IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 1 September 2014; revised 8 December 2014; accepted 30 December 2014.
Date of publication 6 January 2015; date of current version 26 February 2016.

Digital Object Identifier 10.1109/TETC.2015.2389662

Modeling and Analysis of RRC-Based Signalling Storms in 3G Networks

GOKCE GORBIL, OMER H. ABDELRAHMAN, (Member, IEEE), MIHAJLO PAVLOSKI, AND EROL GELENBE, (Fellow, IEEE)

Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K.
CORRESPONDING AUTHOR: G. GORBIL (g.gorbil@imperial.ac.uk)

This work was supported by the European Union FP7 Programme through the Research Project entitled Enhanced Network Security for Seamless Service Provisioning in the Smart Mobile Ecosystem under Grant 317888 within the FP7-ICT-2011.1.4 Trustworthy ICT domain.

ABSTRACT Mobile networks are vulnerable to signaling attacks and storms that are caused by traffic patterns that overload the control plane, and differ from distributed denial of service attacks in the Internet since they directly affect the control plane, and also reserve wireless bandwidth and network resources without actually using them. Such storms can result from malware and mobile botnets, as well as from poorly designed applications, and can cause service outages in 3G and 4G networks, which have been experienced by mobile operators. Since the radio resource control (RRC) protocol in the 3G and 4G networks is particularly susceptible to such storms, we analyze their effect with a mathematical model that helps to predict the congestion that is caused by a storm. A detailed simulation model of a mobile network is used to better understand the temporal dynamics of user behavior and signaling in the network and to show how RRC-based signaling attacks and storms cause significant problems in both the control and user planes of the network. Our analysis also serves to identify how storms can be detected, and to propose how system parameters can be chosen to mitigate their effect.

INDEX TERMS Network attacks, malware, app malfunctions, UMTS networks, 3G, 4G, radio resource control, signalling overload, performance analysis, simulation.

I. INTRODUCTION

Smart devices have not gone unnoticed by cyber-criminals, who have started to target mobile platforms [1], [2], and mobile subscribers and mobile network operators (MNOs) face new security challenges [3], including the identification and mitigation of *signalling attacks and storms*, which overload the control plane through traffic that causes excessive signalling in the network. The susceptibility of mobile networks to such attacks has been identified [4]–[9], and they have now become a reality that MNOs have to face regularly due to side effects of mobile malware, subscribers with high frequency communication sessions [10], poorly designed mobile applications [11], [12] and unwanted traffic from Internet hosts outside the mobile network [13], [14].

While malware and network attacks are common in the Internet, they have not been prevalent in mobile networks until recent times. However, they are quickly becoming a major security concern due to the advent of smart mobile devices and the increasing capacity and use of mobile networks for Internet access [15], [16]. The increasing number of mobile malware and infected devices, together with changing mobile access patterns of users, can create signalling anomalies and overloads, either due to deliberate malicious activity or as a side-effect. Thus signalling attacks and storms are indeed an emerging cyber-security threat in mobile networks, which are a major component of our cyber infrastructure. Smart mobile devices are also increasingly used in emergency management systems, especially in urban environments [17]–[19]. Thus they are likely to be targeted in conjunction with other physical or cyber attacks in order to further compromise the safety and confidentiality of civilians and emergency responders [20], [21].

MNOs have a strong incentive to safeguard mobile users from service outages and degradations due to signalling attacks and storms, and to protect their mobile network infrastructure, market reputation and revenue [3], [22]. It is therefore important to identify how signalling storms are generated, analyze their effect on network performance,

This work is licensed under a Creative Commons Attribution 3.0 License. For more information,
see http://creativecommons.org/licenses/by/3.0/

and develop detection and mitigation methods in this new and dynamic playground of smart devices and new generation mobile networks centered around data services. As we look at the future, we can expect that UMTS and LTE networks will also support major machine-to-machine communications [23] where the human being is not in the loop to identify and remediate against an apparent storm. In the first instance, we can expect that UMTS will have to be secured against such storms and into the future that LTE should be an increasing object of studies to detect and mitigate against signalling storms and attacks [24]–[26].

In our previous work [27], we identified the radio resource control (RRC) protocol of UMTS and LTE networks [28], [29] to be particularly susceptible to creating signalling attacks and storms. In [27], we developed a probability model [30] of signalling state transitions for a single UMTS user, from which we derived analytical results regarding the user's behavior when her device generates user traffic that causes a signalling storm and the impact it has on the network. In the work presented here, we expand upon our earlier work and improve our mathematical model by introducing the effect of congestion in the control-plane. We also design and develop a mobile network simulator that is significantly more complex and realistic than our mathematical model, and present results from large-scale simulation experiments that enable us to better understand the temporal dynamics of user behavior and signalling, and to validate our analytical results. Based on the insights that we gain, we discuss how certain network parameters can help to mitigate against signalling storms, and how signalling storms can be detected.

II. SIGNALLING ATTACKS AND STORMS

Signalling Attacks are caused by traffic patterns that generate excessive signalling in the control plane of mobile networks, and can be launched easily without modification or compromise of the radio or networking stack of mobile devices by generating low volumes of carefully timed user plane traffic. Signalling attacks are in essence distributed denial-of-service (DDoS) attacks [31], but are different than DDoS attacks in the Internet since they directly target the control plane of mobile networks without necessarily generating a high traffic volume at the user plane. RRC-based signalling attacks are further troublesome since they reserve radio resources without actually using them, thereby wasting radio resources.

In this paper, we assume that signalling attacks are due to deliberate malicious activity that aims to disrupt mobile services, as opposed to signalling storms which are discussed below. While we are not aware of any deliberate signalling attacks in operational mobile networks up to now, we should not carelessly dismiss the potential for such attacks since all the ingredients for their realization are already available. For example, the mobile world witnessed its first botnet in 2012 [32], which can be leveraged to launch different types of signalling attacks [33], in addition to other types of malicious activities [34]. Furthermore, there are methods available to an attacker that can be used to improve the efficiency of the attack. For example, the attacker can actively probe the network in order to infer the network's parameters [35]–[37], and also identify IP addresses at specific locations within the network [38]. Indeed, a review of 180 MNOs showed that 51% of them allow mobile devices to be probed from the Internet, by either assigning them public IP addresses, allowing IP spoofing, or permitting mobile-to-mobile probing within the network [38], [39]. Similar attacks can also be launched via compromised femtocells [40], which can further be used to infect other femtocells via Internet-based connections not controlled by the MNO, and thus increase the intensity of the attack.

Signalling Storms are similar to signalling attacks, but they are mainly due to poorly designed or misbehaving mobile applications that frequently establish and tear-down data connections in order to transfer small amounts of data. Many mobile applications are designed and developed by software companies who mainly have an Internet background and thus are not familiar with the control plane of mobile networks. They therefore assume that connectivity is a given and design their applications without taking into account the specifics of mobile networks. This phenomenon was studied early in [41], where a small number of mobile devices were observed to generate an unproportionately high number of PDP context activations and deactivations due to poorly designed application layer software. A good recent example that shows that this trend is still continuing despite earlier work is the case of an Android VoIP application popular in Japan, which used frequent keep-alive messages even when the users were idle, causing a signalling overload and a major outage in the mobile network [42]. In a similar incident, the launch of the free version of the Angry Birds application on Android caused excessive signalling load due to the frequent communications generated by the in-game advertisements [43]. Such problems have prompted the mobile network industry to promote best practices for developing network-friendly applications [11], [12].

Unexpected events in the Internet may also cause signalling storms in mobile networks. For example, an important feature of smartphones is the ability to receive push notifications from cloud services in order to notify the user of an incoming message or VoIP call, which is enabled by having the mobile device send periodic keep-alive messages to a cloud server, typically with a period of five minutes. If the cloud service becomes unavailable, then the mobile device may use a much shorter period, generating significantly higher signalling load. Such incidents have been reported and analyzed in [44] and [45] with outages in Skype and Google's cloud service, respectively.

Signalling storms could also result as a side effect of large-scale malware infections which target the user rather than the network, but generate excessive signalling as a by-product of malicious activity. Examples of malware that would cause signalling storms if many users are infected are

SMS/email spammers, adware, premium service abusers and botclients. All of these malware generate frequent but small amounts of data, requiring repeated signalling to allocate and deallocate radio channels and other resources, and therefore have a negative impact on the control plane of the network. Unfortunately, such malware are among the top threats currently encountered on smart devices [1], [46], [47].

Recent incidents such as the ones described here show that the threat of signalling attacks and storms is very real and that they have the potential to cause major outages in mobile networks. Unlike flash crowds which last for a short time during special occasions such as New Year's Eve, signalling attacks and storms are unpredictable and they persist until the underlying problem is identified and resolved by the MNO. Considering their impact on the availability and security of mobile networks, it is evident that MNOs have a strong incentive to safeguard their users from malware and to proactively detect and mitigate signalling attacks and storms in order to protect their infrastructure and services. Although in principle some of these attacks can be mitigated by smart routing [48] inside the core network, such facilities are currently not available. We also believe that as MNOs progressively take on the role of Internet service provider with 4G networks, we will witness signalling-based DDoS attacks in mobile networks more frequently, and therefore we should be proactive in their analysis and mitigation.

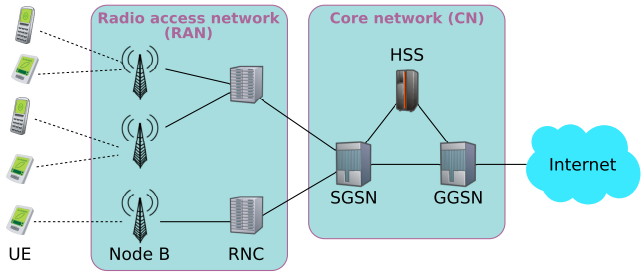


FIGURE 1. The basic architecture of a UMTS network. The user equipment (UEs), e.g., smartphones, are connected to the mobile network via the base stations (Node-Bs), which maintain the radio channels with the UEs. The radio network controller (RNC) manages the radio resources and the Node-Bs in the access network.

III. THE RADIO RESOURCE CONTROL PROTOCOL

In UMTS networks, the radio resource control (RRC) protocol is used to manage resources in the radio access network (RAN) [28]. It operates between the UMTS terminals, i.e., the user equipment (UE), and the radio network controller (RNC). Figure 1 shows the basic architecture of a UMTS network, depicting the RAN and the core network (CN) elements comprising the packet-switched domain of the mobile network. The RNC is the switching and controlling network element in the RAN, and performs radio resource management (RRM) functions in order to guarantee the stability of the radio path and the QoS of radio connections by efficient sharing and management of radio resources. The RRC protocol is utilized for all RRM-related control functions such as the setup, configuration, maintenance and release of radio bearers between the UE and the RNC. The RRC protocol also carries all non-access stratum signalling between the UE and the CN.

