Security attacks against the availability of LTE mobility networks: Overview and research directions

Conference Paper · January 2013

CITATIONS
66

READS
1,841

1 author:

Roger Piqueras Jover
Independent Researcher
32 PUBLICATIONS **597** CITATIONS

SEE PROFILE

Security Attacks Against the Availability of LTE Mobility Networks: Overview and Research Directions

Roger Piqueras Jover
AT&T Security Research Center
New York, NY 10007
roger.jover@att.com

***Abstract*—Modern LTE (Long Term Evolution) cellular networks provide advanced services for billions of users that go beyond traditional voice and short messaging traffic. The recent trend of Distributed Denial of Service (DDoS) attacks impacting the availability of communication systems illustrate the importance of strengthening the resiliency of mobility networks against Denial of Service (DoS) and DDoS threats, ensuring this way full LTE network availability against security attacks. In parallel, the advent of the Advanced Persistent Threat (APT) has capsized the common assumptions about attackers and threats. When it comes to very well planned and funded cyber-attacks, the scale of the threat is not the key element anymore. Instead, scenarios such as a local DoS attack, for example, against the cell service around a large corporation's headquarters or the Stock Exchange become very relevant. Therefore, traditionally overlooked low range threats, such as radio jamming, should not be de-prioritized in security studies.**

In this paper we present an overview of the current threat landscape against the availability of LTE mobility networks. We identify a set of areas of focus that should be considered in mobility security in order to guarantee availability against security attacks. Finally, we introduce potential research directions, including a new attack detection layer, to tackle these problems. The final goal is to rethink the architecture of a mobility network within the current security context and threat landscape and considering the current evolution towards a near future scenario where nearly every electronic device will be connected through Machine-to-Machine (M2M) systems.

I. Introduction

Modern cellular networks support a large number of services that go beyond traditional voice and short messaging traffic to include high bandwidth data communications. These networks are based on 3GPP standards for wireless communications, such as the Universal Mobile Telecommunication System (UMTS) for current 3G access networks. Release 8 of the 3GPP standards resulted in the deployment of Long Term Evolution (LTE). This new technology is characterized by great enhancements in the Radio Access Network (RAN) for capacity improvement in terms of bits per second per Hertz (bps/Hz) as well as a redesign of the cellular core network (Enhanced Packet Core - EPC), moving towards an all-IP system.

Despite the tremendous capacity and system enhancements implemented by LTE, in general, cellular networks are known to be vulnerable to security attacks [1], [2]. In parallel, a recent DDoS campaign against an anti-spam blacklisting service resulted in substantial impact on global communication networks and widespread service degradation [3]. This has sparked the general interest and concern in attacks against the availability of communication networks, which could be either affected despite not being the target (like in the case of the Spamhaus DDoS attack) or be the actual target. However, theoretical vulnerabilities of mobility networks had been known and published for a few years. Billions of users depend on cellular networks on a daily basis. Therefore, the consequences of a Denial of Service (low traffic volume) or a Distributed Denial of Service (high traffic volume) attack against the mobility network could be severe. For example, with the current outbreak of mobile malware, the possibility of a botnet of infected cell-phones launching an attack against the cellular network is closer to reality. An actual milder version of such a scenario was already observed in the wild due to a poorly programmed application which caused severe service degradation on a cellular carrier [4].

The threat scenario against communication networks is drastically changing. Not only the surge of hacktivism is threatening information systems, but also the relevance of the latest player, the Advanced Persistent Threat (APT), in the security ecosystem is growing fast. A recent report released information on a succession of highly sophisticated and long cyber-attacks against all kinds of institutions and enterprises all over the world [5]. The advent of APT brings into the equation the concept of a highly sophisticated and well-funded attacker. In this context, traditionally overlooked threats are very important. When it comes to very well-planned and funded cyber-attacks, even small-scale threats can be important. For example, scenarios such as a local DoS attack against the cell service around a large corporation's headquarters or the Stock Exchange become very relevant. This motivates further the importance of enhancing the security of mobility networks.

GSM (Global System for Mobile Communications) mobility networks were designed two decades ago to address issues in the previous cellular system (Advanced Mobile Phone System, AMPS), namely privacy and authentication. The GSM encryption and authentication algorithms were appropriately enhanced in UMTS and LTE by including new encryption options but most importantly by requiring mutual authentica-

tion. However, the threat landscape and computational power have evolved much faster, with no significant updates in the overall security architecture . A proactive effort is necessary to guarantee the full availability of mobility networks against security attacks.

Analysis is required to determine the conditions that make possible DoS and DDoS attacks against LTE as well as a clear assessment of the impact and severity of such attacks. Based on such a study, solutions and mitigation strategies should be proposed with the overall short-term goal to make mobility networks more secure and resilient to threats ranging from local radio jamming to complex DDoS threats targeting essential EPC elements, such as the Home Subscriber Server (HSS).

Moreover, the long term goal should be to rethink the architecture of a mobility network, originally designed just to guarantee privacy and authentication, for security and with the current threat scenario in mind. Proactive efforts must be taken in order to generate input and recommendations for future standard releases and technologies in order to ensure network availability.

