Department of Computer Science
Series of Publications A
Report A-2021-3

Privacy of User Identities in Cellular Networks

Mohsin Khan

Doctoral dissertation, to be presented for public examination with the permission of the Faculty of Science of the University of Helsinki in Auditorium CK112, in Exactum building, Pietari Kalmin katu 5, on March 5, 2021, at 12 o'clock noon.

University of Helsinki
Finland

Supervisor
Valtteri Niemi, University of Helsinki, Finland

Pre-examiners
Mika Ylianttila, University of Oulu, Finland
Ravishankar Borgaonkar, University of Stavanger, Norway

Opponent
Stig Frode Mjølsnes, NTNU, Trondheim, Norway

Custos
Valtteri Niemi, University of Helsinki, Finland

Contact information

Department of Computer Science
P.O. Box 68 (Pietari Kalmin katu 5)
FI-00014 University of Helsinki
Finland

Email address: info@cs.helsinki.fi
URL: http://cs.helsinki.fi/
Telephone: +358 2941 911

ISSN 1238-8645
ISBN 978-951-51-6991-4 (paperback)
ISBN 978-951-51-6992-1 (PDF)
Helsinki 2021
Unigrafia

Privacy of User Identities in Cellular Networks

Mohsin Khan

Department of Computer Science
P.O. Box 68, FI-00014 University of Helsinki, Finland
mohsin.khan@helsinki.fi

PhD Thesis, Series of Publications A, Report A-2021-3
Helsinki, March 2021, 112 + 88 pages
ISSN 1238-8645
ISBN 978-951-51-6991-4 (paperback)
ISBN 978-951-51-6992-1 (PDF)

Abstract

This thesis looks into two privacy threats of cellular networks. For their operations, these networks have to deal with unique permanent user identities called International Mobile Subscriber Identity (IMSI). One of the privacy threats is posed by a device called IMSI catcher. An IMSI catcher can exploit various vulnerabilities. Some of these vulnerabilities are easier to exploit than others. This thesis looks into fixing the most easily exploitable vulnerability, which is in the procedure of identifying the subscriber. This vulnerability exists in all generations of cellular networks prior to 5G. The thesis discusses solutions to fix the vulnerability in several different contexts.

One of the solutions proposes a generic approach, which can be applied to any generation of cellular networks, to fix the vulnerability. The generic approach uses temporary user identities, which are called pseudonyms, instead of using the permanent identity IMSI. The thesis also discusses another solution to fix the vulnerability, specifically in the identification procedure of 5G. The solution uses Identity-Based Encryption (IBE), and it is different from the one that has been standardised in 5G. Our IBE-based solution has some additional advantages that can be useful in future works. The thesis also includes a solution to fix the vulnerability in the identification procedure in earlier generations of cellular networks. The solution fixes the vulnerability when a user of a 5G network connects to those earlier generation networks. The solution is a hybridisation of the pseudonym-based generic solution and the standardised solution in 5G.

The second of the two threats that this thesis deals with is related to the standards of a delegated authentication system, known as Authentication and Key Management for Applications (AKMA), which has been released in July 2020. The system enables application providers to authenticate their users by leveraging the authentication mechanism between the user and the user's cellular network. This thesis investigates what requirements AKMA should fulfil. The investigation puts a special focus on identifying privacy requirements. It finds two new privacy requirements, which are not yet considered in the standardisation process. The thesis also presents a privacy-preserving AKMA that can co-exist with a normal-mode AKMA.

Computing Reviews (2012) Categories and Subject Descriptors:

Networks → Network types → Mobile networks
Networks → Network properties → Network security
Security and privacy → Security services → Pseudonymity, anonymity and untraceability
Security and privacy → Cryptography → Key management

General Terms:
5G, Privacy, IMSI Catcher, Delegated Authentication, AKMA

Additional Key Words and Phrases:
Authentication, Cryptography, Pseudonym, DTDHP

Acknowledgements

Writing this thesis has been a long journey for me. It took a great deal of patience and tenacity from my part. However, it would not be possible to finish it, had I not received all kinds of support from my supervisor, colleagues, and friends.

I thank my supervisor, Professor Valtteri Niemi, for accepting me as his student. During the years of my PhD studies, he has guided me with tremendous patience. He has allowed my periods of solitude and has offered effective guidance when he felt required. He made a sweet balance. Because of the solitude, I could indulge in studying relatively less pertinent but quite exciting subjects, e.g., algebra, philosophy of science. These studies have deepened my knowledge in ways that, I believe, would be useful in my future research. I am deeply grateful to him. I am also thankful to Huawei Technologies and Business Finland for providing the funding that supported the research work during my PhD studies. The Nokia Scholarship that I received from Nokia Foundation in 2018 made my life quite comfortable. I am also grateful to Nokia Foundation.

This thesis includes four peer-reviewed articles which are all co-authored by me. However, these articles would not exist, had I not collaborated with my co-authors. I express my heartfelt gratitude to my co-authors, Philip Ginzboorg and Kimmo Järvinen, for their cooperation and kind advice. Philip read the first draft of the thesis and gave handy comments.

I also thank both the pre-examiners, Associate Professors Mika Ylianttila and Ravishankar Borgaonkar, for their precious time and effort to examine the thesis and give positive statements. I am grateful to the staff at the department of computer science, who support all PhD students, including me, by providing different administrative and coordination services. I want to especially thank Pirjo Moen and Ritva Karttunen, who have always answered my questions patiently and clearly.

Due to my early-life education at a residential school, friends have always played an unusually influential role in my life. My years as a PhD student

are no exception. I have got a handful of invaluable friends: Eamon, Emmi, Gagan, Jarno, and Saad. I have hanged out with them, travelled with them, and got all kinds of help from them whenever needed. Sometimes I have bored them with things related to my studies, which are not very relevant to them. They have patiently listened, and, at times, have even indulged in the matter. Eamon and Saad have also read the introduction of the thesis and gave useful comments. I have always been warmly welcomed by Eamon and Gagan's whole family, their wives Sharmin and Shanila, and kids, Ninad, Liana, and Ilana. I cannot thank them enough. All these people have been a tremendous support. Without them, my academic journey would be quite lonely, and the thesis would be, most probably, too tiring to finish.

I should make a special mention of my friend Jarno Alanko, who had also been a PhD student at the department and graduated in May 2020. He has helped me in my studies, in understanding many deep technical topics related to computer science. One reason I dared pursuing the PhD studies is that I knew I had Jarno to discuss technical topics when I was at a loss. I have learnt many technical things from him. Jarno and I have also played many games of chess in the coffee room at the department.

Thanks to Amelia and Laura for the refreshing chats we occasionally had. Amelia read the introduction of the thesis, which we meant to be suitable for layman reading, and gave useful comments. Amelia's kind and encouraging words, especially when I expressed my nervousness about defending the thesis, were really calming. Well, the thesis is yet to be defended, and I am still nervous.

I am also grateful to my friends, Dristy, Hasan, and Wali, for inspiring me in one way or another. Many people from the Bangladeshi community in Espoo have been amicable, creating a Bangladeshi ambience in my living in Finland. The list of their names would be too long to mention here. You know who you are. I also thank my fellow students from the computer science department, especially Gizem and Sara. They have helped me with different practical stuff on multiple occasions. I am sure, many more have helped me in one way or another in my studies, research, or social life. Even though I can not remember the names, I am grateful to all of them.

Finally, I thank my parents for their patience. I know that they have been holding their breaths for years. I believe they would be the proudest souls once I finish the PhD.

Espoo, February 2021
Mohsin Khan

List of Original Publications

This thesis is based on four peer-reviewed publications, which are listed below. The publications are added at the end of the thesis as appendices with permission from the copyright holders.

Paper I Khan, M. & Niemi, V., Jul 2017. Privacy Enhanced Fast Mutual Authentication in 5G Network Using Identity Based Encryption. In : Journal of ICT Standardization. Vol: 5, Issue: 1, p. 69-90.

Paper II Khan, M., Järvinen, K., Ginzboorg, P. & Niemi, V., 2 Dec 2017. On De-Synchronization of User Pseudonyms in Mobile Networks. Information Systems Security: 13th International Conference, ICISS 2017, Mumbai, India, December 16-20, 2017, Proceedings. Shyamasundar, R. K., Singh, V. & Vaidya, J. (eds.). Cham: Springer International Publishing AG, Vol. 10717. p. 347-366 20 p. (Lecture Notes in Computer Science; vol. 10717).

Paper III Khan, M., Ginzboorg, P., Järvinen, K. & Niemi, V., 2018. Defeating the Downgrade Attack on Identity Privacy in 5G. Security Standardisation Research : 4th International Conference, SSR 2018, Darmstadt, Germany, November 26-27, 2018, Proceedings. Cremers, C. & Lehmann, A. (eds.). Cham: Springer Nature Switzerland, p. 95-119 25 p. (Lecture Notes in Computer Science; vol. 11322).

Paper IV Khan, M., Ginzboorg, P. & Niemi, V., 2019. Privacy preserving AKMA in 5G. SSR'19: Proceedings of the 5th ACM Workshop on Security Standardisation Research Workshop. New York, NY: ACM, p. 45-56 12 p.

For Paper I, the present author has participated, with the co-author, in generating the initial idea of the presented solution. The present author reviewed the literature and expanded the initial idea into a more detailed solution. The present author has participated, with the co-author, in the detailed analysis of the pros and cons of the solution. The present author has written the first full version of the paper and has received and processed significant comments from the co-author.

For Paper II, the present author has generated the initial idea, reviewed the literature, expanded the initial idea into a more detailed attack and remedy. The present author has participated, with the co-authors, in the detailed analysis of the attack and solution. The present author has written the first full version of the paper and has received and processed significant comments from co-authors.

For Paper III, the present author has generated the initial idea, reviewed the literature, expanded the initial idea into a more detailed solution in the form of concrete algorithms. The present author has participated, with the co-authors, in the detailed analysis of the solution. The present author has written the first full version of the paper and has received and processed significant comments from co-authors.

For Paper IV, the present author has generated the initial idea and reviewed the literature. The present author has participated, with co-authors, in analysing the requirements to be fulfilled by the solution. The present author has also participated with co-authors in expanding the initial idea into a detailed solution. The present author has written the first full version of the paper and has received and processed significant comments from co-authors.

Contents

Acronyms

AApF	AKMA Application Function
AAuF	AKMA Authentication Function
AES	Advanced Encryption Standard
AK	Anonymity Key
AKA	Authentication and Key Agreement
AKMA	Authentication and Key Management for Applications
AMF	Authentication Management Field
AV	Authentication Vector
AUTN	Authentication Token
KDF	Key Derivation Function
BEST	Battery Efficient Security for Very Low Throughput Machine Type Communication Devices
CA	Certificate Authority
CEPT	European Conference of Postal and Telecommunications Administrations
CK	Confidentiality Key
CTR	Counter
C-RNTI	Cell Radio Network Temporary Identifier
DTDHP	Delayed-Target Diffie-Hellman Problem
DoS	Denial-of-Service
DDoS	Distributed Denial-of-Service
EAS	Enterprise Application Server
ECIES	Elliptic Curve Integrated Encryption Scheme
EPS	Evolved Packet System
ET	Expiry Timestamp
ETSI	European Telecommunications Standards Institute
FIP	Fair Information Practice
GBA	Generic Bootstrapping Architecture
GDPR	General Data Protection Regulation
GSM	Groupe Special Mobile
GUTI	Globally Unique Temporary UE Identity
HN	Home Network
HMAC	Hash-based Message Authentication Code
IBE	Identity-Based Encryption
IK	Integrity Key
IMEI	International Mobile Equipment Identity
IMEISV	IMEI and Software Version Number

IMSI	International Mobile Subscriber Identity
IoT	Internet of Things
ITU	International Telecommunication Union
K	Permanent Key
LI	Lawful Interception
LTE	Long-Term Evolution
LU	Location Update
MAC	Message Authentication Code
MCC	Mobile Country Code
ME	Mobile Equipment
MNC	Mobile Network Code
MSIN	Mobile Subscription Identification Number
MSISDN	Mobile Station International Subscriber Directory Number
NAF	Network Application Function
NGMN	Next Generation Mobile Network
NIST	National Institute of Standards and Technology
NTT	Nippon Telegraph and Telephone
NMT	Nordic Mobile Telephony System
PEFMA	Privacy-Enhanced Fast Mutual Authentication
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
PKG	Private Key Generator
PLMN	Public Land Mobile Network
PVT	Private Validation Token
RAND	Random Challenge
SPUID	Service-specific Permanent User Identity
SA	System Aspects
SHA	Secure Hash Algorithm
SIM	Subscriber Identification Module
SN	Serving Network
$\mathrm{SN}_{\mathrm{id}}$	Serving Network Identity
$\mathrm{SN}_{\mathrm{name}}$	Serving Network Name
SMC	Security Mode Command
SQN	Sequence Number
SUPI	Subscription Permanent Identifier
SUCI	Subscription Concealed Identifier
TLS	Transport Layer Security
TS	Technical Specification

TR	Technical Report
TSG	Technical Specification Group
TMSI	Temporary Mobile Subscriber Identity
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
USIM	Universal Subscriber Identity Module
WG	Working Group
3GPP	3rd Generation Partnership Project
5G-GUTI	Globally Unique Temporary UE Identity

Chapter 1

Introduction

Wireless communication technologies have been changing our world in unprecedented ways since the beginning of the last century. This technology has had enormous military influence, enabling fast communication with remote and mobile forces through two world wars. Wartime wireless communication systems were based on long-range radio transmitters, and these systems did not have enough capacity to provide telephony services to millions of users on a commercial scale.

Since the 1980s, telephony services using cell-based wireless networks, known as cellular networks, have become commercially successful. A cell is a relatively compact land area served by at least one radio tower. Typically, a nation-wide large land area is covered by many cells. This cell-based arrangement produces more capacity than using a single large transmitter. This is because the same radio frequency can be used simultaneously for multiple users as long as they are in different cells. Cellular networks also scale well because the network capacity to provide services to an increased number of users can be achieved by replacing one large cell by multiple smaller cells.

Some short-ranged popular wireless technologies emerged in the 1990s. For example, Bluetooth networks for exchanging data between two nearby devices relieved the trouble of using many cables in a personal space, e.g., in a car. Another example is the Wi-Fi network that enables wireless data networking between a set of computers located nearby, e.g., inside a building.

Wireless technology has had a significant influence on society, primarily because it allows communication between devices without the need for a physical wire/cable, which allows users to move around freely. However, for the same reason, wireless communication poses challenges that are of lesser concern in wire-based communication.

In a wire-based communication system, if an attacker wants to eavesdrop on the message that is being communicated, the attacker needs physical access to wires or other parts of the system. In contrast, in a wireless communication system, the attacker can eavesdrop by listening to the radio wave. Indeed, the information travels through the air, and anyone with an appropriate wireless receiver can sniff the information. The attacker's wireless receiver does not need to be expensive.

In a wire-based communication system, the attacker would be treated with suspicion if the attacker starts to play with communication cables in random places. Consequently, the attacker has a higher chance of getting caught. In contrast, using a wireless receiver is much more convenient for stealth attacks. In short-ranged wireless communication systems like Bluetooth or Wi-Fi, the attacker's wireless receiver can be within meters or tens of meters from the sender or receiver. Therefore, the attacker can carry the wireless receiver inconspicuously, e.g., in his backpack, without creating any suspicion.

Things become even easier for the attacker in the cellular networks due to their long-range radio wave. In cellular networks, eavesdropping can be done by placing the wireless receiver in an unnoticeable place chosen by the attacker, e.g., at his home in the same town as the victim. Indeed, in first-generation cellular networks, listening to other people's phone calls became a popular pastime [1, p. 4]. The infamous case of recording and publishing princess Diana's phone call is just one example [2].

An attacker's goal is not always limited to eavesdropping in a passive receiving-only manner. An attacker may actively inject carefully crafted messages in the network to cause harm to his victims, or to gain profit for himself. This kind of attacker is called an active attacker. Such an attacker may, for example, try to put the cost of usage incurred by himself on the bill of others. Active attacks can be mounted in a similar stealthy manner as mentioned above using a wireless transceiver, i.e., transmitter and receiver coupled together. In practice, this was possible in the commercially-used first-generation cellular networks [3] which became popular in the 1980s.

A cellular network needs to know the identity of the subscriber before providing services to a user so that the network can bill the subscriber. The network requests the user's device to send the identity of the subscriber. In response to the request, the user's device sends the subscriber's identity to the network. An active attacker sees opportunities in this arrangement, for example, to know whether a victim is at home or not. Towards that goal, the attacker sets up a wireless transceiver near the victim's home and sends identity requests to all devices within the attacker's range. If a response from the victim is received, the attacker can infer that the victim is still at his home. Otherwise, the attacker can conclude that the victim has either left home or switched off the device. Variants of this kind of location-tracking attacks exist in all generations of cellular networks. This kind of active attack can help in a more serious offence, for example, burglars to break in and steal from the victim's home.

The above discussion may sound bleak and give the impression that wireless communication systems are doomed to be vulnerable to attackers. Fortunately, that is not the case. Introduction of the new digital radio signal (replacing the old analogue signal) has enabled the use of cryptography (a technique used for hiding long-distance military communication since ancient time) to secure wireless communication. In the 1990s, the second-generation commercial cellular networks used digital radio signal and leveraged modern cryptography to fix many vulnerabilities, i.e., opportunities that were exploitable by attackers. Intricate use of cryptographic mechanisms was embedded in users' mobile phones in a transparent manner. A user would only have to turn on his mobile phone, and the phone would be ready to use with all its security properties without the user noticing any interruption or delay.

However, cryptography is not the silver bullet for fixing all the exploitable vulnerabilities of a commercial communication system. Cryptographic techniques require additional computational and infrastructural cost, which may outweigh the benefit of using the techniques. The design process of the security mechanisms of a commercial communication system is guided by some pragmatic principles that may not suggest fixing every vulnerability exploitable by some attackers.

One of these principles says that the benefit-cost ratio of fixing a vulnerability has to be acceptable from a business point of view. The cost of fixing should be smaller than the anticipated loss, in case the vulnerability was not fixed. Another principle says that a system is as strong as its weakest link. This principle stems from the phenomenon that when an attacker sees multiple exploitable vulnerabilities to achieve his attacking goal, the attacker chooses to exploit the vulnerability that takes the least effort. Therefore, the design process should try to fix a vulnerability that is easier to exploit by the attacker before fixing vulnerabilities that are more difficult to exploit.

In the standards for GSM networks, second-generation cellular networks, which have become popular since the 1990s, many vulnerabilities have been fixed. The effort kept attackers from succeeding in their malicious actions. For example, the fraud where attackers try to put their usage on the bill of others were defeated. Also, the passive attackers were not able to eavesdrop on the communicated messages anymore. Both of the fixes were done using cryptographic techniques. However, because of the pragmatic principles explained above, not all the vulnerabilities were fixed. A fake network could still attack a victim.

In GSM, an active attacker could set up a fake network and introduce itself as a network that a victim could trust, i.e., the attacker would try to impersonate a legitimate network to the victim. By doing so, the attacker could do different bad things, for example, eavesdrop on the communication done by the victim or modify the content of a text message sent by the victim or infer the presence/absence of the victim at a location, e.g., the victim's home. Users were left vulnerable to fake networks because the cost of establishing a fake network was thought to be very high. It was estimated that an attacker using a fake network against a victim would not have enough incentive for mounting such attacks. Therefore, the attack was considered highly unlikely, and in consequence, no protection was developed.

Since the publication of GSM standards, the security of the cellular network standards has been periodically re-evaluated and tightened by fixing more and more vulnerabilities. For example, in the standards of 3G networks, which were released in the late 1990s, the active attackers could not eavesdrop or modify messages by using a fake network anymore. This was because before sending or receiving any message, a user's device, by using cryptography, could identify whether a network was fake or not. If fake, the user's device would just stop communicating with the network.

However, the active attackers could still use a fake network to infer a victim's presence at/absence from a place and breach the location privacy. This was because a user's device could identify a fake/legitimate network only after the user's device had sent the identity of the subscriber to the fake/legitimate network. Once the fake network receives the subscriber's identity, it already succeeds in inferring that the user is present at the nearby location. If the fake network does not receive the subscriber's identity, then it can infer that the user is not nearby. Adequate protection against this kind of attack was not developed because the cost involved in maintaining the protection mechanisms outweighed the anticipated gain. Most of the cellular networks prior to the latest 5G network were vulnerable to this attack.

The attack on users' location privacy, as mentioned above, is possible because the users' device sends a long-term identity, i.e., subscriber identity, to the network. Indeed, if the user sent a new identity each time it identified itself, then the attacker would not know who is who. Please note that a user of a cellular network may send different kinds of identities (some are more long-lived than others) over the network. Different network standards and protocols specify these identities. For example, some of the identities (e.g., user names, cookies) are used at the application layer; some identities (e.g., IP addresses) are used when routing the messages to correct destinations on the Internet. A Medium Access Control (MAC) address is used by the user's device in accessing Wi-Fi networks. Also, a unique Bluetooth device address is used in Bluetooth communication. Any of these identities, if captured cleverly, can be used to infer the presence/absence of a user at a location.

This thesis is mostly about the privacy of the user identities in cellular networks. Ensuring the privacy of all identities of a user in a cellular network would require identifying and fixing all the vulnerabilities in all the relevant standards that define those identities. Moreover, the privacy of these identities can also be breached if an attacker can get unauthorised access to the user's device or the network's premises. Dishonest insider personnel from the network's operational team may also cause the privacy breach of a user.

Therefore, ensuring the privacy of all identities of a user is a very complex task with many facets. A pragmatic approach to handle the whole issue is to divide it into many smaller parts and resolve each part separately. Such a divide and conquer approach makes sense because the relevant vulnerabilities are arguably too scattered to take on in a single piece of work. Designing fixes to all the vulnerabilities in one go demands a wide range of technical skills and involves different organisation bodies. Therefore, dividing the whole problem into parts and solving each part separately, possibly in parallel, is pragmatic.

Fixing the vulnerabilities in the cellular network protocols should be prioritised because an active attacker can mount attacks against these protocols in a less conspicuous manner (the wireless transceiver can be placed, e.g., at the attacker's home) than against other protocols. We propose fixes to two vulnerabilities related to the cellular network protocols in this thesis.

The first vulnerability has its root in the user identification protocols (the way the user sends its identity to the network) in GSM, 3G, and LTE. All these generations[1] of cellular networks had left the weakness unresolved due to low benefit-cost ratio [4], i.e., due to low gain in benefit in comparison with the required cost to mitigate the weakness. The second vulnerability is relatively

[1] Please note that in this thesis, we limit our discussion to the most dominant cellular network standards from each generation, i.e., GSM, 3G, LTE, and 5G.

new and relevant only in the second phase of 5G, which has been released in July 2020. This vulnerability appears as a vulnerability in this thesis but not in the design process of 5G. This is because we put less trust in the network than the 5G design process has. One of the reasons to put less trust in the network is the rising trend of insider attacks, i.e., some employees within a company attack the customer of the company [5–7]. In the following, we give a high-level overview of these two vulnerabilities. Rigorous problem definitions are presented in Chapter 3.

The first vulnerability is related to an identity privacy threat posed by International Mobile Subscriber Identity (IMSI) catchers. An IMSI catcher is a rogue device with a radio interface. It impersonates a legitimate network. The IMSI catcher's primary goal is to infer the presence/absence of users at a location of interest. It is called "IMSI catcher" because the most powerful attack it mounts, to achieve this primary goal, reveals (or catches) users' IMSI. This is the most powerful attack because it can attack many users in parallel. The vulnerability that the IMSI catcher exploits in this attack is in the user's identification procedures, and it is the first vulnerability of interest in this thesis.

The IMSI catcher also uses vulnerabilities in other procedures to achieve similar goals. These other procedures include those that the user's device and the network use, to verify each other's identity so that they are not fooled by impostors. Vulnerabilities also exist in the procedures that the network uses to search the user's device for which incoming messages have arrived. Exploiting these other vulnerabilities leads to attacks that are less powerful than the attack that exploits the vulnerability in the identification procedure. This is because each victim has to be targeted separately.

The techniques involved in the defence against IMSI catchers can be put into two categories: detecting and defeating. The detecting techniques help the users to stop communicating with the IMSI catcher or help the police to pursue the IMSI catcher and stop it from attacking. Detecting IMSI catchers may prevent various attacks mounted against a user. However, in many cases, the detection alone does not prevent them from attacking the identity privacy of a user. This is because, by the time a user can detect an IMSI catcher, the IMSI catcher may already have tricked the user into revealing the IMSI. Also, the detectors are not completely accurate; they may produce false positive and false negative detection [8–10]. Defeating IMSI catchers refers to the techniques that propose modifications in vulnerable procedures so that the IMSI catchers cannot exploit them anymore. Therefore, techniques that defeat IMSI catchers are more effective than techniques that use detection as a pre-requisite.

The most easily exploitable vulnerability is in the identification procedure, which we try to fix in this thesis. We try to fix the vulnerability by concealing the IMSI, e.g., using cryptographic measures so that the concealed IMSI is incomprehensible to the IMSI catcher. Thus, our techniques fall under the category of defeating IMSI catchers. The threat of IMSI catchers persists in all the generations of the mobile networks[2] prior to 5G.

We propose solutions to fix the vulnerability in the identification procedure in different contexts, namely in 3G, LTE, or 5G. Our solution to fix the vulnerability in LTE and 3G is a generic solution based on pseudonyms that have the format of IMSI. The solution may be extended to fix the vulnerability in GSM. However, we do not dive into defeating IMSI catchers in GSM because an attacker can mount even more severe attacks on a GSM user by taking a man-in-the-middle position[11] than by simply catching IMSI.[3]

The 5G network has fixed the vulnerability in the identification procedure that IMSI catchers exploit. Even though the vulnerability is fixed, a solution to fix the same vulnerability in 5G is included in this thesis due to the following reasons. We proposed the solution in 2017 when finding a fix to the vulnerability was still an open question. Our solution to fix the vulnerability is different from the one that is standardised. Our solution uses identity-based encryption that offers other advantages (along with disadvantages) on top of fixing the vulnerability. The additional pros could not outweigh the cons in the first phase of 5G. Nevertheless, in future, the additional pros could be so useful that our solution could be re-considered or be useful in works for new releases of standards.

The second vulnerability of interest in this thesis is related to a delegated authentication system. An authentication system aims at guaranteeing that the identity of an entity attempting to access protected resources is genuine. For example, when a traveller arrives at an airport, the immigration control works as an authentication system towards ensuring that only authorised people can enter the country. A passport with a photo of the traveller works as the traveller's credential. The immigration control uses the passport to verify the traveller's identity. A successful verification of a traveller's identity does not guarantee that the traveller is allowed to enter the country. Once the identity of the traveller is confirmed, the immigration control can check the policies to find out whether a person with the confirmed identity is allowed (or authorised) to enter the country.

[2] Cellular networks are frequently referred to as mobile networks. In this thesis, we use the phrases "mobile network" and "cellular network" interchangeably.

[3] The major weakness in 2G is that the user does not authenticate the network, i.e., the user has no way to ensure that it is talking with a legitimate network and not with an impostor.

Similarly, when a user tries to log into a digital service, an authentication system of the digital service verifies the identity of the user based on some user credentials. For example, when a user tries to log into Facebook, the user has to provide the correct password. Another example is the way the cellular networks verify a user's identity. When a cellular network user tries to connect to the network, the identity of the subscriber is authenticated, by the network. The identity of the subscriber is embedded in the so-called Subscriber Identification Module (SIM) card.[4] Without a correct SIM card, the authentication will fail. If the authentication succeeds, it does not mean that the user is authorised to use the service. For example, a pre-paid user may not have enough balance in his/her phone to make a call. However, in this thesis, we do not discuss any authorisation mechanisms.

In a delegated authentication system, the digital service itself does not verify the user's identity. Instead, the digital service takes help from a third-party authentication system to verify the user's identity. The third-party system informs the digital service if the user's identity is authentic. This arrangement of delegating the task of verifying the user's identity is very convenient for the users and the digital service providers, and therefore, a commonplace in today's cyberspace. Users can log into various digital services, for example, in Finland, the tax office's website, using the authentication system of the user's bank. Also, users utilise Facebook or Google's authentication systems to log into different digital services, e.g., a social media of book lovers.

More specifically, the second vulnerability of interest of this thesis is about user privacy in a new delegated authentication system in 5G, known as Authentication and Key Management for Applications (AKMA). The system is based on a user's mobile-phone credentials, namely, the SIM card, where the cellular network takes the role of verifying the identity of the user. The system enables a digital service provider, which is not part of the cellular network, to verify the identity of the digital service's users with the help of the authentication system of the cellular network. The system relieves the service provider from maintaining authentication-related data and the user from memorising many passwords for different digital service providers. The standardisation of AKMA had been on-going for around two years before it was released in July 2020.

Our results include the requirement analysis of AKMA with a particular focus on privacy requirements. We propose to put less trust in the cellular network.

[4] The name "SIM card" became popular from the GSM technology. In newer generation mobile networks, the corresponding element is called UICC. In less technical texts it is still called SIM card.

One of the reasons to put less trust is the rising trend of insider attacks [5–7]. Consequently, we find two new user privacy requirements. The user may use various kinds of services, such as gambling or dating services. We argue that the authentication server, i.e., the cellular network, does not need to know which digital services the user connects to. Also, if two digital service providers collude, they should not be able to link two of their users. That is, if a person uses a dating service with the pseudonym *romeo* and a gambling service with the pseudonym *smeagol* then the dating service and the gambling service, even when colluding, should not be able to figure out that *romeo* and *smeagol* are indeed the same person. In this thesis, we propose a privacy-preserving AKMA, which could be offered as a premium service to privacy-sensitive users.

The thesis includes four academic papers published in an international journal and three conferences. The papers are appended at the end of the thesis. We briefly introduce them in the following.

Paper I This paper is an extension of one of our conference papers [12] into a journal paper. In this paper, we proposed a solution to mitigate IMSI catchers in 5G using Identity-Based Encryption (IBE) to fix the vulnerability in the identification procedure. The solution also has the advantage of faster mutual authentication compared to the existing cellular authentication protocols. However, the standards for the first phase of 5G came out after we published our paper, and the vulnerability in the identification procedure was fixed using public-key cryptography. Nevertheless, the advantages of our solution may become more desirable in the future.

