Counter based Detection and Mitigation of Signalling Attacks

Mihajlo Pavloski, Gökçe Görbil and Erol Gelenbe
Department of Electrical and Electronic Engineering, Intelligent Systems and Networks Group,
Imperial College, London SW7 2AZ, U.K.

Keywords: Signalling Attacks, Detection, Mitigation, Denial of Service.

Abstract: The increase of the number of smart devices using mobile networks' services is followed by the increase of the number of security threats for mobile devices, generating new challenges for mobile network operators. *Signalling attacks* and *storms* represent an emerging type of distributed denial of service (DDoS) attacks and happen because of special malware installed on smart devices. These attacks are performed in the control plane of the network, rather than the data plane, and their goal is to overload the Signalling servers which leads to service degradation and even network failures. This paper proposes a detection and mitigation mechanism of such attacks which is based on counting repetitive bandwidth allocations by mobile terminals and blocking the misbehaving ones. The mechanism is implemented in our simulation environment for security in mobile networks SECSIM. The detector is evaluated calculating the probabilities of false positive and false negative detection and is characterised by very low negative impact on un-attacked terminals. Simulation results using joint work of both detector and mitigator, are shown for: the number of allowed attacking bandwidth allocations, end-to-end delay for normal users, wasted bandwidth and load on the Signalling server. Results suggest that for some particular settings of the mechanism, the impact of the attack is successfully lowered, keeping the network in stable condition and protecting the normal users from service degradations.

1 INTRODUCTION

The use of smart devices and mobile data services in mobile networks record a great increase in the last couple of years. The number of global mobile devices and connections rose for almost half a billion in 2014 out of which smartphones accounted for 88% of the growth. The mobile data traffic grew 69% from 2013 to 2014, reaching 2.5 exabytes per month (Cisco, 2015). In parallel, the number of security threats for mobile devices is rapidly growing with a tenfold increase of mobile malware attacks per month from August 2013 to March 2014 (Kaspersky Lab and INTERPOL, 2014).

Signalling attacks that have emerged as novel security threats to mobile networks are instrumented by such mobile malware. Their purpose is to develop a distributed denial of service (DDoS) attack (Gelenbe et al., 2004) on the control plane of the network rather than the data plane (Gelenbe and Loukas, 2007). The impact of these attacks can be maximised by using groups of mobile devices - *botnets* (Mulliner and Seifert, 2010), and by adapting the attack to the network's parameters (Abdelrahman and Gelenbe, 2014). Similar attacks can also happen due to poor development of smart device applications which use frequent background messages, and are known as *Signalling storms* (Gabriel, 2012; Gorbil et al., 2014). Both of these attacks cannot be detected by traditional flooding-based attack detection systems (Ricciato et al., 2010). Such attacks can be particularly dangerous when coupled with emergency situations (Filippoupolitis et al., 2008) where the physical emergency is exacerbated by the mobile attack that impedes communications, and such attacks also have side effects in significantly increasing the consumption of energy in the mobile operator's backbone and in the hand-held devices (Gelenbe and Morfopoulou, 2011).

The vulnerability that is exploited by these attacks is located in the radio resource control (RRC) part of the system. Whenever a mobile terminal wants to transfer some data, it needs to ask for some communication resources from the network, which triggers a signalling procedure called *connection/radio bearer setup* in 3G Universal Mobile Telecommunications System (UMTS) networks or *random access* in 4G Long Term Evolution (LTE). This procedure involves exchange of up to 22 Signalling messages in the radio access network (RAN) and core network (CN) parts

Pavloski M., Görbil G. and Gelenbe E..
Counter based Detection and Mitigation of Signalling Attacks.
DOI: 10.5220/0005573504130418
In *Proceedings of the 12th International Conference on Security and Cryptography* (SECRYPT-2015), pages 413-418
ISBN: 978-989-758-117-5

in UMTS, while a smaller number of messages are exchanged in LTE. If this behaviour is repeated by a decent number of mobile terminals in the network it can cause overloading of the singling servers which leads to service degradation and even system outages (Gabriel, 2012). For the mobile user this is manifested in high battery consumption (Gupta et al., 2013) and even unwanted billing.