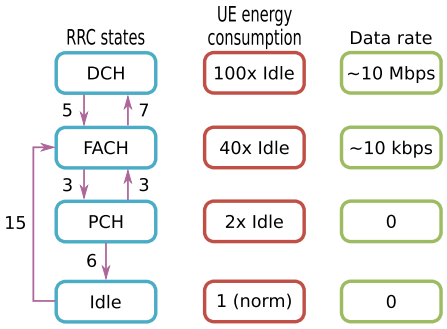


FIGURE 2. RRC states in UMTS. The figure on the left shows the typical number of signalling messages exchanged within the RAN for each transition. The other figures show the approximate energy consumption and maximum data rate at the UE.

In order to manage the radio resources, the RRC protocol associates a *state machine* to each UE, which is maintained synchronized at the UE and the RNC via RRC signalling messages. The RNC controls the transitions between the RRC states based on information it receives from the UEs and the Node-Bs on available radio resources, conditions of the currently used radio bearers, and requests for communication activity. As shown in Fig. 2, there are typically four RRC states, given in order of increasing energy consumption and data rate: *idle, cell-PCH, cell-FACH* and *cell-DCH*. In the rest of this paper, we refer to state *cell-X* simply as *X*. Whenever the UE is not in the idle state, it is in *connected mode* and has a signalling connection with the RNC. In connected mode, the location of the UE is known by the RNC at the level of a single cell, which is maintained by *cell updates* sent by the UE either periodically or when it changes cells. We describe the RRC states in more detail below.

Idle: This is the initial state when the UE is turned on. In this state, the UE does not have a signalling connection with the RNC, and therefore the RNC does not know the location of the UE. Its location is known by the CN at the accuracy of the location area or routing area, which is based on the latest mobility signalling the UE performed with the CN. Any downlink activity destined for a UE in idle mode will require *paging* in order to locate the UE at the cell level. Since the UE does not have an RNC connection, it cannot send any signalling or data until an RNC connection has been established.

FACH: The UE is in connected mode, and the radio connection between the UE and the RNC uses only common channels which allow low-rate data transmission.

DCH: The UE is in connected mode, and the radio connection uses resources dedicated to the UE. While in DCH, the UE may use shared channels, dedicated channels or both. The data rate of the connection is significantly higher than the FACH state, but energy use is also higher.

PCH: This is a low-energy state that allows the UE to maintain its RNC connection and thus stay in connected mode, but it cannot send or receive any traffic while in this state. While in PCH, the UE listens to paging occasions on the paging channel. This state is optional and it can be enabled or disabled by the MNO according to their policies. Although the PCH state is a low-energy state, the UE still consumes more power than in the idle state. Therefore, some MNOs choose to disable the PCH state in order to allow the UE to return to idle mode quickly and thus reduce its energy consumption. We will investigate the effect of the PCH state on signalling load in Sec. VII.

State demotions from a higher to a lower state, e.g., DCH→FACH, occur based on radio bearer inactivity timers at the RNC. The exact order of state demotions is dependent on MNO policy, but a progression as shown in Fig. 2 is common, although some MNOs skip the FACH and/or PCH states. State promotions from the idle and PCH states occur depending on uplink and downlink activity. For example, when the UE has uplink data to send, it sends an *RNC connection request* if in idle, or a *cell update* if in PCH, to the RNC in order to move to a state where it can send and receive data. Whether the UE is promoted to the FACH or DCH state is dependent on MNO policy. A FACH→DCH transition is performed based on buffer occupancy of the uplink and downlink radio links as observed by the RNC.

TABLE 1. RRC state transitions, number of signalling messages exchanged, and related parameters.

Transition	Triggering event	r_{xy}	c_{xy}
Idle→FACH	Uplink or downlink traffic	15	5
PCH→FACH	Uplink or downlink traffic	3	-
FACH→DCH	Radio link buffer threshold (Θ) reached, $\Theta = 1500$ B	7	-
DCH→FACH	Expiry of inactivity timer $T_1 = 6$s	5	-
FACH→Idle	Expiry of inactivity timer $T_2 = 12$s, PCH disabled	5	3
FACH→PCH	Expiry of inactivity timer $T_2 = 4$s, PCH enabled	3	-
PCH→Idle	Expiry of inactivity timer $T_3 = 20$min, PCH enabled	6	3

Table 1 summarizes when RRC state transitions occur and the number of signalling messages exchanged to effect each transition. In our simulations, we assume the RRC state progression given in Fig. 2; whether the UE goes from FACH to PCH, or to idle, depends on whether the PCH state is enabled. For an $x \rightarrow y$ transition, we use r_{xy} and c_{xy} to denote the number of signalling messages exchanged within the RAN and between the RAN and the CN, respectively.

The RRC protocol was designed to manage the limited radio resources among multiple UEs and to decrease energy use at the UE. It is therefore biased towards demoting the UE to a lower state as soon as possible, especially if the UE is in the DCH or FACH state. Indeed, as the number of smartphones accessing UMTS networks has increased, the industry has introduced improvements and changes in order to get more data rate out of limited radio resources, such as HSDPA and HSUPA, and to improve the energy use of smartphones. For example, fast dormancy enables the UE to indicate to the RNC when it has no more uplink data to send for a speedier demotion to the PCH or idle state. In addition, some MNOs choose to disable the PCH state in order to allow the UE to return to idle mode quickly and thus reduce its energy consumption. As we will discuss in Sec. VII, this tendency to perform hasty RRC demotions result in excessive signalling load in the mobile network, especially in the case of signalling attacks and storms.

The RNC will customarily release radio resources for a UE soon after activity ceases in its channel, making those resources available for other UEs. Thus, it uses short inactivity timers, which are in the order of 2–10 seconds (Table 1). These short timers make the RRC protocol susceptible to signalling attacks, as an attacker that approximately determines the values of the T_1 and T_2 timers can then launch a devastating attack from a relatively small number of compromised UEs, as we discuss in Sec. VII. In addition, when combined with the chatty nature of many mobile applications and with emerging mobile trends such as buffering streaming traffic in order to save device energy [49], the tendency to deallocate radio channels quickly necessarily leads to increased RRC signalling in order to reconfigure or setup channels that were released a short time ago, rendering the mobile network vulnerable to RRC-based signalling storms.

We thus focus on the RRC protocol in order to better understand its signalling behavior, and investigate under which conditions signalling load becomes excessive. In the next section, we present a mathematical model of the signalling behavior of a single UE that includes congestion effects in the control plane, and later derive analytical results from it. Section V describes our simulation model of UMTS networks. In Sec. VI, we describe our experimental setup, and discuss our findings on the effect of signalling attacks targeting the RRC protocol in Sec. VII. We discuss related work in Sec. VIII and present a summary of our findings and future work in Sec. IX.

IV. MODELING SIGNALLING BEHAVIOR OF THE UE

Analytical models [50] are a useful way to gain insight into the main performance interactions within a telecommunications system. Thus we will first review the work in [27] for a *single* UE's signalling behavior which focuses on the potential of causing signalling storms. We then extend the analysis to include the effect of congestion which limits the signalling load that a set of misbehaving UEs can impose on the network during a storm.

Consider a UE which generates both normal and malicious connections, and suppose that its RRC state machine is described by Fig. 2. We will represent the state evolution of the UE by a Markov model, presented in Fig. 3, whereby future behavior (residual time in current state and next state) depends only on current state and not on past behavior. Our motivation behind the choice of this modelling

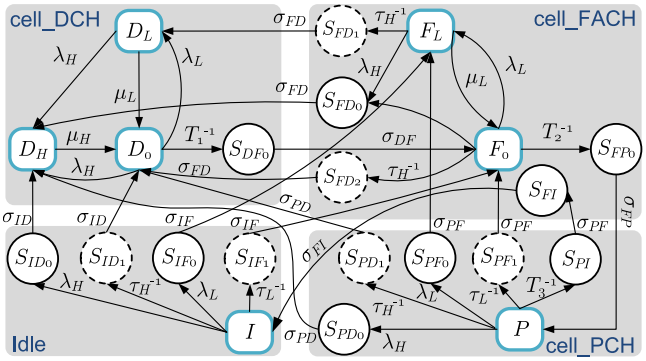


FIGURE 3. Markov model of the signalling behavior of the UE. Up-transitions are caused by either low data rate (*L*) or high data rate (*H*) traffic, while down-transitions are due to timeouts. The model includes the main RRC states, shown as rounded rectangles, as well as intermediate states, shown as circles, some of which represent states where the UE is waiting for a response to a state transition request. The continuous and broken circles represent intermediate states due to normal and malicious traffic, respectively.

approach is that it provides a balance between capturing the interactions between user traffic and the RRC protocol and maintaining analytical tractability, and it can also be extended to a population of users without much technical difficulty. Let λ_L and λ_H be the rates at which low and high data rate connections are *normally* made, and μ_L and μ_H be the rates at which these connections terminate. High bandwidth connections include video streaming, web browsing, VoIP and voice calls, while low bandwidth connections represent small data transfers such as keep-alive messages and location updates. We denote by F_L the state when the UE is using the bandwidth of FACH, and by D_L and D_H the states when low and high rate requests are handled while the UE is in DCH. Since the amount of traffic exchanged in states F_L and D_L is usually very small, we assume that their durations are independent but stochastically identical. At the end of normal usage, the UE transitions from F_L to F_0 or from $\{D_H, D_L\}$ to D_0, where F_0 and D_0 are, respectively, the states when the UE is inactive in FACH and DCH, and before the timers T_2 and T_1 expire. If the UE does not start a new session for some time, it will be demoted from D_0 to F_0, and from F_0 to P, and will then return from P to I (i.e., PCH $\rightarrow$ Idle) when inactivity timer T_3 expires. Since the UE is not able to communicate in P, the transition $P \rightarrow I$ is performed by having the UE first move to FACH, release all signalling connections, and finally move to I.

The attacking or misbehaving connections falsely cause unnecessary up-transitions while the user does not really need to move to a bandwidth using state (F_L, D_L, or D_H), and therefore the UE is soon demoted to a lower state due to inactivity, unless the user starts a new data session before the timeout. Consequently, the attack results in the usage of network resources both by the computation, state transitions and exchange of control messages that occur for session handling, and through bandwidth reservation that remains unutilised.

