An Advanced Persistent Threat in 3G Networks: Attacking the Home Network from Roaming Networks

Christos Xenakis, Christoforos Ntantogian

Department of Digital Systems, University of Piraeus, Greece

xenakis@unipi.gr, dadoyan@unipi.gr

Abstract

The HLR/AuC is considered to be one of the most important network elements of a 3G network. It can serve up to five million subscribers and at least one transaction with HLR/AuC is required for every single phone call or data session. This paper presents experimental results and observations that can be exploited to perform a novel distributed denial of service attack in 3G networks that targets the availability of the HLR/AuC. More specifically, first we present an experiment in which we identified and proved some zero-day vulnerabilities of the 3G network that can be exploited by malicious actors to mount various attacks. For the purpose of our experiment, we have used off-the-shelf infrastructure and software, without any specialized modification. Based on the observations of the experiment, we reveal an Advanced Persistent Threat (APT) in 3G networks that aims to flood an HLR/AuC of a mobile operator. We also prove that the discovered APT can be performed in a trivial manner using commodity hardware and software, which is widely and affordably available.

1 Introduction

Universal mobile telecommunication system (UMTS) is a third generation (3G) mobile service technology that extends the general packet radio service/global system for mobile communication (GPRS/GSM) networks, supporting higher data rates and multimedia services. In UMTS, voice services are provided by the mobile switching center (MSC), while the packet data services of a mobile station (MS) are provided by the serving GPRS support node (SGSN). Both entities (we use the notation MSC/SGSN to refer to both MSC and SGSN in the rest of the paper) communicate with the home location register/authentication center (HLR/AuC), which is a database where information about the operator’s subscribers is stored. MS may access UMTS services either from its home network or from a roaming network. Roaming in UMTS is defined as the ability for an MS to make and receive voice calls, send and receive data, or access other services when traveling outside the geographical coverage area of the home network, by means of using a visited network, which may be located in the same country (i.e., national roaming) or another country (i.e., international roaming). For the operation of roaming, an agreement is required between the home network of the mobile user and the serving network of the visited area, which includes authentication, authorization, and billing services for the roaming users. As mentioned in [1], the majority of the mobile network operators that belong in the GSM association, which currently includes more than 800 mobile network operators from 220 Countries, have signed roaming agreements with each other. The GSM association also outlines the content of such roaming agreements in standardized form for its members [1].

When an MS moves outside its home network, it should perform a registration procedure with the visited/roaming network. In particular, a challenge/response protocol named *authentication and key agreement* (AKA) is executed between the MS and the roaming MSC/SGSN [2]. In the AKA protocol, the roaming MSC/SGSN sends an *authentication data request* (ADR) message to the home HLR/AuC of the MS to fetch fresh authentication credentials, named *authentication vectors* (AVs), for the MS [3]. This message includes the *international mobile subscriber identity* (IMSI) that uniquely identifies the MS in HLR/AuC. The latter generates a batch of L different AVs and sends them to the roaming MSC/SGSN. Having received them, MSC/SGSN provides to the MS one AV (which is used only in the specific authentication and is deleted afterwards) and stores the remaining (L – 1) AVs to serve future authentication requests by the MS. The reason HLR/AuC generates a batch of L different AVs, instead of only one as required for an authentication request, is to avoid executing an ADR and the related burden of generating AVs in HLR/AuC, each time MS access the MSC/SGSN. In this paper, we explore a feasible attack that considerably increases this burden.

The HLR/AuC is considered to be one of the most important network elements of a 3G network. In its database, various information about each operator’s subscriber is stored, such as the permanent key Ki, the IMSI, the last serving MSC/SGSN, the phone number, etc. According to [8], an HLR/AuC can serve up to five million subscribers, and at least one transaction with HLR/AuC is required for every single phone call or data session. Due to its critical functionality, the HLR/AuC is considered to be a single point of failure for every mobile operator [8]. This means that if an HLR/AuC is out of service, then none of its subscribers can be served for calls or data services. Thus, it is motivating for a malicious user to create overload conditions at the level of HLR/AuC that lead to service unavailability.

This paper presents experimental results and observations that can be exploited to perform a novel distributed denial of service (DoS) attack in 3G networks that targets the availability of the HLR/AuC. More specifically, first we present an experiment in which we identified and proved some zero-day vulnerabilities of the 3G network that can be exploited by malicious actors to mount various attacks. For the purpose of our experiment, we have used off-the-shelf infrastructure and software, without any specialized modification. Based on the observations of the experiment, we reveal an Advanced Persistent Threat (APT) [4] in 3G networks that aims to flood an HLR/AuC of a mobile operator. In particular, in this attack, a group of adversaries first collect IMSIs that belong to the same HLR/AuC, using an IMSI catcher [18]. Next, residing in roaming networks, they perform successive registrations using the collected IMSIs that trigger the execution of ADRs to the specific HLR/AuC. For each ADR concerning a different IMSI (i.e., user), the HLR/AuC is forced to generate a batch of L AVs, and send them to the requesting MSC/SGSNs. The continuous execution of ADRs, in a very short-time period incurs the depletion of the computational resources of the HLR/AuC, eventually, leading to system saturation. Moreover, we prove that this attack can be performed in a trivial manner using commodity hardware and software, which is widely and affordably available. The multiple alarming findings of this article should raise awareness of the security risks that threaten the normal operation of the mobile networks and in general critical infrastructures [6].

The rest of the paper is organized as follows. Section 2 presents an overview of the 3G network, and section 3 includes the related work. Section 4 elaborates on the carried out experiments; while section 5 analyses quantitatively and qualitatively the discovered APT in 3G networks, as well as proposes mitigation measures. Finally, section 6 concludes the article.

2 Background

2.1 3G network

2.1.1 Identification

Each 3G subscriber is assigned a unique identity, called IMSI that identifies it, globally. An IMSI is usually presented as a 15-digit number, where the first 3 digits are the mobile country code (MCC), followed by the mobile operator code (MNC). The length of MNC is either 2 digits (European standard) or 3 digits (North American standard). The remaining digits are the mobile subscription identification number (MSIN) within the home network's customer base. However, in the majority of cases, the identification of a user on the radio access link is achieved by means of a temporary mobile subscriber identity (TMSI). A TMSI, which is valid only within the location area, or a packet TMSI (PTMSI), which is valid in the routing area, when available, is employed in paging requests, location update requests, attach requests, service requests, connection re-establishment requests, and detach requests. The association between the permanent and temporary user identities is kept by the MSC/SGSN, in which the user is registered.

2.1.2 Architecture

An outline of the 3G architecture is depicted in Fig. 1, focusing only on the interconnection between the home and roaming network. It consists of the following three components:

a. The MS that comprises a mobile phone or tablet used for radio communication, as well as the UMTS subscriber identity module ((U)SIM) containing the subscriber’s information.
b. The roaming/serving network which includes: a) the Node B that performs the air interface processing; b) the radio network controller (RNC) that is responsible for the radio resource management, handoff decisions, congestion

control and power control; c) the MSC/SGSN that provides voice and data management services; and d) the visiting location register (VLR) that is a database containing temporary/service data for MSs (e.g., TMSI, etc.).

c. The home network that consists of: a) the gateway GPRS support node (GGSN), which is the interface between the 3G network and external packet data networks; and b) the HLR/AuC, which is the central database containing subscribers' information (e.g., IMSI, permanent key Ki, etc.) and generating AVs.

