Annals of Telecommunications (2019) 74:593–603
https://doi.org/10.1007/s12243-019-00721-x

A survey on authentication and access control for mobile networks: from 4G to 5G

Shanay Behrad[1] · Emmanuel Bertin[1] · Noel Crespi[2]

Received: 15 June 2018 / Accepted: 26 June 2019 / Published online: 12 July 2019

Abstract
The next generation of mobile networks, 5G, is expected to support a set of multiple requirements and use cases that will create an improved user experience. 5G will also be able to provide a high level of security by considering a variety of security aspects, such as authentication and access control mechanisms. The current protocol in 4G designed to address security is 4G AKA. It presents some weaknesses and vulnerabilities that negatively affect operators' networks and their subscribers' security. In designing an authentication and access control mechanism for 5G, it is crucial to evaluate both 4G AKA's weaknesses and the new requirements of 5G. In this paper, we survey the vulnerabilities of the 4G AKA protocol, as well as the current 5G architectural answers brought by the 3GPP.

Keywords 5G, mobile network · Authentication and access control · AKA protocol

1 Introduction

Security is one of the most important requirements of any mobile telecommunications system. Providing suitable connectivity services to a network's subscribers, preventing the network from being abused, and protecting the subscribers' privacy and their information are all security issues. Authentication of the users for network access and ensuring a bidirectional trust between users and their network are key elements of building such secured systems. Both secure connectivity and user authentication are related to the authentication and access control mechanisms that provide secure network services for network subscribers.

In 2G, user authentication is based on the SIM (Subscriber Identity Module). A SIM card is a well-known secure element that is provided by the operator to its subscribers and contains the subscriber's IMSI (International Mobile Subscriber Identity) and a permanent key to establish a secure connection between the subscriber and the network. However, its lack of mutual authentication has led to active attacks against subscribers (e.g., an attacker can impersonate itself as a valid network to subscribers). Since 3G, the 3GPP (3rd Generation Partnership Project) has made use of AKA (Authentication and Key Agreement) protocols [1] with mutual authentication feature to address this issue. The AKA mechanism in 4G systems (EPS-AKA) is a complementary form of the AKA mechanism in 3G (UMTS-AKA), with a few differences [2].

It is expected that the forthcoming generation of mobile systems, 5G, will meet the requirements of higher throughput, low latency, and better quality of service. Some additional concepts have also been included in the scope of 5G, such as handling the connectivity for the IoT (Internet of things), providing network slices to specific customers or vertical sectors, and managing heterogeneous network access (e.g., addressing Wi-Fi and cellular access networks from a converged network). All of these requirements and concepts affect the whole network and the associated security needs. Authentication and access control mechanisms for 5G should thus consider the issues and weaknesses of current AKA protocols, as well as these new requirements.

In this paper, we review the challenges of the EPS-AKA procedure for 4G systems and discuss the new needs arising

✉ Shanay Behrad
shanay.behrad@orange.com

Emmanuel Bertin
emmanuel.bertin@orange.com

Noel Crespi
noel.crespi@it-sudparis.eu

[1] Orange Labs, Caen, France

[2] Institut Mines-Telecom, Telecom SudParis, CNRS 5157, Évry, France

from the new 5G use cases, as well as how standards are currently evolving. The remainder of this paper is organized as follows. In section 2, we explain the main nodes of the 4G architecture that participate in EPS-AKA procedures. We study the vulnerabilities of EPS-AKA and summarize them in a survey table in section 3. Section 4 is where we discuss the authentication and access control impact of new 5G use cases and introduce the current architectural answers from the perspective of the 5G.

2 Summary of current (4G) authentication and access control mechanism

2.1 4G architecture

The 4G architecture combines many functional entities to ensure Authentication and Access Control. The 4G network consists of the operator's IP network and all of the entities that are connected to this IP network. This means that all the entities have the same IP protocol and communicate with each other via a typical IP network (through logical interfaces). The main entities of a 4G network are described below and summarized in Fig. 1:

- UE (User Equipment): a mobile device that includes a UICC (Universal Integrated Circuit Card, a SIM card) integrated with the USIM (Universal Subscriber Identity Module). The USIM stores user-related information, such as the IMSI (International Mobile Subscriber Identity) that is used to identify each SIM card in a unique way, and the subscriber's secret key (which is pre-shared with the AuC in the home network and never leaves these two elements). The IMSI uniquely identifies a subscriber and consists of three parts: an MCC (Mobile Country Code), an MNC (Mobile Network Code) that specifies the subscriber's carrier network, and an MSIN (Mobile Subscriber Identification Number) that identifies the subscriber in the mobile network. The USIM participates in the subscriber authentication process.
- eNodeB (evolved Node B): the main component of the E-UTRAN (Evolved Universal Terrestrial Radio Access Network). Each eNodeB consists of an antenna and a set of transceivers. Each UE is connected to the core network via eNodeBs. They are directly linked together; this flat architecture promotes lower latency and better connection performance [3, 4].
- MME (Mobility Management Entity): the main control node of the network. The MME performs authentication and is mainly responsible for the attachment process, bearer handling (in collaboration with the P-GW), the tracking of UE locations, and selecting the gateways (deciding the pathways of the data packets).
- HSS (Home Subscriber Server): a database that stores the subscriber's data (including their identities, rights, and subscription profiles) and the secret keys. HSS contains the AuC (Authentication Center) that holds and generates all the needed cryptographic material. It provides authentication data to the MME.
- S-GW (Serving Gateway): anchors the data bearer and routes data packets to the UE.
- P-GW (Packet Data Network Gateway) connects the packet core network to the external networks, such as the internet, and provides IP addresses to the UE. It is also responsible for policy enforcement, billing, and charging based on the rules provided by a PCRF.
- PCRF (Policy and Charging Rules Function): manages the bandwidth and network resources usage and controls the

