Security Analysis and Enhancements of 3GPP Authentication and Key Agreement Protocol

Muxiang Zhang, *Member, IEEE* and Yuguang Fang, *Senior Member, IEEE*

***Abstract*—This paper analyzes the authentication and key agreement protocol adopted by Universal Mobile Telecommunication System (UMTS), an emerging standard for third-generation (3G) wireless communications. The protocol, known as 3GPP AKA, is based on the security framework in GSM and provides significant enhancement to address and correct real and perceived weaknesses in GSM and other wireless communication systems. In this paper, we first show that the 3GPP AKA protocol is vulnerable to a variant of the so-called false base station attack. The vulnerability allows an adversary to redirect user traffic from one network to another. It also allows an adversary to use authentication vectors corrupted from one network to impersonate all other networks. Moreover, we demonstrate that the use of synchronization between a mobile station and its home network incurs considerable difficulty for the normal operation of 3GPP AKA. To address such security problems in the current 3GPP AKA, we then present a new authentication and key agreement protocol which defeats redirection attack and drastically lowers the impact of network corruption. The protocol, called AP-AKA, also eliminates the need of synchronization between a mobile station and its home network. AP-AKA specifies a sequence of six flows. Dependent on the execution environment, entities in the protocol have the flexibility of adaptively selecting flows for execution, which helps to optimize the efficiency of AP-AKA both in the home network and in foreign networks.**

***Index Terms*—Authentication, privacy, security, third generation (3G), wireless.**

I. Introduction

THE MOVEMENT toward ubiquitous wireless networking has brought about a number of security concerns among service providers and end users. The radio interface and the access to wireless service are two areas where wireless networks do not provide the same level of protection as wired networks unless additional security measures are taken. Two basic threats include interception of data on the radio interface and illegitimate access to wireless services. The interception of user data may result in loss of confidentiality of sensitive user information. The illegitimate use of service is not only of concern with respect to proper billing, but also of concern with respect to masquerading: impersonating a network operator or service provider to intercept user data on the radio interface.

Manuscript received September 22, 2003; revised January 15, 2004; accepted January 27, 2004. The editor coordinating the review of this paper and approving it for publication is G. Cao.

M. Zhang is with the Verizon Communications, Inc., Waltham, MA 02451 USA (e-mail: muxiang.zhang@verizon.com).

Y. Fang is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA.

Digital Object Identifier 10.1109/TWC.2004.842941

Security issues were not properly addressed in the first-generation (1G) analog systems. With low-cost equipment, an intruder could eavesdrop user traffic or even change the identity of mobile phones to gain fraudulent service. Given this background, security measures were taken into account in the design of second-generation (2G) digital cellular systems. The Global System for Mobile (GSM) communications was designed from the beginning with security in mind and has adopted several mechanisms [12] to provide user authentication and user data confidentiality. To prevent fraudulent use of wireless services, the GSM network authenticates the identity of a user through a challenge-response mechanism, that is, the user proves its identity by providing a response to a time-variant challenge raised by the network. When the user roams into a foreign network, the home network transfers a set of authentication data (called triplets) to the foreign network. Based on each triplet, the foreign network can authenticate the user without the involvement of the home network.

The GSM authentication and key agreement (hereafter called *GSM AKA*) is simple and has merits in several aspects. First of all, the cryptographic processing is confined to the mobile station and the home network. The serving network does not require the authentication key to compute the cryptographic response and the cipher key. This helps to minimize the trust that the home network needs to put on the serving network. Second, the home network can select its own algorithms used in the challenge-response protocol; the mobile station only needs to implement those algorithms used by the home network. Third, the home network is not online involved in every authentication process in the serving network. This mitigates the burden on the home network and reduces the overhead caused by the interactions between the serving network and the home network. Nevertheless, the weaknesses of the GSM challenge-response protocol have been uncovered over time [13], [14], [16]. Above all, authentication is only unidirectional; the user cannot authenticate the serving network. The lack of authentication of serving network allows the so-called false base station attack [17]. Furthermore, triplets can be reused indefinitely. There is no assurance provided to the user that authentication information and cipher keys are not being reused.

To address and correct real and perceived security weaknesses in GSM and other 2G systems, the Universal Mobile Telecommunication System (UMTS), an emerging standard for third generation (3G) wireless communications, has adopted an enhanced authentication and key agreement protocol resulted from the Third-Generation Partnership Project (3GPP) [1]. The protocol, known as 3GPP AKA, retains the framework of the GSM AKA and provides significant enhancement to achieve

additional goals such as mutual authentication, agreement on an integrity key between the user and the serving network, and freshness assurance of agreed cipher key and integrity key. As in the GSM AKA, the serving network authenticates the user by using authentication data (called *authentication vectors*) transferred from the user's home network. In each authentication vector, a sequence number is included, which is verified by the user to achieve freshness assurance of agreed cipher and integrity keys. To facilitate sequence number generation and verification, two counters are maintained for each user: one in the mobile station and the other one in the home network. Normally, the counter in the mobile station has a value less than or equal to the counter in the home network. When a mismatch occurs between the two counters, which may be caused by a failure in the home network, the authentication vectors generated by the home network may not be acceptable by the mobile station. Such a phenomenon is called *loss of synchronization* and resynchronization is needed to adjust the counter in the home network.

The 3GPP authentication and key agreement protocol has been scrutinized widely within wireless communications industries. To date, no serious flaw has ever been reported in the public literature. Using a formal method known as *BAN* logic [9], the designers have claimed [2] that "the goals of the protocol as they are stated in…are met by the protocol." In this paper, we show, however, that the protocol 3GPP AKA is vulnerable to a variant of false base station attack. The flaw of 3GPP AKA allows an adversary to redirect user traffic from one network to another. It also allows an adversary to use the authentication vectors corrupted from one network to impersonate other networks, hence the corruption of one network may jeopardize the entire system. The redirection attack represents a real threat since the security levels provided by different networks are not always the same. The redirection attack could also cause billing problem as the service rates offered by different networks are not always the same, either. In addition, the use of synchronization between a mobile station and its home network incurs considerable difficulty for the normal operation of the protocol, and may lead to synchronization attack as occurred in the Internet.