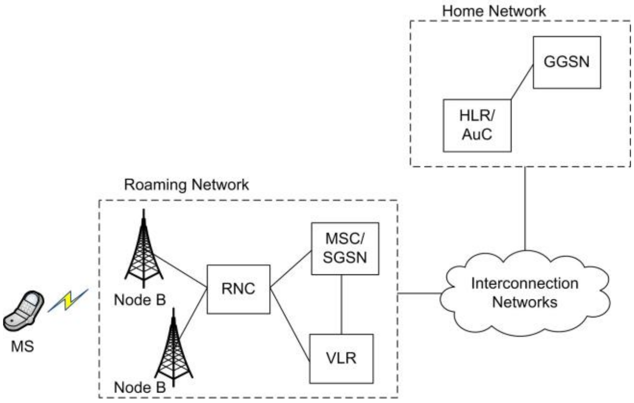


Figure 1: UMTS roaming network architecture

2.1.3 Registration procedure

Before a roaming user initiates a phone call or data session, it should, first, register to the roaming network, which it will serve it. To achieve this, the user's MS establishes a radio resource control (RRC) connection between itself and the RNC [5]. During this, the MS sends an *RRC connection request* message including its TMSI to the RNC, and the latter responds with an *RRC connection setup*. The MS acknowledges this message by forwarding an *RRC connection setup complete* message back to the RNC (see Fig. 2). After establishing the RRC connection, the MS sends a *service request* to the roaming MSC/SGSN including its TMSI. If the MSC/SGSN cannot recognize the received TMSI, it should initiate the registration procedure by sending an *identity request* message back to the MS. The latter responds with an *identity response* message, containing its permanent identity (IMSI). Note that the IMSI is conveyed in plaintext, and, therefore, it can be easily compromised. Next, the roaming MSC/SGSN forwards the received IMSI to the home HLR/AuC, using an *ADR* message. Upon receiving the IMSI, the home HLR/AuC generates a batch of L different AVs. The procedure of generating AVs is specified in [2], and we do not analyze it further. Next, the home HLR/AuC sends an *authentication data response* message back to the roaming MSC/SGSN, containing the array of the generated AVs. The roaming MSC/SGSN selects the first AV and sends it to MS in an *authentication request* message, while it stores the remaining (L-1) AVs for future use. In other words, the roaming MSC/SGSN caches AVs to avoid communicating with the home HLR/AuC for each registration of the user. Upon receiving the *authentication request* message, the MS verifies it authenticating the mobile network and sends back an *authentication response* message to be authenticated by the network. Finally, the roaming MSC/SGSN sends a *cipher mode command* message to MS indicating its successful registration to the network.

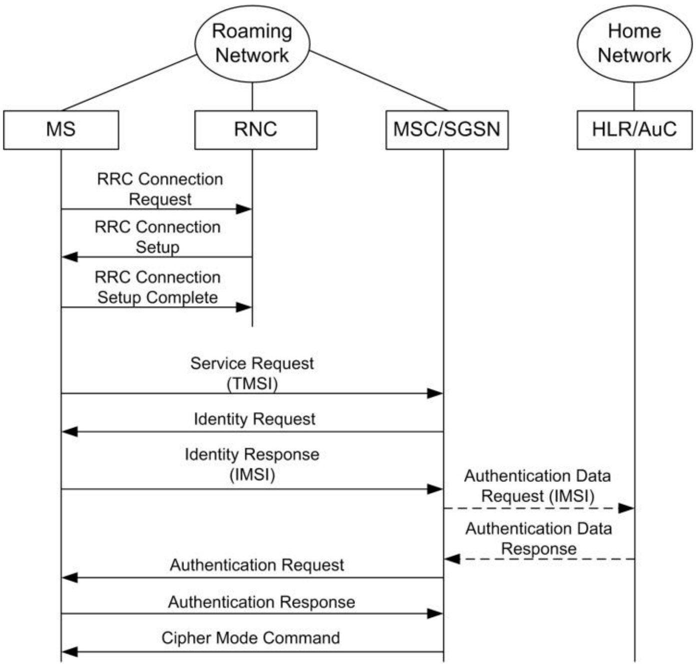


Figure 2: Phone call setup with registration of MS in the roaming network

3 Related Work

The literature includes some previous works, which present discovered vulnerabilities in 3G networks that can be exploited to mount DoS attacks to various segments of 3G networks. In [14], the authors have, collectively, reviewed four different DoS attacks that target 3G networks. The first one is the SMS (short message service) DoS attack [11], in which a high number of SMS are dispatched toward a large number of mobile users, virtually, to all active MS. The procedure of transmitting an incoming SMS through the GSM network is, relatively, complex and consumes resources, such as bandwidth, processing power, memory state, at several network elements and on the radio interface. The second attack is related with the paging procedure in 3G networks [13]. In particular, this attack aims at overloading the paging channel, by causing an exceptionally high rate of paging requests. This is achieved by sending IP packets to a large number of MS, in a short time interval. The goal of the third attack [9] is to maximize the frequency of channel transitions, so as to induce a higher signalling load and, eventually, congest the relevant control channels at the radio interface. Finally, the objective of the fourth attack [10] is to prevent control channels from being released, even in the absence of legitimate traffic, to render them unavailable for other legitimate MSs. In other words, the attacker aims at starving the available control channels, by forcing them to remain assigned to (possibly inactive) terminals.

In [24], the authors have investigated the impact of a parameter named reservation timeout (RT) on the authentication performance in UMTS networks. RT defines the time period that the serving SGSN stores the AVs of a visiting user, before deleting them. A high RT value results in fewer ADR to the home HLR/AuC, but, at the same time, increases the storage overhead in the serving SGSN. Driven by this observation, the authors have developed an analytic model that provides guidelines to mobile operators in order to select an appropriate RT value for each MS, based on its mobility and the occurred authentication rate.

In [27], the authors have discovered and analyzed a low-volume and low-rate DoS attack on UMTS networks. In this attack, an adversary forces, repeatedly, the establishment and release of radio access bearers between MSs and a target RNC. This causes an excessive amount of signalling messages that overload both the RNC and base stations. As a result, RNC is forced to drop service requests from legitimate MSs. The authors have also argued that the same attack can be performed in Wimax networks. To address such an attack, the authors have proposed an early detection mechanism, based on the cumulative sum (CUSUM) test that can, accurately, identify the network source of the described attack, while at the same time it raises very few false positives. A similar study is carried out in [26], where the authors have quantitatively analyzed, from a theoretically point of view, a simple DoS attack in GSM networks. In this attack, adversaries using a number of specialized SIM-less devices, are capable of performing continuous attach requests in order to overload the serving HLR/AuC. However, the paper does not provide any implementation details for this specialized SIM-less device.

