Another Look at Privacy Threats in 3G Mobile Telephony

Mohammed Shafiul Alam Khan^{*} and Chris J Mitchell

Information Security Group, Royal Holloway, University of London Egham, Surrey TW20 0EX, United Kingdom shafiulalam@gmail.com, me@chrismitchell.net

Abstract. Arapinis et al. [1] have recently proposed modifications to the operation of 3G mobile phone security in order to address newly identified threats to user privacy. In this paper we critically examine these modifications. This analysis reveals that the proposed modifications are impractical in a variety of ways; not only are there security and implementation issues, but the necessary changes to the operation of the system are very significant and much greater than is envisaged. In fact, some of the privacy issues appear almost impossible to address without a complete redesign of the security system. The shortcomings of the proposed 'fixes' exist despite the fact that the modifications have been verified using a logic-based modeling tool, suggesting that such tools need to be used with great care.

1 Introduction

The 3GPP/ETSI 3G standards, which incorporate a range of security features [2,3], are the basis for a large part of the world's mobile telephony. As a result, any security or privacy flaws identified in these standards potentially have major implications.

We are primarily concerned with one particular feature of 3G security, namely the service known as *user identity confidentiality*. This service seeks to minimise the exposure of the mobile phone's long term identity (actually the long term identity of the USIM within the phone) on the *air interface*, i.e. the radio path between the phone and the network. The main security feature incorporated into the 3G system designed to provide this service is the use of frequently changing temporary identities, which act as pseudonyms.

A recently published paper by Arapinis et al. [1] describes two novel attacks on this service, which enable user device anonymity to be compromised. As well as describing the two attacks, modifications ('fixes') to the protocol are described which aim to prevent the attacks, and verifications of these fixes using ProVerif are also outlined.

^{*} The first author would like to acknowledge the generous support of the Commonwealth Scholarship Commission.

This paper has the following main objectives. Firstly, the proposed fixes are re-examined, and are found to have significant shortcomings. Secondly, possible alternative approaches to some of the modifications are noted. Thirdly, it is argued that some of the weaknesses in user identity confidentiality are impossible to fix, meaning that making significant system changes to address some of them are unlikely to be worth the effort. Finally, conclusions are drawn about the effectiveness of tools such as ProVerif if not used with appropriate care, and in particular if used without a detailed understanding of the cryptographic primitives being used.

The remainder of the paper is structured as follows. In section 2 the key features of the 3G security architecture are briefly reviewed. The attacks of Arapinis et al. are then summarised in section 3, together with a description of their proposed fixes. Sections 4 and 5 provide an analysis of the 'fixes'. Finally, the findings of the paper are summarised and conclusions are drawn in section 6.

2 Relevant 3G Security Features

The purpose of this section is to introduce those 3G security features of relevance to this paper. Our description follows Niemi and Nyberg [3], and we use their notation.

2.1 The AKA Protocol

At the core of 3G air interface security is a mutual authentication and authenticated key establishment protocol known as AKA (Authentication and Key Agreement). This is regularly performed between the visited network and the mobile phone (the *user equipment (UE)*). It involves the network sending an *user authentication request* to the UE. The UE checks the validity of this request (thereby authenticating the network), and then sends a *user authentication response*. The network checks this response to authenticate the UE. As a result, if successful, the two parties have authenticated each other, and at the same time they establish two shared secret keys.

In order to participate in the protocol, the UE — in fact the User Subscriber Identity Module (USIM) installed inside the UE — must possess two values:

- a long term secret key K, known only to the USIM and to the USIM's 'home network', and
- a sequence number SQN maintained by both the USIM and the home network.

The key K never leaves the USIM, and the values of K and SQN are protected by the USIM's physical security features.

The 48-bit sequence number SQN is used to enable the UE to verify the 'freshness' of the user authentication request. More specifically, the request message contains two values: RAND and AUTN, where RAND is a 128-bit random number generated by the home network, and the 128-bit AUTN consists of the

concatenation of three values: $SQN \oplus AK$ (48 bits), AMF (16 bits), and MAC (64 bits). The value AMF is not relevant to our discussions here, and we do not discuss it further. The MAC is a Message Authentication Code (or tag) computed as a function of RAND, SQN, AMF, and the long term secret key K, using a MAC algorithm known as f1. The value AK, computed as a function of K and RAND, essentially functions as a means of encrypting SQN; this is necessary since, if sent in cleartext, the SQN value would potentially compromise user identity confidentiality, given that the value of SQN is USIM-specific.