To solve the aforementioned problems and provide further enhancement on 3GPP AKA, we present an authentication and key agreement protocol which can defeat the redirection attack and may drastically lower the impact of network corruption. The protocol, called AP-AKA, also eliminates the need of synchronization between the mobile station and the home network. In AP-AKA, the home network does not maintain dynamic states for each individual subscriber. The mobile station can verify whether an authentication vector was indeed requested by a serving network and was not used before by the serving network. The protocol AP-AKA specifies a sequence of six flows. Each flow defines a message type and format sent or received by an entity. How the flows are actually carried out and under what conditions entities accept or reject are dependent on the execution environment. In certain scenarios, only two or three flows are carried out in a protocol execution, while in some other scenarios, all the six flows are carried out in the protocol execution. Dependent on the execution environment, entities have the flexibility of adaptively selecting flows for execution. It is in this sense that we call AP-AKA an adaptive protocol. This is different from a conventional two-party or three-party authentication and key agreement protocol in which entities usually execute all the flows specified by the protocol. It is shown that the adaptability helps to optimize the efficiency of AP-AKA both in the home network and in foreign networks.

This paper is organized as follows. Section II specifies the operation of 3GPP AKA and describes the resynchronization process. In Section III, we present two types of attacks against 3GPP AKA and discuss operational difficulty involved with sequence number management. Section IV presents the protocol AP-AKA and specifies its execution in various environments. In Section V, we analyze the security of AP-AKA against the redirection attack and examine the impact of network corruption. We also provide a comparison with other wireless security protocols. Section VI concludes the paper.

II. Description of 3GPP AKA

In 3GPP AKA, each user U, represented by a mobile station (MS), shares a secret key K and certain cryptographic algorithms with his/her home network, denoted by HN. In addition, the home network HN maintains a counter SQN_{HN} for each individual subscriber, and the user U maintains a counter SQN_{MS} in the mobile station. The initial values for SQN_{HN} and SQN_{MS} are set to zeroes. The cryptographic algorithms shared between HN and MS include three message authentication codes $f1$, $f1^*$, and $f2$ and four key generation functions $f3$, $f4$, $f5$, and $f5^*$. For generic requirements as well as an example algorithm set for these cryptographic algorithms, refer to [3].

There are three goals for the 3GPP AKA: 1) the mutual authentication between the user and the network; 2) the establishment of a cipher key and an integrity key upon successful authentication; and 3) the freshness assurance to the user of the established cipher and integrity keys. To achieve these goals, 3GPP AKA combines a challenge-response protocol identical to the GSM authentication and key agreement protocol and a sequence number-based one-pass protocol derived from the ISO standard ISO/IEC 9798-4 [15]. An overview of the 3GPP AKA protocol is given in Fig. 1.

The 3GPP AKA protocol consists of two phases: 1) the distribution of authentication data (called authentication vectors) from the home network to the serving network; and 2) the authentication and key agreement procedure between the user and the serving network. When the protocol is executed in the home network or when the serving network has unused authentication vectors for the user, the first phase is not executed.

A. Distribution of Authentication Vectors

The serving network SN invokes the procedure by sending an *authentication data request* to the home network HN, including the identity $IMSI$ of the mobile station, where authentication data request is the type of the message. Upon receipt of the message, HN sends an authentication data response back to SN, including an ordered array (denoted by $AV[1\ldots m]$ in Fig. 1) of authentication vectors. Each authentication vector, also called a quintet (the equivalent of a GSM "triplet"), consists of five

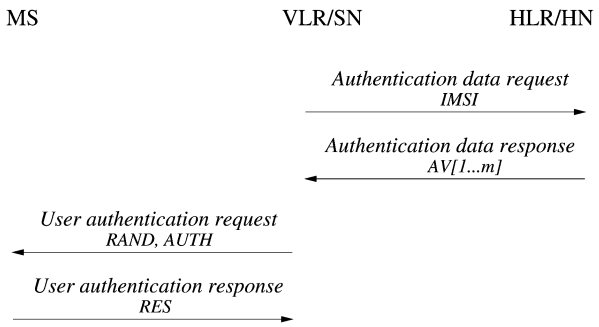


Fig. 1. 3GPP AKA.

components: a random number $RAND$; an expected response $XRES$; a cipher key CK; an integrity key IK; and an authentication token $AUTH$. Each quintet is generated according to the following steps.

1) HN generates a sequence number SQN from the counter SQN_{HN} and also generates an unpredictable random number $RAND$.
2) HN computes the following values: $XRES = f2_K(RAND)$, $CK = f3_K(RAND)$, $IK = f4_K(RAND)$, $AK = f5_K(RAND)$, and $MAC = f1_K(SQN||RAND||AMF)$, where $||$ denotes concatenation.
3) HN assembles the authentication token $AUTH = SQN \oplus AK||AMF||MAC$ and the quintet $(RAND, XRES, CK, IK, AUTH)$, where "$\oplus$" is bit-wise exclusive-or operation.
4) HN increases SQN_{HN} by 1.

The anonymity key AK is used to conceal the sequence number as the latter may expose the location of the user. If no concealment is needed, AK is set to 0. In each authentication vector, an authentication and key management field AMF is included, which serves to define operator-specific options in the authentication process, e.g., the use of multiple authentication algorithms or a limitation of key lifetime.

B. Authentication and Key Agreement

The serving network SN invokes this procedure by selecting the next unused authentication vector from the ordered array of authentication vectors in its database. Each authentication vector is good for one authentication and key agreement between the mobile station and the serving network. The serving network SN sends to MS the random number $RAND$ and the authentication token $AUTH$ from the selected authentication vector.

