A USIM compatible 5G AKA protocol with perfect forward secrecy

Jari Arkko, Karl Norrman, Mats Näslund and Bengt Sahlin
Ericsson Research
{jari.arrko, karl.norrman, mats.naslund, bengt.sahlin} @ ericsson.com

***Abstract*—In this paper, we present constructions for 3GPP Authentication and Key Agreement (AKA) that provides Perfect Forward Secrecy for the session key. Further, the constructs prevents an attacker, with access to the long-term pre-shared key, from simply eavesdropping the challenge RAND in the AKA run, and use the RAND and long-term pre-shared key to derive the session key. We focus on making it possible to re-use large portions of the current structure of 3GPP interfaces and functions, with the rationale that this will make the constructions more likely to be adopted by the industry. In particular, the constructions maintain the interface between the Universal Subscriber Identification Module (USIM) and the mobile terminal intact. As a consequence, there is no need to roll out new credentials to existing subscribers.**

Keywords—5G; 4G; LTE; security; authentication; AKA; mobile networks; telecom; perfect forward secrecy; 3GPP

I. INTRODUCTION

Current 3GPP systems use (U)SIM pre-shared key based protocols to authenticate subscribers. Since the addition of replay protection and mutual authentication in the third generation 3GPP systems, there have been no published attacks that violate the security properties defined for the Authentication and Key Agreement (AKA) in [1], at least not within the assumed trust model. However, there have been attacks using a different trust model [3, 7]. Of course, the protocol was not designed to counter those situations. There have also been attacks against systems where AKA is used in a different setting than initially intended, e.g. [5].

Recent reports of compromised long term pre-shared keys used in AKA [2] indicate a need to look into solutions that allow a weaker trust model, in particular for future 5G systems. It is also noted in [2] that, even if the current trust model is kept, some security can be retained in this situation by providing Perfect Forward Security (PFS) [4] for the session key. If AKA would have provided PFS, compromising the pre-shared key would not be sufficient to perform passive attacks; the attacker is, in addition, forced to be a Man-In-The-Middle (MITM) during the AKA run. Introducing PFS for authentication in 3GPP systems can be achieved by adding a Diffie-Hellman (DH) exchange. While PFS as such does not necessarily protect against passive attackers, DH does.

There are several reasons why adding PFS to AKA may not be a trivial task. The main theme behind the reasons is backwards compatibility with already deployed mobile terminals and networks. Since the functions involved in the infrastructure supporting AKA are distributed globally over several hundreds of operators and are also included in the mobile terminals and USIMs, protocol enhancements has to be made in such a way that not all equipment have to be updated simultaneously. In particular, enhancements which rely on that entities signal support for a new capability, and doing so without proper integrity protection, are a well-known source for bidding-down attacks. Keeping the world-wide installed base of 3GPP 2G, 3G and 4G systems in mind, it is unlikely that the investment in the necessary upgrades to all nodes, mobile terminals and USIMs (and their physical encapsulation) would be seen as worthwhile. However, for 5G, it may be easier to add new functionality from the start. The vast amount of deployed USIMs will most likely lead to support for USIMs based authentication in 5G as well.

A. Our contributions

We present constructions for adding PFS and protection against passive attackers who have compromised the pre-shared key. The constructions respect the 3GPP authentication architecture and maintain the interface between the mobile terminal and the USIM. Therefore, the new constructions can re-use already deployed USIMs. Note that we only consider the core AKA protocol as used for cellular access; we do not cover AKA as used in other technologies such as the Internet Protocol Multimedia Subsystem (IMS) [9], the Generic Bootstrapping Architecture (GBA) [10], or in EAP AKA [11].

B. Related work

There is a large body of work related to attacks and modifications to the 3GPP authentication protocol in general, [3, 5, 7, 13, 14, 15] to name a few. Not all explicitly discuss the specific properties in this paper, i.e., PFS and prevention of passive attacks on the air-interface when the pre-shared key is compromised. However, the following do.

Reference [12] proposes to run a password authenticated DH-based protocol between the mobile terminal and the authenticator node in 4G. The password corresponds to the pre-shared key used in 4G and it has to be accessible in the authenticator node. Since the 4G security architecture intentionally protects the pre-shared key by keeping it only in the subscriber database in the home network domain and in the USIM, this proposal provides a lower level of security than we aim for in this paper.