In previous work (Gorbil et al., 2015; Gelenbe et al., 2014) it is shown that these attacks can be identified by their repetitive pattern and low usage of communication resources in order to evade getting detected by flooding security mechanisms. The low bandwidth usage characteristic is used in attacks detection in (Pavloski et al., 2015). Some previous work from analytical aspect in this field is included in (Gelenbe et al., 2014), where the authors look at the repetitive pattern of attacks and propose a mechanism for their detection and mitigation. This paper proposes a detection mechanism using the repetitive pattern of attacks and shows results obtained from its implementation in a simulation environment. The paper is organised as follows. Section 2 explains the vulnerability used by these attacks and covers the details of the mechanism. In Section 3 we implement and evaluate the detector in the SECSIM simulator (Gorbil et al., 2015) in UMTS networks and show results on the joint work by the detector and mitigator. Section 4 concludes paper and suggests future improvements.

2 MECHANISM DESCRIPTION

The proposed mechanism enables detection and mitigation of Signalling attacks and storms per mobile terminal in real-time. The detection part of the mechanism is based on counting the repetitive bandwidth allocations of same type, while the mitigation part on blocking the misbehaving mobile terminal's communication for some time interval. From an implementation perspective, it is important that the mechanism could be implemented on both mobile terminal and network/operator side. If implemented on the mobile terminal side, due to the terminal's limited resources, some special requirements are needed so it does not impose any processing, storage, and memory difficulties to the terminal. For this purpose, the proposed mechanism is envisioned as lightweight background process requiring only two parameters: the time instances of bandwidth allocation and the type of bandwidth allocation. These two parameters are stored in memory for the duration of a time window of length around one minute. To explain why these parameters are used, first we need to briefly explain how attacks work in UMTS network for example.

2.1 Radio Resource Control in UMTS

The radio resource control (RRC) protocol of UMTS is responsible for managing resources in the radio access network (RAN). Each mobile terminal, called user equipment (UE), is associated a *state machine* which maintains the RRC state that could be one of the following:

- *Idle* - the initial state when UE is turned on and it does not have a connection with the network;
- *cell-FACH* - the UE is connected to the network and is allocated a shared channel for low-speed communication;
- *cell-DCH* - the UE is connected to the network and is allocated dedicated bandwidth for high-speed communication;
- *cell-PCH* - a low energy state in which the UE is connected to the network but cannot send/receive data.

When a UE wants to send or receive some data it needs to have established one *radio connection (RC)* and one or more *radio bearer(s) (RB)* with the base station, which is equivalent to obtaining one of cell-FACH or cell-DCH RRC states. Establishing RC and RB requires exchanging up to 22 Signalling messages between the UE and the radio network controller (RNC) depending on the RRC state. After finishing with communication, the UE keeps the RC/RB active for short period, of a couple of seconds, called *inactivity timeout* before bandwidth is deallocated. Signalling attacks work in such way that they trigger bandwidth allocations, even without having any data to send or receive, then wait for the inactivity timeout to elapse, before repeating the same procedure again.

2.2 Detection

As mentioned earlier, the detection part of the mechanism counts repetitive bandwidth allocations of same type, with 'type' being either: allocation of shared bandwidth (cell-FACH state) or dedicated bandwidth (cell-DCH state). A decision of an attack being detected is simply taken when the number of repetitions reaches a predefined threshold called *repetitions threshold* - n. Repetitions are counted in a sliding time window manner t_w of length suitable to the chosen n. If we denote with t_I the duration of the inactivity timer of bandwidth allocation, then obviously we should take t_w such that $t_w > n \cdot t_I$, i.e. the window should be large enough to collect n repetitions. Contrary, if t_w is too large, the mechanism may use more storage space than needed, and may even have bad influence on the decision process. In the following, we take $t_w = 3nt_I$ which is a suitable value used in the simulations. Research in (Gelenbe et al., 2014) looks at the problem from an analytical perspective and shows a way of finding the optimal threshold n.

2.3 Mitigation

The idea for the mitigation technique is based on (Pavloski and Gelenbe, 2014). The authors of the paper propose a mathematical model of RRC part in UMTS and look at the influence of the system's parameters on the impact of the Signalling attack. Furthermore, they manage to lower the attack impact by adding delay to Signalling messages asking for bandwidth allocation. In a similar fashion, in this paper we will use a fixed time duration t_b called *blocking time* in which all the communication of a misbehaving terminal will be blocked. This approach may have a negative influence on normal mobile users, but it is the detector which is responsible for making the right decisions. For that purpose, in the following we will first evaluate the detector's performance.

