LAHORE UNIVERSITY OF MANAGEMENT SCIENCES

MASTERS THESIS

Cellular Networks under Signalling Attacks

Author:
Wasiq Noor Ahmad Qasmi

Supervisor:
Dr. Zafar Ayyub Qazi
Co-Supervisor:
Dr. Mobin Javed

A thesis submitted in fulfillment of the requirements for the degree of Masters in Computer Science in Syed Babar Ali School of Science and Engineering

October 22, 2019

Declaration of Authorship

I, Wasiq Noor Ahmad Qasmi, declare that this thesis titled, “Cellular Networks under Signalling Attacks” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed:

Date:

Abstract

Wasiq Noor Ahmad Qasmi

Cellular Networks under Signalling Attacks

Cellular Control Plane is a key component of the LTE which carries information necessary to establish and control the data traffic. These control plane operations are expected to grow with rapid growth of IoT devices. This not only imposes a significant workload on the Mobility Management Entity (MME) but also opens new security threats. Though there has been many reported Distributed Denial of Service (DDoS) and Signalling Storm attacks, little attention has been paid to examine its impact on the benign users traffic. In this work, we study and implement various LTE security attacks and simulate them over a high performance MME prototype with real control traffic traces. Our evaluation showed that the response and procedure completion times can be increased up to 40X under malicious traffic

Acknowledgements

First of all, I would like to thank my advisors **Dr. Zafar Ayyub Qazi** and **Dr. Mobin Javed** for their tremendous support and motivation throughout the project. I am grateful that under their supervision, I learned from their vivid spirit. Their constructive feedback and criticism has greatly helped me improve my research abilities. I would also like to thank my great friends Hamad, Arsalan and Ehtasham for their sincere friendship. And lastly, I am thankful to my family for their encouragement and support.

Contents

List of Figures

List of Tables

List of Abbreviations

UMTS	**U**niversal **M**obile **T**elecommunication **S**ystem
RF	**R**adio **F**requency
LTE	**L**ong **T**erm **E**volution
UE	**U**ser **E**quipment
IMEI	**I**nternational **M**obile station **E**quipment **I**dentity
USIM	**U**niversal **S**ubscriber **I**dentity **M**odule
P-GW	**P**acket **G**ateways
EPS–AKA	**Evolved P**acket **S**ystem **A**uthentication and **K**ey **A**greement
MME	**M**obility **M**anagement **E**ntity
RRC	**R**adio **R**esource **C**ontrol
GUTI	**G**lobal **U**nique **T**emporary **I**dentifier
3GPP	**3**rd **G**eneration **P**artnership **P**roject
SDN	**S**oftware **D**efined **N**etworking
RMDA	**R**emote **D**e-Registration **A**ttack
RAN	**R**adio **A**ccess **N**etwork
RDA	**R**esource **D**epletion **A**ttack
BDA	**B**lind **D**oS **A**ttack
PLMN	**P**ublic **L**and **M**obile **N**etwork

Dedicated to my beloved Parents…

Chapter 1

Introduction to Cellular Networks

In this chapter, we will start by describing the basic architecture of UMTS (3G) networks. Then, we will discuss the problems related to 3G that have driven towards the development of a new cellular communication platform, i.e. LTE/4G. Then, we will review the basic architecture and design principles of LTE. Lastly, we will discuss different planes and procedures involved in the LTE operations.

1.1 Architecture of UMTS/3G

1.1.1 High-level Architecture of 3G

Any mobile phone network is known as *Public Land Mobile Network* (PLMN) which is operated by *operators* like AT&T and Verizon. Before LTE was introduced, national and regional telecommunication standard bodies, known as the *Third Generation Partnership Project* (3GPP), developed a wireless telecommunication technology called *Universal Mobile Telecommunication System* (UMTS) as shown in Figure 1.1.

The core network contains two domains: **Circuit switching** and **Packet switching**. *Circuit switch* domain transfer the voice data across the geographical regions covered by the operators similar to that of a traditional fixed-line telephony. It established a dedicated virtual circuit so that users can make phone calls. While the *Packet switch* domain transport data stream into packets, similar to that of the Internet, between the users and external packet data networks.

Both domains manages the resources very differently. Circuit switching dedicates a two-way connection between two ends. Although the data transmission is effective but it is inefficient.

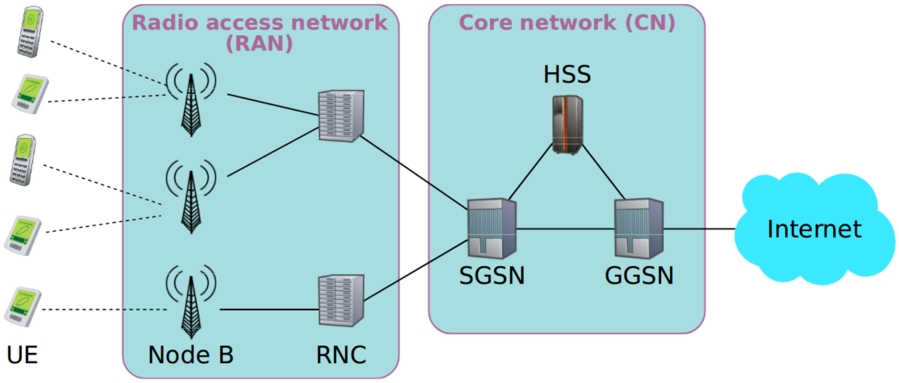

FIGURE 1.1: UMTS Architecture

Because the dedicated channels have enough capacity to carry much more data than being utilized. Packet switching solves this problem by dividing the data streams into packets with source and destination address labeled on them. Packet forwarding routers deployed in the wild read those addresses and forward them towards their destination. This domain is much more efficient for operators in terms of resource utilization.

Figure 1.1 shows the architecture of 3G. It mainly consists of two parts: **Radio Access Network** (RAN) and **Core Network** (CN).

> **RAN** is responsible for handling the radio communication with the end users. The user device is commonly knows as *User Equipment* (UE). It communicates with the RAN over the radio interface, and the direction of moving data from UE to RAN and RAN to UE is called *uplink* and *downlink* respectively. Another important component in RAN is the base-station also called as *Node B* which serves the purpose of providing radio channels to the UE so that the uplink and downlink traffic can be served. Node B are coupled with *Radio Network Controllers* (RNCs). They serves the purpose of maintaining the UE state for efficient radio channel utilization, and also providing information for triggering procedures like *Handover* (See Section 1.2).
>
> **CN** consists of *Gateway GPRS Support Nodes* (GGSN)s which serves as a gateway to packet data networks and outside world. *Serving GPRS Support Nodes* (SGSN) routes the traffic between Node B and GGSN. CN also contains *Home Subscriber Server* which is a central database that contains the information regards to all the subscribers.

1.1.2 The Growth of Mobile Data

The growth of mobile data was initially slow until 2010. But thereafter, data traffic started to dominate voice. To demonstrate this, Figure 1.2 shows the total traffic volume statistics, collected by Ericsson, throughout the world in petabytes [8]. From 2007 to 2013, data traffic volume increased by up to 500 times.

