Response to FCC NOI DA 16-1282

Some key challenges in securing 5G wireless networks

Roger Piqueras Jover
Bloomberg LP, New York, NY
rpiquerasjov@bloomberg.net

I. INTRODUCTION

Long Term Evolution (LTE) is the most recent cellular communication standard being deployed globally. Independent of, and co-existing with, previous generations of different technologies for mobile access, all operators globally have converged towards LTE for the current generation of mobile communication. Meanwhile, the 3rd Generation Partnership Project (3GPP) is already working on designing the next leap in mobile communication systems, generally referred to as Fifth Generation (5G).

LTE is characterized by a redesigned PHYSical layer, which is built upon Orthogonal Frequency Division Multiple Access (OFDMA), thereby providing orders of magnitude higher data rates and lower traffic latency, along with a strong resiliency to multipath fading and highly improved spectrum efficiency. This highly improved Radio Access Network (RAN) is operated by the Enhanced Packet Core (EPC) network to provide connectivity to all types of mobile devices.

Initially, LTE was not specifically intended for deployment of massive numbers of devices. Nevertheless, LTE has become one of the major cornerstones of the global deployment of Machine-to-Machine (M2M) communication systems, interwoven around what the industry defines as the Internet of Things (IoT). This deployment of - generally - low power and rather simple connected embedded devices has resulted in a myriad of new security challenges for mobile networks.

The growing demand for connectivity and fast data transfer, along with new trends such as IoT - for which current mobile architectures are far from appropriate - has triggered a major redesign of mobile standards - at all levels - in the context of 5G. The industry generally highlights five major goals for 5G networks, namely: 1) higher system capacity, 2) higher data rates - with gigabit per second (Gbps) being a common catch phrase, 3) reduced latency - with a rather optimistic, yet promising, target of under 10ms latency, 4) massive device connectivity and 5) energy savings [1]. Some of the most mature technologies already being tested to tackle such 5G demands are millimeter wave (mmWave) communication, with carrier frequencies well above the common 6GHz boundary, and massive MIMO (Multiple Input Multiple Output) arrays with hundreds of antennas.

As the cornerstone of today’s digital and connected society, LTE cellular networks deliver advanced services for billions of users, beyond traditional voice communication and short messaging. Moreover, mobile networks are the connectivity layer for critical communication infrastructure, from first responder systems [2] to ad-hoc military tactical networks [3]. Therefore, the security of mobile systems is of prime importance. After a rather unsuccessful service record, with the first generation lacking support for encryption, GSM networks being vulnerable to several exploits [4] and LTE recently having been found vulnerable to similar exploits [5], [6], the ongoing definition and design of 5G systems is the right time to implement some long overdue security enhancements for wireless networks.

Although there has been outstanding work from generation to generation to optimize the spectral efficiency, increase capacity and improve overall mobile networks, there has yet to be any holistic security redesign of the entire infrastructure. Mobile standards, despite the goal of advancing wireless technology, still rely heavily on outdated legacy technology, operations and equipment. This is periodically improved with new functionality being added and legacy operations being improved. However, there is still demand for a complete redesign using a preventive security approach. As a result, although encryption and authentication algorithms have been beefed up and security functions have been improved, modern mobile networks still carry over certain specific architectural vulnerabilities.

Most of the current protocol security threats at layer 2 in mobile networks span from this legacy security architecture. Despite the addition of sophisticated encryption algorithms, mutual authentication and other functions, mobile networks still implement a rather outdated symmetric key and circuit-switched architecture.

Although there are several areas in which security should be substantially improved, this document highlights those in which the author has been actively involved: security experimentation, exploit analysis and mitigation design, communication protocol security - including privacy and authentication -, and mobile network scalability.