SE-AKA [13] is an interesting authentication scheme that explicitly seeks to provide PFS. However, the scope is much wider than what we cover. For example, SE-AKA can cope with groups and, to this end, requires new server functions, new interfaces and PKI.

Transport Layer Security (TLS) [8] since long supports PFS through the use of ephemeral DH. TLS is, however, a very self-contained protocol with large overhead in terms of number of messages as well as size of messages, at least when compared to pure AKA. TLS also ties the algorithm negotiation and authentication signaling rigidly to the secure channel. Security for 3GPP networks is much related to flexibly being able to change algorithms, evolve keys at mobility events and dynamically derive keys in a hierarchical fashion to support different security associations in different parts of the 3GPP architecture. Hence, TLS would require major reworking to be able to cater for the high-mobile and layered environment of 3GPP systems. Therefore, it does not seem like a ready to use alternative for 3GPP authentication.

II. BACKGROUND

While we believe 5G networks will have, at least somewhat, different architecture compared to previous generations, the main structure of a USIM-based access authentication architecture will probably remain intact. That is, there will be a database in the home network storing the subscriber credentials, the Home Subscriber Server (HSS) in 4G; there will be an authenticator in the serving network, the Mobility Management Entity (MME) in 4G; and there will be a Mobile Entity (ME) equipped with a USIM. For simplicity, we will in the following use the function names from 4G. We use the terms 2G, 3G and 4G to refer to the corresponding generations of the 3GPP defined systems.

A. 3GPP Autentication and Key Agreement

The AKA protocol is a challenge-response protocol based on symmetric key cryptography; it involves multiple parties. The term AKA here refers to 3G AKA [1] (see Fig. 1). 3G AKA can be run in 2G systems and is effectively the same authentication protocol used in 4G.

1) The MME initiates AKA by sending a request for an Authentication Vector (AV) associated with a particular subscriber to the HSS.
2) The HSS responds with an AV consisting of the tuple (RAND, AUTN, XRES, K_{ASME}), where RAND is a random value, AUTN is a network authentication token, XRES is the expected response from the ME and K_{ASME} is the session key that will be established in the ME and MME on completion of the AKA run. The AUTN, XRES and K_{ASME} are derived from the RAND and the pre-shared key K present in the HSS and in the USIM. As an intermediate step in the derivation of K_{ASME}, keying material named CK/IK is produced.
3) When the MME receives the AV, it can initiate the authentication procedure with the ME by forwarding the RAND and AUTN.
4) The ME sends the RAND and AUTN to the USIM, which verifies the authenticity and freshness of the AUTN parameter. If the verification succeeds, the USIM derives a response parameter RES and keying material CK/IK from K stored in the USIM and the received parameters.
5) The USIM then forwards the RES and keying material CK/IK to the ME.
6) The ME derives K_{ASME} from the CK/IK and sends RES to the MME.
7) The MME verifies that RES is equal to XRES and accepts the authentication if so. Otherwise the MME rejects the authentication.

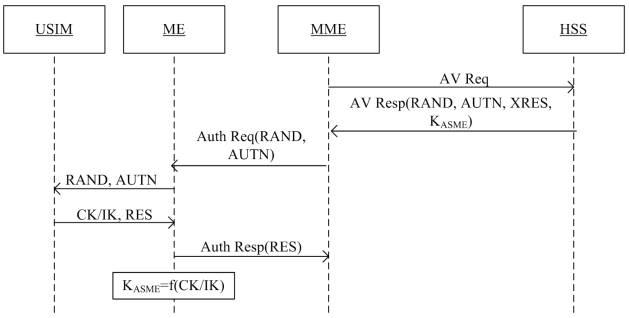

Fig. 1. The flow of AKA in an 4G setting. The difference compared to 3G AKA is the derivation of K_{ASME} from the keying material CK/IK.

AKA uses a sequence number mechanism to guarantee the freshness of the RAND and AUTN parameters to the ME/USIM. The sequence number is carried in the AUTN parameter, and the ME/USIM rejects the authentication should the sequence number be considered old. AKA further include a re-synchronization mechanism for the sequence number. Since our constructions are based on ephemeral DH both ME and network gets assurance of fresh key generation. Our constructions does hence not rely on the AKA sequence number for session key freshness, and we will therefore not discuss the synchronization of the sequence number further.

B. Trust model for 3GPP AKA

We now summarize the trust model under which AKA operates. This is also the trust model that our enhanced AKA follows. After presenting the enhancements we will discuss some implications of this trust model in the analysis section.