3 EVALUATION AND SIMULATION RESULTS

The proposed detection and mitigation mechanisms are implemented in the SECSIM simulator, which is briefly described here. Furthermore, the counter based detector is evaluated in terms of *probability of false positive* and *false negative detection*. Finally, both detector and mitigator are set to work together and some results are obtained regarding the Signalling load, delay and resources allocation in the network.

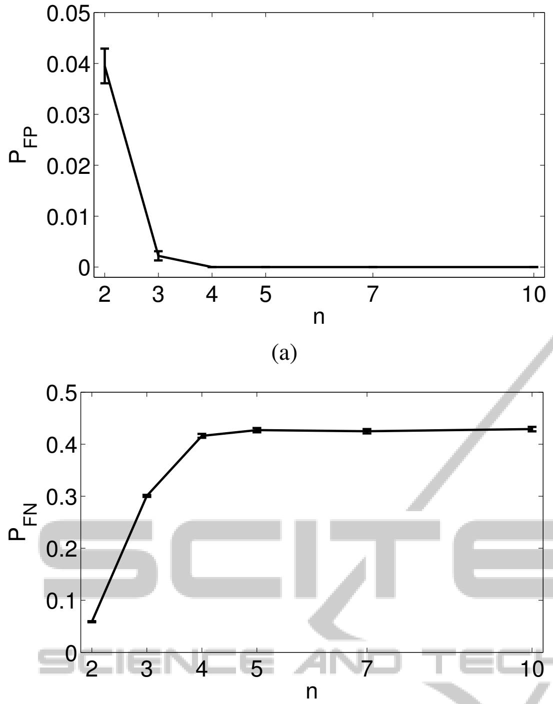
3.1 Simulator Description

The SECSIM simulator focuses on modelling and simulation of the Signalling layer of radio access part of mobile networks. It is developed in Omnet++ (Varga and Hornig, 2008), an object-oriented discrete event simulator. SECSIM is a modular simulator, building network components on top of smaller ones - modules. Its current version contains models of both UMTS and LTE networks, including components such as: UE, RNC, NodeB, SGSN, GGSN, eNodeB, SGW, Internet hosts etc. The UE model consists of the session management (SM), GPRS mobility management (GMM) and RRC layers in the control plane and application layer with both circuit switched and IP applications in the data plane. The transport layer consists TCP and UDP protocols, while there is also a simplified IP layer. MAC and PHY layers are not modelled, while changes in radio conditions are modelled as random variations. The RNC model has the RRC containing a single Signalling server, RANAP, NBAP and GTP protocols. The Signalling server plays a crucial role in the Signalling attacks and their mitigation.

3.2 Detector Evaluation

Since the detector plays central role in the proposed mechanism, it is important to evaluate its performance. The two metrics of interest are: the *false positives probability* P_{FP}, defined as the fraction of time in which an attack is detected but not existing and the *false negatives probability* P_{FN} which is defined as the fraction of time in which an attack is ongoing but is not detected. The P_{FP} metric is particularly important because it shows the error the detector makes because of misclassifying normal data transmissions. In order to protect normal mobile users from being 'punished' it is important to keep P_{FP} value at minimum. To calculate the defined metrics for different values of the threshold $n \in \{2,3,4,5,7,10\}$, we run experiments with 3 simulated hours in a UMTS network of 500 mobile terminals among which 150 perform attacks at random intervals. The selection of 25% attackers is used because it has shown that is enough to cause congestion in the network (Gorbil et al., 2015). All mobile terminals use an application for web browsing, whose parameters come from probability distributions of real world Internet traffic (Ramachandran, 2010), while the 150 attackers have installed an extra application for attack on the DCH state. The attack application is assumed to have estimated the inactivity timeouts of the network with an exponentially distributed error with mean value of 2 seconds. All experiments are repeated 5 times with different seeds for the random number generators. Mitigation is not used in the experiment and $t_w = 3nt_{DCH}$. Figure 1 shows the calculated metrics of interest.