Upon receipt of $RAND$ and $AUTH$, MS computes the anonymity key $AK = f5_K(RAND)$ and retrieves the sequence number $SQN = (SQN \oplus AK) \oplus AK$. Then MS computes $f1_K(SQN||RAND||AMF)$ and compares this with the MAC included in $AUTH$. If they are different, MS sends a *user authentication reject* message back to SN with an indication of the cause and abandons the procedure. Otherwise, MS verifies if the received sequence number SQN is in the correct range, i.e., $SQN > SQN_{MS}$. If MS considers the sequence number to be not in the correct range, it sends a *synchronization failure* message back to SN. In this case, HN may need to resynchronize the counter SQN_{HN} maintained for the mobile station.

If the sequence number SQN is considered to be in the correct range, the authentication of the network is successful. In this case, MS computes $RES = f2_K(RAND)$ and sends it back to SN. Next, MS sets SQN_{MS} equal to SQN if $SQN_{MS} < SQN$. Lastly, MS computes the cipher key $CK = f3_K(RAND)$ and the integrity key $IK = f4_K(RAND)$.

Upon receipt of the user authentication response, SN compares RES with the expected response $XRES$ from the selected authentication vector. If RES is equal to $XRES$, the authentication of the user is successful and SN selects the cipher key CK and the integrity key IK from the selected authentication vector. If RES and $XRES$ are different, SN sends an authentication failure report to the HN and abandons the procedure.

C. Resynchronization

A synchronization failure message sent by MS includes a resynchronization token $AUTS$, which has the form $AUTS = Conc(SQN_{MS})||S_{MAC}$, where $Conc(SQN_{MS}) = SQN_{MS} \oplus f5^*_K(RAND)$, and $S_{MAC} = f1^*_K(SQN_{MS}||RAND||AMF)$. Upon receipt of the synchronization failure message, SN sends $RAND$ and $AUTS$ to the home network with an indication of the synchronization failure.

After receiving the synchronization failure indication, HN retrieves SQN_{MS} and verifies whether the value of SQN_{MS} mandates that SQN_{HN} needs to be changed, i.e., $SQN_{HN} < SQN_{MS}$. If necessary, HN also verifies the correctness of the S_{MAC} included in the resynchronization token $AUTS$ and sets SQN_{HN} equal to SQN_{MS}. Subsequently, HN sends a new batch of authentication vectors to SN. When SN receives a new batch of authentication vectors, it deletes the old authentication vectors stored for the user.

III. Security Analysis of 3GPP-AKA

Through the sequence numbers, the user is ensured that the authentication information (i.e., $RAND$ and $AUTH$) cannot be reused by an adversary or by a serving network. The serving network authenticates the user by verifying the user authentication response, RES, to determine if the user has knowledge of the authentication key. The user, however, can only verify if an authentication vector was generated by the home network. The user cannot determine if an authentication vector was requested by the serving network since the authentication vector could have been requested by any serving network. This fosters the attacks as described in the following.

A. Redirection Attack

Assume that an adversary is operating a device having the functionality of a base station. Such a device is called a false base station and is commercially available, e.g., IMSI catcher. Also assume that the adversary's device is capable of emulating a mobile station. Through the device, the adversary can impersonate as a genuine base station and entices a legitimate mobile

station to camp on the radio channels of the false base station. The adversary can also impersonate as a legitimate mobile station and establishes connection with a genuine base station. This integration of false base/mobile stations allows the adversary to relay messages in between a legitimate mobile station and a genuine base station.

Now, we assume that a user U is in the territory of his home network HN and intends to establish a communication session with the home network. After the adversary intercepts the connection attempt from the mobile station, the adversary entices the mobile station to camp on the radio channels of the false base station. Once the mobile station camps on the false base station, it is out of reach of the paging signals sent by any genuine base stations in the home network. Through the false mobile in her device, the adversary then sends a connection request to a foreign network SN on behalf of the user's mobile station. Next, the adversary faithfully relays messages between the user's mobile station and the foreign network. Authentication will be successful both in the mobile station and in the foreign network and a communication will be protected via established keys. In this way, the adversary can redirect user traffic to an unintended network.

It is worth pointing out that the redirection attack has practical implications. According to [1], data encryption is not mandatory in every network, while data integrity is mandatory in all networks. To intercept user traffic, the adversary may redirect user traffic to a network in which data encryption is either not provided or provided but very weak. The threat of this attack is particularly evident as the user roams into the border of two different networks. It might be argued that the risk could be mitigated if all the networks "trust" each other and use a commonly agreed strong encryption algorithm. Nevertheless, the redirection attack could cause billing problem; the user is in the territory of his home network but gets charged by a foreign network based on a rate higher than that offered by the home network.

B. Active Attack in Corrupted Networks

In 3GPP AKA, authentication vectors are transferred between and within networks. Each network is operated under a different administration. When a network is corrupted, an adversary could forge an authentication data request from the corrupted network to obtain authentication vectors for any user, independent of the actual location of the user. Then, the adversary could use the obtained authentication vectors to impersonate uncorrupted networks and to mount false base station attack against legitimate users. In addition, by flooding authentication data requests to the home network, the adversary could force the counter SQN_{HN} maintained for a subscriber to be set to a high value. As the maximum value of SQN_{HN} is limited, this shortens the lifetime of the mobile station.

Since the corruption of one network may jeopardize the entire system, it is critical that security measures are in place in every network. Although mechanisms are currently being developed to provide security between and within operators' networks, it is unlikely that network-wide security will be implemented in every operator's network at the same time. Situations where network operators A and B have reached agreement on the deployment of network-wide security, but A and C have not, may persist for a long time. This fosters both passive and active attacks in those unprotected networks and impedes the normal operation of 3GPP AKA.