On receipt of these two values, the USIM uses the received RAND, along with its stored value of K, to regenerate the value of AK, which it can then use to recover SQN. It next uses its stored key K, together with the received values of RAND and SQN, in function f1 to regenerate the MAC value; if the newly computed value agrees with the value received in AUTN then the first stage of authentication has succeeded. The USIM next checks that SQN is a 'new' value; if so it updates its stored SQN value and the network has been authenticated.

If authentication succeeds, then the USIM computes another message authentication code, called *RES*, from K and *RAND* using another function f_2 , and sends it to the network as part of the user authentication response. If this *RES* agrees with the value expected by the network then the UE is deemed authenticated.

We note that if the authentication process fails for some reason, then the UE sends an error code (a *Failure Case Code*) as part of an *Authentication failure report*, sent instead of a user authentication response ([2], section 6.3.6). In particular, distinct error codes are sent to indicate an incorrect MAC and an incorrect SQN, i.e. depending whether the authentication process fails at the first or second stage.

Finally observe that the security properties of the AKA protocol itself have been proven to hold [4] — the problems we consider here arise from exchanges not actually part of the AKA protocol. This makes clear the necessity to consider the entirety of a system if robust results about security and privacy are to be achieved.

2.2 Session Keys

As part of a successful AKA procedure, the network and the USIM generate a pair of session keys, known as IK, the *integrity key*, and CK, the *ciphering key*. Both these keys are a function of K and RAND. The USIM exports these two keys to the UE. The IK is used for integrity protection of signalling messages sent across the radio path, and the CK is used for encryption of data sent across the air interface, using a stream ciphering technique.

2.3 User Identity Confidentiality

As mentioned previously, user identity confidentiality is provided by the use of temporary identities. Every USIM has a unique *International Mobile Subscriber Identity (IMSI)*. If this was routinely sent across the network then the UE,

and hence its owner, could be traced. As a result, every UE also possesses a *Temporary Mobile Subscriber Identifier (TMSI)* which is sent instead.

The value of the TMSI, which is chosen by the network the UE is visiting, is changed regularly. A new TMSI is sent by the network to the UE in encrypted form, protected using the CK.

3 Privacy Threats and Fixes

3.1 The Attacks

Arapinis et al. [1] describe two apparently novel attacks that breach user identity confidentiality in 3G mobile telephony. These two threats operate as follows (for further details see [1]).

- IMSI paging attack. This attack exploits a specific type of signalling message known as a Paging message (or, more formally, a PAGING TYPE 1 message see 8.1.2 of ETSI TS 125 331 [5]). Such messages are sent from the network to all mobile devices in a particular area, and can contain either an IMSI or a TMSI. If a UE detects such a message containing its IMSI or its current TMSI then it responds with a message containing its current TMSI. Most importantly, paging messages are not integrity protected (see 6.5.1 of ETSI TS 133 102 [2]), and hence a malicious third party can introduce spurious paging messages into the network. This can be used to both detect the presence of a UE with a specific IMSI, and also to learn the current TMSI for this device. This poses a threat to mobile identity privacy.
- $AKA \ error \ message \ attack$. This attack exploits the error messages incorporated into the AKA protocol, as described in section 2.1 above. Suppose an attacker has intercepted a genuine (RAND, AUTN) pair sent to a particular UE. If these values are relayed to a specific UE, two possible outcomes will arise. If the recipient UE is the device to which the (RAND, AUTN) pair was originally sent then it will respond with an Authentication failure report containing an error code indicating a failed SQN, i.e. to indicate that the pair has been received previously. Otherwise, the UE will respond with a failure report containing an error code indicating an incorrect MAC value. That is, the error code can be used to distinguish between UEs, and this is clearly another means of breaching user identity confidentiality.