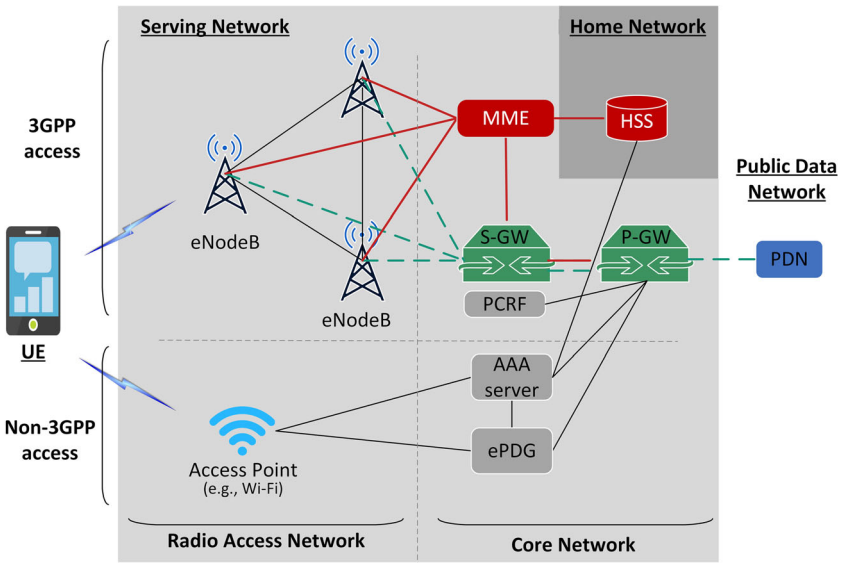

Fig. 1 LTE network architecture. The MME and the HSS are in the control plane and the S-GW and the P-GW are in the user plane. The solid lines show the control plane links and the dashed lines show the user plane links

QoS of the sessions for each subscriber according to the subscription information, the provided services, and the peak usage times.

The 4G network architecture is designed to separate the entities that manage the control (Control Plane) from the entities that take care of traffic (Data Plane). All of the data plane packets in the (public) packet data network that are destined for 4G network subscribers (UEs) are routed to the P-GW of the operator's network. The P-GW sends the data plane packets to the S-GW, the S-GW sends them to eNodeB, and the eNodeB delivers them to the intended UE. S-GWs act as intermediary entities. Each is responsible for a specific geographic area. The movements of UE are usually within the same S-GW. Therefore, thanks to these S-GWs, there is no need to bother the P-GW for UE location updates.

In addition to data plane packets, a set of control functions and signaling messages (control plane packets) are also transmitted in the network to manage network access or the tracking of UEs when they move. Subscribers' authentication and access control processes belong to this category. MME and HSS only take care of control plane packets and do not manage data plane packets. The control messages' path (that is related to the subscribers' authentication and access control) is between UEs, eNodeB, MME, and HSS. An MME is designed to prevent HSSs from being disrupted by the millions of UE requests for each of their activities needing access control (e.g., location update). Each MME manages a very large region. The number of MMEs in a PLMN (Public Land Mobile Network) depends on the operator's decision (e.g., the size of the area that is under the responsibility of the operator). At the first attachment of a UE, the MME obtains the UE's profile and all the security information from the HSS. Then, for all further accesses, the MME will be able to verify the UE's access rights.

The 4G architecture supports multiple access technologies (trusted and untrusted access networks). The operator decides which non-3GPP access networks are trustworthy and which are not. The handling of non-3GPP accesses involves two other entities:

- AAA Server: responsible for the authentication and authorization of the UE in the case of non-3GPP access; and
- EPDG (Evolved Packet Data Gateway): responsible for the establishment of an IPsec tunnel between the operator's core network and the UE in the case of untrusted non-3GPP access.

However, we mainly focus on 3GPP access in the scope of this paper, as it is the main part of the network under the responsibility of the operator.

2.2 EPS-AKA overview

When someone subscribes to a network, the operator provides him/her a connectivity service in the form of a SIM card (with IMSI and a secret key). In 3GPP access networks, the 4G authentication process is supported by the EPS-AKA protocol, which is an authentication and key agreement protocol between the UE (via its SIM card) and the network. In the current 5G specifications, this protocol is reused, with some differences [5].

EPS-AKA provides mutual authentication (the network authenticates the UE and the UE authenticates the serving network), based on symmetric key cryptography. In this protocol, at first, the parties authenticate each other according to the secret key, then some other keys are derived from this secret key to protect data integrity and confidentiality.

There are two sets of protocols in mobile systems that concern the UE, NAS (Non-Access Stratum) and AS (Access Stratum). NAS protocols are for connections between the UE and the core network, while AS protocols cover the link with the radio access network to establish and manage radio connections. While NAS messages are transmitted via eNodeBs, their content is not analyzed by eNodeBs, but the AS messages are analyzed by the eNodeBs. All the messages in EPS-AKA are NAS messages. As mentioned above, every UE should have an IP address to be able to send and receive data. The IP address allocation to a UE is done after the EPS-AKA procedure.

Figure 2 depicts the EPS-AKA procedure and the attacks it may encounter. As indicated in Fig. 2, the EPS-AKA procedure starts by sending the *Attach request* message from the UE to the MME after it finds its operator's eNodeB (each eNodeB broadcasts the operator's identity via a beacon channel). This message contains the UE's IMSI or GUTI (Globally Unique Temporary Identifier) [6]. GUTI is a temporary identifier that the MME allocates to a UE after the initial attachment procedure and after the activation of the radio channel encryption, thereby protecting the IMSI from eavesdropping (i.e., to avoid the IMSI having to be transmitted frequently). A GUTI consists of a TMSI (Temporary Mobile Subscriber Identity) that is allocated by the UE's current MME, and the MME's identifier.

If the MME cannot recognize the GUTI, it sends an *Identity request* message to the UE, which then sends its IMSI in the *Identity response* message. The rest of the EPS-AKA procedure is as follows [2, 7]:

- The MME sends an authentication information request that contains the UE's IMSI and the SNid (Serving Network Identifier to the HSS). The UE trusts the home network about the verification of the serving network's identity (the home network uses the SNid to compute the serving network's specific K_{ASME} key that we will describe below).