Results show that the false positive probability is generally much lower than the false negative probability for all values of n. The P_{FP} values are satisfactory because for $n \geq 3$, P_{FP} is lower than 0.005, which gives the percentage of normal traffic being confused as an attack. It drops with the rise of n because less normal traffic is distinguished as attack. Contrary, the false negatives probability rises with n because less attacks are detected and more normal traffic is misclassified as attack. Its values are generally high be-



(a)

(b)

Figure 1: a) Probability of false positive detection and b) probability of false negative detection; 95% confidence interval used.

cause of the mixture of attack traffic with normal one. Normal bandwidth allocations interrupt attacking repetitions and reset the counter to 0, failing to detect the attack. Normally, it could also happen that attackers use only an attack application on the mobile terminal in order to increase the impact of the attack. In these case, we expect that attacks will be detected more successfully, and the values of P_{FN} will drop significantly.

3.3 Mitigation Results

Now we're interested in using both the detector and the mitigator at the same time. To mitigate the attack we will use blocking of the attacking UEs for a time duration of $t_b = 60$s. The blocking is done immediately when attack is detected. The simulation scenario for this purpose is same as in Section 3.2, only this time the attacking terminals attack during the whole duration of the experiment and the mitigation is switched on in all terminals.

First, we are interested in counting the successful bandwidth allocations that happen due to attacks on Figure 2. The Figure depicts the total number of allocations per mobile terminal for the duration of 3 hours, and the corresponding normalized values. Although the detector successfully detects over 94% of attacks for $n = 2$ ($P_{FN} = 0.059$), the mitigator does not stop all of them. As expected, the number of attack allocations increases with the increase of n because the detector waits for more repetitions to happen before making a decision. For $n \geq 5$ our mechanism shows unsatisfying results.

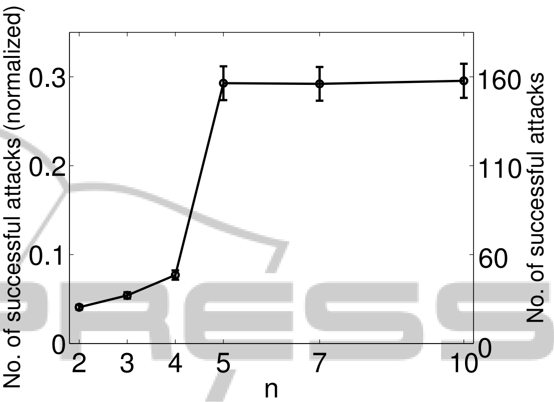


Figure 2: Average number of successful bandwidth allocations due to attack per mobile terminal; 95% confidence interval used.

Furthermore, Figure 3 shows the average end-to-end delay experienced by normal terminals. This is the delay measured on the application layer of the protocol stack. The selected number of attackers (25% of all terminals) is enough to perform a successful Signalling attack and overload the Signalling server. This results in higher delays for the normal terminals in the network. Results suggest that using the proposed mechanism with a threshold of $n \in \{2,3,4\}$, the system is kept stable and normal delays are experienced. Again for $n \geq 5$, the mechanism does not manage to mitigate the attack. The abrupt increase between $n = 4$ and $n = 5$ is due to the type of normal web traffic used, which happens with a rate that usually doesn't allow the attacker to perform more than 4 consecutive repetitions. Furthermore, the delay variability for $n \geq 5$ is much higher than for $n < 5$. This could be possible because in a congested system packets are consecutively lost and retransmitted using the TCP protocol. For longer data bursts, this could result in pretty long delays.

Finally, we are interested in the amount of communication resources (time-frequency blocks) wasted due to the attacks. Figure 4 shows the average allocated bandwidth in uplink direction in Cell-DCH state for attacked and normal mobile terminals in duration of one hour and the corresponding normalized values. Note that in Cell-DCH state a bandwidth allowing high-speed transmission is dedicated exclusively

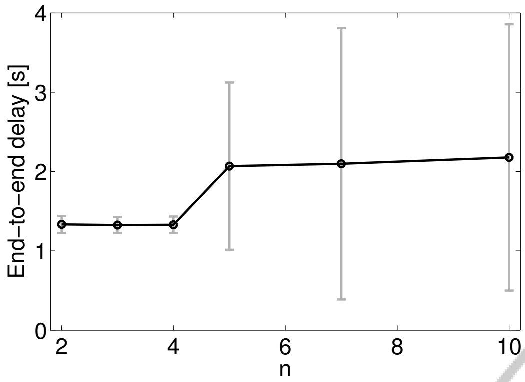


Figure 3: Average end-to-end delay per normal mobile terminal; 99% confidence interval used.

to the requesting terminal, a feature that is excluded in the following generations of mobile networks, like HSPA and LTE. Results show that the amount of resources allocated per attacked terminal is much higher than per normal one, such that for $n = 10$ the attackers are allocated around 600 MB more than normal users in a single hour. Looking at this from a billing perspective, the user containing this kind of malware on his device may be charged much more than usual. Anyway, for $n \in \{2,3,4\}$ the proposed mechanism manages to lower the impact of the attack and the amount of wasted resources drops to 40-90 MB per terminal per hour.