In [25], the authors have discovered and analyzed a set of architectural and protocol vulnerabilities in UMTS that may lead to various signalling-oriented DoS attacks. The impact of these vulnerabilities varies from light attacks that target the quality of service of a specific user, up to massive DoS attacks to the provided services. The authors have also identified vulnerabilities of the EAP-AKA (i.e., extension authentication protocol - authentication key agreement) protocol, employed in integrated UMTS/wireless local area networks, which may lead to DoS attacks. Moreover, [28] analyzes the security issues and vulnerabilities of the 3GPP long term evolution (LTE) networks, as well as evaluates some state-of-the-art solutions that deal with them. The authors, concluding, pinpoint the open issues of this area and provide some future research directions.

In [8], the authors, quantitatively, characterize a distributed DoS attack to an HLR/AuC, coordinated by a botnet of infected mobile devices. This work provides numerical estimations for various parameters to successfully perform the attack, such as the required number of infected mobile phones, the rate of flooding messages, the service requests and network operations that incur the greatest burden to the HLR/AuC, etc. It identifies that the *insert/delete call forwarding requests*, which allow a user to redirect incoming phone calls to other devices, are the most suitable, from an attacker perspective, to flood the HLR/AuC. It reveals that the registration procedure is not so effective to flood the HLR/AuC, due to the caching mechanism of AVs in the serving MSC/SGSN. That is, during an MS registration, the serving MSC/SGSN may provide to the MS an AV, already, stored from a previous ADR, meaning that the MSC/SGSN does not have to perform an ADR to the home HLR/AuC. Finally, the authors have estimated the throughput reduction of an HLR/AuC under DoS attack, using insert call forwarding requests. Due to the fact that the insert call forwarding request have, approximately, the same throughput with the registration procedure (excluding the cashing mechanism of AVs) in HLR/AuC [8], we use the throughput estimations of [8] in the performance analysis carried out in this paper.

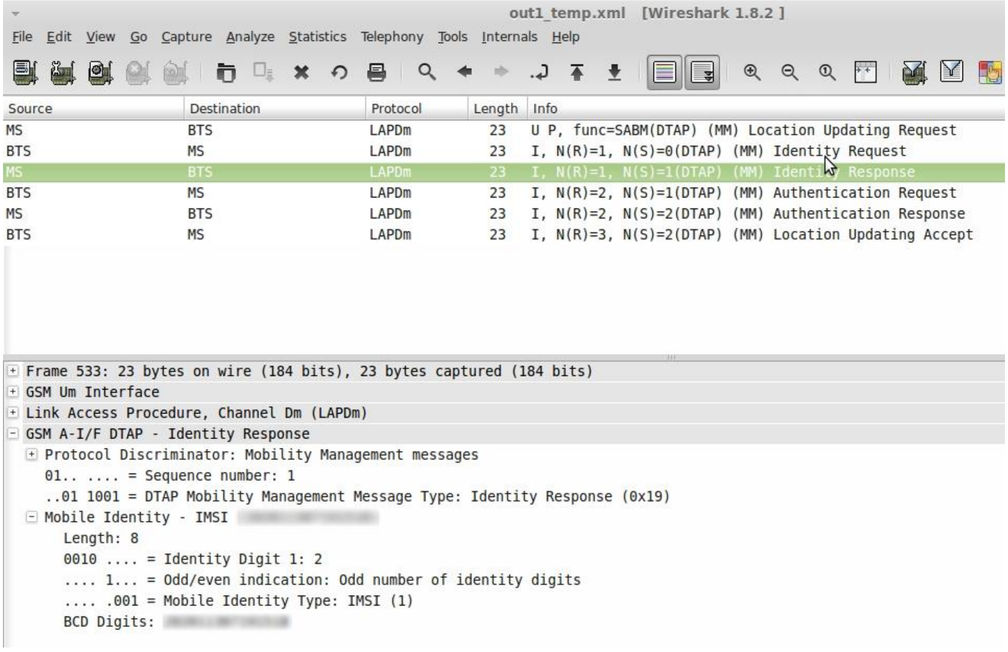
Based on the above analysis, the key contributions of this paper that differentiate it from the related works are summarized as follows:

- The related works do not provide any experimental demonstration that proves the existence of the discovered vulnerabilities, which means that they have been studied only at a theoretical level, and, hence, their feasibility may be questionable. On the contrary, in this paper, except for the theoretical analysis, we also prove the existence of the discovered vulnerability by carrying out an actual experiment on a mobile operator.
- A common limitation of the related works has to do with the fact that they do not provide technical details on how to practically exploit the discovered vulnerabilities. In this work, on the other hand, we have implemented the equipment that can be used by an adversary to launch the presented attack. In this way, we prove that the attack can be performed in a trivial manner using commodity hardware and freely available software.
- Ref. [8] does not discover a security vulnerability that can be exploited to perform a DoS attack to mobile networks. The main contribution of this work is the quantitative characterization of a DoS that can be performed by a botnet of infected mobile devices. In contrast, our paper, first, presents a security vulnerability of mobile networks, and, then, it elaborates on the techniques with which an adversary can exploit this vulnerability to mount a DoS attack.
- Contrary to [8], which argues that the registration procedure is not suitable to flood an HLR/AuC, because of the cashing mechanism; in this paper, we perform a DoS attack to an HLR/AuC using the registration procedure and bypassing the cashing of AVs. To achieve this, the following occur: a) an adversary connects to a roaming MSC/SGSN (i.e., different from the home MSC/SGSN of the captured IMSIs) and uses each captured IMSI only once; (b) multiple adversaries that reside at different roaming networks may, simultaneously, attack the target HLR/AuC, using the same set of captured IMSIs (i.e., for each captured IMSI, one ADR is executed to the target/home HLR/AuC from each visited/roaming MSC/SGSN, where an adversary is connected); and (c) as the attack is launched from roaming networks, it is evident that the roaming MSC/SGSNs do not have cached AVs for the captured IMSIs, except for the rare case where the owner of a

captured IMSI has, recently, visited a roaming network that participates in the attack. The existence of roaming agreements between the home and the roaming networks is a prerequisite.

4 Proven vulnerabilities in 3G networks

In this section, we present an experiment in which we identified and proved some zero-day vulnerabilities of the 3G network that can be exploited by malicious actors to mount various attacks. For the purpose of our experiment, we have used off-the-shelf infrastructure and software, without any specialized modification. First, we cloned a SIM card of a Greek mobile operator, using specialized software named SIM scanner [13]. The home network of this card is located in Athens (Greece). Thus, we had under our possession two different SIM cards: the original card and the cloned one. The two cards share the same permanent key Ki, the same calling number, and the same IMSI identity. In total, we performed two sets of experiments. In the first set, the aim was to verify the fact that the home HLR/AuC always accepts and proceeds an ADR from an MSC/SGSN of a roaming network. In the second set of experiments, the goal was to study the behaviour of the home network to various management procedures that refer to already registered mobile subscribers to the network, which are originated from other serving networks.