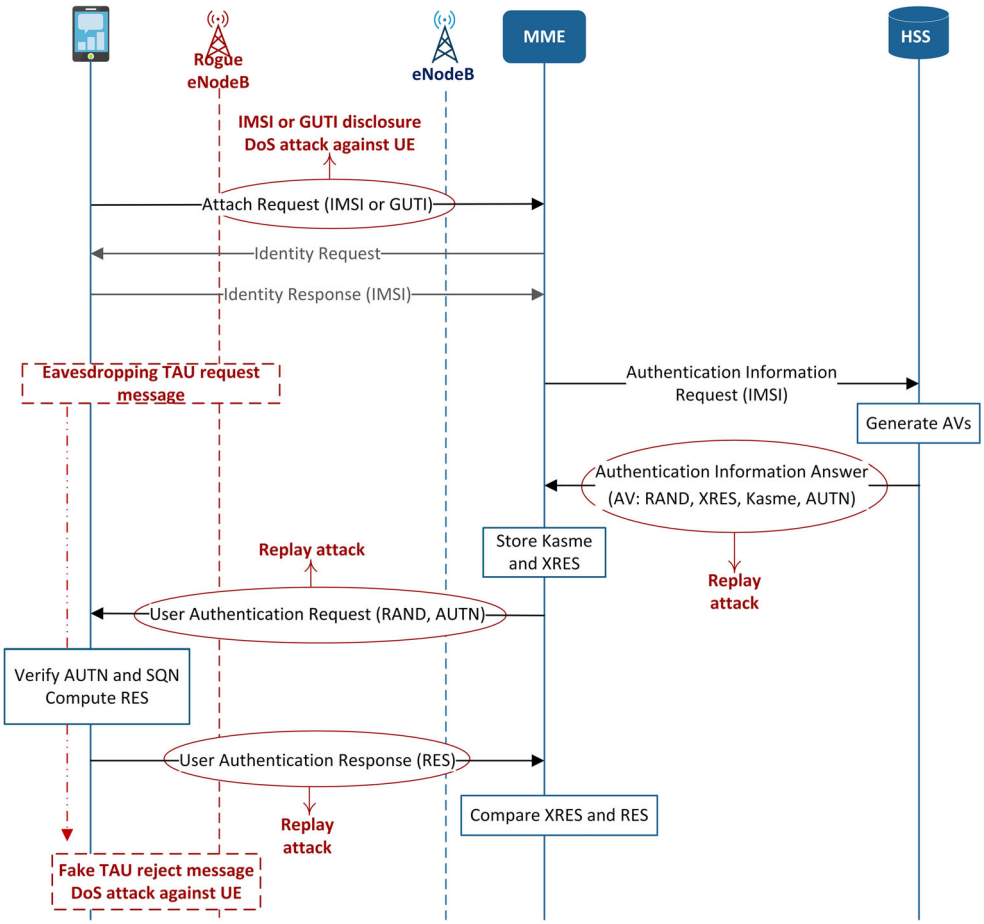

Fig. 2 EPS-AKA procedure and the attacks against it

- The HSS generates a random number RAND. To authenticate the UE, the network needs to be sure about the presence of the secret key in the UE. As explained in subsequent steps, to indicate this presence, the HSS sends this random number (RAND) to the UE. Then, the HSS and the UE will do the same calculation with RAND, and if the results are the same, the presence of the secret key in the UE will be approved. The HSS also finds the UE's secret key K (according to the UE's IMSI) and then inputs the RAND and the K into cryptographic functions to generate AVs (Authentication Vectors). AVs consist of the RAND, an XRES (the MME checks if XRES is equal to the RES from the UE to authenticate the UE), a local master key K_{ASME} (computed by a key derivation function with the SNid as one of its inputs) and an AUTN (Authentication Token). The AUTN is the result of another calculation with the random number and the secret key. The AUTN will be used by the UE to authenticate the network. The UE will also calculate it, and if it gets the same amount, it will trust the network. The other input of the cryptographic functions is SQN (a counter) that is increased with each new authentication. The HSS keeps this counter for each UE, using it to prevent an attacker from impersonating itself to the UE by stealing the AVs and reusing them. Indeed, SQN guarantees the freshness of the AVs.
- The HSS sends the AV to the MME that stores the K_{ASME} and XRES parts of the AV, and sends the RAND and AUTN to the UE.
- The USIM inside the UE retrieves the SQN from the AUTN by using the secret key K and the RAND; next, it computes the XMAC by using the SQN, the RAND, and the AMF part of the AUTN, and compares the XMAC value with the MAC part of the AUTN. Then, it checks if the SQN is in the right range (USIM has its own SQN, and so it checks if the SQN from the HSS is not too far from its own SQN, to ensure synchronization between the HSS and the UE). This is how the UE authenticates the network. Next, the USIM computes the K_{ASME}, so that both the UE and the MME have the same key with which to establish secure connections. The USIM also computes a RES and sends it to the MME. If the SQN is not in the expected range, the UE sends a synchronization failure message, and if the XMAC is not the same as the MAC, the UE sends a MAC failure message.
- The MME checks if XRES and RES are equal and then completes the authentication and key agreement process.

The authentication process described above is also implicitly an access control or authorization process. The authentication of the UE is indeed necessary to provide it access to the network resources. Next, we survey the identified vulnerabilities of this key process.

3 EPS-AKA vulnerabilities

There are various security concerns with LTE security. In the scope of this paper, we only focus on authentication and access control; and so we mainly consider EPS-AKA protocol vulnerabilities, as this protocol plays the main role in securing network access and ensuring the privacy of UEs. Table 1 summarizes these vulnerabilities and their effects on the security of the LTE system. As a general principle, authentication of UEs is needed to avoid the fraudulent use of the network (e.g., by stealing other UEs' IMSIs).

The first vulnerability is IMSI disclosure (IMSI catching), which affects user confidentiality. As mentioned in the previous section, the UE sends the IMSI to the MME in clear text during the first attachment procedure. Furthermore, the IMSI is transmitted in paging messages that are sent from the MME to eNodeBs and from eNodeBs to UEs, in order to locate a specific UE (for example, when a UE has an incoming call). An attacker can trigger a paging procedure without alerting the user, e.g., by using social network applications, and then sniff the paging messages between eNodeB and a UE to decode them and acquire the IMSI [17, 21]. In handover cases between MMEs, if a synchronization failure occurs, the new MME or the previous one request the UE's IMSI, which is then transmitted in clear text again [4, 22–24]. In these cases, an attacker can simply eavesdrop the connection to capture IMSIs.

