Impact of Signaling Storms on Energy Consumption and Latency of LTE User Equipment

Frederic Francois, Omer H. Abdelrahman and Erol Gelenbe
Intelligent Systems and Networks
Department of Electrical and Electronic Engineering
Imperial College, London SW7 2BT, UK
Email: {f.francois, o.abd06, e.gelenbe}@imperial.ac.uk

***Abstract*—Signaling storms in mobile networks, which congest the control plane, are becoming more frequent and severe because misbehaving applications can nowadays spread more rapidly due to the popularity of application marketplaces for smartphones. While previous work on signaling storms consider the processing overhead in the network and energy consumption of the misbehaving User Equipment (UE) only, this paper aims to investigate how signaling storms affect both the energy consumption and bandwidth allocation of normal and misbehaving LTE UEs by constructing a mathematical model which captures the interaction between the UE traffic and the Radio Resource Control state machine and bandwidth allocation mechanism at the eNodeB. Our results show that even if only a small proportion of the UE population is misbehaving, the energy consumption of the radio subsystem of the normal UEs can increase significantly while the time spent actively communicating increases drastically for a normal data session. Moreover, we show that misbehaving UEs have to spend an increasing amount of energy to attack the network when the severity of the signaling storms increases since they also suffer from the attacks.**

***Keywords*-Signaling Storms, LTE, Energy Consumption, Bandwidth Congestion, Radio Resource Control, Misbehaving Mobile Applications, M2M systems, IoT, 5G.**

I. Introduction

Signaling storms [1], [2] are becoming more frequent and severe due to the large number of smartphones which have access to marketplaces where users can download both malicious [3] and non-malicious applications [4] which adversely interact with the control plane of mobile networks to cause signaling storms. Additionally, signaling storms can also originate from the public Internet due to many mobile operators giving public IP addresses to their users without any appropriate protection [5] and non-optimized machine-to-machine (M2M) mobile communication [6]. In this paper, we investigate how a particular type of signaling storms, which leverages the Radio Resource Control (RRC) state machine of a User Equipment (UE) to cause excessive signaling to be exchanged on the control plane of mobile networks, can cause increased energy consumption of both normal and misbehaving UEs, severe data plane (i.e. bandwidth) congestion and excessive signaling overhead at the eNodeB. This holistic investigation of the different impacts that signaling storms can have is different from previous work which either investigate the impact of different RRC-based attacks on the mobile network in terms of signaling overhead and delay only [7]–[11] or the impact of traffic behavior on the energy consumption of the misbehaving UEs only without considering other UEs [12]–[15].

In LTE mobile networks, a UE has an energy saving mechanism in the form of a RRC state machine at the eNodeB which puts the UE in *IDLE* (I) mode with low energy consumption when it is not communicating and in *CONNECTED* (C) mode when the UE needs to communicate with the eNodeB. Each transition between the 2 modes involves the exchange of numerous signaling messages between the UE and the mobile network. A RRC-based signaling storm involves a large number of UEs cycling continuously between the 2 RRC modes to cause excessive signaling messages to be exchanged between the different components of the mobile network, including the UEs. Moreover, LTE UEs have the ability to go into short and long discontinuous reception (DRX) state when in the RRC C mode in order to save energy by sleeping most of the time and intermittently waking up to check if there is data to be communicated. The difference between the short and long DRX is that a UE in long DRX sleeps for a longer period of time between wake-up events.

II. Related Work

Signaling storms in mobile networks have been studied in the literature from three main perspectives: types of signaling storms and their associated processing overhead [7]–[11], power consumption of misbehaving UEs [12]–[14] and detection and mitigation of signaling storms [7], [16].