To perform a signalling attack, the attacker would need to infer the radio network configuration parameters (i.e., the T_i timers and the radio link buffer threshold Θ), and also monitor the user's activity in order to estimate when a transition occurs so as to trigger a new one immediately afterwards. Naturally there will be an error between the actual transition time and the estimated one, and we denote the expected value of the difference between the two time instants by τ_L and τ_H for malicious transitions to FACH and DCH, respectively. In a similar manner, if the storm is caused by a misbehaving mobile application, then τ_L, τ_H represent the level of synchronization between the misbehaving traffic bursts and the UE's state changes; for instance $\tau_H = 0$ indicates the extreme case where a high data rate burst is sent immediately after a demotion from DCH.

Let σ_{xy}^{-1} be the average time needed to establish and/or release network resources during state promotion or demotion $x \rightarrow y$, and S_{xy} be the corresponding state when the UE is waiting in state x for the transition to complete. Note that this overhead is incurred only when the UE moves from one RRC state to another, while changes within the same RRC state (e.g., from inactive to active) occur instantaneously and are seamless to the UE. Denote by π_x the stationary probability that the UE is in state x, and let $\Lambda_H = \lambda_H + \tau_H^{-1}$, $\Lambda_L = \lambda_L + \tau_L^{-1}$, then the state transition model can be described by a set of linear equations:

$$\begin{aligned}
\pi_I[\Lambda_H + \Lambda_L] &= \pi_P T_3^{-1},\\
\pi_P[\Lambda_H + \Lambda_L + T_3^{-1}] &= \pi_{F_0} T_2^{-1},\\
\pi_{F_0}[\Lambda_H + \lambda_L + T_2^{-1}] &= \pi_{F_L}\mu_L + \pi_{D_0} T_1^{-1},\\
\pi_{F_L}[\Lambda_H + \mu_L] &= [\pi_I + \pi_P + \pi_{F_0}]\lambda_L,\\
\pi_{D_0}[\lambda_H + \lambda_L + T_1^{-1}] &= \pi_{D_H}\mu_H + \pi_{D_L}\mu_L,\\
\pi_{D_L}[\lambda_H + \mu_L] &= \pi_{D_0}\lambda_L + \pi_{F_L}\tau_H^{-1},\\
\pi_{D_H}\mu_H &= \sum_{x\in\{I,P,F_0,F_L,D_0,D_L\}} \pi_x \lambda_H, \quad (1)
\end{aligned}$$

The left hand side of (1) represents the steady-state probability of a state x times the total rate of moving out of the state, while the right hand side is the sum of the probabilities of the states from which one can move into x each multiplied by the corresponding transition rate. Similar balance equations can be written for the intermediate states S_{xy}, e.g. $\pi_{S_{ID_H}}\sigma_{ID} = \pi_I\lambda_H$, allowing us to express the normalisation condition $1 = \sum_{x,y\in\{I,P,F_0,F_L,D_0,D_L,D_H\}} \pi_x + \pi_{S_{xy}}$ as:

$$\begin{aligned}
1 = \underbrace{\pi_I[1 + \frac{\Lambda_H}{\sigma_{ID}} + \frac{\Lambda_L}{\sigma_{IF}}]}_{\Pr[\text{user in Idle}]} + \underbrace{\pi_P[1 + \frac{\Lambda_H}{\sigma_{PD}} + \frac{\Lambda_L}{\sigma_{PF}} + \frac{T_3^{-1}}{\sigma_{PF}}}_{\Pr[\text{user in PCH}]}\\
+ \frac{T_3^{-1}}{\sigma_{FI}}] + \pi_{F_0}[1 + \frac{\Lambda_H}{\sigma_{FD}} + \frac{T_2^{-1}}{\sigma_{FP}}] + \pi_{F_L}[1 + \frac{\Lambda_H}{\sigma_{FD}}]\\
+ \underbrace{\pi_{D_0}[1 + \frac{T_1^{-1}}{\sigma_{DF}}] + \pi_{D_L} + \pi_{D_H}}_{\Pr[\text{user in DCH}]}. \quad (2)
\end{aligned}$$

The average signalling load (msg/s) on the RNC generated by the UE due to both normal and malicious traffic is then:

$$\begin{aligned}\gamma_r &= \pi_I[\Lambda_L r_{IF} + \Lambda_H r_{ID}] + \pi_P[\Lambda_L r_{PF} + \Lambda_H r_{PD}] \\ &\quad + [\pi_{F_0} + \pi_{F_L}]\Lambda_H r_{FD} + \pi_{D_0} T_1^{-1} r_{DF} \\ &\quad + \pi_{F_0} T_2^{-1}[r_{FP}\mathbf{1}_{\mathbf{F}\to\mathbf{P}} + r_{FI}\mathbf{1}_{\mathbf{F}\to\mathbf{I}}] + \pi_P T_3^{-1} r_{PI}\mathbf{1}_{\mathbf{F}\to\mathbf{P}}, \quad (3)\end{aligned}$$

where the characteristic function $\mathbf{1}_{\mathbf{x}\to\mathbf{y}}$ takes the value 1 if the transition $x \to y$ is enabled and 0 otherwise. The UE also generates signalling with the CN whenever it moves to/from the Idle state, leading to an average signalling load on the SGSN given by:

$$\begin{aligned}\gamma_c &= \pi_I[\Lambda_L c_{IF} + \Lambda_H c_{ID}] + \pi_{F_0} T_2^{-1} c_{FI}\mathbf{1}_{\mathbf{F}\to\mathbf{I}} \\ &\quad + \pi_P T_3^{-1} c_{PI}\mathbf{1}_{\mathbf{F}\to\mathbf{P}}. \quad (4)\end{aligned}$$

A. MODELING CONGESTION IN THE CONTROL PLANE

The analytical model we just described can be solved in closed-form [27] when the average transition delays are known, allowing to determine the conditions and parameters for which signalling misbehavior has the most serious consequences on the network functioning. In normal circumstances, state promotions and demotions last for few milliseconds that represent only a small fraction of the total lifetime of a session. However, when the mobile network servers become overloaded, as in during a signalling storm, the time needed to establish and release connections also increases, which in turn limits the maximum signalling load that a set of misbehaving UEs can impose on the network. To better understand the effect of a signalling storm, we develop a simple model for the average time σ_{xy}^{-1} needed to perform the transition $x \to y$ as follows:

$$\sigma_{xy}^{-1} = r_{xy} w + \sum_{n=1}^{r_{xy}} (t_{xy}[n] + \delta_{xy}[n]), \quad (5)$$

which consists of three components:

- Communication delay $t_{xy}[n]$ comprising propagation and transmission parts that are subject to the physical characteristics of the links traversed by the n-th signalling message exchanged during the transition. This delay depends only on the path followed by the message, and we ignore queueing at the transmission links, since signalling storms do not affect the data plane, and thus they do not translate into congestion in the wireless or wired links.
- Average queueing delay w at the RNC signalling server, which is a function of the number of normal UEs served by the RNC $M^{\mathcal{N}}$, the number of misbehaving ones $M^{\mathcal{A}}$, and the RNC signalling load (3) of both normal $\gamma_r^{\mathcal{N}}$ and misbehaving $\gamma_r^{\mathcal{A}}$ UEs. Note that we do not represent congestion at the SGSN, since the CN is less susceptible to signalling storms, especially when PCH is enabled.
- Processing time $\delta_{xy}[n]$ at the mobile network servers handling the message, which we assume to be constant per message type[1] such that $\delta_{xy}[n] = \sum_{s\in servers} \delta_{xy,s}[n]$.

The aggregate load that the RNC signalling server needs to handle is then:

$$\Gamma_r = M^{\mathcal{N}}\gamma_r^{\mathcal{N}} + M^{\mathcal{A}}\gamma_r^{\mathcal{A}}.$$

Note that Γ_r is a function of w, which itself is determined by Γ_r. Using a simple $M/M/K$ system to model the RNC signalling server, the average queueing delay becomes [51]:

$$w = \frac{(K\rho)^K}{K!(1-\rho)(K\nu - \Gamma_r)}\left[\sum_{i=0}^{K-1}\frac{(K\rho)^i}{i!} + \frac{(K\rho)^K}{K!(1-\rho)}\right]^{-1}, \quad (6)$$

where $\rho = \frac{\Gamma_r}{K\nu}$, and ν is an "equivalent" average service rate which depends on the composition of the signalling messages processed by the RNC:

$$\nu^{-1} = \Gamma_r^{-1} \sum_{\mathcal{C}\in\{\mathcal{N},\mathcal{A}\}} M^{\mathcal{C}} \sum_{x,y} a_{xy}^{\mathcal{C}} \sum_{n=1}^{r_{xy}} \delta_{xy,r}[n],$$

where $a_{xy}^{\mathcal{C}}$ is the rate at which a UE of type $\mathcal{C} \in \{\mathcal{N}, \mathcal{A}\}$ triggers the transition $x \to y$ (i.e. $\gamma_r^{\mathcal{C}} = \sum_{x,y} a_{xy}^{\mathcal{C}} r_{xy}$), and $\delta_{xy,r}[n] \geq 0$ is the RNC's processing time of the n-th signalling message exchanged during the transition. Finally, w is obtained by solving the system of equations (1), (2), (5) and (6), from which the steady state probabilities and average signalling loads follow directly.

V. SIMULATION OF UMTS NETWORKS AND SIGNALLING ANOMALIES

The mathematical user model we have developed and described in Sec. IV differentiates between normal and attack or misbehaving traffic, but it aggregates all the different user plane applications and services, and other control plane events carried by RRC such as mobility management updates, into a few representative traffic rates assuming Poisson arrivals. Therefore, this model is necessarily an approximation of the overall signalling behaviour of the UE, and the traffic parameters of the user need to be carefully selected based on the scenario of interest and the real-life behaviour of users as they interact with various mobile applications and services. This process would normally involve the aggregation of all user plane activity into the few traffic rates of the model and an approximate translation of non-Poisson traffic patterns into Poisson arrivals, which introduces some discrepancy between the mathematical model and the actual behaviour of the UE.

Although the model enables us to quickly derive analytical results in order to investigate the effect of signalling storms and the values of the various network parameters, such as the T_i timers, on signalling load, it cannot represent the user plane behaviour at the application level in detail, e.g., it cannot

[1]Note that signalling message types are defined by the 3GPP standards and known a priori.

differentiate between traffic patterns due to web traffic and instant messaging. Another assumption of the mathematical model is that we know the (aggregate) normal and attack traffic patterns and therefore can select the corresponding traffic parameters accordingly. In cases when the misbehaving traffic pattern is not known, or if we cannot clearly distinguish between normal and attack traffic, the mathematical model is still useful for improvised evaluations, but it is significantly more difficult to choose the correct model parameters for a more realistic analysis.

