

Feasibility of Signaling Storms in 3G/UMTS Operational Networks

Frederic Francois^(✉), Omer H. Abdelrahman, and Erol Gelenbe

Intelligent Systems and Networks Group, Department of Electrical and Electronic Engineering, Imperial College, London SW7 2BT, UK
{f.francois,o.abd06,e.gelenbe}@imperial.ac.uk

Abstract. Signaling storms are becoming prevalent in mobile networks due to the proliferation of smartphone applications and new network uses, such as machine-to-machine communication, which are designed without due consideration to the signaling overheads associated with the de/allocation of radio resources to User Equipment (UE). In this work, we conduct a set of experiments on a 3G operational mobile network to validate previous claims in literature that it is possible to significantly change the signaling behavior of a normal UE so that the UE has an adverse impact on the mobile network. Our early results show that it is possible to increase by 0.330 *signaling messages/s* the signaling rate of a normal 3G UE loaded with popular applications when it is not in active use by the owner. In addition, we explore the different factors which can either increase or decrease the effectiveness of signaling attacks on mobile networks.

Keywords: Signaling storms · Radio resource control · 3G/UMTS · Malicious mobile applications · 4G/LTE · M2M

1 Introduction

Mobile networks are increasingly susceptible to Radio Resource Control (RRC) based signaling storms because of the proliferation of smartphone applications [5, 8, 16] and new network uses such as machine-to-machine and Internet of Things communication [22] which are not optimized in terms of signaling load and are not tested by mobile operators. In this paper, we present experiments to verify the feasibility of RRC-based signaling storms in operational 3G networks by measuring the number of successful RRC transitions that an attacker can trigger on a targeted User Equipment (UE). The attack is performed over the public Internet where a computer, acting as the attacker, is used to ping the targeted UEs at a variable interval to observe the relationship between the ping frequency and number of successfully triggered RRC transitions. This setup emulates both deliberate RRC-based signaling attacks as well as signaling storms caused by either misbehaving or malicious applications that frequently establish and tear-down data connections in order to transfer small amounts of data.

In 3G/UMTS networks, each UE has a RRC state machine which controls the amount of bandwidth resources that it is currently allocated [6]. The RRC state machine has 4 states ordered in terms of increasing energy consumption and bandwidth allocation: *IDLE*, *PCH*, *FACH* and *DCH*. The RRC state of a UE in a 3G network is controlled by a mobile network element called *Radio Network Controller* (RNC) where the transition between the different RRC states requires different number of signaling messages [13] to be exchanged between the UE and the RNC. A UE can move from a higher-bandwidth RRC state to a lower one after a network operator specified timeout if no data traffic is communicated between the UE and the mobile network during this timeout.

In current literature, there are numerous prior experimental work [3, 14, 15, 18, 19, 23] which looks mainly at how to infer the RRC timeouts and the impact of applications on RRC signaling load. In [19], the authors infer the type and parameters of the RRC state machine of 2 operational 3G networks by probing the network through the transmission of different amount of data between a UE and a server on the public Internet. In contrast, [3] assumes only one type of RRC state machine and infer its parameters by using ICMP packets as probe packets. The main author of [19] developed a new RRC state inference algorithm in [18] which provides better accuracy and then uses the algorithm to characterize the signaling, energy and bandwidth utilization of mobile applications by analysing their packet traffic traces only. The authors of [23] carried experiments to measure the impact of RRC timeouts on the power consumption, signaling load and web quality of experience. [15] analyzes the impact of the frequency of keep-alive messages on the energy consumption of the UE in 3G networks while we concentrate on the signaling load and use a UE which is a modern smartphone loaded with popular applications that most users have installed on their phone nowadays. In [14], the authors develop an android application which can measure the RRC signaling, radio resources and energy efficiency of background applications by logging the data packets and corresponding RRC state on a targeted UE.

Our previous work on signaling storms in the context of the NEMESYS project [2, 10] has involved the mathematical modeling, simulation and analysis of the impact of different RRC-based signaling storms in 3G/UMTS networks [1, 11, 12] and 4G/LTE networks [7]. In our recent work, we also investigated methods for the detection and mitigation of signaling storms through the use of RRC timeout adjustment [17] and counters [9].