USIM and the baseband of the ME are assumed to be honest. No attacker is assumed to have access to the ME-USIM interface. MME and HSS are also assumed to be honest. Since security needs to be terminated in the visited network to avoid home-routing of all traffic, the MME has to be trusted with the data it is given, i.e., the AV. Neither ME nor MME is trusted with K. The key K is kept secret in the USIM and the HSS. The communication between HSS and MME is assumed to be integrity, replay and confidentiality protected. The link between the MME and the ME is assumed to be without any kind of protection. Clearly, the confidentiality and integrity of K, have to be ensured when it is provisioned into the HSS and

the USIM. This implies that if K is compromised, then an attacker only needs to eavesdrop RAND to be able to derive the encryption and integrity key for the air-link protection. It is a goal of this paper to prevent this passive attack.

III. ENHANCEMENTS TO 3GPP AKA

The enhancements presented in in this paper are compatible with the signaling flow and other basic structures of AKA. More precisely, the purpose of the protocol is to achieve mutual authentication between the MME and the ME, and to establish keying material for secure communication between the two. For simplicity we refer to this keying material as K'_{ASME}. The enhancements are in the form of new properties for the K'_{ASME} compared to the properties of the K_{ASME}. Furthermore, the protocol follows the AV generation and distribution approach from MME and HSS point of view. Finally, it is based on a pre-shared key and may re-use already deployed USIMs.

In addition to the security goals stated for AKA in [1], we introduce the goals to achieve PFS and infeasibility of passive attacks on the interface between ME and MME. PFS is defined as the property that a session key remains secret even though the long-term key is compromised in the future [4].

A. Protocol constructions

The basic idea is to run DH between ME/USIM and MME, and authenticating the DH-exchange using session key material derived from K. Depending on how the authentication of the DH-exchange is done, the HSS may participate in calculating the ephemeral DH-parameters g^x and g^y. The DH-exchange is embedded in the Authentication request/Authentication response messages of the Non Access Stratum (NAS) protocol that run between the MME and ME. The parameters defining the DH-group, e.g., the generator g and prime modulus p for a mod p DH-group need to be known to all parties prior to running this protocol. In the sequel, we assume this is so and leave the provisioning for future work. We now describe the two main options, A and B. Option A requires no modification to the HSS, whereas option B does.

1) In option A, depicted in Fig. 2, the MME request a regular AV from the HSS. The MME then generates an ephemeral DH-parameter g^x and calculates a Message Authentication Code (MAC_g^x) over the g^x using K_{ASME} as a key.
2) The MME next sends g^x and MAC_g^x together with the RAND and AUTN from the AV to the ME.
3) The ME forwards the RAND and AUTN to the USIM, which verifies the AUTN and calculates the keying material CK/IK and the response parameter RES; these parameters are then forwarded to the ME.
4) Next, the ME calculates the K_{ASME} and generates g^y and a corresponding MAC_g^y. The ME concludes by sending the RES and MAC_g^y to the MME so that the MME can verify the RES. Both the ME and MME now calculate a K'_{ASME} from the DH-key g^{xy}. The K'_{ASME} can at this point be used as a basis for a key hierarchy. Before using the K'_{ASME}, both parties verify the MAC of the other party's ephemeral DH-parameter.

Option B, depicted in Fig. 3, uses a slightly different approach and roots the network ephemeral DH-parameter in the HSS rather than in the MME. Option B works as follows.

1) The MME requests a different type of AV from the HSS. This type of AV also contains an x generated by the HSS, and the RAND value is generated by computing a hash over g^x. Because the HSS includes x in the AV, the MME will be able to calculate $(g^y)^x$ in step 5.
2) Upon receipt of the AV, the MME transmits the RAND, AUTN and g^x to the ME.
3) The ME forwards the RAND and AUTN to the USIM which verifies the AUTN, calculates CK/IK and RES, before forwarding these to the ME.
4) Since the AUTN is calculated in dependence of the RAND, which in turn contains the hash of g^x, this provides authentication of g^x to the USIM, and hence indirectly to the ME based on the trust model.
5) Next, the ME generates g^y, calculates the corresponding MAC_g^y, and sends the g^y, MAC_g^y and RES to the MME. The ME and MME calculates a K'_{ASME} from the g^{xy} as in option A.