In this paper we introduce the current research initiative being carried out in order to address the problem of availability of LTE mobility networks. We first present an overview of DoS and DDoS attacks against LTE cellular networks. Based on this analysis, we identify the main research areas that should be addressed to guarantee full mobility availability. On one hand, fast detection algorithms and mitigation strategies are being defined. On the other hand, new network architectures are being proposed as potential input and recommendations for future technologies and standards, rethinking the network with a security perspective. All these are being done within the context of the new mobility ecosystems. We are progressing to a near future scenario where most electronic devices will be connected to the network through Machine to Machine (M2M) systems [6], giving shape to the Internet of Things (IoT) [7]. At the same time, wireless access networks are becoming highly heterogeneous and complex, combining cellular deployments with other advanced access schemes (such as metropolitan micro- and pico-cells and user deployed femto-cells) and Radio Access Technologies (RATs), such as Wireless Local Area Networks (WLANs).

New research directions are proposed to tackle the major security concerns and architectural challenges of LTE, covering all the network layers. For example, the Physical Layer (PHY) should be revisited to address the growing threat of new sophisticated radio jamming attacks [8], [9]. Moreover, the mobility network architecture should be flattened and distributed to prevent large loads of signaling traffic in the LTE EPC as a result of common NAS (Non-Access Stratum) operations, such as idle-to-connected and connected-to-idle Radio Resource Control (RRC) state transitions. Such signaling overloads are known to be a potential way to attack a mobility network [10]. Note that this challenge becomes highly important with the expected rapid increase of the number of connected devices.

The remainder of this paper is organized as follows. First, some important background details are briefly discussed in Section II. Section III reviews the threat landscape and related attacks against the availability of mobility networks. In Section IV we discuss methods to improve the security in mobility networks along with further research directions in this area. Finally, in Section V we sketch potential attack detection strategies and in Section VI we present the concluding remarks.

II. Background

Mobile communication networks are rapidly evolving into complex systems both in terms of the network architecture and the types of connected devices. This increasing complexity naturally results in an increasing number of security threats. This section presents an overview of aspects relevant to mobility network security.

A. Mobility network vulnerabilities

Security research on wireless systems has been increasing throughout academia and industry over the last few years. One possible reason for this is the widespread availability of open source platforms that support wireless protocols. For example, the open source GSM project OpenBTS [11] has significantly decreased the cost to research GSM and thus has spurred a large amount of security research focused on GSM networks [12] along with the implementation of certain attacks [13], [14]. However, at this time, not much research has been focused on LTE networks.

B. M2M traffic scalability and signaling load impact

The convergence of the Internet and cellular mobility networks is enabling new M2M communication systems as part of the Internet of Things [7]. The industry consensus is that there will be drastic growth in mobile cellular connectivity due to M2M and embedded mobile applications. The majority of M2M systems currently operate on 2G and 3G networks but, in the long term, everything is expected to transition to LTE. Some predict that 50 billion non-personal data-only mobile devices will be on existing networks in the near future [6].

The emergence of the IoT and the spike in the number of connected devices could have signaling load implications on the cellular core network [15]. It will be necessary to optimize how M2M nodes utilize network resources. Even with the enhancements made to LTE, Machine-to-Machine traffic is expected to significantly affect the network [16]. The expected number of devices trying to connect wirelessly may be sufficient to overwhelm the network due to high signaling traffic volume.

In the same context, the combination of the IoT and the transition towards IPv6 could drive this trend to a point where every single consumer item is IP addressable. Mobile devices are not normally directly addressable from the Internet; mobile traffic is routed through the P-GW (Packet Gateway) which has NAT (Network Address Translation) functionality and thus effectively firewalls off incoming connections. However, it is

possible that future M2M services will require to be addressable from the Internet in order to deploy new services. This will certainly open new attack vectors, especially for multi-homed devices, and is an area that needs to be investigated.

C. Heterogeneous networks (metrocells, femtocells and WiFi)

Mobility networks have evolved over the last few years to become highly-complex heterogeneous systems such as the one depicted in Figure 1. One example of heterogeneity is found within cellular networks, where radio resources are optimized by deploying smaller cells. High density metropolitan areas with large traffic demands are provisioned with micro-, pico, and femtocells that provide coverage to areas less than 100 meters in radius.

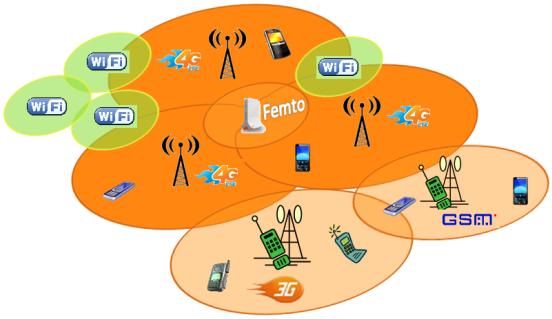

Fig. 1. Example of mobile network heterogeneity

In particular, femtocells are rapidly becoming a popular and low cost solution to enhance the coverage and capacity of mobile systems. Femtocells are low-power base stations that are installed by end users and backhauled to the EPC over the broadband IP connection available at the user's premises (cable, fiber, DSL, etc). A new network node, the Femtocell Gateway (FGW), enables the interconnection of femtocells with the network over an encrypted IP channel. Such a connection must be both highly secure and well authenticated in order to avoid misuse of this alternative entry into the internal mobility network.

Further heterogeneity of mobility networks is caused by the number of different RATs that they support. In the absence of LTE coverage, User Equipment (UE) will establish a connection to UMTS or even GSM networks. While service availability is extremely important, falling back to these other protocols may open up the device to attack [14]. Moreover, some cellular operators deploy WiFi access points on very dense areas and configure UEs to automatically camp on them when available to spare as much load as possible from the highly congested cellular systems.