Paper II This paper presents a generic solution to fix the vulnerability in the identification procedure in LTE and 3G. The solution introduces pseudonyms that are exchanged between the user and the network piggybacking on the messages of the existing authentication protocols. Pseudonym-based solutions have the advantage that they require relatively less patching effort during implementation in the already deployed LTE or 3G networks. Existing pseudonym-based solutions had a common vulnerability. We showed that an attacker could exploit the vulnerability to mount a "distributed denial of service" attack. Our solution proposed a fix to the vulnerability.

Paper III This paper presents a pseudonym-based solution to fix the vulnerability in the identification procedure in the 3G and LTE networks for 5G users. A

5G user, when 5G network coverage is not available in a place, may get an offer to connect to a legacy network, e.g., 3G or LTE, which can be seen as a downgraded offer. The 5G user would want to accept the downgraded offer because the user would have to remain disconnected otherwise. Because of this compliance to downgrading, IMSI catchers remain a threat against 5G users, even if the 5G network itself is immune to the threat. The solution leverages 5G AKA to get rid of some downsides of the pseudonym-based solution presented in Paper II.

Paper IV The paper analyses the requirements that AKMA should fulfil. The requirement analysis includes a comparison of AKMA requirements with the requirements of similar existing systems. Along with the existing long-term identity privacy requirement, the paper also proposes two new privacy requirements. The paper presents a solution that fulfils these privacy requirements and most other requirements too. The solution includes an outline of how the privacy-preserving AKMA can be combined with a normal mode AKMA so that they can co-exist. This paper uses the assumption that DTDHP (a mathematical problem) is hard to solve [13].

Chapter 2

Big Picture

In this chapter, we present a bird's-eye view of this thesis in the realm of relevant technologies and concepts. Most of the contributions to the thesis lie at the intersection of cellular technology, identity privacy, and cryptography. Identity privacy is a sub-concept of privacy; we discuss their differences in Section 2.1. A small portion of this thesis lies at the intersection of cellular technology, privacy (not identity privacy), and cryptography. Another small portion lies at the intersection of cellular technology and cryptography (see Figure 2.1).

In the following, we discuss aspects of these concepts – (identity) privacy, cellular technology, cryptography – that are relevant for explaining the big picture sketched above. In this chapter, we do not present details of the specific problems that we solve; they are presented in Chapter 3.

2.1 Privacy and Identity Privacy

At the turn of every technological leap forward, privacy threats become more invasive and encompass broader contexts of human society. At the beginning of the second industrial revolution, which is at the end of 19th century, invasions of privacy were perceived due to the emergence of instantaneous photography and widespread circulation of newspapers.[1] During the third industrial revolution, which is in the second half of the 20th century, the advent of contraceptives and computers sparked a greater concern for the invasion of privacy.[2]

[1] The concern is articulated in the influential law review article "The Right to Privacy" written by Samuel Warren and Louis Brandeis [14].

[2] In 1965, the Griswold decision [15] against contraception prohibition on the ground of the right to "marital privacy" gave us a glimpse of how the understanding of privacy had changed

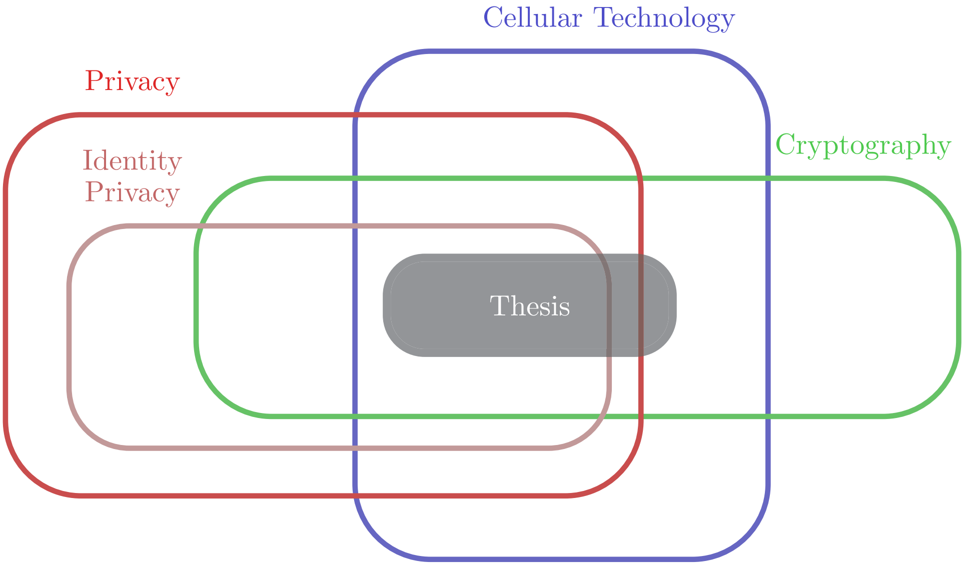

Figure 2.1: Position of the thesis.

The concern mentioned above has turned into public outcries in the Internet-shaped modern world (the fourth industrial revolution) [17, 18]. Today's privacy threats are invasive to the point of threatening one of the fundamental pillars of the modern world – democracy.

Though the debate of privacy is old and pervading in our society, an overarching conceptualisation of privacy is elusive. In 2008, Daniel J. Solove, in his book "Understanding Privacy" [19] has claimed that privacy as a concept is in disarray. In favour of his claim, Solove has quoted philosophers, legal theorists, and jurists who have frequently lamented the great difficulty in reaching a satisfying conception of privacy. For example, Solove quotes legal theorist Robert Post, "Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings that I sometimes despair whether it can be usefully addressed at all".

In 1976, Paul Sieghart, in his book *Privacy and Computers* [20] has discussed the difficulties that lie around defining the concept of privacy. Sieghart writes that discussions of privacy revolve around three facets: (i) the condition of privacy

with the advent of new technologies. In 1967, Alan F. Westin, in his book "Privacy and Freedom" [16], had discussed how the low-cost electronic surveillance devices were threatening privacy of individuals.

(ii) the desire for privacy (iii) the right to privacy. According to Sieghart, the condition of privacy has more or less a common denominator, i.e., seclusion. However, he explains that the difficulty comes from the subjective nature of the desire for privacy, the conflict between privacy and other socio-economic values, and the necessity of defining a firm boundary of privacy as a legal right.

Many theories (e.g., the right to be let alone, limited access to the self, secrecy, intimacy) have been proposed by the theorists to capture the essence of privacy. Solove, in his book, has argued that each of the theories is either too broad or too narrow. He claims, the reason the theories fail lies in the method used by theorists to conceptualise privacy. Solove proposes a pragmatic method that suggests treating privacy more contextually.

In this thesis, we do not try to obtain a universal definition of privacy. We turn our attention to the 3GPP community for a definition of privacy in the context of cellular networks. This approach is consistent with Solove's pragmatic approach. We searched and found just one 3GPP document with an explicit definition of privacy. This document is the 3GPP Technical Report (TR) 33.899 [21].[3]

2.1.1 A Working Definition of Privacy in Cellular Networks

In 3GPP TR 33.899 [21], privacy is defined as the right to the protection of Personally Identifiable Information (PII). In the document, PII is defined as any information that (a) can be used to identify a subscription to whom such information relates, or (b) is or might be directly or indirectly linked to a subscription. According to 3GPP TR 21.905 [22], a subscription describes the commercial relationship between the subscriber and the service provider.

It is worth discussing the difference between a subscriber and a user. A subscriber is an entity (associated with one or more users) that is engaged in a subscription with a service provider. The subscriber is allowed to subscribe and unsubscribe services, to register a user or a list of users authorised to enjoy these

[3] 3GPP TR 33.899 [21] is a collection of proposed security requirements and solutions that were collected during the study of the 5G architecture. The study phase ended in August 2017, and the document was withdrawn in September 2017. The reason for withdrawal is perhaps due to the big effort required to evaluate every proposed requirement and solution rigorously and make the whole document coherent. Nevertheless, the content of 3GPP TR 33.899 [21] reflected the state of the work when the study ended. The privacy definition presented in the TR appears to be reasonable, and we use it as a working definition of privacy of users in the cellular network.

services, and also to set the limits relative to the use that associated users make of these services [22].

So, in summary, a subscriber is an entity that pays, and a user is an entity that uses the services. These two entities, in most cases, are the same but may be different. The network authenticates the subscriber's identity (not the user's identity) which is embedded in the user's device. Therefore, in principle, the network remains oblivious about the actual user, i.e., the network cannot tell who (the subscriber or someone else) is using the subscribed services. This arrangement, i.e., not authenticating the user's identity, keeps the authentication infrastructure relatively simple and provides relatively easy usability experiences. For example, a mother may become the subscriber so that her child can use the services.

In the legal context of the USA, sometimes, the term PII refers to a smaller set of information than that in 3GPP vocabulary [23]; information that is not considered to be PII in North America may be considered to be PII in 3GPP. In 2007, in the USA, the Executive Office of President - Office of Management and Budget (OMB), defined PII in a memorandum for safeguarding against and responding to the breach of PII. Later, the National Institute of Standards and Technology (NIST) used the definition in one of their recommendations for protecting PII [24]. The OMB definition of PII is in the following [25]:

> *Information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc.*

In the OMB definition, the term PII refers to information that is limited to human. In contrast, the 3GPP definition refers to information related to subscriptions; both an individual and an organisation can have subscriptions. The OMB definition talks about information that would reveal a long-term (social) identity of a person. In contrast, it appears, the scope of the 3GPP definition extends to any information that is linkable even if it does not reveal a long-term (social) identity of the related person.

In the light of the above discussion, it appears that the 3GPP definition of PII is more aligned with the European understanding of personal information as defined as "personal data" in the General Data Protection Regulation (GDPR). The definition of personal data in GDPR includes data about both identified and identifiable natural persons. The GDPR definition is the following [26]:

> *Any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.*

The European understanding of personal information includes a wide range of data related to an identifiable person. However, the 3GPP definition differs from the European definition too – 3GPP refers to a subscription, but GDPR refers to a natural person only. Perhaps the difference stems from 3GPP's focus on the subscriber's identity instead of the user's identity.

In 3GPP, the definition of PII includes information related to both natural persons and organisations as long as they are subscribers. Maybe 3GPP does not want to differentiate between subscribers and natural persons only for the purpose of the privacy definition. This is indeed not a problem from the privacy point of view because the scope of personal information in 3GPP's definition is wider than in the GDPR definition. However, it is usually more important to protect the privacy of a natural person than an organisation. For example, a municipality may subscribe to services which are used by machines like traffic signal posts; here, neither the municipality nor the signal post's privacy is much of a concern.

It is known that an adversary equipped with relevant but crucial partial information may be able to link non-personal-looking data to an individual [27]. Thus, identifying the adversaries and their capabilities are central towards recognising privacy threats. We follow the seven principles of privacy by design [28] towards identifying privacy threats. In conjunction with the 3GPP definition of privacy and the principles of privacy by design, we identify privacy problems in cellular networks. We put a special focus on the "collection limitation" and "data minimisation" FIPs (Fair Information Practices) that inform the "privacy as the default" principle of privacy by design [28]. The "collection limitation" and "data minimisation" FIPs as defined by Cavoukian [28] are presented below.

Collection Limitation: The collection of personal information must be fair, lawful and limited to that which is necessary for the specified purposes.

Data Minimisation: The collection of personally identifiable information should be kept to a strict minimum. The design of programs, information and communications technologies, and systems should begin with non-identifiable interactions and transactions, as the default. Wherever possible, identifiability, observability, and linkability of personal information should be minimised.

Here is a summary of both FIPs. An entity in an electronic communication system should not know anything about a user that the entity does not need, to provide the expected functionality.

2.1.2 Identity Privacy

The "identity" of an individual is a broad concept that refers to values of different attributes – e.g., name, address, ethnicity, religion – of the individual. The term "identity privacy" is used in discussions of user privacy in the context of electronic communication systems [29–36]; and the term "identity" in "identity privacy" refers to the digital identity, i.e., the identifier that represents the external agent (usually a natural person) that participates in the electronic communication system.

Some of the digital identities are used to recognise an individual for a long duration of time; for example, email addresses, or Facebook ID. Because of their permanent nature, by learning these long-term identities, an adversary may be able to track the footprint of the natural person inside the digital system. However, not all identities are of a long-term nature. Some identities are used only for a short duration of time. For example, when a person arrives at a service centre, the individual is assigned a queue number or a case number. The assigned number works as a temporary identity for the individual during his time waiting for the service. In a digital system, a person may be assigned many such temporary identities.

In several publications [29–36], the authors define (or implicitly assume) "identity privacy", in the context of a communication system, as concealing the digital identities of a sender (or receiver) of a message from anyone else but the intended receiver (or sender) of the message. Therefore, the concept of identity privacy is equivalent to the concept of identity confidentiality.

Usually, the temporary identities are not meant to be linkable with a long-term identity of the person, hence they do not threaten the privacy of the person. However, if not designed carefully, temporary identities may become linkable and be used to identify the associated natural person participating in the system [37].

Please note that the privacy of an individual may be breached due to the undesired exposure of other information (apart from digital identities) about the participating natural person. We discuss such privacy breaches in the following.

2.1.3 Privacy Issues Due to Exposure of Non-identity Data

The privacy of an individual can be breached without an attacker knowing the true identity of the individual, i.e., without knowing the specific natural person involved. For example, an advertising agency can track a user, e.g., by using third-party cookies or browser fingerprinting, without learning the true identity of the user, and be able to send targeted-marketing content. Though the true identity may remain unknown to the agency, being a victim of such a marketing target is considered to be a breach of privacy. Reasons behind such perception include the risk of receiving a tailored price of the advertised product, and the perceived intrusiveness of the advertising [38].

The privacy of an individual can also be breached by compiling an anonymised dataset with a de-anonymised dataset. For example, an adversary can breach an individual's financial privacy by analysing anonymised credit-card data, which include date and place of spendings for each card, in conjunction with the knowledge of the individual's travel destinations extracted from his twitter posts [27]. The financial privacy breach, in this case, is possible due to the attacker knowing both the victim's travel destinations and the credit-card data that includes, among other things, the victim's data.

2.2 Cellular Networks

In this section, we give a short description of the evolution of cellular networks. We briefly explain the working procedures of the 3GPP community and present an overview of the architecture of cellular networks. Finally, we discuss user identities in cellular networks.

2.2.1 A Brief History of Cellular Networks

In the late 1970s and early 1980s, all around the world, many automated, and large-scale mobile-network standards emerged [39, p. 250-251]. For example, Bell Labs in the USA developed the Advanced Mobile Phone System (AMPS) in 1979. Japan's Nippon Telegraph and Telephone (NTT) company developed mobile-network standards in 1979. The Nordic countries developed the Nordic

Mobile Telephony (NMT) system in 1981. These were the first generation, 1G networks, and all of them were based on the analogue radio signal. Throughout the 1980s, operators in many countries of the world deployed these networks. Some of these standards became commercially successful [40, p. 24-30].

Following the success of 1G networks, the European Conference of Postal and Telecommunications Administrations (CEPT) conceived the need for a common mobile-network standard. The European Commission showed interest in CEPT's desire for the common standard [41]. Within CEPT, a new group called Groupe Special Mobile (GSM) [42] was created with the specific task to create the common standard [43, Section 1.3]. In 1987, the GSM group delivered the first standard of the GSM network [44]. In 1988, CEPT transferred the GSM group (and all the standards) to the European Telecommunications Standards Institute (ETSI) [42], which was also created by CEPT in the same year. Finnish operator OY Radiolinja was the first to commercially deploy the GSM network [45, p. 529]. Since then, GSM networks have been deployed in many countries and became the most successful in the history of telecommunication. However, other second-generation network standards also exist; e.g., CDMA-One and D-AMPS from the USA, and PDS from Japan [40].

The data rate provided by the GSM networks was not fast enough for many multimedia mobile applications [46, p. 440]. Also, the idea to ensure fully global roaming (users to use the mobile system services anywhere in the world) was pushing for another technological leap forward. Standards bodies from Europe, Asia, and North America established a collaborative project known as the 3rd Generation Partnership Project (3GPP) in 1998, under the scope of the International Telecommunication Union (ITU). This new collaborative project developed the first truly global cellular technologies (3GPP Release-1999) based on the GSM standards [47]. The new third generation, 3G technology, is also known as Universal Mobile Telecommunications System (UMTS). Other third-generation standards also emerged; e.g., CDMA2000 in the USA [48]. However, in the rest of this thesis, we use the terms UMTS and 3G interchangeably.

Since its inception, 3GPP has managed the evolution and maintenance of the GSM and 3G standards. The first version of specifications for the 3GPP-defined fourth generation, 4G network (3GPP Release-8) was released in 2008. This 4G network is also known as Long-Term Evolution (LTE) or Evolved Packet System (EPS). In the rest of this thesis, we use the terms LTE and EPS interchangeably. The first version of specifications for the 3GPP-defined fifth generation, 5G network (3GPP Release-15) was released in 2018. This 5G network is also known as the Next Generation Mobile Network (NGMN).

One major improvement from one generation to the next has always been a significantly higher data rate. The maximum downlink data rate in GSM Release-96 was 14.4 kbit/s [49], which improved to 42 Mbit/s in 3G [50]. In LTE, the maximum downlink data rate reached 300 Mbit/s [51], which is improved to more than 1 Gbps in 5G [52].

2.2.2 Working Procedures of 3GPP

In 3GPP, a specification work follows a three-stage model. Stage 1 specifications define the requirements of new services. Stage 2 specifications contain architectural descriptions, e.g., what the functional entities are and what information flows between them, that meet the requirements. Finally, Stage 3 specifications include bit-level descriptions of protocols that realise the architecture.

The entire specification work is divided between different Technical Specification Groups (TSG). Each TSG consists of multiple working groups. Typically different working groups carry out different stages for the same features. For example, one working group of the TSG Service and System Aspects (TSG-SA), which is known as SA WG1, concentrates purely on requirements, i.e., Stage 1 specifications. Another working group of the same TSG, which is known as SA WG2, creates system architecture, i.e., Stage 2 specifications. An example of a working group that specifies Stage 3 specifications would be Working Group 1 of the TSG Core Network and Terminals (CT), which is known as CT WG1. The security features are specified in a working group known as SA WG3 (or SA3 for short) under TSG-SA. The working group SA3 mostly produces Stage 2 specifications. However, it also produces some Stage 3 specifications, such as the bit level descriptions of cryptographic algorithms.

We mentioned in Subsection 2.2.1 that 3GPP is a collaborative project. Specification work on a feature can go forward only if some of the participant members are interested in investing necessary resources. As a general rule, the relevant TSG has to approve a work item before a working group can start working on the specification of a feature. The working group often opens a study document known as a Technical Report (TR). The participants contribute to the TR document towards analysing the feasibility of the feature and selecting an optimal solution. The insights gained from a TR document usually used in writing the Technical Specification (TS) document.

The TR documents are informative, for example, they inform about the relevant state-of-the-art prior to writing the related technical specifications, and therefore, also inform the rationale of the technical specifications. Therefore, TR documents assist an interested person with regard to a particular subject area. An implementer can completely ignore the TR documents and still build equipment compliant with 3GPP specifications. Some TR documents may have a ”withdrawn” status because of, e.g., incomplete studies.

The TS documents have normative status, i.e., the specifications mentioned in these documents are necessary for the application of the standard in which they are mentioned. If an implementer ignores some specifications mentioned in a TS document, then the implementer may end up building equipment that is incompatible with equipment built by other implementers. Forsberg et al. [47, p. 22] present a discussion of the working procedure in 3GPP.

2.2.3 A Simplified View of Cellular Network Architecture

Cellular networks have complex architectures. Moreover, across the generations, the architecture has changed many times. To be able to discuss the problems in cellular network security and privacy that are relevant in this thesis, it is convenient and possible to keep most of the intricate parts of the architecture out of focus. By doing so, we can get rid of most of the jargon, obtain a birds-eye view, and gain more leverage to concentrate on the real essence of the problems. However, as a consequence of such a simplification, this architecture would not shed much light on other aspects of mobile networks, e.g., how the user makes a voice call to another user or uses Internet services.

A mobile network consists of three parts (see Figure 2.2): (i) User Equipment (UE) – a user usually carries this device as the user moves, (ii) Serving Network (SN) – that the UE connects to, and (iii) Home Network (HN) – the user has a subscription with this network. Both the SN and HN are connected to the public Internet. A UE comprises a Mobile Equipment (ME) with a radio interface and a tamper-resistant smart card known as UICC.[4] The ME has a slot to insert the UICC inside it. The UICC hosts a Universal Subscriber Identification Module (USIM).[5]

A UE runs many protocols with the SN to provide services to its user. For example, once turned on, the UE scans and finds a suitable radio interface to camp on; then the UE runs an authentication protocol to authenticate itself to

[4] According to ETSI TR 102 216 [53], a UICC is a smart card that conforms to the specifications written and maintained by the ETSI Smart Card Platform project. The TR document also states that UICC is neither an abbreviation nor an acronym.

[5] In GSM network, the card is limited to the functions of a Subscriber Identification Module (SIM). Therefore, the card itself is known as the SIM card. However, in the newer generations, the UICC may run other functionalities than a USIM.

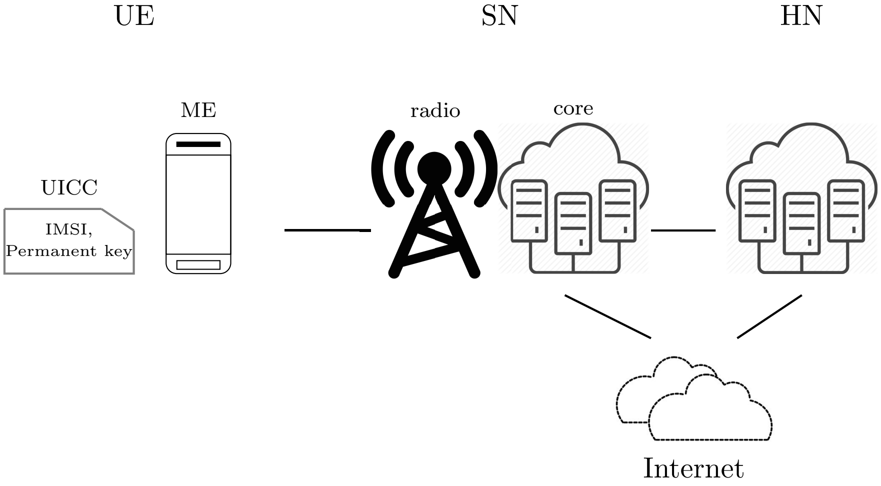

Figure 2.2: Cellular Network.

the SN. Another example is the paging procedure – when there is incoming traffic to a UE, the SN broadcasts paging messages to a bunch of UEs in order to find the right receiver. Most of these protocols are transparent to the user, i.e., they are silently run between the UE and the SN without the user noticing anything.

Furthermore, both the SN and HN themselves have two parts: (a) a radio network that covers a large geographical area and (b) a core network that the radio network connects to, is centred in a small area like a data centre, and can itself connect to another core network or the Internet. When the user is roaming, the core network of the SN connects to the core network of the HN. However, when the user is not roaming, the SN and the HN are the same networks; hence, we do not show the radio network part of the HN in the picture.

The UICC contains a permanent identity of the user. This permanent identity is known as IMSI in GSM, 3G, and LTE, and as SUPI in 5G. The UICC includes a symmetric cryptographic key that is also known only by the HN and no one else. In GSM, 3G, LTE, and 5G specifications, this key is referred to as "Individual Subscriber Authentication Key", "authentication key", "permanent key" and "long-term key" respectively. In this thesis, we call it the permanent key. The key is used only in the authentication server and the USIM. Both of these entities are configured by the same authority, i.e., the HN. Therefore, the length of the permanent key does not need to be standardised. However, 3GPP specifications provide example use of these keys which require them to be only 128 bits long.

The 3GPP specifications also provide example use of these keys which require them to be either 128 or 256 bits long [54, Clause 6.3.7].

Using an API of the USIM, an ME can read the IMSI. The UICC is tamper-resistant; therefore, reading the content of the UICC using other means than using a USIM API is not feasible. However, there is no API to read the permanent key. Nevertheless, the USIM has APIs that the ME can use to compute different functions based on the permanent key. The security of most of the cryptographic techniques used in mobile networks (in all the generations) depends on the randomness and secrecy of this pre-shared permanent key.

In 5G, the HN has a public-private key pair which is used during the identification phase of a user. The public key of the HN is provisioned in a 5G-enabled UICC.

2.2.4 User Identities in Cellular Networks

A mobile user has various identities. Some identities are more volatile than others. Some identities are assigned and maintained by the SN, and some by the HN. Some identities are permanently embedded in the ME by the manufacturer; hence, they can be associated with different users if multiple users reuse the ME. This thesis does not focus on the privacy of all of these identities. In this section, we shed light on some of these identities that are relevant in this thesis.

International Mobile Subscriber Identity (IMSI) is a long-term identity of a mobile user that is embedded in the user's UICC. An IMSI is usually presented as a 15 decimal-digit number but can be shorter. The first three digits are the Mobile Country Code (MCC), which are followed by the Mobile Network Code (MNC), either two or three digits. The length of the MNC depends on the length of the MCC. The remaining digits are the Mobile Subscription Identification Number (MSIN) within the network's customer base [55]. In GSM, 3G, and LTE, a mobile user may use IMSI to identify itself to the network.

The 5G users have a globally unique Subscription Permanent Identifier (SUPI). A SUPI may contain an IMSI or other identifiers as defined in Clause 5.9.2 in 3GPP TS 23.501 [56]. The Subscription Concealed Identifier (SUCI) is a privacy-preserving identifier containing the concealed SUPI. The construction of SUCI is specified in 3GPP TS 33.501 [57]. The structure of SUPI and SUCI is specified in 3GPP TS 23.003 [55]. In 5G, the user sends SUCI to the network to connect, and never sends its SUPI unless connecting in an emergency mode [57, p. 96].[6]

A mobile user also has an identity called Mobile Station International Subscriber Directory Number (MSISDN), usually known as a mobile phone number. The MSISDN is used for routing calls. To make a phone call, a caller uses the receiver's MSISDN in dialling up the receiver.

International Mobile Equipment Identity (IMEI) is a unique identity permanently assigned to mobile phones by their manufactures. The ME sends the IMEI so that the network can know the ME's expected behaviour. When the UE has no IMSI or valid GUTI, during emergency attach, the UE sends the IMEI to the network to identify itself [58, p. 18]. Sometimes the IMEI is accompanied by a Software Version Number (SV) in which case the identity is called International Mobile Equipment Identity and Software Version Number (IMEISV). The IMEI is used as a key identifier for blocking stolen MEs and also as a possible target identity of LI [21, 59]. Therefore, the UE usually sends the IMEI of the ME to the network.

Once a mobile user connects to an SN and establishes a security context[7], the SN assigns a temporary identity to the user over the radio network. The message that carries the temporary identity is confidentiality protected. This identity is known as Temporary Mobile Subscriber Identity (TMSI)[8] in GSM, and 3G. In LTE and 5G, the corresponding identity is called Globally Unique Temporary UE Identity (GUTI) and 5G-GUTI respectively [55]. In the case of roaming, the HN does not know a user's TMSI, GUTI, or 5G-GUTI. The SN also assigns temporary identities, e.g., C-RNTI (Cell Radio Network Temporary Identifier) to manage radio resources assigned to the user.

At the application layer, a mobile user may be identified by different identities by different application service providers. In this thesis, we refer to these identities with the term Service-specific Permanent User Identity (SPUID). These user identities are not concerned with the cellular networks, and the SN or HN do not need to know any SPUID of a user. Also, a user has identities that appear publicly in the domain of particular services. We refer to these domain-specific public identities with the term "screen-name".

[6] It is worth mentioning here that in 5G, the SN never sends SUPI over the air to the UE in any form of communication procedures or protocols.

[7] A security context refers to the set of security parameters, such as cryptographic keys and algorithm identifiers that are used to protect the data/signal exchanged between the UE and the SN [47, p. 129].

[8] A user may be assigned multiple TMSIs [55].

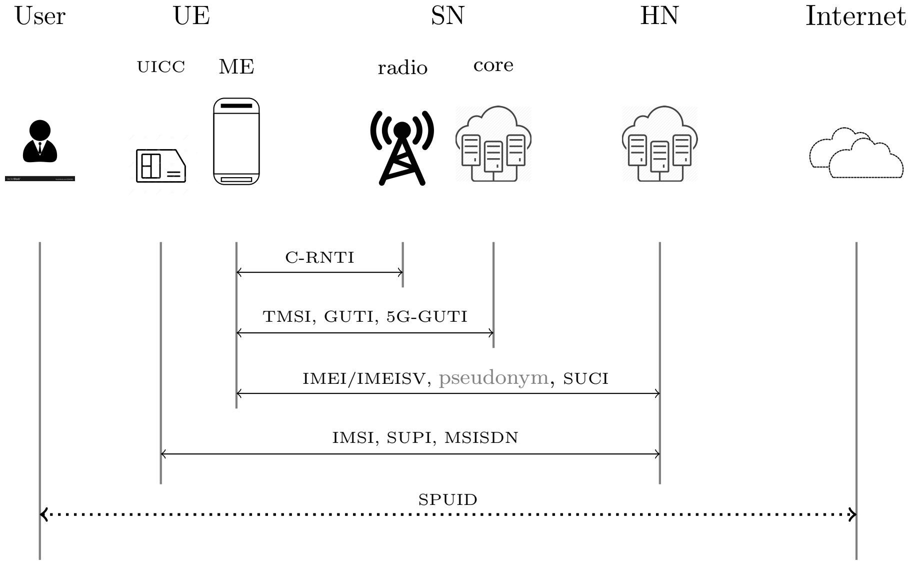

Figure 2.3: User Identities in Cellular Network.

In this thesis, we use a new temporary identity known as a pseudonym in a similar way as proposed in several publications [4, 60–62]. Pseudonyms have the same format as IMSI. Unlike other temporary identities (TMSI, GUTI, 5G-GUTI), these pseudonyms are assigned by the HN[9].

Figure 2.3 shows different identities of cellular network users along with their scope; some identities, e.g., C-RNTI, TMSI, are local only to the SN, and some other identities, e.g., IMSI, reach up to the HN. A UICC can host one or more USIMs [63, 64], and every USIM has a unique IMSI/SUPI. In most cases, one IMSI/SUPI is associated with one MSISDN. An IMEI can be associated with multiple IMSIs/SUPIs and MSISDNs, for example, when the ME has a UICC that hosts multiple USIMs or has multiple UICCs, e.g., in a dual-SIM phone [65].