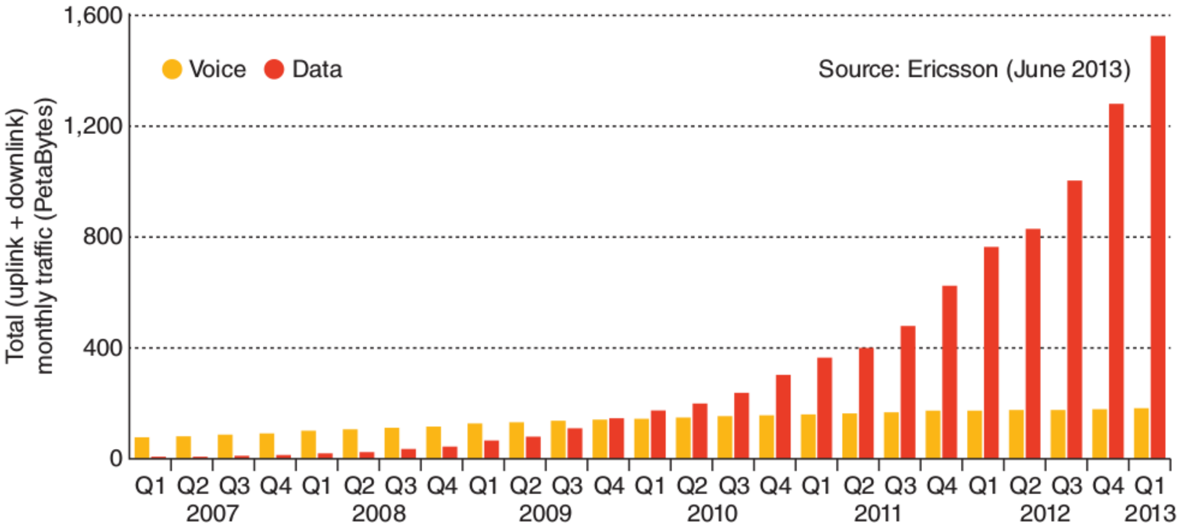

FIGURE 1.2: Ericson Measurements on Voice and Data traffic

In 2007, Google's **Android** operating system attracted third party developers to implement more user-friendly applications than ever before. This resulted in explosion is number of mobile phone applications, and consequently neither users nor developers were motivated to limit their data consumption. This started to congest the 3G radio resources, and therefore the need of a more flexible communication system was felt.

1.2 Architecture of LTE/4G

Unlike 3G, that supported Circuit as well as Packet switching, LTE was designed to support the Packet Switching only. So far, Long Term Evolution (LTE) is the most advanced telecommunication technology deployed by cellular networks. It not only provides low latency fast data transmission but also ensures robustness against failures. In this section, we will discuss the high

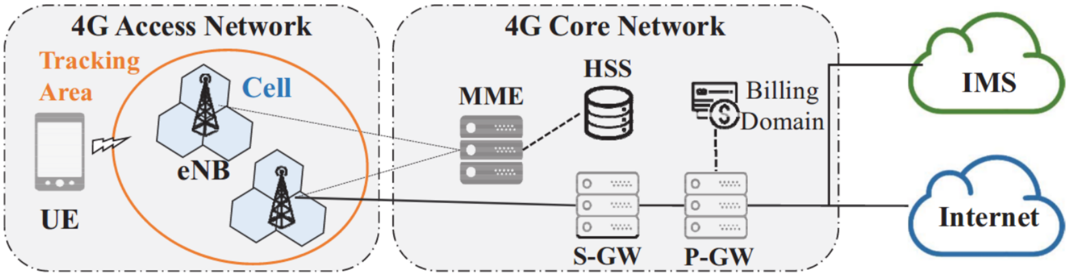

FIGURE 1.3: LTE Architecture

level architecture of LTE, the hardware involved, different planes of LTE, and its communication protocols.

1.2.1 High-level Architecture of LTE/4G

Figure 1.3 shows the architecture of an LTE network. It mainly consists of two components: **4G Access Network** and **4G Core Network**.

- **4G Access Network** is the front-end of any LTE infrastructure and includes the resources like UE and eNB (Base-station).
  - **UE** enables a user with legitimate services such as voice and data. UE can be uniquely identified by an International Mobile station Equipment Identity (IMEI) number. It uses a Universal Subscriber Identity Module (USIM), also know as SIM card, provided by the telecom provider to connect to the nearby eNB to access the subscribed services.
  - **eNB** enable UEs to establish a wireless connection with LTE core network. Each eNB is responsible for processing the reception and transmission of RF signals in a sector, also called cell. There can be multiple cells under one eNB as shown in the Figure 1.3. eNB is connected to Mobility Management Entity (MME) for control plane, and Packet Gateways (P-GW) for data plane communication.
- **4G Core Network** is the backend of any LTE deployment consisting of MMEs and Gateways.
  - **MME** is the key component that handles the signalling traffic that a user/UE generates. It is responsible for the authentication and mobility of the subscribers. UE authentication is followed through the authentication protocol knows as Evolved Packet

System–Authentication and Key Agreement (EPS–AKA). MME is also responsible for setting up the EPS context bearers, a session tunnel created between UE and Gateways for data traffic.

- **Gateways** ensures mobility and provide Internet service to the UE. They allocate IP addresses and also maintain the accounting of data traffic for the UE.

1.2.2 LTE Planes

There are mainly two major communication planes in LTE: **Control Plane** and **Data Plane**.

Control Plane carries the information necessary to establish and control the data traffic. It primarily includes the context establishment at multiple resources of the network.

Data Plane is the user generated traffic e.g. voice or data, which is transmitted over RF tunnels setup by the control plane.

1.2.3 LTE Procedures

A procedure is a set of request response messages that are exchanged between two or more entities in an LTE network. Some commonly used LTE procedures are as follow:

Attach is the procedure initiated by the UE to request the services when it is turned on. Firstly, UE listens for the nearby base-stations and requests a *Radio Resource Control (RRC)* connection. This connection is established in plain text without any security or integrity protection. Once the connect is established, **Attach** procedure is triggered by the eNB and forwarded to the MME. Thereafter, MME generates the encryptions keys and integrity protection algorithms to use. Lastly, MME allocates a Global Unique Temporary Identifier (GUTI) to uniquely identify UE for future communication.

S1 –Handover is the procedure initiated by the eNB when a user context needs to be switched between the MMEs. It is primarily caused when a user moves from one cell to another which is controlled by some other remote MME at the network core.

PDN Connect is the procedure that establishes the UE context at the packet gateway once it has been attached to the network. This context is then used to tunnel the UE uplink and downlink traffic to the Internet.

Detach procedure is triggered once a UE is turned off. It causes the UE context to be released at the eNB and packet gateways.

1.3 Chapter Summary

In this chapter, we learned the basic design architecture of 3G including RAN and CN. We then studied how, over the years, the gradual adoption of **Android** has led to the increase in data traffic. Then, we examined the architecture of 4G networks followed by different planes and procedure involved in an operational LTE deployment.

Chapter 2

Background

Over the years, LTE has become the standard mode of communication for many applications including IoT devices like connected cars, augmented reality and remote surgery. In this chapter, we will discuss the increasing dependency on LTE architecture, and study its impact on data plane due to the signalling traffic generated by such devices. Then, we will discuss different threats imposed due to growing control traffic in LTE. And lastly, we will highlight some of the recent DDoS and Signalling attacks to understand the importance and impact of these imposed threats.