Several existing research papers [12]–[14] have investigated the energy consumption of UEs due to different application traffic patterns which can also lead to signaling storms in LTE networks. In [12], the authors model the DRX mechanism using a semi-Markov chain to obtain the trade-off between power consumption and various DRX parameters such as timeout, for bursty packet data traffic. The analytical results were also verified against simulations. [13] also investigates the same impact factors of

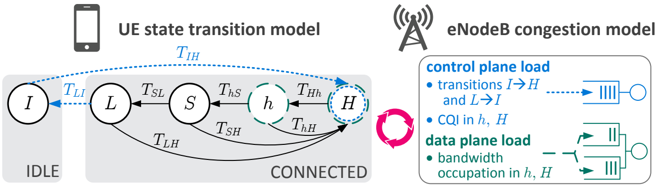


Figure 1. State transition diagram representing the behavior of a UE in the different LTE RRC states (left) and its interactions with the data and control planes of eNodeB (right). T_{XY} denotes the transition delay from state X to state Y; different states will also correspond to different power consumption levels of the radio subsystem.

the DRX mechanism but it also takes into account the various signaling messages that are exchanged during RRC mode transitions. The most important contribution of [14] is the measurement of the power consumption of LTE UEs in different operational networks around the world during different RRC and DRX states, which we use in this work. In addition, the authors of [14] infer the different DRX parameters used by operators from their power and traffic measurements which they then use to build a power model for a LTE UE so that they can compare the power and delay performance of a LTE UE against a 3G and WIMAX UE.

Our previous work on signaling storms in the context of the NEMESYS project [2], [17] has involved the mathematical modeling, simulation and analysis of the impact of different RRC-based signaling storms in 3G/UMTS networks [9]–[11]. In our recent work, we also investigated methods for the detection and mitigation of signaling storms through the use of RRC timer's adjustment and counters [16].

Our main contribution in this paper is the investigation of the adverse effect of signaling storms on the energy consumption of normal UEs during a data session, rather than focusing only on misbehaving UEs. We also obtain results for the average connection setup and data transfer times, taking into account congestion in both the control plane and data plane.

III. The Model

A. Energy Model of the UE's Radio Subsystem during One Data Session

We use the state transition diagram in Fig. 1 to model the behavior of a LTE UE in terms of the energy consumption of the radio subsystem for different types of communications, e.g. Voice over LTE (VoLTE), instant messaging and web browsing. The diagram shows the transitions that a UE can make inside the RRC state machine [18] during a data session.

RRC in LTE differs from 3G [9] in that the UE may either be in I or C mode, and transitions between these 2 modes will cause signaling messages to be exchanged between the LTE UE and the mobile network [18]. In the C mode, the radio subsystem of a LTE UE has four decreasing power states: H, h, S and L as shown in Fig. 1 where:

- In the H state the UE is actively communicating with the eNodeB.
- In the h state the UE has finished communicating and if new data is received before a certain timeout T_h, the UE will become active again (i.e. be in H) otherwise the UE will be moved to state S.
- The short DRX state S corresponds to a UE which is in short intermittent reception mode, where it sleeps most of the time and wakes up periodically to check whether data is buffered at the eNodeB for its reception or it needs to send data to eNodeB and if so, the UE moves to state H. If no data is communicated by the UE for a certain timeout T_S, the UE will be moved to state L.
- The long discontinuous reception mode L is similar to S except that the UE wakes up less frequently and if the UE does not communicate within T_L seconds, it finally enters the idle mode I.

Through the use of either statistical observations or mathematical models, the following UE behavior can be captured:

- The *average relative number of times* $p_{X,Y}$ that a state Y is visited from state X, where $X, Y \in \{I, H, h, S, L\}$, $X \neq Y$.
- The average time τ_X spent in state X each time it is visited.
- The average time T_{XY} (shown in Fig. 1) that it takes to accomplish the transition from state X to state Y.
- The power consumption of the radio subsystem of the UE during each state occupation and state transition, which are denoted Π_X and Π_{XY}, respectively, for $X, Y \in \{I, H, h, S, L\}$. Note that the radio subsystem of a UE consumes more power when transmitting compared to when receiving data [19] since transmission to the eNodeB involves sending energy over the radio waves; this is addressed in Section III-D.