II. AUTHENTICATION, PRIVACY AND PROTOCOL EXPLOITS IN THE CONTEXT OF 5G

The first generation of mobile networks (1G) lacked support for encryption and legacy 2G networks lack mutual authentication and implement an outdated encryption algorithm. Combined with the wide availability of open source implementations of the GSM protocol stack, this has resulted in the discovery of many possible exploits on the GSM insecure radio link [4].
Specific efforts were made to substantially enhance confidentiality and authentication in mobile networks, with much stronger cryptographic algorithms and mutual authentication having been explicitly implemented in both 3G and LTE. Because of this, LTE is generally considered secure given this mutual authentication and strong encryption scheme. As such, confidentiality and authentication are wrongly assumed to be sufficiently guaranteed. As it has been recently demonstrated, LTE mobile networks are still vulnerable to protocol exploits, location leaks and rogue base stations [5], [6].

It is of great importance that such exploits are addressed in the context of 5G mobile networks. In order to do so, the root cause of such security threats must be addressed. Although there are other areas where security should be enhanced, this manuscript focuses on the following:

- **Implicit trust in pre-authentication messages**: The security and integrity of mobile systems is vulnerable today due to the mere fact that mobile devices inherently trust all downlink pre-authentication messages coming from anything that appears to be a legitimate base station.

- **Legacy symmetric key security architecture**: The latest mobile standards still leverage their entire security infrastructure on a rather outdated, legacy symmetric key architecture. Symmetric key systems allow for strong authentication and encryption, but are not flexible enough to provide new security features to prevent basic downgrade or Denial of Service (DoS) attacks or address the aforementioned implicit trust in pre-authentication messages.

- **Unnecessary disclosure of location and radio configuration parameters**: Due to the fact that the overall security architecture of cellular systems has not been modified since legacy 2G networks\(^1\), it still carries over architectural security flaws and privacy leaks inherent to the actual day-to-day operation of a cellular system. Mobile networks should be rethought from top-to-bottom with a strong security focus in order to address such inherent architectural security vulnerabilities.

### A. Implicit trust on pre-authentication messages

Despite the strong cryptographic protection of user traffic and mutual authentication of LTE, a very large number of control plane (signaling) messages are regularly exchanged over an LTE radio link in the clear. Before the authentication and encryption steps of a connection are executed, a mobile device engages in a substantial conversation with *any* LTE base station (real or rogue) that advertises itself with the correct broadcast information. This results in a high threat due to the implicit trust placed, from the mobile device point of view, on the messages coming from the base station. A large number of operations with critical security implications are executed when triggered by some of these implicitly trusted messages, which are neither authenticated nor validated. It is rather obvious that, in the age of large scale cyber-attacks, one of the largest civilian communication systems should rely on privacy protocols far more sophisticated than just basic implicit trust anchored on the fact that the base station "looks like a legitimate base station".

Table I summarizes some of the pre-authentication messages that are implicitly trusted by any LTE mobile device, as well as some critical functions they can trigger. By exploiting such messages, one can set up a rogue access point that, despite not being capable of full "man-in-the-middle" connections, can render a mobile device useless (DoS), track its whereabouts (privacy threat), and instruct it to switch to an insecure GSM connection (downgrade attack\(^2\)). Note that security based on implicit trust is simply unacceptable in the context of wireless systems applied to first responders, national security and military tactical networks.

<table>
<thead>
<tr>
<th>Types of message</th>
<th>Messages</th>
<th>Critical functions triggered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radio Resource Control (RRC)</td>
<td>RRC Connection Request, RRC Connection Setup, RRC Connection Setup Response, RRC Connection Reconfiguration, etc.</td>
<td>Radio connection characteristics, mobility to a new cell, downgrade to legacy radio protocol, etc.</td>
</tr>
<tr>
<td>Non Access Stratum</td>
<td>Attach Request, Attach Response, Attach Reject, Location Update Request, Location Update Reject, etc.</td>
<td>Connection blocking, connection throttling down to legacy protocol, etc</td>
</tr>
<tr>
<td>Other</td>
<td>Paging, Measurement Update, etc.</td>
<td>Location measurements and location information</td>
</tr>
</tbody>
</table>

**Table I. Unprotected Pre-authentication Messages Implicitly Trusted in LTE Mobile Networks**

As discussed above, any mobile device trusts and obeys the messages listed in Table I as long as the base station advertises itself with the right parameters. As long as the mobile device decodes the expected broadcast information from the MIB (Master Information Block) and SIB (System Information Block) messages (i.e., the right MCC [Mobile Country Code] and MNC [Mobile Network Code]), the end point implicitly trusts the legitimacy of the base station. Note that both the MIB and SIB messages are

\(^1\)Except for the addition of mutual authentication.