1.1 Motivation

Nowadays, smartphones often run many applications that communicate over the Internet even when users are inactive in order to enable always-on connectivity which allows users to receive promptly new data such as social media updates, VoIP calls and messages and location-based services. This mix of applications may hinder the ability of either a deliberate RRC-based signaling attacker or a malfunctioning application to cause high signaling load since some of the attacking traffic will not trigger changes in the RRC state because other normal appli-

cations have already performed unknowingly the required changes in RRC state to carry the attacking traffic.

Although the impact of RRC-based signaling overload on mobile networks has been evaluated extensively in [1, 7, 12] using mathematical and simulation models, the assumption therein is that the attacker is able to control to a great extent the severity of the attack so that the resulting load in the network is proportional to the rate at which either attacking or misbehaving traffic is generated. The set of experiments designed in this paper aims to validate this assumption in a realistic setting by carrying out RRC-based signaling attacks on an operational mobile network. Furthermore, an additional objective of this set of experiments is to evaluate whether attacks can be optimized in the presence of active mobile applications by modifying the frequency of either malicious or misbehaving transmissions. The results of such experiments will help in the design of more accurate normal UE and attack models which will lead to the ability of running more realistic simulation experiments.

2 Description of the Experiments

2.1 Equipment Used

The experiments were conducted on the 3G network of a large operator in the UK, and included the following components which are connected as shown in Fig. 1:

- **Samsung Galaxy SII (GT-I9100):** A 3G phone which acts as the targeted UE. The phone runs stock Android 4.1.2 (carrier branded) and has been rooted to allow applications to run with root privileges. In addition, the phone has the popular packet capture utility *tcpdump* installed on it. Several popular mobile applications, which communicate over the Internet even when the user is not interacting with the device, have been installed on the UE to emulate normal UEs in existing mobile networks. It is expected that the overall communication pattern of these installed applications will reduce the number of successful RRC transitions that can be triggered by the attacker.
- **3G SIM:** To allow the UE to connect to the 3G network of the selected operational mobile operator. The Access Point Name (APN) used by the SIM card was modified so that the phone appears to the network as a cellular WiFi router (known as MiFi) and is allocated a public IP address, rather than a private one, by this particular network operator. Having a public IP allowed us to ping the smartphone from the public Internet to conduct our experiments, which would not have been possible otherwise since direct mobile-to-mobile communication over the cellular network is blocked on this particular network. It should be noted that many mobile operators across the world provide public IP address by default to their customers [20] and therefore, signaling attacks based on pinging the public IP address of UEs can occur on these mobile networks without any APN change. In addition, signaling attacks can still occur when UEs are allocated private IP addresses only if the UEs have either

malicious or misbehaving applications which regularly send user traffic to the mobile network.

- **Linux-based computer:** To act as attacker and to record the changing RRC state and packets communicated over time by the targeted UE. The RRC attacks are carried by a ping generator which pings the public IP of the targeted UE at regular interval. In addition, the targeted UE is connected via a USB cable to the computer and the Android Debug Bridge (ADB) command line tool is used to connect to the targeted UE to retrieve logged information about the changes in RRC state and the packets that are communicated by the UE over the mobile network during the duration of the experiments.