Let us take some concrete examples. If a user changes to a new phone but keeps using the same UICC, then the IMEI and temporary identities such as C-RNTI or TMSI change, but the other identities such as IMSI, SUPI, or MSISDN remain the same. If a user removes a UICC from the ME and inserts a new UICC, then the IMSI, MSISDN, and associated temporary identities such as C-RNTI or TMSI change, but the IMEI remains the same. A user may lose a current

[9] The SN assigns TMSI, GUTI, and 5G-GUTI; the HN does not know them.

UICC, or the current UICC may become incompatible with the new phone, and he/she will get a new UICC from his mobile operator while keeping the associated MSISDN unchanged. A user's SPUIDs are independent of the UICC and phone, i.e., the SPUIDs of a user remain the same no matter which UICC and phone the user uses. Conversely, a user could lend his phone to somebody else in order to check Facebook. In such a case, the SPUID related to Facebook changes, but everything else remains the same.

2.2.5 Delegated Authentication in Cellular Networks

Through a delegated authentication system, a third-party application can authenticate its users and establish secret keys with the help of an authentication server. Such a system is useful for both users and third-party applications. Users can get rid of memorising many application-specific credentials. Third parties can get rid of maintaining authentication-related data.

Two delegated authentication systems exist in the previous generations of cellular networks. One is the Generic Bootstrap Architecture (GBA) [66, 67] and the other is Battery Efficient Security for Very Low Throughput Machine Type Communication Devices (BEST) [68]. In both of them, the user's HN takes the role of the authentication server. A new study on a 5G-suitable delegated authentication system known as AKMA (Authentication and Key Management for Applications) [69] is in the standardisation phase at the time of writing this.

2.3 Identity Privacy in Cellular Networks

In light of the discussion about privacy and identity privacy in Section 2.1, the long-term identities, e.g., IMSI, SUPI, MSISDN, IMEI, IMEISV, should be protected from third parties. Indeed, these identities should be required only by the SN/HN to provide network services to the user. For Lawful Interception (LI) purposes, the SN needs to know permanent identities, e.g., IMSI. For routing purposes, the SN/HN needs to know the MSISDN of the dialled number.[10]

Towards ensuring identity privacy in cellular networks, it is crucial to ensure that while sending the identities mentioned above to the network, a third party (an active/passive attacker) does not get access to them. Malicious third parties known as IMSI catchers are well-known threats against the identity privacy of

[10] The UE never sends/receives its own MSISDN to/from the network [70].

mobile users. An IMSI catcher tricks user devices to reveal IMSIs, and can track and monitor the location of a user [60].

A major portion of this thesis discusses mitigating IMSI catchers. An IMSI catcher may track/monitor a user by exploiting vulnerabilities in different protocols that the UE runs. In this thesis, we focus on the most easily exploitable vulnerability that IMSI catchers exploit. The vulnerability is due to the way the UE identifies itself to the network, explained in Chapter 3. Other vulnerabilities exist in the way paging messages are broadcast by the network or authentication reject messages are sent by the UE [71]. Due to vulnerabilities in paging and authentication reject messages, an attacker may track/monitor a known target by simply sending the target an SMS or an instant message in any instant messaging app [72].

The other identities mentioned in Subsection 2.2.4 are either concealed or short-lived. The short-lived identities do not usually pose threats to the identity privacy of mobile users. However, if those identities are not changed at the suggested intervals, privacy might be compromised [72]. Also, the temporary identities such as GUTIs, need to be unlinkable, i.e., two GUTIs assigned to the same UE should not be linkable. If the temporary identity assignment procedure is not designed carefully, the temporary identities might become linkable with each other, as shown in the 3G network [37].

A user may have many other identities at different layers of the protocol stack. For example, a user has many identities at the application layer to use various applications. Also, a user is usually assigned cookies by many web service providers. An advertisement agency may drop third-party cookies when a user loads a web page that comes with an advertisement of the advertisement agency. The third-party cookies give great leverage to the advertisement agencies to do targeted marketing; therefore breach the privacy of the users [73, 74]. A user's device also has permanent and unique identities for different radio interfaces such as Bluetooth or WiFi. A user's device has IP address(es), which sometimes can be used for a long period of time. Identity privacy may be compromised due to an attacker exploiting undesired exposures of the aforementioned identities, possibly by linking with other personal data [75–77]. However, usernames, cookies, IP addresses are artefacts that are not specified by the cellular network protocols.

In this thesis, we primarily focus on identity privacy threats due to undesired exposures of digital identities that are exchanged during the run of cellular network protocols. Moreover, we do not try to fix vulnerabilities related to all the cellular-network identities. This is because trying to fix every vulnerability would be too big a task to handle in a single work. We prioritise the vulnerabilities based on how easily and effectively an adversary can exploit a vulnerability. One vulnerability related to IMSI is the most easily and effectively exploitable by an adversary; the vulnerability is explained in Subsection 3.2.1. Most of our effort is invested in fixing this vulnerability. Please note that identity privacy in cellular networks can be compromised due to other reasons, e.g., unauthorised access to the servers of the network [78].

2.4 AKMA-related Problems

Authentication and Key Management for Applications (AKMA) is a delegated authentication system [69] that is under standardisation in the 3GPP community at the time of writing this. In the evolved ecosystem of 5G, use cases of AKMA vary from that of similar systems, e.g., GBA. In this thesis, we present a requirement analysis of AKMA.

The AKMA requirements include privacy requirements too. Some of the privacy requirements lie in the domain of identity privacy, and some do not. We discuss the AKMA-related privacy problems in Chapter 3.

2.5 Lawful Interception in Cellular Networks

Lawful Interception (LI) involves selectively intercepting communications of individual subscribers by legally authorised agencies. Lawful interception requirements vary from one jurisdiction to the other. A user, when roaming, is legally accountable to the jurisdiction of the country where the user is travelling. The cellular networks are designed so that they can fulfil the LI requirements of the country where the networks are operating.

For example, the authorised agency typically provides the long-term identifier of a target UE to the network operator, which must have the means to intercept communications of the correct target based on long-term or permanent identifiers associated with that target. The requirement also says that a network operator should be able to intercept without the need to rely on another network operator or jurisdiction. In particular, an SN does not need to share LI target identities of roaming UE with an HN and vice versa [79, 80].

The additional constraints of LI make the design of privacy-preserving protocols challenging in some cases. In Paper I, Paper II, and Paper III, we have discussed one of the LI requirements and how the requirement has created complications in designing solutions defeating IMSI catchers.

2.6 Cryptographic Techniques

All the solutions presented in this thesis are based on cryptographic techniques. Moreover, the solutions related to AKMA are all about establishing keys that would be used mostly for cryptographic purposes.

We use well-known cryptographic constructs towards solving privacy problems. Our solutions depend on the cellular network authentication protocols known as Authentication and Key Agreement (AKA): UMTS AKA [54], LTE AKA [58], and 5G AKA [57].

The AKA protocols are symmetric-key-based authentication protocols that include cryptographic operations such as encryption, message authentication, hashing, key derivation, and more. In 5G, the user identification protocol uses public-key-based cryptography [57] without a need for Public Key Infrastructure (PKI). Both the symmetric-key and public-key cryptography depend on a secure random number generator.

Details of all of these cryptographic constructs can be found in many standard cryptographic textbooks. For example, Douglas R. Stinson's book "Cryptography: Theory and Practice" [81] is a good read. A complexity-theory-based treatment of the security of these constructs can be found in the two parts of Oded Goldreich's book "Foundations of Cryptography" [82].

Our solutions also utilise the aforementioned cryptographic constructs on top of the AKA protocols. In addition, in Paper I, we use the Identity-Based Encryption (IBE) [81, 83–85]. In the solution, we also use digital signatures that may require a PKI. The privacy-related claims of the solution depend on the hardness of the Delayed-Target Diffie-Hellman Problem (DTDHP) [13].

Chapter 3

Research Problems

The thesis looks into privacy (mostly identity privacy) problems of users in cellular networks. In a communication system, generally, most identity privacy issues stem from the system's user authentication requirements. Had the system not required user authentication, the user would not need to identify himself; therefore, many privacy problems would disappear.

3.1 Background on Authentication

The users pay for the mobile network services they use. Therefore, it is unavoidable to authenticate a user so that the user can be billed correctly. To study the privacy problems of mobile users, it is essential to first look into the authentication mechanisms of mobile networks. Every generation of mobile networks has their versions of Authentication and Key Agreement (AKA) protocols: (i) GSM AKA, (ii) UMTS AKA, (iii) EPS AKA, and (iv) 5G AKA.

Before we present the problems that this thesis solves, we present the basic principles of these authentication protocols. The basic principles would be useful to understand the problems. We use greater details of these protocols in our solutions. Hence, in order to discuss the solutions, we need to recall details of the AKAs. In Chapter 4, the UMTS, EPS and 5G AKA are presented in greater detail so that solutions presented in the successive chapters can be read with fewer cross-checks in the literature.

The basic principles of all these authentication protocols are the same. They all hinge on a pre-shared permanent key that only the UE and the HN know. All the AKA protocols can be characterised by having two phases: (i) identification

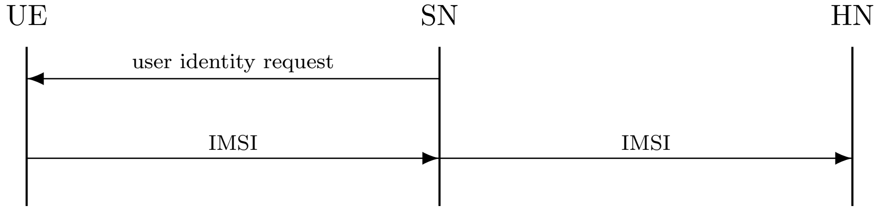

Figure 3.1: User Identification (GSM, 3G, LTE).

and (ii) authentication. In the following, we explain the two phases and then also briefly explain what happens after a successful run of an AKA protocol.

3.1.1 Identification

In the *identification* phase, the UE sends the user identity to the SN. The UE may use a temporary identity – e.g., TMSI or GUTI – that the SN assigned to the UE. The UE uses its IMSI, for example, when connecting for the first time with the SN or when the UE has lost the TMSI or GUTI. In case the UE sent a temporary identity, and the SN could not recognise the temporary identity, the SN may send an identity request to the user.

Once the UE receives the identity request, it responds to the request by sending the IMSI in cleartext to the SN (see Figure 3.1). This mechanism is the same in GSM, 3G and LTE [54, 58, 86, 87], and represents a breach of user identity privacy. The reason for sending IMSI in cleartext is the absence of a security context before the identification. Every UE has a different symmetric key pre-shared with the HN. If the IMSI could be encrypted, then the HN would not know what key to use to decrypt the IMSI.

If a 5G SN sends an identity request to a 5G user, in the response, the UE includes a SUCI instead of the SUPI. The SN forwards the SUCI that was sent by the UE to the HN. Later, the HN sends the SUPI of the user to the SN (see Figure 3.2). The SUCI is constructed using the public key of the user's HN [57].

If the subsequent authentication becomes successful and a security context established between the UE and SN, then the SN assigns a temporary identity (TMSI, GUTI, or 5G-GUTI) to the UE. The message that carries the temporary identity from the SN to the UE is confidentiality protected. In the next identification, the UE may use the temporary identity instead of IMSI or SUCI.

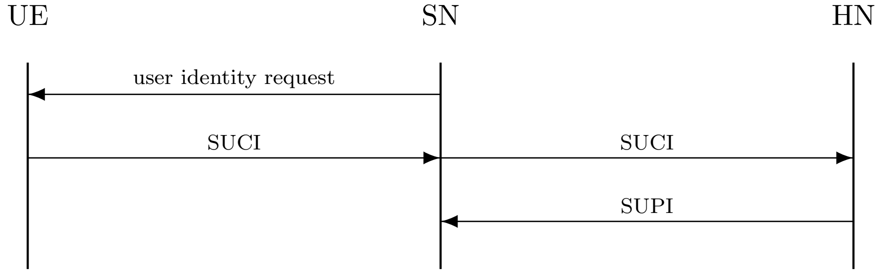

Figure 3.2: User Identification (5G).

3.1.2 Authentication

In the *authentication* phase, the HN finds the permanent key K of the user (see Figure 3.3). The HN chooses a random challenge RAND to be solved in the UICC inside the user's device. Then the HN computes a message authentication code MAC of the random challenge RAND using an one-way function f_1 that takes the RAND and K as inputs.[1] Then, the HN computes the expected response XRES to the challenge RAND based on the permanent key K using the one-way function f_2. The HN also computes a session key K_S based on the RAND and K using a one-way function h.[2] The HN forwards the random challenge RAND, its solution XRES, and MAC to the SN. The SN forwards the RAND and MAC to the user.[3]

The UE knows the required permanent key K. It computes the solution RES of the challenge RAND using the one-way function f_2 in the same way the HN computed the XRES. The UE also verifies the message authentication code by computing the MAC using the function f_1 in a similar way as the HN computed and comparing it with the MAC sent by the SN. If the verification of the MAC is successful, the UE considers that the involved HN is authentic and the HN has authorised the SN. The UE forwards the solution RES to the SN. The UE also computes a session key K_S based on the permanent key and the random

[1] In 3G, LTE, and 5G, the function f_1 also takes some other inputs which we do not mention here in the discussion of the basic principle of the authentication mechanism.

[2] The session key K_S, and one-way function h are just placeholders for the sake of the discussion in this section. In 3G, there are two such session keys called CK and IK, which are computed using one-way functions f_3 and f_4 respectively. In LTE and 5G, the session keys are called K_{ASME}, and K_{AUSF} respectively, which are computed using a key derivation function that takes more inputs than functions h, f_3, or f_4.

[3] In 3G, LTE, and 5G, the MAC is accompanied with some other parameters.

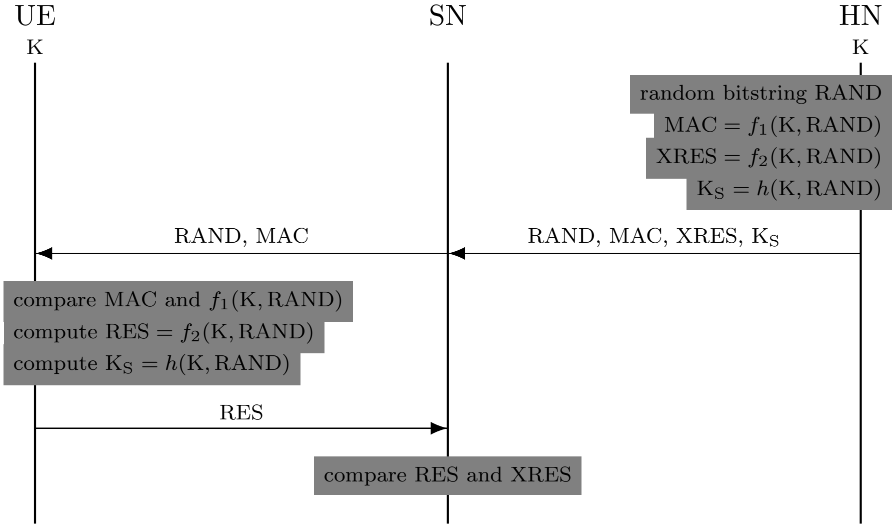
UE
K
SN
HN
K
random bitstring RAND
MAC = $f_1(\mathrm{K},\mathrm{RAND})$
XRES = $f_2(\mathrm{K},\mathrm{RAND})$
$\mathrm{K_S} = h(\mathrm{K},\mathrm{RAND})$
RAND, MAC
RAND, MAC, XRES, $\mathrm{K_S}$
compare MAC and $f_1(\mathrm{K},\mathrm{RAND})$
compute RES = $f_2(\mathrm{K},\mathrm{RAND})$
compute $\mathrm{K_S} = h(\mathrm{K},\mathrm{RAND})$
RES
compare RES and XRES

Figure 3.3: Principle of Authentication in Mobile Networks.

challenge using the one-way function h_3. The SN compares XRES and RES. If they are equal, the SN considers the UE to be authentic.

The idea is that one must know the permanent key K of the user to solve the challenge, compute the MAC, or $\mathrm{K_S}$. Since only the UE and HN know the permanent key K, the authentication would be successful if and only if the legitimate UE and correct networks participate.

The description above is the basic principle involved in every AKA except in GSM. One example of how GSM AKA is different is that no MAC is involved in GSM AKA. All the parameters related to AKAs are not discussed in this section. For example, SQN, the sequence number parameter which is used in UMTS, LTE, and 5G AKA to defeat replay attacks, is not discussed in this section. In Chapter 4, we give a more elaborate presentation of UMTS, LTE and 5G AKA.

3.1.3 Security after a Successful AKA Run

Following the AKA, the UE and the SN run additional procedures (known as security mode command procedures) to establish a security context and start cryptographic data protection. The UE and the SN may not use the session key directly to protect data. In LTE and 5G, during the security mode command procedure, the SN and UE generate more keys for different kinds of protections of different kinds of data. After the activation of confidentiality protection, the SN may send a temporary identity (TMSI, GUTI, 5G-GUTI) to the UE. Please see 3GPP TS 33.102 [54, Clause 6.1.2], TS 33.401 [58, Clause 7.1], and TS 33.501[57, Clause 6.12.3] for details.

In this thesis, the solutions to the privacy problems do not propose changes in the existing post-AKA security mechanisms. Consequently, existing post-AKA security procedures are not very relevant in this thesis. Therefore, we limit our discussion only up to the end of the AKA protocols.

3.2 Our Research Problems

This thesis looks into two problems. The first problem is related to defeating IMSI catchers. An IMSI catcher can exploit various vulnerabilities. In Subsection 3.2.1, we present a summary of these vulnerabilities. However, we do not try to fix all those vulnerabilities. Instead, we try to fix the most easily exploitable vulnerability. The second problem is related to developing AKMA standards. The research problem includes analysing requirements that AKMA should fulfil. We put a particular focus on identifying (potentially new) privacy requirements. The research problem also includes developing a solution mechanism that fulfils the identified requirements. In Subsection 3.2.2, we present a brief overview of AKMA-related research problems.

3.2.1 IMSI Catchers

When turned on, a UE scans for legitimate SNs and selects one SN to connect. The process is called PLMN (Public Land Mobile Network) selection process [88]. Once a PLMN is selected, the UE selects a suitable radio interface, also called a cell, and camp on it. The process is known as cell selection process [88]. The UE may switch from a radio interface, on which it is currently camped, to a newly discovered and more suitable radio interface. The process is known as cell re-selection process [88]. Different types of measurements are used in different radio access technologies and modes for cell selection and re-selection [88]. The performance requirements for the measurements are specified in 3GPP TS 25.123 [89] and TS 25.133 [90].

The primary purpose of cell re-selection is to ensure that the UE camps on or connects to the best radio interface (cell) in terms of the quality of radio condition [91]. This tendency of switching to a more suitable radio interface is perhaps motivated by the goal of achieving an uninterrupted radio connection, especially when the user is moving. When a UE is moving, the appearance of a more suitable radio interface implies that the current radio interface, on which the UE is currently camped, may become unavailable soon.

An IMSI catcher is a rogue device with a radio interface; it impersonates a legitimate SN and presents a very tempting radio connection to its near neighbourhood. Consequently, the nearby UEs camp on the rogue device and start listening to its broadcast channel. At this point, the rogue device can run various 3GPP-defined procedures with the UE and try to exploit vulnerabilities in those procedures. The most easily exploitable vulnerability that the rogue device exploits is in the identification procedure, which is explained in the following.

The rogue device sends an *identity request* to all the UEs that are listening. The security architecture of cellular networks is designed in a way that it does not facilitate a mechanism that could be used at this stage by the UEs to distinguish a rogue device from a legitimate network. The security architectures, except in 5G, also do not facilitate a mechanism to send an encrypted IMSI to the (potentially) rouge device. Currently, the procedure specifies only one possible response to the identity request, i.e., sending the IMSI in plaintext (encrypted in 5G). Consequently, according to the protocol, all the UEs that received the identity request respond by giving their IMSIs in plaintext (encrypted only in 5G) to the rogue device, thus the name "IMSI catcher". Please see a pictorial depiction of the attack in Figure 3.4. By observing the presence and absence of an IMSI at a location, and using some other auxiliary information about the IMSI, the IMSI catcher can infer the presence of a user in nearby places.[4]

To breach the location privacy of a user, an IMSI catcher may exploit other vulnerabilities too. For example, an IMSI catcher can exploit the fact that the authentication reject messages (sent by the UE to the SN) are sent in plaintext and be able to monitor the presence/absence of a specific target user; the attack is explained by Arapinis et al. [71]. Also, vulnerabilities in other procedures, e.g., the paging and GUTI allocation procedures, can be exploited by an IMSI catcher in order to breach the identity privacy of a user. Exploiting these vulnerabilities usually enables the IMSI catchers to link different temporary identities, or link between a known long-term identity and a temporary identity of the victim. Khan and Martin have summarised a list of these vulnerabilities and the associated attacks in their survey paper on subscription privacy in 5G [92].

[4] More sophisticated attacks start with IMSI catching, and those attackers are also sometimes called IMSI catchers [60]. However, in this thesis, we will use the name "IMSI catchers" only to refer to their ability to breach the identity privacy of mobile users.

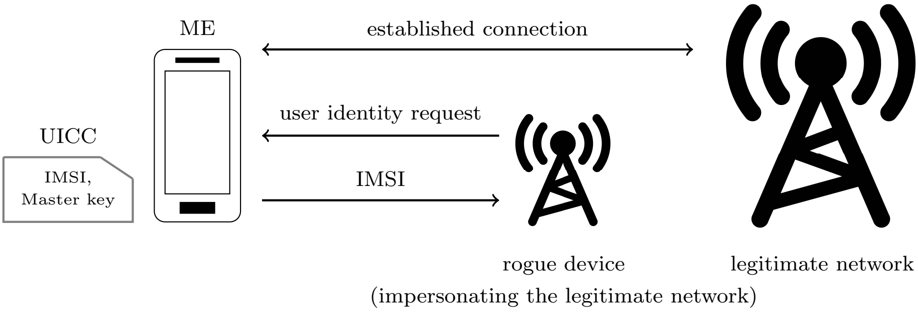

Figure 3.4: IMSI catching in GSM, 3G, and LTE.

The vulnerability in the identification procedure is the most easily exploitable because, among other reasons, the IMSI catcher does not need to know any long-term identity of a victim before mounting the attack. Therefore, the attack can be mounted simultaneously against all the nearby users that find the IMSI catcher's radio signal suitable. Other vulnerabilities cannot be exploited in such a simultaneous fashion because the IMSI catcher would have to target a particular victim and analyse the information sniffed in from the radio channel before arriving at any conclusion about the victim's presence or absence at the location.

This thesis focuses on fixing the vulnerability in the identification procedure to defend against IMSI catchers, and it keeps the other vulnerabilities out of its scope. We look for fixes to the vulnerability from three angles:

1. in the 5G network,
2. in legacy networks (3G and LTE), and
3. when a 5G user tries to connect to a 3G or LTE network.

The vulnerability in the identification procedure has been fixed in 5G by using public-key cryptography. So, it is worth explaining why this thesis includes a solution to fix a vulnerability in 5G that has already been solved in 5G.

During the early phase (in 2016 and 2017) of our research, 5G was undergoing the process of standardisation, and a solution to fix the vulnerability in 5G was an open question. Therefore, we contributed to the literature by presenting a potential solution based on the LTE authentication and key agreement protocol and identity-based encryption. Besides, our solution has other benefits on top of fixing the vulnerability. Due to the other benefits, it might be considered to be a preferred solution in some use cases, for example, in a smart factory or be useful in future releases of standards.

Fixing the vulnerability in the identification procedure of the legacy networks is challenging. This challenge is due to the existing widespread installation of those networks across the world.

3.2.2 Privacy-preserving AKMA

The new delegated authentication system named AKMA has been released in the second phase of 5G in July 2020. It supports authentication and key management for applications and 3GPP services, including the IoT use cases based on 3GPP credentials in the 5G system [69, 93]. In AKMA, the application provider delegates the authentication of its user to the Home Network (HN) of that user.

Our research problem includes an investigation towards finding requirements that AKMA should fulfil. Our investigation is mostly based on a comparative requirement analysis of AKMA, GBA, and BEST; we compared what requirements were fulfilled by GBA and BEST, with what requirements AKMA is envisioned to fulfil. We have identified some requirements that were fulfilled by GBA/BEST but not yet considered to be requirements of AKMA. Our requirement analysis puts a particular focus on finding (potentially new) privacy requirements. We have identified two privacy requirements in AKMA: (i) privacy of usage – the identity provider (i.e., the HN) should not know what application providers a user is using, and (ii) identity unlinkability – two network application providers, by colluding, should not be able to link two of their users. The research problem includes finding a privacy mode AKMA solution that can co-exist with a normal mode AKMA solution.

Chapter 4

Cellular Authentication Protocols

In this thesis, our solutions to the privacy problems leverage messages of UMTS, EPS and 5G AKA protocols. So, we present details of these protocols that are needed to understand our solutions. We do not describe parts of these protocols that are not important to understand our solutions. For example, we skip the error/failure messages that are sent by the UE or the network when something does not happen as expected.

4.1 UMTS and EPS AKA

The UMTS and EPS AKA[1] are defined in 3GPP TS 33.102 [54] and 33.401 [58], respectively. The EPS AKA is built on top of the UMTS AKA and reuses many cryptographic building blocks of the UMTS AKA. In addition, the EPS AKA has some new functionality and uses new cryptographic building blocks on top of those in UMTS AKA. The TS 33.401 does not include a stand-alone description of EPS AKA. It gives only the changes over UMTS AKA.

In this section, we give a high-level description of both UMTS and EPS AKA together. We explicitly mention their similarities and differences when necessary. We point out the relevant building blocks and refer to documents that define those building blocks. We have shown both the identification and authentication phases in Figure 4.1. The first six steps (from Step 1 to Step 6) comprise the identification phase. The rest of the steps comprise the authentication phase.

Every UMTS/LTE user has a USIM that contains an $\langle \mathrm{IMSI}, \mathrm{K} \rangle$ pair. An HN maintains a database of such pairs, one pair for each of its subscribers. One

[1] The UMTS and EPS AKA are also known as 3G and LTE AKA, respectively.

important assumption is that the permanent key K is known only by the USIM and the HN.[2]

4.1.1 Identification

In the following, we describe the identification phase step-by-step, as depicted in Figure 4.1:

(1) In a radio connection request message, the UE sends one of its identities in an attempt to connect to the SN. If the UE has a TMSI or a GUTI that the SN assigned previously, the UE sends the TMSI or GUTI, otherwise the UE can send its IMSI [94, 95].

(2) Using the received TMSI or GUTI in Step 1, the SN tries to resolve the IMSI from its lookup table. If the SN cannot find the TMSI or GUTI in the lookup table, then the SN continues from Step 3. If the SN received the IMSI itself in Step 1 or can resolve the IMSI using the TMSI or GUTI, the SN jumps to Step 5.

(3) The SN sends an identity request to the UE.

(4) The UE sends a response to identity request by sending the IMSI in cleartext. Please note that sending the response in cleartext is a weakness that is exploited by the IMSI catchers to attack users' identity privacy.

(5) The SN checks if it has a pre-fetched unused Authentication Vector (AV) for the user. If yes, then the SN jumps to Step 9. Otherwise, the SN continues from Step 6.

(6) The SN resolves the HN name from the MCC and MNC part of the IMSI and sends an AV request to the HN. The AV request includes the IMSI. In case of requesting for an EPS AV (sometimes called LTE AV), the SN includes the fields $\mathrm{SN}_{\mathrm{id}}$ and "network type" to indicate the LTE network.[3] The $\mathrm{SN}_{\mathrm{id}}$ consists of MCC and MNC of the SN, and is encoded as described in 3GPP TS 33.401 [58, Annex A.2].

[2] It is not feasible for an attacker to read the permanent key K from the USIM due to the tamper resistance of the UICC that hosts the USIM. Also, the HN employs state-of-the-art security mechanisms to protect the database of $\langle \mathrm{IMSI}, \mathrm{K} \rangle$ pairs from unauthorised access.

[3] In 5G, $\mathrm{SN}_{\mathrm{name}}$ is used instead of $\mathrm{SN}_{\mathrm{id}}$. The $\mathrm{SN}_{\mathrm{name}}$ includes the network type, e.g., 5G (see Subsection 4.2.1).