out1_temp.xml [Wireshark 1.8.2]
File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help

```
Source          Destination     Protocol    Length  Info
MS              BTS             LAPDm       23      U P, func=SABM(DTAP) (MM) Location Updating Request
BTS             MS              LAPDm       23      I, N(R)=1, N(S)=0(DTAP) (MM) Identity Request
MS              BTS             LAPDm       23      I, N(R)=1, N(S)=1(DTAP) (MM) Identity Response
BTS             MS              LAPDm       23      I, N(R)=2, N(S)=1(DTAP) (MM) Authentication Request
MS              BTS             LAPDm       23      I, N(R)=2, N(S)=2(DTAP) (MM) Authentication Response
BTS             MS              LAPDm       23      I, N(R)=3, N(S)=2(DTAP) (MM) Location Updating Accept
```

```
⊞ Frame 533: 23 bytes on wire (184 bits), 23 bytes captured (184 bits)
⊞ GSM Um Interface
⊞ Link Access Procedure, Channel Dm (LAPDm)
⊟ GSM A-I/F DTAP - Identity Response
  ⊞ Protocol Discriminator: Mobility Management messages
    01.. .... = Sequence number: 1
    ..01 1001 = DTAP Mobility Management Message Type: Identity Response (0x19)
  ⊟ Mobile Identity - IMSI [redacted]
      Length: 8
      0010 .... = Identity Digit 1: 2
      .... 1... = Odd/even indication: Odd number of identity digits
      .... .001 = Mobile Identity Type: IMSI (1)
      BCD Digits: [redacted]
```

Figure 3: Successful registration of the cloned SIM card as shown in Wireshark (the IMSI number has been blurred on purpose)

The initial experiment was carried out as follows. First, we powered on a mobile device using the original SIM in Athens/Greece, and we initiated some phone calls to register the IMSI of the SIM card in the HLR/AuC of its home network. Next, we powered on a mobile device with the cloned SIM in Lisbon/Portugal (the choice of this location was arbitrary). Using a tool named Nokia Net monitor [22], we captured the network traffic that this mobile device had generated. In the sequel, we analysed the captured packets using the protocol analyser Wireshark, which is capable of decoding GSM protocol messages. The analysis revealed that the cloned SIM performed a successful registration procedure in the home network, through the roaming network (see Fig. 3), despite the fact that the original SIM had performed a registration in Greece/Athens few moments ago. We repeated the experiment five times at different hours to ensure our findings. From this experiment, we can deduce the following observation:

Observation 1*:* *a roaming/serving MSC/SGSN accepts all IMSIs and initiates an ADR with their home networks (i.e., HLR/AuC) in order to service the corresponding subscribers.*

Observation 2: *the home HLR/AuC accepts an IMSI received from a roaming network, without taking into account whether this IMSI is active and registered in its home network.*

Observation 3: *upon a request from a roaming network, the home HLR/AuC generates fresh AVs for each of its subscribers, regardless if the latter are already registered in the home or another serving network.*

Observation 4: *the home HLR/AuC does not validate the elapsed time between two successive ADRs for the same IMSI, which are originated from different MSC/SGSNs.*

In the second set of experiments, first, we simultaneously made outgoing calls using the two SIM cards. After performing outgoing calls, we made an incoming call to the phone number of the SIM card (as mentioned previously, the phone number is the same for both the original and the cloned SIM card). We noticed that the mobile device that rang (i.e., received the call) was the one that had used to make the last outgoing call. We repeated this experiment ten times in a row and in all cases we observed the same result. *That is, the mobile device that rang was the one that had made the last outgoing call.* This can be justified as follows. As mentioned previously, the home HLR/AuC stores for each user the last MSC/SGSN that served it. As a result, when we performed an outgoing call using the original SIM, the home HLR/AuC stored the home MSC/SGSN that served the call, which is located in Athens. On the other hand, when we performed an outgoing call using the cloned SIM, the home HLR/AuC stored the roaming MSC/SGSN that operates in Lisbon. Thus, when an incoming call to the number of the SIM card occurred; the home HLR/AuC directed the call to the MSC/SGSN (i.e., home or roaming) that had served the SIM last, deducing to the following observation.

Observation 5: *the HLR/AuC does not validate the elapsed time between two successive phone calls of the same SIM, which are originated from different MSC/SGSNs.*

Based on the above, we can conclude to a more generic observation that characterizes the behaviour and operation of 3G networks.

Observation 6*: the home HLR/AuC trusts and processes all signalling and management requests, updates, etc., concerning its subscribers, which are received from potential roaming MSC/SGSNs that maintains with them roaming agreements.*

Our experiments and observations revealed and proved that 3G networks present some zero-day vulnerabilities, which may lead to a series of possible attacks. An evident attack is related to the fact that a malicious user, who owns a cloned SIM card, may direct incoming calls to the cloned card, causing in this way the legitimate user to lose incoming calls (i.e., service unavailability). Moreover, the malicious user can overcharge the legitimate owner of the SIM card, simply, by making phone calls using the cloned SIM card. Finally, two malicious users may share the original and the cloned SIM card and perform calls from two different countries at the same time. Next, they can deny paying the phone calls originated from one of the two countries, based on the fact that a SIM card cannot be in two countries at the same time (i.e., repudiation).

5 Advanced Persistent Threat in 3G networks

5.1 General

In this section, we present and analyze an APT in 3G networks that aims to flood an HLR/AuC of a mobile operator, resulting in system saturation. As the HLR/AuC is queried in the delivery of all phone calls and text messages, acts as the authentication server for the network, records data for the purposes of billing, and generally assists in a wide range of management jobs, such an attack would, potentially, devastate nearly all services in the network of the mobile operator. In essence, the attack is a distributed DoS attack, orchestrated by a group of adversaries that perform continuous registrations from roaming networks, using a set of collected IMSIs. The attack is considered to be an APT in 3G networks, since: (i) it exploits a series of vulnerabilities of 3G networks; (ii) the adversary can easily evade detection, and (iii) once launched it cannot be effectively confronted without seriously affecting the normal network operation (see section 5.5). More specifically, the discovered APT has the following characteristics:

- It exploits the following identified vulnerability. That is, an HLR/AuC accepts from the home MSC/SGSN an ADR and generates AVs for the related IMSI, and after that, it may accept from a roaming MSC/SGSN another ADR for the same IMSI and, again, generate a batch of AVs, regardless if the IMSI is, already, registered in the home network.
- The considered attack exploits another well-known vulnerability of mobile networks to capture and collect the IMSI identities, which are required to perform the DoS attack. That is, in a registration procedure the permanent

identity (IMSI) of a USIM/SIM is conveyed in plaintext and an adversary can easily read it (see section 2.1.3).

- The adversaries cannot be detected, since they reside in roaming networks far away from the country of the home network. In addition, they can change their location area and mount the attack from different Node Bs, fact that hinders their physical tracking.
- To partially defend from this attack, the mobile operator may, temporally, block the roaming functionality to specific networks, where the attack is originated. However, this would have a great negative impact on the reliability of the mobile operator, since no phone calls or data services will be allowed for the users of the home network that roam to these networks. A more robust protection against the attack is the mobile operator to find and blacklist all the captured IMSIs or block their roaming. As a result, the HLR/AuC would not produce AVs for the blacklisted IMSIs at all, or only for their presence in roaming networks, respectively. However, this means that the legitimate owners of the blacklisted IMSIs have to obtain new USIM cards. Again, this solution would reduce the reliability of the operator.