2.1 Growing Control Plane Traffic in LTE

LTE have recently gained significant attention due to its compatibility with modern IoT devices. According to Ericson, Cellular IoT is becoming the standard of choice for wide area networks, and therefore IoT devices are expected to reach 4 billion mark by 2024 [4]. With this increase in IoT devices, we expect to see an increase in amount of signalling and data traffic as well. In fact, Cisco whitepaper showed that the amount of signalling traffic has 50% more growth than the data traffic [6]. Therefore, it is important to understand how this growing signalling traffic can have an impact the data traffic. To answer this, a recent study tried to introduce parallelism in traditional sequential control plane procedures [13]. They also conducted various testbed experiments to show how access to data is dependent on the timely completion of the control plane procedures. Figure 2.1 shows two different app-level latencies experienced with and without DPCM. Since any failure in LTE operation due to link failure or rejected requests re-initiates the procedure, added latency is experienced at the application. The results in Figure 2.1 shows that web page loading and YouTube pause experienced at the application level is lower when control plane operations are completed faster.

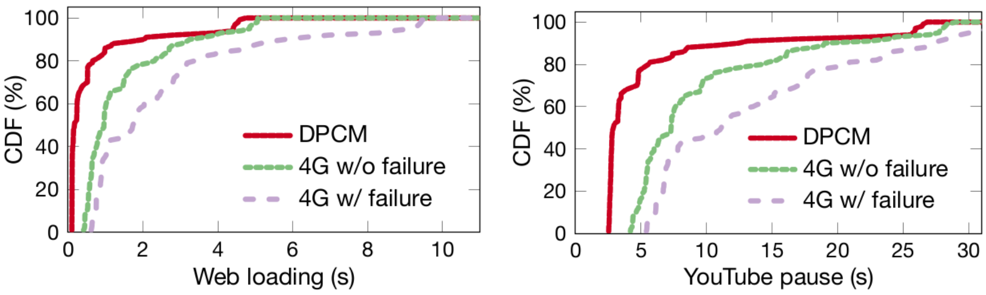

FIGURE 2.1: Web/YouTube latency improvement in DPCM

2.2 Threats to Cellular Control Plane

One of the most important factors in the wide adoption of the LTE is its always-on connectivity for IoT devices like autonomous vehicles and augmented reality. Since such devices need high availability, there is a serious threat of malfunctionality if disconnected from LTE service.

From our discussion in Section 2.1, we know that the increasing number of IoT devices lead to an increase in amount of control traffic. Therefore, there is a potential threat to the availability of the network if traffic volume surpasses a certain threshold. Although the 3rd Generation Partnership Project (3GPP) maintains the standards including security aspects for the LTE communication, there have been many reported DDoS and Signalling Storm attacks that had led to the failure of the cellular networks. Following is a list of few incidents and published studies that narrates the failure of control plane is recent past:

2.2.1 Reported Signalling Attacks

1. In 2012, T-Mobile's network was brought to knees by an Android instant messaging app. Director of T-Mobile explained that the network was temporarily brought down when a local developer released an instant messaging android app into the market. This app would make frequent refresh connections which overloaded the T-Mobile's network and the services were unavailable in an area for a while. Studies showed that the load increased to 1200% from this application alone. This was fixed by contacting the developer and directly working out the fix. Reference can be found at [7] reported by New York Times.

2. In 2011, Nokia Siemens Smart Labs conducted tests and reported that Angry Birds on an Android-based Samsung Galaxy smartphone, with mobile advertising, generated 2,422

signals in one hour of play, a 352 percent increase on the base level as determined by Smart Labs of 688 signals per hour. In contrast, Angry Birds on an iPhone with no advertising generated only eight percent more signalling traffic than the base level. This overloading was caused by the frequent acquiring/releasing of new connections. Light Reading published this report at [21].

3. In 2012, Japan's famous telecom provider, *DoCoMo*, suffered a major network outage due to the signalling storm. This was attributed to an Android application and, more generally, the way Android constantly polls the network as referred by [24]. *Verizon* also blamed similar factors for several network outages in its LTE network in 2011 [24]. DoCoMo, according to the Nikkei news agency, has demanded that Google revisit the signalling and data loads imposed by Android, particularly the habit of handsets transmitting control signals to the network, and pinging the servers, automatically rather than as-needed. They also reported that this outage has affected 2.5m subscribers in Tokyo, and was the fifth such incident in last six month.

4. In 2011, Open Mobile Summit was conducted in London where representative from the Europe's largest cellular network operators called for actions to prevent mobile applications from overloading their network with signalling traffic. Clearly, operators wanted to crack down the generators of signalling traffic - this is, applications that are designed to ping the network very frequently for updates. Vice president of *Orange* stated that there is a need for issuing guidelines; and a collective action is required from operators as well as developers. Full Light Reading report is available at [14].

5. Online services rely entirely on the service availability in large data centers provided by companies like Google, Amazon and Microsoft. Nokia Siemens Smart Labs verified that a cloud service outage can trigger smart phones to generate signaling loads up to 20 times greater than normal. A smart phone maintains always-on connection by sending keep alive messages. But, when the service is unavailable, previously connected smart phone begins frequent reconnect attempts every few seconds. This generates a signalling traffic peak up to 20 times higher than the normal traffic. Source of the report is available at [20].

2.2.2 Literature Review on Signalling Attacks

1. In May 2007, Patrick et al [11] introduce a novel DoS attack termed the *signaling attack*, which seeks to overload the control plane of a 3G wireless network using low-rate, low-volume attack traffic. This attack can be triggered by repeatedly accruing and releasing

radio channel. To accomplish this, an attacker first sends a low-volume packet burst to a mobile. If the mobile does not currently have a radio channel, the network will allocate a new one to complete the data transfer. After an inactivity timeout, the radio channel is torn down to recycle it back for others' use and help preserve the mobile energy that will otherwise be wasted on maintaining the channel. Immediately after the channel release, the attacker sends another low-volume packet burst to the mobile so as to trigger another radio channel establishment

2. It is well known that a simple event on the phone side triggers substantial number of request/response messages on several EPC components. Bassil et al [3] conducted a study and showed that this message exchange protocol can be exploited using botnets to amplify the traffic load and launch a DoS attack. Relevant to that, the authors of [16] demonstrated the availability of various platforms to exchange command and control messages that can be potentially used as the botnets.

3. Prasad et al [19] presented that the EPC could potentially be saturated legitimately due to the overwhelming traffic and frequent re-connections of millions of M2M devices. Also, [3] showed how infected mobile devices could create an amplified signalling attack by establishing and releasing IP connections to an external server.

4. There are several EPC components that solely perform their purpose such as HSS and HLR. In 2009, Traynor et al [23] published a paper demonstrating the possibility of overloading the 3G HLR. It is important to note that the HSS is involved in many signalling events in the EPC and could also suffer from the signalling amplification attacks.

5. Roger et al from AT&T Research Center published a paper and narrated different type of attacks on the availability of LTE Mobility Networks at [18]. He also provided some research direction for the research community that can potentially help prevent such attacks on the availability of the network. He proposes three major security directions in the context of current threat model. As a first step, the capabilities of the current network should be re-configured and adopted to prevent against security attacks. In the mid-term, architectural changes such SDN should be adopted with full or partial deployment of the EPC in the cloud. Finally, the complete design of the EPC should be considered to dealt with the scenario where almost every electronic device will be connected to the network.

2.3 Chapter Summary

In this chapter, we learned that the amount of control traffic is increasing much faster than the data traffic. Therefore, it not only effects the data plane operations but also poses different threats to the LTE control plane e.g. Signalling Attacks. We also understand the impact of signalling storms by reporting different signalling attacks from the recent past, along with the literature review from the research community.