In order to capture such aspects of the mobile network not explicitly represented in the mathematical model, we have developed a discrete event simulation model of the UMTS network, focusing on the signalling layer in the RAN. The simulation models were developed independent of the mathematical model, and are indeed a more realistic approximation of the UMTS protocol stacks of both the control and user planes. Each node of the mobile network is represented as a self-contained and independent entity in the simulation, and nodes communicate through message exchanges, which are modeled based on the 3GPP standards for mobile protocols. We have developed models of the UE, Node-B, RNC, SGSN and GGSN, and also models of the Internet cloud and Internet hosts (i.e., servers). While we do not model the circuit-switched (CS) domain explicitly, the SGSN model contains aspects of the MSC server necessary to establish and tear-down CS calls, i.e., voice calls and SMS; our SGSN model is therefore a hybrid of the SGSN and the MSC server.

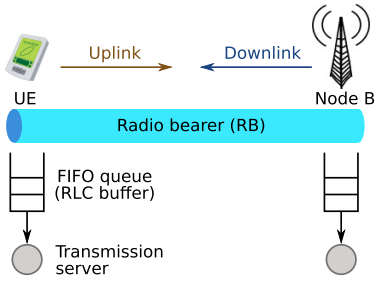


FIGURE 4. The simulation model of a radio bearer, consisting of a *(single server, single FIFO queue)* pair in each direction. The uplink and downlink servers are located at the UE and the Node-B, respectively.

In the control plane, we model the session management (SM), GPRS mobility management (GMM) and RRC layers in significant detail. In the user plane, we model different applications at the application layer, which includes CS and IP applications and allows us to differentiate between different types of user activity. We also realistically model the transport layer (TCP and UDP) and the IP layer. We have a simplified model of the RLC layer, but we do not explicitly model the MAC and PHY layers; effects of changes in radio conditions are modeled as random variations in the data rate of the radio channels. Uplink and downlink radio transmissions over a radio bearer (RB) are modeled by two single server, single FIFO queue pairs, one for each direction as shown in Fig. 4. The service time at the transmission server, i.e., radio bearer, is calculated based on the length of the currently transmitted RLC packet and the current data rate for the RB. Changes in the RB data rate are reflected on the service time of the current packet. Each UE has one signalling RB and one data RB. In addition to the transmission delays for the RBs, propagation and processing delays are also modeled. We also model the usual communication delays (i.e., transmission, propagation and processing delays) over wired links connecting the different network elements, e.g., between the RNC and the SGSN.

In order to improve the performance of simulations and to be able to realistically evaluate large scale mobile networks, we combine *packet-level* and *call-level* representation of user plane communications in our simulation model. Communications that are message-based or bursty in nature are represented at the packet level; these include communications for SMS, email, web browsing, and instant messaging. Other types of communications are represented at the call level: examples include voice and VoIP calls, and multimedia streaming. Furthermore, our simulation models support *distributed simulation*, allowing us to leverage multiple hosts and processors in a single simulation.

In addition to the control plane protocols discussed above, we model the RANAP, NBAP and GTP protocols. The RRC model in the RNC consists of a single signalling server and a single FIFO queue, used to model the processing time $\delta_{xy,r}$ for RRC signalling messages. The server handles two classes of signalling messages, where one class consists of signalling messages that effect a state transition $x \rightarrow y$ (e.g., the RB setup message), and the second class includes all other signalling messages, including mobility updates. The service time assigned to the first class reflects the time taken to allocate and deallocate radio resources by the RNC, whereas a default and smaller service time is used for the second class (one ms in our simulations). In the analytical results presented in the next section, $K = 1$, and ν is calculated based on the $\delta_{xy,r}$ values, which are given in Table 2. These values were chosen based on the typical processing required to effect a change that the signalling involves, for example setting up a radio bearer, and reflects the complexity of the procedure based on 3GPP standards. It should be noted that while these values are realistic, they are by no means definitive since the exact values are vendor-dependent. The signalling server at the RNC is one of the main points of interest in our simulations, and as we will discuss in Sec. VII, it will become overloaded as the severity of the signalling storm increases.

VI. EXPERIMENTS

In order to understand the effect of RRC-based signalling attacks in UMTS networks, we implemented our simulation model in the OMNeT++ simulation framework [52]. We present results from simulation experiments and analytical results derived from our mathematical model. The UMTS network topology used in the simulations closely resembles the architecture shown in Fig. 1. In the simulations, we have 1,000 UEs in an area of 2×2 km^2, which is covered by

TABLE 2. Service times at the RNC signalling server for handling RRC signalling messages.

Initial state x	Resulting state y	Service time $\delta_{\mathbf{xy,r}}$ (ms)
Idle	FACH	75
PCH	FACH	15
FACH	DCH	35
DCH	FACH	25
FACH	PCH	10
PCH	Idle	30

seven Node-Bs connected to a single RNC. The CN consists of the SGSN and the GGSN, and the GGSN is connected to ten Internet hosts acting as web servers. All UEs attach to the mobile network at the start of the simulation, and remain attached. We simulate a high level of web browsing activity in a two and a half hour period. Our web browsing model is based on 3GPP recommendations [53], and is described below.

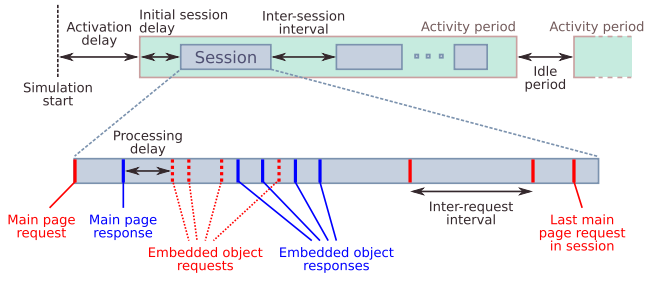


FIGURE 5. Web traffic model representing interactive user browsing in our simulations. The traffic model is self-similar, consisting of active browsing sessions and inactive intervals. This user model is independent of and significantly different than the simpler aggregate signalling model of the user presented in Sec. IV. Time is not drawn to scale.

A. THE WEB BROWSING MODEL

We model interactive web browsing behavior using a self-similar traffic model as shown in Fig. 5. The parameters of the web traffic model are random variables from probability distributions; Table 3 gives the values we used in our simulations, which are based on web metrics released by Google [54]. This simulation model of the user is significantly more complex than in the mathematical model, and allows us to represent user behaviour more realistically and without assuming Poisson arrivals.

The day-night cycle of the user is represented by the activity period, which is the time the UE is actively generating web traffic during a 24-hour period. The user starts its first activity period after an activation delay $\mathbf{d_a}$, and the period consists of one or more browsing sessions. The first session within an activity period starts after an initial session delay $\mathbf{d_s}$, and the inter-session interval $\mathbf{i_s}$ is the time between the last and the first main request in one session and the next.

Within a session, the user generates main page requests and embedded object requests for web pages and the web

TABLE 3. Parameters of the web traffic model used in the simulation experiments.

Name	Description	Value
p_a	Activity period	constant, 24 hours
d_a	Activation delay (min.)	uniform(1, 10)
d_s	Initial session delay	$i_s/2$
n_s	Number of main page requests in session	truncated normal $\mu = 10$, $\sigma = 5$, min $= 2$
i_s	Inter-session interval (min.)	truncated normal, $\mu = 20$, $\sigma = 10$, min $= 2$
i_r	Inter-request interval (sec.)	truncated exponential, $\lambda^{-1} = 60$, min $= 10$, max $= 600$
l_r	Request size (B)	truncated normal, $\mu = 600$, $\sigma = 100$, min $= 300$
l_m	Main page size, excluding embedded resources	histogram [54]
l_{img}	Size of image resources (KB)	truncated exponential, $\lambda^{-1} = 50$, min $= 1.2$, max $= 400$
l_{txt}	Size of text resources	histogram [54]
n_e	Number of embedded objects in page	histogram [54]
R_{img}	Ratio of image resources to all embedded resources in page	uniform(0.1, 0.5)
d_{p_c}	Processing delay, client (ms)	truncated normal, $\mu = 50$, $\sigma = 10$, min $= 0$
d_{p_s}	Processing delay, server (ms)	truncated normal, $\mu = 4$, $\sigma = 1$, min $= 1$

objects embedded within the main page, respectively. The first main page request is scheduled at the start of the session, which results in a page response from the web server. This response is subject to a processing delay $\mathbf{d_{p_c}}$ at the client, which represents the time it takes for the web client at the UE to process the received response. A web page contains zero or more embedded objects, and the client generates an embedded object request for each one. We assume that HTTP version 1.1 is used and that each embedded object request is pipelined over a single TCP connection. The length of a request is denoted by $\mathbf{l_r}$. The inter-request interval $\mathbf{i_r}$ is the time between the generation of two consecutive main page requests, and it is independent of the reception of the responses. The session length is controlled by the number of main page requests $\mathbf{n_s}$ in the session.

The web server generates a response for each request it receives after a processing delay $\mathbf{d_{p_s}}$. The length of a main page response is $\mathbf{l_m}$, and it excludes the sizes of any embedded objects and TCP/IP headers. The number of embedded objects per page is $\mathbf{n_e}$, and we model two types of objects: images and text (e.g., CSS documents, scripts). The size of an embedded object is $\mathbf{l_{img}}$ and $\mathbf{l_{txt}}$ for image and text objects, respectively. $\mathbf{R_{img}}$ gives the ratio of image objects to all embedded objects in a page. In the simulations, a client selects a web server uniformly at random for each main page request.

B. THE ATTACK MODEL

We consider two different attack strategies, or equivalently, misbehaviour patterns in our evaluation: FACH and

DCH attacks. Note that in the rest of this paper, we will use the terms *attack* and *misbehaviour* interchangeably. In *FACH attacks*, the attacker aims to overload the control plane by causing superfluous promotions to the FACH state, and therefore needs to know when a demotion from FACH occurs in the UE. In *DCH attacks*, the demotion of interest is from the DCH state. As introduced in Sec. IV, the error between the actual transition time and the estimated one is denoted by τ_L and τ_H in the FACH and DCH attack scenarios, respectively. Consequently, $1/\tau$ is a measure of the *aggressiveness* of the misbehaving application.