5.2 Attack analysis

In this section, we analyze the exploitation of the identified vulnerabilities that lead to the DoS attack in 3G networks. A requirement to perform this attack is that the adversaries capture IMSIs identities, in order to execute ADRs. The number of the captured IMSIs is a key parameter for the duration of the DoS attack. In particular, the higher the number of captured IMSIs, the longer the attack duration is. For example, if the adversaries capture 20.000 IMSIs, they can perform the attack for a duration of 10.000 seconds (see section 5.4). IMSIs can be captured using is a special equipment, named IMSI catcher, which masquerades as a base station and logs the IMSI identities of all the MSs in the area, as they attempt to connect to it. Note that in order to flood the targeted HLR/AuC, the adversaries should utilize IMSIs that belong to the targeted HLR/AuC. Despite the fact that the adversaries do not know to which HLR/AuC an IMSI is subscribed, they can identify if an IMSI belongs to the mobile operator of the targeted HLR/AuC, based on the MCC/MNC codes of the IMSI. In this way, the adversaries can utilize IMSIs that belong to the mobile operator of the targeted HLR/AuC, increasing the probability a utilized IMSI reach the targeted HRL/AuC. Apart from the IMSI catcher, the adversaries own a special device that we name it *mal-MS,* which is capable of, consequently, executing a registration procedure, using a different IMSI for each registration attempt.

The studied DoS attack is carried out in a geographically distributed manner. More specifically, cooperative adversaries that reside in different countries, or in the same country but in different location areas (i.e., areas served from different MSC/SGSN), initiate at the same time registration procedures. This guarantees that each adversary uses a different MSC/SGSN to flood the targeted HLR/AuC. This is crucial to perform the attack, because if the adversaries tried to flood the targeted HLR/AuC, only, from one MSC/SGSN, the latter would become a bottleneck, and the malicious registration messages would never reach the targeted HLR/AuC. Another advantage of using multiple MSC/SGSNs to perform the attack is that the same IMSI can be used multiple times to perform registrations and flood the targeted HLR/AuC. Recall from section 2.1.3, that each IMSI can be used, only, once to reach the targeted HLR/AuC, due to the caching mechanism of the AVs in MSC/SGSN. By utilizing multiple MSC/SGSNs, we avoid this limitation, because each IMSI can be used more than once (specifically, equal to the number of participating MSC/SGSN) to flood the HLR/AuC, since they will be originated from different MSC/SGSN. Finally, it is important to notice that to successfully perform the DoS attack, the consecutive registration messages should not strain the radio network elements, including Node B and RNCs. For this reason, each adversary can establish parallel RRC connections with different Node Bs and RNCs that belong to different mobile operators. In this way, the malicious registration messages traverse through multiple radio paths to reach and flood the targeted HLR/AuC.

The considered attack is performed by each participating mal-MS, which executes the attacking protocol, analyzed below (see also Fig. 2). First, the mal-MS establishes an RRC connection with Node B, using a non-valid TMSI (see section 2.1.3). Next, the mal-MS initiates a phone call by sending a *service request* message to MSC/SGSN, using the same TMSI. Since the TMSI is not valid, the MSC/SGSN does not recognize it and requests from the mal-MS a valid IMSI, using an *identity request* message (see Fig. 2). After that, the mal-MS chooses and sends to MSC/SGSN a captured IMSI in an *identity response* message. The MSC/SGSN does not have any stored AVs for the specific IMSI, since the mal-MS is located in a roaming network. This means that the MSC/SGSN, which serves the mal-MS, has to contact and obtain subscriber's information from the targeted HLR/AuC, which is located in the home network of the IMSI (see Fig. 2). Thus, the MSC/SGSN sends to the targeted HLR/AuC an ADR message, including the IMSI, to generate fresh AVs. The targeted HLR/AuC is forced to generate a batch of L AVs for the specific IMSI and send them to the MSC/SGSN in an *authentication data response* message.

This procedure is carried out, repeatedly, from each mal-MS, initiating a phone call to perform a registration in a very short time. In this way, a great amount of ADR messages is directed from the roaming MSC/SGSN to the targeted HLR/AuC, which is successively forced to generate AVs for each received IMSI, depleting its computational resources. Eventually, the targeted HLR/AuC reaches a saturation point and cannot serve new requests (legitimate or malicious).

It is important to mention that the underlying attack, in any case, will not trigger resynchronization [30]. A resynchronization occurs, when a USIM rejects an AV received from the serving MSC/SGSN, during the authentication procedure, as outdated. As referred in the 3GPP 33.102 [2] specifications, USIM verifies that the AV is not too “old,” using two sequence numbers, one maintained by itself and the other stored within the received AV. However, during our attack, the resynchronization procedure cannot be triggered for two fundamental reasons: (i) the participating USIMs (i.e., included in mal-MSs) do not proceed to the verification of the received AVs freshness; and (ii) the attack originates, only, from roaming networks (i.e., MSC/SGSNs) that do not have any “old” (i.e., stored) AV for the captured IMSIs, except for the rare case where the owner of a captured IMSI has, recently, visited a roaming network that participates in the attack. In this case, the only result is that the ADR for this IMSI is not executed, since the particular roaming MSC/SGSN already possesses, at least one, AV for it (i.e., IMSI).

5.3 Implementation of the equipment

To further prove the feasibility of the presented attack, we have implemented the functionality of mal-MS, which is the essential equipment to perform the attack. It is alarming that we were able to implement the functionality of the mal-MS using commodity hardware, which was easily and affordably available. In our implementation, the mal-MS is composed of a smartphone with a rooted Android OS (v2.3.6) [21], a SIM card and SIMtrace [15]. The latter is a software and hardware system for passively sniffing the communication between the SIM card and the modem of a mobile phone. For our purposes, we have developed an enhanced version of the SIMtrace firmware that allows SIMtrace to act as a man-in-the-middle, and, actively, modify the data exchanged between the SIM card and the phone’s modem. We have also developed an application, implemented with the C programming language for the Android OS, which instructs the phone’s modem to initiate and stop phone calls, using AT commands [20]. To achieve this functionality, first we had to stop the radio interface layer (RIL) daemon. RIL is a layer in the Android OS that provides an interface to the phone’s radio and modem. We observed that when the RIL daemon was active, the phone’s modem was delaying to answer to our own AT commands. For this reason, we rooted the Android OS and stopped the RIL daemon. Finally, it is worth mentioning that we could have also built the functionality of the mal-MS using Arduino [16], which is an open-source electronics prototyping platform, based on a programmable microcontroller. The functionality of an Arduino board can be easily extended using interchangeable add-on modules, known as shields. One such shield is the GSM shield [17], which allows an Arduino board to connect to the internet, make/receive voice calls and send/receive SMS messages, using the GSM/GPRS modem.