One of the problems of IMSI disclosure is the theft of services with session mix-up attacks. This can be an inside attack, where the attacker is a subscriber of the network but impersonates itself as another subscriber to use the services that the victim should get from the network [8, 22, 25]. It could also be an outside attack, in which the attacker is not a network subscriber network and swaps services between network subscriber network [25]. Theft of service attacks can also happen between the UE and the IMS parts of the network (IP multimedia subsystems that provide multimedia services such as voice calls) and thus affect the operator's revenue [18]. It is also possible for an attacker to force a UE to repeatedly send IMSIs, thereby expending both the computational power of the HSS and the memory of the MME [9, 26]. In addition to the above problems, IMSI disclosure can cause a service disruption for the UE. As mentioned in the previous section, the UE checks the SQN range after getting the AV. If a malicious UE sends attach requests several times by using a victim UE's IMSI, the SQN amount increases in an uncounted way on the HSS side. Then, if the victim UE sends a real attach request to the network, it will get an AV with an out of range SQN and so the UE will face a synchronization failure.

To solve the IMSI disclosure problem, some solutions based on public key cryptography have been proposed [11–16, 20, 27]. Some of these encrypt all the messages between the UE and the network, and some only encrypt the IMSIs. Most of the public key-based solutions increase the computational and communication costs for UEs (with limited capabilities and energy) and for network elements. Pseudonym-based solutions to the IMSI disclosure problem were also proposed [24, 28], but these require additional capabilities in UEs or additional entities in the network [4, 12, 20, 28].

As mentioned in the previous section, GUTI is a temporary identifier and should be fresh. The main purpose of using this

Table 1 Summary of EPS-AKA vulnerabilities and attacks, the goal of these attacks, and the current solutions

Vulnerability	Attacks	Attacks goals	Proposed solutions
• IMSI disclosure • GUTI persistence	• Impersonating UEs [4, 8–10] • MitM	• Weaken subscriber confidentiality • DoS attacks against the HSS and the MME • Theft of service	• Public key-based solutions [11–16]
• SNid disclosure	• Rogue eNodeB [9, 17–19]	• Disclosure of the subscriber's location • Weaken UE's data security • Intercepting connections between the UE and the network • DoS against the MME	• Public key-based solutions [13, 20]
• Acceptance of TAU reject, service reject, attach reject messages without integrity protection	• DoS attack	• DoS against a UE [17]	• Public key + digital signature [17]
• UE's network and security capabilities disclosure	• Bidding down attack	• DoS against a UE	• Public key + digital signature [17]
• Synchronization failure	• Replay attack • Impersonating UEs	• Disclosure of the subscriber's location • DoS against UEs	

temporary identity is to have a protection against the UE's location disclosure (if a UE sends its IMSI to the network frequently, an attacker can detect it and determine that the UE is nearby). However, in reality, GUTIs are not changed frequently (the operator does not configure its network to refresh the GUTI frequently), and so their disclosure may cause the same problems as IMSI disclosure [17, 20, 29]. An attacker can also change GUTIs. In this scenario, the server cannot recognize GUTIs and so requests UEs to send their IMSIs [16].

One of the most severe types of attacks is to use a rogue eNodeB that pretends to be a legitimate eNodeB. By operating with high power, a false eNodeB can force UEs to connect to it [9, 17–19, 30]. A rogue eNodeB can redirect UEs to another network that provides weak data encryption instead of the UE's home network [22]. It can cause man-in-the-middle attacks (MitM, where the attacker impersonate itself to the network as a legitimate UE) [22] and also the disclosure of a UE's location. A rogue eNodeB can compromise session keys during handover processes as well (de-synchronization attacks) [4, 23], or hijack the paging channel (blocking the UE's incoming calls or creating paging messages with a victim UE's IMSI and forcing it to disconnect from the current legitimate eNodeB and send and attach request to the rogue eNodeB). Leakage of the SNid, because of clear transmission from the MME to the UE, may also cause rogue eNodeB attacks [13, 23]. SNid disclosure may cause traffic on the MME as well, as an attacker can force UEs to attach to an MME [20]. Furthermore, LTE systems support femtocells and HeNodeBs and operators do not control them, so an attacker can use them as rogue eNodeBs to collect IMSIs [12, 27].

The next type of vulnerability is related to the TAU (Tracking Area Update) procedure. Mobile operators divide their service area into tracking areas and each tracking area consists of a number of cells. UEs inform the MME about their locations by sending TAU messages. Some network services are not accessible in some tracking areas, or some UEs are not authorized to access them; as a result, the network sends TAU reject message to UEs. This message is not encrypted and integrity protected (if the UE performs the TAU procedure after changing location in idle mode, it does not contain the keys for the encryption and the integrity protection purposes). In this case, an attacker can cause DoS (Denial of Service) attacks against a UE by getting TAU request messages from a UE via a rogue eNodeB and sending TAU reject message to the UE with "LTE services not allowed" or "LTE and non-LTE services not allowed" content [17, 26, 30]. It is also possible for an attacker to use the location information of a UE to find a link between its IMSI and GUTI and then trace the UE across the network [12].

DoS attacks against UEs can also happen during an attachment procedure when the UE sends its network and security capabilities to the network. An attacker can change this message, causing the MME to reject some of the UE's requests [17, 22].

Unprotected AVs' vulnerability can be used to determine if a specific UE is in a particular area or not, and thus track its movements. AVs are sent in clear text between the HSS and the MME and between the MME and the UE [13]. If an attacker gets these AVs (using User Authentication requests) by eavesdropping the connection between the MME and the UE, it can replay them. The attacker will then send these AVs to the UEs in a specific area. The UE that the AVs belong to will send synchronization failure message and the other UEs will send MAC failure messages, allowing the attacker to determine the presence of the UE in that location [16, 18, 20, 21, 31–33].

Finally, EPS-AKA is based on symmetric key cryptography, and all the keys that are used to prevent data integrity are derived from the secret key (in the key hierarchy); therefore, the leakage of this key would cause serious problem to the whole network [13, 22].

In addition to the aforementioned vulnerabilities, some security issues are due to the interworking with non-3GPP access networks. The UE uses EAP-AKA and EAP-AKA′ as the authentication and key agreement protocol when trying to access the LTE core network via a non-3GPP access network, as well as during handover procedures between 3GPP access networks and non-3GPP access networks [6]. These protocols are similar to the EPA-AKA protocol (instead of MME, they work with an AAA server; the needed keys are driven from the AVs that the AAA server gets from the HSS) and so they have similar vulnerabilities, such as attacks against UE privacy and location, DoS attacks, UE impersonation, and billing mechanism attacks [34–36].