Each different method that allows UE to access the mobility network increases the complexity of the system. New attack vectors potentially appear if this interconnectivity and heterogeneity is not properly addressed. For example, researchers recently presented ways to hack into femtocell access points to gain root access to the device [17]. In this situation, the common assumption that an attacker does not own or control network elements is not valid anymore.

III. Attacks against the availability of LTE

DoS and DDoS attacks in LTE mobility networks can be classified based on the traffic load maliciously generated: one single attacker or low traffic volume (DoS) and a large combination of multiple simultaneous attackers or high traffic volume (DDoS). Note that a special class of attack is defined for the case of the attacker being already within the network perimeter and, therefore, not requiring a charge of malicious traffic. This is the case of an insider attack.

A parallel classification is based on the impact of the attack. Some attacks have a local scope, disrupting service at the RAN level and blocking service for a single cell or sector. Other attacks can have a much wider scope, and are capable of disrupting a large portion of the mobility network. Note that, in some situations, local attacks may be combined so that they affect a larger area. The combination of both attack categories plus the insider attack is depicted in Figure 2.

It is important to note that several other types of attacks could be launched from or through a mobility network, such as malware spreading, phishing or even data exfiltration in the context of an APT [18]. Also, a botnet of UEs could be leveraged to enhance the severity of a DDoS attack against a specific target in the Internet. These threats are added to Figure 2 for completion but do not fall in the category of attacks against the availability of LTE and are, therefore, out of the scope of this paper.

DDoS against the LTE EPC can exploit either single points of failure or amplification effects inherent to the operation of mobility networks. For example, the successful operation of the LTE network depends on a central authentication node, the Home Subscriber Server (HSS). In parallel, some theoretical studies have pointed out the potential risk that amplification attacks present against the EPC. Specifically, it is well known that a single simple event on the phone side (a state transition in the RRC state machine) requires a substantial number of messages exchanged among several EPC nodes. This could theoretically be exploited to become a DDoS attack [10], [19]. Such an overloading has already been observed in the wild with a non-malicious origin. A major US cellular operator had part of its network highly saturated due to an instant messaging app update - installed on many smartphones - that checked very often with a server [4]. This resulted in a large number of connect/disconnect events at the RRC engine of the EPC which, as a result, generated a very large load on the EPC.

Local attacks, such as radio jamming and saturation of the wireless interface, can be launched from a single device or radio transmitter and have recently become more sophisticated with an increased impact [8], [9]. Other more complex local threats against the RAN originating at multiple cell phones should be considered as well, such as Theft of Service (ToS) and protocol misbehavior. Although these are not availability attacks per se, they can potentially degrade the Quality of Service (QoS) for legitimate customers.

Finally, insider attacks defy the common assumption that an attacker cannot have access to any node of the network.

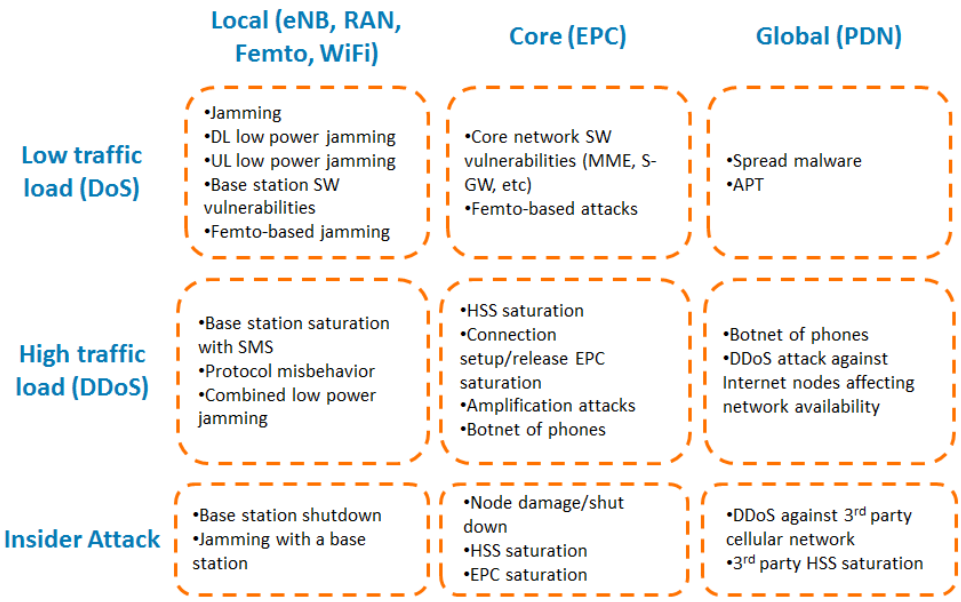

Fig. 2. DoS and DDoS attacks in mobility networks

An insider is an individual with elevated rights and access to specific network elements. Such privileges could be maliciously exploited to disrupt communications locally or at a global network scale. This attack category is unlikely but should be included in a complete analysis given the current threat scenario.

A. Denial of Service attacks (DoS)

Following the overview in Figure 2, this sub-section briefly introduces the details of DoS attacks against a mobility network. The focus is mostly on local attacks, i.e. jamming, as well as threats against the RAN that could be leveraged from a single attacker.

1) ***Radio jamming****:* Radio jamming is the deliberate transmission of radio signals to disrupt communications by decreasing the signal to noise ratio. This attack has been studied in the literature in the context of cellular communications and essentially consists of blasting a high power constant signal on the entire target band [20], [21]. Although one way to block this attack is to locate and stop the jamming device, the large amount of power required reduces the effectiveness of the attack.