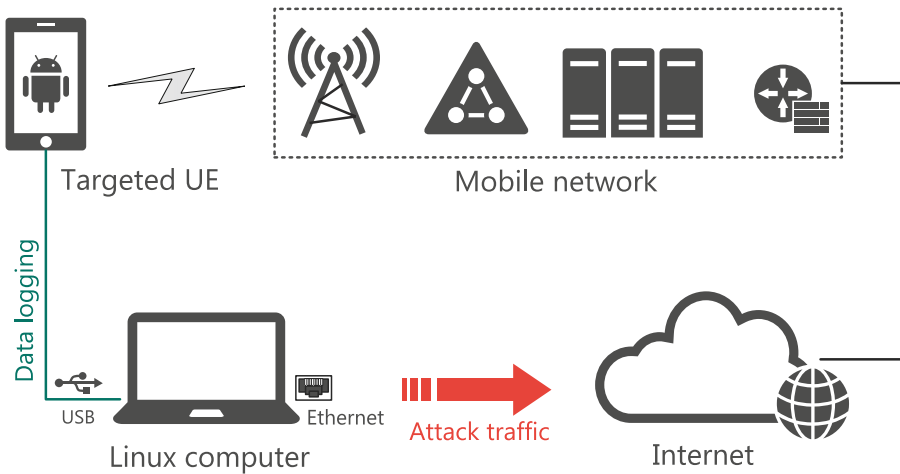


Fig. 1. RRC experiment setup in an operational 3G network.

2.2 Methodology

The RRC experiments are carried by following the following steps: in the first step, Wi-Fi is first deactivated on the targeted UE and then 3G is activated on it so that the UE connects via cellular connection to the Internet and RRC-based attacks can be carried out. The public IP address of the UE is noted so that the ping generator can be configured to attack the UE.

In the second step, two ADB terminals are open on the computer attached to the targeted UE. The first terminal is used to issue commands to the *tcpdump* utility on the targeted UE to start capturing all packets that the UE is receiving through the radio interface; *tcpdump* also records the time when its filters capture the packets. The second terminal is used to record the RRC state of the UE at regular interval. For the UE to start logging the RRC state, it must be put into *ServiceMode* by dialling **#0011#* on the Samsung Galaxy SII phone (this

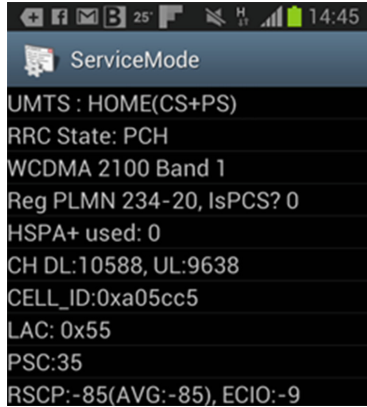


Fig. 2. Screenshot showing the RRC state on the UE after it enters the *ServiceMode* state.

number is known to work only on Samsung Galaxy phones, and other manufacturers may use other codes). When the phone is in *ServiceMode*, it will show the current RRC state as shown in Fig. 2. The second ADB terminal is used to record the RRC state displayed on the UE through the filtering of the system log of the UE for *ServiceMode* information only.

In the third step, the targeted UE is pinged from the attacking computer for the duration of T seconds with a time interval of P seconds between successive pings. The duration of each experiment run T is chosen according to the inter-ping time P such that the total number of attack pings T/P is sufficiently large to provide statistically significant results. In this set of experiments, the default size of the *ICMP Echo Request* packets of the *ping* utility is not changed and was measured to be 100 bytes in the *tcpdump* capture. The impact of the payload size of the *ICMP Echo Request* packets on the type and frequency of RRC transitions triggered is left for future work.

In the final step, the packet and RRC records are then analyzed to obtain the number of successful transitions due to the attack: an attacking ping is deemed to have triggered a successful RRC attack if there is a RRC promotion within $\pm\epsilon$ seconds from the time the attacking ping is recorded by *tcpdump* and there is a RRC demotion within D seconds from the time a successful promotion has been triggered and completed.

2.3 Metrics of Interest

We use two performance measures in order to quantify the effectiveness of the attack. The first metric S captures the proportion of ping messages that successfully trigger a RRC attack:

$$S = \frac{\#\text{successfully triggered RRC attacks}}{\#\text{ping messages captured at the UE}} \times 100\% \quad (1)$$

Clearly, higher values of S reflect higher attack success probability. However, each type of RRC state transition causes a certain number of signaling messages to be exchanged in the network as shown in Table 1. Therefore, the impact of the attack is better characterized by taking into consideration the number of successful transitions of each type $r_{x \rightarrow y}$ and using the values $n_{x \rightarrow y}$ in Table 1 to compute the effective attack rate A as follows:

$$A = \frac{\sum_{\forall x \rightarrow y} r_{x \rightarrow y} \times n_{x \rightarrow y}}{T} \quad (2)$$

Note that not all possible RRC transitions are shown in Table 1, since only certain transitions were observed in the experiment.