From the information above, the total average energy consumption of the radio subsystem of a UE for each complete cycle that starts when the UE exits state I and ends at its next departure from I (refereed to here as a *data session*) can be expressed as follows:

$$E = \sum_{X \in \{I,H,h,S,L\}} n_X \tau_X \Pi_X + \sum_{X,Y \in \{I,H,h,S,L\}} n_X p_{X,Y} T_{X,Y} \Pi_{X,Y} \quad (1)$$

where n_X denotes the total number of visits to state X during the cycle, which can be calculated from the relative frequencies $p_{X,Y}$ using the system of linear equations:

$$n_Y = \sum_{X,Y \in \{I,H,h,S,L\}} n_X p_{X,Y} \quad (2)$$

Notice that $n_I = 1$ since we are considering a single data session only, and $p_{IH} = p_{Hh} = 1$ because there is only one possible transition out of I and H. In order to obtain the remaining n_X values in terms of the relative frequencies, Eq. (2) and the state transition diagram in Fig. 1 can be used leading to:

$$n_H = n_h = \frac{1}{p_{hS}p_{SL}p_{LI}},\ n_S = \frac{1}{p_{SL}p_{LI}},\ n_L = \frac{1}{p_{LI}} \quad (3)$$

Since energy is also consumed when moving between power states, it is necessary to calculate the number of times n_{XY} that a single UE makes the transition from X to Y during one data session, which can be calculated from the relationship $n_{XY} = n_X p_{XY}$ and using (3):

$$\begin{aligned} &n_{IH} = 1, \qquad n_{hH} = \frac{p_{hH}}{p_{hS}p_{SI}p_{LI}}, \quad n_{SH} = \frac{p_{SH}}{p_{SL}p_{LI}}, \\ &n_{LH} = \frac{p_{LH}}{p_{LI}}, \quad n_{Hh} = \frac{1}{p_{hS}p_{SL}p_{LI}}, \quad n_{hS} = \frac{1}{p_{SL}p_{LI}}, \\ &n_{SL} = \frac{1}{p_{LI}}, \qquad n_{LI} = 1 \end{aligned} \quad (4)$$

Now the rate λ_n at which a *normal* UE makes data sessions is simply the inverse of the average session duration, i.e.:

$$\lambda_n^{-1} = \sum_{X \in \{I,H,h,S,L\}} n_X \tau_X + \sum_{X,Y \in \{I,H,h,S,L\}} n_X p_{XY} T_{XY} \quad (5)$$

During signaling storms, the time τ_H that a UE spends in each visit to state H will be affected since the number of UEs in states H or h which are connected to the same eNodeB determines the bandwidth allocated to a UE for communication; thus, signaling storms may lead to congestion in the data plane. On the other hand, control plane congestion affects mainly T_{IH} and T_{LI} because significant signaling is involved in establishing and releasing RRC connections; the other transition times T_{hH}, T_{Hh}, T_{hS}, T_{SL}, T_{SH} and T_{LH} involve negligible signaling between the UE and eNodeB and are therefore marginally affected by the attack. This coupling between the state transitions of UEs and signalling and bandwidth congestion at the eNodeB is illustrated in Fig. 1.

B. Signaling Behavior of Misbehaving UEs

We assume that a misbehaving UE will cause excessive signaling through a continuous cycle of promotions $I \rightarrow H$ and demotions $L \rightarrow I$ which is triggered by sending a small amount of data when the UE is in I and waiting for demotion back to I after undergoing the timeouts. The rate λ_a at which a misbehaving UE can attack is the inverse of the cycle duration:

$$\lambda_a^{-1} = \sum_{X \in \{h,S,L\}} T_X + \sum_{XY \in \{IH,Hh,hS,SL,LI\}} T_{XY} + \tau_I^A + \tau_H^A \quad (6)$$

where the sum of the timeout periods T_X and the transition times T_{XY} represents the minimum duration that a misbehaving UE must spend before returning to I. We assume that misbehaving UEs spend negligible amount of time $\tau_I^A \simeq 0$ in I during the storm, and τ_H^A is the amount of time a misbehaving UE spends in H, typically to send a small amount of data to trigger a promotion from I to H.