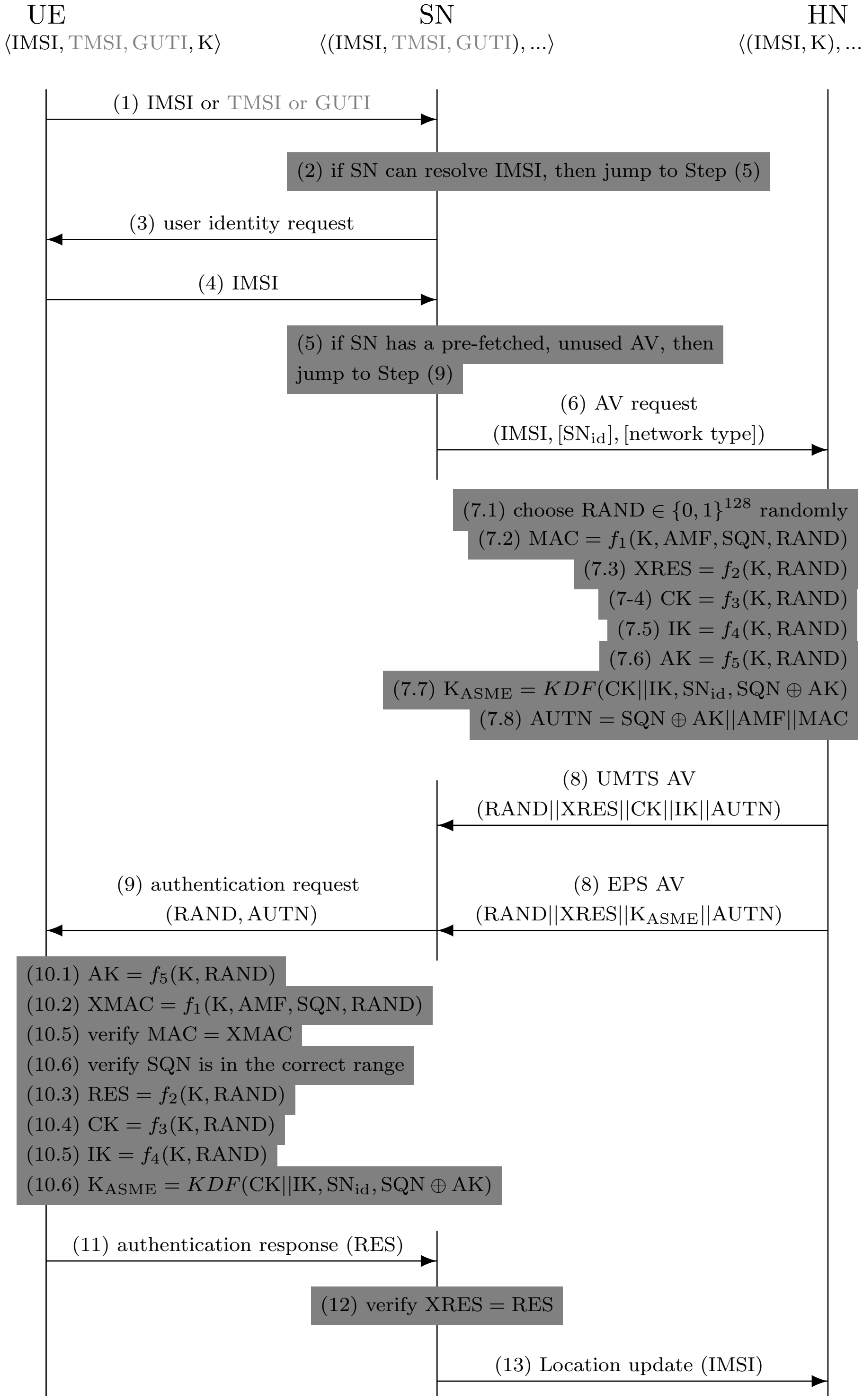
UE
$\langle$IMSI, TMSI, GUTI, K$\rangle$
SN
$\langle$(IMSI, TMSI, GUTI), ...$\rangle$
HN
$\langle$(IMSI, K), ...$\rangle$
(1) IMSI or TMSI or GUTI
(2) if SN can resolve IMSI, then jump to Step (5)
(3) user identity request
(4) IMSI
(5) if SN has a pre-fetched, unused AV, then jump to Step (9)
(6) AV request
(IMSI, [SN$_{\mathrm{id}}$], [network type])
(7.1) choose RAND $\in \{0,1\}^{128}$ randomly
(7.2) MAC $= f_1$(K, AMF, SQN, RAND)
(7.3) XRES $= f_2$(K, RAND)
(7-4) CK $= f_3$(K, RAND)
(7.5) IK $= f_4$(K, RAND)
(7.6) AK $= f_5$(K, RAND)
(7.7) K$_{\mathrm{ASME}}$ $= KDF$(CK||IK, SN$_{\mathrm{id}}$, SQN $\oplus$ AK)
(7.8) AUTN $=$ SQN $\oplus$ AK||AMF||MAC
(8) UMTS AV
(RAND||XRES||CK||IK||AUTN)
(9) authentication request
(RAND, AUTN)
(8) EPS AV
(RAND||XRES||K$_{\mathrm{ASME}}$||AUTN)
(10.1) AK $= f_5$(K, RAND)
(10.2) XMAC $= f_1$(K, AMF, SQN, RAND)
(10.5) verify MAC $=$ XMAC
(10.6) verify SQN is in the correct range
(10.3) RES $= f_2$(K, RAND)
(10.4) CK $= f_3$(K, RAND)
(10.5) IK $= f_4$(K, RAND)
(10.6) K$_{\mathrm{ASME}}$ $= KDF$(CK||IK, SN$_{\mathrm{id}}$, SQN $\oplus$ AK)
(11) authentication response (RES)
(12) verify XRES $=$ RES
(13) Location update (IMSI)

Figure 4.1: UMTS/EPS AKA.

4.1.2 Authentication

To describe the authentication phase, we continue the numbering of the steps from the identification phase. In the following, we describe the authentication phase step-by-step, as depicted in Figure 4.1:

(7) In this step, the HN prepares an AV. If the “network type” field is not present in the authentication request sent by the SN, then the HN prepares a UMTS AV (sometimes called 3G AV). Otherwise, the HN prepares an EPS AV. We explain these two AVs in the following:

- ***UMTS Authentication Vector:*** A UMTS AV includes the following: (i) a 128-bit long random challenge called RAND, (ii) a bit string XRES which is the expected response to the challenge (iii) two keys called CK and IK to be used as confidentiality and integrity keys between the UE and the SN, and (iv) an authentication token AUTN. The construction of an UMTS AV is described in 3GPP TS 33.102 [54, Clause 6.3.2]. The message authentication code MAC, expected response of the challenge XRES, confidentiality key CK, integrity key IK, and anonymity key AK are computed using one-way functions f_1, f_2, f_3, f_4, and f_5 respectively. The AUTN is a concatenation of a bunch of parameters (explained later in this step): $\mathrm{SQN} \oplus \mathrm{AK}$, AMF, and MAC.

  The input parameters of function f_1 are the following: $\mathrm{K}, \mathrm{AMF}, \mathrm{SQN}$ and RAND. The AMF is a 16-bit long string, used to indicate different meta-information about the AV to the USIM [54]. For example, the HN sets the most significant bit (also called the separation bit) of the authentication management field (AMF) to “0” to indicate the UE that the associated AV is not meant for LTE uses [47, p. 117]. The sequence number SQN is a counter maintained by the HN for each of its users. The counter increases each time the HN generates an AV for the user. See Annex C.1.1 in 3GPP TS 33.102 for details on SQN.[4] The other four functions (f_2, f_3, f_4, f_5) take two inputs: the permanent key K and the random challenge RAND.

  The 3GPP standards do not require functions f_1, f_2, f_3, f_4, and f_5 to be standardised; this is because these functions are used in the USIM and HN, both of which are managed by the same entity, i.e.,

[4] The SQN plays a central role in defeating replay attacks.

the HN. However, a set of informative example algorithms for these functions have been developed by a special task force within 3GPP. This set is called MILENAGE, and the specifications are documented in 3GPP TS 35.205, TS 35.206, TS 35.207, TS 35.208 [96–99]. Niemi and Nyberg [1] have discussed the MILENAGE algorithm.

– ***EPS Authentication Vector:*** An EPS AV includes (i) a 128-bit long random challenge called RAND, (ii) a bit string XRES which is the expected response to the challenge, (iii) a key called $\mathrm{K}_{\mathrm{ASME}}$ to be used as a local master key in the SN, and (iv) an authentication token AUTN. The RAND and XRES are computed in the same way as they are computed in the above discussion. Unlike UMTS, the separation bit of AMF in the AUTN is set to “1”. All the other parameters in the AUTN, i.e., $\mathrm{SQN}, \mathrm{AK}$, and MAC are constructed in the same way as in UMTS.

  An EPS AV does not include the keys CK, IK. Instead, it includes the key $\mathrm{K}_{\mathrm{ASME}}$ derived (explained later in this step) from the keys CK, IK and SN identity $\mathrm{SN}_{\mathrm{id}}$. The introduction of the key $\mathrm{K}_{\mathrm{ASME}}$ was necessary due to the decision of using 3G USIM in LTE, and the new requirement of cryptographic network separation. The key $\mathrm{K}_{\mathrm{ASME}}$ has an additional desirable effect: it reduces the frequency with which AVs need to be fetched from the HN. Reasons for introducing $\mathrm{K}_{\mathrm{ASME}}$ are discussed by Forsberg et al. [47, p. 101].

  To derive the key $\mathrm{K}_{\mathrm{ASME}}$, the HN uses the key derivation function KDF, which is defined in 3GPP TS 33.220 [67]. The KDF function computes a hash of an input string based on the input key. In Figure 4.1, we present the KDF function in a simple way by showing only those parameters which are involved in our discussion. More parameters are involved in the actual computation. In our simplified notation, the first parameter is the key for the hashing, and the rest of the parameters are used to construct the message according to 3GPP TS 33.401 [58, Annex A.2].

  The lengths of different authentication parameters that are used to construct an AV are defined in Subsection 6.3.7 of 3GPP TS 33.102 [54]. The key $\mathrm{K}_{\mathrm{ASME}}$ is 256 bits long, but it still has a key entropy of only 128 bits when it is derived from the permanent key K which has 128 bits.[5]

[5] The permanent key can be of different lengths, as discussed in Subsection 2.2.3.

(8) The HN forwards the AV to the SN. If the SN sent an authentication request for a UMTS AKA, then the HN forwards the UMTS AV. If the SN sent an authentication request for an EPS AKA, then the HN forwards the EPS AV.

(9) The SN stores the $(\mathrm{XRES}, \mathrm{CK}, \mathrm{IK})$ in case of UMTS AKA. In the case of EPS AKA, the SN stores $(\mathrm{XRES}, \mathrm{K}_{\mathrm{ASME}})$. Then the SN forwards RAND and AUTN to the UE.

(10) Upon receiving the RAND and AUTN, the ME transfers RAND and AUTN to the USIM. The USIM does a series of computations, similar to the computations that the HN did to construct the AV in Step 7.

The USIM first computes the anonymity key $\mathrm{AK} = f_5(\mathrm{K}, \mathrm{RAND})$, retrieves the sequence number $\mathrm{SQN} = (\mathrm{SQN} \oplus \mathrm{AK}) \oplus \mathrm{AK}$, and computes XMAC in the same way the HN computed MAC. It compares this XMAC with MAC which is included in AUTN. If the two are different, the UE sends an authentication failure message back to the SN with an indication of the cause and abandons the procedure.

The USIM then verifies the freshness of the AV by checking whether the received sequence number SQN is in the correct range, as described in the Annex C.2.2 in 3GPP TS 33.102 [54]. If the verification fails, the UE sends an authentication failure message back to the SN with an indication of the cause, and the UE abandons the procedure. The synchronisation failure message contains the parameter AUTS. Algorithms for constructing AUTS are described in 3GPP TS 33.102 [54].

If the verifications mentioned above are positive, the USIM computes RES, CK, and IK in the same way that the HN computed $\mathrm{XRES}, \mathrm{CK}$, and IK in Step 7. The USIM returns $\mathrm{RES}, \mathrm{CK},$ and IK to the ME. At this point, the ME treats the involved HN and SN as authentic.

If the separation bit is set to “0” then the AV is usable only in legacy contexts such as GSM and UMTS [58, Clause 6.1.2]. If the separation bit is set to “1”, then the AV is usable in LTE [58, Clause 6.1.2] and 5G [57, Clause 6.1.3.2] contexts.

The ME checks whether the separation bit of the AMF is set correctly. If the separation bit is set to “1” and the UE is connecting to an LTE network, then the ME computes $\mathrm{K}_{\mathrm{ASME}}$ using key $\mathrm{CK}||\mathrm{IK}$ using the KDF function in the same way as the HN computed $\mathrm{K}_{\mathrm{ASME}}$ in Step 7. If the separation bit is not set correctly, then the ME sends an authentication failure message back to the SN with an indication of the cause [47, p. 122].

(11) The ME responds to the SN by sending a user authentication response message that includes RES.

(12) The SN checks if the RES equals XRES. If so, the SN considers the user to be authentic. If not, depending on the type of identity used by the UE in the identification phase, the SN may initiate further identity requests, i.e., go back to Step 3, or send an authentication-reject message towards the UE (see 3GPP TS 24.301 [100]).[6]

(13) If the UE connected to the SN for the first time and the authentication was successful, then the SN sends a Location Update (LU) message to the HN by including the IMSI of the subscriber [101–103]. Please note that the message is called update location request in 3GPP TS 33.401 [102]. Due to the LU message, the HN knows the SN in which the user can be found.

After successful and affirmative completion of the AKA, more algorithm-specific cryptographic keys between the UE and SN are established based on $\mathrm{K}_{\mathrm{ASME}}$. To indicate the selected algorithms and to start ciphering and integrity protection, two Security Mode Command (SMC) procedures are used. Some cryptographic data protection already starts during the run of SMC procedures. Other cryptographic protection starts after the SMC procedures complete. Forsberg et al. [47, Chapter 8] have discussed the SMC procedures and starts of cryptographic protection.

4.2 5G AKA

The primary purpose of the 5G AKA is the same as LTE AKA, i.e., authentication and key agreement between the UE and the network. Nevertheless, 5G AKA has two main additions. One is privacy-preserving identification: a 5G user sends a SUCI – a privacy-preserving identity, instead of its long-term identity SUPI. The other is increased home control: the HN obtains cryptographically secure confirmation that the authentication between the SN and UE has been successful, i.e., the SN cannot lie about the success of the authentication.

[6] We do not present the authentication failure/reject messages in Figure 4.1.

In this section, we describe first the privacy-preserving identification phase of 5G AKA. Then we describe the authentication part of 5G AKA and explain how it achieves increased home control. Since the privacy-preserving identification is a major improvement in 5G, we present the identification phase using a separate figure before explaining the authentication phase.

4.2.1 Identification

The identification process of a 5G user is defined in 3GPP TS 33.501 [57, Clause 6.1.2]. We have written out the process in Figure 4.2. In the following, we explain the process step-by-step:

(1) In a radio connection request message, the UE sends one of its identities in an attempt to connect to the SN. The UE first checks if it has a valid 5G-GUTI that can be used to identify itself to the SN. If not, it jumps to Step 6.

(2) The UE sends the 5G-GUTI to the SN.

(3) The SN checks if it has a SUPI mapped to the 5G-GUTI. If not, it jumps to Step 5.

(4) The SN sends an authentication request message to the HN. The request includes the SUPI of the user and the $\mathrm{SN}_{\mathrm{name}}$. The $\mathrm{SN}_{\mathrm{name}}$ is the concatenation of a service code and the $\mathrm{SN}_{\mathrm{id}}$ with a separation character ":" such that the service code prepends the $\mathrm{SN}_{\mathrm{id}}$. The value of the service code is set to "5G" to indicate that the authentication is intended between a 5G core and a UE [57, Clause 6.1.1.4]. At this point, the identification process ends.

(5) The SN sends an identity request to the UE.

(6) The UE computes a SUCI of the user. The UE constructs a SUCI by encrypting the SUPI, using the public key of the HN. The details of how a SUCI is constructed are specified in 3GPP TS 33.501, in Annex C [57]. It is also summarised by Khan and Martin [92]. We presented a brief overview of the mechanism in Paper III[7].

[7] The UE conceals the 5G user's long-term identities with Elliptic Curve Integrated Encryption Scheme (ECIES) [104, 105] before sending them to the SN. The ECIES is a hybrid encryption scheme that combines elliptic-curve-based public-key cryptography with secret key

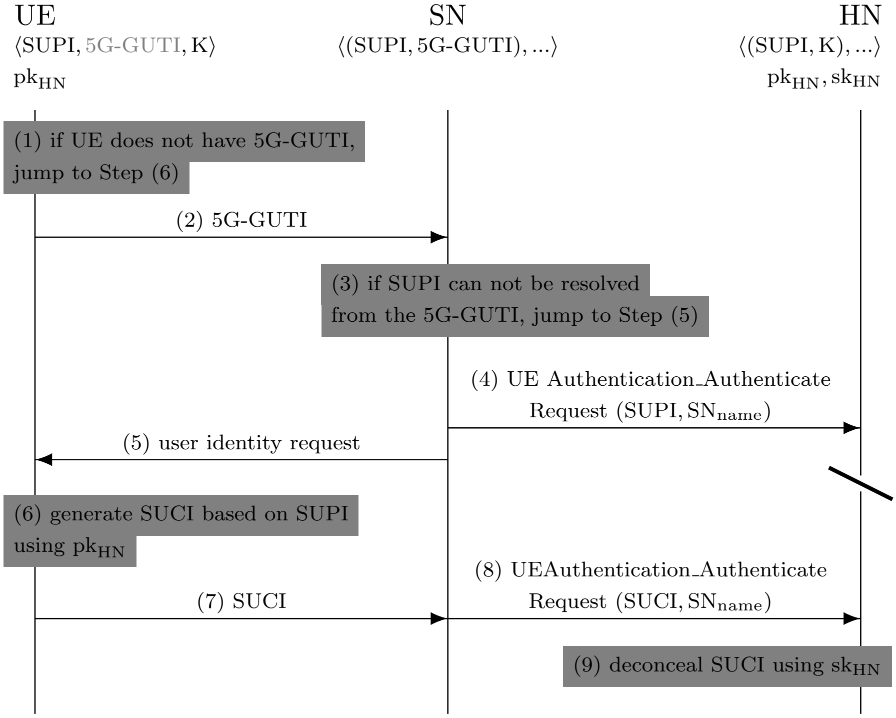
UE
⟨SUPI, 5G-GUTI, K⟩
$\mathrm{pk_{HN}}$
SN
⟨(SUPI, 5G-GUTI), ...⟩
HN
⟨(SUPI, K), ...⟩
$\mathrm{pk_{HN}}$, $\mathrm{sk_{HN}}$
(1) if UE does not have 5G-GUTI, jump to Step (6)
(2) 5G-GUTI
(3) if SUPI can not be resolved from the 5G-GUTI, jump to Step (5)
(4) UE Authentication_Authenticate Request (SUPI, $\mathrm{SN_{name}}$)
(5) user identity request
(6) generate SUCI based on SUPI using $\mathrm{pk_{HN}}$
(7) SUCI
(8) UEAuthentication_Authenticate Request (SUCI, $\mathrm{SN_{name}}$)
(9) deconceal SUCI using $\mathrm{sk_{HN}}$

Figure 4.2: Identification of a 5G User.

(7) The UE sends the SUCI to identify itself in the registration request message to the SN.

(8) The SN sends an authentication request message to the HN. The message includes the SUCI and the $\mathrm{SN}_{\mathrm{name}}$. The construction of $\mathrm{SN}_{\mathrm{name}}$ is explained in Step 4 above.

(9) The HN decrypts the SUPI of the user from the SUCI, using the HN's private key. Details are specified in Annex C of 3GPP TS 33.501 [57]. At this point, the identification process ends.

Introduction of SUCI fixes the vulnerability in the identification procedure in the 5G network. This is because the UE would never send its SUPI to the SN in plaintext except when the UE chooses to use the null-scheme to compute the SUCI. The null-scheme uses the identity function instead of a real encryption/decryption function. In practice, it is equivalent to skipping the encryption. For technical reasons, it is more convenient to have an explicit scheme instead of bypassing the encryption step. It is unlikely that an IMSI catcher would be able to trick a UE to use the null-scheme by sending an identity request. This is because the UE shall generate a SUCI using "null-scheme" only in the following cases [57, Section 6.12]:

- the UE is setting up an unauthenticated emergency session, and it does not have a 5G-GUTI to the chosen PLMN, or
- the home network has configured "null-scheme" to be used, or
- the home network has not provisioned the public key needed to generate a SUCI.

Please note that the SUCI contains parts of the SUPI in plaintext. For example, MCC and MNC, which are needed for routing the SUCI to the HN, are not encrypted [111]. Due to this routing-related information sent in plaintext, SUCI may not provide strong privacy. For example, when only one user, who is

cryptography; it is a semantically secure probabilistic encryption scheme ensuring that successive encryptions of the same plaintext with the same public key result in different ciphertexts with very high probability. Specifically, [57] includes two ECIES profiles, both for the approximately 128-bit security level. Both profiles use AES-128 [106] in CTR mode [25] for confidentiality and HMAC-SHA-256 [107, 108] for authenticity in the secret key cryptography part but use either Curve25519 [109, 110] or secp256r1 [105] elliptic curves for the public key cryptography part.

known to an IMSI catcher, of an HN is roaming to a particular place, the IMSI catcher can track the user. The IMSI catcher would send an identity request to all the UEs that try to camp on it; when the users respond to the identity request by sending SUCI, the IMSI catcher observes only the MCC and MNC parts disregarding the encrypted part of the SUCI. If there is only one UE that sends the MCC and MNC of the targeted user's HN, then the IMSI catcher may infer that they most probably come from the UE of the targeted user.

4.2.2 Authentication

The authentication process in 5G is defined in 3GPP TS 33.501 [57]. We have presented the 5G authentication process in Figure 4.3. In the following, we explain the process step-by-step:

(1) First, the HN generates an AV as specified in 3GPP TS 33.102 [54]. So, the RAND, MAC, XRES, CK, IK, AK, and AUTN are generated in the same way they are generated in 3G/LTE. In addition, in 5G, the HN generates a few more strings that are described below:

   - The HN generates key $\mathrm{K}_{\mathrm{AUSF}}$ using the KDF function.[8] The KDF function computes a hash of an input message using the input key. The input key to the KDF function is CK||IK. Some other inputs are $\mathrm{SN}_{\mathrm{name}}$ and $\mathrm{SQN} \oplus \mathrm{AK}$. The construction of the input message to the KDF function is specified in 3GPP TS 33.501 [57, Annex A.2].
   - The expected response XRES^* is also computed using the KDF function. The input key is CK||IK. The other inputs are $\mathrm{SN}_{\mathrm{name}}$, RAND, XRES. The construction of the input message to the KDF function is specified in 3GPP TS 33.501 [57, Annex A.4].
   - The key $\mathrm{K}_{\mathrm{SEAF}}$ is also computed using the KDF function. The input key is $\mathrm{K}_{\mathrm{AUSF}}$. The other input is $\mathrm{SN}_{\mathrm{name}}$. The construction of the input message to the KDF function is specified in 3GPP TS 33.501 [57, Annex A.6].
   - A hash digest HXRES^* of $\mathrm{RAND}||\mathrm{XRES}^*$ is also computed using the hash function SHA-256, as specified in 3GPP TS 33.501 [57, Annex A.5].

(2) The HN forwards $\mathrm{RAND}||\mathrm{AUTN}||\mathrm{HXRES}^*$ to the SN.

[8] The KDF function is explained in Section 4.1.

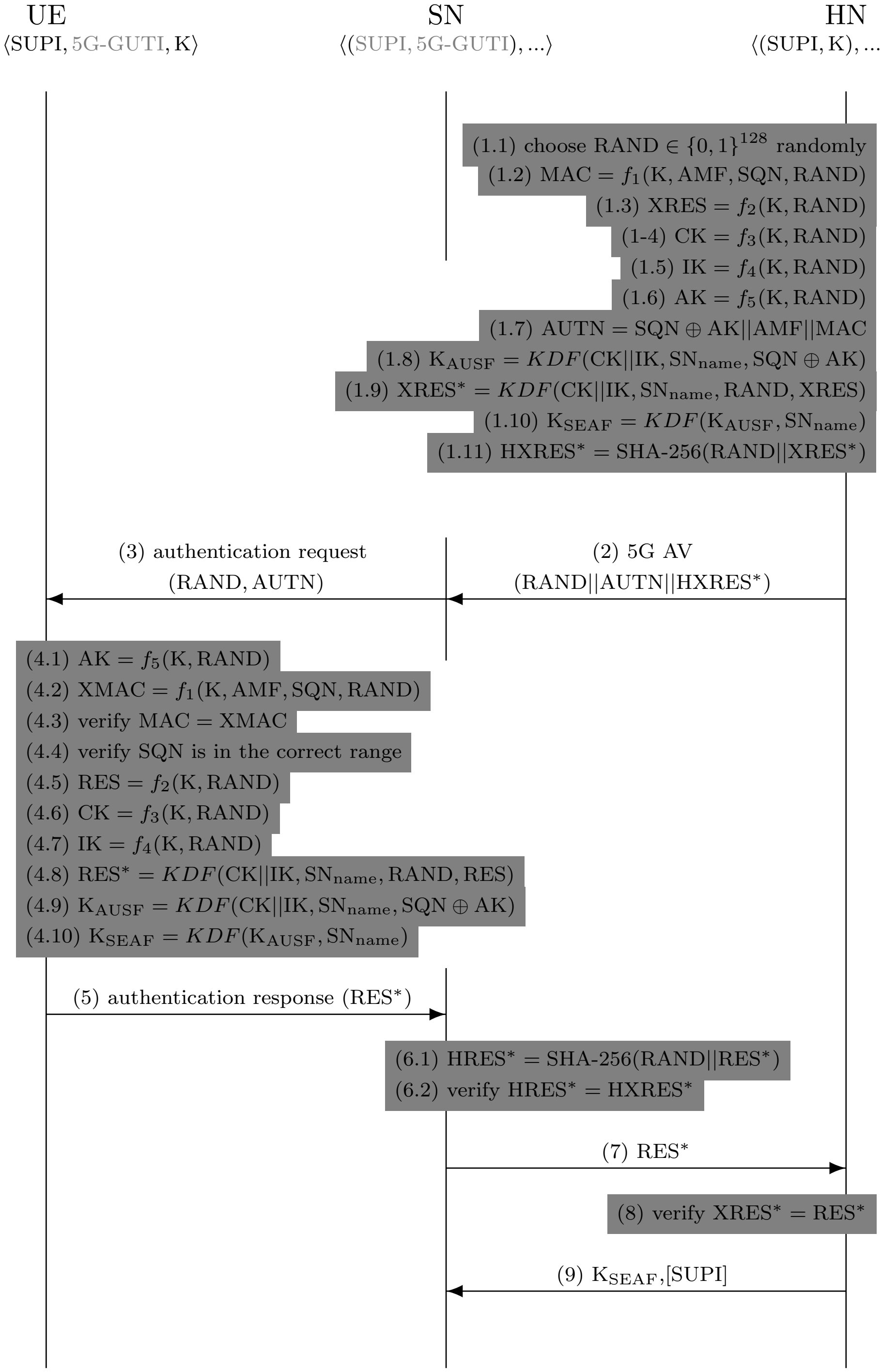
UE
⟨SUPI, 5G-GUTI, K⟩
SN
⟨(SUPI, 5G-GUTI), ...⟩
HN
⟨(SUPI, K), ...⟩
(1.1) choose RAND $\in \{0,1\}^{128}$ randomly
(1.2) MAC $= f_1(\mathrm{K}, \mathrm{AMF}, \mathrm{SQN}, \mathrm{RAND})$
(1.3) XRES $= f_2(\mathrm{K}, \mathrm{RAND})$
(1-4) CK $= f_3(\mathrm{K}, \mathrm{RAND})$
(1.5) IK $= f_4(\mathrm{K}, \mathrm{RAND})$
(1.6) AK $= f_5(\mathrm{K}, \mathrm{RAND})$
(1.7) AUTN $= \mathrm{SQN} \oplus \mathrm{AK}||\mathrm{AMF}||\mathrm{MAC}$
(1.8) $\mathrm{K}_{\mathrm{AUSF}} = KDF(\mathrm{CK}||\mathrm{IK}, \mathrm{SN}_{\mathrm{name}}, \mathrm{SQN} \oplus \mathrm{AK})$
(1.9) $\mathrm{XRES}^* = KDF(\mathrm{CK}||\mathrm{IK}, \mathrm{SN}_{\mathrm{name}}, \mathrm{RAND}, \mathrm{XRES})$
(1.10) $\mathrm{K}_{\mathrm{SEAF}} = KDF(\mathrm{K}_{\mathrm{AUSF}}, \mathrm{SN}_{\mathrm{name}})$
(1.11) $\mathrm{HXRES}^* = \text{SHA-256}(\mathrm{RAND}||\mathrm{XRES}^*)$
(3) authentication request
(RAND, AUTN)
(2) 5G AV
$(\mathrm{RAND}||\mathrm{AUTN}||\mathrm{HXRES}^*)$
(4.1) AK $= f_5(\mathrm{K}, \mathrm{RAND})$
(4.2) XMAC $= f_1(\mathrm{K}, \mathrm{AMF}, \mathrm{SQN}, \mathrm{RAND})$
(4.3) verify MAC = XMAC
(4.4) verify SQN is in the correct range
(4.5) RES $= f_2(\mathrm{K}, \mathrm{RAND})$
(4.6) CK $= f_3(\mathrm{K}, \mathrm{RAND})$
(4.7) IK $= f_4(\mathrm{K}, \mathrm{RAND})$
(4.8) $\mathrm{RES}^* = KDF(\mathrm{CK}||\mathrm{IK}, \mathrm{SN}_{\mathrm{name}}, \mathrm{RAND}, \mathrm{RES})$
(4.9) $\mathrm{K}_{\mathrm{AUSF}} = KDF(\mathrm{CK}||\mathrm{IK}, \mathrm{SN}_{\mathrm{name}}, \mathrm{SQN} \oplus \mathrm{AK})$
(4.10) $\mathrm{K}_{\mathrm{SEAF}} = KDF(\mathrm{K}_{\mathrm{AUSF}}, \mathrm{SN}_{\mathrm{name}})$
(5) authentication response (RES^*)
(6.1) $\mathrm{HRES}^* = \text{SHA-256}(\mathrm{RAND}||\mathrm{RES}^*)$
(6.2) verify $\mathrm{HRES}^* = \mathrm{HXRES}^*$
(7) RES^*
(8) verify $\mathrm{XRES}^* = \mathrm{RES}^*$
(9) $\mathrm{K}_{\mathrm{SEAF}}$,[SUPI]

Figure 4.3: 5G AKA.

(3) The SN forwards RAND||AUTN to the UE.

(4) The UE forwards the RAND||AUTN to the USIM. The USIM performs a series of computation similar to those in the HN, i.e., the USIM generates $\mathrm{AK}, \mathrm{XMAC}, \mathrm{CK}, \mathrm{IK}, \mathrm{RES}$ and verifies MAC, and SQN in the same way as described in Section 4.1. The USIM forwards CK, IK, and RES to the ME. The ME computes $\mathrm{RES}^*, \mathrm{K}_{\mathrm{AUSF}}$, and $\mathrm{K}_{\mathrm{SEAF}}$, in the same way as $\mathrm{XRES}^*, \mathrm{K}_{\mathrm{AUSF}}, \mathrm{K}_{\mathrm{SEAF}}$ are generated in the HN.

(5) The ME forwards RES^* to the UE.

(6) The SN computes HRES^* in the same way the HN computed HXRES^*. Then the SN verifies whether HRES^* and HXRES^* are equal or not. If they are equal, the authentication is considered to be successful from the SN point of view. If not equal, authentication is unsuccessful from SN point of view. However, in either case, the SN continues from Step 7 to Step 9. In case the authentication is unsuccessful, the SN sends a message to the UE as explained below after Step 9.