2) ***(Low Power) Smart Jamming****:* Smart jamming consists of attacks that aim to locally disrupt the communications of an LTE network without raising alerts. This can be done by saturating one or more of the essential control channels required by all mobile devices to access the spectrum. Saturation of these channels would make the network appear unresponsive. Moreover, given that this attack requires low transmitted power and requires no authentication, detection and mitigation are very difficult.

This type of attack can be launched against essential control channels in both the downlink and the uplink. Instead of saturating the entire channel, this attack concentrates on the much narrower control channels and so consumes less power. The fact that the radio resource allocation of the main LTE downlink synchronization and broadcast channels (PSS, SSS and PBCH) is known a priori makes this a very simple improvement over basic jamming. By tuning a regular off the shelf radio jammer at the central frequency of the LTE band and transmitting at a bandwidth of at least 1.08MHz, an attacker would block reception of the aforementioned downlink control channels [9].

Uplink smart jamming targets LTE uplink control channels. The required bandwidth is substantially lower than Downlink Smart Jamming and, given that the attacker is competing against low-power UEs, the required power is very low, too. However, this category of attack might require an advanced jammer that is able to fully synchronize with the LTE signal.

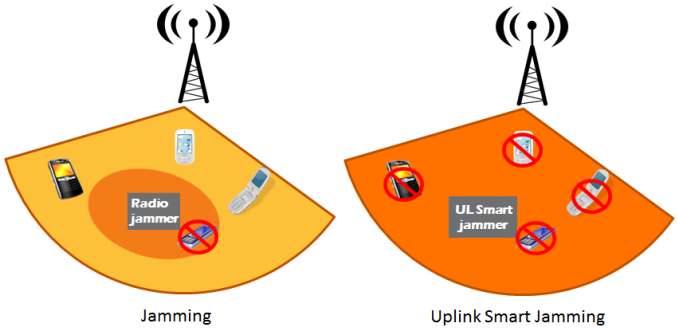

Fig. 3. Impact range of radio jamming vs UL smart jamming

Note that, unlike a traditional jamming attack, which is localized around the actual attacker, an uplink smart jamming attack would deny the service to all the users within the target cell or sector. Therefore, as it can be seen in Figure 3, while still being a local DoS attack, smart jamming has a wider

range.

Based on this vulnerability, a well organized group of attackers could simultaneously activate jamming devices in order to block access to the network over several contiguous cells. This attack could potentially be optimized by equipping each radio jammer with multiple directive antennas and targeting this way multiple sectors or cells.

In jamming mitigation studies, the goal is to force any sophisticated attack to be just as efficient as basic jamming [22]. Therefore, LTE security systems should be able to mitigate or block smart jamming attacks. We are proposing certain mitigation strategies based on leveraging the multiple antennas per cell/sector available at most cell sites [8]. These arrays of antennas are currently only being used to improve the performance of the PHY layer in terms of bit error rate and minimum received power. Following similar ideas as the spatial multiplexing concepts planned for Advanced LTE, multiple receiving antennas can be leveraged to perform beamforming in order to detect interference and block its effects.

The first instance of smart jamming was proposed a few years ago in the context of GSM networks [13]. Also, the authors of [12] suggested the theoretical feasibility of locally jamming a GSM base station with a constant load of text messages. The fact that, in legacy GSM networks, text messages share resources with signaling channels makes such an attack possible. Recent results indicate that it is theoretically also applicable to UMTS-based networks, but with less intensity [23]. However, it is important to note that, in LTE systems, SMS traffic does not share resources with system signaling messages. Therefore, an LTE SMS-based flood of RAN signaling channels is not possible.

A basic form of smart jamming against LTE was presented recently as well [9] as a response to developing a public safety LTE-based network.

*3) **Classic computer vulnerabilities**:* Cellular equipment and the software running on mobility networks are similar to any other computer system and so can be affected by the same large class of vulnerabilities. For example, there is no a-priori reason to expect that firmware and software on telecom equipment does not have the class of vulnerability referred to as buffer overflows. Although normally one would expect an attacker to use a buffer overflow to cause code to execute on the attacked device, a failed attack or even exploratory investigations of a buffer overflow could cause equipment to crash and, hence, cause a DoS condition.

A good example of this is a recent effort on baseband fuzzing of mobile devices [24]. It identified zero day bugs that caused the baseband or the entire device to crash when parsing malformed text messages.

B. Distributed Denial of Service attacks (DDoS)

This section provides an overview of potential DDoS attacks launched by a botnet of mobile devices or a high volume of malicious traffic against the mobility network. The case of a similar botnet of mobile terminals leveraged to attack an external network or node is considered as well for completion.

*1) **Botnet of mobile devices**:* Research studies on botnet detection and tracking are starting to consider the possibility of a mobile botnet conformed by a large number of smartphones. The authors of [25] stated that, given the potential economic incentives and impact of a mobile botnet, these are likely to appear and spread on current cellular networks. This study demonstrates the availability of multiple platforms for command and control messaging exchange and the feasibility of the successful operation of such a botnet. The surge in malware instances and successful spread infections enhances the likelihood of such a scenario. In fact, a recent study reported the discovery of an Android malware-based botnet that activates smart phones as text message spam platforms [26].