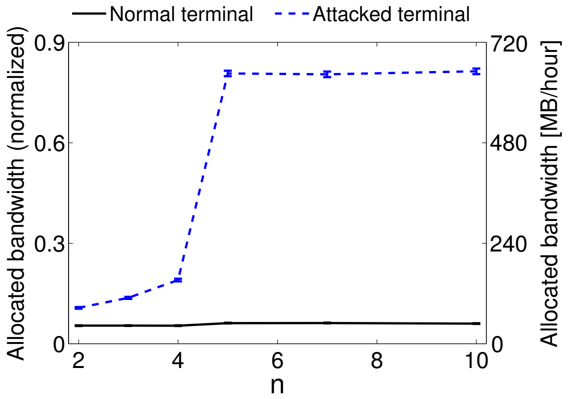
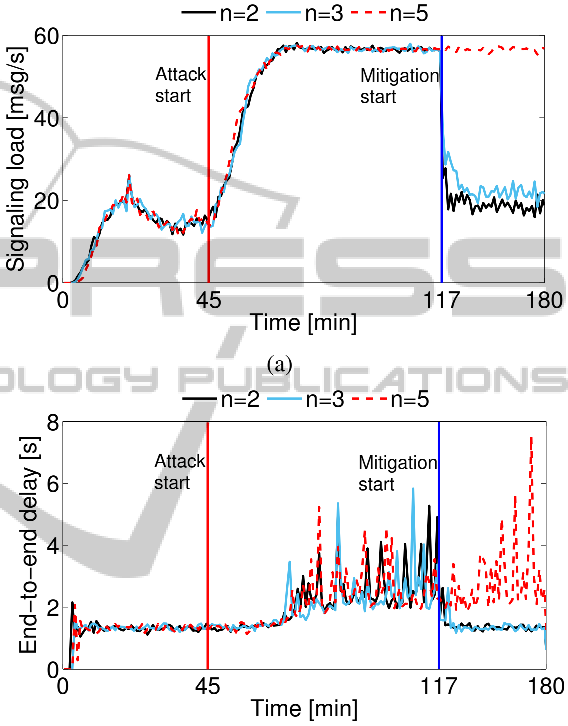


Figure 4: Average allocated uplink bandwidth in Cell-DCH for attacked and normal terminals; 95% confidence interval used.

To show how the mechanism works in time domain, we will conduct another small experiment. The scenario setup is again similar to the described in Section 3.2, only this time the mitigation starts in the 117 minute, a randomly chosen value. Figure 5 shows the load on the Signalling server in the radio network controller, in terms of received messages per second, and the average end-to-end delay experienced by normal mobile users. From the moment of start of the attack, the load on the RNC is constantly increasing and after it reaches some maximum value the normal users start experiencing communication delays. Starting the mitigation with $n \in \{2,3\}$ helps in stabilizing both the network load and the experienced delay. The variation in the delay in congested system is again due to TCP retransmission of packets.



(a)

(b)

Figure 5: a) Signalling load on the RNC and b) End-to-end delay per normal terminal.

4 CONCLUSION

Signalling attacks and storms are a reality in the last couple of years, forcing many mobile operators to look for solutions. These attacks threaten the stability of networks, and on many occasions reduce the quality of offered services and even cause complete network outages (Gelenbe et al., 2013b; Gelenbe et al., 2013a). We have distinguished some basic characteristics of these attacks and used the 'repetitive pattern' to define a detection technique which is capable of detecting attacks in real-time. The technique can be implemented on both mobile terminal and network sides of the system. We evaluated the proposed detector calculating the probability of false positive and false negative detection. Furthermore, we used the detector together with a simple attack mitigation technique and provided some simulation results on network load, end-to-end delay and wasted communication resources. Certain settings of the mechanism manage to detect attacks and lower their impact. Further improvements could be done in combining the proposed mechanism with the one based on 'low bandwidth usage' characteristic to obtain better results.