C. Operational Difficulty With Sequence Numbers

In 3GPP AKA, it is required that the home network maintain a counter for each subscriber. Unlike an authentication key whose value is fixed, the value of a counter is dynamic. If there is a crash in the database storing the counters in the home network, it will affect all the mobile stations subscribed to the home network. In such an instance, the cost will be tremendous to resynchronize the counters maintained for each individual subscriber. Furthermore, even if there is no fault in the database, the normal operation of the 3GPP AKA protocol could cause resynchronization requests to the home network. As described in Section II-C, a resynchronization is requested by the mobile station, not by the home network. Whenever a sequence number is considered to be not in the correct range, the mobile station decides that a synchronization failure has occurred in the home network and consequently initiates a resynchronization request to the home network. This may produce spurious resynchronization requests, as the fact that a sequence number is not in the correct range does not necessarily mean a failure in the counter SQN_{HN}. It might be caused by an adversary by replaying a pair of used $RAND$ and $AUTH$. The out-of-order use of authentication vectors in the serving network could also cause synchronization failure. An array of authentication vectors may arrive at the SN out of order, i.e., the initial ordering of the authentication vectors may be disturbed on their way from the HN to the SN. In addition, the user movement between different $VLRs$ which do not exchange authentication information could cause synchronization failures, too. When the user returns to a previously visited VLR_o from the newly visited VLR_n, the authentication and key agreement procedures based on unused authentication vectors in VLR_o may cause synchronization failures. The mobile station, however, cannot discern the actual reason for the sequence number being not in the correct range. Spurious resynchronization adds extra cost to the signaling between the serving network and the home network and may cause deletion of those unused authentication vectors.

IV. Adaptive Protocol for Mobile Authentication and Key Agreement

To address the security issues involved with 3GPP AKA, in this section, we present a new authentication and key agreement protocol, called AP-AKA, which can defeat the redirection attack and may drastically lower the impact of network corruption. An overview of AP-AKA is given in Fig. 2.

The protocol AP-AKA retains the framework of the 3GPP AKA protocol but eliminates the synchronization between the mobile station and its home network. In AP-AKA, each mobile station and its home network share an authentication key K and three cryptographic algorithms F, G, and H, where F and H are message authentication codes and G is a key generation function. In practice, the authentication key is usually generated by the home network and programmed into the mobile station during service provisioning. Unlike in 3GPP AKA,

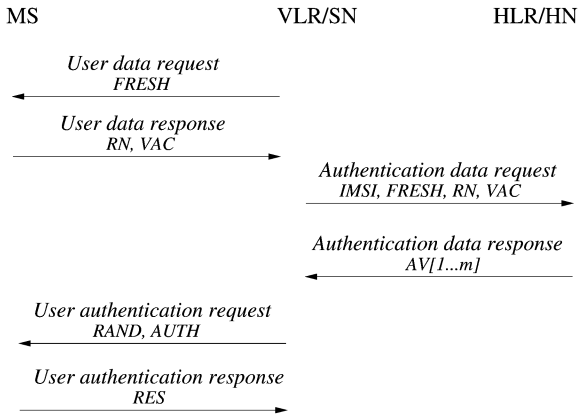


Fig. 2. Overview of AP-AKA.

however, the home network in AP-AKA does not maintain a dynamic state, e.g., the counter, for each individual subscriber. The mobile station can verify whether an authentication vector was indeed requested by a serving network and was not used before by the serving network. The protocol AP-AKA specifies a sequence of six flows. Each flow defines a message type and format sent or received by an entity. Depending on the execution environment, entities have the flexibility of adaptively selecting flows for execution. It is in this sense that we call AP-AKA an adaptive protocol. In the following, we specify the operation of AP-AKA in various execution environments. In Section V, we analyze the security of AP-AKA against both passive and active attacks.

A. Protocol Execution in Foreign Networks

Assume that a user, represented by a mobile station MS, roams into a foreign network SN. Depending on whether SN has unused authentication vectors for the user, the protocol execution may be carried out in two different ways.

1) If SN does not have unused authentication vectors for the user, all the six flows in Fig. 2 will be carried out. SN starts the protocol by sending a user data request to the user. A random number, denoted by $FRESH$, is included in the request. After receiving the request, the mobile MS generates a random number RN and computes a message authentication code, denoted by VAC

$$VAC = F_K(FRESH||RN||ID_{SN})$$

where ID_{SN} denotes the identity of SN. The mobile station MS then sends a user data response back to SN including RN and VAC. Subsequently, SN sends an authentication data request to the home network HN, including $IMSI$, $FRESH$, RN, and VAC. Upon receipt of the authentication data request from SN, HN retrieves the secret key K of the user and verifies the correctness of the received VAC. If the verification fails, HN sends back a *reject notification* including $IMSI$, $FRESH$, and RN, and SN aborts the connection. If the verification succeeds, HN generates a batch of m authentication vectors, denoted by $AV[1 \dots m]$ in Fig. 2, and sends them back to SN. Each authentication vector consists of four components $(RAND, XRES, SK, AUTH)$ and is indexed by an integer idx, $1 \leq idx \leq m$. The index number describes the order of the authentication vector in the batch. To generate an authentication vector, HN proceeds as follows:

   a) HN allocates an index number, $1 \leq idx \leq m$, for the authentication vector and generates a random number $RAND$;
   b) HN computes the following values: an expected response $XRES = F_K(RAND)$; a session key $SK = G_K(RAND)$; and two message authentication codes

$$RN_{idx} = H_K(idx||RN),$$
and
$$MAC = F_K(RAND||idx||RN_{idx});$$

   c) HN assembles the authentication token $AUTH = idx||RN_{idx}||MAC$ and the authentication vector $(RAND, XRES, SK, AUTH)$.

After receiving the authentication vectors from HN, SN takes out one of them from the batch and stores the rest in its database. Then SN sends a user authentication request to the mobile station, including $RAND$ and $AUTH$ from the selected authentication vector. After receiving the user authentication response from the mobile station, SN compares if $RES = XRES$. If not, the authentication of the user fails and SN abandons the connection. Otherwise, the user authentication is successful and the agreed session key is the SK from the selected authentication vector, where SK may be a concatenation of a cipher key and an integrity key.

2) If SN has unused authentication vectors for the user, it selects an unused authentication vector from its database and starts the protocol by executing the fifth flow in Fig. 2, i.e., sending a user authentication request to the mobile station, including $RAND$ and $AUTH$ from the selected authentication vector. After receiving the user authentication response, SN compares if $RES = XRES$ and proceeds as described earlier.