\(^2\)Once the connection is dowgraded to GSM, an attacker could set up a rogue GSM base station and achieve a full "man-in-themiddle" control of the connection.
broadcasted in the clear by every base station and they can be eavesdropped using low-cost radios and basic open-source tools [7].

The overall 5G security architecture must take a leap forward and move away from implicit trust of downlink signaling messages. There must be a method such that a mobile device can determine the legitimacy of a base station prior to engaging in any communication with it. For example, cryptographic primitives could be leveraged to verify the legitimacy of broadcast messages, such as MIB and SIB packets, which could be signed with a secret key from the cellular operator. Moreover, the system should guarantee the freshness of such broadcast messages in order to prevent an adversary from intercepting legitimate broadcast messages and replaying them from a rogue access point. For example, critical downlink signaling messages, as well as MIB-SIB configuration messages, could be enhanced with a signature and a hash of multiple values, including a time stamp. However, note that, as further discussed in Section II-B, such solutions would only be possible by leveraging a public key architecture.

This is a challenging problem and would likely require expensive computation and processing. It should be designed to place most of its computational and cryptographic complexity on the infrastructure side. Meanwhile, the still computationally-demanding operations on the device side would occur only in the event of a connection to a new access point. As an alternative, the 5G protocols should guarantee that certain critical Radio Resource Control (RRC) functions [8], such as downgrading the connection to public key architecture.

B. Legacy symmetric key security architecture

Despite the constant evolution of mobile protocols, cellular networks still rely on an inflexible legacy symmetric key architecture [9]. Although modern LTE mobile networks implement mutual authentication, the mobile device is not truly authenticating the network (i.e., the cell network operator). Instead, it is verifying that the network has a copy of the user’s secret key.

Given this symmetric-key implementation, the cryptographic protocols of current mobile networks do not provide, as opposed to public key systems, a means to uniquely identify each party. There is a need to define and store a secret identifier for each subscriber. This secret identity, the IMSI (International Mobile Subscriber Identifier), is verified via the symmetric-key authentication handshake and a temporary identifier, the TMSI (Temporary Mobile Subscriber Identifier), is derived.

Although the IMSI should always be kept private and never transmitted over the air, it is intuitive that it will be required to transmit it over the air - unprotected and unencrypted - at least once. The very first time a mobile device is switched on and attempts to attach to the network, it only has one possible unique identifier to use in order to identify itself and authenticate with the network: the IMSI. The alternative would be for the authentication process to be initiated with no explicit identity, with the network checking, one-by-one, every subscribers (symmetric) secret keys in order to find the right one.

5G wireless systems should move away from legacy infrastructure exclusively based on symmetric-key cryptography and embrace the possibilities of a public key system. Although this would result in substantially higher computational complexity, one could argue that, on one hand, such cryptographic handshakes occur infrequently and, on the other hand, the great majority of wireless SoC (System on Chip) modules are commonly equipped with public key hardware accelerators. Embracing public key cryptography for future mobile systems has indeed been argued for many years already [10].

Although there are ongoing efforts to derive methods to never disclose the IMSI in the clear, an evolved 5G public key-based architecture would address the challenge of IMSI catchers in a simple way [11], [12]. For example, the mobile device could transmit its IMSI encrypted with one of the mobile operators public keys and sign the message with its own private key.

A public-key infrastructure for 5G radio access systems could also be leveraged to authenticate broadcast messages without an actual handshake. As discussed in Section II-A, broadcast messages could be signed with a private key from the network operator to verify their legitimacy prior to establishing any connection. Moreover, a hash of certain features could be included in the message - and signed as well - in order to guarantee freshness of the message and prevent replay attacks.