3.2 Observations

We start by observing that the first threat, whilst apparently novel, is closely related to another threat to user identity privacy. As described in section 6.2 of ETSI TS 133 102 [2], 'when the user registers for the first time in a serving network, or when the serving network cannot retrieve the IMSI from the TMSI by which the user identifies itself on the radio path', the serving network must obtain the IMSI from the UE — this is performed using a User identity request/User identity response message pair, where the latter message contains the IMSI. 'This represents a breach in the provision of user identity confidentiality'. This attack, called *user identity catching* (or *IMSI catching*), is further mentioned in A.1 of ETSI TS 121 133 [6], and is also noted by Arapinis et al. ([1], section 2.2).

Given that this attack exists, i.e. an active attacker can obtain the IMSI of any UE by impersonating the network, neither of the new attacks appear to significantly weaken the user privacy service any further. That is, neither of the new attacks appear to be any easier to launch than the IMSI catching attack in particular, they both require active impersonation of the network.

Most interestingly, the second attack seems to be an issue that has not previously been discussed in the literature. It is just one example of a very broad class of threats arising from poorly designed error messages that reveal information of value to an attacker — see, for example, Vaudenay [7].

3.3 The Fixes

As well as describing the two privacy issues, Arapinis et al. [1] give three separate modifications to the operation of 3G mobile telephony designed to fix the two newly identified problems as well as the well known user identity catching attack. We next briefly describe these proposed modifications.

- Fixing the IMSI paging attack. This modification is not described in complete detail ([1], section 5.2), and as a result some suppositions need to be made. It involves cryptographically protecting the paging message using a secret key UK known only to the network and the UE. Like the CK and IK, this additional key is generated as a function of the RAND and K during the AKA protocol.

The paging message format is modified to incorporate two additional fields, namely a sequence number SQN and a random challenge CHALL. It is not clear whether SQN is in the same 'series' as the SQN sent in the AUTN of whether this is a distinct sequence number used for this purpose only. This issue is discussed further in section 4 below.

The entire paging message is then encrypted using UK. However, the method of encryption is not specified. This issue is also discussed further in section 4 below.

Since this message is broadcast, it is received by all UEs currently attached to a base station. Each UE must use its current UK to decrypt the message. By some (unspecified) means the recipient UE decides whether the decrypted message is intended for it or not — Arapinis et al. simply state ([1], section 5.2) that each UE 'has to decrypt and check all the received IMSI paging to determine if it is the recipient' (sic). If it is the intended recipient, then the UE checks the SQN against its stored value to verify its freshness (as in AKA). If it is fresh then the USIM updates its stored SQN, and sends a paging response containing the TMSI and the received value of CHALL; otherwise, if the freshness check fails, the paging message is ignored.

 Fixing the AKA error message attack. This fix involves leaving the 'normal' operation of AKA unchanged; the only modification is to require (asymmetric) encryption of authentication failure report messages, thereby hiding the nature of the embedded error message. This encryption is performed using a public encryption key belonging to the visited network. Providing a reliable copy of this key to the UE requires the pre-establishment of a Public Key Infrastructure (PKI) involving all the 3G network operators, in which each network operator has an asymmetric encryption key pair *and* a signature key pair. Each operator must use its private signature key to create a certificate for every other network's public encryption key. Every USIM must be equipped with the public signature verification key of the issuing (home) network.

In order for the UE to obtain a trusted copy of the appropriate public encryption key, the visited network must send a copy of a certificate for its public encryption key, signed using the private signature key of the USIM's home network (this could be achieved by modifying an existing signalling message or by introducing a new such message). The USIM exports its trusted copy of the public verification key of its home network to the phone, and the phone can use this to verify the certificate, thereby obtaining the required trusted public encryption key. The phone can perform the encryption of the failure report message, obviating the need for the USIM to perform any computationally complex asymmetric encryption operations.

A further modification to the failure report message is proposed by Arapinis et al. [1], namely to include the USIM's current value of SQN. This change is designed to enable resynchronisation of this value by the network, but is not explained further.

- Fixing user identity catching. Finally, Arapinis et al. [1] also propose modifying the procedure by which a UE identifies itself when first joining a network. They propose that the UE asymmetrically encrypts the User identity response message containing the IMSI. As in the previous modification, this encryption is performed using the public encryption key of the visited network.

4 IMSI Paging Re-Examined

There are a number of significant issues with the fix proposed to mitigate IMSI paging attacks. We enumerate some of the most serious.