(7) The SN forwards RES^* to the HN.

(8) The HN verifies whether XRES^* and RES^* are equal or not. If they are equal, then HN considers that the authentication was successful between the SN and the UE from the HN point of view.

(9) The HN responds to the SN in an authentication response message whether the authentication was successful from the HN point of view or not. If the authentication was successful, the HN includes the anchor key $\mathrm{K}_{\mathrm{SEAF}}$ in the authentication response. The HN may optionally include the SUPI in the authentication response if the SN sent a SUCI in the authentication request message, as shown in Step 8 of Figure 4.2.

If the HN, in Step 9, has indicated that the authentication was not successful in the HN, or if the authentication was not successful in the SN in Step 7, then the SN shall do the following. The SN shall reject the authentication by sending an authentication-reject message to the UE if the SUCI was used by the UE in the previous identification procedure. If the UE did not use the SUCI in the previous identification procedure, then the only alternative identity that the UE could have used was 5G-GUTI. The SN shall initiate an identification procedure with the UE if the UE used 5G-GUTI in the previous identification procedure.

Once the UE responds to the SN with a SUCI in the identification procedure, the SN may initiate an additional authentication procedure.

After successful and affirmative completion of the AKA, the SN uses the key K_{SEAF}, and the SUPI to derive a key named K_{AMF} according to 3GPP TS 33.501 [57, Annex A.7]. The key K_{AMF} is used to derive further keys, which are used in providing ciphering and integrity protection for various layers of communication. If the UE used a SUCI during identification, then the SN would provide communication services to the UE only after the HN sent SUPI to the SN [57, Clause 6.1.3.2].

To indicate the selected algorithms and to start ciphering and integrity protection, two Security Mode Command (SMC) procedures are used [57]. Some cryptographic data protection already starts during the run of SMC procedures. Other cryptographic protection starts after the SMC procedures complete. The reason for not starting all protection at the same time is that sometimes the cryptographic keys needed for the protection can already be determined during the SMC run-time, but sometimes the keys cannot be determined before the SMC procedure is complete. The SMC procedures are explained in 3GPP TS 33.501 [57].

Increased Home Control

In UMTS or LTE, an SN can send a Location Update (LU) message to an HN; please see Step 13 in Figure 4.1. The LU message includes the IMSI of a subscriber. According to the specification [101, 102], the SN should send an LU message in some defined situations, for example, when the user uses IMSI to connect to the SN and the subsequent authentication becomes successful. However, a malicious SN may send a fake LU message after an unsuccessful AKA run. The rationale of such malicious behaviour is discussed in Paper II. If the SN sends a fake location update message, the HN cannot detect the malicious behaviour. This is because the location update message does not include any cryptographic measure that can enable such detection at the HN end.

In 5G, the use of $XRES^*$ and RES^* enables the HN to be convinced that the correct UE has indeed run the AKA with the SN and that the authentication result is positive. This is because the right UE is the only party besides the HN that can compute RES^*. Therefore, if the SN returns the right RES^* in Step 7, then it would imply that the correct UE participated in the AKA. At this stage, the HN would also update the location of the UE. However, if the RES^* returned in Step 7 is wrong, then the HN would notice it in Step 8 and consequently, would not update the UE's location or return the anchor key $\mathrm{K}_{\mathrm{SEAF}}$ in Step 9.

Chapter 5

Defending Against IMSI Catchers

The threat of IMSI catchers, as explained in Subsection 3.2.1, is not just a theoretical possibility. They are real entities that have become a commonplace in large cities [112] and airports [113]. Besides, IMSI catchers have increasingly become cheaper and easier to use for an adversary [60].

5.1 Related Work

A significant volume of work has been dedicated in the last three decades mitigating the threat of IMSI catchers. Most of the efforts on mitigating the threat of IMSI catchers can be put in two categories: (i) detecting and (ii) defeating. However, not all efforts fall into one of these two categories. An example of such an effort is the legal prohibition of using IMSI catchers. This thesis focuses on defeating techniques. In the following, we briefly refer to the detecting techniques and then point to a relatively large number of studies on defeating techniques.

5.1.1 Techniques for Detecting IMSI Catchers

Techniques in this category mainly operate by gathering information about a cellular network, analysing the gathered information for anomalies and warning users/operators about the presence of an IMSI catcher [10]. Park et al. [10] mentioned a list of such techniques. The techniques involve additional measures on top of the standard cellular network protocols, e.g., using applications installed in the user device. Once warned, the user abstains from communicating with the detected IMSI catcher.

Although detection of IMSI catchers typically prevents severe attacks such as downgrading attacks, they can still be used to violate identity privacy. This is because, by the time a user can detect an IMSI catcher, the IMSI catcher may already have tricked the user revealing the IMSI. Also, the detectors are not completely accurate. They may produce false positive and false negative detection [8–10].

5.1.2 Techniques for Defeating IMSI Catchers

Techniques that fall under the category of defeating IMSI catchers typically involve cryptographic measures so that IMSI catchers can neither trick users into revealing their IMSI nor deduce the presence of a specific target user. Typically, these techniques are transparent to the user, i.e., they do not require any extra measures on top of running the standard cellular network protocols.

The IMSI catchers would be completely defeated when all vulnerabilities that IMSI catchers can exploit are fixed. A few of these vulnerabilities are mentioned in Subsection 3.2.1, and a summary of these vulnerabilities is presented by Khan and Martin [92]. However, this thesis focuses on the most easily exploitable vulnerability (in the identification procedure) and keeps the other vulnerabilities out of the primary focus.

Most of the efforts towards defeating IMSI catchers have been on fixing the vulnerability, as explained in Subsection 3.2.1, in the identification procedure. We identified trends of solutions based on the following three mechanisms: (i) group-key, (ii) pseudonyms, and (iii) public-key cryptography. In Paper III, a list of such works is presented. However, for the sake of fewer interruptions in reading this thesis, in the following, we reuse some of the text from Section 3 of Paper III.

Solutions Based on Group-key In this mechanism, a group of users use the same symmetric key to encrypt the response of the “user identification request”. The UE sends the resulting chiphertext along with the identity of the group key, MCC, and MNC in plaintext so that the SN can send the chiphertext and the identity of the group key to the correct HN. The HN uses the identity of the group key for resolving the key and decrypts the ciphertext.

If the number of users in a group is k, then a user in the group is expected to achieve k-anonymity. So, a large k may appear to be desirable. However, if a group is too large, then the symmetric key would be shared by many users, and anyone belonging to the group can decrypt the encrypted response. Ginzboorg and Niemi [4] have explained the group-key-based solution, its challenges, and why it was eventually not adopted although 3GPP has seriously considered it for a long time for 3G already.

Solutions Based on Pseudonyms This kind of solutions uses users' temporary identities, known as pseudonyms. A user uses its pseudonyms to identify itself so that no one else but the HN can resolve a long-term identity of the user who uses the pseudonym. The detailed workings of pseudonyms are discussed in Subsection 5.4.1.

In a proposal by Herzberg et al. [114], one-time pseudonyms are created by probabilistic encryption of the user's identity using a key that is known only by the HN. The HN sends a set of pseudonyms to the UE when the UE and the HN have a secure communication channel.

Several publications [4, 60–62] propose using pseudonyms in the same format as IMSI, but with randomised and frequently changing MSIN.[1] In solutions proposed by Broek et al. [60], and Khan and Mitchell [62], the pseudonym's update is embedded in a random challenge, RAND of AKA. Khan and Mitchell [103] identified a weakness in solutions proposing cellular network pseudonyms in the same format as IMSI, which could be exploited to de-synchronize pseudonyms in the UE and the HN. Khan and Mitchell [103] also proposed a fix.

In Paper II, we discuss how the weakness could be exploited to mount a Distributed Denial-of-Service (DDoS) attack against the HN. Paper II has also found a weakness in the fix that is proposed by Khan and Mitchell [103]. In Paper II, we have proposed a different fix that is simpler than that by Khan and Mitchell [103].

Solutions Based on Public-key Cryptography Asokan [115] described how public key encryption could be used to achieve identity privacy in mobile environments. Arapinis et al. [71, Figure 8] also proposed similar public-key-based solution. In these solutions, only the HN has a public/private key pair, and the UE is provisioned with the public key of the HN. The UE encrypts identity information using the public key of the HN before sending it to the HN. Køien [116] suggests using Identity-Based Encryption (IBE) to defeat IMSI catchers in LTE.

[1] An IMSI has three parts: MCC, MNC, and MSIN. These parts are explained in 2.2.4.

5.2 Outline of Our Contributions

Our contributions to mitigate the threat of IMSI catchers fall in the category of defeating IMSI catchers. The threat of IMSI catchers is mitigated in the standards of the first release of 5G. The 5G standards have adopted the Elliptic Curve Integrated Encryption Scheme (ECIES), which is a public-key cryptosystem, towards defeating the IMSI catchers (see Subsection 4.2.2). Nevertheless, Paper I includes a complementary solution that could be used in certain use cases.

This thesis also includes solutions to defeat IMSI catchers in 3G and LTE.[2] Defeating IMSI catchers in the legacy networks involves a huge additional challenge on top of the challenges researchers faced while developing a new solution to defeat IMSI catchers in 5G. This challenge is due to the existing widespread deployment of LTE and 3G networks. If an IMSI-catcher-immune 3G/LTE user sends a privacy-protected message to a 3G/LTE SN that is not yet patched to handle such a message, the successive authentication may not succeed at all even if both the UE and SN were legitimate. In Paper II, we discuss a solution based on pseudonyms, which requires only a relatively small effort to patch the existing 3G/LTE SNs.

Even though IMSI catchers are defeated in 5G, it is important to notice that a 5G user still remains vulnerable to IMSI catchers when the user uses LTE, 3G, or GSM networks. In Paper III, we discuss a mechanism that can defeat IMSI catchers in 5G, as well as in all legacy networks. This mechanism is a hybrid of the solution presented in Paper II and the 3GPP-adopted solution to defeat IMSI catchers in 5G.

5.3 Defending against IMSI Catchers in 5G

This section is based on Paper I. In 5G, the most easily exploitable vulnerability (the one in the identification procedure) and some of the other vulnerabilities that IMSI catchers exploit are fixed. Khan and Martin [92] have summarised a list of these fixes in their survey paper on subscription privacy in 5G. However, IMSI catchers may still exploit other vulnerabilities that are not yet fixed in 5G [92]. These left-out vulnerabilities are related to (i) the authentication failure messages sent by the UE to the SN [71], and (ii) paging messages triggered by sending

[2] We do not delve into defeating IMSI catchers in 2G because an attacker can mount even more serious attacks on a 2G user by taking a man-in-the-middle position[11]. The major weakness in 2G is that the user does not authenticate the network.

(preferably hidden) SMS-like content to the victim [72]. These vulnerabilities are arguably less harmful than the vulnerability in the identification procedure. This is because an IMSI catcher has to know some identity of the target victim before exploiting these vulnerabilities and can mount the attack on one victim at a time. In contrast, to exploit the vulnerability in the identification procedure, the IMSI catcher does not need to know any identity of a victim beforehand and can mount the attack on all the UEs present in its radio coverage. This thesis keeps the arguably less harmful vulnerabilities out of scope and focuses on the vulnerability in the identification procedure.

During the standardisation of the first phase of 5G, different solutions to fix the vulnerability in the identification procedure were proposed. These proposals were made both in the academic literature (as discussed in Subsection 5.1.2) and in the standardisation study document [21].

In Paper I, we briefly outlined principles of eight different candidate solutions based on different cryptographic techniques to fix the vulnerability in the identification procedure. The techniques include the use of pseudonyms (temporary identity assigned to a user by the HN) and different public-key cryptography schemes, e.g., certificate-based or identity-based.

In Paper I, we have discussed three variants of the certificate-based technique: Variant 1, Variant 2 and Variant 3. These variants differ from each other in the way the Certificate Authorities (CAs) constitute the Public Key Infrastructure (PKI).

Variant 1 is similar to the contemporary certificate-based solution used in Transport Layer Security (TLS). The SN presents a digital certificate to the UE, which the UE verifies. If the verification is successful, the UE uses the public key of the SN to encrypt the IMSI and sends the encrypted IMSI to the SN. For certificate verification, a root CA is provisioned to the UE beforehand.

In Variant 2, a UE considers only its HN as the root CA, which means the chain of signatures in the certificate presented by the SN has to include a signature signed by the HN. Otherwise, the certificate verification would not succeed.

In Variant 3, only the HN can act as a CA; this means a certificate would have a chain of only one signature, which is signed by the HN. Therefore, in Variant 3, it suffices if the HN provisions the UE with the set of trusted SNs with their respective public-keys.

In all of the three variants, the SN can decrypt the encrypted response (to “identity request”) sent by the UE. Therefore, the SN can know the identity of the user without the assistance of the HN.

We also have discussed one public-key-based solution (we call it root-key-based solution), which is in principle similar to those proposed by Arapinis et al. [71], and Asokan [115]. The root-key-based solution does not need any PKI at all. Please note that the root-key-based solution is in principle the same as the SUCI-based solution (see Subsection 4.2.1) adopted by 3GPP to defeat the vulnerability in 5G.

In Paper I, we have briefly discussed two IBE-based schemes: (i) JPL, which is named after the initials of its authors, and (ii) Privacy-Enhanced Mutual Authentication protocol (PEMMA). In the following, we explain the principle of our IBE-based solution. In Paper I, we have presented this solution in greater detail compared to the other candidate solutions (for which only the core principles are stated). In this section, we present a qualitative comparative analysis (in greater detail than in Paper I) between the aforementioned candidates.

5.3.1 IBE-based Solution

One of the two main contributions in Paper I is presenting a new solution to fix the vulnerability in the identification procedure using Identity-Based Encryption (IBE). The fix in the identification procedure is extended into a mutual authentication protocol that we call the Privacy-Enhanced Fast Mutual Authentication (PEFMA) protocol, which can co-exist along with the native AKA protocol. We sketched out the co-existence mechanism using a flow diagram. The protocol is faster than AKA when the HN is located far. This is because PEFMA does not require fetching an AV from the HN each time it runs, and consequently, saves the time of a round-trip from the SN to HN. The other main contribution of Paper I is a qualitative comparison between all the candidate solutions based on thirteen measuring criteria.

IBE [83] is a public-key cryptographic technique. In the following, we present an overview of the workings of a general IBE system. Such an overview is also presented with a figure in Section 4.2 of Paper I. Then, we introduce the principle of PEFMA and present a summary of the qualitative comparison between the candidate solutions.

How IBE Works

To use IBE in a communication system, a third party, which is known as a Private Key Generator (PKG), has to be trusted by all the participants of the communication system. A sender computes a receiver's public key as a function of the identity of the receiver and the public key of the PKG. So, the PKG has to securely provide the sender with the PKG's public key before the sender sends an encrypted message to the receiver. The PKG computes the receiver's private key as a function of the identity of the receiver and the private key of the PKG. The receiver obtains its private key from the PKG over a secure channel, which implies that an IBE system does not require a Public Key Infrastructure (PKI). Indeed, the sender does not need to authenticate the public key of the receiver (PKG authenticates the receiver). Therefore, no CA is required.

A fair amount of trust is required to be put in the PKG. This is because unlike CAs in the certificate-based public-key cryptography, the PKG knows all the private keys used in the system.

In a public-key-based cryptographic protocol, changing a key is trickier than in symmetric-key-based ones. When a symmetric key is compromised, establishing a new symmetric key is usually sufficient to defeat the attacker who possesses the compromised key. However, generating a new public/private key pair is not sufficient to defeat an attacker who possesses the compromised private key – the attacker can still impersonate the receiver whose private key has been compromised. The compromised key must be revoked so that a sender can verify whether the public key presented by the receiver is revoked or not.

In IBE, since the keys of a receiver are functions of its identity, revoking the key of the receiver needs some extra care compared to that needed in the certificate-based system. This is because if revoking the key of a receiver requires revoking the identity of the receiver, then for an operational system, key revocations may cause too many interruptions to function smoothly. Usually, this difficulty of key revocation in IBE is resolved by concatenating a Private Validation Token (PVT) to the identity of the receiver in computing the keys of the receiver. When a revocation is required, instead of revoking the whole identity, only the PVT is revoked. This allows a new PVT to be assigned to the receiver.

PEFMA

We call our solution towards defeating IMSI catchers in 5G PEFMA, which is based on IBE. In PEFMA, the HN takes the role of the PKG. The HN chooses an Expiry Timestamp (ET) as the PVT for the UE and computes a private key for the UE using the IMSI, ET, and the private key of the HN. Then the HN securely provisions the UE with the HN's public key, ET, and the UE's private key. The HN also provisions the UE with a list of all the trusted SNs, marking the SNs which are PEFMA-enabled. When a roaming agreement is signed between the

HN and an SN, the HN chooses an ET for the SN and provision the SN with the HN's public key, ET, and the SN's private key.

These keys, ETs, and list of SNs need to be re-provisioned on different occasions, e.g., when ET is due, or when an associated key is compromised/revoked. In some cases, even though provisioning is needed, it cannot be done. For example, provisioning is not allowed when a UE is roaming. To enable the use of a 3G/LTE USIM in 5G, the provisioning needs to be done to the ME instead of the USIM. In Section 4.4 of Paper I, we have presented an outline of when these provisionings are needed, and when they can be done.

When an SN sends the "identity request" (written as "IMSI inquiry" in Paper I), the UE computes the public key of the SN using the identity of the SN (and an ET that is specified by the HN) and the public key of the HN. Then the UE encrypts the intended response (to the identity request), using the SN's public key. The response includes IMSI and two ETs – one for UE (encrypted) and one for SN (in plaintext). The UE also sends the MCC, MNC, and public-key identity of the HN in plaintext so that the SN can query the correct HN to obtain the SN's private key (if it does not have it yet).

Once the SN obtains its private key, it decrypts the encrypted response that the UE sent. The SN is now equipped with required materials (IMSI and ET of the user, and the public key of the HN) to compute the public key of the UE. Thus, neither the public key nor the identity of the UE is sent in cleartext over the radio channel. Therefore, the vulnerability in the identification procedure is fixed.

At this point, both the UE and SN know their respective private keys and each other's public keys. The UE and SN, in principle, can use these keys in different ways to exchange signatures or encrypted messages or both for achieving mutual authentication and establishing symmetric cryptographic keys.

In PEFMA, the UE and SN ensure authenticated and confidentiality protected exchanges of two random strings RAND1 (chosen by UE) and RAND2 (chosen by SN). They sign RAND1 and RAND2 using their respective private keys. They verify the signatures using each other's public keys. The UE and SN can use a Key Derivation Function (KDF) using the two random strings as parameters to compute different symmetric keys. Thus, a privacy-enhanced mutual authentication protocol is obtained. Please look into Paper I for details.

The primary motivation for running PEFMA is to fix the vulnerability in the identification procedure. In our solution, it is the UE that initiates PEFMA. So, when the SN obtains the identity of the UE, the SN is free to choose to discontinue running the PEFMA protocol and instead may choose to run the native AKA protocol. In some cases, the SN may prefer running AKA to PEFMA.

The SN's preference to AKA could be due to the relatively low computation cost (because of symmetric-key operations) and communication cost (because of shorter ciphertext and message authentication code instead of digital signature) incurred by AKA. Nevertheless, the SN could choose to run PEFMA when obtaining an AV is expensive due to, e.g., the HN being located at a far distance. Another reason could be that the HN has too big a subscriber base to compute an AV with reasonable latency each time a user needs to be authenticated. Such a big subscriber base is plausible when the HN supports smart factories.

In PEFMA, the UE, SN, and HN each have their own pairs of public/private keys. Compromising these keys have different levels of security and privacy impacts. Different revocation strategies are needed depending on the impacts of compromising a key. For example, when the key of an SN is compromised, revoking the key does not help because the sender, i.e., the UE, cannot look up the revocation list before accepting the key. Therefore, the ETs used for SNs need to be kept short-lived. In this way, an attacker who possesses the private key of the SN can mount an attack only as long as the ET is not expired. In Subsections 4.4.1 and 4.5.2 of Paper I, we have discussed the key revocation and impact of compromised/lost keys in greater detail.

5.3.2 Comparative Evaluation

In Paper I, we have discussed principles of eight candidate solutions to defend against IMSI catchers. We have presented a comparative analysis between these candidate solutions based on thirteen measuring criteria (see Table 5.1). This table is copied from Table 1 of Paper I. Scores given to candidate solutions on particular measuring criteria are qualitatively estimated based on 3GPP requirements and established views of the underlying cryptographic techniques of the candidate solutions.

In Paper I, we have not explicitly discussed the essence of the measuring criteria. The scores given to candidate solutions were based on the state-of-the-art in 2017. The justifications of the given scores are not explicitly articulated in Paper I. In 2018, a re-interpretation of the Lawful Interception (LI) requirement caused some changes in the given scores, as discussed by Khan et al. [111]. So, towards a clearer understanding of the comparative analysis as of 2017, we make the discussion explicit in the following. The scores are marked by (i) "++" meaning very good, (ii) "+" meaning good, and (iii) "−" meaning not

Table 5.1: Comparative Evaluation (as presented in Paper I).

Criteria	**Pseudonym**	**Public-key-based Solutions**						
		Certificate-based			**Root-key**	**IBE-based**		
		V1	**V2**	**V3**		**JPL**	**PEMMA**	**PEFMA**
Immunity to AIC	+	+	+	+	+	+	+	+
Concealing *hnid*	-	+	-	-	-	+	-	-
Lawful interception	+	+	+	+	-	-	+	+
Mutual Authentication	-	+	+	+	-	+	-	+
Signaling overhead	+ +	-	-	-	+	+	+	+
Computational overhead	+	-	-	-	+	+	+	+
Latency while roaming	-	-	-	-	-	+	-	+
Latency while at home	+ +	-	-	-	+	+	+	+
PKI effort	+ +	-	+	+	+	+	+	+
Key revocation	++	-	-	-	+	-	-	-
Provisioning effort	+	+	+	-	+	+	+	+
Using existing gear	+	-	+	+	+	-	+	+
Maturity	-	+	+	-	+	-	-	-

desirable but acceptable. Every solution is at least acceptable for every criterion. Therefore, we do not use the mark "−−" at all.

Immunity to Active IMSI Catcher (AIC)

This measuring criterion is about how effective a solution is in fixing the vulnerability in the identification procedure. It is also the most important criterion because the vulnerability in the identification procedure is the most easily exploitable vulnerability by an active IMSI catcher, and the whole effort is towards defeating the active IMSI catchers. All the eight candidate solutions fix the vulnerability in the identification procedure, i.e., an active IMSI catcher cannot trick the UE to reveal its IMSI by impersonating a legitimate SN and sending "identity request". Hence, none of the solutions is given the score "−".

Concealing Home Network Identity (*hnid*)

This measuring criterion is about whether the solutions can conceal the MCC and MNC parts of the SUPI or not. The SN needs to know MCC and MNC parts of the IMSI of a user so that it can route identification-related messages to the right HN [111]. To ensure that the SN can learn this information correctly, most of the eight candidate solutions require sending MCC and MNC in plaintext with the exception of Version 1 of the certificate-based solution and JPL. Consequently, these two solutions are given the score "+" and the rest are given "−".

Effort Needed for Lawful Interception

This measuring criterion is about how much effort is required in the SN and HN to fulfil the requirements related to Lawful Interception (LI). The most relevant LI requirement is that the SN should be able to identify a user with the IMSI/SUPI without relying on the HN. In the root-key-based solution and JPL, the SN has to rely on the HN to identify a user with the IMSI – HN has to send the IMSI to the SN. This is because these two schemes encrypt the MSIN part of the IMSI using a key such that only the HN can decrypt the ciphertext.

In the 5G standards, the SN uses SUPI, which is sent by the HN, to compute the key K_{AMF} (see Subsection 4.2.2). The UE has to compute the key K_{AMF} in the same way as the SN to get services from the SN. Indeed, otherwise, the UE and SN would have different K_{AMF}, and therefore, the Security Mode Command Procedures (SMC) will fail. The failure of SMC procedures would imply the termination of the communication establishment attempt. Therefore, if the SMC procedures succeed, the SN can be sure that the HN has sent the same SUPI as the UE is using, i.e., the SN does not need to simply trust that the HN would not lie about the SUPI.

In all the other public-key-based solutions, the SN does not need the HN's assistance for resolving the IMSI. This is because the SN itself can decrypt the encrypted MSIN. In Paper I, we assumed that a pseudonym would be considered as a valid IMSI for the purposes of LI since it looks like an IMSI.[3] In light of the above discussion, the root-key-based solution and JPL are given the score "−" and the rest are given the score "+".

Mutual Authentication Without Contacting the HN Each Time

This measuring criterion is about whether it is possible to run mutual authentication between the UE and the SN without the assistance of the HN each time the authentication runs. Such a mutual authentication may be useful in certain use cases in which obtaining AV from the HN is expensive. The pseudonym-based solution, root-key-based solution, and IBE-based protocol PEMMA cannot be used for such mutual authentication. This is because, in these cases, the SN cannot resolve the IMSI without consulting with the HN. Therefore these three solutions are given the score "−".

In contrast, in other solutions, once the right keys are provisioned to the UE and SN, a mutual authentication as discussed above can be achieved. Therefore, these solutions are given the score "+".

Signalling Overhead

This measuring criterion is about how much signalling between the UE and the network is required by the solutions. The pseudonym-based solution does not use public-key encryption. The use of only symmetric key encryption produces shorter ciphertext, which keeps the signalling overhead significantly low. So, the score given to the pseudonym-based solution is "++".

Certificate-based solutions require an extra round trip between the UE and the SN to exchange the certificate, which creates signalling overhead. Besides, certificates can be quite long. So, the score given to all the three versions of certificate-based solutions is "−".

[3] Later, we no more considered that pseudonyms would substitute IMSI for the purposes of LI. Therefore, the pseudonym-based solutions (Paper II and Paper III) too needed the HN's assistance to resolve the IMSI.

Neither the root-key-based nor IBE-based solutions need to exchange certificates, which is a significant relief. However, these solutions produce encrypted IMSI, which is quite a lot longer than pseudonyms. So, compared to the pseudonym-based solution, we give a lower score to the root-key-based solution, which is "+".

Computational Overhead

This measuring criterion is about how much computation is required by the solutions in the UE and the network. Pseudonym-based solutions require some extra computation in the HN to generate the next pseudonyms of a user. Since the computation is symmetric-key-based, the overhead is small. However, the solution also requires other systems, e.g., the billing system, to maintain a history of pseudonym-IMSI mapping. So, the score given is "+".

Certificate-based solutions require verifying certificates and compute public-key encryption and decryption. This creates additional computational overhead. So, the score given to all the three versions of the certificate-based solutions is "−".

Both root-key and IBE-based solutions require public-key encryption and decryption but do not require verifying certificates. So, the score given to them is "+'.

Authentication Latency While Roaming

This measuring criterion is applicable when a user is roaming. The criterion is about the network delay in authenticating the roaming user.

In the pseudonym-based solution, when the user is roaming, in order to resolve the IMSI, the pseudonym has to travel all the way to the HN, which increases latency. So we rate it with "−".

Certificate-based solutions require an extra round trip between the UE and SN to exchange and verify the certificates. These solutions compute public key encryption and have messages of longer length. All these affect the latency. So we rate the certificate-based solution with "−".

In the root-key-based solution and in PEMMA, when the user is roaming, in order to resolve the IMSI, the encrypted IMSI has to travel a relatively long distance, i.e., up to the HN, which increases latency. So we evaluate it with "−". The solutions in JPL and PEFMA do not require any extra round trips or certificates. The encrypted IMSI does not need to travel up to the HN. So we rate them with "+".

Authentication Latency While at Home

This measuring criterion is about the network delay in authenticating a user while the user is not roaming. The latency due to the distance between the SN and HN disappears when the user is not roaming. Moreover, due to symmetric-key operations, pseudonym-based solutions are relatively fast. Therefore, we rate the pseudonym-based solution with "++".

On the other hand, due to the added computational and communication overhead in certificate-based solutions, we give them the score "–". The rest of the solutions are given a better score than the certificate-based solution and worse than the pseudonym-based solution, i.e., "+".

Required PKI Effort

This measuring criterion is about how much effort is needed in setting up and maintaining a PKI. Except for Version 1 of the certificate-based solution, none other needs a PKI. So, Version 1 of the certificate-based solution is given the score "–".

In the pseudonym-based solution, no public key is involved. Consequently, it has no need for a PKI. Therefore, we give it the score "++".

In other certificate-based solutions, only the HN can be a root CA. Therefore, a full-scale PKI would not be needed because the HN itself can provide the services (signing, revocation, etc.) expected from the PKI. Also, in the root-key-based and IBE-based solutions, the HN is the root of trust. The public key of the HN is provisioned to the UE. Therefore, a PKI is not needed. All these solutions are given the score "+".

Required Effort in Key Revocation

This measuring criterion is about how much effort is required by the network and UE to ensure that an old or compromised key does not continue to be used. The pseudonym-based solution is symmetric-key-based. So, it does not require any key revocation. Therefore, we rate it with "++".

In certificate-based solutions, to access the revocation list, a user has to connect to an SN. It is difficult to design a feasible solution so that before connecting to the SN, the UE can know whether the public key of the SN is revoked or not.

On the other hand, in IBE-based solutions, revocation is inherently complicated (as explained before in this chapter). One solution around the complication is to use short-lived ET for the public keys of the SN. So we rate certificate-based and IBE-based solutions with "–".

In root-key-based solution, no revocation is required because there is only one public key. However, if the corresponding private key is compromised, the UE has to be re-provisioned with the new public key. So, we rate the root-key-based solution with "+".

Required Provisioning Effort

This measuring criterion is about how much effort in the HN is required by the solutions to provision the UE and SNs with necessary information. In the pseudonym-based solution, the UE needs to be provisioned with initial pseudonyms. If pseudonyms get de-synchronised for some reason (discussed in Paper II), the users may need to physically visit a shop of the HN to be re-provisioned with fresh pseudonyms.

In the certificate-based solutions, the UE needs to be provisioned with a root of trust, i.e., the public key and its certificate of a trusted entity. In these solutions, also the SNs need to be provisioned with the certificates of their public keys.

So, we give both the pseudonym-based and certificate-based solutions the same score: "+". However, in Version 3 of certificate-based solutions, UEs need to be additionally provisioned with the certificates of all the trusted SNs. Due to this additional need, Version 3 is given the score "–".