Table 1. No. of signaling messages exchanged per RRC transition type [1]

Start state, x	End state, y	No. of messages, $n_{x \rightarrow y}$
PCH	FACH	3
FACH	DCH	7
PCH	DCH	10
DCH	PCH	5
FACH	PCH	2
DCH	FACH	5

3 Results

The parameter D in our set of experiments controls the maximum time limit that a demotion must happen after a promotion occurred near an attacking ping for the transitions to be considered as an attack. During a successful attack, the time that it takes for the UE to demote back to a lower RRC state depends on many factors, the first one being the time it takes to finish communicating all the packets related to the attack, this includes the varying end-to-end delay between the UE and the attacking computer. The second factor is the timeout that the network operator has set for RRC demotion to occur if no data traffic is being transmitted between the UE and the mobile network. It should be noted that normal UE data traffic can cause the UE to either stay in the higher RRC state that the attacking ping caused it to promote to or promote to an even higher RRC state. When normal traffic happens after an attacking promotion but before its associated demotion, the attack is deemed to have failed. Hence, the value of D helps to identify RRC promotion and demotion transition pairs that are affected by normal traffic and should not be counted as attacks. In order to help us identify the appropriate value of D , we compare in Fig. 3 the histogram of the time between promotion and demotion when there are ping attacks occurring and not. It can be observed that most of the change in the

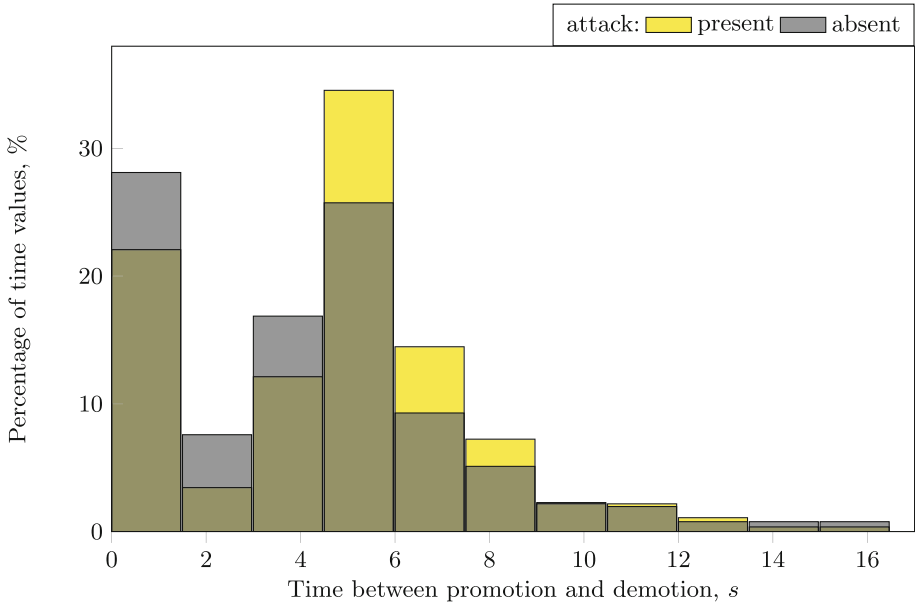


Fig. 3. Bar chart showing the percentage of “time spent in promoted RRC state before demotion” values falling within a specified time interval when RRC-based signaling attack is present and absent. The bin size is set to 1.5 s.

distribution of the time spent in promoted RRC state happens before 9 s and hence, in this work we choose D to be 9. We also set D to ∞ to have a measure of RRC transitions that were not counted.