C. Unnecessary disclosure of location and radio configuration parameters

Mobile networks operate, protocol-wise, at the PHY layer and layer 2 in a similar fashion to that of legacy technologies. As such, the use of temporary and perpetual identifiers is not designed with a security and privacy preservation goal. A good example of this is the RNTI (Radio Network Temporary Identifier), a PHY layer ID used to identify and address traffic at the cell level. This is the ID that the base station assigns to each mobile terminal upon their very first Random Access Process to request radio resources [13]. Used simply to tell mobile devices apart within a cell, there is no need for this ID to be persistent or semi-persistent, but neither do the standards explicitly require the contrary.

The RNTI is included in the header of every single packet at the PHY layer, both in user data and control plane messages. It has been observed in the wild to be assigned both sequentially and kept constant per device for up to several hours at a time [5], [14]. This results in a trivial privacy leak, allowing a passive eavesdropper to, for example, determine how long a given user stays in a given cell. Moreover, since the RNTI preambles every single message, an adversary can trivially estimate the amount of both UL and DL traffic per mobile node3. In the context of an ad-hoc military tactical network, this could be trivially

---

3The size and format of the header of all types of LTE messages, independently of whether they are encrypted or not, is publicly defined in the 3GPP standard documents. The header of a message also allows us to tell the difference between user data and control plane payloads.
leveraged to determine the node managing the highest traffic load - potentially identifying the backhaul hotspot of the ad-hoc deployment.

The unnecessary disclosure of location information from the RNTI would be highly mitigated by something as simple as a pure randomization of the RNTI and the assignment of a new value during each Random Access Procedure. However, the RNTI is a good example of why the security architecture must be redesigned from top-to-bottom by means of a full redesign of mobile protocols. And the evolution towards 5G presents the perfect chance to do so.

Paging architecture has also been known for years to be vulnerable to privacy attacks and location leaks [15]. The fact that the network pages endpoints always, or for an extended period of time, with the same identifier (TMSI) results in a trivial location leak [16]. 5G mobile systems should introduce a redesigned paging protocol to mitigate such a threat. For example, standard cryptographic primitives could be used in order to generate a new random pTMSI (paging TMSI) every certain time. By means of some pre-shared key material, such as the current secret key, both the network and the mobile device could ensure they always generate the same new pTMSI, which would then be used to anonymously page devices.

### III. Mobile network architecture

One of the main goals for 5G wireless networks is massive connectivity, aiming to be the main enabler for the IoT. There is a common catchphrase in the industry about billions or even trillions of connected embedded devices by 2020, and the 5G guidelines and goals also envision wireless connectivity of a similar order of magnitude. This rapid growth of the IoT, which resulted in the number of connected devices topping the world’s population in 2013 [17], is one of the main foreseeable challenges for communication networks, both in terms of capacity and security.

The massive and unsupervised deployment of low computational power embedded devices over mobile wireless networks results in a highly challenging security landscape. Exploits that are difficult to monetize or that generate a large impact, such as blocking the connectivity of mobile endpoints [6], are substantially more concerning in the context of critical M2M applications. On top of that, sophisticated authentication and encryption techniques, such as the ones discussed in Section II-B, are very challenging to implement in low-power embedded devices. Despite this being one of the main security challenges of 5G (in the context of IoT), there is a lot of excellent work in both industry and academia aimed at tackling these problems. This document aims to highlight the main architectural challenges of 5G mobile networks in the context of the IoT surge – a challenge that will only be exacerbated if mobile connectivity indeed reaches billions or trillions.

The concept of mobile core overloading due to control plane signaling was first introduced in [18], which described a theoretical signaling overload threat against cellular networks. A low-volume attack, consisting of small data packets addressed to a large number of mobile devices, would theoretically induce a large number of RRC state transitions and, theoretically, overload the packet core of a mobile network.