In the root-key-based solution, the UE needs to be provisioned with the public key of the HN. In IBE-based solutions, both the UE and SN need to be provisioned with their respective public/private key pairs and the public key of the HN. However, these solutions do not require to provision the UE with the public keys of the SN. So, they are given the score "+".

Opportunity of Using Existing Gear

This measuring criterion is about the needs for additional infrastructure on top of the existing. Six of the eight solutions do not need any additional gears. Version 1 of certificate-based solutions and JPL need additional gears. Therefore we rate Version 1 and JPL with "–", and the rest of the solutions with "+".

Maturity of the Underlying Cryptographic Techniques

This measuring criterion is about the maturity of the cryptographic techniques used in a solution. We look at the maturity of the cryptographic concepts and the success stories of using those concepts in solving real-life problems.

Use of pseudonyms (assigned by the HN) for privacy purposes was not yet a matured technology. So we rate the pseudonym-based solution with "–".

Certificate-based public-key encryption technology is widespread and matured. Use of root-key can be viewed as a special case of a certificate-based public-key cryptosystem. So, we rate both certificate-based and root-key-based solutions with "+".

However, Version 3 of the certificate-based solution is an unusual case of certificate-based solutions and not yet studied carefully. Therefore, the score given to it is"–".

Use of IBE is not yet widespread. So we rate the IBE-based solutions too with "–".

Discussion

If concealing MCC and MNC is essential, then the only applicable solution is Version 1 of the certificate-based solution, i.e., the solution with a full-scale PKI. If revealing MCC and MNC is found to be acceptable in benefit-cost analysis, then the choice of the solution depends on the importance of obtaining mutual authentication between the UE and SN without the assistance of the HN. If such mutual authentication is conceived to be necessary, then PEFMA is the most suitable candidate. One potential use case of such necessity is the smart factory where numerous devices are connected to the network. If such mutual authentication is not necessary, the root-key-based or the pseudonym-based solution will suffice.

Please note that while roaming, if a user moves its USIM from one ME to another, then the user may not be able to use PEFMA in the new ME. This is because the PEFMA-related keys are provisioned to the ME, not to the USIM. If the USIM is moved to another ME, then the new ME either would not work at all or would lose protection against IMSI catchers that exploit the vulnerability in the identification procedure.[4] A future avenue of research would be to solve this usability issue.

The SUCI-based solution adopted by 3GPP, as discussed above, is in principle the same as the root-key-based solution. A pseudonym-based solution was not adopted by 3GPP, perhaps because of the additional effort needed to maintain the mapping between pseudonyms and IMSI at the HN.

5.4 Fixing Vulnerability in the Identification Procedures of 3G and LTE

This section is based on Paper II. One main contribution of this paper is showing that a pseudonym synchronisation issue [103] of the proposed pseudonym-based solutions [4, 60, 62, 103] can be exploited by attackers to cause fatal harm to an HN on a large scale. The other main contribution is providing a solution to mitigate the weakness. Our solution fixes the vulnerability in the identification procedures of 3G and LTE. The basic principle of our solution could be used to fix the vulnerability in 5G, had the vulnerability not already been fixed in 5G.

Pseudonym-based solutions to defeat IMSI catchers have been published in the recent past [4, 60, 62, 103]. The most promising aspect of these solutions is that they are transparent to the SN. Therefore, if an HN implements a pseudonym-based solution, then its users already can go roaming to an SN that has not yet implemented the solution, and the users remain protected against IMSI catchers while roaming.

The basic idea of a pseudonym-based solution is to use frequently-changing temporary identities called pseudonyms that look like IMSI, instead of using actual IMSI. A pseudonym looks like an IMSI when the format of the IMSI is preserved in the pseudonym. More specifically, a pseudonym looks like an IMSI when (i) the MCC and MNC parts of the IMSI remain unchanged (ii) the length and alphabet of the MSIN part remain unchanged. Khan et al. [111] has explained that the MCC and MNC need to remain fixed in order to be able to route the IMSI to the correct HN.

[4] If the new ME is PEFMA aware and provisioned with the public key of an HN that is different from the HN of the USIM, then the authentication would not succeed. If the new ME is not PEFMA aware, then it may send IMSI in plaintext to the SN and consequently loses protection against IMSI catchers.

5.4.1 How Pseudonyms Work

In pseudonym-based solutions [4, 60, 62, 103], the UE would never send the IMSI in response to the identity request. Instead, the UE would send a pseudonym. Pseudonyms are different from TMSI or (5G) GUTI. The most important difference is that pseudonyms are assigned to a UE by the HN, but TMSI and (5G) GUTI are assigned by the SN. Also, pseudonyms change less frequently than TMSI and (5G) GUTI. A UE is provisioned with pseudonyms in the beginning. The UE uses these pseudonyms in response to the very first “identity request” that it receives from an SN. The SN forwards the pseudonym (that the UE sent) to the HN.

New pseudonyms that the UE uses in response to the successive identity requests can be allocated in different ways. For example, the UE could generate new pseudonyms based on rules previously provisioned to the UE by the HN. Alternatively, the UE and HN could generate identical pseudonyms independently as a bi-product of running AKAs, for example, by hashing the IMSI and a salt using the shared permanent key. The third example is that the HN sends a new pseudonym (confidentiality and integrity protected) to the UE (via the SN) by piggybacking extra information on the existing messages of the AKA protocol. In our solutions, we have adapted this last example, which we explain more in the following.

Using cryptographic techniques, the HN embeds a new pseudonym into the random challenge RAND in the process of preparing an authentication vector AV. In 3GPP standards, the length of the RAND is 128 bits. Embedding of pseudonyms in RAND is done in such a way that the entropy of RAND does not decrease too much. When the AV arrives, the UE first verifies the authenticity of the AV. If the verification result is positive, i.e., the AV is authentic, then the UE can perform the inverse cryptographic operation that the HN used to embed a pseudonym into the RAND. Thus the UE can extract the pseudonym. The next time, the UE uses the new pseudonym in response to an “identity request”.

Once the UE has received a new pseudonym, an IMSI catcher will get a different response to the subsequent “identity request”. Therefore, the UE appears as a new UE to the IMSI catcher, causing the IMSI catching attack to fail. At the end of Section 4 of Paper III, we have presented a summary of different alternative mechanisms of allocating pseudonyms.

One challenge of using a pseudonym-based solution is to inform the SN the true IMSI of the subscriber that is linked to the used pseudonym. The need for informing the SN about the true IMSI of a user stems from a Lawful Interception (LI) requirement as discussed in Section 2.5.

Another complexity of the solution is that the pseudonym space is finite, and due to scalability reasons, the HN has to reuse pseudonyms. Therefore, both the UE and HN have to release old pseudonyms when new pseudonyms are assigned. The releasing of old pseudonyms has to be synchronised between the UE and the HN. Otherwise, it may happen that the HN has released a pseudonym from one user and assigned it to another user when the first user continues to use the pseudonym. In that situation, when the first user uses the pseudonym, the HN may mistake the first user for the second user.

If the HN incorrectly recognises every pseudonym that a UE sends, then UE cannot succeed in authenticating itself. Therefore, the UE remains locked out of the network. It is crucial to ensure that an attacker cannot trick the HN into the following de-synchronised state: the HN retains a set of pseudonyms for one of its subscribers that has no common elements with the set of the pseudonyms that the subscriber's UE has retained. For many pseudonym-based solutions existing in the literature [4, 60, 62], it is possible for a malicious UE to trick the HN into the de-synchronised state, as explained by Khan and Mitchell [103] and Paper II.

For billing purposes, the HN has to maintain a history of pseudonyms assigned to an IMSI over different time periods. Such maintenance of pseudonym history may also be dictated by the data retention policy of the overseeing jurisdiction.

5.4.2 Pseudonym De-synchronisation and a DDoS Attack

In 2017, Khan and Mitchell [103] first identified a weakness in the pseudonym-based solutions that had been published until then. The weakness can be exploited by a malicious UE just by trying to connect to an SN. All that the malicious UE has to do is send pseudonyms to an SN in an attempt to connect. We have explained the attack in Section 4 of Paper II. The attack succeeds mostly due to the fact that the HN releases an old pseudonym from one of its users before getting the confirmation that the user has received the new pseudonym.

In Paper II, we have argued that the attack can be converted into a DDoS attack. We have characterised the attack using a probabilistic argument (supported by simulation) and showed that a DDoS attack requiring only a modest amount of effort could do significant harm to an HN. For example, by sending three fake pseudonyms an hour from each bot, a botnet of one million bots can push out (of the network) approximately two per cent of the users of a large mobile operator (ten million subscribers), creating a massive strain on the customer service points of the operator.

Khan and Mitchell [103] proposed a solution by leveraging the Location Update (LU) message. The LU message is sent by an SN to the HN when a UE first connects to the SN (see Subsection 4.1.2). We name the solution "KM17 scheme". The solution uses six temporary identities (three pseudonyms and three recovery identities) at the HN. The solution is vulnerable to a pseudonym de-synchronisation attack by a malicious or faulty SN. The solution provides a re-synchronisation procedure. In case pseudonyms get de-synchronised due to an attack by a malicious or faulty SN, the re-synchronisation procedure can be used to get the pseudonyms synchronised.

In Section 3.2 of Paper II, we have argued that the recovery technique itself is vulnerable to IMSI catchers. In the same section we have explained, we also have explained that in the KM17 scheme, that a user has to use one pseudonym at least twice before it can receive a new pseudonym from the HN. In our solution, we fix these issues.

5.4.3 Our Solution

We propose a solution in Section 5 of Paper II that is simpler than the KM17 scheme. Our solution also leverages the LU message. Unlike the KM17 scheme, our solution does not include a re-synchronisation procedure, therefore, does not add any additional vulnerability. Also, in our solution, a user does not need to use a pseudonym more than once to receive a new pseudonym.

We have put a particular focus on analysing the synchronisation issue of pseudonyms. We have classified the states of pseudonyms in the HN and the UE. Then we have identified how one state of pseudonyms transforms into another state of pseudonyms. Based on these observations, we have sketched a state-diagram of the pseudonym states. Using the state diagram, we argue that the pseudonym de-synchronisation can not happen unless the SN sends wrong LU messages (see Sections 6.1 and 6.2 of Paper II). Thus, like the KM17 scheme, our solution remains vulnerable to the pseudonym de-synchronisation attack by a malicious or faulty SN.

However, unlike the KM17 scheme, our solution does not include any re-synchronisation procedure. Instead, we argue in Section 6.2 of Paper II that a malicious SN can be very quickly identified and be blocked from communicating with the HN before it can cause any serious damages to the HN and that the incentive to mount such an attack is not sufficient.

In Section 6.2 of Paper II, we have also discussed other properties of our solution. For example, we have presented a list of unusual but possible situations and discussed the behaviour of our solution in those situations. The discussion also includes how the solution is compatible with the legacy SNs and is immune to replay attacks (to de-synchronise pseudonyms) by malicious SNs. We also have discussed the implication of using pseudonyms in lawful interception processes in the HN. The HN would need to maintain a history of pseudonyms used by a particular user.

5.4.4 Discussion

In Paper II, we have presented an almost SN-transparent (small patching in the SN is required for LI purposes) solution to fix the vulnerability in the identification procedure. The solution is the first of its kind that is immune to pseudonym de-synchronisation, except that a malicious SN can cause pseudonym de-synchronisation by sending fake LU messages. However, currently, a malicious SN can mount other attacks (perhaps more fatal) by sending fake LU messages. Thus, 3G and LTE networks already put trust in SNs that they would not send fake LU messages. In that line of argument, it is acceptable that the pseudonym-based solution is mildly vulnerable to malicious SNs.

5.5 Defeating Downgrade Attack by IMSI Catchers Against 5G Users

This section is based on Paper III. The main contribution of Paper III is combining the SUCI-based 5G mechanism of defeating IMSI catchers with the pseudonym-based mechanism (as presented in Paper II) of defeating them in LTE networks. The outcome is a hybrid solution that defeats IMSI catchers that attack a 5G user by downgrading to the LTE network and exploiting the vulnerability in the identification procedure of the LTE network. In Paper III, we discuss the hybridisation process of 5G AKA and LTE AKA. Nevertheless, similar hybridisation can be done using 5G AKA and UMTS AKA, too.

The hybrid solution, which is presented in Paper III, makes two improvements in the pseudonym-based solution, which is presented in Paper II. First, a malicious SN cannot attack anymore by sending fake LU messages to de-synchronise pseudonyms between a UE and the HN. However, pseudonyms may still go de-synchronised due to unlikely errors, e.g., corrupted memory, at the UE or HN end. The second improvement is that re-sychnronisation of pseudonyms happens automatically when the UE connects to a 5G SN.

Use of pseudonyms for a 5G user has an additional issue. A 5G UE may connect to multiple SNs simultaneously, causing the UE to use different pseudonyms to connect to different SNs. Hence, it is a challenge for the HN to know when it can disassociate a pseudonym from a current user and reallocate to a new user. In our solution, we solve this challenge by embedding pseudonym-related information in the SUCI.

5.5.1 Solution

Our solution uses a subscriber-specific strictly monotonically increasing counter d associated with a pseudonym, which is maintained by the HN (see Section 4 of Paper III). Along with a pseudonym, the HN also sends the associated counter to the UE using a similar technique as in Paper II. The UE may receive a new pseudonym by running the LTE AKA or 5G AKA. The UE and HN also maintain sets of pseudonyms in addition to those in Paper II. The UE deletes pseudonyms from its set of pseudonyms based on the policy provided by the HN that could utilise, e.g., the pseudonym lifetime or maximum size of the set.

The HN disassociates a pseudonym from a user only if it receives sufficient information about the in-use pseudonyms from the right UE. The required information about pseudonyms in the UE, along with integrity protection, is piggybacked on the SUCI. The UE embeds the counter of the newest and oldest in-use pseudonyms into the SUCI, see Section 4.2 and Algorithm 4 in Paper III. The HN does not disassociate a pseudonym from a user in any other situations. Thus, the LU messages from LTE networks do not have any impact on disassociating a pseudonym and in turn, do not have any impact on pseudonym de-synchronisation.

The above discussion implies that the set of pseudonyms for a user in the HN may reduce in size only when the user sends a SUCI. Thus, if a user keeps using LTE networks without connecting to 5G, the set of pseudonyms keeps growing in size. Such an unrestricted growth of the set would shrink the number of free pseudonyms and would threaten the exhaustion of all the free pseudonyms. The state where there are no free pseudonyms would cause the system not to provide the desired privacy any more. In Section 4 of Paper III, we have discussed a mechanism based on a cap on the size of the sets of pseudonyms in the HN. This mechanism prevents sets of pseudonyms from growing beyond an upper limit, e.g., 100 pseudonyms in one set. In the paper, we present an argument that shows that the HN can allow the average size of the sets of pseudonyms to be as low as a dozen.

Pseudonym de-synchronisation happens when the UE has not a single pseudonym at its disposal that is associated with the UE's IMSI at the HN. We have defined de-synchronisation more formally in Section 5.1 of Paper III. We also have argued that de-synchronisation of pseudonyms cannot happen as long as the HN and UE function correctly. However, in an unlikely error case, if the UE and HN end up in de-synchronised states of pseudonyms, re-synchronisation can be done by simply sending a SUCI to a 5G SN and participating in the subsequent 5G AKA protocol.

Based on the information about pseudonyms piggybacked on the SUCI, the HN can know if de-synchronisation has happened. If yes, the HN sends a signal to the UE by piggybacking on the RAND. Upon receiving the signal, the UE deletes all the existing pseudonyms and accepts the new pseudonym extracted from the RAND. The re-synchronisation process is explained in Section 5.1 of Paper III.

Our solution would need patching in the LTE SNs to support an LI requirement. The LI requirement is that the SNs need to be able to recognise a user using the IMSI of the user without relying on another network or jurisdiction, for example, the HN. In Section 5.2 of Paper III, we have outlined the design of the required patch.

A pseudonym would be reused for different users. The billing systems of the mobile operators would have to implement a pseudonym-to-IMSI mapping before charging a subscriber. The reuse of pseudonyms makes it challenging to implement a pseudonym-to-IMSI mapping that can be efficiently used by the billing system.

5.5.2 Discussion

In Paper III, we have presented an almost SN-transparent (small patching needed in the SN for LI purposes) solution that (i) protects a downgrade attack on the identity privacy of a 5G user, and (ii) unlike other pseudonym-based solutions, is immune to the pseudonym de-synchronisation attack by a malicious/faulty SN. If pseudonym de-synchronisation happens due to some error, e.g., corrupted memory, at the UE or the HN end, re-synchronisation happens automatically the next time the UE connects to a 5G SN.

Chapter 6

Privacy Preserving AKMA

This chapter is based on Paper IV that discusses AKMA — an authentication and key management system for applications. Specifications for the system AKMA have been published in 3GPP Release 16 in July 2020. One of the two main contributions of the paper is an analysis of AKMA requirements with a particular focus on finding new privacy requirements. The other main contribution is the development of a privacy mode of AKMA that can co-exist with a normal mode.

In AKMA, authentication of a user to an "application service provider", also known as AKMA Application Function (AApF) in the context of AKMA,[1] would be based upon the user's 3GPP credentials. For example, a gambling service provider can take the role of the AApF. The user's HN takes the role of the authentication server. The endpoint of the AKMA protocol in the HN is called AKMA Authentication Function (AAuF). The application providers would not require storing authentication relation data. The users would not need to remember passwords for different applications, as long as the user uses the mobile phone to connect.

Similar systems also exist in the older generations of mobile networks. For example, 3GPP published the Generic Bootstrap Architecture (GBA) in Release-6 in 2004, and Battery Efficient Security for Very Low Throughput Machine Type Communication Devices (BEST) in Release-15 in 2017. So, the concept of AKMA is not new but has evolved from GBA and BEST. Naturally, in 3GPP, the study on AKMA [69] is influenced by the standards of GBA and BEST in identifying

[1] The entity has different names in different 3GPP contexts, e.g., Network Application Function (NAF) or Enterprise Application Server (EAS).

the requirements that AKMA should fulfil and developing solutions to meet those requirements.

In Paper IV, we compare these three systems and look into their similarities and differences more closely. Such an investigation produces a birds-eye view of the envisioned AKMA. The birds-eye view gives more leverage in revealing requirements that might have been overlooked in the study of AKMA and provides a more precise objective for the solution design activities.

The comparative requirement analysis of GBA, BEST, and AKMA is a nontrivial task. There is no single document that enlists the requirements that GBA or BEST fulfils. This is because the requirement analysis is not done upfront. Requirements were made in parallel with developing the solutions. However, during our analysis in Paper IV, it was possible to sift out the essential characteristics of GBA and BEST by studying their specifications [67, 68]. We consider these characteristics to be the requirements of the systems. For AKMA, we study the potential security requirements from the AKMA study document [69].

The details of requirement analysis can be found in Paper IV. In this chapter, we do not represent the requirement analysis in detail. Instead, we only present a summary of the comparison.

6.1 Summary of AKMA Requirement Analysis

We have identified 22 different requirements. Paper IV summarises all these requirements in Table 1 and explains them in Section 3. Most of these requirements are common in all three systems: GBA, BEST, and AKMA. Five of the requirements are AKMA requirements, but they are not GBA or BEST requirements. Three of the requirements are either GBA or BEST requirements, but they are not yet identified as AKMA requirements (see Requirements 12, 13, and 14 in Table 1 of Paper IV). We anticipate the 3GPP community will discuss these requirements in the future.

We propose to put less trust in the HN than the 3GPP community has put in the AKMA study [69]. This is partly because of the rising trend of insider attacks [5–7]. A recent GSMA Intelligence report [117] has presented the result of a survey on consumers' concerns about privacy. The survey has found that despite the growing device sales, consumers are seriously concerned about privacy and the report also mentions that regulators are still working on devising new data regulation. The report suggests that beyond complying with regulation, tech companies that pre-emptively adopt the strictest measures for data privacy will benefit from a competitive advantage by meeting consumer expectations.

As a consequence of putting less trust in the HN, we have identified two privacy requirements which are not fulfilled by GBA or BEST (see Requirements 21 and 22 in Table 1 of Paper IV). We believe these two privacy requirements are worth considering for AKMA. The first one is "Privacy of usage history" and the second one is "Identity unlinkability". Paper IV discusses these two requirements in Section 3.1.

"Privacy of usage history" says that the AAuF, in the HN, should not know the name of the AApFs, i.e., the service providers, that a user connects to. Such a privacy requirement may be desired by a user, especially when the AApF provides privacy-sensitive services, e.g., a gambling or dating service.

"Identity unlinkability" says, two SPUIDs[2] (that may belong to the same or different AApF) should not be linkable by one or more AApFs. In simple words, given two of its users, an AApF cannot tell if they belong to a single mobile phone user. Also, given two users from two AApFs, the AApFs (even if they collude) cannot tell if those two users belong to a single mobile phone user. For example, if a mobile phone user uses a dating service using the SPUID *romeo* and a gambling service using the SPUID *smeagol* then the dating service and the gambling service (even when colluding) should not be able to figure out that the SPUIDs *romeo* and *smeagol* belong to the same mobile phone user.

Please note that the aforementioned privacy requirements are not new for delegated authentication systems. They have also been identified and discussed in the context OpenID and OpenID Connect [118, 119].

6.2 AKMA Solution

Section 4 of Paper IV presents our solution in detail. Our solution has two modes: (i) normal mode and (ii) privacy mode. The normal mode does not fulfil the privacy requirements, i.e., Requirements 21 and 22 in Table 1 of Paper IV. The privacy mode fulfils these privacy requirements and as many of the other requirements as possible. Our solution has two separate procedures: (i) AKMA bootstrapping and (ii) AKMA bootstrap usage. In both the normal and privacy modes, the AKMA bootstrapping procedure is the same. However, the "AKMA bootstrap usage" procedures are different in different modes. The UE chooses

[2] A SPUID is a user identity that is used by a user to identify himself to a specific service provider. SPUID is defined in Subsection 2.2.4

which mode it wants to use in the first message of the AKMA bootstrap usage procedure.

A UE first runs the AKMA bootstrapping procedure (see Figure 6 of Paper IV) and then the AKMA bootstrap usage procedure. The AKMA bootstrapping procedure is run between the UE and the AAuF. The bootstrapping procedure is designed based on the 5G AKA protocol. As a result of the bootstrapping procedure, the UE and the HN establish a key called K_{AKMA}. The HN also sends a temporary identity `TID` and a key lifetime to the UE. The UE uses the `TID` later when it runs an AKMA bootstrap usage procedure.

In the normal-mode AKMA bootstrap usage procedure (see Figure 7 of Paper IV), the UE sends the `TID` and the AAuF identity to the AApF so that the AApF can obtain a symmetric key (derived from K_{AKMA}) from the AAuF. During the normal-mode AKMA bootstrap usage procedure, if the AApF and AAuF communicate with each other, then the AAuF is able to mount a "correlation attack" (see Section 4.3 of Paper IV) against the "privacy of usage history". In the "correlation attack", the AAuF correlates the time of "AKMA bootstrapping" with the time of the "AKMA bootstrap usage" procedure. In our normal-mode "AKMA bootstrap usage" procedure (also in all the proposed solutions in the AKMA study [69]), the AApF and AAuF communicate with each other. Therefore, the normal-mode remains vulnerable to the "correlation attack".

6.2.1 Privacy-mode AKMA Bootstrap Usage

Our privacy-mode "AKMA bootstrap usage" procedure is designed in a way such that the AApF and AAuF do not communicate (see Figure 8 of Paper IV) directly. The AApF, with the assistance of the UE directly and AAuF indirectly, computes a one-way function that takes two secret constant inputs a, and b that are known only by the AAuF and the UE respectively. No one else but the AAuF knows a and no one else but the UE knows b. All the parties participate in a series of computations, initiated by the AApF who first generates a random number. In sequence, the UE and the AAuF transform the random number using secrets b and a with the means of one-way cryptographic functions. The result is eventually delivered back to the AApF by the UE. On the result, the AApF performs a cryptographic inverse operation related to the initial random number that the AApF chose. As an outcome of the inverse operation, the AApF obtains a value which is the output of a one-way function that takes inputs a, and b, i.e., secrets unknown to AApF but known only to the AAuF and UE respectively. We make some assumptions in our solution design:

1. the UE and AApF establish a TLS connection before running AKMA, based on a certificate provided by the AApF,
2. the AAuF has a public/private key pair and the AApF can somehow authenticate the public key of the AAuF,
3. the UE has a secret key that no one else (not even the AAuF or any other module in the HN) knows, and
4. the AAuF has a user-specific secret key that no one else (not even the UE) knows.

The UE first establishes a TLS connection with the AApF. As part of this procedure, the UE authenticates the AApF. The UE chooses its own SPUID and sends it to the AApF over the TLS connection. However, in other connection attempts that happen in future with the same SPUID, the AApF needs to authenticate the SPUID. To that end, the AApF sends a random challenge g^m to the UE, where g is a generator of a group of prime order, and m is a randomly chosen integer. No one but the AApF can know m due to the hardness of discrete logarithm problem.

The UE generates the non-ephemeral key b, which is also discussed earlier in this section, based on the SPUID, AApF id, and the secret key that no one else but the UE knows. The UE blinds the random challenge g^m using this non-ephemeral key b by computing $(g^m)^b$ and forwards it to the AAuF. The forwarding of g^{mb} is integrity and confidentiality protected using key $\mathrm{K}_{\mathrm{AKMAae}}$, which is generated during the "AKMA bootstrapping" procedure.

The AAuF generates the non-ephemeral key a based on the SUPI and the secret key that no one else but the AAuF knows. The AAuF blinds g^{mb}, which it received from the UE, using the non-ephemeral key a by computing $(g^{mb})^a$. The AAuF then signs g^{mba} using its private key and forwards g^{mba} and the signature to the UE. This forwarding is also confidentiality and integrity protected. The UE extracts g^{mba} and the signature by decrypting the message received from the AAuF. Then, the UE forwards g^{mba} and the signature to the AApF. The AApF verifies the signature and computes $(g^{mba})^{m^{-1}} = g^{ab}$.

If the SPUID appeared for the first time, the AApF stores (SPUID, $W_{\mathrm{SPUID}} = g^{ab}$) in its database and allows the SPUID to connect. However, if the SPUID is already present in the database, then the AApF compares g^{ab} with W_{SPUID}.

Amid above computations, optionally, the UE and AApF also perform an ephemeral Diffie-Hellman key exchange. At the end of the AKMA bootstrap usage procedure, the UE and the AApF possess a shared secret key. See Steps 5.1, 5.2, 7.5, and 11.6 in Figure 8 of Paper IV. Please note that the TLS handshake protocol itself facilitates an ephemeral Diffie-Hellman key exchange [120, 121]. Therefore, while establishing a TLS connection, the UE and AApF can exchange a secret key using TLS's native ephemeral Diffie-Hellman key exchange mechanism. In this Diffie-Hellman key exchange, the UE's contribution would not be signed because the UE does not have a public/private key pair. In contrast, in the ephemeral Diffie-Hellman key exchange of the AKMA bootstrap usage procedure, the UE's contribution is signed by the public key of the AAuF. The secret key established by the ephemeral Diffie-Hellman key exchange of the AKMA bootstrap usage procedure can be useful, for example, when the TLS connection expires and cannot be re-established for some reasons.

6.2.2 Security and Privacy analysis of AKMA Bootstrap Usage Procedure

In a nutshell, in the privacy-mode AKMA bootstrap usage procedure, the AApF challenges a UE with a random number of the form $g^{m'}$. The UE computes a response to the challenge with the assistance of the AAuF. Let us assume that in response to the AApF's random challenge $g^{m'}$, the UE returns a number of the form g^{α}, for some unknown α, to the AApF. The AApF computes $(g^{\alpha})^{m'^{-1}}$. If the SPUID that the UE used was not seen before by the AApF, then the AApF considers $(g^{\alpha})^{m'^{-1}}$ equal to g^{ab} and stores g^{ab}. If the SPUID that the UE used was seen before, then the AApF compares $(g^{\alpha})^{m'^{-1}}$ with the stored g^{ab}. If the AApF's computation $(g^{\alpha})^{m'^{-1}}$ becomes equal to g^{ab}, then it means that $g^{\alpha} = g^{m'ba}$.

Authentication-Related Security Claim

In AKMA, the secret b is chosen and kept protected by the UE so that no one else can know b. Hence, the party that knows b can be considered to be the authentic UE. Secret a is chosen and kept protected by the AAuF so that no one else can know a. Hence, the party that knows a can be considered to be the authentic AAuF.

Claim: When challenged with $g^{m'}$, a UE would be able to return $g^{\alpha} = g^{m'ba}$ to the AApF if and only if the UE had access to parties that knew b and a.

We present a proof of the authentication-related security claim in this thesis. To prove the claim, we provide a reduction from a variant of the Diffie-Hellman problem known as the Delayed-Target Diffie-Hellman problem (DTDHP) [13, 122].

DTDHP

Freeman [122] has first defined DTDHP though he called it One-More Computational Diffie Hellman Problem. Koblitz and Menezes [123] called the problem "Delayed Target" One-More Diffie-Hellman Problem (DTDHP). In this thesis, we simply call it Delayed Target Diffie-Hellman Problem (DTDHP). In the following, a definition of the DTDHP problem is presented. The DTDHP problem can also be seen in the form of the DTDHP Game presented in Figure 6.1.

Definition of DTDHP: Let G be a group of prime order. The solver is given $X \in G$, and a one-sided Diffie-Hellman oracle. The oracle works as follows: given an element $Y \in G$, the oracle answers $Z \in G$ such that $Z = g^{xy}$ where $X = g^x$ and $Y = g^y$. The solver must compute $Z' = g^{xy'}$ with input $X = g^x, Y^{'} = g^{y'}$ where $Y^{'}$ is a random group element that is given to the solver only after the solver has made all the queries to the oracle.