In FACH attacks, the attacker sends a small data packet to a random Internet server in order to cause a promotion to FACH. Higher rate data traffic is generated in DCH attacks in order to cause the buffer threshold to be reached and therefore result in a promotion to DCH. For simulation purposes, our RRC model at the UE informs all registered malicious applications when an RRC state transition occurs. Before launching the next attack, the attacker waits for a period of τ_L or τ_H after a suitable demotion is detected, e.g., from FACH to PCH in the FACH attack case, where τ_L, τ_H are random variables. In our experiments, we assume that τ_L, τ_H are exponentially distributed with mean $= \{0, 1, 2, 4, 6, 10, 14, 20, 30\}s$ to simulate varying degrees of error on behalf of the attacker. For signalling storms, τ represents the synchronization between the RRC state machine of the UE and the misbehaving application, while the attack scenario represents whether the misbehaving application generates low-rate or high-rate traffic. We present results from the DCH attack scenario only since the FACH attack scenario produces similar behaviour in most cases.

VII. MODELING AND SIMULATION RESULTS

We performed simulation experiments in order to investigate the effect of signalling attacks and storms due to the RRC protocol on the RAN and the CN. We vary the number of compromised or misbehaving UEs from 1% to 20% of all UEs. Both normal and misbehaving UEs generate *normal traffic* based on the web browsing model described above. The misbehaving applications are activated gradually between 20 and 30 minutes from the start of the simulation in order to prevent artifacts such as a huge spike of signalling load due to many malicious applications coming online at the same time. We collect simulation data only from the period when all misbehaving UEs are active. Each data point in the presented results is an average of five simulation runs with different random seeds, resulting in different mobility and traffic patterns. The relevant RRC protocol parameters are as given in Tables 1 and 2. The simulation results do not capture signalling due to mobility and session management, but we have observed from other experiments that these signalling activities have negligible effect on the resulting signalling load in the network since the rate of signalling messages exchanged for these activities is minor compared to RRC signalling, especially in the case of a signalling storm.

We present analytical results derived from our mathematical model together with the simulation results. However, we do not present analytical results for Figs. 8 and 9 to prevent repetition of similar results, and for Fig. 8 since the mathematical model does not capture quality-of-experience. The parameters of the mathematical model were chosen based on an initial set of simulation experiments, from which we derived the aggregate normal and misbehaving user patterns for the UE. This enabled us to validate the mathematical model using simulation experiments in similar settings and parameters.

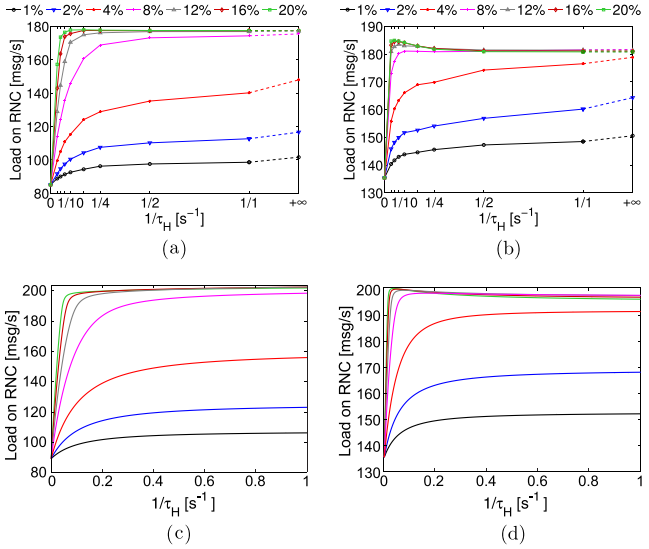


FIGURE 6. Signalling load (sum of the rates of the incoming and outgoing signalling messages) on the RNC vs. aggresiveness ($1/\tau_H$) under DCH attacks. Each line represents a different number of misbehaving devices. The $1/\tau_H = 0$ case corresponds to a *no attack* scenario. We present analytical and simulation results with the PCH state enabled or disabled in the network, and observe that the analytical model can produce accurate results given that the parameters of the model are correctly chosen. (a) PCH enabled (simulation). (b) PCH disabled (simulation). (c) PCH enabled (analytical). (d) PCH disabled (analytical).

Figure 6 shows the signalling load in the RAN under DCH attacks, with PCH enabled or disabled; the signalling load is calculated as the sum of the rate of incoming and outgoing signalling messages to and from the RNC, and therefore it is not a direct measure of the capacity of the RNC. We observe that the rate of increase of the signalling load is significantly higher when the number of attackers is high, and that enabling the PCH state slightly decreases the signalling load in the RAN. A worrying observation is that when PCH is disabled, there is a possibility to induce a maximum signalling load on the RNC without requiring a high level of synchronization between the misbehaving application and the RRC state machine. Enabling the PCH state resolves this issue. Another useful observation is that given a fixed number of attackers, RRC attacks are *self-limiting*: as signalling load on the RNC increases, this prevents attackers from being able to attack the network at a high rate since they are themselves subject to longer waits for channel allocations. We will re-visit this issue

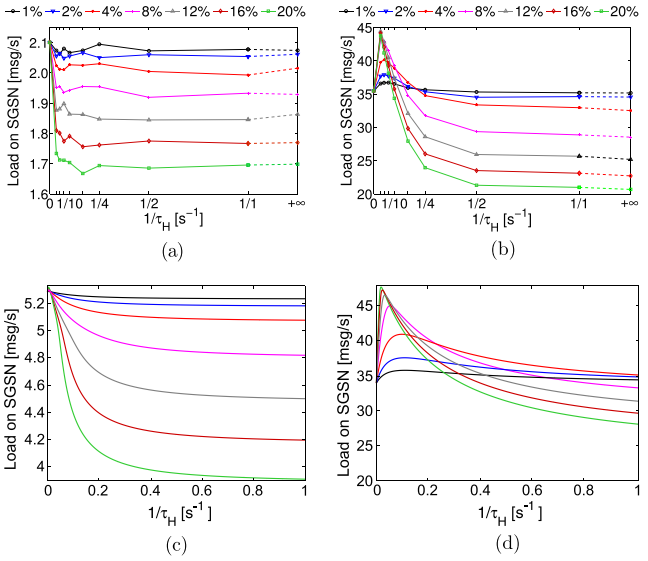


FIGURE 7. Signalling load (sum of the rates of the incoming and outgoing signalling messages) on the SGSN vs. aggresiveness ($1/\tau_H$) under DCH attacks. Each line represents a different number of misbehaving devices. The $1/\tau_H = 0$ case corresponds to a *no attack* scenario. We present analytical and simulation results with the PCH state enabled or disabled in the network, and observe that enabling it significantly reduces signalling load on the SGSN. The analytical and simulation results still show a high degree of agreement. (a) PCH enabled (simulation). (b) PCH disabled (simulation). (c) PCH enabled (analytical). (d) PCH disabled (analytical).

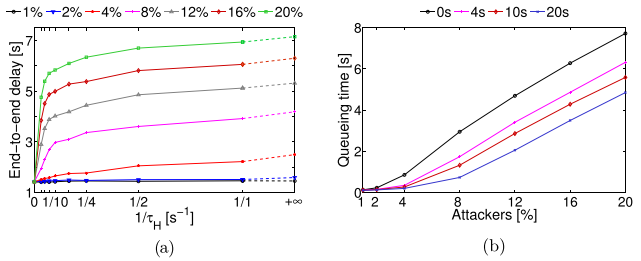


FIGURE 8. Effect of signalling storms on application response time at *normal* devices, and on queueing time at the RNC signalling server under DCH attacks, with PCH disabled. (a) Application response time (s) vs. aggresiveness ($1/\tau_H$) under DCH attacks, with PCH disabled. Each line represents a different number of misbehaving devices. (b) Average queueing time (s) at the RNC signalling server vs. percentage of misbehaving devices under DCH attacks, with PCH disabled. Each line represents a different τ_H value.

when we discuss congestion at the RNC signalling server below.

Figure 7 shows the signalling load in the CN under DCH attacks, with PCH enabled or disabled, and demonstrates the advantage of enabling the optional PCH state. Most RRC-induced signalling with the CN occurs when the UE enters and exits the *idle* state. With PCH enabled, signalling load on the SGSN drops with decreasing τ_H since more frequent messages prevent the UE from entering the idle state and thus reduce the signalling load on the SGSN. Therefore, our recommendation would be to enable PCH as a first step in the mitigation of RRC-based signalling attacks and storms. Enabling the PCH state also eliminates the problem of the maximum signalling load observed in Fig. 7 for high values of τ_H, which is due to the interaction between τ_H and the RRC inactivity timers T_1 and T_2. When $\tau_H > T_1+T_2$, the UE enters the *idle* state as a result of inactivity, and then the misbehaving application causes the UE to go into FACH or DCH in order to send data, resulting in excessive signalling with the CN. The long T_3 timer for demotion from the PCH state solves this issue.

Our results so far demonstrate how the mobile network infrastructure is seriously affected by RRC-based signalling anomalies. These anomalies also have an appreciable impact on the quality-of-experience (QoE) of the mobile user. Figure 8a shows the application response time, which is defined as the time between when the user requests a web page and when all of the web page is received, at a normal UE. The response time is not greatly affected when there are very few misbehaving UEs and when τ_H is high. But delay increases by up to 400% as the severity of the attack increases with increasing number of attackers and $1/\tau_H$. Users normally tolerate a wait of 2–10 seconds for a web page to download [55], and therefore the observed response times are significant from a QoE view. The affected mobile users are highly likely to attribute the bad QoE to the MNO, so the MNO has one more incentive to detect and mitigate signalling problems in its network.

The main reason for the increase in application response time is the time it takes for the UE to acquire a radio channel in order to send and receive data, which includes, in addition to the communication delays between the UE and the RNC, the service and queueing times experienced by the RRC signalling messages effecting the channel acquisition. Figure 8b shows that queueing time at the RRC signalling server component of the RNC greatly increases as the number of attackers increase. We observe that effects of congestion at the server become significant when the percentage of attackers is $\geq$8%, affecting application response time for normal users, and also placing a limit on the impact of signalling attacks on the network since the attackers themselves are subject to longer delays for channel acquisition. This *self-limiting behaviour* imposes a maximum signalling load of around 200 msgs/s on the RNC (Fig. 6). Note that the service time for RRC messages effecting a FACH→DCH transition, which is the transition exploited in the DCH attack scenario, is 35 ms, meaning that the RNC would be congested by an incoming rate of 30 msgs/s of such messages. However, the signalling load observed on the RNC is significantly higher than this (around 200 msgs/s) since it is (mostly) the rate of incoming signalling messages to the RNC, which is only loosely based on the service capacity of the RNC because the congestion at the RNC signalling server does not prevent the UEs from sending channel requests until they are blocked waiting for a reply to their previous request. This behaviour is the main cause of the self-limiting nature of the signalling storm: if all the UEs in the area are blocked due to congestion, no more signalling requests are received by the RNC until it has processed some of the requests and therefore has allowed those UEs to send subsequent requests.