<table>
<thead>
<tr>
<th>Cause</th>
<th>Event</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatty app</td>
<td>IM app checking for new messages too frequently caused outage</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>at U.S. carrier</td>
<td></td>
</tr>
<tr>
<td>Signaling spike</td>
<td>Outage for 3 million users at the world’s 6th largest operator</td>
<td>[20]</td>
</tr>
<tr>
<td>Smartphone native apps</td>
<td>Native apps from one of the main mobile OSES caused signaling overloads on Japanese mobile network</td>
<td>[21]</td>
</tr>
<tr>
<td>Chatty apps</td>
<td>Operators at Open Mobile Summit discuss actions to mitigate signaling spikes from chatty apps</td>
<td>[22]</td>
</tr>
<tr>
<td>Adds in popular app</td>
<td>Signaling spikes caused by ads displayed in a popular mobile game</td>
<td>[23]</td>
</tr>
<tr>
<td>LTE-connected tablet</td>
<td>Connectivity from popular tablet substantially increases control plane signaling</td>
<td>[24]</td>
</tr>
<tr>
<td>Mobile cloud service</td>
<td>Frequent reconnect attempts to a cloud service under outage resulted in signaling spike</td>
<td>[25]</td>
</tr>
</tbody>
</table>

**TABLE II. SAMPLE OF KNOWN SIGNALING OVERLOAD EVENTS [26]**

This doomsday scenario, initially conceived as an adversarial threat, started to become a reality in the wild with the advent of the smartphone and mobile app stores. This is when the term *signaling storm*, also referred to as *signaling tsunami* in eye-catching media headlines, was coined. This availability threat against the mobile core infrastructure arises, yet again, from the long overdue need for mobile networks to completely reinvent themselves and take a leap away from legacy technologies.

4Known exploits block the connectivity of mobile devices until rebooted or forced a connection reestablishment. Launching such an attack against a large deployment of M2M sensors could result in a lengthy DoS, as it is costly and time consuming to send technicians to manually reboot all the impacted M2M sensors. Security concerns and the stakes are especially high if such deployments belong to either physical security or healthcare systems.
Specifically, the circuit-switched core architecture of mobile systems, nicely suited to carry circuit-switched traffic - phone calls, that fits a Poisson arrival process with exponentially distributed call duration, is not the right architecture to transport the bursty IP-based traffic from billions of modern devices.

The more the Internet traffic from a consumer device differed from a classic Markov process, the more this issue became apparent. A simple Google search for the term \textit{signaling storm} returns dozens of major outages at large cell operators due to spikes in control plane signaling traffic originated by constantly bringing up and tearing down circuits for the bursty packetized IP traffic from mobile devices. Table III summarizes a few instances of signaling overloads over the past few years.

As discussed in [18], the impact of a malicious spike of control plane traffic could result in a direct threat against the availability of mobile systems. In the current mobile security landscape, it is common to see malicious applications (i.e., malware) infecting tens of thousands of mobile devices, or more, within a short span of time. Often, malware is written with a monetary or fraudulent goal. However, there are concerns across the industry and academia about the potential impact of a botnet of malware-controlled smartphones triggering, for example, a spike of attack procedures by repeatedly toggling airplane mode.

The traffic characteristics of most IoT applications are substantially different from the traffic of smartphones and tablets. It has been known for years that IoT traffic is a potential source for large mobile resource utilization inefficiencies [27]. Given that signaling overloads are already impacting the availability of mobile systems, the challenge is obvious especially when one mixes in an even more heterogeneous mix of billions of devices with bursty traffic flows or, even worst, periodic small flows of data [28].

The current evolution of LTE mobile systems has been addressing the signaling overload problem for years [29]. Moreover, the push towards virtualization and Software Defined Networks (SDNs) provides more enhanced scalability for mobile networks. Upon a spike of signaling traffic, the mobile core elements can flexibly have their computing resources expanded to handle the increased control plane load.

Despite these aforementioned technology enhancements to improve mobile connectivity for massive IoT deployments, the architecture of mobile communication systems must be fully redesigned. It has been argued for years that, in the long run, 5G wireless systems must move away from the rigid and non-scalable legacy circuit-switched architecture and aim for a fully packet-switched infrastructure, not too different from today's IP networks [30].