Chapter 3

Signalling Traffic and Mobility Management Entity (MME)

From our discussion in Chapter 1, we know that MME is the main component responsible for handling the control traffic in LTE. We also understand, from Chapter 2, that the increasing amount of signalling traffic impacts the data plane operations significantly. Therefore, in order to understand the impact of signalling attacks, it is important to first measure the signalling load at an MME due to the benign traffic.

In this chapter, we leverage the work of two different published studies to estimate the signalling workload at a commercial deployed MME. We will then quantify the signalling load in terms of procedures and number of messages experienced at an MME of a Tier 1 telecom provider (AT&T) in USA.

3.1 Signalling Load at an MME

3.1.1 Considerations for re-designing the cellular infrastructure exploiting software-based networks

This paper [15] was aimed at proposing new cellular infrastructure for low latency applications using Software Defined Networking (SDN). In order to stress test their proposed solution, authors gathered the infrastructural insights from a Tier 1 telecom provider (AT&T). Table 3.1 shows the details of cellular components in USA. Stress Vector shows the quantitative dimension vector of wireless infrastructures, while the system impact are the relative derived workloads on LTE components.

Stress Vector		**System Impact**
Number of UEs	400M	1 eNB : 2000 UEs
Number of eNB	200K	1 S−GW : 4000 eNB
Number of EPC	50	1 S−GW : 8*M* UEs
Number of P−GW	50	1 MME : 1340 eNB
Number of S−GW	50	1 MME : 2.7*M* UE
Number of MME	150	

TABLE 3.1: Cellular Infrastructure in USA

Paper also mentions that the MME experience 3X more interactions as experience by a S−GW. Also, that there is at least 1 interaction every 1 minute per UE at the S−GW. The most important matrix to observe is UEs to MME ratio i.e. **2.7M**.

3.1.2 An IoT control plane model and its impact analysis on a virtualized MME for connected cars

Authors of this paper wanted to propose a virtualized MME solution for highly mobile IoT devices like connected cars [2]. Since connected cars are associated with performing frequent handovers due to their mobility, authors gathered the traffic traces during busy hours for 40 days from a Tier 1 telecom provider (At&T). Table 3.2 shows the workload measured from the collected dataset during the busy hours for 5 most common LTE procedures.

Event Type	**Average UE procedure count per Busyhour**	**Average car procedure count per Busyhour**
Attach	0.096	0.082
Detach	0.899	0.152
PDN Connect	0.043	0.368
PDN Disconnect	0.087	0.45
S1 HO	0.158	0.262

TABLE 3.2: Average Procedures per MME in busy hour

3.2 MME during Peak Hours

Details and insights shown in the Section 3.1.1 and 3.1.2 can be combined to derive the number of procedures that happen at an MME in a busy hour. Table 3.3 summarizes all the events that

MME handles during a busy hour. Note that the number of UE and connected cars per MME is **2.7M** and **48** from first paper.

Event Type	**Average UE Procedure count per Busyhour**	**Average Car procedure count per Busyhour**	**Total procedures per MME per Busyhour**
Attach	0.096	0.082	$0.096 * 2.7M + 0.082 * 48 = 259K$
Detach	0.899	0.152	$0.899 * 2.7M + 0.152 * 48 = 2427K$
PDN Connect	0.043	0.368	$0.043 * 2.7M + 0.368 * 48 = 116K$
PDN Disconnect	0.087	0.45	$0.087 * 2.7M + 0.45 * 48 = 234K$
S1 HO	0.158	0.262	$0.158 * 2.7M + 0.262 * 48 = 426K$

TABLE 3.3: Signalling workload on MME in busy hour

From Table 3.3, we are able to estimate the number of procedures at an MME to be **3.4M** procedures or **13M** messages per busy hour.

3.3 Chapter Summary

In this chapter, we leveraged the statistical measures of two studies to derive the signalling workload at an MME during any peak hours. We find that during peak hours, the MME approximately handles **13M** messages. This finding is important in not only measuring the signalling delays for a benign user but will also help in quantifying the increase in delay factor for a user during signalling attacks.

Chapter 4

Signalling Attacks

Now that we have an approximation of signalling workload at an MME, we can discuss different signalling attacks and their threat models that can effect the MME response and procedure completion times. In this chapter, we will start by discussing the threat model of signalling attacks. Then, we shall study the root cause that allows the attacker to generate signalling traffic, and thereafter, we will extend its implication by studying different signalling attacks in 3G and 4G networks.

4.1 Threat Model of Signalling Attack

Signalling attacks are referred to as class of attacks where one or more attackers generates a flood of control traffic from many sources to compromise the availability of the cellular network [17]. Since cellular networks consist of many components, there can be different threat models having different impacts and victim types. In this study, we shall be focusing on signalling attacks that are targeted towards MME.

In order to understand the threat model that targets the MME, Figure 4.1 demonstrate the threat model for signalling attack where a *storm* of signalling traffic is targeted towards the MME from malicious devices that are under attacker's control. This model aims to **IMPACT** the MME performance where as the **VICTIM** can be any benign UE which gets affected by the degraded performance of a serving MME. In this threat model, an attacker is able to generate malicious signalling traffic mainly due to the vulnerable RRC protocol. Therefore, we discuss the RRC protocol and its implications in the following sub-section.

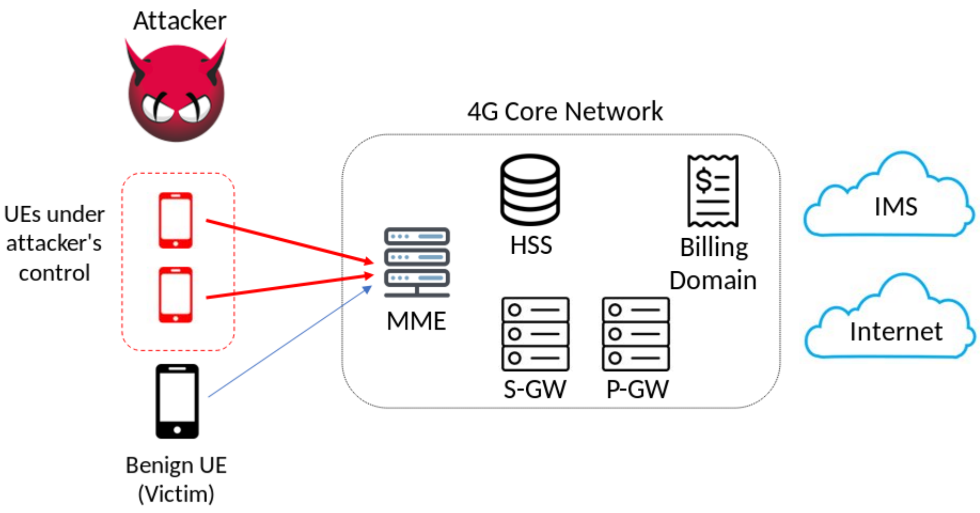

FIGURE 4.1: Signalling Attack Threat Model

4.1.1 RRC Protocol

In 3G networks, RRC protocol is used to manage the Radio Access Network(RAN) resources efficiently. This protocol maintains a state machine between the UE and RNC server (See Figure 1.1) so that the device's state can be monitored and bandwidth allocations is utilized efficiently. Possible state machine values are as follow:

- **IDLE**: When the UE is turned on, and it has no connection with the base station.
- **FACH**: When the UE tries to transmit or receive limited bandwidth resources, it is allocated a shared transmission channel.
- **DCH**: When the UE transmission exceeds a certain threshold (1500 Bytes), a dedicated channel is allocated.
- **PCH**: UE maintains a connection with base station but listens for *paging* requests for any incoming requests from the Internet.