Tables 2 and 3 show the results of the RRC experiment for different ping intervals P and maximum time between an attacking promotion and demotion for an attack to be considered successful D , with the parameter ϵ set to 2 s to match the longest sampling interval (which is not under our control) of the RRC state of the targeted UE during the whole set of experiments. The results show that the effective attack rate A is highest when the ping interval $P = 10$ s for all considered values of D . In all the experiments, it can be observed that most of the attacking transitions are of type $PCH \rightarrow FACH$ and vice versa because ping attacks are low traffic volume attacks and therefore, the $FACH$ RRC state is enough in most cases for the targeted UE to handle the traffic linked with the ping attacks. The targeted UE can move to the highest RRC state DCH when the ping attacks happen at the same time as when legitimate user traffic is being communicated by the targeted UE and therefore, a transition to DCH state is required in order to carry the additional attack traffic.

Figure 4 shows that the success rate of the attack S increases with the time interval between two consecutive pings P up to a level ($P = 10$ s) after which S stays almost constant. When carrying out a RRC-based signaling attack, attackers may be tempted to set the ping frequency interval P to be slightly larger

Table 2. Results of signaling attacks when $D = 9s$

Ping interval P_s	Exp. Duration T_s	# successful transitions							# Attack pings logged	Attack success rate, $S\%$	Effective attack rate, A msg/s
		Total	PCH → FACH	PCH → DCH	FACH → DCH	FACH → PCH	DCH → PCH	DCH → FACH			
2	1080	0	0	0	0	0	0	0	487	0	0
3	1620	36	0	0	18	0	15	3	445	4.04	0.133
5	2700	194	81	1	15	81	14	2	376	25.8	0.222
10	5400	620	278	2	30	278	31	1	464	66.8	0.330
20	10800	608	266	2	36	266	37	1	491	61.9	0.166
30	16200	698	307	3	39	307	40	2	509	68.6	0.126
40	21600	696	311	1	36	311	35	2	522	66.7	0.093

Table 3. Results of signaling attacks when $D = \infty s$

Ping interval D_s	Exp. Duration T_s	# successful transitions							# Attack pings logged	Attack success rate, $S\%$	Effective attack rate, A msg/s
		Total	PCH → FACH	PCH → DCH	FACH → DCH	FACH → PCH	DCH → PCH	DCH → FACH			
2	1080	0	0	0	0	0	0	0	487	0	0
3	1620	48	1	0	23	1	19	4	445	5.39	0.173
5	2700	332	146	1	19	146	18	2	376	44.1	0.360
10	5400	736	333	2	33	333	34	1	464	79.3	0.387
20	10800	736	326	2	40	326	40	2	491	74.9	0.198
30	16200	784	346	3	43	346	43	3	509	77.0	0.141
40	21600	786	355	1	37	355	36	2	522	75.3	0.103

than the timeouts for demotion in order to maximize the severity of the attacks. Unfortunately, this way of carrying out attacks is not the most efficient one since only a fraction of the attack pings can trigger RRC transitions due to variations in the end-to-end delay between the attacker and the targeted UE, which can decrease the inter-arrival times of ping messages at the UE and do not allow the targeted UE to have enough time to undergo RRC demotions so that the attack pings can trigger malicious RRC promotions again. Therefore, the attacker should add additional time between his attacks to take into consideration the variation in end-to-end delay.

When P is very small, the majority of the attack pings fails to trigger RRC transitions because the targeted UE does not have enough time to undergo the RRC timeout(s) and demote to lower RRC states before the next attack ping arrives. The targeted UE stays most of the time in the higher RRC states where it is able to handle the additional traffic associated with the attack pings without further promotions. Indeed, a large scale attack with very small P can potentially cause bandwidth starvation in the network, which is a data plane type of attack, since the UE remains active throughout the duration of the attack, effectively reducing the available bandwidth in the affected base stations and depleting the battery of the targeted UEs.