4 New 5G needs

The fifth generation of mobile communications has a number of goals, such as achieving low latency, high data rates, increased convergence, accessibility, and dense connectivity. 5G will also support IoT (Internet of Things) services and address the needs of different vertical markets, such as healthcare, automotive, and transport. The 5G-PPP (Fifth Generation Public Private Partnership) has defined several different use cases for 5G, including enhanced mobile broadband and critical communications [37].

These different goals and use cases have important impacts on the security aspects of the system, and service-specific security requirements should be considered when designing appropriate authentication and access control mechanisms for 5G networks, e.g., fast communications need fast AKA procedures [38]. As another example, in the IoT, numerous devices may access the network at the same time, and so the network should have the ability to control this large amount

of signaling traffic and authenticate the devices correctly to avoid DDoS (Distributed Denial of Service) attacks. IoT devices have low power capacity and cannot support strong authentication procedures. In addition, they are usually able to connect to the network via non-3GPP access options (some of them will not have 5G radio access and will use Wi-Fi or Bluetooth) [39]. In light of these limitations, some solutions based on group-based authentications with an IoT gateway have been proposed to decrease the number of full AKA procedure executions [40, 41]. But these group-based AKA solutions have their own weaknesses. While some of these include the traditional AKA weaknesses mentioned in the previous section, some are specific to the group-based nature of these approaches. For example, an attacker can pose as a member of a group and get access to the network [42].

The aforementioned requirements of 5G have also produced new concepts, and thus new security issues:

- Network slicing, which is a solution to meet heterogeneous requirements from different vertical markets [43]. Networks slices are logical networks relying on a single physical network [44]. Each network slice is composed of various network functions to provide specific capabilities and to satisfy a specific type of usage [44]. For example, in some IoT cases (e.g., a smart factory), mobility will not be very high, so it may not need mobility handling functions [44]. There can be different approaches in providing network slicing (for example, we can have a slice per service or a slice per vertical market). Different technologies like SDN (Software-defined Network) and NFV (Network Function Virtualization) will be used to deploy slicing. The references [45–47] present some proposals for network slicing architecture and implementations. Concerning security, network slicing also adds some issues out of the scope of this paper, such as slice isolation to prevent threat propagation through slices, authentication and integrity protection of input data, and access control between slices [39].
- Heterogeneous network access, an important capability because radio technologies may be used to access 5G networks. As mentioned above, one of the 5G goals is to provide better accessibility to users; therefore, when users do not have 5G connectivity, they may connect to a 5G network through other types of accesses, e.g., via satellite access. In the case of the IoT, devices may also use different radio access technologies. In these situations, the enterprises or satellite providers may have their own AAA servers, and so the management of the connection between different AAA servers, especially in roaming scenarios, will be very important [39, 48]. It will also be a major task to protect the network against unauthorized access in this heterogeneous infrastructure [49].

5 First standardized solutions: 5G phase 1 architecture

5G phase 1 is published in the 3GPP specifications release 15 in December 2017 for first deployments in 2019. It will be later followed by phase 2, for which many options are still open. Concerning 5G phase 1, 3GPP has provided a technical specification to define the first architecture of 5G systems and to specify the main nodes and their responsibilities [50]. In this architecture, control planes and user planes are separated as much as possible to achieve more flexible and scalable deployment. Instead of network entities grouping many functions, 3GPP attempted to define NFs (Network Functions) with more atomistic roles (i.e., one specific responsibility per function). However, most of these NFs are somehow a mapping of existing 4G entities. Two representations are possible for NF interactions; one of them is based on the service-oriented architecture (SOA) viewpoint and the other is based on traditional reference points. In service-based representation, an NF exposes a set of services it offers to other NFs, and it uses the services provided by other NFs. All interactions are carried by the same protocol for API invocations. Each time a new NF needs to be plugged in, only its new API should be declared to other components. In reference point representation, specific protocol links are kept between pairs of network functions. Figure 3 shows the current 5G phase 1 architecture and its network functions.

The two first defined NFs can be seen as an evolution of the HSS:

- AUSF (Authentication Server Function) provides a unified framework for authentication issues (for 3GPP access as well as non-3GPP access); and
- UDM (Unified Data Management) contains data that is related to the HSS (i.e., user data). The UDM stores only some part of the data (such as a user's subscription data) and not all of it. It also supports authentication credential processing, user identification handling, and access authorization.

Indeed, we should observe that the concept of the data in 5G is a little bit different than it is in 4G, with the differentiation between structured data and unstructured data. Structured data is exchanged between NFs in a standardized way, to enable communication between equipment from different vendors. Unstructured data is vendor-specific data that can be hidden from other network functions. Three new functions are defined in this context:

- SDSF (Structured Data Storage network function);
- UDSF (Unstructured Data Storage network function); and
- UDR (Unified Data Repository), which is responsible for storing or retrieving subscription and policy data.

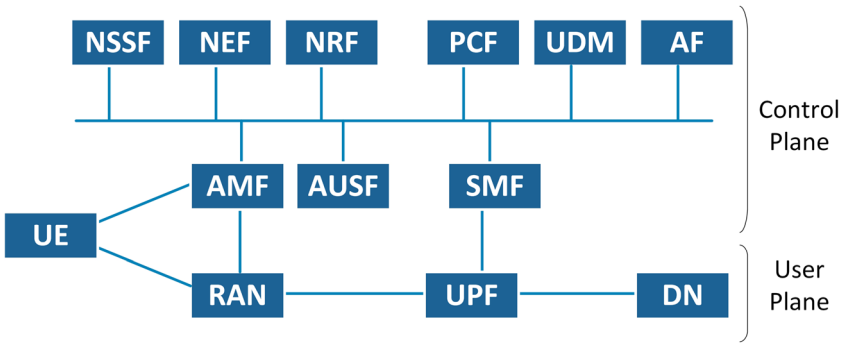

Fig. 3 5G phase 1 architecture and its main network functions. All of the NFs can connect to the UDSF, NEF, and NRF; therefore, they are not shown in the figure. RAN stands for radio access network

Two other NFs can be seen as a division of the 4G MME:

- AMF (Core Access and Mobility Management Function) has different functionalities, including access authentication and authorization, registration management, and mobility management. Since different access technologies will be used, 5G needs a common framework for access management, as well as for handling mobility between different types of access. Therefore, AMF will support both 3GPP access networks and non-3GPP access networks. Unlike 4G (where MME is used for 3GPP access and ePDG for non-3GPP access), the structure of the core network will be common for 3GPP access and non-3GPP access in the 5G system.
- SMF (Session Management Function) is responsible for session management and some other functionalities, such as the allocation of IP addresses and control of the policy enforcement and QoS (establishment of a session is totally separated from mobility management in 5G).

