Signaling Attacks in Mobile Telephony

Mihajlo Pavloski and Erol Gelenbe
Department of Electrical & Electronic Engineering, Intelligent Systems and Networks Group, Imperial College, London SW7 2AZ, UK
{*m.pavloski13, e.gelenbe*}*@imperial.ac.uk*

Keywords:
Signaling Attacks, UMTS, Mitigation, Denial of Service

Abstract:
Mobile telephony based on UMTS uses finite-state control schemes for wireless channels and for signaling across the network. These schemes are used systematically in various phases of the communication and are vulnerable to attacks that can bring down the network through unjustified bandwidth allocation and excessive signaling across the control plane. In this paper we identify those system parameters which are critical to the success of such attacks, and propose changes that can limit the effect of the attack. The approach is based on establishing a mathematical model of a UMTS system that is undergoing attacks, and on showing how parameters can be optimally modified to minimise the effect of the attack as experienced by the mobile device and the network.

1 INTRODUCTION

Smartphone and tablet use has increased rapidly in the last few years, allowing users to access the Internet at any time and place (Marck, 2013). The popularity of smartphone applications has also rapidly increased and in order to improve user experience and provide real-time services, application developers tend to provide always-on connectivity in their applications by including background traffic and “keep-alive” messaging. All of this would have functioned perfectly well if mobile phones had PC-like Internet connectivity.

However in mobile networks such as Universal Mobile Telecommunications System (UMTS), connections are dynamically created and teareddown to optimise the use of resources. The creation of each connection requires exchanging of additional signaling messages to setup and reserve network resources. Malicious or poorly developed smartphone apps generate traffic which triggers frequent connections and produces excessive signaling in the control plane of the network.

In order to avoid developer’s mistakes, the Global System for Mobile Communications Association (GSMA) has issued a guideline for network-friendly application development (GSMA, 2012). But malicious apps may be intentionally developed to attack the network stability and interrupt it’s services. This type of attack can lead to network overload and represents a Denial of Service (DoS) attack known as a *Signaling Attack* (Abdelrahman and Gelenbe, 2014).

Indeed, it is reported that most of todays networks are vulnerable to signaling attacks that can lead to documented network failures (Wireless, 2012). While from a network perspective such an attack via many mobile devices using the network, would congest the wireless bandwidth and the signaling servers in the backbone, it would also have a negative effect for the user. An infected device would successively trigger unwanted communication with the network and drain the battery of the device, and perhaps also create undesirable billing for services that are being accessed within the UMTS network and outside via web sites.

2 AIMS OF THIS RESEARCH

This paper proposes a probabilistic approach towards optimising network’s parameters in order to lower the impact of signaling attacks. We are interested to see how the network (referred also as system) reacts to a signaling attack. Then we investigate whether it can maintain its stability under an attack by changing some specific state transition time constants, such as it's inactivity timers, or by adding delay in responding to bandwidth request messages.

For this purpose we first propose a mathematical model of a an individual mobile's states in the UMTS system, that includes the effects of a signaling attack. We introduce an attack cost function that incorporates the probability of presence of the mobile's states in the attack and normal states. This model is then used for optimisation. The results obtained suggest that selecting correct values for the parameters can improve the overall system performance even when attacks occur.

The paper is organised as follows. Section 2.1 briefly reviews the related work in the field, while Section 3 discusses the Radio Resource Control (RRC) mechanism in UMTS. In Section 4 we describe our model and its parameters, and in Section 5 we introduce the cost function to be optimised. Section 6 presents and discusses the results that are obtained, while Section 7 presents some conclusions and suggests further research directions.

2.1 Related Work

Security in general (Gelenbe and Wu, 2012) has come to the forefront of much of the research in information technology in recent years, and cybersecurity in particular (Gelenbe et al., 2013b) is viewed as an integral part of security in general. Indeed it is impossible today to address physical security (Gelenbe et al., 1997; Cao and Gelenbe, 1998) without including the impact of cybersecurity. In particular UMTS based mobile network infrastructures which are universally available constitute an essential component of today's secure infrastructures.