The service capacity in the RAN can be increased by installing more RNCs to handle the same number of subscribers or by using a node with more capacity. Installing more RNCs is very cost-ineffective, and thus would be shunned by MNOs. Installing a higher-capacity RNC also does not address the inherent signalling problem since the RNC would then be provisioned to handle a larger number of base stations, and thus more subscribers, due to cost efficiency reasons. We therefore need to understand the nature of signalling storms so that we may develop cost-effective detection and mitigation methods, which could be installed as part of the admission control component in the RNC and prevent the signalling storm from occurring in the first place.

We observe that while RRC-based attacks have a significant impact on the RAN, they do not greatly affect the CN. This is due to the nature of the RRC protocol, which is essentially an access network protocol between the UE and the RNC. Therefore, an attacker that wishes to attack the CN directly needs to adopt other strategies, such as authentication attacks [56]. The advantages of the investigated attack for the attacker is its ease of implementation since it only requires that the attacker estimates the RRC-related parameters of the network, which is easily attainable [35], and then listens to user activity in order to estimate when RRC state transitions will occur on the infected device. A simpler attack that would have a similar impact would be sending frequent and periodic messages in order to induce unnecessary state transitions, and this is indeed the type of behaviour we observe with misbehaving or poorly designed applications which cause signalling storms rather than deliberate signalling attacks.

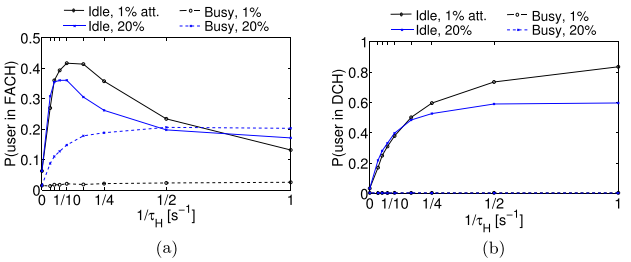


FIGURE 9. Radio channel utilization vs. aggresiveness ($1/\tau_H$) under DCH attacks, with PCH disabled. We observe that normal ($1/\tau_H = 0$) and misbehaving ($1/\tau_H \neq 0$) devices exhibit markedly different channel utilizations, which suggests that channel utilizations and busy and idle times can be used as representative features for efficient detection of signalling storms. (a) Ratio of time spent in the FACH state while idle and busy to total time spent in all RRC states. (b) Ratio of time spent in the DCH state while idle and busy to total time spent in all RRC states.

Our final results relate to how the UE utilizes its allocated radio resources, and provide a useful feature that we aim to exploit in our future work on the detection of signalling attacks. Figure 9 shows the ratio of time the UE is in the FACH or DCH state while busy (i.e., ending or receiving data) and idle. The most important observation is that a normal UE, represented with $1/\tau_H = 0$, has a markedly different behaviour than a misbehaving UE ($1/\tau_H > 0$), and the discrepancy increases with $1/\tau_H$. Normal UEs do not spend a significant time in FACH or DCH as busy or idle, but attackers spend a long time as idle while in FACH and DCH, i.e., their session tails are comparatively *longer* than their session body. This is because normal users only acquire the channel when they have legitimate traffic, and they send larger chunks of data and therefore use the channel for longer than attackers, resulting in a low ratio of idle to busy time. Attackers, on the other hand, frequently acquire the channel to send only a small amount of attack traffic and therefore waste most of the radio channel as reflected in their high ratio of idle to busy time. The exception to this is the FACH state when there is congestion in the control plane due to the signalling attack: we observe that attackers spend significantly long times as busy in the FACH state when there is congestion, e.g., with 20% of attackers, which is due to the long delay it takes the UEs to acquire the channel as discussed above.

VIII. RELATED WORK

The vulnerability of mobile networks to different types signalling attacks and storms have been recognized even prior to 3G networks. Pre-3G signalling attacks include the *SMS flooding attack* [57] and the *paging attack* [4]. Enck et al. [57] show that an SMS attack originating from GSM-capable Internet hosts can significantly degrade, and in the worst case prevent voice and SMS services on the cellular network. Two countermeasures are proposed in [7]: providing differentiated services via queue management, and resource provisioning to preferentially allocate channel resources over the air interface. In [9], the possibility of SMS attacks originating from mobile devices within the cellular network is considered, and the authors show the feasibility of such an attack by implementing it using feature phones on a 2G network.

The paging attack exploits the paging mechanism which is used to locate and connect to idle devices in the mobile network for incoming calls. Serror et al. [4] addressed the problem of paging attacks due to Internet-originating data calls on a CDMA2000 network, and showed that the paging channel exhibits a sharp rather than a graceful degradation under load. Similar problems still exist in 4G networks as discussed in [58].

RRC-based signalling attacks and storms have been investigated in [6], where the authors consider a remote host-based attack on UMTS networks and propose an online detection method based on the statistical cumulative sum test. The detector is located at the GGSN, and uses a packet sniffer to look at IP metrics such as destination addresses and the estimated radio access bearer setup time in order to detect the intention of launching an attack, even though the activity may not actually have an effect on the signalling load. Our investigation of signalling storms suggests that a better method would be to install the detector at the RNC, possibly as part of the existing admission control mechanisms, since then an effective mitigation mechanism can be combined with the detector to jointly identify and solve the problem.

RRC-based signalling attacks [59] and storms [60] effect LTE networks as well. In [59], the authors evaluate the effect

of an RRC-based signalling attack on an LTE network using simulation experiments, and show the resulting performance degradation in the eNode-Bs and the evolved packet core (EPC). The utilization of LTE radio channels such as PUSCH and PUCCH due to keep-alive messages is studied in [60]. We are currently investigating the effect of signalling storms in LTE networks, paying special attention to machine-to-machine communications, which are a considerable source of signalling problems [23].

RRC-based signalling attacks are not the only possible attacks targeting the control plane of mobile networks. Other attacks typically target the core network, aiming to overload the Home Location Register (HLR) or the Authentication Center (AuC). Various types of authentication attacks exploiting the authentication mechanism between the UE and the mobile network in UMTS networks have been discussed in [56], and the signalling load of authentication messages in LTE networks has been evaluated using renewal process theory and analytical modeling in [61]. An interesting attack that exploits the network attach procedure in UMTS networks is described in [62], where SIM-less devices are used to overload the HLR and the AuC.

The IP Multimedia Subsystem (IMS) in 3G and 4G networks has also been the target of signalling attacks. Early work in this area has looked at the signalling load due to the Session Initiation Protocol (SIP) used in the IMS [63]. Zhao et al. [64] have identified an IMS attack that overloads the presence servers by exploiting SIP, and have proposed a detection mechanism based on the Girshick-Rubin-Shiryaev algorithm that looks at the CPU usage at the presence servers in order to detect the attack.

Other work has looked at how signalling attacks can be mitigated. A detailed review of signalling attacks in 3G networks is presented in [8], where the authors identify the system design decisions that result in such attacks, and convincingly argue that the design focus should move from optimality to robustness and elasticity of mobile networks. The methods that they propose to achieve this change are randomization of the radio resource management (RRM) and mobility management (MM) procedures, device-specific adaptive state transitions based on profiles, and prioritization of devices. Wu et al. [65] evaluate one such method, the randomization of the RRM and MM procedures in 3G networks, and show that it can indeed mitigate against certain attacks while acceptably degrading normal performance. We are currently developing a signalling storm detector and mitigator (SSDM) based on our investigation of the signalling behaviour of UMTS networks under signalling storms. Our SSDM adopts the device-specific adaptive state transitions approach discussed in [8], and mitigates the storm by adaptively controlling the state transitions of devices that are identified to be misbehaving, and thus will impact normal users less than network-wide solutions such as randomization. The SSDM can be implemented as part of the admission control mechanism in the RNC, or it can implemented on the mobile devices, for example as part of a virtualization solution designed to mitigate against a wide variety of device-originating problems as proposed in [66].

The signalling attacks and storms discussed here are not specific to UMTS and LTE networks, and WiMAX networks are also vulnerable to such problems. Kolias et al. [67] provide an in-depth review in this area. Such works highlight the importance of analyzing and understanding the root causes and the dynamic behaviour of signalling anomalies in mobile networks as they evolve with emerging application patterns and new network technologies. Recent work [68] shows that this task is not trivial since the interactions between the control plane and the user plane are more complex than previously thought. Thus, further work is necessary in this old but still emerging field in order to stay ahead of changes in the mobile landscape.

IX. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the effect of signalling attacks and storms in mobile networks, focusing on signalling anomalies that exploit the radio resource control (RRC) protocol in UMTS networks. We presented a Markov model of the signalling behaviour of the UE and extended the model for effects of congestion in the control plane. The analytical model provides an accurate representation of the RRC signalling behaviour and allows us to reach quick analytical results, but its parameters need to be carefully selected using user traffic models built based on either real-life data or on simulation results. Without being able to choose representative parameters for the mobile network under investigation and the user plane behaviour of the UE, the results provided by the mathematical model will necessarily be speculative.

In order to validate the mathematical model and to select representative parameters, we developed a realistic simulation model of the UMTS network, which is comprised of the relevant user plane and control plane protocols represented at various abstraction levels. The simulation model captures the interactions between the network elements and protocols in a UMTS network. We implemented the simulation model in a distributed network simulator, and conducted simulation experiments to evaluate the effect of signalling storms on the signalling servers and the mobile devices.

Our analytical and simulation results show that RRC-based signalling storms can cause significant problems in both the control plane and the user plane in the network, and provide insight into how such attacks and storms can be detected and mitigated. While we have focused on UMTS networks in this work, the RRC protocol is also employed in LTE networks, and any RRC related anomalies would have a more severe impact in LTE networks since they employ only two RRC states (connected and idle), and the mitigating effect of the long T_3 timer used in the PCH state are non-existent in LTE networks.

While this work has employed mathematical modelling and simulation experiments to evaluate the effect of signalling storms, it is important to validate these findings using data from operational mobile networks. We are in the process of

negotiating the release of data relevant to signalling storms from our telecommunication partners, which is inevitably a lengthy process due to legal and privacy issues. As future work, we plan to use charging data records from mobile subscribers to build user models, which will result in the adjustment of the parameters of the mathematical model and the development of new simulation models. We will also conduct experiments on signalling storms on a small-scale physical mobile network test-bed, and use these results to design more realistic simulation experiments which can scale up to larger networks.