The functionality of the developed mal-MS is rather simple. First, we activate the Android application that enables the initiation of phone calls via AT commands. This application executes the *dial command* (see sect. 6 of [20]) to trigger the phone’s modem to start a phone call. In order to perform the call, the phone’s modem communicates with the SIM to obtain the TMSI. As mentioned previously, the SIMtrace intercepts the communication channel between the SIM and the modem. Thus, it obtains the TMSI, when it transmitted from the SIM to the phone modem, and modifies it, making it invalid. Next, the SIMtrace conveys the modified TMSI to the modem. Afterwards, the phone's modem initiates the phone call, by sending a *service request* that includes the modified TMSI to the 3G network. Since the TMSI is invalid, the 3G network sends an *identity request* message to obtain the IMSI. The phone's modem now requests from the SIM the IMSI. Again, the SIMtrace intercepts the communication between the modem and the SIM card, and replaces the IMSI with one of the collected IMSIs. The phone's modem now sends an *identity response* message to the 3G network that includes the captured IMSI. Immediately after this message, the Android application sends a *stop command* to the modem to stop the current procedure, and initiate a new phone call, using another collected IMSI.

5.4 Attack characterization

In this section, we evaluate the impact of the attack under various network implementations, network traffic conditions and attack parameters. First, we estimate the attack rate that the mal-MS can perform successive registration attempts with the collected IMSIs. Recall that the mal-MS executes a phone call setup with registration until it sends an *identity response* message (see section 2.1.3). This means that in order to estimate the attack rate of mal-MS, it is sufficient to estimate the elapsed time between an *RRC connection request* and an *identity response* message (see Fig. 2). Using the developed mal-MS, we performed *service requests* a total of 1,300 times during low traffic hours, in order to avoid overwhelming the HLR/AuC. As shown in Fig. 4, we measured that the elapsed time between an *RRC connection request* and an *identity response* message is around 450 msec (~ 0.5 second). This means that the attack rate that our implementation of the mal-MS supports is, approximately, two registration attempts per second. Note that we could further increase the attack rate, by optimizing the software and/or hardware of the mal-MS. However, this was omitted, since it is not the goal of our experiments.

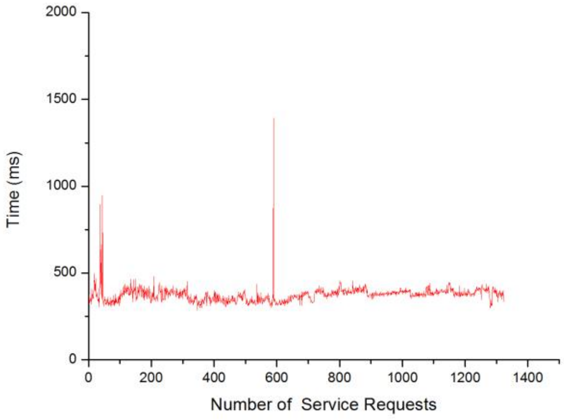


Figure 4: Number of service requests vs elapsed time (ms) using mal-MS

Now, we elaborate on the throughput reduction of the HLR/AuC under the considered DoS attack. To estimate such a reduction, caused by a directly implemented attack, the ideal approach would have been to perform the attack against an operated 3G mobile network and observe the performance degradation or even the network failure. However, such experiments on a real 3G network are not feasible, due to both economic and legal reasons. Another approach could have been to collect data traces from a deployed 3G networks, and then on these data emulate the attack occurrence using hypothetical attack traffic. However, such operational data from a live 3G network are not available, because of the users' privacy and the competition among operators. Based on these facts, we used the methodology described below. First, we determined the attack parameters and performance metrics that influence the severity and the impact of the considered attack (see Table 1). These include: (i) the HLR/AuC throughput, (ii) the HLR/AuC database implementation, (iii) the intensity of legitimate traffic, (iv) the number of captured IMSIs, (v) the number of participating roaming networks, (vi) the number of mal-MS devices in each roaming network, and (vii) the registration attempts per second of each mal-MS.

Table 1: Attack parameters

Attack parameters
HLR/AuC throughput
HLR/AuC database implementation
Intensity of legitimate traffic
of captured IMSIs
of participating roaming networks
of mal-MS in each roaming network
Registration attempts per second of mal-MS

The HLR/AuC throughput is the mean number of database transactions per second (tps) in HLR/AuC. The latter can be deployed using either MySQL or SolidDB [8]. MySQL can use the InnoDB storage engine, which maintains a buffer pool for caching data and indexes in main memory. On the other hand, SolidDB stores the entire database in memory, offering significantly higher performance, compared to MySQL. The intensity of the legitimate traffic can be categorized as high or low, and its exact value depends on the HLR/AuC implementation. Based on the experiments of [8], for MySQL implementation, the high intensity traffic value is, approximately, 4000 tps and the low 2500 tps. On the other hand, for SolidDB implementation, these values are 5500 tps and 3000 tps, respectively. The captured IMSIs are obtained by the adversaries using an IMSI catcher. The higher the number of captured IMSIs, the longer the attack duration is (see section 5.4). The attack is performed throughout a number of participating roaming networks. The more involved roaming networks, the higher the severity of the performed DoS attack and the less number of captured IMSIs is required, because each IMSI can be used multiple times (once in each network). In each roaming network, a number of mal-MS may reside (each one uses a different subset of the captured IMSIs of the home/target network), capable of performing continuous registration attempts. The more participating mal-MSs and the higher the number of registration attempts per second that each mal-MS supports, the faster the HLR/AuC is saturated. Our mal-MS implementation supports two registrations attempts per second.

After determining the attack parameters, we acquired data from the related work [8], and we projected them to the number of the registration attempts that our implementation of mal-MS supports (i.e., two registration attempts per second). In this way, we obtained numerical results that allow the quantitative characterization of the considered attack. In particular, we estimated the HLR/AuC throughput under attack in regard to the number of mal-MSs for MySQL and SolidDB implementation and for high and low intensity traffic. As shown in Fig. 5, using MySQL implementation under normal conditions (i.e., no attack occurrence), in a low intensity traffic stream, the HLR/AuC achieves a throughput of 2,427 tps; while for the same implementation in a high intensity traffic stream, the throughput reaches up to 4,132 default tps. In this implementation, in a low intensity traffic scenario, the participation of 500 mal-MSs (i.e., each one performs two registration attempts per second) reduces the HLR/AuC throughput to ~900 tps, which means that the throughput of the legitimate traffic is reduced by approximately 63%. Similarly, 1,250 mal-MSs shrink the HLR/AuC throughput to 146 tps (93% decrease of the throughput). In a high intensity traffic scenario, the participation of 1,250 mal-MSs reduces the throughput of the legitimate traffic by approximately 66% (from 4,132 to ~1,400 tps). In order to reach a reduction level of 93% (from 4,132 to 273 tps), a group of 2,500 mal-MSs is required.