The security of wireless networks has been of great interest in recent years (Yu et al., 2014), leading to many research projects in Europe and elsewhere (Gelenbe et al., 2013a). In particular, signaling DoS attacks and their mitigation (Gelenbe and Loukas, 2007) has been a popular research topic in wireless and mobile communications. Publications in the field range from analytical algorithms, simulations using real world data to complex systems for inspecting attacks on mobile networks.The authors in (Kambourakis et al., 2011) present an extensive survey of possible attacks in mobile networks.

A large Markov chain model is used in (Abdelrahman and Gelenbe, 2014) for the mathematical evaluation of signaling attacks parameters, with the objective of identifying the system parameters which should be avoided, namely those that, from an attacker's perspective, produce the largest amount of damage through load in the network. The work in (Wang et al., 2004) regards the detection of traditional flooding-based DoS attacks as a change-point problem and applies the non-parametric CUSUM test for detection. Similarly, in (Lee et al., 2007) a CUSUM test in the early detection algorithm of low-rate, low-volume signaling attacks is suggested and simulations driven by real traces are used to demonstrate the impact of a signaling attack. Good points of the approach are the simplicity, dynamism and small detection time of the method, although the emulation of 3G signaling on WLAN is doubtful and some unrealistic assumptions are made.

The work in (Wu et al., 2010) proposes a randomization of the Radio Resource Management (RRM) and Mobility Management (MM) procedures to hide the parameters which are important to attackers. The analysis of signaling traffic in real-world UMTS network is presented in (Choi et al., 2014). The paper shows a comparison of signaling traffic by different type of mobile applications and its influence on the RRC part of the network. It also explores some application and network layer solutions for controlling application signaling traffic. The authors in (Gupta et al., 2013) inspect the influence of high signaling volumes in LTE networks on the energy consumption in mobile phones. Other modeling approaches of DoS attacks in 3G cellular networks are reviewed in (Ricciato et al., 2010).

3 UMTS RADIO RESOURCE CONTROL

The management of communication resources in UMTS is regulated by the Radio Resource Control (RRC). In general, there are two RRC connectivity modes: Idle and Connected. In Idle mode there aren't any radio resources used between the User Equipment (UE) and the Radio Network Controller (RNC). The few tasks a UE performs in Idle mode are related to neighbor cell monitoring, cell re-selection, paging and broadcast data reception. In this state, the UE consumes the least amount of energy. RRC's Connected mode is further divided in four states:

- CELL_DCH - a state where a dedicated connection exists in UL and DL direction. Radio resources are dedicated exclusively to the UE allowing it to send and receive data at higher speeds;
- CELL_FACH - there aren't any dedicated connections but data can be transferred via common channels. This state is suitable for transfer of small amount or bursty data. This state preserves the use of radio resources in the cell;
- CELL_PCH - similarly to Idle state the UE monitors only the paging and broadcast channels. The difference is that the logical RRC connection still exists;
- URA_PCH - a state similar to CELL_PCH where every cell change does not trigger a cell update procedure in order to decrease the signaling activity.

In UMTS the concept of connection is separated from the concept of Radio Bearer (RB). When an idle UE wants to make a data call it needs to establish a connection and obtain communication resources. The UE first initiates establishment of a RRC connection and then the network creates one or more RBs depending on the requested and available resources. There can be only one RRC connection per data call or per UE but many RBs within one connection. The RB defines the properties of the connection depending on the requested QoS parameters. For instance, to transfer low volume data the UE will obtain a common physical channel (CELL_FACH state) and a dedicated physical channel (CELL_DCH state) for a higher volume, delay-restricted data. The network then revokes allocated resources after an inactivity timeout t_L in CELL_FACH state or t_H in CELL_DCH state (3GPP, 2002; Korhonen, 2003).

The RRC mechanism, as described, is vulnerable to attacks triggering an excessive number of transitions between states. A single user request for connection/resources triggers multiple signaling messages that are transferred in the access and core part of the network. If requests are repeated regularly by many malicious UEs, the network will overload.