C. Control Plane Congestion at eNodeB

Following our approach in [9], [11] for 3G networks, the times T_{IH} and T_{LI} that it takes for a single LTE UE to promote from I to H and to demote from L to I depend on the signaling queueing delay w_e at the eNodeB, and can be characterized by the following expression:

$$T_{XY} = r_{XY} w_e + \sum_{n=1}^{r_{XY}} (t_{XY}[n] + \delta_{XY}[n]), \quad XY \in \{IH, LI\} \quad (7)$$

where r_{XY} is the number of messages sent between the UE and eNodeB (given in Table I) during the transition $X \rightarrow Y$ [8], [18], and $t_{XY}[n]$ and $\delta_{XY}[n]$ are the propagation delay and processing time of the n-th signaling message, respectively.

The signaling queueing delay w_e at the eNodeB is calculated as follows: we first obtain the rate of signaling messages generated by a single normal UE on the eNodeB:

$$\gamma_e^N = \lambda_n [r_{IH} + r_{LI} + n_H \lambda_c (\tau_H^N + \tau_h)] \quad (8)$$

where λ_n, given in (5), is the rate at which a normal UE makes data sessions, and λ_c is the rate at which a UE sends Channel Quality Information (CQI) messages while in H and h [20] so that $n_H \lambda_c (\tau_H^N + \tau_h)$ is the total number of CQI messages sent within a session. Similarly, the total signaling rate of a misbehaving UE during the storm is given by:

$$\gamma_e^A = \lambda_a [r_{IH} + r_{LI} + \lambda_c (\tau_H^A + T_h)] \quad (9)$$

From (8) and (9), it is possible to obtain the total rate of signaling messages at the eNodeB due to a population of β UEs connected to the eNodeB as follows:

$$\Gamma_e = (1 - Z)\beta\gamma_e^N + Z\beta\gamma_e^A \quad (10)$$

where Z is the effective proportion of misbehaving UEs in the population which are sending only misbehaving traffic and no normal traffic.

The average signaling queueing delay w_e at the eNodeB can be obtained by approximating the signaling server at the eNodeB by a $M/M/1$ system:

$$w_e = \frac{1}{v - \Gamma_e} \quad (11)$$

Here v is an "equivalent" service rate which depends on the composition of the signaling messages processed by the eNodeB and is given by:

$$\begin{aligned} v^{-1} = \Gamma_e^{-1} \Big[&\beta Z \lambda_a \Big\{ \sum_{XY} \sum_{n=1}^{r_{XY}} \delta_{XY}[n] + \lambda_c (\tau_H^A + T_h) \delta_c \Big\} \\ &+ \beta (1 - Z) \lambda_n \Big\{ \sum_{XY} \sum_{n=1}^{r_{XY}} \delta_{XY}[n] + n_H \lambda_c (\tau_H^N + \tau_h) \delta_c \Big\} \Big] \end{aligned} \quad (12)$$

where $XY \in \{IH, LI\}$ and δ_c is the processing time of the CQI messages at the eNodeB.

D. Data Plane Congestion at eNodeB

We now proceed with obtaining the time and energy that a UE spends in state H, which require modeling the bandwidth congestion at the data plane between the UE and the serving eNodeB in both the uplink and downlink directions. This type of congestion occurs because there is a limited amount of bandwidth which is shared among all UEs that are attached to the eNodeB and in state H or h. When a UE is in H, it actually uses the bandwidth which is allocated to it by the eNodeB, while when the UE is in h, it has a reserved bandwidth but does not use it for active communication. Note however that the allocated bandwidth for a UE in h might be less than that when it is in H depending on the bandwidth *allocation and retention priority* policy of the mobile operator.