A function is also dedicated to policy management, as the PCRF (Policy and Charging Rules Function) in 4G was:

- PCF (Policy Control Function) is related to policy framework and provides policy rules to NFs in the control plane.

New functions are introduced to manage the instantiation of network functions and the interactions between them, in an NFV (Network Function Virtualization) approach:

- NEF (Network Exposure Function) handles all the information and services that can be exposed by NFs, for example, to 3rd parties, and the information exchanges between different NFs in the control plane.
- NRF (NF Repository Function) stores the NFs available in the system and informs other NFs about new NFs. In service-based representation, each time a new NF is added to the system, it needs to be discovered by all the other NFs.

A new function is also dedicated to network slicing:

- NSSF (Network Slice Selection Function) determines the serving AMF for the UE and selects network slice instances for it (in addition to the network slicing concept, network slice instances provide specific services to different enterprises).

Finally, generic functions represent the application plane, transfer plane, and external data network:

- AF (Application Function) provides services to 3rd parties (e.g., it establishes the QoS and some charging aspects for a service in IMS).
- UPF (User plane Function) is responsible for everything related to user data and acts as a high-performance forwarding engine for user traffic. It is geographically located closer to the end users to achieve the latency requirements.
- DN (Data Network) handles internet access or services from operators and 3rd parties.

6 Authentication and access control choices for 5G phase 1

This new 5G architecture comes with some new design choices for authentication and access control, but also brings much continuity. The most important continuity concerns the symmetric key-based authentication through a secure element. In phase 1 of the 5G standards, it was decided to keep a secure element in the UE (like the UICC in 4G and 3G and the SIM card in 2G) to process subscription credentials [5], which could also be an ESIM (Embedded SIM) provided by device makers and with which operators can provision their profile over-the-air at subscription time.

As for the differences, 5G introduces a new type of identifier, the SUPI (Subscriber Permanent Identifier), which is somehow equivalent to the IMSI but with a more global

footprint, as it can be used not only for cellular service subscribers but for different environments like the IoT. The SUPI can have different formats: IMSI and NAI (Network Access Identifier). NAI is more flexible than IMSI and it can include different identifiers (including IMSI). To protect user privacy, the MSIN part of the identifier will be encrypted with the public key of the subscriber's home network (the IMSI disclosure vulnerability is limited). This choice can be justified as follows: if all parts of the identifier were encrypted, the decryption would have to be done in the serving network in order to route the messages to the right home network. This would impose the need for a global mechanism to distribute and manage certificates as well as to control multiple public keys for different serving networks. The SUCI (Subscription Concealed Identifier) contains the concealed SUPI. The public key of the home network could be stored in the secure element of the UE. We will also have 5G-GUTI as the temporary identifier, like the GUTI in 4G systems.

Figure 4 depicts the detailed message flow in 5G-AKA procedures. As mentioned in the previous section, the authentication mechanisms in 5G systems will be done along with the same principle as in 4G systems (AKA mechanism, 5G-AKA, and EAP-AKA′) with some minor differences. These differences in AKA mechanisms will be from the network perspective only, and not from the UE perspective. AKA mechanisms in 5G systems, like those in 4G systems, use a "serving network name" (like SNid in 4G) to derive the anchor key (K_{SEAF}); thus, the anchor key will belong to the specific serving network and this serving network cannot pretend to be another serving network. AKA mechanisms offer secondary protection for 5G systems; the visited network will provide an Authentication Confirmation message to the home network and confirm that the UE's authentication is successful. Another difference in AKA mechanisms for 5G systems is that the anchor key (K_{SEAF}) that is derived in a 3GPP access can also be used in a non-3GPP access without a new authentication process. As mentioned in the previous section, 4G systems use EPS-AKA for 3GPP access and EAP-AKA for non-3GPP access, but in 5G systems, both 5G-AKA and EAP-AKA′ can be used in 3GPP access and non-3GPP access. The NAS context is needed for 5G-AKA, which is not present for non-3GPP access, and so, at the beginning of non-3GPP access, only EAP-AKA′ is foreseen.