In particular, we can distinguish between two different types of signaling attacks:

- FACH attacks. A FACH attack occurs when the attacker makes a low bandwidth request in repetitive intervals. This attack triggers signaling messages by transitioning between CELL_PCH and CELL_FACH states or between Idle and CELL_FACH states.
- DCH attacks. The attacker performs a DCH attack with repetitive high bandwidth requests. This type of attack generates signaling traffic by alternating between CELL_DCH and CELL_FACH, CELL_PCH or Idle states.

The most common signaling attacks are CELL_PCH state triggered FACH attacks and CELL_FACH state triggered DCH attacks. Excessive signaling has negative influence also on the user experience because of high power consumption in the UE.

4 SYSTEM MODEL

The model used in this research is based on conventional stochastic modeling techniques (Gelenbe, 1979) and focuses on a single user's RRC part of the UMTS system. It is described by the state diagram on Figure 1. The figure depicts a model derived from the conventional UMTS model with added 'attack' states in the system. The idle state is represented by D - Dormant. CELL_PCH and URA_PCH are represented by a single P state. L (the low state) represents CELL_FACH and H (the high state) represents the CELL_DCH state. The corresponding states when attacks occur, or the attack states, are denoted with subscripts L_A and H_A for allocated FACH and DCH channels because of an attack.

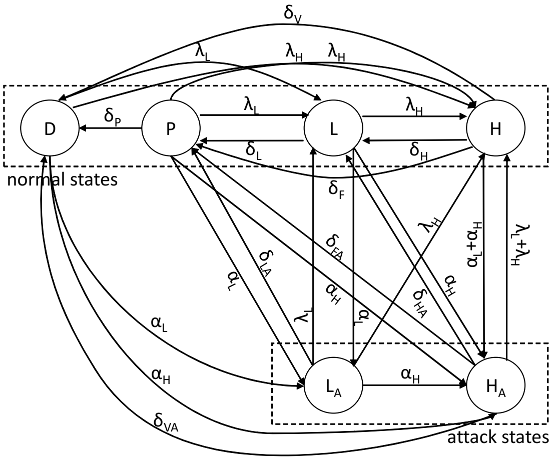

Figure 1: RRC model of UMTS under signaling attack

At any given time and state, the system may receive one of the following four requests triggered by the UE: normal FACH, normal DCH, attack FACH and attack DCH request which trigger the promotion transitions in the system. Namely, λ_i denoting the rate of normal and α_i the rate of attack requests for state i, where $i \in \{L, H\}$. We define the *attack ratio* parameter k as

$$k = \frac{\alpha_L}{\lambda_L} = \frac{\alpha_H}{\lambda_H}, \tag{1}$$

which gives the impact of the attack. State demotion rates from normal states are denoted by $\delta_P = \frac{1}{t_P}$, $\delta_L = \frac{1}{t_{FACH}+t_L}$ and $\delta_H = \delta_F = \delta_V = \frac{1}{t_{DCH}+t_H}$, where t_{FACH} and t_{DCH} represent the average duration of data transmission in the respective states while t_P is the inactivity timeout period in CELL_PCH state. Transitions denoted by δ_F and δ_V represent the *fast dormancy* mechanisms which were introduced in later versions of UMTS standards.

During a signaling attack, the attacker usually does not transmit any data because the purpose of the attack is solely to trigger the signaling transitions. Therefore, the demotion rates from the attack states are selected as $\delta_{LA} = \frac{1}{t_L}$ and $\delta_{HA} = \delta_{FA} = \delta_{VA} = \frac{1}{t_H}$. Two specific cases are included when low-bandwidth (FACH) requests are served in dedicated channel states, represented by the transitions from H to H_A and vice-versa.