During signaling storms, the misbehaving UEs will first occupy state H and subsequently state h after sending or receiving only a small amount of data which triggers a transition from I to H and causes signaling messages to be exchanged with the eNodeB. In most scenarios, the misbehaving traffic comprises of keep-alive messages, instant messaging, or M2M communications, which are typically very small. On the other hand, deliberate attackers are likely to send only a small amount of packets in order to preserve their network usage quota and minimize their network data charges. Therefore, misbehaving UEs will reserve some bandwidth that is not used for any useful communication for the timeout period T_h. This wasteful reservation of bandwidth by the misbehaving UEs will result in less bandwidth being available for the normal UEs in H and therefore, the amount of time τ_H^N that the normal UE spends in H per visit will increase. In turn, the number of CQI messages sent to the eNodeB by the normal UEs will also increase according to Eq. (8) and as a result the severity of the signaling storm may increase.

A bandwidth congestion model can be developed based on the following generic characteristics of mobile networks:

- The total uplink and downlink bandwidth of an eNodeB are different and are shared between the UEs in states H and h according to the *allocation and retention priority* (ARP) of the mobile operator.
- The bandwidth occupation in H is different from that in h.
- The traffic flow between a UE and its serving eNodeB can be asymmetric, e.g. as in web browsing and video streaming.

Furthermore, we can make the following assumptions to simplify the analysis:

- The data link between a UE and eNodeB is shared using time division duplexing.
- Requests for data communication are assumed to be made up of a fixed proportion of uplink U and downlink $1-U$ traffic.
- There are two types of bandwidth requests that UEs make during visits to H: (i) normal bandwidth requests with D_N amount of data to transfer on average, and (ii) misbehaving bandwidth requests with average data transfer size of D_A, with $D_A << D_N$.
- The total data bandwidth B_e of the eNodeB (both uplink and downlink) is shared equally among all the UEs which are currently in states H and h.

The bandwidth sharing assumption above is accurate enough to capture the different effects of signaling storms on UEs but it should be noted that in an operational mobile network, users usually obtain different data rates depending on various factors such as the specific ARP used by the network operator, the Quality of Service Class Identifier of data bearers and the agreed data plan with the network operator. It is left for future work to investigate if a more complex bandwidth sharing model is necessary to fully characterize the effect of signaling storms on different types of normal users.

Based on the above assumption, an egalitarian processor sharing model [21], [22] with two classes of bandwidth requests, each requiring a different service demand, can be used to represent bandwidth congestion at the eNodeB. Specifically, the service demand μ_i of a bandwidth request of type $i \in \{N, A\}$, measured in seconds, consists of two components: (i) the time it takes to actually transfer D_i bytes of data, and (ii) the time that the UE waits after active communication while reserving bandwidth, before either going into short DRX after timeout or going back into H because new data is available before timeout. Hence, the service demand of bandwidth requests can be calculated as:

$$\mu_N = \frac{D_N}{B_e} + \tau_h, \quad \mu_A = \frac{D_A}{B_e} + T_h \tag{13}$$

The average number of concurrent normal and misbehaving bandwidth requests in an egalitarian processor sharing system are [21], [22]:

$$G_N = \frac{\rho_N}{1-\rho_A-\rho_N}, \quad G_A = \frac{\rho_A}{1-\rho_A-\rho_N} \tag{14}$$

where $\rho_i = \lambda_i\mu_i$, with λ_i denoting the total rate of bandwidth requests of type i:

$$\lambda_N = (1-Z)\beta n_H \lambda_n, \quad \lambda_A = Z\beta\lambda_a \tag{15}$$

Note that data plane congestion occurs when $\rho_N + \rho_A$ is close to 1. Using Little's theorem, the time that a UE of type i spends in *both* H and h per visit is $\frac{G_i}{\lambda_i}$, from which the times spent in each visit to H become:

$$\tau_H^N = \frac{G_N}{\lambda_N} - \tau_h, \quad \tau_H^A = \frac{G_A}{\lambda_A} - T_h \tag{16}$$