ACKNOWLEDGEMENTS

This work is part of the EU FP7 project NEMESYS (Enhanced Network Security for Seamless Service Provisioning in the Smart Mobile Ecosystem), under grant agreement no.317888 within the FP7-ICT-2011.1.4 Trustworthy ICT domain.

REFERENCES

Abdelrahman, O. H. and Gelenbe, E. (2014). Signalling storms in 3G mobile networks. In *Proceedings of IEEE International Conference on Communications (ICC'14), Communication and Information Systems Security Symposium*, pages 1017–1022, Sydney, Australia.

Cisco (2015). Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019. White Paper.

Filippoupolitis, A., Hey, L., Loukas, G., Gelenbe, E., and Timotheou, S. (2008). Emergency response simulation using wireless sensor networks. In *Proceedings of the 1st international conference on Ambient media and systems*, page 21. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Gabriel, C. (2012). DoCoMo demands Google's help with signalling storm.

Gelenbe, E., Abdelrahman, O. H., and Görbil, G. (2014). Time-outs and counters against storms.

Gelenbe, E., Gellman, M., and Loukas, G. (2004). Defending networks against denial of service attacks. In Carapezza, E., editor, *Proceedings of the Conference on Optics/Photonics in Security and Defence (SPIE), Unmanned/Unattended Sensors and Sensor Networks*, volume 5611, pages 233–243, London, UK.

Gelenbe, E., Görbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., and Lyberopoulos, G. L. (2013a). NEMESYS: enhanced network security for seamless service provisioning in the smart mobile ecosystem. In Gelenbe, E. and Lent, R., editors, *Information Sciences and Systems 2013 - Proceedings of the 28th International Symposium on Computer and Information Sciences, ISCIS 2013, Paris, France, October 28-29, 2013*, volume 264 of *Lecture Notes in Electrical Engineering*, pages 369–378. Springer.

Gelenbe, E., Görbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., and Lyberopoulos, G. L. (2013b). Security for smart mobile networks: The NEMESYS approach. In *2013 International Conference on Privacy and Security in Mobile Systems, PRISMS 2013, Atlantic City, NJ, USA, June 24-27, 2013*, pages 1–8. IEEE.

Gelenbe, E. and Loukas, G. (2007). A self-aware approach to denial of service defence. *Computer Networks*, 51(5):1299–1314.

Gelenbe, E. and Morfopoulou, C. (2011). A framework for energy-aware routing in packet networks. *The Computer Journal*, 54(6):850–859.

Gorbil, G., Abdelrahman, O. H., and Gelenbe, E. (2014). Storms in mobile networks. In *Proceedings of the 9th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet'14)*, pages 119–126.

Gorbil, G., Abdelrahman, O. H., Pavloski, M., and Gelenbe, E. (2015). Modeling and analysis of RRC-based signaling storms in 3G networks. *IEEE Transactions on Emerging Topics in Computing, Special Issue on Emerging Topics in Cyber Security*, PP(99):1–14.

Gupta, M., Jha, S., Koc, A., and Vannithamby, R. (2013). Energy impact of emerging mobile internet applications on LTE networks: issues and solutions. *Communications Magazine, IEEE*, 51(2):90–97.

Kaspersky Lab and INTERPOL (2014). Mobile Cyber Threats. Joint Report.

Mulliner, C. and Seifert, J.-P. (2010). Rise of the iBots: Owning a telco network. In *Proc. 5th Inter. Conf. on Malicious and Unwanted Software (MALWARE'10)*, pages 71–80.

Pavloski, M. and Gelenbe, E. (2014). Signaling attacks in mobile telephony. In *Proceedings of the 11th International Conference on Security and Cryptography (SECRYPT'14)*, pages 206–212.

Pavloski, M., Gorbil, G., and Gelenbe, E. (accepted, 2015). Bandwidth Usage - Based Detection of Signaling Attacks. In *30th International Symposium on Computer and Information Sciences*. Springer Verlag.

Ramachandran, S. (2010). Web metrics: Size and number of resources.

Ricciato, F., Coluccia, A., and DAlconzo, A. (2010). A review of dos attack models for 3g cellular networks from a system-design perspective. *Computer Communications*, 33(5):551 – 558.

Varga, A. and Hornig, R. (2008). An overview of the OMNeT++ simulation environment. In *Proc. 1st Inter. Conf. on Simulation Tools and Techniques for Communications, Networks and Systems W'shops (Simutools'08)*, pages 60:1–60:10.