B. Protocol Execution in the Home Network

If HN has unused authentication vectors for the user, the protocol execution is carried out in the same way as in the foreign network, i.e., HN selects an unused authentication vector from its database and starts the protocol by sending a *user authentication request* to the mobile station. If HN does not have unused authentication vectors for the user, it executes a three-flow protocol as described in Fig. 3.

In the three-flow protocol, HN starts by sending a *user data request* to the mobile station, including a random number $FRESH$. After receiving the request, the mobile station MS generates a random number RN and computes $VAC = F_K(FRESH||RN||ID_{HN})$. Next, MS sends a *user data response* back to HN, including RN and VAC.

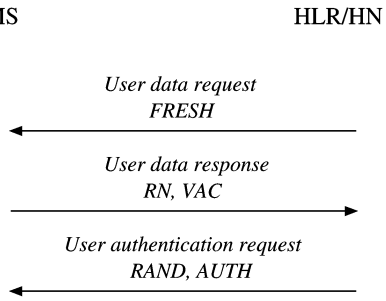


Fig. 3. Execution of AP-AKA in the home network.

After receiving the *user data response*, HN verifies if the received VAC is equal to $F_K(FRESH||RN||ID_{HN})$. If the verification fails, HN abandons the connection. Otherwise, the authentication of the user is successful and HN proceeds as follows:

1) HN computes the session key $SK = G_K(RAND)$;
2) HN generates a random number $RAND$ and computes the following values:

$$RN_0 = H_K(0||RN),$$

and

$$AUTH = 0||RN_0||F_K(RAND||0||RN_0);$$

3) HN sends a user authentication request to MS, including $RAND$ and $AUTH$;
4) HN may also generate a batch of authentication vectors for future use. Each authentication vector consists of a different index number greater than 0.

Note that in the three-flow protocol, the authentication token $AUTH$ has the same format as that included in an authentication vector except that an index value of zero is used. Also note that the mobile station does not send the *user authentication response* when network authentication is successful.

C. User Actions

The user (i.e., the mobile station) acts as a *responder* in the protocol. When the mobile station receives a user data request, it generates a random number RN, and computes a message authentication code $VAC = F_K(FRESH||RN||ID_{SN})$. Then, it replies back with RN and VAC. When the mobile station receives a *user authentication request*, it retrieves $RAND$ and $AUTH$ from the request and verifies the correctness of the MAC included in $AUTH$. If it is incorrect, the authentication of the network fails and the mobile station aborts the connection. Otherwise, the mobile station further verifies if the RN_{idx} included in $AUTH$ is acceptable, that is, $RN_{idx} = H_k(idx||RN)$ and RN_{idx} was not used before by the network. If RN_{idx} is not acceptable, the mobile station aborts the connection. Otherwise, the authentication of the network is successful. In the case of successful network authentication, the mobile station computes the session key $SK = G_K(RAND)$ and proceeds as follows:

1) if $idx > 0$, the mobile station updates its internal state and replies back with $RES = F_K(RAND)$;
2) if $idx = 0$, the mobile station does not send the user authentication response.

To facilitate fast verification of RN_{idx}, the mobile station maintains the usage information on RN_{idx}. The usage information is used to verify that the authentication vector containing RN_{idx} was indeed requested by the serving network and was not used before by the serving network. This helps to prevent redirection attacks. It also helps to defeat replay attacks by exploiting previously used authentication vectors.

D. Verification of RN_{idx}

In AP-AKA, each number RN_{idx} can only be used once. For convenience, we refer to RN_{idx} as a nonce. To support fast verification of RN_{idx}, the mobile station maintains a list of unused nonces for every visited network. Each list is pointed by the identity of the network. Assume that the home network generates a batch of m authentication vectors in response to each authentication data request. After receiving a user data request from a network N_i with identity ID_{N_i}, the mobile station replies back with RN and VAC. Then the mobile station computes a sequence of nonces $RN_r = H_K(r, RN)$, $r = 0, 1, \ldots, m$, and adds the computed nonces to the list pointed by ID_{N_i}. When the mobile station verifies a nonce RN_{idx} received from the network N_i, the mobile station only needs to check if RN_{idx} is in the list pointed by ID_{N_i}. If not, RN_{idx} is not acceptable; otherwise, it is acceptable. In the case of acceptance, the mobile station removes RN_{idx} from the list.

E. Discussions

The protocol AP-AKA may seem more efficient if it is started by the mobile station. For instance, when the mobile station initiates a connection request to a serving network SN, it also includes a random number RN and a message authentication code U_{MAC} in the connection request, where $U_{MAC} = F_K(RN||ID_{SN})$. Then, the serving network SN requests authentication vectors by sending ID_U, RN, and U_{MAC} to HN. This modification, however, makes it difficult for the serving network to reauthenticate the user, since the serving network cannot request new authentication vectors during a communication session. Moreover, the modified protocol is susceptible to a denial-of-service attack, that is, an adversary impersonates the mobile station and starts the protocol by sending a used pair (RN, U_{MAC}) to SN. In subsequent protocol execution, the mobile station will reject a newly requested authentication vector by SN since RN_{idx} included in the authentication vector was used before. The denial-of-service attack causes considerable overhead both on HN and on SN and impedes the normal operation of the modified protocol.

In AP-AKA, the protocol execution in the home network HN is different than the protocol execution in a foreign network SN. The purpose is to make AP-AKA efficient both in the home network and in foreign networks. Note that the protocol executions in both types of networks could be made identical by slightly modifying the protocol execution in HN, that is, HN sends a user data request, the mobile station replies with a user data response, HN then sends a user authentication request and the mobile station replies back with a user authentication response. This arrangement, however, needs to carry out four flows and is less efficient than the three-flow protocol. Also note that the home network does not have to generate authentication vectors

for its subscribers. As the use of authentication vectors involves only two flows, the home network may decide to generate authentication vectors for those frequently visiting subscribers.