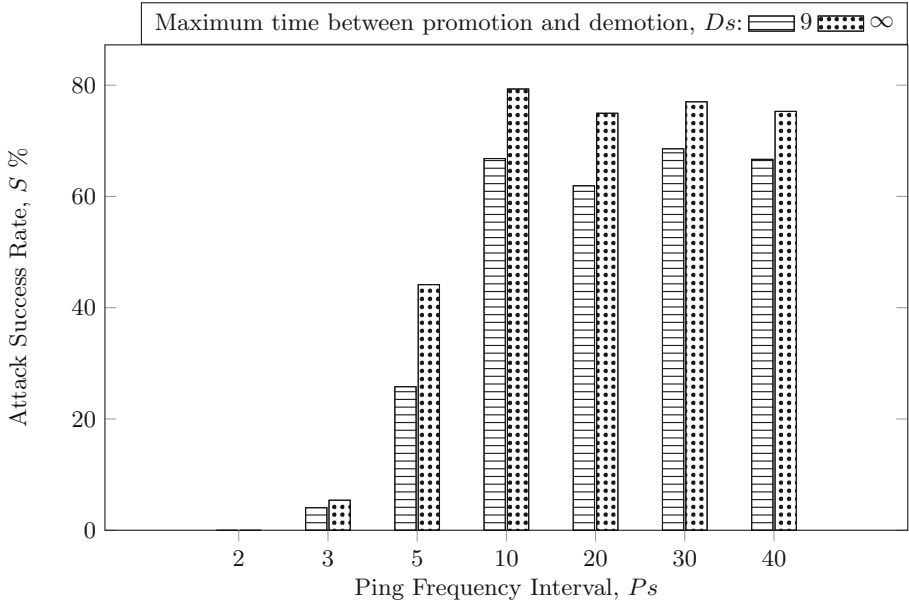


Fig. 4. Change in the attack success rate S when the ping interval P is increased and the maximum time between a successful attacking promotion and demotion is changed D .

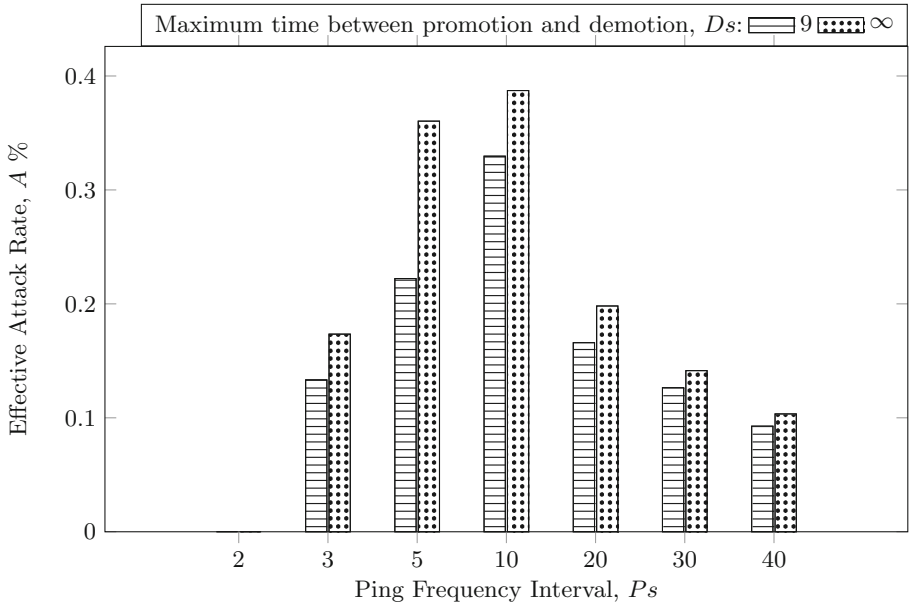


Fig. 5. Change in the effective attack rate A when the ping interval P is increased and the maximum time between a successful attacking promotion and demotion is changed D .

Note again that the success rate of the attack S only measures the probability of an attack ping triggering an RRC transition, but does not reflect the impact of these triggered transitions in the network, which is captured by the effective attack rate A shown in Fig. 5 where the number of signaling messages required for each type of triggered transitions is taken into consideration. We see that A increases with P up to a maximum value when $P = 10$ s (or the attack rate = 0.1 pings/s), then drops monotonically as P is increased as one would expect from a low rate attack.