4.1.2 Vulnerability in RRC Protocol

As per the 3GPP standards, RRC protocol is performed in plain text without any security or integrity protection [1]. Once the context is established, from thereafter, all the communication is protected via AKA security protection algorithms. Because the initial RRC connection is not protected, any adversary can spoof the content of the message while establishing the connection. Exploiting this vulnerability, research community has demonstrated various signalling attacks in 3G and 4G which we shall discuss in the following sections.

4.2 Signalling Attacks in 3G

4.2.1 3G RRC Transition Attack

A recent study by Gorbil [9] conducted simulated experiments to explain how RRC protocol can be exploited to generate signalling storms in network core. Authors of the paper mention that the networks operator deliberately disable PCH state, and allow the UE to go in IDLE mode to benefit the UE battery life. Now this misconfiguration can lead to serious signalling attacks in the network as shown by Table 4.1 . It is important to note that the messages are only exchanged with Network Core if UE either makes a transition to or from **IDLE** state.

Transition	Triggering Event	Messages exchanged in RNS	Messages exchanged in Network Core
IDLE -> FACH	Uplink or Downlink	15	5
PCH -> FACH	Uplink or Downlink	3	0
FACH -> DCH	Radio link threshold reached (1500B)	7	0
DCH -> FACH	Expiry of inactivity timer (6s)	5	0
FACH -> IDLE	Expiry of inactivity timer (12s)	5	3
FACH -> PCH	Expiry of inactivity timer (4s)	3	0
PCH -> IDLE	Expiry of inactivity timer (20min)	6	3

TABLE 4.1: RRC Transitions

Authors conducted the experiment to observes the queuing delay experienced at the application level where 150 malicious users would change their states frequently. Figure 4.2 shows the queuing time before and after the attack. Expectedly, the queuing time sharply increases by up to 7X.

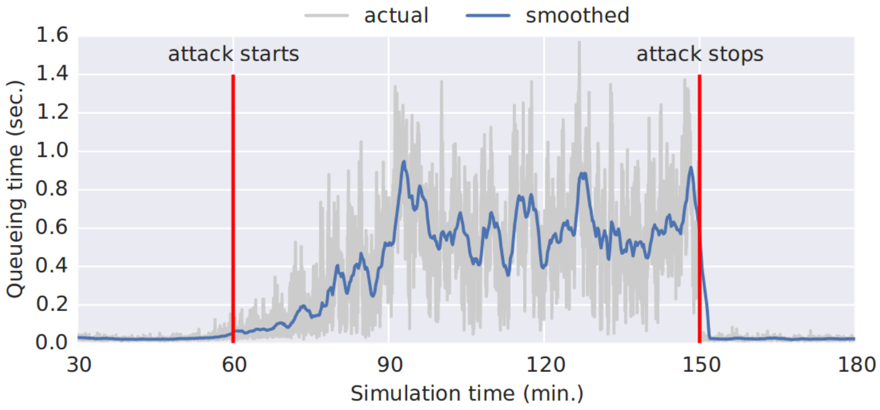

FIGURE 4.2: Queuing Time during 3G RRC attack

4.3 Signalling Attacks in 4G/LTE

Section 4.2 showed us that even a small number of state transitions can impact the queuing delay significantly. But this attack could not be performed with LTE, since RRC is fully managed at the eNB, and no signalling messages are exchanged with network core. However, an important observation from the experiment in Section 4.2 is that an adversary is able to get hold of a small number of devices, it is possible to impact the response time of an MME.

Fortunately, a study named **Touching the Untoucahables** has enlisted 34 vulnerabilities in LTE design and implementation [10]. From the study, we have extracted out all the LTE attacks that generated some amount of traffic at the MME. In the following sub-sections, for each attack, we will discuss the attack model and the number of devices required to meet the threshold discussed in Chapter 3.

4.3.1 Resource Depletion Attack (RDA)

Every commercial eNB has a certain capacity to support RRC connections based on the hardware and software specifications. The purpose of the RDA attacks is to deplete the resources of an eNB such that other user cannot connect to the eNB. In RDA, an attacker repeatedly performs RRC connections with eNB. In normal situations, an eNB will *piggyback* a **NAS Attach Request** over the completion of the RRC connection. In RDA, attacker spoofs the IMSI number in the initial attach request, and once the MME responds back to the attach request, attacker never responds back to the MME and instead restart the RRC connection establishment procedure. The reason for not replying to the MME is to preserve the RRC state at eNB while it waits for the valid **NAS authentication response**. Authors also mention that they were able to make 20 RRC connections per second with a commercial eNB.

Note that while this attack is performed, at least one message is exchanged with the MME i.e. **NAS Attach Request**. Leveraging this fact, Table 4.2 shows the number of devices required to perform the attack such that it overwhelms (meet 13M message mark) the MME.

Number of connections per hour per device	$3600 * 2$ = 72K Connection
Number of messages delivered to MME per hour	$1 * 72$K = 72K Messages
Number of messages in peak hour at MME	13M
Number of devices to meet threshold	13M / 72K = 180 Devices

TABLE 4.2: RDA Attack Vector

4.3.2 Blind DoS Attack (BDA)

Unlike RDA attack that denies the normal user from connecting to eNB, BDA denies the UE by establishing spoofed RRC connection as victim UE. Attacker needs to know the TMSI of the victim to perform this attack which can be obtained by one of the following methods:

1. Attacker can get the victim's phone number and perform *silent Paging attack* [22] to retrieve TMSI number.
2. Attacker can operate a rogue eNB and sniff **NAS Tracking Area Update** procedure to retrieve TMSI number.

Once the attackers has the TMSI number, it established a spoofed RRC connection with **ueIdentity** field set to the victim's TMSI number. It causes the eNB to release the context of the victim at the eNB. The impact of this attack can be separated into two type according to the victim's RRC state at eNB.

- **Victim is in IDLE state**: If the adversary established a spoofed RRC connection as the victim's UE, RRC state will be changed to **CONNECTED** at the eNB. Thus, the MME will not trigger the *paging* for any incoming traffic for the victim's phone until it established new RRC connection.
- **Victim is in CONNECTED state**: In this case, spoofed connection from the attacker causes the context of the victim's phone to be released at the eNB. While the victim continues to

transmit the traffic, but it fails since the context has already been released. Multiple attempts to transmit data initiate the **NAS Service Request** procedure before the connectivity is established again.

Case 1 of this attack is similar to that of RDA, so we focused on performing the Case 2. Table 4.3 shows the number of malicious devices required to meet the **13M** threshold.

Number of connections per hour per device	$3600 * 2$ = 72K Connection
Number of messages delivered to MME per hour	(Service Request) $4 * 72$K = 288K Messages
Number of devices to meet threshold	13M / 288K = 45 Devices

TABLE 4.3: BDA Attack Vector

4.3.3 Remote De-Registration Attack (RMDA)

To perform this attack, adversary sends a malicious **NAS** message to the MME in which victim's context is registered. To perform this attack, attacker opens a spoofed RRC connection and sends a crafted NAS or invalid security protected or replayed message e.g. **PDN Disconnect** to the MME. In response, MME inappropriately handles the message and consequently releases the victim's context from gateways. This change is also reflected back to the appropriate eNBs. Thus, the connectivity of the victim's UE is lost without any notification to the victim. Table 4.4 shows the stress vector of the attack.