Future work can exploit the insight gained in this paper for the detection and mitigation of signalling attacks in mobile networks. One aspect that requires attention is the identification of possible locations, such as specific cells, where attacks may originate, and methods related to search and smart traffic routing may prove valuable in this context [69], [70]. Another important aspect relates to identifying sets of representative features for the detection of signalling attacks and storms, and of the misbehaving UEs. An important consideration is to prevent false positives as much as possible so as not to punish normal heavy users. We will also develop system-wide models based on queueing theory [71] that represent a single user in a simple manner, to study mitigation methods that involve randomization and adaptively introducing artificial delays in the state transitions of the UEs so that they may automatically reduce the negative impact of attacks and signalling storms.

REFERENCES

[1] A. P. Felt, M. Finifter, E. Chin, S. Hanna, and D. Wagner, "A survey of mobile malware in the wild," in *Proc. 1st ACM Workshop Secur. Privacy Smartphones Mobile Devices (SPSM)*, 2011, pp. 3–14.

[2] M. Chandramohan and H. B. K. Tan, "Detection of mobile malware in the wild," *Computer*, vol. 45, no. 9, pp. 65–71, Sep. 2012.

[3] E. Gelenbe *et al.*, "Security for smart mobile networks: The NEMESYS approach," in *Proc. IEEE Global High Tech Congr. Electron. (GHTCE)*, Nov. 2013, pp. 63–69.

[4] J. Serror, H. Zang, and J. C. Bolot, "Impact of paging channel overloads or attacks on a cellular network," in *Proc. 5th ACM Workshop Wireless Secur. (WiSe)*, Sep. 2006, pp. 75–84.

[5] H. Yang, F. Ricciato, S. Lu, and L. Zhang, "Securing a wireless world," *Proc. IEEE*, vol. 94, no. 2, pp. 442–454, Feb. 2006.

[6] P. P. C. Lee, T. Bu, and T. Woo, "On the detection of signaling DoS attacks on 3G/WiMax wireless networks," *Comput. Netw.*, vol. 53, no. 15, pp. 2601–2616, Oct. 2009.

[7] P. Traynor, W. Enck, P. McDaniel, and T. La Porta, "Mitigating attacks on open functionality in SMS-capable cellular networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 40–53, Feb. 2009.

[8] F. Ricciato, A. Coluccia, and A. D'Alconzo, "A review of DoS attack models for 3G cellular networks from a system-design perspective," *Comput. Commun.*, vol. 33, no. 5, pp. 551–558, Mar. 2010.

[9] C. Mulliner, N. Golde, and J.-P. Seifert, "SMS of death: From analyzing to attacking mobile phones on a large scale," in *Proc. 20th USENIX Conf. Secur. (SEC)*, Aug. 2011, pp. 363–378.

[10] J. Li, W. Pei, and Z. Cao, "Characterizing high-frequency subscriber sessions in cellular data networks," in *Proc. IFIP Netw. Conf.*, May 2013, pp. 1–9.

[11] (Feb. 2012). *Smarter Apps for Smarter Phones! GSMA*. [Online]. Available: http://www.gsma.com/technicalprojects/wp-content/uploads/2012/04/gsmasmarterappsforsmarterphones0112v.0.14.pdf

[12] S. Jiantao, "Analyzing the network friendliness of mobile applications," Huawei, Shenzhen, China, Tech. Rep. M3-001034414-20120731-C-2.0, Jul. 2012. [Online]. Available: http://www.huawei.com/ilink/en/download/HW_146595

[13] F. Ricciato, "Unwanted traffic in 3G networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 2, pp. 53–56, Apr. 2006.

[14] F. Ricciato, E. Hasenleithner, P. Svoboda, and W. Fleischer, "On the impact of unwanted traffic onto a 3G network," in *Proc. 2nd Int. Workshop Secur., Privacy Trust Pervasive Ubiquitous Comput. (SecPerU)*, Jun. 2006, pp. 49–56.

[15] Trend Micro. (Jan. 2013). *TrendLabs 2012 Annual Security Roundup: Evolved Threats in a Post-PC World*. [Online]. Available: http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/reports/rpt-evolved-threats-in-a-post-pc-world.pdf

[16] C. Raiu and D. Emm. (Dec. 2012). *Kaspersky Security Bulletin 2012: Malware Evolution*, Kaspersky Lab. [Online]. Available: http://www.securelist.com/en/analysis/204792254/Kaspersky_Security_Bulletin_2012_Malware_Evolution

[17] A. Filippoupolitis and E. Gelenbe, "A distributed decision support system for building evacuation," in *Proc. 2nd Conf. Human Syst. Interactions (HSI)*, May 2009, pp. 323–330.

[18] E. Gelenbe and F.-J. Wu, "Large scale simulation for human evacuation and rescue," *Comput. Math. Appl.*, vol. 64, no. 12, pp. 3869–3880, Dec. 2012.

[19] A. Filippoupolitis, G. Gorbil, and E. Gelenbe, "Spatial computers for emergency support," *Comput. J.*, vol. 56, no. 12, pp. 1399–1416, Dec. 2013.

[20] G. Gorbil and E. Gelenbe, "Opportunistic communications for emergency support systems," *Proc. Comput. Sci.*, vol. 5, pp. 39–47, Aug. 2011.

[21] G. Gorbil and E. Gelenbe, "Resilience and security of opportunistic communications for emergency evacuation," in *Proc. 7th ACM Workshop Perform. Monitor. Meas. Heterogeneous Wireless Wired Netw. (PM2HW2N)*, Oct. 2012, pp. 115–124.

[22] O. H. Abdelrahman, E. Gelenbe, G. Görbil, and B. Oklander, "Mobile network anomaly detection and mitigation: The NEMESYS approach," in *Information Sciences and Systems* (Lecture Notes in Electrical Engineering), vol. 264, E. Gelenbe and R. Lent, Eds. Berlin, Germany: Springer-Verlag, Oct. 2013, pp. 429–438.

[23] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: Potential, challenges, and solutions," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 178–184, Mar. 2012.

[24] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: Overload control," *IEEE Netw.*, vol. 26, no. 6, pp. 54–60, Nov./Dec. 2012.

[25] Y. Chang, C. Zhou, and O. Bulakci, "Coordinated random access management for network overload avoidance in cellular machine-to-machine communications," in *Proc. 20th Eur. Wireless Conf.*, May 2014, pp. 1–6.

[26] H.-L. Fu, P. Lin, H. Yue, G.-M. Huang, and C.-P. Lee, "Group mobility management for large-scale machine-to-machine mobile networking," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1296–1305, Mar. 2014.

[27] O. H. Abdelrahman and E. Gelenbe, "Signalling storms in 3G mobile networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, Australia, Jun. 2014, pp. 1017–1022.

[28] 3GPP. *3GPP TS 25.331: Universal Mobile Telecommunications System (UMTS) Radio Resource Control (RRC) Protocol Specification*. [Online]. Available: http://www.3gpp.org/DynaReport/25331.htm, accessed Aug. 19, 2014.

[29] 3GPP. *3GPP TS 36.331: Evolved Universal Terrestrial Radio Access (E-UTRA) Radio Resource Control (RRC) Protocol Specification*. [Online]. Available: http://www.3gpp.org/DynaReport/36331.htm, accessed Aug. 19, 2014.

[30] E. Gelenbe and R. R. Muntz, "Probabilistic models of computer systems—Part I (exact results)," *Acta Inf.*, vol. 7, no. 1, pp. 35–60, 1976.

[31] E. Gelenbe and G. Loukas, "A self-aware approach to denial of service defence," *Comput. Netw.*, vol. 51, no. 5, pp. 1299–1314, Apr. 2007.

[32] D. Maslennikov and Y. Namestnikov. (Dec. 2012). *Kaspersky Security Bulletin 2012: The Overall Statistics for 2012*, Kaspersky Lab. [Online]. Available: http://www.securelist.com/en/analysis/204792255/Kaspersky_Security_Bulletin_2012_The_overall_statistics_for_2012

[33] P. Traynor *et al.*, "On cellular botnets: Measuring the impact of malicious devices on a cellular network core," in *Proc. 16th ACM Conf. Comput. Commun. Secur. (CCS)*, Nov. 2009, pp. 223–234.

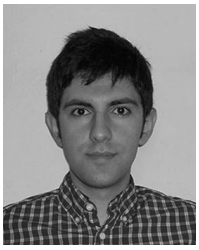
[34] C. Mulliner and J.-P. Seifert, "Rise of the iBots: Owning a telco network," in *Proc. 5th Int. Conf. Malicious Unwanted Softw. (MALWARE)*, Oct. 2010, pp. 71–80.

[35] A. Barbuzzi, F. Ricciato, and G. Boggia, "Discovering parameter setting in 3G networks via active measurements," *IEEE Commun. Lett.*, vol. 12, no. 10, pp. 730–732, Oct. 2008.