In comparison with 3GPP AKA, the protocol AP-AKA does not mandate that the home network maintains a dynamic state for each individual subscriber. This relieves the burden on the home network and resolves operational difficulty involved with resynchronization. During the distribution stage of authentication vectors, however, the protocol AP-AKA introduces two additional local-exchanges between the mobile station and the serving network. Considering that multiple authentication vectors are transferred from the home network to the serving network, the overhead on each authentication vector is increased only to a small extent. To see this, assume that the home network generates a batch of m authentication vectors in response to each authentication data request. Also assume that all the authentication vectors will be used by the serving network. In a foreign network, the average number of flows carried out in each protocol execution is $\eta_{SN} = (6 + 2(m-1))/m = 2 + 4/m$, e.g., $\eta_{SN} = 2.8$ when $m = 5$. In the home network, $\eta_{HN} = 3$ if the home network does not generate authentication vectors; otherwise, $\eta_{HN} = (3+2m)/(m+1)$, e.g., $\eta_{HN} \approx 2.17$ when $m = 5$. For 3GPP AKA, $\eta_{SN} = (4+2(m-1))/m = 2+2/m$ and $\eta_{HN} = 2$. So, AP-AKA is slightly more costly in terms of signaling than 3GPP AKA.

V. Analysis of AP-AKA

In the protocol AP-AKA, the network authenticates the mobile station either by verifying VAC included in the user data response or by verifying RES included in the user authentication response. The mobile station authenticates the network through the MAC included in the user authentication request. If the verification succeeds, the mobile station is ensured that the network is either the home network or a serving network authorized by the home network to provide the service. In addition, by verifying the RN_{idx} included in $AUTH$, the mobile station is assured that the authentication vector was requested by the serving network and was not used before by the serving network. In the following, we analyze the security of AP-AKA against the redirection attack and examine the impact of network corruption. We also provide comparison with other wireless security protocols.

A. Security Against Redirection Attack

In AP-AKA, an authentication vector generated by the home network can only be used by a specific serving network since the identity of the serving network is involved in the generation and verification of the authentication vector, while in 3GPP-AKA, an authentication vector can be used by any serving network. Whenever a mobile station receives a *user data request* and a random number $FRESH$ from a serving network SN, it replies back with another random number RN and a message authentication code VAC providing integrity for $FRESH$, RN, and ID_{SN}. In addition, the mobile station maintains a record (RN, ID_{SN}) in its database. By verifying VAC, the home network is ensured that the user is indeed in the territory of SN. When generating authentication vectors, the home network inserts a number RN_{idx}, which is derived from RN and an index, into each authentication vector. After receiving a user authentication request, the user can determine if the request is sent by the serving network SN, not by other serving networks, since the user can verify if the number RN_{idx} included in the request can be derived from the random number RN sent to the serving network SN. To show how this helps to defeat the redirection attack, let us assume that, when the user roams to a foreign network SN, the mobile station displays an indication to the user if he/she is roaming. If the user answers yes, the mobile station starts initiating a connection request to SN; otherwise, the mobile station disconnects. This feature was implemented in some of the 2G phones, e.g., Qualcomm PDQ phones. It is expected that 3G phones will even display the brand name of the serving network during roaming. This makes the user aware of the roaming services provided to him/her.

Now, assume that the user is in the territory of the home network HN, which usually has national footprint. To hook up the user's mobile station to a false base station, the false base must broadcast the identity ID_{HN} of HN in a broadcasting channel. Otherwise, the mobile station will indicate to the user that he/she is roaming and displays the name of the serving network. A vigilant user can detect this fraud and the mobile station will disconnect. After receiving a connection request from the mobile station, the adversary forward the connection request to a foreign network SN via a false mobile station. After SN receives the connection request, it generates a random number $FRESH$ and sends it back to the adversary. The adversary then forwards $FRESH$ to the user's mobile station. The mobile station generates a random number RN and computes $VAC = F_K(FRESH, RN, ID_{HN})$ and sends RN, VAC back to the adversary. The adversary then forwards RN and VAC to SN. Next, SN sends $IMSI$, $FRESH$, RN, VAC to the home network HN to request for authentication vectors. After receiving the request, HN computes $XVAC = F_K(FRESH, RN, ID_{SN})$. Note that ID_{SN} is used in the computation of $XVAC$ since the request of authentication vectors is from SN. It is obvious that $XVAC$ is not equal to VAC. So HN will decline the request for authentication vectors.

In this discussion, we assume that the user interacts with the mobile station to detect a fraudulent network identity. This is actually not needed when the mobile station is equipped with global positioning system (GPS). A GPS-enabled mobile station can be provisioned such that it has knowledge of the identity (or identities) of the designated serving network(s) in each geographic location. In addition, the mobile station can also include the position information POS in the user data response, that is, the mobile station replies with RN, POS, and VAC when it receives a user data request from SN, where $VAC = F_K(FRESH, RN, POS, ID_{SN})$. After receiving an authentication data request from SN including $IMSI$, $FRESH$, RN, POS, and VAC, the HN decides if SN is the authorized serving network in the region specified by POS. If not, HN drops the request. Otherwise, HN further verifies the correctness of VAC. If the value of VAC is correct, HN sends back a batch of authentication vectors to SN.

B. Impact of Network Corruption

For 3GPP AKA, we have shown that the corruption of one network (either a foreign network or a home network) affects the security of the entire system. We now examine the impact of network corruption on AP-AKA. Assume that every network has established a communication channel with every other network, that is, communications between two networks go through a dedicated circuit.

Assume that a serving network SN is corrupted. The adversary can eavesdrop any message sent or received by SN. In addition, the adversary can concoct a message and sends it to any other network through SN. Since the adversary can obtain the authentication vectors generated or received by SN, it certainly can impersonate SN to establish a communication session with a user roaming into SN. It can also impersonate any subscriber of SN. Let us assume that a user U, whose home network HN is not corrupted, roams into a network SN' which is not corrupted, either. To impersonate the user U, the adversary must send back a correct response RES corresponding to a user authentication request sent by SN'. The adversary, however, cannot derive RES from the corrupted network SN since the authentication vector containing RES was transferred between HN and SN'. So the adversary cannot impersonate the user U. Next, let us see if the adversary can impersonate the network SN'. The following are the two possible scenarios.