Number of connections per hour per device	$3600 * 2$ = 72K Connection
Number of messages delivered to MME per hour	(PDN Disconnect) $2 * 72$K = 144K Messages
Number of devices to meet threshold	13M / 144K = 90 Devices

TABLE 4.4: RMDA Attack Vector

4.3.4 SMS Phishing Attack

In this attack, an adversary sends an SMS message to the victim's phone UE_1 spoofed with the messages sender as victim UE_2. In order to perform this attack, attacker establishes spoofed RRC connection. Then, the SMS content is generated and forwarded to MME using **Uplink NAS Transport**. Upon receiving the messages, MME forwards the message to the victim's UE. This attack can be performed repeatedly to generate large volume of signalling traffic at the MME. This attack is similar to that of **RDA Attack**, so we do not perform explicitly for this work.

4.4 Chapter Summary

In this chapter, we learned the threat model of signalling attacks that aim to degrade the MME performance by generating malicious traffic. We then studied the vulnerability in RRC protocol that allows an adversary to generate signalling traffic from malicious devices under its control. We also discussed some of the reported signalling attacks in 3G and 4G networks, and using our findings from Chapter 3, we also derived the number of malicious devices that are required to generate as much workload as during a peak hour for each LTE signalling attack.

Chapter 5

Evaluation and Results

In Chapter 4, we studied various signalling attacks in LTE that exploit the vulnerable RRC protocol to degrade the MME performance. In this chapter, we will start by explaining the experiment setup that we build to conduct various LTE attacks as discussed in Chapter 4. Then, we will demonstrate a baseline experiment to observes the MME response and procedure completion time under normal scenarios. Following that, we will evaluate the MME response and procedure completion time under different signalling attacks.

5.1 Experiment Setup

Our test setup consisted of two servers running Ubuntu 16.04 with Linux kernel 4.4.0. Each server has 14 Intel Xeon E2-2699 v4 CPU cores clocked at 2.2GHz (hyperthreading disabled) and single 64GB NUMA node. Both the server were connected via 1 Gbps link. One of the servers would generate control traffic, called **PacketGen** while the other would be serving as the MME, called **TinyMME**. Figure 5.1 show the testbed setup used for the experiments.

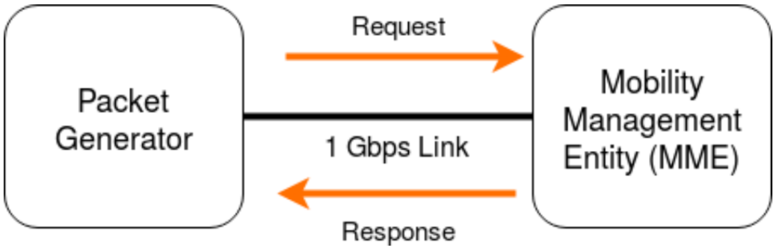

FIGURE 5.1: Experiment Setup

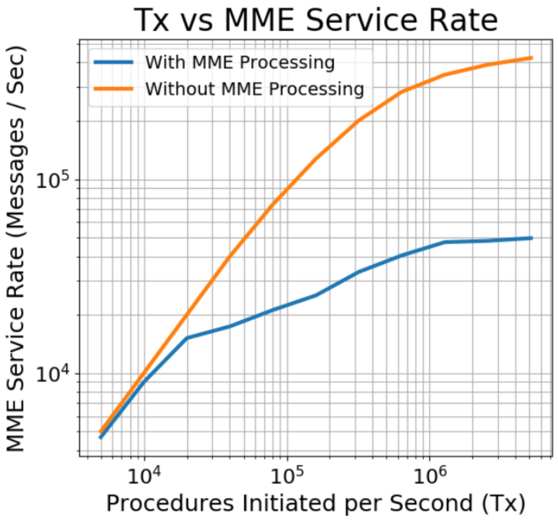

FIGURE 5.2: MME Service Rate with respect to Procedures per Second

For testing control traffic, we also implemented S1AP and NAS protocol, and the handling of request response messages between **PacketGen** and **TinyMME**. We also implemented the **ASN1** encoding and decoding scheme which is the default scheme used in LTE messages. Experiments were run with real signalling traces from a commercial traffic generator provided by ng4T [5].

5.2 MME Service Rate

In order to stress test the MME capacity, we wanted to see if our experiment setup is not the bottleneck to high rate traffic. To verify the claim, we gradually increased the number of procedures initiated per second, and measured the MME service rate with and without the MME processing enabled. Figure 5.2 shows the results of the experiment. From here, it can be concluded that the MME processing is becoming the bottleneck although we were able to transmit more traffic.

5.3 Uniform Benign Traffic

Before we examine the MME behavior in different attack scenarios, it is first important to see how MME responds to uniform benign traffic. It is also important as the reference point to compare it against different attack scenarios. Figure 5.3 and 5.4 shows the graphs for MME response time and procedure completion time under Benign traffic.

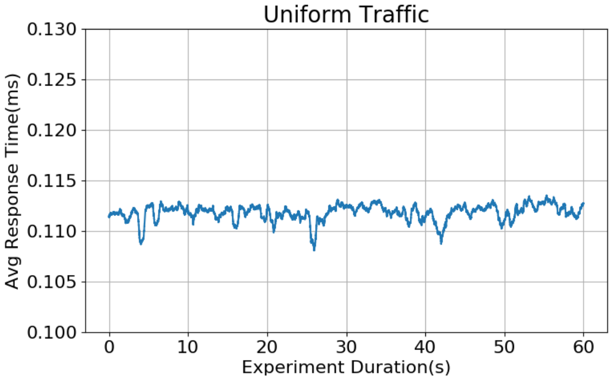

FIGURE 5.3: MME Response Time for Uniform Benign Traffic

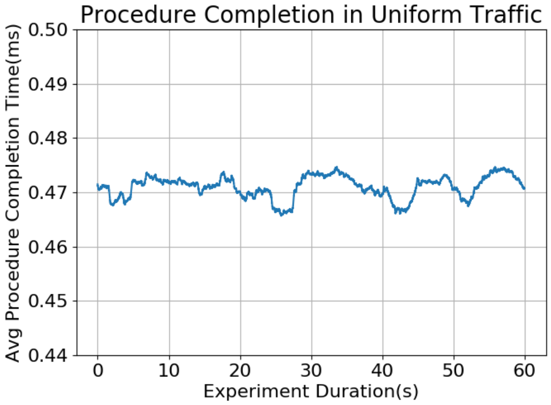

FIGURE 5.4: Procedure Completion Time for Uniform Benign Traffic

5.4 Resource Depletion Attack (RDA)

As discussed in Chapter 4, RDA attack was implemented and performed for varying number of devices. At the start of the experiment, Benign traffic was generated for 60 seconds. While the experiment reached the 10th second, RDA attack was launched by the PacketGen which lasted for 30 seconds. Experiment was conducted with 200, 300 and 400 malicious devices. Lastly, MME response time and procedure completion time were recorded. For each experiment, we recorded the MME response and the procedure completion time. Figure 5.5 and 5.6 shows the results for each experiment. Note that the MME response and procedure completion time increased as the attach started.