To save signaling over the air-interface the ME can calculate RES' = f(RES, g^y) for a suitable function f, and send RES' to the MME instead of RES; the ME would omit sending MAC_g^y in this case. The MME can do the same calculation of RES' and hence verify both RES and the authenticity of g^y simultaneously.

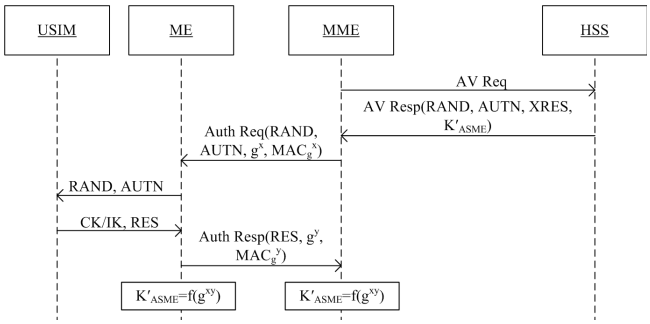

Fig. 2. Option A for DH-integration in 3GPP authentication.

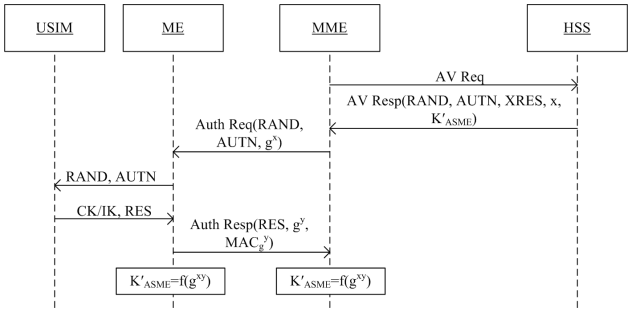

Fig. 3. Option B for DH-integration in 3GPP authentication

IV. ANALYSIS

We now analyze the proposed constructions in terms of security, impact on existing functions and residual threats.

A. Security

Both option A and B derives the K'_{ASME} partially based on data that was freshly generated during the run of the authentication protocol between the MME and the ME. Assuming honest participants, both parties are therefore assured of the K'_{ASME}'s recent generation. This is not the case for AKA, where K_{ASME} can have been generated by the HSS long before the run of the NAS part of the authentication protocol. In option A, both g^x and g^y will be recent by the assumption on ME and MME being honest. Each party is also guaranteed on freshness of K'_{ASME} since they each have control over their contribution to the key derivation in form of their DH-parameter. In option B, only g^y generated by the client has recentness guarantees, but assuming an honest ME, the attacker still needs to be a MITM to obtain K'_{ASME}.

It may be tempting to avoid generating y for g^y as a random value in the ME and rather use a function of one of the parameters output from the USIM, e.g., y = f(CK, IK) for some function f. However, the CK and IK are completely determined from the RAND and the K, so a passive attacker with access to these could eavesdrop g^x and calculate g^{xy}. Therefore it would be completely insecure to generate g^y in this way.

The authenticity of the DH-parameters differs between the two options. In option A, the USIM/ME gets the assurance that it is the MME that generated g^x both freshly and recently. In option B, the USIM/ME gets the assurance that it is the HSS that freshly, but not necessarily recently, generated g^x. Although we have not identified any attacks based on either observation, this is an interesting asymmetry that deserves further study.

B. Impact on invovled entities

As mentioned above, we believe the structure of the AKA protocol, in terms of a home network subscription database, a serving network authenticator and a USIM equipped mobile terminal will be used in 5G. We now analyze the impact on the corresponding functions in the 4G system, should they be evolved to support the new constructions.

The HSS-MME interface requires no updates for option A, but a new type of AV is required to carry g^x for option B. The latter option also requires the HSS to generate the RAND in a specific way. The interface between the ME and the USIM remains intact even though the procedure for generating the RAND in option B is given a specific implementation. Therefore, legacy USIMs can be re-used. The ME and MME require updates to cater for the new information elements required to carry the new parameters in the authentication procedure of the NAS protocol. However, no new NAS procedures are required.

C. Residual threats

As previously pointed out, active MITM attacks on the interface between the MME and ME are still possible in case K is compromised. While not being a goal of this study, we note that the proposed constructions do not prevent an attacker from directly requesting the current K'_{ASME} from the MME once authentication and key establishment is completed, similar to [6]. The constructions neither protect from an attacker requesting new AVs from the HSS and eavesdropping and re-routing AVs in transit from the HSS to MMEs [7]. The two latter attacks are assumed to not be possible according to the trust model defined above. It could be argued that this trust model needs to be updated for 5G; this is interesting future work.