Based on these assumptions, a smart-phone botnet presents a new and very powerful attack vector against mobility networks. As a result, a new set of DDoS attacks is possible when large volumes of traffic and signaling messages can be generated from within the network. We categorize such threats in the following subsections.

*2) **Signaling amplification attacks**:* Mobility networks do not have sufficient radio resources to provide service to every single customer at the same time. Typically, resources are deployed to be able to sustain peak traffic hours and, in the event of load spikes known a priori (e.g. a large technology festival) extra capacity can be temporarily deployed [27].

The scarcity of bandwidth requires advanced techniques to reuse idle resources in an efficient manner. The RRC engine of the network reassigns radio resources from a given user when the connection goes idle for a few seconds. When an inactivity timer expires, the radio bearer between the mobile device and the core network is closed and those resources become available to be reassigned to another UE. At this stage, the UE moves from connected to idle state.

Each instance of bearer disconnection and setup involves a significant number of control messages exchanged among nodes within the EPC. This signaling load, if not properly managed, can result in large-scale saturation of the network which could be exploited in the context of a DDoS attack [10]. Such impact has already been seen in the wild. An instant messaging application that was poorly designed checked for new messages with a server too often and flooded portions of the cellular network of one of the major providers in America [4].

A botnet of infected mobile devices could be used to generate a signaling amplification attack by forcing each terminal to constantly establish and release IP connections with an external server [19]. A piece of malware could also trigger mobile phones to reboot at the same time, thereby potentially overloading the EPC with registrations once they come back up. Such saturation of the EPC could potentially also occur legitimately due to the overwhelming amount of traffic and frequent reconnections of billions of M2M nodes [16].

*3) **HSS saturation**:* The HSS is a key node of the EPC that stores information for every subscriber in the network. Some of the parameters stored per user are the phone number (”pub-

lic" id), the International Mobile Subscriber Identity (IMSI) ("private" id), billing and account information, cryptographic primitives and keys to perform authentication of subscribers and also the last known location of the user.

This essential node is the provider of authentication of users and is the cornerstone of the paging infrastructure. Therefore, a DDoS attack against this node could potentially prevent the network from being operated. Some research work in the academia has explored the possibility of overloading the 3G Home Location Register (HLR) leveraging a botnet of mobile terminals [28]. It is important to note that the HSS is involved in a substantial number of signaling events in the EPC and could suffer as well the consequences of the aforementioned signaling amplification attack.

*4) **DDoS against external nodes/networks**:* Recently, major banking institutions were the target of some of the largest DDoS attacks ever seen in communication networks [29]. These attacks originated from a number of servers that were remotely controlled by an attacker and were able to inject large traffic loads into the network. Although the target of DDoS attacks is not the network itself, the recent DDoS attacks against Spamhaus resulted in substantial impact against the availability of communication networks [3].

In this context, the high volume of traffic aimed to a specific target during a DDoS attack could originate at a botnet of mobile phones and, therefore, potentially impact the performance of the mobility network.

C. Insider attacks

A complete security analysis of mobility communication networks should address the concept of an insider threat, which is often ignored or assumed unlikely. This has many opportunities to architect sophisticated intrusion detection techniques for this attack category. However, with the current security threat landscape and the advent of the APT [5], an insider attack becomes highly relevant.

A very well funded attacker could persuade an insider to perform an attack against the availability of the mobility network . This would change a common assumption in mobility network security: the attacker could own and control an internal network node.

An insider could physically or remotely shut down a network node. Unless this node was the HSS, the impact of the attack would not be global. Beyond a strong security perimeter and remote access control to the HSS, the privileges and access granted to employees should be carefully planned and designed in order to minimize, if not totally deny, the reachability of the HSS and other important EPC nodes.

However, access control protection to essential elements of the EPC is not the optimal solution. To guarantee full availability of a mobility network against an insider attack, this new dimension of security threat must be analyzed carefully and addressed by specific access control policies, perimeter delimitation, and new advanced techniques.

D. Overview of threats against mobility availability

In this section we present a brief overview of the aforementioned attacks against the availability of LTE mobility networks. The main threats are summarized in Table I based on the attack platform, the scope, the difficulty or cost to launch such an attack and, finally, an estimate of the impact against the availability of an LTE network. The larger the impact, the more the availability is affected within the scope of the attack.

A smart jamming attack, despite being local, can potentially block one or multiple sectors at a very low cost. The cost of a signaling amplification attack or a threat against the HSS is larger because it requires a large botnet of infected devices. In the latter case of a DDoS against the HSS, the range of the attack could be potentially global.

The case of an insider threat is also considered. Once an insider goes rogue, the potential attacks have low cost and potentially high impact. Finally, a botnet of UEs could be leveraged to launch or enhance a DDoS attack against an external target.

IV. New security-oriented network architecture

Wireless cellular networks were originally designed to provide ubiquitous access for communication. Although the second generation of mobility networks was designed with some security aspects in mind, GSM just featured cryptographic algorithms to guarantee privacy and authentication. The GSM security architecture, proposed two decades ago, is nowadays known to be insufficient given current computational power [14]. UMTS-based 3G networks enhanced the system by implementing stronger encryption and a two-way authentication scheme. Both encryption and authentication are further enhanced in LTE. However, with the current threat landscape and the increasing sophistication of attacks, such security architecture is not enough to guarantee the availability of mobility networks.

