3GPP TSG SA WG3 Security — SA3#36 **S3-041015**
23 - 26 November 2004
Shenzhen, China

Source: **Ericsson**

Title: **Access Security Review**

Document for: **Discussion**

Agenda Item: **6.6 GERAN network access security**

1 Introduction

In a previous contribution, [15], Ericsson proposed a new work item to study long-term security enhancements to GERAN and UMTS access security. The following is a draft TR *[ed note: currently on "outline" level]* for discussion and an invitation to further contributions towards finalizing the study.

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.

2 Definitions and Terminology

GERAN security: tbd

UMTS security: tdb

GSM security context: see [19].

UMTS security context: see [19].

AG: 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. *[ed note: should we be even more general and open up for terminating GSM security in another place? Seems a quite tough change…]*

AN: Access Network.

CN: Core Network.

DoS: Denial of Service.

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

SSO: Single Sign-on.

TE: 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.

~~2~~3 Scope

Editor's note: this section should define the scope of the study

The high-level objective is to analyze the potential vulnerabilities and threats coming from the re-use of security context between GSM, GPRS (and other access networks) in the absence of security features such as strong algorithms, network authentication, key separation, etc. Interaction between GERAN/UMTS during hand-over is also to be considered. Finally, the security objectives under which UMTS security was designed are to be re-evaluated to assess the long-term security of UMTS in new attack scenarios and threat models, and effects of re-use of UMTS security context.

Details in the scope of the study are to

- Re-assess the security objectives for GERAN and UMTS security, i.e. consider whether the security objectives that were deemed required at the time GERAN/UMTS security was designed is still sufficient or not.
- Perform a threat, risk and vulnerability analysis of GERAN and UMTS 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].
- 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 GERAN/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 Y. 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 UMTS) that reduces relevant risks/realistic threats.

Specific issues to consider include (but are not limited to):

- The need and feasibility for network authentication, replay protection and key separation
- Need and feasibility for integrity protection of important signaling messages
- Effects of a near-future break of A5/1, GEA1 and/or other algorithms.
- Risk assessment of implications of "two-time-pads"
- Ensure protection both for the PS and CS domains, in particular that possible insecurity does not spread across domains.
- Study effects relating to inter-system handover and security context re-use or re-mapping.
- Consider new threats, e.g. caused repudiation scenarios and effects of using GSM/UMTS security context for other accesses, e.g. WLAN
- Consider security "bottlenecks" arising from different sizes of keys and other security parameters in the various access types, in particular enhancing the GERAN radio interface ciphering mechanism so that it supports key lengths of up to 128 bits.
- Study the possibility of using AKA and AKA based applications for enhancing security.

3.1 Non-issues

The following issues are explicitly left outside of the scope:

- DoS attacks of pure radio interference or "jamming" nature.

- Cryptographic analysis of individual algorithms.
- …

4 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 GSMGPRS//UMTS networks that are part of the study. Also, we may need to make *assumptions* about changes that may have been introduced in the time frame of Release Y, e.g. the disabling of A5/2 in R6.

Next, we define which *assets* that are desirable to protect and classify them according to sensitivity/value.

The consequences of the attack on A5/2 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, we define which *attackers/threat agents* we consider relevant, and classify them into categories.

Finally, we define a t*rust model* reasonable for use in Release Y time frame. 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.

5 System Description

5.1 Assumptions

<Editor's note: list assumptions, e.g. that A5/2 have been removed in this time-frame>

We shall make the assumption that the following holds for 3GPP Release Y:

1. A5/2 has been disabled from TEs.
2. Any possible new access technology that uses GSM/UMTS security context information has a well-defined AG.
3. …

5.2 System

The system under study consists of

- the AN user and control plane traffic (GERAN/UTRAN/WLAN,…), from TE to the AG,
- the security processing in AuC, MSC/VLR, BSS, AG, and TE,
- any application service relying on GSM/UMTS security context, e.g. a server using SIM based SSO,
- …

5.3 Assets

<Editor's note: what are the assets we may want to protect>

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

Security context data: authentication vectors (triplets/quintuplets) anywhere (i.e. not necessarily restricted to AN).

Security signalling: signalling in the AN directly related to security: algorithm selection/cipher mode command, authentication exchange, etc. *[ed note: can we make an explicit list, at least in the case of GSM/UMTS ?]*

Control signalling: any other signalling in the AN, e.g. radio resource management, etc. *[ed note: should this be further classified, e.g. signalling "releasing" the phone from the network ?]*

…

<Editor's note: a "value" classification is missing>

5.4 Actors and Threat Agents

<Editor's note: who is the attacker (mafia, terrorists ,what resources?)>

The following are the main actors, which to varying extent (see next section) are trusted by each other:

Subscriber: …

Home network:

Visited (access) network:

…

The (untrusted) threat agents are classified as follows:

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.