[36] P. H. J. Perala, A. Barbuzzi, G. Boggia, and K. Pentikousis, "Theory and practice of RRC state transitions in UMTS networks," in *Proc. IEEE Global Commun. Conf. Workshops (GLOBECOM Workshops)*, Nov./Dec. 2009, pp. 1–6.
[37] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "Characterizing radio resource allocation for 3G networks," in *Proc. 10th ACM SIGCOMM Conf. Internet Meas. (IMC)*, Nov. 2010, pp. 137–150.
[38] Z. Qian, Z. Wang, Q. Xu, Z. M. Mao, M. Zhang, and Y.-M. Wang, "You can run, but you can't hide: Exposing network location for targeted DoS attacks in cellular networks," in *Proc. 19th Annu. Netw. Distrib. Syst. Secur. Symp. (NDSS))*, Feb. 2012, pp. 3.3:1–3.3:16.
[39] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, "An untold story of middleboxes in cellular networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 374–385, Aug. 2011.
[40] N. Golde, K. Redon, and R. Borgaonkar, "Weaponizing femtocells: The effect of rogue devices on mobile telecommunication," in *Proc. 19th Annu. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, Feb. 2012, pp. 1–16.
[41] F. Ricciato *et al.*, "Traffic monitoring and analysis in 3G networks: Lessons learned from the METAWIN project," *e&i Elektrotech. Informationstech.*, vol. 123, nos. 7–8, pp. 288–296, Aug. 2006.
[42] C. Gabriel. (Jun. 2012). *DoCoMo Demands Google's Help With Signalling Storm*, Rethink Wireless. [Online]. Available: http://www.rethink-wireless.com/2012/01/30/docomo-demands-googles-signalling-storm.htm
[43] S. Corner. (Jun. 2011). *Angry Birds + Android + ADS = Network Overload,* IT Wire. [Online]. Available: http://www.itwire.com/business-it-news/networking/47823
[44] A. Coluccia, A. D'Alconzo, and F. Ricciato, "Distribution-based anomaly detection via generalized likelihood ratio test: A general maximum entropy approach," *Comput. Netw.*, vol. 57, no. 17, pp. 3446–3462, Dec. 2013.
[45] G. Reddig. (Sep. 2013). *OTT Service Blackouts Trigger Signaling Overload in Mobile Networks*, Nokia Solutions and Networks. [Online]. Available: http://blogs.nsn.com/mobile-networks/2013/09/16/ott-service-blackouts-trigger-signaling-overload-in-mobile-networks/
[46] (Jan. 2013). *TrendLabs 2012 Mobile Threat and Security Roundup: Repeating History*, Trend Micro. [Online]. Available: http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/reports/rpt-repeating-history.pdf
[47] Y. Zhou and X. Jiang, "Dissecting Android malware: Characterization and evolution," in *Proc. IEEE Symp. Secur. Privacy*, May 2012, pp. 95–109.
[48] E. Gelenbe, "Sensible decisions based on QoS," *Comput. Manag. Sci.*, vol. 1, no. 1, pp. 1–14, Dec. 2003.
[49] M. Siekkinen, M. A. Hoque, J. K. Nurminen, and M. Aalto, "Streaming over 3G and LTE: How to save smartphone energy in radio access network-friendly way," in *Proc. 5th Workshop Mobile Video (MoVid)*, Feb. 2013, pp. 13–18.
[50] E. Gelenbe, "Probabilistic models of computer systems," *Acta Inf.*, vol. 12, no. 4, pp. 285–303, 1979.
[51] E. Gelenbe and G. Pujolle, *Introduction to Queueing Networks*, 2nd ed. New York, NY, USA: Wiley, Apr. 1998.
[52] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. 1st Int. Conf. Simulation Tools Techn. Commun., Netw. Syst. Workshops (Simutools)*, Mar. 2008, pp. 60:1–60:10.
[53] *CDMA2000 Evaluation Methodology—Revision A*, document 3GPP2 C.R1002-A, May 2009. [Online]. Available: http://www.3gpp2.org/public_html/specs/C.R1002-A_v1.0_Evaluation_Methodology.pdf
[54] S. Ramachandran. (May 2010). *Web Metrics: Size and Number of Resources*, Google. [Online]. Available: https://developers.google.com/speed/articles/web-metrics
[55] F. F.-H. Nah, "A study on tolerable waiting time: How long are Web users willing to wait?" *Behaviour Inf. Technol.*, vol. 23, no. 3, pp. 153–163, 2004.
[56] G. Kambourakis, C. Kolias, S. Gritzalis, and J. H. Park, "DoS attacks exploiting signaling in UMTS and IMS," *Comput. Commun.*, vol. 34, no. 3, pp. 226–235, Mar. 2011.
[57] W. Enck, P. Traynor, P. McDaniel, and T. L. Porta, "Exploiting open functionality in SMS-capable cellular networks," in *Proc. 12th ACM Conf. Comput. Commun. Secur. (CCS)*, Nov. 2005, pp. 393–404.
[58] A. Baraev, U. Ayesta, I. M. Verloop, D. Miorandi, and I. Chlamtac, "Technical vulnerability of the E-UTRAN paging mechanism," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 2247–2252.
[59] R. Bassil, I. H. Elhajj, A. Chehab, and A. Kayssi, "Effects of signaling attacks on LTE networks," in *Proc. 27th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, Mar. 2013, pp. 499–504.
[60] Z. Zhang, Z. Zhao, H. Guan, D. Miao, and Z. Tan, "Study of signaling overhead caused by keep-alive messages in LTE network," in *Proc. 78th IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–5.
[61] C.-K. Han, H.-K. Choi, J. W. Baek, and H. W. Lee, "Evaluation of authentication signaling loads in 3GPP LTE/SAE networks," in *Proc. 34th IEEE Conf. Local Comput. Netw. (LCN)*, Oct. 2009, pp. 37–44.
[62] A. Merlo, M. Migliardi, N. Gobbo, F. Palmieri, and A. Castiglione, "A denial of service attack to UMTS networks using SIM-less devices," *IEEE Trans. Dependable Secure Comput.*, vol. 11, no. 3, pp. 280–291, May/Jun. 2014.
[63] D. S. Tonesi, L. Salgarelli, Y. Sun, and T. F. La Porta, "Evaluation of signaling loads in 3GPP networks," *IEEE Wireless Commun.*, vol. 15, no. 1, pp. 92–100, Feb. 2008.
[64] B. Zhao, C. Chi, W. Gao, S. Zhu, and G. Cao, "A chain reaction DoS attack on 3G networks: Analysis and defenses," in *Proc. 28th IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2009, pp. 2455–2463.
[65] Z. Wu, X. Zhou, and F. Yang, "Defending against DoS attacks on 3G cellular networks via randomization method," in *Proc. Int. Conf. Edu. Inf. Technol. (ICEIT)*, Sep. 2010, pp. V1-504–V1-508.
[66] C. Mulliner, S. Liebergeld, M. Lange, and J.-P. Seifert, "Taming Mr Hayes: Mitigating signaling based attacks on smartphones," in *Proc. 42nd Annu. IEEE/IFIP Inter. Conf. Dependable Syst. Netw. (DSN)*, Jun. 2012, pp. 1–12.
[67] C. Kolias, G. Kambourakis, and S. Gritzalis, "Attacks and countermeasures on 802.16: Analysis and assessment," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 487–514, Mar. 2013.
[68] S. Rosen *et al.*, "Discovering fine-grained RRC state dynamics and performance impacts in cellular networks," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Sep. 2014, pp. 177–188.
[69] E. Gelenbe and Y. Cao, "Autonomous search for mines," *Eur. J. Oper. Res.*, vol. 108, no. 2, pp. 319–333, Jul. 1998.
[70] E. Gelenbe and Z. Kazhmaganbetova, "Cognitive packet network for bilateral asymmetric connections," *IEEE Trans. Ind. Informat.*, vol. 10, no. 3, pp. 1717–1725, Aug. 2014.
[71] E. Gelenbe, "The first decade of G-networks," *Eur. J. Oper. Res.*, vol. 126, no. 2, pp. 231–232, 2000.

GOKCE GORBIL received the Ph.D. degree in electrical and electronic engineering from Imperial College London, London, U.K., in 2013, where he is currently a Research Associate working in the areas of mobile network security and cloud computing. He is an Organizing Committee Member of the ISCIS'15 Conference, and a Technical Program Committee Member of the IEEE ISSNIP'15 Conference and the IEEE PerNEM'15 Workshop. His research interests include wireless and mobile networks, distributed systems, cloud computing, modeling and simulation of computing systems and networks, and network security.

OMER H. ABDELRAHMAN (M'14) received the B.Sc. degree in electrical and electronic engineering from the University of Khartoum, Khartoum, Sudan, in 2005, and the M.Sc. degree in communications and signal processing and the Ph.D. degree in computer networks from Imperial College London, London, U.K., in 2007 and 2012, respectively, where he is currently a Research Associate with the Intelligent Systems and Networks Group. His research interests include stochastic analysis and queuing theory, search techniques in random environments, and network security.



MIHAJLO PAVLOSKI received the B.Sc. degree in telecommunications and the M.Sc. degree in wireless and mobile communications from the Saints Cyril and Methodius University of Skopje, Skopje, Macedonia, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree in electrical and electronic engineering with Imperial College London, London, U.K. His research interests include queuing networks, statistical analysis, and machine learning.

EROL GELENBE (F'86) is currently the Dennis Gabor Chair Professor with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. He is a fellow of the French National Academy of Engineering, and the Science Academies of Hungary, Poland, and Turkey. He is an expert on the performance and security of large-scale computer and network systems. He was born in Istanbul, Turkey, and graduated from Ted Ankara Koleji, Ankara, Turkey. He received the B.Sc. (Hons.) degree in electrical and electronic engineering from Middle East Technical University, Ankara, and the M.Sc. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of New York University, Brooklyn, NY, USA. He joined the University of Michigan, Ann Arbor, MI, USA, as an Assistant Professor. In 1972, he joined the French Institute for Research in Computer Science and Automation (INRIA), Paris, France, where he established the Modeling and Performance Evaluation of Computer Systems and Networks Research Group, which is still today one of INRIA's strongest research areas. He received the Doctorat d'Etat degree in mathematical sciences from Université Pierre et Marie Curie, Paris, in 1973. He was appointed as the Chaired Professor of Computer Science with the University of Liege, Liège, Belgium, in 1974. In 1979, he became a Professor of Computer Science with Université Paris-Sud, Orsay, France, while continuing his association with INRIA, and served as a Lecturer in Applied Mathematics with École Polytechnique, Paris. From 1984 to 1986, he was a Science and Technology Advisor to the Minister for Universities (France), and then moved to Université Paris V, Paris, where he started the Department of Computer Science. From 1993 onwards, he was on leave from the University of Paris, Paris, first as the Chaired Professor and the Department Head at Duke University, Durham, NC, USA, and then as the Director of the School of Electrical Engineering and Computer Science at the University of Central Florida, Orlando, FL, USA, and currently at Imperial College London since 2003. He is currently a Principal Investigator (PI) of the 2.9M Euro EU FP7 grant on Mobile Network Security. He is also PI of two grants (EPSRC and DSTL) regarding energy savings in ICT, and the EU FP7 grant PANACEA regarding resilient Cloud Computing with Imperial College London. His research addresses biologically inspired neural networks, gene regulatory networks, and certain aspects of statistical physics. His papers appear in the top journals in the fields of electrical engineering, computer science, applied mathematics, and physics. He is a fellow of the Association for Computing Machinery (ACM) and the Institution of Engineering and Technology (IET). He received the Science Prize of the Parlar Foundation, the Grand Prix France Telecom of the French Academy of Sciences, and the ACM SIGMETRICS Life-Time Achievement Award. His honors include the Chevalier de la Légion d'Honneur and the Officier de l' Ordre du Mérite (France). He is the Commander of Merit and Grand Officer of the Order of the Star of Italy.