1. Introducing a new type of session key, i.e. the UK, has major ramifications. To see why we first need to consider some issues surrounding the use of AKA. The long term K is not passed to a visited network. Instead, the home network of the USIM will generate, on request, sets of authentication vectors, i.e. 5-tuples (RAND, XRES, CK, IK, AUTN), which are passed to the visited network. Each 5-tuple contains a random RAND value and a distinct SQN value embedded in the AUTN. Note that several such 5-tuples will be passed to the visited network at the same time (to reduce the number of inter-network signalling messages), and the visited network must use them in the correct order, i.e. in ascending order of SQN values.

When it wishes to authenticate a UE, the visited network sends the (RAND, AUTN) pair from the 'next' authentication vector, and receives back *RES*, which it compares with the *XRES* value from the authentication vector (the 'expected value of *RES*) to authenticate the UE. Introducing an additional key type means that the authentication vectors will need to become 6-tuples to include the *UK* value, which will involve changing the formats of messages sent between networks (this is, in itself, a significant change).

- 2. As noted in section 3.3 above, there are two possible ways in which the SQN might be generated and managed. It could be generated and verified using the same mechanism as employed for the AKA protocol, or a separate sequence number scheme could be involved. Unfortunately, there are major implementation difficulties with both options.
 - (a) Using the same SQN values as are used in the AKA protocol is problematic. The visited network does not have a means of finding out these values, as they are not included in the authentication vectors sent to the visited network. Even if the current SQN value was sent as part of the authentication vector (which would affect the inter-network signalling infrastructure), two major problems remain. Firstly, if the visited network is permitted to generate new SQN values and have them accepted by the USIM, then this means that the visited network is able to modify the SQN value stored by the USIM. This could have the effect of invalidating any unused authentication vectors that the visited network retains for the UE. Secondly, giving the visited network the power to change the SQN value held by the USIM is a major change in the current trust model, and would give the visited network the power to deliberately or accidentally completely block the operation of the USIM by sending it a very large SQN value.
 - (b) Using a different SQN value also raises major issues, as there is no obvious mechanism to keep multiple networks aware of the current value of the SQN for a particular UE. This would require the home network to maintain the current value, and for visited networks to exchange messages with the home network to maintain synchronisation between the value held by the USIM and the home network.
- 3. The 'encryption' of the paging message appears to be intended to provide two distinct security services: (a) guarantees to the recipient regarding the origin and integrity of the message, and (b) confidentiality of the contents so that passive interceptors cannot observe the link between an IMSI and a TMSI. It is well known that simple encryption cannot guarantee property (a), especially if that means use of a stream cipher (see, for example, section 9.6.5 of Menezes, van Oorschot and Vanstone [8]). However, stream cipher encryption is the only encryption primitive available in the current 3G security architecture. Clearly what is really required is the application of an authenticated encryption technique [9], which would provide the necessary security guarantees. However, this is never made explicit by Arapinis et al. [1]. Their success in proving the security of the modification using ProVerif suggests that their input to ProVerif implicitly assumed the provision of

properties (a) and (b), whereas their description of the necessary modifications to the system did not make these requirements explicit. This shows the danger of not carefully considering and making explicit all the properties of the cryptographic primitives being employed.

Of course, the visited network and UE share a pair of keys (CK and IK) designed explicitly for confidentiality and integrity protection of data and signalling messages. A much simpler solution, which achieves precisely the same objectives, would be to first encrypt the paging message using CK and then generate an accompanying MAC using IK. This would both achieve the security objectives and avoid the need to introduce an additional key type.

4. Finally, we note that, even if it could somehow be repaired, the fix imposes very significant burdens on the system. As stated by the authors (final sentence of 5.2 of [1]) the overheads of the proposed modification are non-trivial. This is because every UE that receives a paging message is required to decrypt it and somehow verify whether or not it is intended for them.

In conclusion, the number and seriousness of the issues identified with the fix, especially relating to the use of the *SQN* sequence number, suggest that it cannot work in practice. Moreover, finding an alternative fix without completely redesigning the 3G system appears highly problematic. As a result it would appear that accepting that user identity confidentiality is imperfect seems inevitable, a point we return to below.