IV. ATTACKING MOBILE NETWORKS WITH LOW-COST AND OPEN-SOURCE TOOLS

The security redesign of 5G mobile networks should be strongly motivated by the current availability of offensive tools. Over the last few years, a number of open source projects have been developed which provide the right tools for sophisticated LTE security research. Running on off-the-shelf software radio platforms, these open source libraries provide the functionality of a software-based base station and, in some cases, the implementation of the endpoint software stack as well. With some rather simple modifications of the code, these tools can easily be turned into LTE protocol analyzers, stingrays and rogue base stations.

The two main LTE open source implementations being actively developed can be summarized as follows:

- **openLTE** [31]: Currently the most advanced open source implementation of the LTE stack, it provides a fully-functional LTE access network, including the features of the LTE packet core network. With proper configuration, it can operate NAS protocols and provide access to the Internet for mobile devices. It implements the HSS functionality on a text file storing IMSI-key pairs. It only requires a few lines of code to turn it into a stingray or a device that will block access to all smartphones and mobile devices in its vicinity [5].

- **srsLTE** [32]: This partial, yet fully functional, implementation of the LTE stack provides full access to PHY layer features and metrics and full access to decoded broadcast messages. It can be used to perform sophisticated scans of LTE mobile networks. The srsLTE project recently introduced srsUE, an implementation of the UE stack that allows one to emulate the communication between a mobile device and a base station. It is simply a matter of time before someone leverages srsUE to build a protocol fuzzer for security experimentation of threats against the mobile infrastructure. Based on the srsLTE engine, AirProbe is a fully functional LTE scanner that captures over the air DL LTE traffic, which can be analyzed offline using Wireshark and other standard software.

Most open source implementations can be run using standard off-the-shelf software radios, such as the USRP from Ettus Research [33]. This tool allows both passive and active experimentation, as it provides both transmit and receive features. A full, advanced set-up for LTE radio experimentation can be acquired for under $2,000. Passive traffic and eNodeB broadcast information capture can also be performed with much simpler platforms, such as a RTL-SDR radio for around $25 [34].

With a budget of under $2,000 and a powerful Linux computer, one can run a custom LTE IMSI catcher or rogue base station. On one hand, such wide availability of low-cost open-source tools for mobile network experimentation is positive, as it opens the doors for brilliant security researchers to improve the security of communication systems used by billions of people [35]. On the other hand, such tools also substantially lower the bar for attacks on mobile communication systems and should be taken into consideration when designing the security architecture of 5G systems.
V. CONCLUSIONS

Despite the significant technology improvements from legacy 2G networks to current LTE systems, the overall architecture and functionality of cellular networks still contains strong ties to outdated legacy technologies. Also, certain simple features are now long overdue for a systematic redesign that also considers the current cyber-security landscape and the low-cost availability of tools that can be leveraged to attack a mobile network (e.g., the unnecessary disclosure of location information from the PHY layer identifiers and privacy leaks linked to the paging protocol and the implicit trust on messages that come from a node that seems to be a legitimate base station).

In parallel, the legacy circuit-switched architecture of mobile networks still poses a great challenge for massive connectivity of embedded devices in the context of IoT. Although this challenge can currently be addressed through virtualization, this is not an appropriate long-term solution. In the era of packet-switched traffic and global IP networks, mobile systems should be redesigned accordingly to scale towards the massive connectivity goal of 5G systems.

As the next evolutionary step in wireless communications is taken, the industry has the perfect chance to embrace a holistic approach to security, as opposed to a set of functionalities and procedures attached to the overall architecture. This document summarizes some of the security challenges that must be addressed as mobile technology transitions towards 5G. Along with some of the key goals for future wireless systems, such as massive connectivity and sub-millisecond latency, the industry, academia and standards bodies should join forces to spearhead a true overall architecture redesign to address inherent vulnerabilities.

REFERENCES