…

5.5 Trust Model

<Editor's note: in the time-frame of Release Y, how much do operator trust subscriber, does home trust visited,WLAN access operator, UMA access,… etc>

TBD.

46 Security Objectives

<Editor's note: list the "standard" security objectives, then those that were considered when designing GERAN/UMTS, also list those that may now have shown up as new requirements>

In the following, we re-asses the need for the most common security objectives, some of which were considered irrelevant or found to be met implicitly when designing GSM security.

6.1 Confidentiality

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

6.2 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. The need for more robust charging may be reason to re-assess this requirement.

6.3 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).

6.4 Non-repudiation

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. This model has changed slightly in IMS access, where the home network performs the authentication, and there may be need to re-assess also if this objective should be more strongly enforced.

6.5 Availability

TBD.

47 Known Vulnerabilities

<Editor’s note: list the known GERAN/UMTS vulnerabilities and possible impact>

7.1 Cryptographic Algorithm Vulnerabilities

7.1.1 Weak ciphering algorithms

Any access or application security solution which uses GSM/UMTS security context and a weak algorithm potentially jeopardizes confidentiality and possibly also a spread of this problem to other accesses using these contexts. 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 (e.g. [4]: about 20 seconds of known plaintext, and a ten minutes computational effort) raises the question how long we can trust A5/1. The main thing that protects GEA1 is probably the fact that the algorithm is still not known to the general public.

7.1.2 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 possible 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 generally recommendation for secure key size for the foreseeable future would be around 100 bits, and 128 bits may be a practical choice.

7.1.3 Weak AKA algorithms

Weak (GSM or UMTS) AKA algorithms could make responses or cipher keys predictable or even reveal parts of the subscriber key, Ki. Though the AKA algorithms are not standardized, effects of proprietary weak AKA algorithms (e.g. COMP128) needs consideration.

7.2 Cryptographic Protocol Vulnerabilities

7.2.1 Lack of network authentication and replay protection

7.2.2 Lack of integrity protection

<Editor's note: e.g. signaling/algorithm selection>

7.2.3 Key (in)separation

<Editor's note. Also WLAN use of SIM etc>

By key separation we mean 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 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.

7.2.3.1 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 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 had been specified:

Kc = CK XOR IK.

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 deduce from CK. However, suppose an active attacker fools the TE into making an hand-over to GSM. The ciphering there will take place using Kc as above, which can similarly be retrieved by breaking A5/4. However, 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 hijack the UMTS session.

Admittedly, this is a highly hypothetical example, but it again shows that key derivation choices are not arbitrary.

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".

7.2.4 Two-time Pads

<Editor's note. GPRS PS h.o. issues etc>

GSM/UMTS (for good reasons) relies on stream ciphers. These are vulnerable to replay attacks, causing so-called two-time-pads. On 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. email) 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 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, and may not reach the expected 128-bit level.

7.3 Repudiation scenarios

<Editor's note. Access network falsely claims it has a roaming subscriber>

There is a trend towards decreased trust in the visited network. E.g. 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. Note that non-repudiation can in this case be achieved with lightweight symmetric (SIM based) techniques without the need for PKI.

7.4 Potential Vulnerabilities with Suggested Enhancements

<Editor's note. Some problems with special RAND etc>

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

7.4.1 Special RAND

The special RAND solution decreases the maximum entropy of the RAND from 128 to about 80 bits. Thus, it also decreases the theoretically effective key-space of Kc by the same amount. However, assuming a secure A3/A8 implementation, it cannot be exploited in practice. 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.

One issue has been observed though. Namely that special RAND may open up attacks based on the Wagner et al attack on COMP128, [20].

7.5 Other potential vulnerabilities

A potential threat scenario is discussed in [11]. TBD.

48 Threat and Risk Analysis

<Editor's note: pretty standard>

8.1 Threat Analysis

8.2 Risk Analysis

9 Overview of Possible Enhancements

10 Feasibility Study

~~4~~11 Conclusions and Proposal

~~5~~12 References

<editor's note: should be set in alphabetical order>

[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 “, S3-04xxxx, 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] http://www.isaac.cs.berkeley.edu/isaac/gsm.html