5 Addressing the Error Message and Identity Catching Attacks

In evaluating the fix proposed to address the AKA error message attack, we start by considering the practicality of introducing a brand new PKI. Whilst the required PKI is relatively small scale, involving only the network operators, introducing such a PKI would nevertheless involve significant changes to the operation of the system. In particular, over and above requiring changes to all phones, all USIMs and all networks, every USIM would need to be equipped with a public key, every network would need to exchange public keys and certificates with every other network, certificates (potentially quite large) would need to be routinely sent across the air interface, and the USIM would need to routinely transfer a public key to its host phone (across a smart card interface with a very limited data transfer capability). That is, whilst the PKI itself might be relatively small-scale, the changes to the air interface protocol to allow its use would require fundamental changes to the system infrastructure. It is not even clear how a phased deployment could be undertaken, and changing the entire system (including all mobile phones) at a single point in time is clearly infeasible.

It is interesting to note that the difficulty of providing robust identity privacy without asymmetric cryptography has long been known — see, for example, Mitchell ([10], section 4.1). Indeed, this point is also made by Arapinis et al. ([1], section 5.5) who make similar remarks. This suggests that modifications analogous to the proposed fix have been considered in the past, and rejected for reasons of complexity and low pay off (a point we return to below).

Moreover, deploying the required PKI requires all networks to possess two key pairs, one for encryption/decryption and one for signature generation and verification. This is because, in general, the widely accepted principle of *key separation* (see, for example, 13.5.1 of Menezes, van Oorschot and Vanstone [8]) requires that different keys are used for different purposes. However, if sufficient care is taken, sometimes the same key pair can be securely used for both encryption and signature, although this is not without risks (see, for example, Degabriele et al. [11]).

We further note that if the private decryption key of any network is ever compromised, then security is compromised. The usual solution in a PKI is to deploy a revocation system, e.g. in the form of Certificate Revocation Lists (CRLs). However, deploying CRLs on the scale necessary would appear to be very challenging in a 3G system. Indeed, the difficulties of deploying CRLs across large networks are well-established, [12, 13].

One alternative to the proposed solution would simply be to remove the error code from the error message, or, to minimise protocol modifications, to program mobile phones to always return the same error message regardless of how AKA actually fails. This is, in any case, clearly best practice for any security protocol, i.e. if an authentication procedure fails then the only information that should be provided is that the process has failed, and not how.

Finally we note that implementing the proposed fix to mitigate IMSI catching is problematic. Requiring a UE to encrypt the IMSI it sends to the network requires the phone to have a reliable copy of the network's public key. This will, in turn, require the network to send the UE a certificate — but which one? The UE will only be able to verify a certificate signed by the USIM's home network, but the visited network will not know what this is until it has seen the IMSI. That is, the UE will not be able to encrypt the IMSI for transmission to the network until the network knows the IMSI, and hence we have a classic 'chicken and egg' problem.

6 Summary and Conclusions

It would appear that the modifications proposed to address the identified privacy threats either do not work or impose a very major overhead on the network, over and above the huge cost in modifying all the network infrastructure. Very interestingly, the failures in the fixes arise despite a detailed analysis using formal techniques.

Of course, making significant changes to a protocol as widely deployed as the 3G air interface protocol is unlikely to be feasible, so the discussion here is perhaps rather moot. However, even where the fixes appear to work, in two cases significantly simpler approaches appear to have been ignored. That is, removing the error messages would mitigate the AKA error message attack (and would also conform to good practice), and it would appear that the introduction of a new key UK is unnecessary. If changes are to be made, then it is vital to try to minimise their impact on the operations of the system.

Most significantly in any discussion of whether it might be worth trying to implement 'fixed up' versions of the fixes, there exist 'passive' attacks on user identity confidentiality other than those discussed thus far. For example, a malicious party wishing to discover whether or not a particular phone is present in a cell could simply inaugurate a call to the phone or send it an SMS, simultaneously monitoring messages sent across the network. If such a procedure is repeated a few times, then it seems likely to be sufficient to reveal with high probability whether a particular phone is present, especially if the network is relatively 'quiet'. Such an attack only requires passive observation of the network, and hence would be simpler to launch than attacks requiring a false base station (which is the case for all the attacks we have discussed previously). Moreover, addressing such an attack would be almost impossible.