1) Assume that the user visited the corrupted network SN before and the adversary has an unused authentication vector which was requested by SN. Then the adversary may decide to use the authentication vector to impersonate the network SN'. As has been discussed in Section V-A, the impersonation attempt will fail as the user can verify that the authentication vector was not requested by SN'.
2) Assume that adversary does not have an authentication vector for the user. The adversary starts the protocol by sending $FRESH$ and $ID_{SN'}$ to the mobile station and the mobile station replies back with RN and VAC. Then, the adversary requests authentication vectors for the user through the corrupted network SN. By verifying VAC, however, the home network can determine that the user is not in SN and hence refuse the request.

From this analysis, we see that the corruption of a serving network only affects those users who either subscribe to or roam into the corrupted network; the adversary cannot impersonate the uncorrupted serving networks. Thus, the impact of network corruption is drastically mitigated on AP-AKA in comparison with 3GPP AKA.

C. Comparison With Other Proposals

Over the last decade, numerous security schemes specifically designed for wireless networks have been proposed, e.g., [4]–[7], [18]–[21]. Many of the proposed protocols were designed based on *ad hoc* approaches (i.e., breaking and fixing) and have been found containing various flaws (see, e.g., [8], [10], and [11]). Based on the underlying cryptographic primitives, the proposed protocols can be divided into two classes: public-key protocols and symmetric-key protocols. In the public-key protocols, it is typically assumed that the user or the network can verify its partner's public key certificate. This requires the construction of large-scaled (or even, global) public key infrastructure in order to support global roaming. There are also concerns on the processing power of mobile stations and the bandwidth consumption in exchanging public keys of large size.

In most of the proposed symmetric-key protocols, e.g., [7], [14], [18], and [20], a foreign network basically acts as a proxy of the home network. The home network is online involved in every authentication process in the foreign network. In addition, most of the symmetric-key protocols were designed in the three-party setting and might not be efficient in the home network. In [18], the authors suggested a construction which consists of two separate protocols: one protocol is designed in the two-party setting and is running in the home network, another protocol is designed in the three-party setting and is running in foreign networks. With this construction, the protocol can be optimized both for the home network and for foreign networks. This construction, however, requires that the mobile station recognizes the identity of the home network. When the home network is merged into another network, the mobile station may not be able to authenticate the merged network although it has knowledge of the user's authentication key. Another concern on this construction is the interaction between the two component protocols. Such interaction may induce security implications (see [22]) on either of them.

As described in Section IV, the protocol AP-AKA can be executed efficiently both in the home network and in foreign networks. The mobile station does not need to determine whether it is in a foreign network or in the home network. Based on each authentication vector, the mobile station and the serving network can perform authentication and key agreement without the involvement of the home network. In addition, the protocol AP-AKA allows the home network to select its own algorithms used in the protocol. The benefit is that the mobile station does not need to implement all the algorithms used by serving networks, it only needs to implement those algorithms used by its home network.

VI. Conclusion

This paper investigates the security of the 3GPP AKA protocol and examines the operational difficulty involved with the sequence number management. The 3GPP AKA protocol is based on the framework of GSM AKA and intends to defeat real and perceived attacks, especially the so-called false base station attack. In this paper, we demonstrate that 3GPP AKA is vulnerable to a variant of false base station attack. The vulnerability allows an adversary to redirect user traffic from one network to another. It also allows an adversary to use authentication vectors from a corrupted foreign network to impersonate other networks. In addition, the use of synchronization between a mobile station and its home network incurs considerable difficulty for the normal operation of 3GPP AKA.

To resolve the security problems and provide further enhancement on 3GPP AKA, we present a new authentication and key

agreement protocol which can defeat the redirection attack and drastically mitigates the impact of network corruption. The protocol AP-AKA also eliminates the synchronization between the mobile station and the home network. Depending on the execution environment, entities in the protocol have the flexibility of adaptively selecting flows for execution. We show that the adaptability helps to optimize the efficiency of AP-AKA in both the home network and foreign networks.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] 3rd Generation Partnership Project; Technical Specification Group SA; 3G Security, "Security Architecture, version 4.2.0, Release 4," 3GPP, TS 33.102, 2001.

[2] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security, "Formal analysis of the 3G authentication protocol, version 3.1.0," 3GPP, TR 33.902, 1999.

[3] 3rd Generation Partnership Project; Technical Specification Group SA; 3G Security, "Report on the evaluation of 3GPP standard confidentiality and integrity algorithms, version 1.0.0, 2000-12," 3GPP, TR 33.909, 1999.

[4] A. Aziz and W. Diffie, "Privacy and authentication for wireless local area networks," *IEEE Personal Commun.*, vol. 1, no. 1, pp. 25–31, 1st Qtr. 1994.

[5] M. J. Beller, L.-F. Chang, and Y. Yacobi, "Privacy and authentication on a portable communication system," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 6, pp. 821–829, Aug. 1993.

[6] M. Beller and Y. Yacobi, "Fully-fledged two-way public key authentication and key agreement for low-cost terminals," *Electron. Lett.*, vol. 29, pp. 999–1001, 1993.

[7] R. Bird, I. Gopal, A. Herzberg, P. Janson, S. Kutten, R. Molva, and M. Yung, "The kryptoknight family of light-weight protocols for authentication and key distribution," *IEEE/ACM Trans. Networking*, vol. 3, pp. 31–41, 1995.

[8] C. Boyd and A. Mathuria, "Key establishment protocols for secure mobile communications: A selective survey," in *Proc. Australian Conf. Information Security and Privacy (ACISP'98)*, vol. 1438, Lecture Notes in Computer Science, 1998, pp. 344–355.