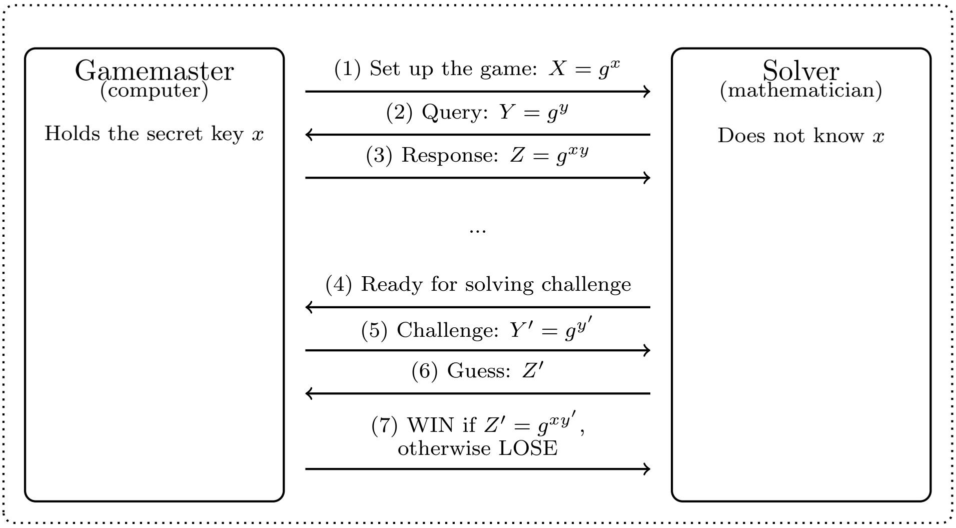
Gamemaster
(computer)
Holds the secret key x
Solver
(mathematician)
Does not know x
(1) Set up the game: $X = g^x$
(2) Query: $Y = g^y$
(3) Response: $Z = g^{xy}$
...
(4) Ready for solving challenge
(5) Challenge: $Y' = g^{y'}$
(6) Guess: Z'
(7) WIN if $Z' = g^{xy'}$,
otherwise LOSE

Figure 6.1: DTDHP Game.

Freeman [122] has presented an identification scheme and given its security proof based on the assumption that DTDHP is hard. Koblitz and Menezes [123] further analyzed how to interpret the difficulty of DTDHP. We also assume that DTDHP is hard and give proof of the authentication-related security claim.

Proof of the Authentication-Related Security Claim

It is straightforward to see that if the UE has access to parties that know b and a, then the UE is able to return g^{α} that is equal to $g^{m'ba}$. Conversely, we have to show that if it is feasible for the participating UE to return g^{α} that is equal to $g^{m'ba}$, then the UE has access to parties that know b and a. Let us assume that it is feasible for an adversary to return g^{α} that is equal to $g^{m'ba}$, but the adversary does not have access to parties that know b and a. In the following, we prove that such an adversary could be used to the benefit of a mathematician to win the DTDHP game. However, since winning the DTDHP game is hard, it is also hard for the adversary to return g^{α} that is equal to $g^{m'ba}$. Therefore, our claim holds under the assumption that DTDHP is hard. In our proof, we partition the set of all adversaries into three parts:

> **(1) Adversary does not know b but has access to somebody who knows a:** Let us consider a real-world AKMA adversary that does not know b but knows g^b and has access to somebody who knows a. The adversary impersonates the UE that is the legitimate owner of a SPUID. That is, the UE with secret b is the victim of the attack by the adversary. To succeed in the attack, when the AApF sends a challenge $Y = g^{m'}$, the adversary has to return $Z = g^{m'ab}$.
>
> We model such a real-world AKMA adversary as a Dolev-Yao adversary, and we assume that the adversary does not have access to any side-channel information such as time or power used for cryptographic computations by AKMA elements. Based on the AKMA protocol, we design a simulated AKMA game, called"AKMA Game I". The game is presented in Figure 6.2. The game is designed in a way so that the adversary can win if the adversary can be successful in the real-world AKMA attack as described above. Soon, we will show that if the adversary can win AKMA Game I, then by simulating AKMA Game I and using the adversary as an assistant, a mathematician can win the DTDHP Game. Before showing so, we first explain how AKMA Game I works.

In the AKMA Game I (Figure 6.2), the Gamemaster simulates the AKMA world. It knows all the secrets of the AKMA world and can play the role of all the honest parties of AKMA. The Gamemaster gives the adversary the secret a, but not b. The adversary is a Dolev-Yao adversary, and therefore, can be seen as a mailman who is responsible for carrying all the messages between different parties in the AKMA world. A message carried by the adversary to a party in the AKMA world is called a query. The response from a party in the AKMA world is first given to the adversary. The adversary may modify a received response from a party and send it as a query to another party in the AKMA world. Also, the adversary is allowed to forge and send new queries. In Figure 6.2, in Step 2 (and respectively in Step 5), these queries and responses are presented in a simplified manner using a single double-headed arrow between the Gamemaster and the adversary.

We further explain the queries and responses in AKMA Game I using a few examples. The adversary could consider the Gamemaster to be the AApF and try to connect to the AApF using a SPUID. To do this, the adversary would have to forge and send an AKMA message as a query to the Gamemaster. The AKMA message, which includes the SPUID, is described in Step 2 of Figure 8 in Paper IV. The Gamemaster simulates AApF and sends an AKMA message as a response to the adversary. This AKMA message is a random challenge g^m (see Step 4 in Figure 8 of Paper IV). Also, the adversary can forward the random challenge g^m as a query to the Gamemaster considering the Gamemaster to be the victim UE. The Gamemaster would simulate the victim UE and would return an AKMA message as a response to the adversary. Please see Step 6 in Figure 8 of Paper IV. This message includes g^{mb}. Similarly, the adversary could consider the Gamemaster to be the AAuF and send messages that the AAuF can process and respond to.

In a nutshell, the adversary can freely interact with the Gamemaster by sending and receiving AKMA messages as queries and responses. The AKMA messages that the adversary sends as queries can be exactly the same as received previously as responses from the Gamemaster or can be forged. The queries are allowed to be repetitive and can happen in an interleaved manner.

After the adversary has done enough interactions with the Gamemaster, the adversary informs the Gamemaster that the adversary is ready to go for an attack against the victim UE. Then, the Gamemaster would act as an AApF and sends an AKMA message to the adversary. The AKMA message is $Y' = g^{m'}$ for a randomly chosen m'. The message corresponds to Step 4 in Figure 8 of Paper IV. In AKMA Game I, we consider this message as a challenge to the adversary. After this, the adversary and the Gamemaster could do further interactions. However, at this point, the Gamemaster would not play the role of the victim UE anymore. Then, the adversary sends an AKMA message to the Gamemaster of the AKMA Game I. The AKMA message corresponds to Step 10 of Figure 8 in Paper IV. This message includes a response Z' to the challenge Y'. The adversary wins the AKMA Game I if $Z' = g^{m'ab}$. Otherwise, the adversary loses the game.

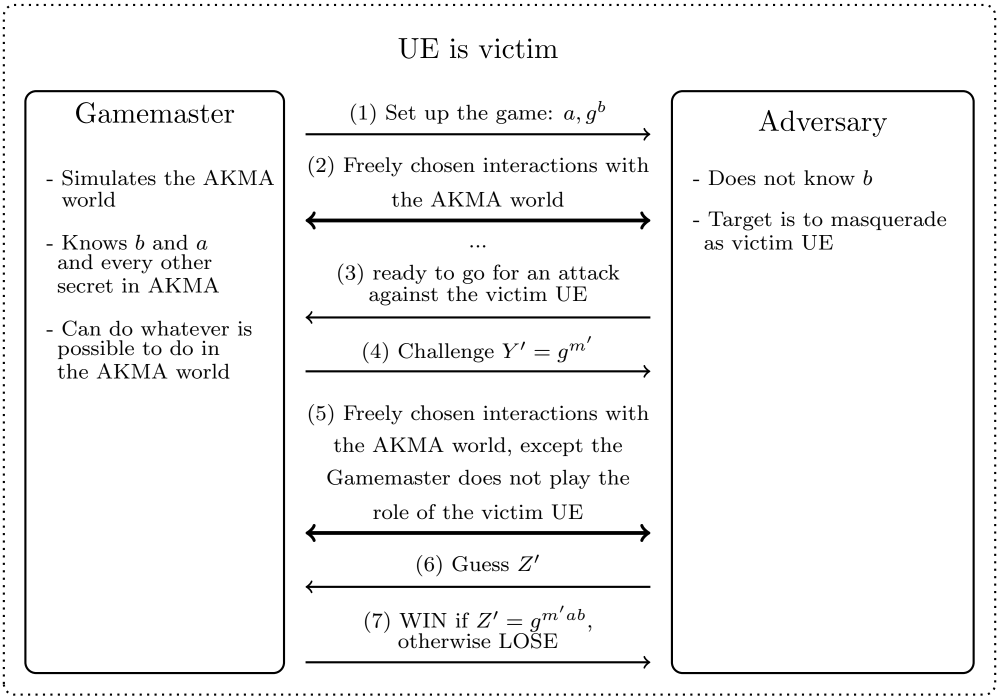
UE is victim
Gamemaster
- Simulates the AKMA world
- Knows b and a and every other secret in AKMA
- Can do whatever is possible to do in the AKMA world
(1) Set up the game: a, g^b
(2) Freely chosen interactions with the AKMA world
...
(3) ready to go for an attack against the victim UE
(4) Challenge $Y' = g^{m'}$
(5) Freely chosen interactions with the AKMA world, except the Gamemaster does not play the role of the victim UE
(6) Guess Z'
(7) WIN if $Z' = g^{m'ab}$, otherwise LOSE
Adversary
- Does not know b
- Target is to masquerade as victim UE

Figure 6.2: AKMA Game I.

A solver mathematician can win the DTDHP game by simulating AKMA Game I and taking help from the adversary, who can win the AKMA Game I. The simulation is presented in Figure 6.3. In the simulation, the DTDHP solver mathematician plays the role of the Gamemaster of AKMA game I without knowing b, which is the secret of the victim UE. In the simulation, "Adversary Assistant I" plays the role of the adversary in AKMA Game I. The Gamemaster of AKMA Game I performs the computations with input $b = x$ by leveraging the Gamemaster of the DTDHP game. Whenever the Adversary Assistant I sends a query g^m (Step 4 in Figure 8 of Paper IV), which is meant for the victim UE, the Gamemaster of AKMA Game I forwards the query to the Gamemaster of the DTDHP Game. The response (g^{mx}) of the Gamemaster in the DTDHP Game is included in an AKMA message (Step 6 in Figure 8 of Paper IV) by the Gamemaster of AKMA Game I and sent to the Adversary Assistant I. All the other queries that the Adversary Assistant I sends are responded to by the Gamemaster of AKMA Game I without the assistance of the Gamemaster of the DTDHP Game.

In Step 6 of Figure 6.3, the Adversary Assistant I declares that it is ready to go for an attack against the victim UE. In Step 7, the solver mathematician in the DTDHP Game declares that it is ready to solve a challenge. In Step 8, the Gamemaster of the DTDHP Game sends a challenge ($g^{y'}$) to the solver mathematician. In Step 9, the Gamemaster of AKMA Game I forwards the same challenge to Adversary Assistant I. This step corresponds to Step 4 in Figure 8 of Paper IV. To win AKMA Game I, in Step 11, the Adversary Assistant I must return an AKMA message, which corresponds to Step 10 of Figure 8 in Paper IV. The AKMA message must include a response $Z' = g^{y'ab}$. In Step 12, the solver mathematician forwards $(Z')^{a^{-1}}$ as a response to the original challenge in the DTDHP Game. It is straightforward to see that the Adversary Assistant I wins AKMA Game I if and only if the solver mathematician wins the DTDHP Game.

Since DTDHP is hard, the DTDHP solver mathematician cannot win the DTDHP game. Therefore, there cannot be any adversary who can win the AKMA Game I. Consequently, no adversary can exist against the real AKMA that does not know b and has access to somebody who knows a and can return $g^{m'ab}$ when challenged with $g^{m'}$.

(2) Adversary does not know a but has access to somebody who knows b: Let us consider a real-world AKMA adversary that does not know a but knows g^b and has access to somebody who knows b. The adversary impersonates the AAuF, which means the AAuF is the victim. To succeed in the attack, when the AApF sends a challenge $Y = g^{m'}$, the adversary has to return $Z = g^{m'ab}$.

Similarly as before, we model such a real-world AKMA adversary as a Dolev-Yao adversary, and we assume that the adversary does not have access to any side-channel information. Based on the AKMA protocol, we design a simulated AKMA game, called "AKMA Game II". The game is designed in a way so that the adversary can win if the adversary can be successful in the real-world AKMA attack as mentioned above. Soon, we will show that if the adversary can win AKMA Game II, then by simulating AKMA Game II, a mathematician can win the DTDHP Game. AKMA Game II is presented in Figure 6.4. AKMA Game II is similar to AKMA Game I except that the adversary does not know a and knows b.

A solver mathematician can win the DTDHP game by simulating AKMA Game II and taking help from the adversary, who can win AKMA Game II. The simulation is presented in Figure 6.5. The simulation of AKMA Game II is similar to the simulation of AKMA Game I. Nevertheless, for better readability, we describe it fully and not just the difference to the simulation of AKMA Game I.

In the simulation, the DTDHP solver mathematician plays the role of the Gamemaster of AKMA game II without knowing a, which is the secret of the victim AAuF. In the simulation, "Adversary Assistant II" plays the role of the adversary in AKMA Game II. The Gamemaster of AKMA Game II performs the computations with input $a = x$ by leveraging the Gamemaster of the DTDHP game. Whenever the Adversary Assistant II sends a query that includes g^{mb} (Step 6 in Figure 8 of Paper IV), which is meant for the victim AAuF, the Gamemaster of AKMA Game II forwards the query to the Gamemaster of the DTDHP Game. The response (g^{mbx}) of the Gamemaster in the DTDHP Game is included in an AKMA message (Step 8 in Figure 8 of Paper IV) by the Gamemaster of AKMA Game II and sent to the Adversary Assistant II. All the other queries that the Adversary Assistant II sends are responded to by the Gamemaster of AKMA Game II without the assistance of the Gamemaster of the DTDHP Game.

In Step 6 of Figure 6.5, the Adversary Assistant II declares that it is ready to go for an attack against the victim AAuF. In Step 7, the solver mathematician in the DTDHP Game declares that it is ready to solve a challenge. In Step 8, the Gamemaster of the DTDHP Game sends a challenge ($g^{y'}$) to the solver mathematician. In Step 9, the Gamemaster of AKMA Game II forwards the same challenge to Adversary Assistant II. This step corresponds to Step 4 in Figure 8 of Paper IV. To win AKMA Game II, in

Step 11, the Adversary Assistant II must return an AKMA message, which corresponds to Step 10 in Figure 8 of Paper IV. The AKMA message must include a response $Z' = g^{y'ab}$. In Step 12, the solver mathematician forwards $(Z')^{b^{-1}}$ as a response to the original challenge in the DTDHP Game. It is straightforward to see that the Adversary Assistant II wins AKMA Game II if and only if the solver mathematician wins the DTDHP Game.

Since DTDHP is hard, the DTDHP solver mathematician cannot win the DTDHP game. Therefore, there cannot be any adversary who can win AKMA Game II. Consequently, no adversary can exist against the real AKMA that does not know a and has access to somebody who knows b and can return $g^{m'ab}$ when challenged with $g^{m'}$.

(3) Adversary knows neither a nor b: If the adversary knows neither a nor b then the adversary can be modelled by either of the two other adversaries discussed previously in this section. This is because anything such an adversary can do can also be done by an adversary in AKMA Game I and also in AKMA Game II.

"Privacy of Usage" Requirement Fulfilled

The "Privacy of Usage" requirement is fulfilled due to the following reasons:

1. The AApF and AAuF do not interact.
2. A user does not reveal AApF-specific information (e.g., the identity of the AApF, SPUID) to the AAuF.
3. The AAuF cannot link the content of messages exchanged between the UE and AApF with the content of the messages exchanged between the UE and AAuF.

The reasons are explained in Section 5.3 of Paper IV in more detail.

"Identity Unlinkability" Requirement Fulfilled

The "Identity Unlinkability" requirement is fulfilled due to the following reasons:

1. The SPUID is not associated with any alternative identity/artefact known by the AApF; if two SPUIDs are different, everything else (e.g., g^{mb}, g^{mba}, and g^{ab}) in the protocol associated with those SPUIDs are different.

2. The AApF cannot link the content of messages exchanged between the UE and AApF with the content of the messages exchanged between the UE and AAuF.

These reasons are explained in Section 5.3 of Paper IV in more detail.

6.3 Discussion

Both the normal-mode and privacy-mode AKMA can fulfil most of the AKMA requirements that are listed in Paper IV. However, a few of the requirements are fulfilled only by the normal-mode but not the privacy-mode AKMA. Also, a few requirements can be fulfilled only by the privacy-mode but not by the normal-mode AKMA. Neither of the modes can fulfil Requirement 19: "Pushed bootstrapping". This is because new modes of AKMA are required to fulfil this requirement. There are some requirements which neither our normal-mode nor privacy-mode AKMA can fulfill, but separate procedures can be embedded in our normal-mode and/or privacy-mode AKMA to fulfil those requirements. We have discussed which requirements can be fulfilled by which mode at length in Section 6 of Paper IV.

Having a solution for privacy-mode AKMA, mobile network operators can offer premium service to privacy-cautious consumers. Our scheme could be extended in the future by, e.g., including re-keying of secret keys that are known only by the UE or by the AAuF. Future work could also include formal security analysis, implementation, and standardisation of our scheme.

6.3.1 Current Status of AKMA Study

Since the writing of Paper IV, the AKMA study [69] has evolved further. Version 16.0.0 of the AKMA study, which came out in December 2019, presents the evaluation of proposed solutions and a list of recommendations (as conclusions of the study) for devising a normative AKMA solution. The full list of recommendations and conclusions can be found in Clause 6 of the AKMA study [69]. Following the recommendations, in July 2020, the first version of AKMA specification was published in 3GPP TS 33.535 [124]. Please note that in AKMA specification, corresponding entities of AAuF, AApF, and TID are known by different names.

At the time of writing this, the AKMA study, 3GPP TR 33.835 [69], has not introduced any new key issues (i.e., no new requirements are added) in addition to those considered in Paper IV. The study recommended using implicit bootstrapping (Solutions 15, 19, and 24 in 3GPP TR 33.835 [69]) as the basis of normative work. The AKMA specification in TS 33.535 [124] is created based on implicit bootstrapping. The use of implicit bootstrapping has addressed two of the AKMA requirements (Requirements 13 and 14 in Table 1 of Paper IV) that we identified as missing in the AKMA study. Requirement 12, which we identified as missing in Table 1 of Paper IV, is also fulfilled due to the use of the Network Exposure Function (NEF) of 5G.

The resulting AKMA, as specified in TS 33.535, does not address the two additional privacy requirements we have identified in Paper IV, i.e., Requirements 21 and 22 in Table 1. This is because, in AKMA specification, the AApF and the AAuF directly communicate. Consequently, the AAuF knows which AApF the user uses, i.e., requirement 21 in Table 1 of Paper IV is not fulfilled. According to the specifications, the `TID` may be reused across multiple AApFs. Therefore, the requirement of identity unlinkability (Requirement 22 in Table 1 of Paper IV) would not be fulfilled. The specifications also do not give any guidelines about what other parameters, along with application key and expiry time, the AAuF is allowed to send to the AApF. The requirement of identity unlinkability may also not be fulfilled if the AAuF sends linkable information of one user to two AApFs.

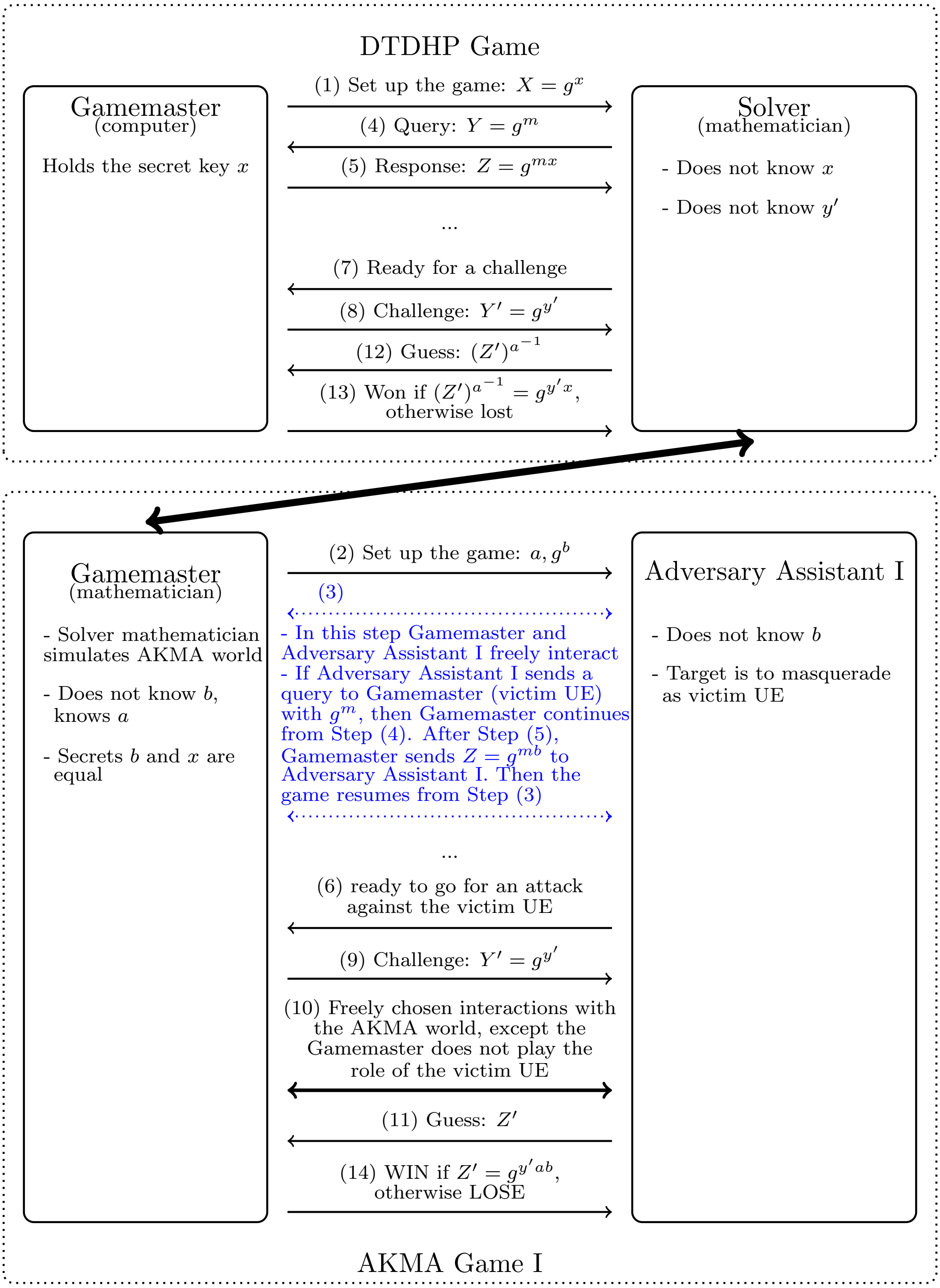
DTDHP Game
Gamemaster
(computer)
Holds the secret key x
(1) Set up the game: $X = g^x$
(4) Query: $Y = g^m$
(5) Response: $Z = g^{mx}$
...
(7) Ready for a challenge
(8) Challenge: $Y' = g^{y'}$
(12) Guess: $(Z')^{a^{-1}}$
(13) Won if $(Z')^{a^{-1}} = g^{y'x}$, otherwise lost
Solver
(mathematician)
- Does not know x
- Does not know y'
Gamemaster
(mathematician)
- Solver mathematician simulates AKMA world
- Does not know b, knows a
- Secrets b and x are equal
(2) Set up the game: a, g^b
(3)
- In this step Gamemaster and Adversary Assistant I freely interact
- If Adversary Assistant I sends a query to Gamemaster (victim UE) with g^m, then Gamemaster continues from Step (4). After Step (5), Gamemaster sends $Z = g^{mb}$ to Adversary Assistant I. Then the game resumes from Step (3)
...
(6) ready to go for an attack against the victim UE
(9) Challenge: $Y' = g^{y'}$
(10) Freely chosen interactions with the AKMA world, except the Gamemaster does not play the role of the victim UE
(11) Guess: Z'
(14) WIN if $Z' = g^{y'ab}$, otherwise LOSE
Adversary Assistant I
- Does not know b
- Target is to masquerade as victim UE
AKMA Game I

Figure 6.3: Winning DTDHP Game by Simulating AKMA Game I.

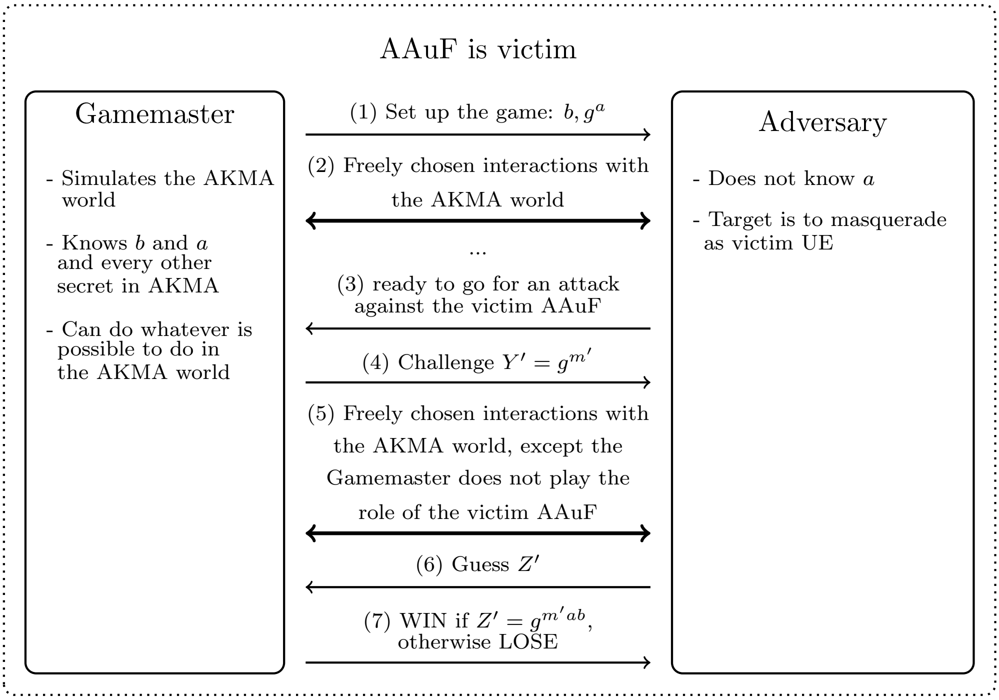
AAuF is victim
Gamemaster
- Simulates the AKMA world
- Knows b and a and every other secret in AKMA
- Can do whatever is possible to do in the AKMA world
(1) Set up the game: b, g^a
(2) Freely chosen interactions with the AKMA world
...
(3) ready to go for an attack against the victim AAuF
(4) Challenge $Y' = g^{m'}$
(5) Freely chosen interactions with the AKMA world, except the Gamemaster does not play the role of the victim AAuF
(6) Guess Z'
(7) WIN if $Z' = g^{m'ab}$, otherwise LOSE
Adversary
- Does not know a
- Target is to masquerade as victim UE

Figure 6.4: AKMA Game II.

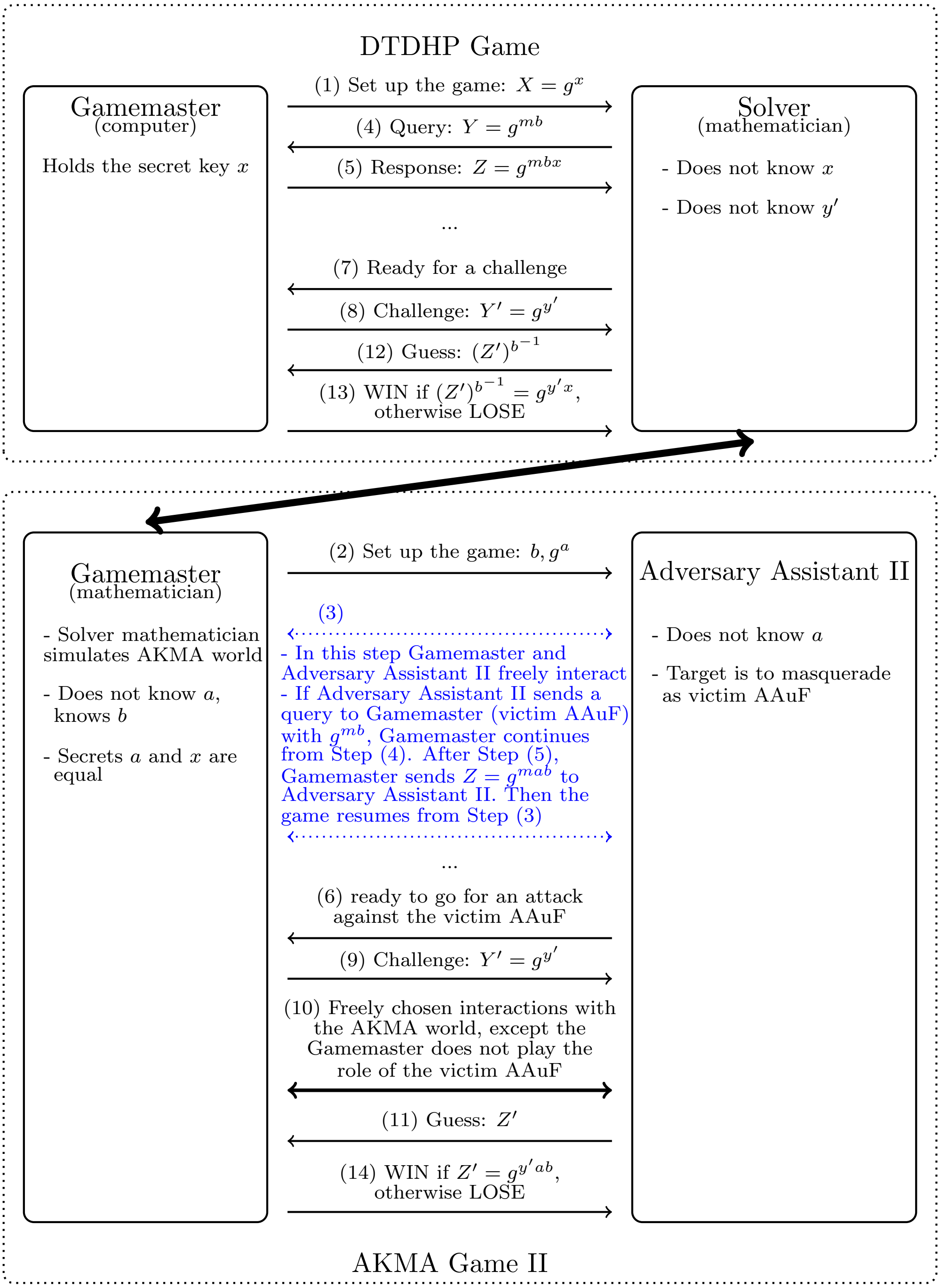
DTDHP Game
Gamemaster
(computer)
Holds the secret key x
(1) Set up the game: $X = g^x$
(4) Query: $Y = g^{mb}$
(5) Response: $Z = g^{mbx}$
...
(7) Ready for a challenge
(8) Challenge: $Y' = g^{y'}$
(12) Guess: $(Z')^{b^{-1}}$
(13) WIN if $(Z')^{b^{-1}} = g^{y'x}$, otherwise LOSE
Solver
(mathematician)
- Does not know x
- Does not know y'
Gamemaster
(mathematician)
- Solver mathematician simulates AKMA world
- Does not know a, knows b
- Secrets a and x are equal
(2) Set up the game: b, g^a
(3)
- In this step Gamemaster and Adversary Assistant II freely interact
- If Adversary Assistant II sends a query to Gamemaster (victim AAuF) with g^{mb}, Gamemaster continues from Step (4). After Step (5), Gamemaster sends $Z = g^{mab}$ to Adversary Assistant II. Then the game resumes from Step (3)
...
(6) ready to go for an attack against the victim AAuF
(9) Challenge: $Y' = g^{y'}$
(10) Freely chosen interactions with the AKMA world, except the Gamemaster does not play the role of the victim AAuF
(11) Guess: Z'
(14) WIN if $Z' = g^{y'ab}$, otherwise LOSE
Adversary Assistant II
- Does not know a
- Target is to masquerade as victim AAuF
AKMA Game II

Figure 6.5: Winning DTDHP Game by Simulating AKMA Game II.