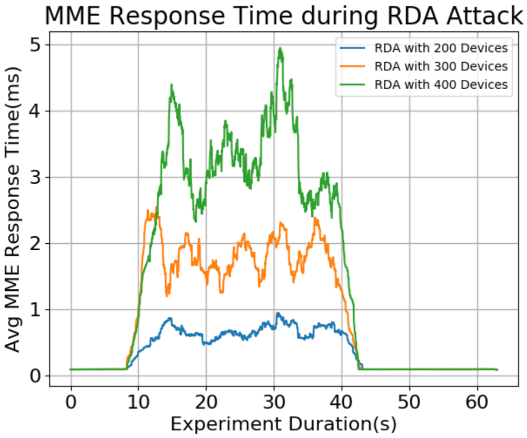

FIGURE 5.5: MME Response Time during RDA Attack

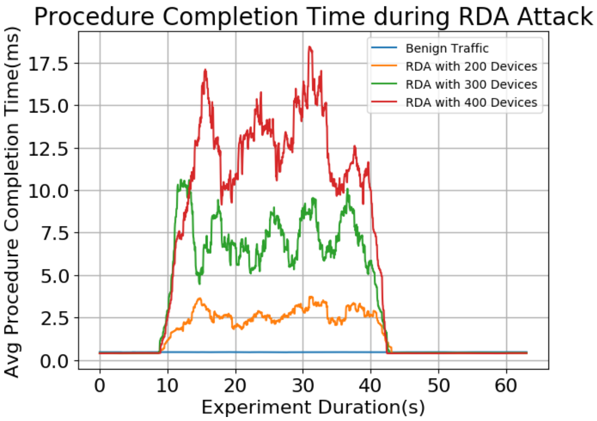

FIGURE 5.6: Procedure Completion Time during RDA Attack

Take away:
Procedure completion time can be increased up 20X if attacker can get hold of 400 malicious devices only.

5.5 Blind DoS Attack (BDA)

As discussed in Chapter 4, BDA attack was implemented and performed for varying number of devices. At the start of the experiment, Benign traffic was generated for 60 seconds. While the experiment reached the 10th second, BDA attack was launched by the PacketGen which lasted for 30 seconds. Experiment was conducted with 50, 100 and 150 malicious devices. Lastly, MME response time and procedure completion time were recorded. For each experiment, we recorded the MME response and the procedure completion time. Figure 5.7 and 5.8 shows the results for each experiment.

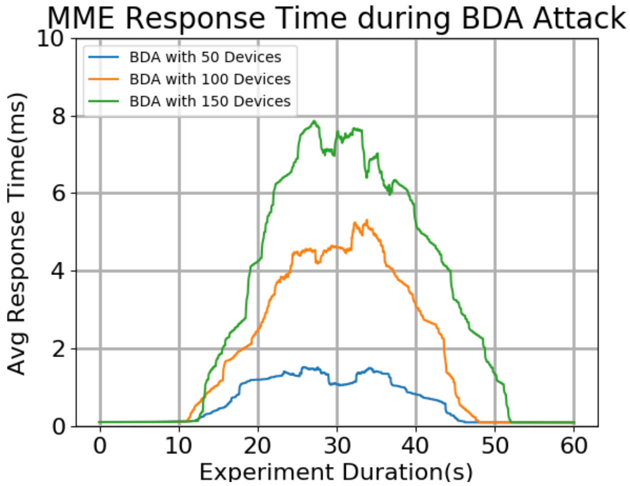

FIGURE 5.7: MME Response Time during BDA Attack

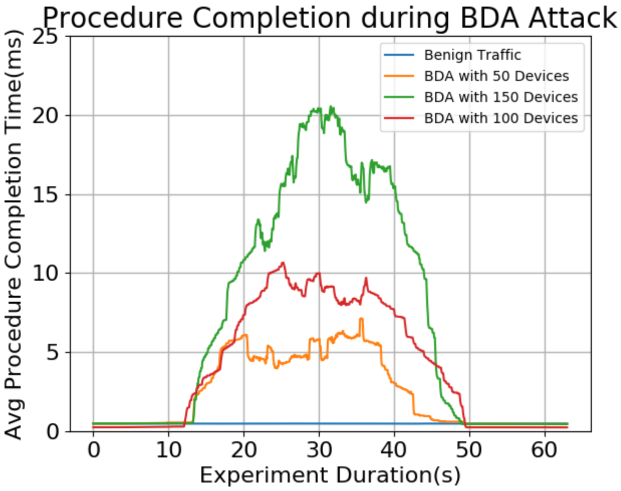

FIGURE 5.8: Procedure Completion Time during BDA Attack

Take away:
Procedure completion time can be increased up 20X if attacker can get hold of 150 malicious devices only.

5.6 Remote De-Registration Attack (RMDA)

As discussed in Chapter 4, RMDA attack was implemented and performed for varying number of devices. At the start of the experiment, Benign traffic was generated for 60 seconds. While the experiment reached the 10th second, RMDA attack was launched by the PacketGen which lasted for 30 seconds. Experiment was conducted with 50, 100 and 150 malicious devices. Lastly, MME response time and procedure completion time were recorded. For each experiment, we recorded the MME response and the procedure completion time. Figure 5.9 and 5.10 shows the results for each experiment.

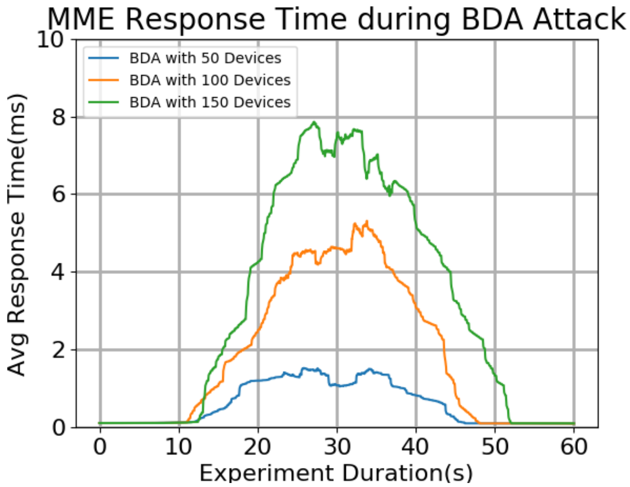

FIGURE 5.9: MME Response Time during RMDA Attack

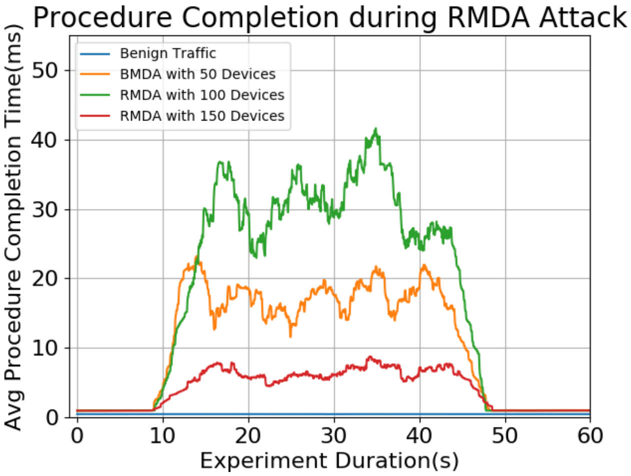

FIGURE 5.10: Procedure Completion Time during RMDA Attack

Take away:

Procedure completion time can be increased up 40X if attacker can get hold of 150 malicious devices only. Also note that the procedure completion time in RMDA is double than the rest of the attacks. This is because in RMDA attack, victim tries to re-connect to the MME immediately after the disconnection. Therefore, the aggregate traffic volume builds up the queues at MME. Thus, we expect to see more delays in procedure completion time when compared to other attacks.