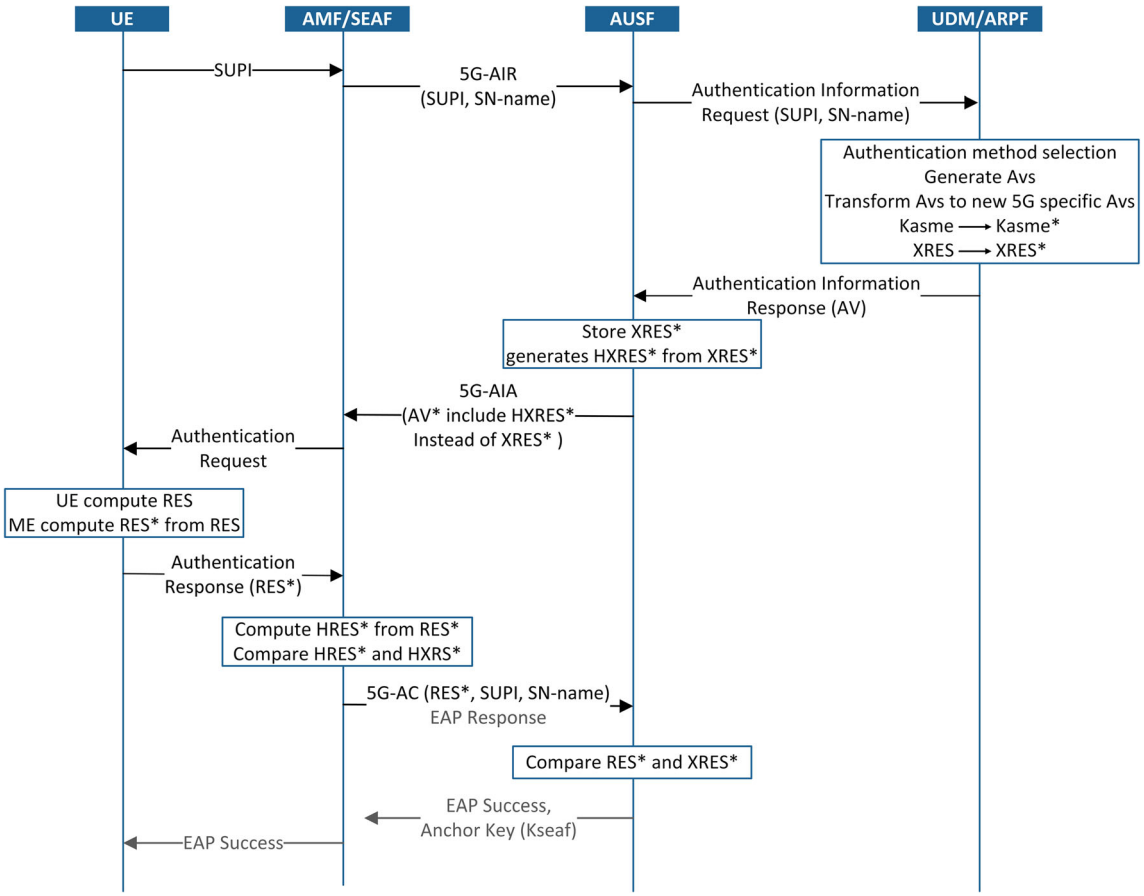

Fig. 4 5G-AKA and EAP-AKA′. The main focus is on the 5G-AKA. The computation of RES* in the ME (Mobile Equipment) is in the same way as the computation of XRES* in the ARPF and the computation of HRES* in the SEAF is in the same way as the computation of HXRES* in the AUSF

The authentication process will involve the UE, the SEAF (Security Anchor Function), the AUSF, and the UDM/ARPF (Authentication Repository and Processing Function) [51]. The SEAF will be included in the AMF, and interact with the AUSF to obtain authentication data from the UDM. It accomplishes UE authentication for different access networks. The ARPF stores subscribers' profiles and the information related to security. At the beginning of the authentication process, the UE will send its SUPI to the SEAF. Next, the SEAF will send the 5G-AIR (Authentication Initiation Request) to the AUSF. The 5G-AIR contains the SUCI or SUPI of the UE, the name of the serving network. This message also indicates that the UE uses a 3GPP access or a non-3GPP access. After receiving the authentication information request from the AUSF, the UDM/ARPF generates an AV as in 4G, and then transforms them to new AVs that are specific to 5G systems (this transformation will be different in EAP-AKA′ and 5G-AKA). In the case of the UE's successful authentication, the SEAF will send a 5G-AC (Authentication Confirmation) message in the 5G-AKA process.

These messages are useful but not adequate to protect the system against some frauds, such as fraudulent Update Location requests for subscribers [5].

It is important to observe that the authentication process shall be done outside the slice. This means that the UE should authenticate with the network and not with the slice. A UE can access to a specific slice instance through the NSSF only when its authentication with the home network is completed [5, 50, 52].

7 Conclusion

As reviewed in this paper, authentication mechanisms for 4G networks have weaknesses that make them vulnerable to various attacks. While the new AKA procedures for 5G will solve IMSI disclosure problems and mitigate the consequences of SNid disclosure, other 4G vulnerabilities will remain in 5G (GUTI potential persistence, acceptance of reject messages, capabilities' disclosure, and synchronization failure). In addition, new vulnerabilities appear with new 5G use cases (e.g., group-based authentication).

While 3GPP have provided the AKA protocols for the first phase of the 5G networks, more research is needed to design innovative authentication and access control mechanisms to better support new 5G needs (e.g., the huge number of objects in IoT connectivity, heterogeneous network access, D2D connections, as well as issues related to network slicing and openness to 3rd parties through a wholesale-oriented model), in order to ensure both network operators' and customers' security.

References

1. 3GPP (2017) Security Architecture, TS 33.102, Tech. Spec. 14.1.0
2. 3GPP (2017) Security Architecture, TS 33.401, Tech. Spec. 15.1.0
3. 3GPP (2017) Network Architecture, TS 23.002, Tech. Spec. 14.1.0
4. Cao J, Ma M, Li H, Zhang Y, Luo Z (2014) A survey on security aspects for LTE and LTE-A networks. IEEE Commun Surv Tutorials 16(1):283–302
5. 3GPP (2017) Security Architecture and Procedures for 5G System, TS 33.501, Tech. Spec. 995985
6. 3GPP (2018) Numbering, Addressing and Identification, TS 23.003, Tech. Spec. 15.6.0
7. Forsberg D, Horn G, Moeller W-D, Niemi V (2012) LTE security. Wiley
8. Tsay J-K, Mjølsnes SF (2012) A vulnerability in the umts and lte authentication and key agreement protocols. In: International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, pp 65–76
9. Abdrabou MA, Elbayoumy ADE, El-Wanis EA (2015) LTE authentication protocol (EPS-AKA) weaknesses solution. In: Intelligent Computing and Information Systems (ICICIS), 2015 IEEE Seventh International Conference on, pp 434–441
10. Park Y, Park T (2007) A survey of security threats on 4G networks. In: Globecom Workshops, 2007 IEEE, pp 1–6
11. Abdo JB, Demerjian J, Ahmad K, Chaouchi H, Pujolle G (2013) EPS mutual authentication and crypt-analyzing SPAKA. In: Computing, Management and Telecommunications (ComManTel), 2013 International Conference on, pp 303–308
12. Haddad ZJ, Taha S, Saroit IA (2017) Anonymous authentication and location privacy preserving schemes for LTE-A networks. Egypt Inform J 18:193–203
13. Li X, Wang Y (2011) Security enhanced authentication and key agreement protocol for LTE/SAE network. In: Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on, pp 1–4
14. Franklin JV, Paramasivam K (2011) Enhanced authentication protocol for improving security in 3GPP LTE networks. In: Proc. International Conference on Information and Network Technology (ICINT 2011)
15. Abdo JBB, Chaouchi H, Aoude M (2012) Ensured confidentiality authentication and key agreement protocol for EPS. In: Broadband Networks and Fast Internet (RELABIRA), 2012 Symposium on, pp 73–77
16. Fouque P-A, Onete C, Richard B (2016) Achieving better privacy for the 3GPP AKA protocol, IACR Cryptology ePrint Archive, vol 2016, p 480
17. Shaik A, Borgaonkar R, Asokan N, Niemi V, Seifert J-P (2015) Practical attacks against privacy and availability in 4G/LTE mobile communication systems, arXiv preprint arXiv:1510.07563
18. Bhasker D (2013) 4G LTE security for mobile network operators. Cyber Secur Inf Sys Inf Anal Cent(CSIAC) 1(4):20–29
19. Cichonski J, Franklin JM, Bartock M (2016) LTE architecture overview and security analysis. NIST Draft NISTIR, vol 8071
20. Hamandi K, Sarji I, Chehab A, Elhajj IH, Kayssi A (2013) Privacy enhanced and computationally efficient HSK-AKA LTE scheme. In: Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on, pp 929–934
21. Khan MSA, Mitchell CJ (2014) Another look at privacy threats in 3G mobile telephony. In: Australasian Conference on Information Security and Privacy, pp. 386–396
22. Degefa FB, Lee D, Kim J, Choi Y, Won D (2016) Performance and security enhanced authentication and key agreement protocol for SAE/LTE network. Comput Netw 94:145–163