5 SYSTEM OPTIMISATION

The following section describes a probabilistic approach towards minimizing the impact of signaling attacks on the system. Denoting the *probability of state i* with P_i we can describe the system model with the following linear equations:

$$\begin{aligned}
P_D(\lambda_L + \lambda_H + \alpha_L + \alpha_H) &= \\
= P_P\delta_P + P_H\delta_V + P_{HA}\delta_{VA}&, \\
P_P(\lambda_L + \lambda_H + \alpha_L + \alpha_H + \delta_P) &= \\
= P_L\delta_L + P_H\delta_F + P_{LA}\delta_{LA} + P_{HA}\delta_{FA}&, \\
P_L(\lambda_H + \alpha_L + \alpha_H + \delta_L) &= \\
= (P_D + P_P + P_{LA})\lambda_L + P_H\delta_H + P_{HA}\delta_{HA}&, \\
P_H(\delta_V + \delta_H + \alpha_L + \alpha_H + \delta_F) &= \\
= (P_D + P_P + P_L + P_{LA})\lambda_H + P_{HA}(\lambda_H + \lambda_L)&, \\
P_{LA}(\lambda_L + \lambda_H + \alpha_H + \delta_{LA}) &= \\
= (P_D + P_P + P_L)\alpha_L&, \\
P_{HA}(\lambda_L + \lambda_H + \delta_{HA} + \delta_{FA} + \delta_{VA}) &= \\
= (P_D + P_P + P_L + P_{LA})\alpha_H + P_H(\alpha_L + \alpha_H)&.
\end{aligned} \tag{2}$$

The probabilities of each state can be found by solving this system of equations with the normalization condition $\sum_i P_i = 1$.

The goal of our optimisation will then be to minimise the time spent in the attack states, i.e. to minimise P_{LA} and P_{HA} and maximise the time spent in normal states, and is reflected by the cost function C:

$$C = \frac{P_{LA} + P_{HA}}{P_L + P_H} \tag{3}$$

which needs to be minimised.

To minimise C the influence of the following two parameters is inspected:

- inactivity timers t_L and t_H;
- call setup delay in promotion transitions to FACH and DCH, denoted with t_{xL} and t_{xH} respectively.

Changing the inactivity timers is fairly straightforward, while inserting delay in call setup should be looked at from the system's perspective.

Denote by $\theta_i = \lambda_i + \alpha_i$ the total rate of requests for state i seen by the system, where $i \in \{L, H\}$. The average interval between requests is $t_{\theta_i} = \frac{1}{\theta_i}$. We insert a setup delay at transitions to state i and get $t'_{\theta_i} = t_{\theta_i} + t_{xi}$, where t'_{θ_i} is the new interval between requests. Solving for the new arrival rates we get:

$$\lambda'_i = \frac{\lambda_i}{1 + t_{xi}\lambda_i(k+1)} \quad \alpha'_i = \frac{k\alpha_i}{k + t_{xi}\alpha_i(k+1)} \tag{4}$$

which represent the "delayed" normal and attack rates at state i.

The partial derivative of $C(t_i, t_{xi})$ with respect to the inspected parameters does not lead to a closed form solution. To find the numerical solution for the problem we make the following assumptions:

- the arrival rates of normal requests are set to $\lambda_L = \lambda_H = 0.2$ [req/sec], as a typical value in real-world networks (Inc., 2013);
- the average duration of data transmission in FACH and DCH states is $t_{FACH} = t_{DCH} = 1.28\,\text{sec}$ which is calculated as the time to transmit an average 320 KB web page in DCH channel at 2 Mbps (Ramachandran, 2010) or 512 B (upper limit of FACH data size) of background data at 3.2 kbps in FACH channel;
- the inactivity timers are set to $t_P = 20\,\text{min}$, $t_L = 4\,\text{sec}$ and $t_H = 6\,\text{sec}$ as usually used values by mobile operators;

- the attack ratio is set to $k = 1$. Note that k is only a measure of attack strength, and as so it does not influence the form of C, only its amplitude.

We also assume that the fast dormancy mechanism is not implemented ($\delta_F = \delta_V = \delta_{FA} = \delta_{VA} = 0$) and we inspect the system under both FACH and DCH attacks. The next section presents and discusses the obtained results.