Chapter 7

Conclusion

A user in a cellular network needs to be allowed to roam around with a wireless user device seamlessly and to be billed accurately for the services the user uses. These characteristics imply that the networks need to handle identities of a user in ways which are distinctive from other networks. For example, the network must know the long-term identity, e.g., IMSI or SUPI, before the network allows a user to use any services. Such requirements from the side of networks pose threats to the privacy of the user.

Since the early days of cellular networks, designers have been aware of many threats. Most of the known threats were mitigated since the second generation of cellular networks by leveraging the use of digital radio signal and modern cryptography. Also, throughout the evolution of cellular networks, some new threats have emerged. Most of these threats have been mitigated. However, some of the known threats were not mitigated due to a low benefit-cost ratio.

As a whole, the thesis contributes to the effort of making cellular networks more privacy-friendly. Our contributions can be put into two categories. The first category focuses on solving an old privacy problem – defeating IMSI catchers. We contribute by proposing a novel solution, fixing problems in existing proposals, and proposing a combination of existing proposals with a standardised solution. The second category focuses on privacy in a new sub-system of the 5G network. We contribute by pointing out privacy problems in the new sub-system. We also propose solutions to these problems.

More specifically, in the first category of our contributions, we have looked into one of the vulnerabilities related to the privacy of long-term identity, IMSI, of a cellular network user. The vulnerability lies in the procedures that user equipment uses to identify itself in GSM, 3G, and LTE – the user sends the long-term identity, IMSI, in plaintext. Adversaries, who are known as IMSI catchers, exploit this vulnerability in the identification procedures. Mechanisms towards fixing the vulnerability in the identification procedures have constituted a major part of this thesis – Paper I, Paper II, and Paper III.

In 5G standards, the identification procedure does not have the same vulnerability as the identification procedures in GSM, 3G, and LTE have. This is because the 5G standards have fixed the vulnerability. The fix uses public-key cryptography – the user equipment encrypts the long-term identity of the subscriber using the public key of the home network and sends the resulting ciphertext to the serving network. Paper I presents an alternative solution to fix the vulnerability using a different approach. The 5G standards were finalised after Paper I was published. However, the identification procedures of GSM, 3G, and LTE still have the vulnerability. Paper II and Paper III propose solutions to fix the vulnerability. Furthermore, we have identified a new privacy problem in a new sub-system, known as AKMA, of the 5G network. We have proposed a solution to solve the identified privacy problem in Paper IV.

In Paper I, we have presented a solution towards fixing the vulnerability in the identification procedure of 5G using identity-based encryption. The solution to defeating IMSI catchers is extended into a mutual authentication protocol between the user equipment and its serving network. The mutual authentication protocol is suitable in certain use cases, e.g., smart factories, of 5G. In Paper I, we have also discussed the situations in which the mutual authentication is preferable to the traditional AKA protocol. One usability issue of the solution is that a user cannot move its USIM from one mobile equipment to another while roaming. An avenue of future work would be to solve this usability issue.

Paper II improves pseudonym-based solutions [4, 60, 62, 103] towards defeating IMSI catchers in 3G and LTE networks. In pseudonym-based solutions that we improved, the home network frequently chooses new pseudonyms and sends to the user equipment. The user equipment uses pseudonyms (instead of IMSI) to identify itself to the serving network. The user equipment prefers using a pseudonym that was received later to a pseudonym that was received earlier. The pseudonym-based solution relies on symmetric-key cryptography, and the solution is transparent to the serving network, which reduces the computational cost in the home network and also reduces the effort needed to patch serving networks while deploying the solution.

In Paper II, we have shown that one of the vulnerabilities of hitherto-published pseudonym-based solutions can be exploited into mounting a DDoS attack that can cause serious damage to the HN by using a modest amount of resources. The attack de-synchronises pseudonyms between the user equipment and its home network, and hence, when the user equipment sends a pseudonym, the HN cannot identify the user equipment correctly. In Paper II, we have presented a solution that fixes the vulnerability and defeats the DDoS attack.

In Paper III, we have presented a solution to defeat IMSI catchers that attack a 5G user by downgrading the user to LTE or 3G networks. This solution is an extension of the solution presented in Paper II. Besides, it leverages 5G AKA to fix the vulnerability that a malicious SN could use to mount an attack on a home network to de-synchronise pseudonyms of users of the home network. Leveraging 5G AKA also enables a user equipment to re-synchronise pseudonyms if they get de-synchronised due to some erroneous situation, e.g., memory corruption in the user equipment or in the home network. Some of the implementation issues of this solution are subject to further studies. For example, it needs to be studied which parts of the solution should be implemented in the mobile equipment and which parts should be implemented in the USIM.

In the second category of our contributions (Paper IV), we have looked into a sub-system of the 5G network, known as Authentication and Key Management for Applications (AKMA), which extends and evolves two earlier services that 3GPP specified for previous generations of mobile systems. Those earlier services are GBA (Generic Bootstrapping Architecture) and BEST (Battery Efficient Security for Very Low Throughput Machine Type Communication Devices). When Paper IV was published, the first version of AKMA was still under the process of standardisation. In Paper IV, we have first analysed potential AKMA requirements vs GBA and BEST requirements. We have identified two new privacy requirements that could be useful to protect the privacy of user transactions with the AKMA Application Function (AApF) against, e.g., an insider attacker in the home network of that user. Second, we have developed a privacy mode for AKMA that fulfils those new requirements. Having this kind of solution, mobile network operators can offer premium service to privacy-cautious consumers. Our scheme could be extended in the future by, e.g., including rekeying of master keys. Future work could also include formal security analysis, implementation, and standardisation of our scheme.

References

[1] Silke Holtmanns, Valtteri Niemi, Philip Ginzboorg, Pekka Laitinen, and N. Asokan. *Cellular Authentication for Mobile and Internet Services*. Wiley, 2008.

[2] Wikipedia. Squidgygate. `https://en.wikipedia.org/wiki/Squidgygate`. Last accessed May 4, 2020.

[3] Upasna Sonal. Mobile Phone Cloning. In *International Journal of Engineering Research & Technology (IJERT) NCETEMS - 2015 (Volume 3 - Issue 10)*, 2015.

[4] Philip Ginzboorg and Valtteri Niemi. Privacy of the long-term identities in cellular networks. In *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*, MobiMedia '16. ICST, 2016.

[5] Peter Shaojui Wang, Feipei Lai, Hsu-chun Hsiao, and Ja-ling Wu. Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems. *IEEE Access*, 4:2244–2255, 2016.

[6] Tripwire Guest Authors. Insider Threats as the Main Security Threat in 2017. `https://www.tripwire.com/state-of-security/security-data-protection/insider-threats-main-security-threat-2017/`, Tripwire, April 2017. Last Accessed April 29, 2020.

[7] Nicole Lindsey. Insider Threats: Not Just That Unhappy Employee. `https://www.cpomagazine.com/cyber-security/insider-threats-not-just-that-unhappy-employee/`, CPO Maganize, September 2018. Last Accessed April 29, 2020.

[8] Bauke Brenninkmeijer. Catching IMSI-catcher-catchers: An effectiveness review of IMSI-catcher-catcher applications. `https://www.cs.ru.nl/bachelors-theses/2016/Bauke_Brenninkmeijer___4366298___Catching_IMSI-catcher-catchers.pdf`, Radboud University, July 2016. Last accessed April 29, 2020.

[9] Ravishankar Borgaonkar, Andrew Martin, Shinjo Park, Altaf Shaik, and Jean-Pierre Seifert. White-Stingray: Evaluating IMSI Catchers Detection Applications. In *Proceedings of the 11th USENIX Conference on Offensive Technologies*, WOOT'17, page 21, USA, 2017. USENIX Association.

[10] Shinjo Park, Altaf Shaik, Ravishankar Borgaonkar, and Jean-Pierre Seifert. Anatomy of Commercial IMSI Catchers and Detectors. In *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society*, WPES19, pages 74–86, New York, NY, USA, 2019. Association for Computing Machinery.

[11] Ulrike Meyer and Susanne Wetzel. On the impact of GSM encryption and man-in-the-middle attacks on the security of interoperating GSM/UMTS networks. In *2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No.04TH8754)*, volume 4, pages 2876–2883 Vol.4, Sep. 2004.

[12] Mohsin Khan and Valtteri Niemi. Concealing IMSI in 5G Network Using Identity Based Encryption. In *Network and System Security*, pages 544–554. Springer International Publishing, 2017.

[13] Neal Koblitz and Alfred Menezes. Another look at non-standard discrete log and Diffie-Hellman problems. *IACR Cryptology ePrint Archive*, 2007:442, 2007.

[14] Samuel Warren and Louis Brandeis. The Right to Privacy, 4 Harv. L. Rev. 193. `https://www.stetson.edu/law/studyabroad/spain/media/Wk3.Stuart.Day1-1-THE-RIGHT-TO-PRIVACY-(excerpt).pdf`, December 1890. Last accessed April 29, 2020.

[15] Oyez. Griswold v. Connecticut 381 U.S. 479 (1965). `https://www.oyez.org/cases/1964/496`. Last Accessed April 29,2020.

[16] Alan F. Westin. *Privacy and Freedom.* Ig Publishing, New York, NY 10163, 1967.

[17] Dan Swinhoe. The 15 biggest data breaches of the 21st century. https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html, CSO, April 2020. Last accessed April 29, 2020.

[18] Carole Cadwalladr and Emma Graham-Harrison. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election, The Guardian, March 2018. Last accessed April 29, 2020.

[19] Daniel J. Solove. *Understanding Privacy.* Harvard University Press, 2008.

[20] Paul Sieghart. *Privacy and Computers.* Latimer New Dimensions, 1976.

[21] 3GPP. Study on the security aspects of the next generation system. 3GPP TR 33.899 V 1.3.0 (withdrawn before change control), August 2017. Available on https://www.3gpp.org/.

[22] 3GPP. Vocabulary for 3GPP Specifications. 3GPP TR 21.905 V 16.0.0, November 2019. Available on https://www.3gpp.org/.

[23] David Zomaya. GDPR Compliance: What is PII? https://www.cbtnuggets.com/blog/technology/data/gdpr-compliance-what-is-pii, CBT Nuggets, December 2019. Last accessed April 29, 2020.

[24] Erika McCallister, Tim Grance, and Karen Scarfone. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf, NIST, April 2010. Last accessed April 29, 2020.

[25] Clay Johnson III. Memorandum for the heads of executive departments and agencies. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2007/m07-16.pdf, Office of Management and Budget, May 2007. Last accessed April 29, 2020.

[26] European Union. Article 4 EU GDPR “Definitions” . https://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm, 2016. Last accessed April 29, 2020.

[27] Rob Matheson. The privacy risks of compiling mobility data. http://news.mit.edu/2018/privacy-risks-mobility-data-1207, Massachusetts Institute of Technology, December 2018. Last accessed April 29, 2020.

[28] Ann Cavoukian. Privacy by Design The 7 Foundational Principles Implementation and Mapping of Fair Information Practices. https://iapp.org/media/pdf/resource_center/Privacy%20by%20Design%20-%207%20Foundational%20Principles.pdf, 2020. Last Accessed April 29, 2020.

[29] Hiten Choudhury, Basav Roychoudhury, and Dilip Kr. Saikia. Enhancing User Identity Privacy in LTE. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 949–957, June 2012.

[30] Guangyang Yang, Jia Yu, Wenting Shen, Qianqian Su, Zhangjie Fu, and Rong Hao. Enabling public auditing for shared data in cloud storage supporting identity privacy and traceability. *Journal of Systems and Software*, 113:130 – 139, 2016.

[31] Giorgos Karopoulos, Georgios Kambourakis, Stefanos Gritzalis, and Elisavet Konstantinou. A framework for identity privacy in SIP. *Journal of Network and Computer Applications*, 33(1):16 – 28, 2010.

[32] Giorgos Karopoulos, Georgios Kambourakis, and Stefanos Gritzalis. Privasip: Ad-hoc identity privacy in sip. *Computer Standards & Interfaces*, 33(3):301 – 314, 2011.

[33] Hiten Choudhury, Basav Roychoudhury, and Dilip Kr. Saikia. Article: A New Trust Model for Improved Identity Privacy in Cellular Networks. *International Journal of Computer Applications*, 56(14):1–8, October 2012. Full text available.

[34] Fotios Papaodyssefs, Costas Iordanou, Jeremy Blackburn, Nikolaos Laoutaris, and Konstantina Papagiannaki. Web identity translator: Behavioral advertising and identity privacy with wit. In *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, HotNets-XIV, New York, NY, USA, 2015. Association for Computing Machinery.

[35] Didier Samfat and Refik Molva. A Method Providing Identity Privacy to Mobile Users During Authentication. In *1994 First Workshop on Mobile Computing Systems and Applications*, pages 196–199, Dec 1994.

[36] Zhaohui Cheng, Liqun Chen, Richard Comley, and Qiang Tang. Identity-Based Key Agreement with Unilateral Identity Privacy Using Pairings. *IACR Cryptology ePrint Archive*, 2005:339, 2005.

[37] Myrto Arapinis, Loretta Ilaria Mancini, Eike Ritter, and Mark Ryan. Privacy through pseudonymity in mobile telephony systems. In *NDSS*, 2014.

[38] Catherine Tucker. The Economics of Advertising and Privacy. *International Journal of Industrial Organization*, 30(3):326 – 329, 2012.

[39] Yezid Donoso. *Network Design for IP Convergence.* CRC Press, 2009.

[40] Ari T. Manninen. Elaboration of NMT and GSM standards: From idea to market. `https://jyx.jyu.fi/bitstream/handle/123456789/13478/951391786x.pdf?sequence=1`, September 2002. Last Accessed April 29, 2020.

[41] Svenolof Karlsson and Anders Lugn. Launch of the NMT 900. `https://www.ericsson.com/en/about-us/history/changing-the-world/world-leadership/launch-of-the-nmt-900`, Ericsson. Last Accessed April 29, 2020.

[42] 3GPP. GSM Spec history: Drafting and publication of GSM Specs. `https://www.3gpp.org/specifications/gsm-history`. Last Accessed April 29, 2020.

[43] Liansheng Tan. *Resource Allocation and Performance Optimization in Communication.* CRC Press, 2018.

[44] 3GPP. Historical Information (Incl. Closed groups). `https://www.3gpp.org/specifications-groups/31-historical-information-incl-closed`. Last Accessed April 29, 2020.

[45] Anton A. Huurdeman. *The Worldwide History of Telecommunications.* John Wiley & Sons, 2003.

[46] Bharat Bhasker. *Electronic Commerce: Framework, Technologies and Applications.* Tata McGraw-Hill Education, 2013.

[47] Dan Forsberg, Günther Horn, Wolf-Dietrich Moeller, and Valtteri Niemi. *LTE Security.* Wiley, 2008.

[48] CDG. CDMA History. `http://www.cdg.org/resources/cdma_history.asp`. Last Accessed April 29, 2020.

[49] 3GPP. Functionality in early GSM releases. `https://www.3gpp.org/specifications/releases/78-functionality-in-early-gsm/`. Last Accessed April 29, 2020.

[50] 3GPP. UMTS detailed description. `https://www.3gpp.org/news-events/3gpp-wiki/2-uncategorised/1984-umts-detailed-description`. Last Accessed April 29, 2020.

[51] 3GPP. LTE. `https://www.3gpp.org/technologies/keywords-acronyms/98-lte`. Last Accessed April 29, 2020.

[52] 3GPP. Release description; Release 15. `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3389`, October 2019.

[53] ETSI. Smart Cards; Vocabulary for Smart Card Platform specifications. ETSI TR 102 216 V 4.0.0, August 2019.

[54] 3GPP. 3G security; Security architecture. 3GPP TS 33.102 V 15.1.0, December 2018. Available on `https://3gpp.org/`.

[55] 3GPP. Numbering, addressing and identification. 3GPP TS 23.003 16.2.0, March 2020. Available on `https://3gpp.org`.

[56] 3GPP. System Architecture for the 5G System. 3GPP TS 23.501 V 16.4.0, March 2020. Available on `https://3gpp.org/`.

[57] 3GPP. Security architecture and procedures for 5G System. 3GPP TS 33.501 V 16.2.0, March 2020. Available on `https://3gpp.org/`.

[58] 3GPP. 3GPP System Architecture Evolution (SAE); Security architecture. 3GPP TS 33.401 V 16.2.0, March 2020. Available on `https://3gpp.org/`.

[59] IMEI.info. Check IMEI Number to Get to Know Your Phone Better. `https://www.imei.info/`. Last accessed April 30, 2020.

[60] Fabian Van den Broek, Roel Verdult, and Joeri de Ruiter. Defeating IMSI Catchers. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS'15. ACM, 2015.

[61] Karl Norrman, Mats Näslund, and Elena Dubrova. Protecting IMSI and User Privacy in 5G Networks. In *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*, MobiMedia'16. ICST, 2016.

[62] Mohammed Shafiul Alam Khan and Chris J. Mitchell. Improving Air Interface User Privacy in Mobile Telephony. In *Second International Conference, SSR 2015, Proceedings*. Springer International Publishing, 2015.

[63] 3GPP. Characteristics of the Universal Subscriber Identity Module (USIM) application. 3GPP TS 31.102 V 16.4.0, July 2020. Available on `https://3gpp.org/`.

[64] ETSI. Smart Cards; UICC-Terminal Interface; Physical AND Logical Characteristics. ETSI TS 102 221.

[65] Gizem Akman, Philip Ginzboorg, and Valtteri Niemi. AKMA Support in Multi SIM User Equipment. In Sergey Balandin, Valtteri Niemi, and Tatiana Tuytina, editors, *Proceedings of the 25th Conference of Open Innovations Association FRUCT, Helsinki, Finland*, Proceedings of the ... Conference of Open Innovations Association FRUCT, pages 15–24, Finland, 2019. FRUCT Oy.

[66] Valtteri Niemi and Kaisa Nyberg. *UMTS Security.* Wiley, 203.

[67] 3GPP. Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture (GBA): TS 33.220 V 16.0.0, September 2019. Available on `https://3gpp.org/`.

[68] 3GPP. Battery Efficient Security for very low throughput Machine Type Communication (MTC) devices (BEST), TS 33.163 V 16.1.0, June 2019. Available on `https://3gpp.org/`.

[69] 3GPP. Study on authentication and key management for applications based on 3GPP credential in 5G. 3GPP TR 33.835 V 16.0.0, December 2019. Available on `https://3gpp.org/`.

[70] EFRAM. SIM Cards: What is the difference between ICCID, IMSI and IMEI numbers? `https://m.blog.naver.com/framkang/220363349346`, May 2015. Last accessed April 29, 2020.

[71] Myrto Arapinis, Loretta Mancini, Eike Ritter, Mark Ryan, Nico Golde, Kevin Redon, and Ravishankar Borgaonkar. New Privacy Issues in Mobile Telephony: Fix and Verification. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS'12, pages 205 – 216, New York, NY, USA, 2012. Association for Computing Machinery.

[72] Altaf Shaik, Jean-Pierre Seifert, Ravishankar Borgaonkar, N. Asokan, and Valtteri Niemi. Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems. In *23rd Annual Network and Distributed System Security Symposium, NDSS 2016, San Diego, California, USA, February 21-24, 2016*, 2016.

[73] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and Defending Against Third-Party Tracking on the Web. In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pages 155–168, San Jose, CA, 2012. USENIX.

[74] Aaron Cahn, Scott Alfeld, Paul Barford, and S. Muthukrishnan. An Empirical Study of Web Cookies. In *Proceedings of the 25th International Conference on World Wide Web*, WWW'16, pages 891–901, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee.

[75] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, and Sándor Imre. User tracking on the web via cross-browser fingerprinting. In Peeter Laud, editor, *Information Security Technology for Applications*, pages 31–46, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[76] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 878–894, May 2016.

[77] Ting-Fang Yen, Yinglian Xie, Fang Yu, Roger Peng Yu, and MartÃn Abadi. Host fingerprinting and tracking on the web: Privacy and security implications. In *NDSS*. The Internet Society, 2012.

[78] Zack Whittaker. Hackers are stealing years of call records from hacked cell networks: At least 10 cell networks have been hacked over the past seven years. `https://techcrunch.com/2019/06/24/hackers-cell-networks-call-records-theft/`, June 2019. Last accessed April 30, 2020.

[79] 3GPP. 3G security; Lawful interception requirements. 3GPP TS 33.106 V 15.1.0, June 2018. Available on `https://3gpp.org/`.

[80] 3GPP. Lawful Interception Requirements. 3GPP TS 33.126 V 16.1.0, September 2019. Available on `https://3gpp.org/`.

[81] Douglas Stinson. *Cryptography: Theory and Practice.* Champman & Hall/CRC, third edition, 2006.

[82] Oded Goldreich. *Foundations of Cryptography: Volume 1.* Cambridge University Press, USA, 2006.

[83] Adi Shamir. Identity-Based Cryptosystems and Signature Schemes. In *Advances in Cryptology: Proceedings of CRYPTO 84*, pages 47–53. Springer Berlin Heidelberg, 1985.

[84] Dan Boneh and Matthew Franklin. Identity-Based Encryption from the Weil Pairing. In *Advances in Cryptology CRYPTO 2001: 21st Annual International Cryptology Conference, 2001 Proceedings*, pages 213–219. Springer Berlin Heidelberg, 2001.

[85] Clifford Cocks. An Identity Based Encryption Scheme Based on Quadratic Residues. In *Proceedings of the 8th IMA International Conference on Cryptography and Coding*, pages 360–363, London, UK, 2001. Springer-Verlag.

[86] ETSI. Digital cellular telecommunications system (Phase 2); Security management (GSM 12.03). `https://www.etsi.org/deliver/etsi_i_ets/300600_300699/300614/01_60/ets_300614e01p.pdf`, 1996. Last accessed April 30, 2020.

[87] ETSI. Digital cellular telecommunications system (Phase 2); Security aspects (GSM 02.09 version 4.5.1). `https://www.etsi.org/deliver/etsi_`

i_ets/300500_300599/300506/03_60/ets_300506e03p.pdf, 2000. Last accessed April 30, 2020.

[88] 3GPP. User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode: TS 25.304 V 15.0.0, June 2018. Available on https://3gpp.org/.

[89] 3GPP. Requirements for Support of Radio Resource Management (TDD). 3GPP TS 25.123 V 16.0.0, January 2019. Available on https://3gpp.org/.

[90] 3GPP. Requirements for Support of Radio Resource Management (FDD). 3GPP TS 25.133 V 16.0.0, January 2019. Available on https://3gpp.org/.

[91] Sassan Ahmadi. Chapter 4 - System Operation and UE States. In Sassan Ahmadi, editor, *LTE-Advanced*, pages 153 – 225. Academic Press, 2014.

[92] Haibat Khan and Keith M. Martin. A Survey of Subscription Privacy on the 5G Radio Interface - The Past, Present and Future. *IACR Cryptology ePrint Archive*, 2020:101, 2020.

[93] 3GPP. Study on authentication and key management for applications based on 3GPP credential in 5G (Work Item Description). https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionUid=SP-180443, June 2018. Last accessed April 30, 2020.

[94] 3GPP. Radio Resource Control (RRC); Protocol specification. 3GPP TS 25.331 V 15.4.0, April 2020. Available on https://3gpp.org/.

[95] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification. 3GPP TS 36.331 V 16.0.0, April 2020. Available on https://3gpp.org/.

[96] 3GPP. 3G Security; Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 1: General: TS 35.205 V 15.0.0, October 2018. Available on https://3gpp.org/.

[97] 3GPP. 3G Security; Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation

functions f1, f1*, f2, f3, f4, f5 and f5*; Document 2: Algorithm specification. 3GPP TS 35.206, 15.0.0, October 2018. Available on `https://3gpp.org/`.

[98] 3GPP. 3G Security; Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 3: Implementors' test data: TS 35.207, 15.0.0, December 2018. Available on `https://3gpp.org/`.

[99] 3GPP. 3G Security; Specification of the MILENAGE algorithm set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*; Document 4: Design conformance test data. 3GPP TS 35.208, 15.0.0, October 2018. Available on `https://3gpp.org/`.

[100] 3GPP. Non-Access-Stratum (NAS) Protocol for Evolved Packet System (EPS); Stage 3. 3GPP TS 24.301 V 16.4.0, March 2020. Available on `https://3gpp.org/`.

[101] 3GPP. Location Management Procedures. 3GPP TS 23.012 V 15.0.0, June 2018. Available on `https://3gpp.org/`.

[102] 3GPP. General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access. 3GPP TS 23.401 V 16.6.0, March 2020. Available on `https://3gpp.org/`.

[103] Mohammed Khan and Chris J Mitchell. Trashing IMSI Catchers in Mobile Networks. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec 2017), Boston, USA, July 18-20, 2017*, United States, May 2017. Association for Computing Machinery (ACM).

[104] Certicom Research. SEC 1: Elliptic Curve Cryptography. Standards for Efficient Cryptography, V 2.0, `http://www.secg.org/sec1-v2.pdf`, May 2009. Last accessed April 30, 2020.

[105] Certicom Research. SEC 2: Recommended Elliptic Curve Domain Parameters. Standards for Efficient Cryptography, V 2.0, `http://www.secg.org/sec2-v2.pdf`, January 2010. Last accessed April 30, 2020.

[106] NIST. Advanced Encryption Standard (AES). FIPS PUB 197, 2001.

[107] Hugo Krawczyk, Mihir Bellare, and Ran Canetti. HMAC: Keyed-Hashing for Message Authentication. RFC 2104, February 1997.

[108] NIST. Secure hash standard (SHS). FIPS PUB 180-4, 2012.

[109] Daniel J Bernstein. Curve25519: New Diffie-Hellman Speed Records. In *Public Key Cryptography (PKC'06)*, volume 3958 of *Lecture Notes in Computer Science*, pages 207–228. Springer, 2006.

[110] Adam Langley, Mike Hamburg, and Sean Turner. Elliptic Curves for Security. RFC 7748, January 2016.

[111] Mohsin Khan, Philip Ginzboorg, and Valtteri Niemi. IMSI-based Routing and Identity Privacy in 5G. In *Proceedings of the 22st Conference of Open Innovations Association FRUCT*, Jyvaskyla, Finland, May 2018.

[112] Ashkan Soltani and Craig Timberg. Tech firm tries to pull back curtain on surveillance efforts in Washington, The Washington Post, 17 September 2014. Last accessed April 30, 2020.

[113] CBC News. Devices that track, spy on cellphones found at Montreal's Trudeau airport. `https://www.cbc.ca/news/canada/montreal/trudeau-airport-spying-1.4055803`, April 2017. Last accessed April 30, 2020.

[114] Amir Herzberg, Hugo Krawczyk, and Gene Tsudik. On Travelling Incognito. In *1994 First Workshop on Mobile Computing Systems and Applications*. IEEE Xplore, 1994.

[115] N. Asokan. Anonymity in a Mobile Computing Environment. In *1994 First Workshop on Mobile Computing Systems and Applications*, Santa Cruz, California, USA. IEEE.

[116] Geir M. Køien. Privacy Enhanced Mutual Authentication in LTE. In *2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 614–621, Oct 2013.

[117] Christina Patsioura. Despite the tech paradox, privacy is far from dead. `https://data.gsmaintelligence.com/research/research/research-2020/despite-the-tech-paradox-privacy-is-far-from-dead`, 2020. Last accessed June 02, 2020.

[118] Arkajit Dey and Stephen Weis. PseudoID: Enhancing Privacy for Federated Login. 2010.

[119] Sven Hammann, Ralf Sasse, and David Basin. Privacy-Preserving OpenID Connect, March 2020. Available on `https://people.inf.ethz.ch/rsasse/pub/poidc-asiaccs.pdf`.

[120] Eric Rescorla. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446, August 2018.

[121] Eric Rescorla and Tim Dierks. The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246, August 2008.

[122] David Freeman. Pairing-based Identification Schemes, 2005.

[123] Neal Koblitz and Alfred Menezes. Another Look at Non-standard Discrete Log and Diffie-Hellman problems. `https://eprint.iacr.org/2007/442.pdf`, 2007. Last Accessed April 29, 2020.

[124] 3GPP. Authentication and Key Management for Applications (AKMA) Based on 3GPP Credentials in the 5G System (5GS). 3GPP TS 33.535 V 16.0.0, July 2020. Available on `https://3gpp.org/`.