[9] M. Burrows, M. Abadi, and R. Needham, "A logic of authentication," *ACM Trans. Comput. Syst.*, vol. 8, pp. 18–36, 1990.

[10] L. Buttyan, C. Gbaguidi, S. Sttmann, and U. Wilhelm, "Extensions to an authentication technique proposed for global mobility network," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 373–376, Mar. 2000.

[11] U. Carlsen, "Optimal privacy and authentication on a portable communications system," *Oper. Syst. Rev.*, vol. 28, pp. 16–23, 1994.

[12] *European Telecommunications Standards Institute (ETSI), GSM 02.09: Security Aspects*, June 1993.

[13] L. Harn and H. Lin, "Modifications to enhance the security of GSM," in *Proc. 5th Nat. Conf. Information Security*, Taipei, Taiwan, R.O.C., May 1995, pp. 74–76.

[14] C. H. Lee, M. S. Hwang, and W. P. Yang, "Enhanced privacy and authentication for the global system for mobile communications," *Wireless Networks*, vol. 5, pp. 231–243, 1999.

[15] "Information technology-security techniques-entity authentication-Part 4: Mechanisms using a cryptographic check function," ISO/IEC, 9798-4.

[16] H. Lin and L. Harn, "Authentication protocols for personal communication system," in *Proc. ACM Special Interest Group on Data Communications (SIGCOMM'95)*, Aug. 1995, pp. 256–261.

[17] C. Mitchell, "The security of the GSM Air Interface Protocol," Univ. of London, Royal Holloway, RHUL-MA-2001-3, 2001.

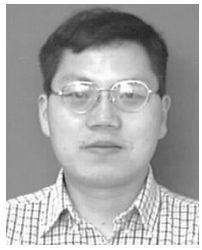
[18] R. Molva, D. Samfat, and G. Tsudik, "Authentication of mobile users," *IEEE Network*, vol. 8, no. 2, pp. 26–34, Mar./Apr. 1994.

[19] Y. Mu and V. Varadharajan, "On the design of security protocols for mobile communications," in *Proc. Australian Conf. Information Security and Privacy (ACISP'96)*, vol. 1172, Lecture Notes in Computer Science, 1996, pp. 134–145.

[20] C. Park, K. Kurosawa, T. Okamoto, and S. Tsujii, "On key distribution and authentication in mobile radio networks," in *Proc. Advances in Cryptology-Eurocrypt*, vol. 765, Lecture Notes in Computer Science, 1993, pp. 461–465.

[21] M. Tatebayashi, N. Matsuzaki, and D. B. J. Newman, "Key distribution protocol for digital mobile communication systems," in *Proc. Advances in Cryptology-Crypto '89*, vol. 435, Lecture Notes in Computer Science, 1989, pp. 324–334.

[22] W. Tzeng and C. Hu, "Inter-protocol interleaving attacks on some authentication and key distribution protocols," *Inf. Processing Lett.*, vol. 69, pp. 297–302, 1999.



Muxiang Zhang (M'01) received the Ph.D. degree in computer science from Northeastern University, Boston, MA, in June 2000.

He is currently a Distinguished Member of the Technical Staff of Verizon Communications, Inc. Waltham, MA, where he has been involved in the design of security architectures for next generation telecommunications networks, broadband fixed wireless, wireless hotspots, and fiber to the premise. His research interests include analysis and design of efficient encryption algorithms, and authentication and key exchange protocols in wireless communication networks.

Yuguang Fang (S'92–M'94–SM'99) received the B.S. and M.S. degrees in mathematics from Qufu Normal University, Qufu, Shandong, China, in 1984 and 1987, respectively, the Ph.D degree in systems and control engineering from the Department of Systems, Control and Industrial Engineering, Case Western Reserve University, Cleveland, OH, in 1994, and the Ph.D degree in electrical engineering from the Department of Electrical and Computer Engineering, Boston University, Boston, MA, in 1997.

From 1987 to 1988, he held research and teaching positions in both the Department of Mathematics and the Institute of Automation at Qufu Normal University. From September 1989 to December 1993, he was a Teaching/Research Assistant in the Department of Systems, Control and Industrial Engineering, Case Western Reserve University, where he held a Research Associate position from January 1994 to May 1994. From June 1994 to August 1995, he held a Postdoctoral position in the Department of Electrical and Computer Engineering, Boston University. From September 1995 to May 1997, he was a Research Assistant in the Department of Electrical and Computer Engineering, Boston University. From June 1997 to July 1998, he was a Visiting Assistant Professor in the Department of Electrical Engineering, University of Texas, Dallas. From July 1998 to May 2000, he was an Assistant Professor in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark. In May 2000, he joined the Department of Electrical and Computer Engineering at University of Florida, Gainesville, Florida, where, since August 2003, he has been an Associate Professor. He has published over 100 papers in refereed professional journals and conferences. He is an Editor for *ACM Wireless Networks*, an Area Editor for *ACM Mobile Computing and Communications Review*, and an Associate Editor for *Wiley International Journal on Wireless Communications and Mobile Computing*. He is also actively involved with many professional conferences such as ACM MobiCom'01, IEEE INFOCOM'98, INFOCOM'00, INFOCOM'03, INFOCOM'04, INFOCOM'05 (Technical Program Vice-Chair), IEEE Globecom'02, Globecom'03, Globecom'04 (Program Vice Chair), IEEE WCNC'99, WCNC'00 (Technical Program Vice-Chair), WCNC'02, WCNC'04, and the IC3N'98 (Technical Program Vice-Chair). His research interests span many areas including wireless networks, mobile computing, mobile communications, automatic control, and neural networks.

Dr. Fang is an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, an Editor for IEEE JOURNAL ON WIRELESS COMMUNICATIONS. He is Feature Editor for Scanning the Literature in IEEE PERSONAL COMMUNICATIONS. He has received the prestigious National Science Foundation Faculty Early Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002. He is listed in Marquis *Who's Who in Science and Engineering*, *Who's Who in America*, and *Who's Who in World*. He is a Member of ACM.