6 RESULTS AND DISCUSSION

First, we investigate the influence of inactivity timers in FACH and DCH states on the security of the system. Three scenarios are inspected for both FACH and DCH types of attacks: t_L and t_H are changed together; the timer in DCH is fixed to 6s and we change the timer in FACH; the timer in FACH is fixed to 4s and we change the timer in DCH only. Then we insert delay t_{xL} and t_{xH} in setup rates and inspect three similar scenarios: inserting delay in both FACH and DCH requests; inserting delay only in FACH requests; and inserting delay only in DCH requests.

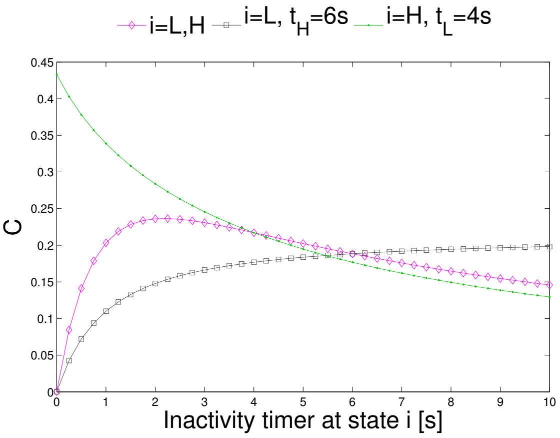

Figure 2: C as a function of inactivity timeout period at state i for FACH attacks

Figure 2 depicts the three scenarios of inactivity timers in a system under FACH attack. For fixed $t_L = 4$ sec the cost function decreases with the increase of t_H. This simply shows that the longer the system stays in H state the lower the probability of attack on FACH state. For fixed $t_H = 6$ sec the cost function increases with the increase of t_L meaning that the quicker the system returns to normal state, the lower probability of attack. The cost function for changing both t_L and t_H together rises to a certain point after which it starts to decline. Of course, the cost function has a minimum at $t_L = t_H = 0$ but selecting low values for the timeout periods would mean larger number of transitions (attacks) although the time spent in attacking states is minimised. Therefore a better choice is selecting higher values for the two timers.

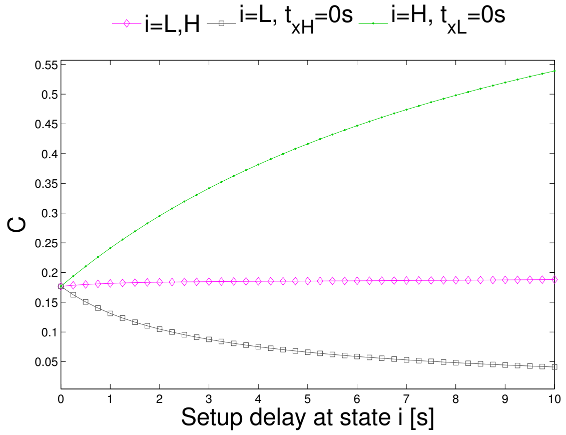

Figure 3: C as a function of setup delay at state i for FACH attacks

Figure 3 shows the influence of inserting delay in state promotion transitions in system under FACH attack. Setting $t_{xH} = 0$ and increasing t_{xL} is a good choice for lowering the attack. In contrast to that, increasing the delay of DCH requests while an attack is ongoing on FACH state sharply increases the probability of attack states. Increasing the delay in both FACH and DCH requests at the same time does not introduce any improvements.

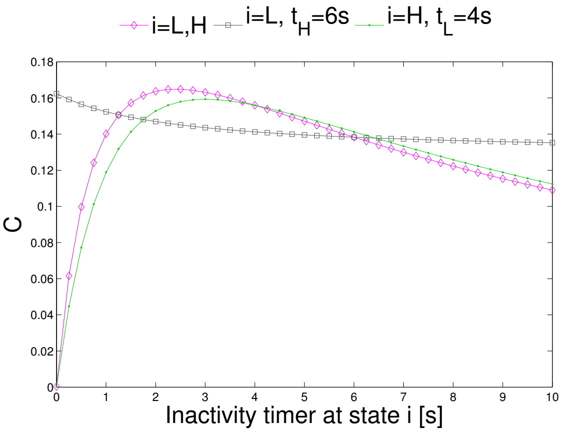

Figure 4: C as a function of inactivity timeout period at state i for DCH attacks