4 Conclusions and Future Work

This paper has shown that it is possible to carry RRC-based signaling attacks at a very high success rate, i.e. around 70 %, in operational mobile networks by optimizing the frequency at which attacks are carried out so that the attacks are not heavily reduced by the communication pattern of normal applications and by variations in the end-to-end delay between the attacker and targeted UEs. While it was possible to maximize the impact of the attack for a single UE, with a high effective attack rate of 0.330 signaling messages/s at a ping rate of 0.1 pings/s, it may be difficult to optimize the attack for all UEs using the same attack interval because different users have different communication pattern.

An important finding in this paper is that, in contrast to previous belief, RRC-based signaling attacks cannot be optimized based only on the configurations of the networks (i.e. timeout values) since variations in the end-to-end delay between the attacker and targeted UEs and the communication patterns of normal applications on the targeted UEs can significantly reduced the effectiveness of the attacks.

Finally, the experiment has shown that mobile network operators are now following best practices by setting the timeout in state *PCH* to be very large and the buffer threshold at the RNC for transitioning from *FACH* to *DCH* to be also large, thus significantly reducing the effect of chatty mobile applications on the control plane of the network but this comes at the expense of higher energy consumption for the radio subsystem of the UEs.

In future work, there are a number of limitations to the current setup of the experiments which can be improved. First, the RRC state can only be logged when the targeted UE is in *ServiceMode*, i.e. when the RRC state is displayed on the UE screen. This limits the ability to carry out the experiment when the user is actively using the UE. Thus, we expect that the values of A , shown in Fig. 5, represent an upper bound for the load on the network. In practice, this load will be reduced by the activities of the user which will generate normal traffic more frequently.

Second, we found that the most reliable way to capture packets, with ping still working, was through *tcpdump* which is controlled by an ADB terminal. Future work will involve using a terminal directly on the phone, which will be running in the background, to run *tcpdump* so that together with the logging of RRC states in the background, the collection of both packets and RRC data

can be done without preventing users from actively using the phone and also be mobile. This will provide more realistic data about when and at what frequency the attacks can be performed against the phone.

In our current work, the attacking ping messages cause the RRC transitions to be mostly of type $PCH \rightarrow FACH$ and vice versa because ping packets are small and can be handled with the $FACH$ RRC state. If the UE is already in $FACH$, the UE has in most cases enough capacity to handle the ping packets without transitioning to the DCH state. In future work, it might be useful to perform more volumetric attacks, by for e.g. increasing the payload of the $ICMP Echo Request$ packets to the maximum, to try to trigger more $PCH \rightarrow FACH \rightarrow DCH$, $FACH \rightarrow DCH$ and vice versa transitions which may lead to higher attack success ratio and also higher induced signaling in the mobile network.

Finally, we aim to repeat the RRC-based signaling attack experiments carried in this paper in the context of an operational 4G/LTE network where a simplified RRC state machine is used and new enhancements such as Machine Technology Communication (MTC) are being introduced to alleviate the impact of machine-to-machine communications on 4G/LTE networks [4, 21].

Acknowledgments. The work presented in this paper was supported by the EU FP7 research project NEMESYS (Enhanced Network Security for Seamless Service Provisioning in the Smart Mobile Ecosystem), under grant agreement no. 317888 within the FP7-ICT-2011.1.3 Trustworthy ICT domain.

References

1. Abdelrahman, O.H., Gelenbe, E.: Signalling storms in 3G mobile networks. In: Proceedings of IEEE International Conference on Communications (ICC), pp. 1017–1022 (2014)
2. Abdelrahman, O.H., Gelenbe, E., Gorbil, G., Oklander, B.: Mobile network anomaly detection and mitigation: the NEMESYS approach. In: Gelenbe, E., Lent, R. (eds.) ISCIS 2013. LNEE, vol. 264, pp. 429–438. Springer, New York (2013)
3. Barbuzzi, A., Ricciato, F., Boggia, G.: Discovering parameter setting in 3g networks via active measurements. *IEEE Commun. Lett.* **12**(10), 730–732 (2008)
4. Cheng, M.Y., Lin, G.Y., Wei, H.Y., Hsu, A.C.C.: Overload control for machine-type-communications in lte-advanced system. *IEEE Commun. Mag.* **50**(6), 38–45 (2012)
5. Corner, S.: Angry Birds + Android + ads = network overload. *IT Wire* (2011). <http://www.itwire.com/business-it-news/networking/47823>
6. ETSI 3GPP: 3Gpp. TS 25.331: Universal mobile telecommunications system (UMTS) radio resource control (RRC) protocol specification (2015)
7. Francois, F., Abdelrahman, O.H., Gelenbe, E.: Impact of signaling storms on energy consumption and latency of LTE user equipment. In: Proceedings of the 7th IEEE International Symposium on Cyberspace safety and security (CSS 2015), New York (2015)
8. Gabriel, C.: DoCoMo demands Google's help with signalling storm. *Rethink Wireless* (2012). <http://www.rethink-wireless.com/2012/01/30/docomo-demands-googles-signalling-storm.htm>