The energy spent by the radio subsystem of a UE per visit to H is then given by:

$$E_H^i = \tau_H^i[\Pi_h + U\frac{D_i}{\tau_H^i}\pi_H^{tx} + (1-U)\frac{D_i}{\tau_H^i}\pi_H^{rx}] \tag{17}$$

where power consumption is assumed to follow a linear relationship with throughput based on the experimental evaluation in [14]. Here Π_h, the intercept, is the power consumed when the radio subsystem is awake but not transmitting or receiving data, while π_H^{tx} (resp. π_H^{rx}) is the increase in power consumption when the bit rate increases by 1Mbps in the uplink (resp. downlink). Notice that since we only consider the average throughput and energy consumption, τ_H^i remains only in the intercept of (17).

E. Model of the Traffic behavior of Normal UEs

In this section, we describe a simple user traffic model which allows us to estimate the transition probabilities p_{XY} and the average time τ_X spent in each visit to state $X \neq H$ of Fig. 1. We model the data transfer between the UE and eNodeB as traffic bursts of average size D_N bytes, with the idle time between traffic bursts (i.e. time when there is no data plane communication between the UE and eNodeB) being a random variable Ω which is independent of the control and data plane congestion. Thus we assume that users spend random time consuming the data received (e.g. reading a web page) before making further requests, independently of the time taken to download the data. In future work, we will also consider a model which captures the changes in user behavior due to degraded quality of service during signaling storms.

We first obtain the time $\boldsymbol{\tau_X}$ that a UE spends in state X between two consecutive traffic bursts, by modeling the LTE DRX mechanism as follows: one short DRX period is divided into N_s cycles of duration t_s, which consists of a t_s^{on} period followed by a t_s^{off} period, and let $t_s = t_s^{on} + t_s^{off}$. During t_s^{on}, the UE wakes up to send uplink traffic or to listen for notification from the eNodeB about any buffered downlink data available for the UE to receive; t_s^{off} is the period during which the UE's radio subsystem is totally asleep. The time $\boldsymbol{\tau_S}$ spent in the short DRX state S between two consecutive traffic bursts takes the following form:

$$\boldsymbol{\tau_S} = \begin{cases} 0, & \text{if } \ \Omega \leq T_h, \\ it_s, & \text{if } \ \Omega > T_h + (i-1)t_s + t_s^{on} \\ & \quad \wedge \ \Omega \leq T_h + it_s, \\ (i-1)t_s + t_s^{on}, & \text{if } \ \Omega > T_h + (i-1)t_s \\ & \quad \wedge \ \Omega \leq T_h + (i-1)t_s + t_s^{on}, \\ N_s t_s, & \text{if } \ \Omega > T_h + T_S \end{cases} \tag{18}$$

where in the first case, the inter-arrival time Ω between traffic bursts is smaller than the timeout of state h, so that the UE spends no time in S. The second case illustrates the scenario where the next traffic burst happens during the t_s^{off} period of the $(i-1)^{th}$ DRX cycle and therefore the UE will receive the traffic during the t_s^{on} period of the next cycle and will spend it_s time in state S, where $i = 1, \ldots, N_s$. In the third case, the next traffic burst happens during the t_s^{on} period of the $(i-1)^{th}$ DRX cycle and therefore the UE will send or receive immediately and will spend a maximum of $(i-1)t_s + t_s^{on}$ time in state S. The last case corresponds to the scenario where the next traffic burst occurs in time greater than $T_h + T_S$ and as a result the UE spends $N_s t_s$ time in state S which is equal to T_S.