We can thus conclude that not only are the proposed fixes highly problematic, but providing a robust form of user identity confidentiality is essentially impossible in practice. That is, if robust identity confidentiality is not achievable, then it is very unlikely to be worth the huge cost of making changes of the type proposed. The 'pay off' in mitigating some threats but not others is small relative to the overall cost of implementing them.

Finally, the practical and security issues encountered in considering the detailed implementation of the proposed modifications suggests that the use of formal tools to try to guarantee security and privacy properties should be used with great care. In particular, any such analysis should always be accompanied by an analysis of the practical working environment for the security protocol.

References

- Arapinis, M., Mancini, L., Ritter, E., Ryan, M., Golde, N., Redon, K., Borgaonkar, R.: New privacy issues in mobile telephony: Fix and verification. In Yu, T., Danezis, G., Gligor, V.D., eds.: ACM Conference on Computer and Communications Security, CCS '12, Raleigh, NC, USA, October 16–18, 2012, ACM (2012) 205–216
- European Telecommunications Standards Institute (ETSI): ETSI TS 133 102 V11.5.1 (2013-07): Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); 3G Security; Security architecture (3GPP TS 33.102 version 11.5.1 Release 11). (2013)
- Niemi, V., Nyberg, K.: UMTS Security. John Wiley and Sons, Chichester, England (2003)
- Lee, M.F., Smart, N.P., Warinschi, B., Watson, G.J.: Anonymity guarantees of the UMTS/LTE authentication and connection protocol. Cryptology ePrint Archive: Report 2013/27 (2013)
- European Telecommunications Standards Institute (ETSI): ETSI TS 125 331 V11.6.0 (2013-07): Universal Mobile Telecommunications System (UMTS); Radio Resource Control (RRC); Protocol specification (3GPP TS 25.331 version 11.6.0 Release 11). (2013)
- European Telecommunications Standards Institute (ETSI): ETSI TS 121 133 V4.1.0 (2001-12): Universal Mobile Telecommunications System (UMTS); 3G Se-

curity; Security threats and requirements (3GPP TS 21.133 version 4.1.0 Release 4). (2001)

- Vaudenay, S.: Security flaws induced by CBC padding Applications to SSL, IPSEC, WTLS.... In Knudsen, L., ed.: Advances in Cryptology — EUROCRYPT 2002, International Conference on the Theory and Applications of Cryptographic Techniques, Amsterdam, The Netherlands, April 28 – May 2, 2002, Proceedings. Volume 2332 of Lecture Notes in Computer Science., Springer-Verlag, Berlin (2002) 534–545
- Menezes, A.J., van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1997)
- International Organization for Standardization Genève, Switzerland: ISO/IEC 19772:2009, Information technology — Security techniques — Authenticated encryption mechanisms. (2009)
- Mitchell, C.J.: The security of the GSM air interface protocol. Technical Report RHUL-MA-2001-3, Mathematics Department, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK (2001) Available at http://www.ma.rhul.ac.uk/techreports.
- Degabriele, J.P., Lehmann, A., Paterson, K.G., Smart, N.P., Strefler, M.: On the joint security of encruption and signature in EMV. In Dunkelman, O., ed.: Topics in Cryptology — CT-RSA 2012 — The Cryptographers' Track at the RSA Conference 2012, San Francisco, CA, USA, February 27–March 2, 2012. Proceedings. Volume 7178 of Lecture Notes in Computer Science., Springer-Verlag, Berlin (2012) 116– 135
- Kocher, P.C.: On certificate revocation and validation. In Hirschfeld, R., ed.: Financial Cryptography, Second International Conference, FC '98, Anguilla, British West Indies, February 23–25, 1998, Proceedings. Volume 1465 of Lecture Notes in Computer Science., Springer-Verlag, Berlin (1998) 172–177
- Myers, M.D.: Revocation: Options and challenges. In Hirschfeld, R., ed.: Financial Cryptography, Second International Conference, FC '98, Anguilla, British West Indies, February 23–25, 1998, Proceedings. Volume 1465 of Lecture Notes in Computer Science., Springer-Verlag, Berlin (1998) 165–171