In case of DCH attacks, Figure 4 suggests that increasing the inactivity timeout in FACH state t_L introduces small improvements, analogous to the case of increasing the DCH timer for FACH attack. Changing both timers at the same time is similar to the case of changing t_L and t_H timers together under FACH attack (Figure 2) - the cost function has a minimum value for $t_H = 0$ but selecting higher values for t_H is a safer choice. Although we would expect constant increase in C with the increase of t_H when t_L is fixed to 4s, C drops after a certain point. This is due to normal FACH requests being served in high bandwidth channels, thus the transition H_A to H being more probable than H to H_A.

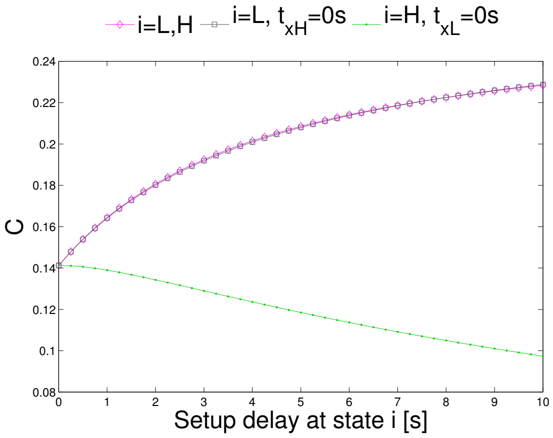

Figure 5: C as a function of setup delay at state i for DCH attacks

Finally, Figure 5 depicts the case of inserting delay in promotion transitions in a system under DCH attack. Analog to the case of FACH attack, inserting delay in DCH requests in this case lowers the cost function. Inserting delay in FACH requests increases the probability of attack states, same as inserting delay at FACH and DCH requests at the same time. The increase in C at this case is a result of lowering the probability of L state which is in the denominator of function C.

In general, the cost function for changing the inactivity timer of FACH state under FACH attack or the inactivity timer of DCH state under DCH attack has a minimum at $t_L = 0$ and $t_H = 0$ respectively. This result is correct by means of lowering the probability of attack state. Anyway, selecting small timers in both cases means larger number of transitions, i.e. larger number of attacks. In case of FACH attack, setting the inactivity timer of DCH state to higher values is a good choice. Similarly, selecting higher values for the inactivity timer in FACH state slightly improves the security of the system under DCH attack. Adding delay in setup transitions for FACH state in a system under FACH attack, or setup transitions for DCH state in a system under DCH attack provides good results by lowering the probability of attack states. The other cases of adding delay in setup transitions have negative influence.

7 CONCLUSION

The increasing use of smartphone applications has created new security issues for mobile cellular networks. In order to provide better user experience and real-time services, mobile applications are usually developed assuming they have an “always on” connectivity to the Internet. However mobile networks such as UMTS are originally designed for voice calls and browsing type of data traffic, and do not provide a continuous access to the network. Connections are created and teared down dynamically by demand. This “creation and tear down” characteristic, which is also meant to save bandwidth capacity for the network as a whole, introduces some interesting liabilities that can be easily exploited for signaling attacks.

In this paper we have analysed the influence of parameters such as the inactivity timeouts and call setup delay on the impact of signaling attacks in UMTS networks. We proposed a model of UMTS Radio Resource Control mechanism under attack and defined a cost function based on the probability of attack states in the model. Results show that in system under FACH/DCH attack it is a good choice to extend the duration of inactivity timers and to add delay in requests for the corresponding states.

Future work may include optimising the system in terms of transitions’ probability, analysis in a simulation environment as well as obtaining new mechanisms for mitigation of attacks. It would also be of interest to evaluate how such signaling attacks may actually affect a realistic security setting, for instance when spectators at a sports or cultural venue have to be evacuated rapidly with the help of instructions distributed through smartphones because of an emergency (Filippoupolitis and Gelenbe, 2009; Dimakis et al., 2010; Görbil and Gelenbe, 2013) and a signaling attack is simultaneously launched by malicious individuals who wish to further disrupt the emergency situation.