9. Gelenbe, E., Abdelrahman, O.H.: Time-outs and counters against storms (2014)
10. Gelenbe, E., Gorbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., Lyberopoulos, G.: Security for smart mobile networks: the NEMESYS approach. In: Proceedings of IEEE Global High Tech Congress on Electronics (GHTCE), pp. 63–69, Shenzhen (2013)
11. Gorbil, G., Abdelrahman, O.H., Gelenbe, E.: Storms in mobile networks. In: Proceedings of 10th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet), pp. 119–126, Montreal, Canada (2014)
12. Gorbil, G., Abdelrahman, O.H., Pavloski, M., Gelenbe, E.: Modeling and analysis of RRC-based signalling storms in 3G networks. *IEEE Trans. Emerg. Topics Comput.* **4**(1), 113–127 (2015)
13. GSMA: Fast dormancy best practises. White paper (2011). <http://www.gsma.com/newsroom/ts18-v10-tsg-prd-fast-dormancy-best-practices>
14. Gupta, S., Garg, R., Jain, N., Naik, V., Kaul, S.: Android phone based appraisal of app. behavior on cell networks. In: Proceedings of the 1st International Conference on Mobile Software Engineering and Systems, MOBILESoft 2014, pp. 54–57 (2014)
15. Haverinen, H., Siren, J., Eronen, P.: Energy consumption of always-on applications in wcdma networks. In: IEEE 65th Vehicular Technology Conference, 2007, VTC2007-Spring, pp. 964–968 (2007)
16. Jiantao, S.: Analyzing the network friendliness of mobile applications. Technical report (2012)
17. Pavloski, M., Gelenbe, E.: Signaling attacks in mobile telephony. In: Proceedings of the 11th International Conference on Security and Cryptography (SECRYPT 2014), pp. 206–212 (2014)
18. Qian, F., Wang, Z., Gerber, A., Mao, Z., Sen, S., Spatscheck, O.: Profiling resource usage for mobile applications: a cross-layer approach. In: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, MobiSys 2011, pp. 321–334 (2011)
19. Qian, F., Wang, Z., Gerber, A., Mao, Z.M., Sen, S., Spatscheck, O.: Characterizing radio resource allocation for 3g networks. In: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, IMC 2010, pp. 137–150 (2010)
20. Qian, Z., Wang, Z., Xu, Q., Mao, Z.M., Zhang, M., Wang, Y.M.: You can run, but you can't hide: exposing network location for targeted DoS attacks in cellular networks. In: Proceedings of 19th Annual Network and Distributed System Security Symposium (NDSS), San Diego, CA (2012)
21. Qualcomm: LTE MTC: optimizing LTE advanced for machine-type communications (2014). <https://www.qualcomm.com/media/documents/files/lte-mtc-optimizing-lte-advanced-for-machine-type-communications.pdf>
22. Research, R.: Gsma seeks to avert chaos on mobile iot networks (2014). <http://www.rethinkresearch.biz/articles/gsma-seeks-avert-chaos-mobile-iot-networks/>
23. Schwartz, C., Hofffeld, T., Lehrieder, F., Tran-Gia, P.: Angry apps: the impact of network timer selection on power consumption, signalling load, and web QoE. *J. Comput. Netw. Commun.* **2013**, 13 (2013). Article ID 176217