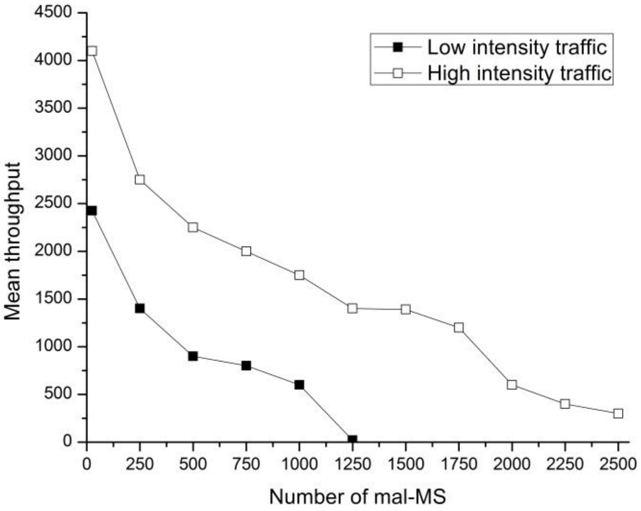


Figure 5: Mean throughput vs number of using mal-MS for MySQL implementation of HLR/AuC

On the other hand, using SolidDB implementation, in a low intensity traffic stream, the HLR/AuC presents a throughput of 3,075 tps, while in a high intensity traffic stream, a throughput of 5,424 tps (see Fig. 6). In this case, the participation of 15,000 mal-MSs results in the reduction of the legitimate traffic throughput by approximately 75% (e.g., in a low intensity scenario from 3075 to 803, while in a high intensity scenario from 5424 to 1340). Given the fact that the attack rate of the mal-MS is two registration attempts per second, the duration of the DoS attack depends, directly, on the number of the collected IMSI. For example, if the adversaries have collected 20,000 IMSIs, then we conclude that they can flood the HLR/AuC for 10,000 seconds (2 hours and 40 minutes), since each IMSI can be used once. However, the collected IMSIs can be separated in groups and distributed among mal-MSs, with the condition of not reusing the same IMSI by two or more mal-MSs that reside at the same roaming network. Such an approach reduces the number of participating roaming networks, required for the attack occurrence; but on the other hand, decreases the time-period that the attack can take places, or increases the required number of collected IMSIs.

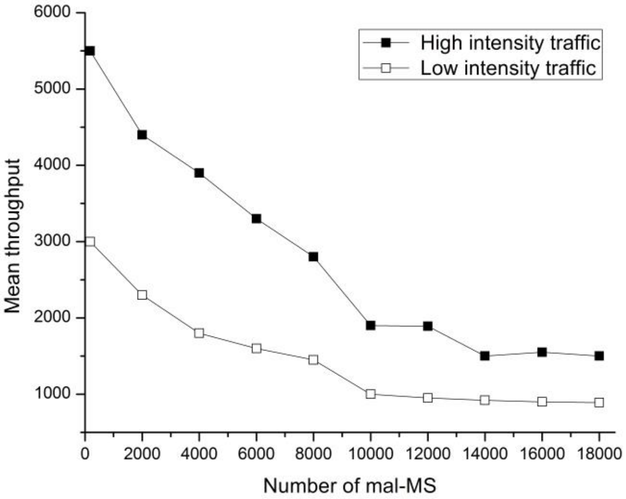


Figure 6: Mean throughput vs number of mal-MS for SolidDB implementation of HLR/AuC

It is important to mention that the discovered attack can be launched also in LTE networks [23]. In fact, the attack is considered to be more harmful in LTE networks, compared to 3G networks. In particular, in LTE networks there is no caching mechanism of AVs, in order to reduce the signaling load between the home and the serving network, as well as the memory requirements of the serving network components for storing the AVs. As a result, for each authentication/registration of an MS, the home network should respond with exactly one AV. Thus, the adversaries can deplete the computational resources of the home network for longer time periods, using the collected IMSIs more than once. Moreover, the deployment of the new evolved Node Bs and femtocells at public local service points, such as schools, universities, theaters, malls, etc., will further increase the intensity of the attack, enabling the activation of more mal-MS in parallel.

5.5 Attack mitigation

To confront the discovered APT, the mobile operators may either employ, at the level of HLR/AuC, early intrusion detection systems (IDS) or load balancing schemes. An IDS monitors the total number of ADR messages, and when this number exceeds a predetermined threshold, it raises an alarm for further operation and maintenance (O&M) actions. The IDS mechanism can be based on the non-parametric CUSUM test [31], which detects the cumulative effect of the deviation from normal protocol and traffic behavior, caused by a DoS attack. The CUSUM test is suitable for such an attack, presented in this paper, since it provides a simple solution with low computational overhead [31]. Thus, the overall performance of the underlying HLR/AuC will not be considerably deteriorated. Moreover, the non-parametric CUSUM test is insensitive to traffic patterns with unknown distribution, making the detection mechanism robust, generally applicable, and its overall deployment easy and straightforward. Another advantage of using the CUSUM test is related to the fact that, given an appropriate threshold value, it detects the attack at the earliest possible. This is required to confront the considered attack, because the sooner a mobile operator reacts to it, the less damage will be caused. Taking into account the above characteristics of CUSUM, it is evident that it provides an effective and efficient solution to detect the considered DoS attack.

As soon as the attack is detected, the home network may either, directly, proceed to some temporary measures to confront the attack, or collaborate with the serving/roaming networks, where the registration requests are initiated, to diminish them. In the first option, the home network may:

a. Reject/block ADRs from specific roaming networks, which are flooding the home network. This will stop flooding the home network, but the legitimate users of the home network that are served by the blocked roaming networks cannot consume fresh AVs. This means that if there is no stored AVs for them, they cannot be authenticated, and, hence, receive 3G services. Otherwise, if there are some stored AVs, as soon as they have been consumed, the roaming networks may either stop providing services to the legitimate home users, or use the already established keys for longer, which degrades the provided security services [2].
b. Blacklist all the captured IMSIs and stop serving roaming registration requests for them. However, this means that the legitimate owners of the blacklisted IMSIs are not allowed to roam, or have to obtain new USIM cards.

In the second option, i.e., collaboration between the home and roaming networks, the former may instruct the latter to:

a. Deactivate the base stations and/or RNCs where the attack is originated; but this cannot be accepted by the roaming networks, since it results in network unavailability at specific parts of their coverage area.
b. Deactivate ADR for the roaming users of the home network. This will stop flooding the home network, but it will also stop serving either the legitimate

home users that are, newly, visiting the roaming networks, or the legitimate home users already served by the roaming networks and have consumed their locally stored AVs. To overcome the latter, the roaming networks may decide to use the already established keys for longer, but this fact results in the degradation of the supported security level.

c. Disable roaming services for the users of the home network. This will stop flooding the home network; but it will also stop providing services to the legitimate home users that visit the roaming networks.

Summarizing the above, we may deduce that after detecting the flooding attack, the allowable O&M action may either result in service unavailability (i.e., blocking of legitimate users), which is a type of DoS, or the degradation of the supported security level, which is a serious threat that may lead to the occurrence of a series of other well known attacks [2].

Apart from, early, detecting and trying to confront the considered DoS attack, another solution that could mitigate its negative effects is the adoption of a load balancing scheme for HLR/AuC. More specifically, the mobile operator of the home network can replicate and distribute the HLR/AuC, in order to improve its reliability and capacity. As a result, the availability of HLR/AuC will be higher, in case the home network is under the DoS attack. On the other hand, replicating an HLR/AuC increases the cost to setup and maintain database copies. Also, the database load will increase, due to the extra overhead of update operations, which are required to keep the contents of the HLR/AuC database consistent across the replicas. In [29], the authors have evaluated the performance of a load balancing scheme for HLR/AuC with respect to the number of replicas.