Mobility networks are not designed to guarantee availability beyond redundancy and resiliency to network outage. The inherent characteristics and operation modes of mobility networks are strongly bound to centralized nodes and essential control channels. In other words, large clusters of mobility users are tightly dependent on certain specific nodes (e.g. the HSS in LTE) and all the devices within a given cell or sector transmit and receive essential control traffic on shared channels (such as the Physical Broadcast Channel -PBCH).

In this section we introduce potential directions for a redesign of the mobility network architecture for security, with a goal of full mobility availability against security attacks.

A. Main areas of focus for security enhanced network architecture

The analysis of the threat scenarios presented in this paper, DoS and DDoS attacks against mobility networks, allows us to identify specific areas of focus. In order to achieve full availability and resiliency of cellular networks against security attacks, efforts should be carried out in the following areas.

Threat	Platform	Range	Difficulty	Impact
Smart Jamming	1 RF/SW-defined radio device	Local (cell/sector)	Low cost and complexity	High (but local)
Signaling amplification	Botnet of infected UEs	Large portion of a national network	Medium (10K-100K infected UEs)	High
HSS saturation	Botnet of infected UEs	Potentially global	Medium (10K-100K infected UEs, avoid attack throttled at EPC)	Very high
External DDoS	Botnet of infected UEs	DDoS target	Medium	Potentially high
APT	Insider	Local to global	Low	Very high

TABLE I
OVERVIEW OF ATTACKS AGAINST THE AVAILABILITY OF MOBILITY NETWORKS

Broadcast and Control Channel protection for enhanced jamming resiliency: Radio jamming is a common threat for all kinds of wireless network. On top of designing jamming mitigation and blocking techniques, it is important to ensure that the main control and broadcast channels of a mobility network are protected against radio jamming. This can prevent smart jamming attacks, through which an attacker could block the access over up to an entire cell by means of a low power and low bandwidth signal. An initial proposal on security solutions tackling this problem is presented in [8].

Initial access to the network: A random access procedure to request resources is the first step in the initial access to the network and the transition from idle to connected state. Such procedure is carried out on an uplink shared control channel, the (Random Access Channel) RACH. This resource is often the first source of network congestion due to a legitimate traffic spike. The characteristics of random access procedure can be leveraged in the context of malicious DoS attacks. The increasing sophistication of attacks plus the increasing number of connected devices and network load require the design of distributed network initial access techniques. Concepts of cognitive radio and reuse of legacy networks could be applied to design more robust radio resource allocation architectures. Optimization of the RACH procedure and a flexible adaptation to changes in traffic and channel conditions are already within the scope of the design of Self Organizing Networks [30].

RRC bearer management: The scarcity of spectrum results in complex radio resource management techniques at the RAN. LTE Physical Resource Blocks (PRBs) cannot be permanently allocated to a given device so, by default, UEs are moved to a disconnected state after they have been idle for a certain time. This on-demand resource reuse strategy, combined with the cost of cellular infrastructure, results in cellular networks provisioning enough capacity at the RAN and the EPC to handle the traffic during the busiest times of the day. RRC algorithms activate and deactivate traffic bearers depending on the UE traffic activity. Each bearer establishment and release generates a substantial number of signaling among the nodes in the EPC, which can potentially saturate network elements or connectivity links with bursty spikes of traffic. Initial efforts of designing LTE networks with a more distributed bearer management procedure have to be continued in order to distribute EPC signaling load and minimize its impact [15]. This way the scalability of the IoT over mobility networks could be achieved.

Core network signaling: The evolution to LTE has made great efforts in designing a flat and flexible network . Nevertheless, the NAS signaling load at the EPC has strong relevance on network security [15]. This very complex problem results in large amounts of traffic exchanged among EPC entities and between the EPC and the RAN each time a NAS network function is executed. This results in a sub-optimal network architecture that can be exploited maliciously in the context of DDoS attacks. A flexible and rapidly adapting architecture is required to minimize the NAS signaling load at the EPC and provide mitigation features to balance, re-route or filter network control traffic. Such functionality can potentially be achieved by means of strong SoN and software-based network nodes.

Central Node Dependency: Mobility networks strongly depend on specific logically centralized nodes. For example, essential authentication and billing functions are carried out by the HSS periodically. This way, the consequences of a DDoS or an insider threat against such a node could be severe, potentially denying access to mobile communications over very large geographical areas. Distributed solutions should be implemented in order to reduce the dependance on centralized nodes. Two possible solutions to be explored include partial local replicas of the content of the HSS closer to the RAN and optimization of the signaling handshake for connection management and authentication procedures. Software-defined cellular networks running in the cloud also offer promising solutions by adaptive tuning the capacity and processing of centralized nodes (i.e. assigning more CPU resources to the HSS or distributing it among several virtual machines) as a result of a legitimate traffic spike or an attack.