The trust model assumes that the baseband implementation of the ME is honest. Even though the baseband is often much better protected than the execution environment running applications in the mobile terminal, an attacker may still affect the baseband code. With the progress of hacking attempts against mobile terminals this is an indication that a higher level of platform security may be needed to maintain the trust model. Further, the trust model assumes that the ME-USIM interface is protected from eavesdropping and manipulation. By default, this interface is only protected by the physical proximity of the two entities. Devices such as TurboSIMs can be used to eavesdrop and manipulate data transmitted over this interface. It is less clear how easily accessible the interface is without having physical access to the mobile phone. Even so, standardized techniques exist to protect the interface [8]. It may be wise to mandate the use of similar technology to protect the ME-USIM interface by default in 5G, at least for access authentication signaling.

V. CONCLUSIONS

We have shown that it is possible to introduce PFS and protection against passive attacks on the air-interface protection keys into the existing architecture and protocol structures shared by previous generations of 3GPP networks. The constructions do not affect the ME-USIM interface, and therefore allows re-use of the existing deployed USIMs.

Though threats still remain to the session key K'_{ASME}, we conclude that it is possible to limit the effects of a compromise of K (even if it happens already at manufacturing time), and that doing so is possible with relatively little impact on the traditional 3GPP architectures. However, due to backwards compatibility issues, it is probably not practically possible to introduce this in legacy systems, but it is rather functionality for 5G.

REFERENCES

[1] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; "Formal Analysis of the 3G Authentication Protocol," 3GPP TR 33.902 version 4.0.0, September 2001.

[2] J. Scahill and J. Begley, "The great SIM heist," February 19, 2015. [Online]. Available: https://firstlook.org/theintercept/2015/02/19/great-sim-heist/. Accessed on: April 17, 2015.

[3] A. Choudhary and R. Bhandari, "3GPP AKA Protocol: Simplified Authentication Process," International Journal of Advanced Research in

Computer Science and Software Engineering, Volume 4, Issue 12, December 2014.

[4] W. Diffie, P. van Oorschot and M. Wiener, ”Authentication and Authenticated Key Exchanges,” Designs, Codes and Cryptography 2 (2): pp. 107–125, June 1992.

[5] J. G. Beekman and C. Thompson, ”Breaking Cell Phone Authentication: Vulnerabilities in AKA, IMS and Android,” 7th USENIX Workshop on Offensive Technologies, WOOT '13, August 2013.

[6] K. Nohl, ”Lecture: Mobile self-defense,” December 27, 2014. [Online]. Available: http://events.ccc.de/congress/2014/Fahrplan/events/6122.html. Accessed on: April 27, 2015].

[7] S. F. Mjølsnes and J-K Tsay, ”A vulnerability in the UMTS and LTE authentication and key agreement protocols,” Proceedings of the 6th international conference on Mathematical Methods, Models and Architectures for Computer Network Security: computer network security, October 2012.

[8] Internet Engineering Task Force, ”The Transport Layer Security (TLS) Protocol Version 1.2,” RFC 5246, August 2008.

[9] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; ”Access security for IP-based services,” 3GPP TS 33.203 V12.8.0, December 2014.

[10] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Generic Authentication Architecture (GAA); ”Generic Bootstrapping Architecture (GBA),” 3GPP TS 33.220 V12.3.0, June 2014.

[11] Internet Engineering Task Force, ”Improved Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA'),” RFC 5448, May 2009.

[12] C-E. Vintila, V.V Patriciu and I. Bica, ”Security analysis of LTE access network,” ICN 2011, The Tenth International Conference on Networks, 2011.

[13] N.Saxena, N.S. Chaudhari, ”NS-AKA: An Improved and Efficient AKA Protocol for 3G (UMTS) Networks,” CSEE 2014 , Proc. of the Intl. Conf. on Advances in Computer Science and Electronics Engineering, 2014.

[14] M. Purkhiabani and A. Salahi, ”Enhanced Authentication and Key Agreement Procedure of next Generation 3GPP Mobile Networks,” International Journal of Information and Electronics Engineering, Vol. 2, No. 1, January 2012.

[15] M. Zhang and Y. Fang, ”Security Analysis and Enhancements of 3GPP Authentication and Key Agreement Protocol,” IEEE Transactions On Wireless Communications, Vol. 4, No. 2, March 2005.