Both IDS mechanisms and load balancing techniques provide a defense line that can mitigate the effects of the considered DoS attack. However, in order to eliminate the threat, a strong collaboration is required between the home network and roaming networks. More intelligent mechanisms should be developed, capable of swiftly detecting when an ADR message originated from a roaming network is legitimate or not. In general, responding to a large-scale DoS attack in 3G networks remains challenging, and much work remains to be done in the development of defenses.

6 Conclusions

This paper presented experimental results and observations that can be exploited to perform a novel distributed DoS attack in 3G networks that targets the availability of the HLR/AuC. First, we analysed an experiment in which we identified and proved some zero-day vulnerabilities of the 3G network, which can be exploited by malicious actors to mount various attacks. For the purposes of our experiment, we used off-the-shelf infrastructure and software, without any specialized modification. Based on the observations of the experiment, we revealed an APT in 3G networks that aims to flood an HLR/AuC of a mobile operator. In particular, in this attack, a group of adversaries, first, collect IMSIs that belong to the same HLR/AuC. Next, residing in roaming networks, they perform successive registrations using the collected IMSIs that trigger the execution of ADRs to the specific HLR/AuC. For each ADR concerning different IMSI (i.e., user), the HLR/AuC is forced to generate a batch of L AVs and send them to the requesting MSC/SGSNs. The continuous execution of ADRs in a very short time period incurs the depletion of the computational resources of the HLR/AuC, eventually, leading to system saturation. Moreover, we proved that this attack can be performed in a trivial manner using commodity hardware and software, which is widely and affordably available.

References

[1] GSM Association, available at http://www.gsma.com

[2] 3GPP TS 33.102: "3rd Generation Partnership Project (3GPP); Technical Specification Group Services and System Aspects; 3G Security; Security architecture."

[3] 3GPP TS 29002: "3rd Generation Partnership Project (3GPP); Technical Specification Group Core Network; Mobile Application Part (MAP) specification."

[4] N. Virvilis, D. Gritzalis, "The Big Four - What we did wrong in Advanced Persistent Threat detection?", 8th International Conference on Availability, Reliability and Security (ARES-2013), September 2013, Germany.

[5] 3GPP TS 25331: "3rd Generation Partnership Project (3GPP); Technical Specification Group Radio Access Network; Resource Control (RRC) protocol specification."

[6] P. Kotzanikolaou, M. Theoharidou, D. Gritzalis, "Cascading effects of common-cause failures on Critical Infrastructures, 7th IFIP International Conference on Critical Infrastructure Protection (CIP-2013), March 2013, USA.

[7] G. Kambourakis, C. Kolias, S. Gritzalis, J.H. Park, "Signaling-oriented DoS attacks in UMTS Networks", 3rd International Conference on Information Security and Assurance (ISA 2009), Jun 25-27, Seoul, Korea.

[8] P. Traynor, M. Lin, M. Ongtang, V. Rao, T. Jaeger, P. McDaniel, T. La Porta, "On cellular botnets: measuring the impact of malicious devices on a cellular network core", 16th ACM conference on Computer and communications security (CCS 2009), Nov. 09-13, 2009, Chicago, Illinois, USA.

[9] P. Lee, T. Bu, T. Woo, "On the detection of signaling DoS attacks on 3G wireless networks", 26th IEEE International Conference on Computer Communications (INFOCOM 2007), May 6-12, Anchorage, Alaska, USA.

[10] A. Barbuzzi, F. Ricciato, G. Boggia, "Discovering parameter setting in 3G networks via active measurements", IEEE Communications Letters, Vol. 12, No 10, pp. 730-732, Oct. 2008.

[11] W. Enck, P. Traynor, P. McDaniel, T. La Porta, “Exploiting open functionality in SMS-capable cellular networks” 12th ACM conference on Computer and communications security (CCS 2005), Nov 7-11, Alexandria, Virginia, USA.

[12] J. Serror, H. Zang, J.C. Bolot, “Impact of paging channel overloads or attacks on a cellular network”, ACM Workshop on Wireless Security in conjunction with the 12th ACM International Conference on Mobile Computing and Networking (ACM WiSe 2006), Sept. 24-19, Los Angeles, USA.

[13] http://www.nowgsm.com/download/SIM-Scanner.pdf

[14] F. Ricciato, A. Coluccia, A. D'Alconzo, “A review of DoS attack models for 3G cellular networks from a system-design perspective”, Computer Communications, Elsevier Science, Vol. 33, No. 5, pp. 551-558, March 2010,.

[15] http://bb.osmocom.org/trac/wiki/SIMtrace

[16] http://www.arduino.cc/

[17] http://arduino.cc/en/Main/ArduinoGSMShield

[18] F. Joachim, R. Bott, “Method for identifying a mobile phone user or for eavesdropping on outgoing calls”, EPO Patent EP1051053, Jul. 2003.

[19] http://www.qualcomm.com/solutions/testing/diagnostics-software

[20] 3GPP TS 27.007: “3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; AT command set for User Equipment”

[21] Samsung Galaxy mini 2 6500 full phone specifications: http://www.gsmarena.com/samsung_galaxy_mini_2_s6500-3883.php

[22] http://en.wikipedia.org/wiki/Nokia_network_monitor

[23] 3GPP TS 33.401: “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3GPP System Architecture Evolution (SAE); Security architecture.

[24] Lin-Yi Wu, Yi-Bing Lin, “Authentication Vector Management for UMTS”, IEEE Transactions on Wireless Communications, Vol. 6, No. 11, 2007.

[25] G. Kambourakis, C. Kolias, S. Gritzalis, J. H. Park,“DoS Attacks Exploiting Signaling in UMTS and IMS”, Computer Communications, Vol. 34, No. 3, pp. 226-235, 2011, Elsevier.

[26] N. Gobbo, A. Merlo, M. Migliardi “A Denial of Service Attack to GSM Networks via Attach Procedure”, 8th International Conference on Availability, Reliability and Security (AReS 2013), University of Regensburg, Germany, Sept. 2013.

[27] Patrick P. C. Lee, Tian Bu, Thomas Y. C. Woo, “On the Detection of Signaling DoS Attacks on 3G/WiMax Wireless Networks”, Elsevier Science, Computer Networks Volume 53 Issue 15, October 2009.

[28] Cao, J., Ma, M., Li, H., Zhang, Y., “A Survey on Security Aspects for LTE and LTE-A Networks”, IEEE Communications Surveys and Tutorials, vol.PP, no.99, pp.1-20, 2013.

[29] Guan-Chi Chen and Suh-Yin Lee, “Evaluation of Distributed and Replicated HLR for Location Management in PCS Network”, Journal of Information Science and Engineering, Vol.19 No.1, pp.85-101, Jan. 2003.

[30] C. Ntantogian, C. Xenakis, I. Stavrakakis, “Reducing False Synchronizations in 3G-WLAN Interworking Networks”, IEEE Transactions on Wireless Communications, Vol. 10, No. 11, pp: 3765–3773, Nov. 2011.

[31] H. Wang, D. Zhang, K.G. Shin, “Change-point monitoring for detection of DoS attacks”, IEEE Transactions on Dependable and Secure Computing, Vol. 1 No. 4, 2004.