A redesign of the network architecture based on software-based mobile network might not require an intense modifica-

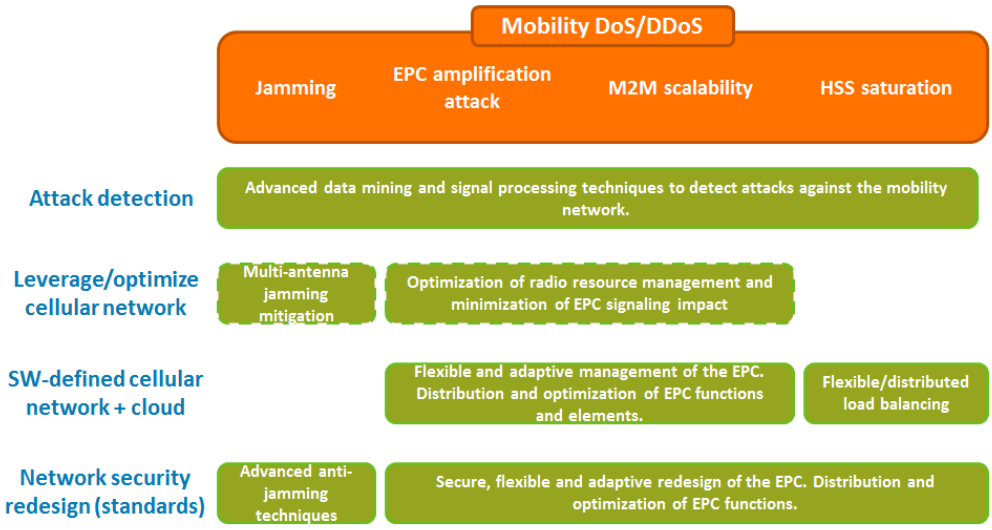

Fig. 4. Research directions for mobility availability against security attacks

tion of basic LTE standards. Instead, new security-oriented proposals and recommendations would be included in the standards for self organization and self healing. Nevertheless, new RAN security solutions to enhance the resiliency against radio jamming and to optimize the initial access to the network would require modifications to the standards.

B. Security research directions

The mobility network redesign for security described in Section IV-A should follow three main directions, with increasing security benefit but with increasing complexity as well. These three directions are summarized in Figure 4 with their impact on four examples of availability threats against mobility (radio jamming, EPC signaling based threats, M2M scalability and HSS saturation). These four cases have been selected as a sample of the overall security of mobility availability problem. Note that as indicated in the figure, a successful enhancement of mobility network security must be built together with a strong and effective attack-detection engine. More details on detecting mobility attacks are given in Section V.

The first building block of a mobile security architecture should leverage the power of the network as is. For example, the availability of multiple antennas at the eNodeB enables the possibility of advanced anti-jamming techniques based on multi-antenna and beam forming technology [8]. Such a modification at the PHY layer would require minimum changes to the network, from a standards and deployment point of view, and would offer strategies to tackle radio jamming threats.

The traffic scalability and signaling load should be analyzed, through simulations, by stressing the network with unbounded traffic and device growth to give insights on the EPC signaling load architecture. Based on these results, the configuration of the current network could be modified to optimally handle NAS messaging and bearer-related signaling traffic, providing a certain degree of mitigation against DDoS attacks.

Note that leveraging the current network design can provide a first layer of protection from mobility attacks, but does not fully address the problem. With a more forward-looking approach, security research should propose both a new implementation of cellular networks and an actual redesign of the network architecture for security enhancements.

As the next building block for a new mobility security architecture, software-defined cellular networks offer many potential benefits by fully or partially deploying the EPC in the cloud. Such flexible approaches would provide new means for a secure, flexible and adaptive management of the EPC. A more efficient load processing and balancing among nodes would be possible. Moreover, the processing capabilities of network nodes could flexibly be enhanced in the event of a legitimate or malicious spike in user or signaling traffic. In the context of a DDoS threat, the nodes under attack could be either replicated or assigned more processing capacity.

Finally, security efforts should impact upcoming standards and technologies. The goal is to rethink the architecture of a mobility network, originally designed only to guarantee encryption and authentication but without the current security context and threat landscape in mind. Such redesign should also consider the current evolution of mobility networks, progressing to a near future scenario where nearly every electronic device will be connected to the network through M2M systems and the IoT.

Similar directions have been proposed for the development of efficient and flexible mobility network architectures, able to serve the still unknown needs and preferences of future users [31].

V. Mobility Network Attack Detection

As depicted in Figure 4, on top of all the security enhancements will lay an advanced attack detection layer. Efficient data mining and machine learning techniques will ensure the rapid and accurate detection of security attacks against the

mobility network, automatically triggering the appropriate defenses and self-healing functionalities. Some of these might be designed and proposed within the scope of SoN, establishing the context of secure and self recovering networks.

The network attack detection capability should be able to sense both large scale DDoS-type of attacks as well as more subtle threats, such as much localized low volume DoS attacks or other network and traffic anomalies. An effective detection engine should leverage the power and computational resources of the network and the cloud and leverage data that is available at all stages and layers of the network.

In parallel, a localized detection layer should monitor for local low traffic threats against the RAN. This RAN-level detection layer should generate constant feedback to the network-based global detection engine because the aggregation and correlation of localized attacks could be indicative of a larger attack at a higher layer.

VI. Conclusions

In this paper we presented an overview of the current availability threat landscape of LTE mobility networks, covering both local DoS attacks against the RAN and largely distributed DDoS attacks aiming to saturate the EPC or to simultaneously block multiple cells at the RAN level. In parallel, considering the recent drastic changes on the security assumptions resulting from the advent of the APT, we include the insider threat in our study.

Mobility networks were initially designed primarily to guarantee privacy and authentication. However, in the context of the current threat landscape, security research efforts are necessary in order to achieve full mobility availability.

To achieve this goal, we propose three major security research directions, as well as an effective and efficient network-based attack detection layer. As a first step, the capabilities of current mobility networks should be leveraged, reconfigured and adapted for security enhancement. In the mid term, architecture changes such as, software-defined cellular architectures, with full or partial deployments of the EPC in the cloud, potentially provide a strong enhancement of resiliency against DDoS attacks. Finally, the mobility network and security architectures should be completely rethought to support a scenario in the not-too-distant future when nearly every electronic device will be connected to the network.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Gustavo de Los Reyes and Dr. Joshua Lackey for their valuable help, comments and suggestions.