5.7 Chapter Summary

In this chapter, we describe the experimental setup that we used in order to conduct various LTE signalling attacks. Then, we demonstrated the MME behavior under normal/benign traffic.

Having this baseline, we then conducted and evaluated the response and procedure completion time under RDA, BDA and RMDA signalling attacks over a high performance MME prototype with real control traffic traces that were encoded and decoded using default serialization scheme, ASN1. Our evaluation shows that the MME response and procedure completion time increases by to 40X using 150 malicious devices only.

Chapter 6

Future Work

With the growth in adoption of IoT devices, cellular networks are becoming a standard choice for providing always-on connectivity. Although, the traditional infrastructure of cellular networks was intended to support UEs only, but the compatibility of LTE modems has attracted many modern systems like connected cars as an important use case. Therefore, the reliability and availability of the cellular network is becoming a vital challenge for data hungry applications.

From our discussions above, we examined various signalling attacks and concluded that an attacker can very easily exploit open vulnerabilities in LTE design and implementation using very few number of malicious devices to overwhelm the MME resources. This problem becomes more challenging as we move towards 5G where the smaller cell sizes will cause more signalling traffic due to frequent **Handovers**.

Given these insights, we look for quick and robust solutions that can detect and mitigate the signalling attacks on the fly. One of the extensions of this work can be to look for **Learning Based** solutions. Since machine learning has been actively used in classifying traffic patterns [12], it would be useful if we can classify malicious traffic and take appropriate actions without effecting the benign users.

We hope to extend this work in future that provides robust and on-the-fly detection and mitigation solution to the signalling storms caused by malicious devices.

Bibliography

[1] *3GPP Standards*. https://www.3gpp.org/. Accessed: 2019-10-14.

[2] R. Archibald et al. "An IoT control plane model and its impact analysis on a virtualized MME for connected cars". In: *2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. 2016, pp. 1–6. DOI: 10.1109/LANMAN.2016.7548864.

[3] Ramzi Bassil et al. "Signaling Oriented Denial of Service on LTE Networks". In: *Proceedings of the 10th ACM International Symposium on Mobility Management and Wireless Access*. MobiWac '12. Paphos, Cyprus: ACM, 2012, pp. 153–158. ISBN: 978-1-4503-1623-1. DOI: 10.1145/2386995.2387024. URL: http://doi.acm.org/10.1145/2386995.2387024.

[4] *By 2024, over four billion IoT connections on LTE and 5G, Ericsson forecasts*. https://iot.eetimes.com/by-2024-over-four-billion-iot-connections-on-lte-and-5g-ericsson-forecasts/. Accessed: 2019-10-14.

[5] *Cellular Wireshark Traces*. https://www.ng4t.com/wireshark.html. Accessed: 2019-10-14.

[6] *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html. Accessed: 2019-10-14.

[7] Mike Dano. *The Android IM app that brought T-Mobile's network to its knees*. 2010. URL: https://www.fiercewireless.com/wireless/android-im-app-brought-t-mobile-s-network-to-its-knees.

[8] *Ericsson Mobility Report*. https://www.ericsson.com/assets/local/news/2014/11/ericsson-mobility-report-november-2014.pdf. Accessed: 2019-10-14.

[9] Gokce Gorbil, Omer H. Abdelrahman, and Erol Gelenbe. "Storms in Mobile Networks". In: *Proceedings of the 10th ACM Symposium on QoS and Security for Wireless and Mobile Networks*. Q2SWinet '14. Montreal, QC, Canada: ACM, 2014, pp. 119–126. ISBN: 978-1-4503-3027-5. DOI: 10.1145/2642687.2642688. URL: http://doi.acm.org/10.1145/2642687.2642688.

[10] H. Kim et al. "Touching the Untouchables: Dynamic Security Analysis of the LTE Control Plane". In: *2019 IEEE Symposium on Security and Privacy (SP)*. 2019, pp. 1153–1168. DOI: 10.1109/SP.2019.00038.

[11] P. P. C. Lee, T. Bu, and T. Woo. "On the Detection of Signaling DoS Attacks on 3G Wireless Networks". In: *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*. 2007, pp. 1289–1297. DOI: 10.1109/INFCOM.2007.153.

[12] W. Li and A. W. Moore. "A Machine Learning Approach for Efficient Traffic Classification". In: *2007 15th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*. 2007, pp. 310–317. DOI: 10.1109/MASCOTS.2007.2.

[13] Yuanjie Li, Zengwen Yuan, and Chunyi Peng. "A Control-Plane Perspective on Reducing Data Access Latency in LTE Networks". In: *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. MobiCom '17. Snowbird, Utah, USA: ACM, 2017, pp. 56–69. ISBN: 978-1-4503-4916-1. DOI: 10.1145/3117811.3117838. URL: http://doi.acm.org/10.1145/3117811.3117838.

[14] M.Denegan. *Operators Urge Action Against Chatty Apps*. 2011. URL: https://www.lightreading.com/operators-urge-action-against-chatty-apps/d/d-id/687399.

[15] A. Mohammadkhan et al. "Considerations for re-designing the cellular infrastructure exploiting software-based networks". In: *2016 IEEE 24th International Conference on Network Protocols (ICNP)*. 2016, pp. 1–6. DOI: 10.1109/ICNP.2016.7784474.

[16] C. Mulliner and J. Seifert. "Rise of the iBots: Owning a telco network". In: *2010 5th International Conference on Malicious and Unwanted Software*. 2010, pp. 71–80. DOI: 10.1109/MALWARE.2010.5665790.

[17] *Performance Analysis of Mobile Networks Under Signalling Storms*. https://san.ee.ic.ac.uk/publications/PhD_Mihajlo.pdf. Accessed: 2019-10-14.

[18] R. Piqueras Jover. "Security attacks against the availability of LTE mobility networks: Overview and research directions". In: *2013 16th International Symposium on Wireless Personal Multimedia Communications (WPMC)*. 2013, pp. 1–9.

[19] Anand R. Prasad. *3GPP SAE-LTE Security*. 2011. URL: https://niksun.com/presentations/day2/NIKSUN_WWSMC_July26_AnandRPrasad.pdf.

[20] G. Redding. *OTT service blackouts trigger signaling overload in mobile networks*. 2013. URL: https://web.archive.org/web/20140609053424/blogs.nsn.com/mobile-networks/2013/09/16/ott-service-blackouts-trigger-signaling-overload-in-mobile-networks/.

[21] S.Corner. *Angry Birds + Android + ads = network overload*. 2011. URL: https://www.itwire.com/business-it-news/networking/47823f.

[22] Altaf Shaik et al. *Practical Attacks Against Privacy and Availability in 4G/LTE Mobile Communication Systems*. 2015. arXiv: 1510.07563 [cs.CR].

[23] Patrick Traynor et al. “On cellular botnets: measuring the impact of malicious devices on a cellular network core”. In: *ACM Conference on Computer and Communications Security*. 2009.

[24] Rethink Wireless. *DoCoMo demands Google's help with signalling storm*. 2012. URL: https://web.archive.org/web/20120202054002/http://www.rethink-wireless.com/2012/01/30/docomo-demands-googles-signalling-storm.htm.