhancements (3) (for domain separation) does not require (1).

11.12 Secure implementation strategies for security enhancements

It is required that **a new mobile should still work in an old network**. Thus, when an upgraded mobile is roaming in an old network, the choices are to either accept the possibility of downgrading attacks or the mobile refusing connection, contradicting the requirement. If a mobile after a certain Release can assume that the enhanced security features are implemented in all networks, then communication not using these features (e.g. signalling integrity) can be rejected. This raises the same type of questions (e.g. shall the connection be rejected by user intervention, or automaticly) which came up when discussing the rejection of unciphered connections.

One solution is to upgrade the core/access network first before security enhanced mobiles can be used. (See footnote 2), and agreeing on a cut-off date, to guarantee that all necessary network/access features are available before upgraded mobiles need them.

Other possibility is to use the special-RAND to indicate the network capabilities E.g. by indication that an extension field follows e.g. in the Authentication and ciphering Request or in the cipher mode command. This provides a means to flexibly upgrade serving networks, with the additional advantage that much more data can be carried within such an extension. Impacts however are higher than the special-RAND solution as known from S3-030588 (with algorithm restriction list only). Could be used to protect against network supported security feature-set bidding down protection, as a vehicle to support more than secure algorithm restriction list transfer only. The Home network would need to store the Serving Network support, while the Serving network will control extention field values e.g. the permitted algorithm settings or other required fields.

It is also required that **an old mobile should still work in a new network**, and users cannot be forced to upgrade to new mobiles. It has to be ensured that a MITM does not pretend to be an old mobile not supporting some of the enhanced features.

12 Conclusions and Proposal

TBD

Annex A (informative): Change history

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New
2005-04					Initial version	0.0.0	0.0.1
2005-05					Updated to reflect changes agreed at SA3#38.	0.0.1	0.1.0
2005-07					Updated to reflect changes agreed at SA3#39 (including pCR:s S3-050373 and S3-050398, both after minor modifications).	0.1.0	0.2.0
2005-09					Merged in pCR S3-050512, modified according to comments at SA3#40.	0.2.0	0.3.0