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

3GPP (June 2002). Utran functions, examples on signaling procedures (release 1999). TR 25.931 v3.7.0.

Abdelrahman, O. H. and Gelenbe, E. (2014). Signalling storms in 3G mobile networks. In *IEEE International Conference on Communications (ICC'14), Communication and Information Systems Security Symposium*, Sydney, Australia. Accepted for publication.

Cao, Y. and Gelenbe, E. (1998). Autonomous search for mines. *European Journal of Operational Research*, 108(2):319–333.

Choi, Y., Yoon, C.-h., Kim, Y.-s., Heo, S., and Silvester, J. (2014). The impact of application signaling traffic on public land mobile networks. *Communications Magazine, IEEE*, 52(1):166–172.

Dimakis, N., Filippoupolitis, A., and Gelenbe, E. (2010). Distributed building evacuation simulator for smart emergency management. *Comput. J.*, 53(9):1384–1400.

Filippoupolitis, A. and Gelenbe, E. (2009). A distributed decision support system for building evacuation. In *Human System Interactions, 2009. HSI'09. 2nd Conference on*, pages 323–330.

Gelenbe, E. (1979). Probabilistic models of computer systems. *Acta Inf.*, 12:285–303.

Gelenbe, E., Gorbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., and Lyberopoulos, G. (2013a). Security for smart mobile networks: The NEMESYS approach. In *Proceedings of the 2013 IEEE Global High Tech Congress on Electronics (GHTCE'13)*.

Gelenbe, E., Görbil, G., Tzovaras, D., Liebergeld, S., Garcia, D., Baltatu, M., and Lyberopoulos, G. L. (2013b). Nemesys: Enhanced network security for seamless service provisioning in the smart mobile ecosystem. In Gelenbe, E. and Lent, R., editors, *ISCIS*, volume 264 of *Lecture Notes in Electrical Engineering*, pages 369–378. Springer.

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

Gelenbe, E., Schmajuk, N., Staddon, J., and Reif, J. (1997). Autonomous search by robots and animals: A survey. *Robotics and Autonomous Systems*, 22(1):23–34.

Gelenbe, E. and Wu, F.-J. (2012). Large scale simulation for human evacuation and rescue. *Computers & Mathematics with Applications*, 64(12):3869–3880.

Görbil, G. and Gelenbe, E. (2013). Disruption tolerant communications for large scale emergency evacuation. In *PerCom Workshops*, pages 540–546. IEEE.

GSMA, S. A. P. (Apr 2012). Smart apps for smarter phones.

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.

Inc., S. (2013). Charting the signaling storms.

Kambourakis, G., Kolias, C., Gritzalis, S., and Park, J. H. (2011). Dos attacks exploiting signaling in umts and ims. *Comput. Commun.*, 34(3):226–235.

Korhonen, J. (2003). Introduction to 3g mobile communications.

Lee, P., Bu, T., and Woo, T. (2007). On the detection of signaling dos attacks on 3g wireless networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 1289–1297.

Marck, P. (2013). Iab - focus on mobile.

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.

Wang, H., Zhang, D., and Shin, K. (2004). Change-point monitoring for the detection of dos attacks. *Dependable and Secure Computing, IEEE Transactions on*, 1(4):193–208.

Wireless, R. (2012). Docomo demands google's help with signalling storm.

Wu, Z., Zhou, X., and Yang, F. (2010). Defending against dos attacks on 3g cellular networks via randomization method. In *Educational and Information Technology (ICEIT), 2010 International Conference on*, volume 1, pages V1–504–V1–508.

Yu, C.-M., Ni, G.-K., Chen, I.-Y., Gelenbe, E., and Kuo, S.-Y. (2014). Top-k query result completeness verification in tiered sensor networks. *IEEE Transactions on Information Forensics and Security*, 9(1):109–124.