References

[1] E. Gadaix, "GSM and 3G security," in *In BlackHat Asia*, 2001, http://tinyurl.com/85plhlv.

[2] P. Traynor, P. McDaniel, and T. La Porta, "On attack causality in internet-connected cellular networks," in *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*, 2007, pp. 21:1–21:16.

[3] J. Markoff, "Firm Is Accused of Sending Spam, and Fight Jams Internet," The New York Times, March 2013, http://goo.gl/76ipU.

[4] M. Dano, "The Android IM app that brought T-Mobile's network to its knees," Fierce Wireless, October 2010, http://goo.gl/O3qsG.

[5] D. Alperovitch, "Revealed: Operation Shady RAT," Threat Research, McAfee, 2011, http://goo.gl/ATL2X.

[6] "More than 50 billion connected devices," Ericsson, Ericsson White Paper, February 2011, http://goo.gl/KGaVg.

[7] A. Iera, C. Floerkemeier, J. Mitsugi, and G. Morabito, "Special Issue on the Internet of Things," in *IEEE Wireless Communications*, vol. 17, December 2010, pp. 8–9.

[8] R. Piqueras Jover, J. Lackey, and A. Raghavan, "Enhancing the security of LTE networks against jamming attacks," February 2013, Under submission.

[9] Talbot, David, "One Simple Trick Could Disable a Citys 4G Phone Network," MIT Technology Review, November 2012, http://goo.gl/PnqHf.

[10] P. Lee, T. Bu, and T. Woo, "On the detection of signaling dos attacks on 3g wireless networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, May 2007.

[11] Kestrel Signal Processing, Inc., "The OpenBTS Project," http://openbts.sourceforge.net/.

[12] P. Traynor, W. Enck, P. Mcdaniel, and T. La Porta, "Exploiting open functionality in sms-capable cellular networks," in *J. Comput. Secur.*, vol. 16. Amsterdam, The Netherlands, The Netherlands: IOS Press, December 2008, pp. 713–742.

[13] D. Spaar, "A practical DoS attack to the GSM network," in *In DeepSec*, 2009, http://tinyurl.com/7vtdoj5.

[14] K. Nohl and S. Munaut, "Wideband GSM sniffing," in *In 27th Chaos Communication Congress*, 2010, http://goo.gl/wT5tz.

[15] 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects, "Study on Core Network overload solutions. 3GPP TR 23.843," vol. v0.7.0, 2012.

[16] A. Prasad, "3GPP SAE-LTE Security," in *NIKSUN WWSMC*, July 2011.

[17] N. Golde, K. Redon, and R. Borgaonkar, "Weaponizing femtocells: The effect of rogue devices on mobile telecommunications," in *Annual Network & Distributed System Security Symposium*, 2012.

[18] R. Piqueras Jover and P. Giura, "How vulnerabilities in wireless networks can enable Advanced Persistent Threats," in *International Journal on Information Technology (IREIT)*, 2013.

[19] R. Bassil, A. Chehab, I. Elhajj, and A. Kayssi, "Signaling oriented denial of service on lte networks," in *Proceedings of the 10th ACM international symposium on Mobility management and wireless access*. ACM, 2012, pp. 153–158.

[20] M. Stahlberg, "Radio jamming attacks against two popular mobile networks," in *Helsinki University of Technology. Seminar on Network Security. Mobile Security*, 2000.

[21] W. Xu, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *ACM MOBIHOC*, 2005, pp. 46–57.

[22] T. Clancy, "Efficient ofdm denial: Pilot jamming and pilot nulling," in *Communications (ICC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–5.

[23] I. Murynets and R. Piqueras Jover, "How an SMS-Based malware infection will get throttled by the wireless link," in *IEEE ICC 2012 - Communication and Information Systems Security Symposium (ICC'12 CISS)*, Ottawa, Ontario, Canada, June 2012.

[24] M. Vuontisjarvi and T. Rontti, "SMS fuzzing," Codenomicon, http://goo.gl/C0pXj.

[25] C. Mulliner and J.-P. Seifert, "Rise of the ibots: 0wning a telco network," in *Proceedings of the 5th IEEE International Conference on Malicious and Unwanted Software (Malware)*, 2010.

[26] "Security Alert - SpamSoldier," The Lookout Blog, December 2012, http://goo.gl/7lkRM.

[27] J. Donovan, "Helping Fans Connect and Share at SXSW," AT&T Innovation Space, March 2013, http://goo.gl/vy5w8.

[28] P. Traynor, M. Lin, M. Ongtang, V. Rao, T. Jaeger, P. McDaniel, and T. La Porta, "On cellular botnets: measuring the impact of malicious devices on a cellular network core," in *Proceedings of the 16th ACM conference on Computer and communications security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 223–234.

[29] D. Goldman, "Major banks hit with biggest cyberattacks in history," CNN Money, September 2012, http://goo.gl/4qCXG.

[30] M. Kottkamp, A. Rossler, J. Schlienz, and J. Schutz, "LTE Release 9: Technology Introduction," Rhode & Schwarz, White Paper, 2011, http://goo.gl/QeGPO.

[31] B.-j. Kim and P. Henry, "Directions for future cellular mobile network architecture," *First Monday*, vol. 17, no. 12-3, 2012.