23. Mavoungou S, Kaddoum G, Taha M, Matar G (2016) Survey on threats and attacks on mobile networks. IEEE Access 4:4543–4572
24. Choudhury H, Roychoudhury B, Saikia DK (2012) Enhancing user identity privacy in LTE," in Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, pp 949–957
25. Mjølsnes S, Tsay J-K (2012) Computational security analysis of the UMTS and LTE authentication and key agreement protocols
26. Qiang L, Zhou W, Cui B, Na L (2014) Security analysis of TAU procedure in LTE network," in P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2014 Ninth International Conference on, pp 372–376
27. Escudero-Andreu G, Raphael CP, Parish DJ (2012) Analysis and design of security for next generation 4G cellular networks. In: The 13th annual post graduate symposium on the convergence of telecommunications, networking and broad-casting (PGNET)
28. 3GPP (2009) Rationale and Track of Security Decisions in Long Term Evolved (LTE) RAN / 3GPP System Architecture Evolution, TR 33.821, Tech. Report. 1031871
29. Hamandi K, Sarji I, Elhajj IH, Chehab A, Kayssi A (2013) W-AKA: privacy-enhanced LTE-AKA using secured channel over Wi-Fi. In: Wireless Telecommunications Symposium (WTS), 2013, pp 1–6
30. Bikos AN, Sklavos N (2013) LTE/SAE security issues on 4G wireless networks. IEEE Secur Priv 11(2):55–62
31. Alt S, Fouque P-A, Macario-Rat G, Onete C, Richard B (2016) A cryptographic analysis of UMTS/LTE AKA. In: International Conference on Applied Cryptography and Network Security, pp 18–35
32. Arapinis M et al (2012) New privacy issues in mobile telephony: fix and verification. In: Proceedings of the 2012 ACM conference on computer and communications security, pp 205–216
33. Lee M-F, Smart NP, Warinschi B, Watson GJ (2014) Anonymity guarantees of the UMTS/LTE authentication and connection protocol. Int J Inf Secur 13(6):513–527
34. Othmen S, Zarai F, Obaidat MS, Belghith A (2013) Re-authentication protocol from WLAN to LTE (ReP WLAN-LTE) In: Global Communications Conference (GLOBECOM), 2013 IEEE, pp 1446–1451
35. El Idrissi YEH, Zahid N, Jedra M (2012) Security analysis of 3GPP (LTE)—WLAN interworking and a new local authentication method based on EAP-AKA. In: Future Generation Communication Technology (FGCT), 2012 International Conference on, pp 137–142
36. Mun H, Han K, Kim K (2009) 3G-WLAN interworking: security analysis and new authentication and key agreement based on EAP-AKA. In: Wireless Telecommunications Symposium, WTS 2009, 2009, pp 1–8
37. Alliance N (2015) 5G white paper, Next generation mobile networks, white paper
38. Schneider P, Horn G (2015) Towards 5G security. In: Trustcom/BigDataSE/ISPA, 2015 IEEE, vol 1, pp 1165–1170
39. 5G Ensure Project (2016) Deliverable D2.4 Security Architecture (draft)
40. Li J, Wen M, Zhang T (2016) Group-based authentication and key agreement with dynamic policy updating for MTC in LTE-A networks. IEEE Internet Things J 3(3):408–417
41. Su W-T, Wong W-M, Chen W-C (2016) A survey of performance improvement by group-based authentication in IoT. In: Applied System Innovation (ICASI), 2016 International Conference on, pp 1–4
42. Giustolisi R, Gerhmann C (2016) Threats to 5G group-based authentication. In: 13th International Conference on Security and Cryptography (SECRYPT 2016), 26–28 July 2016, Madrid, Spain
43. Foukas X, Patounas G, Elmokashfi A, Marina MK (2017) Network slicing in 5G: survey and challenges. IEEE Commun Mag 55(5): 94–100
44. Chatras B, Kwong UST, Bihannic N (2017) NFV enabling network slicing for 5G. In: Innovations in Clouds, Internet and Networks (ICIN), 2017 20th Conference on, pp 219–225
45. Ordonez-Lucena J, Ameigeiras P, Lopez D, Ramos-Munoz JJ, Lorca J, Folgueira J (2017) Network slicing for 5G with SDN/NFV: concepts, architectures, and challenges. IEEE Commun Mag 55(5):80–87
46. Katsalis K, Nikaein N, Schiller E, Ksentini A, Braun T (2017) Network slices toward 5G communications: slicing the LTE network. IEEE Commun Mag 55(8):146–154
47. Rost P, Mannweiler C, Michalopoulos DS, Sartori C, Sciancalepore V, Sastry N, Holland O, Tayade S, Han B, Bega D, Aziz D, Bakker H (2017) Network slicing to enable scalability and flexibility in 5G mobile networks. IEEE Commun Mag 55(5):72–79
48. 5G Ensure Project (2016) Deliverable D2.1 Use Cases
49. 5GPP (2017) 5G PPP Phase1 Security Landscape, white paper
50. 3GPP (2017) System Architecture for the 5G System, TS 23.501, Tech. Spec. 4356743
51. 3GPP (2017) Study of Security Aspects of the Next Generation System, TR 33.899, Tech. Report. 19482209
52. Han C-K, Choi H-K (2014) Security analysis of handover key management in 4G LTE/SAE networks. IEEE Trans Mob Comput 13(2):457–468