Similarly to the short DRX period, one long DRX period is divided into N_l cycles of duration t_l which is further divided into one t_l^{on} and one t_l^{off} period. Usually the network operator configure the UE to have $t_l^{on} = t_s^{on}$ and $t_l^{off} >> t_s^{off}$ so that less energy is consumed by a UE during a long DRX cycle compared to a short one, since the UE is spending more time sleeping in the former cycle. Thus, we can compute $\boldsymbol{\tau_L}$ in a similar manner as for the short DRX:

$$\boldsymbol{\tau_L} = \begin{cases} 0, & \text{if } \Omega \leq T_h + T_S, \\ it_l, & \text{if } \Omega > T_h + T_S + (i-1)t_l + t_l^{on} \\ & \wedge \ \Omega \leq T_h + T_S + it_l, \\ (i-1)t_l + t_l^{on}, & \text{if } \Omega > T_h + T_S + (i-1)t_l \\ & \wedge \ \Omega \leq T_h + T_S + (i-1)t_l + t_l^{on}, \\ N_l t_l, & \text{if } \Omega > T_h + T_S + T_L \end{cases} \tag{19}$$

Next, the time spent in h between consecutive traffic bursts is:

$$\boldsymbol{\tau_h} = \min\left(\Omega, T_h\right) \tag{20}$$

and it is also straightforward to compute $\boldsymbol{\tau_I}$:

$$\boldsymbol{\tau_I} = \max(0, \Omega - T_h - T_S - T_L) \tag{21}$$

The average time τ_X that a UE spends in state X during each visit to that state can now be obtained as follows:

$$\begin{aligned} \tau_S &= \mathbb{E}[\boldsymbol{\tau_S} | \Omega > T_h] \\ \tau_L &= \mathbb{E}[\boldsymbol{\tau_L} | \Omega > T_h + T_S] \\ \tau_h &= \mathbb{E}[\boldsymbol{\tau_h}] \\ \tau_I &= \mathbb{E}[\boldsymbol{\tau_I} | \Omega > T_h + T_S + T_L] \end{aligned} \tag{22}$$

Furthermore, the transition probabilities p_{XY} become:

$$\begin{aligned} p_{hS} &= \Pr(\Omega > T_h), & p_{hH} &= 1 - p_{hS}, \\ p_{SH} &= \Pr(\Omega < T_h + T_S | \Omega \geq T_h), & p_{SL} &= 1 - p_{SH}, \\ p_{LH} &= \Pr(\Omega < T_h + T_S + T_L | \Omega \geq T_h + T_S), & & \\ & & p_{LI} &= 1 - p_{LH} \end{aligned} \tag{23}$$

A quality of service metric of interest for normal users is the average time to complete transmission or reception of a single traffic burst, which is the sum of two delay components: (i) the transition delay from any of the states $\{I, h, S, L\}$ to state H where communication takes place, and (ii) the actual communication time τ_H^N. This metric captures the average response time of the network, and is given by:

$$R = \frac{\sum\limits_{XY \in \{IH, hH, SH, LH\}} n_{XY} T_{XY}}{\sum\limits_{XY \in \{IH, hH, SH, LH\}} n_{XY}} + \tau_H^N \qquad (24)$$

Note that the transition time T_{hH} is negligible since the radio subsystem is already fully awake in state h, while T_{IH} and T_{LI} depend on signalling load and were derived in (7). Thus, it remains to calculate $T_{SH} \in [0, t_s^{off}]$ and $T_{LH} \in [0, t_l^{off}]$ which represent the average durations from the arrival instant of a traffic burst in S and L to the time when the radio subsystem is scheduled to wake up. A simple approximation for these transition times would be to assume that the traffic bursts occur uniformly at random during sleep periods so that:

$$T_{SH} \approx \frac{t_s^{off}}{2}, \qquad T_{LH} \approx \frac{t_l^{off}}{2} \qquad (25)$$

We can follow the approach used for τ_X to compute more accurate estimates of these transition times, which will still be an approximation since we neglect the small effect of signaling overload. Moreover, the worst-case values of these transition times are much smaller than the transition delays between I and C, and so (25) should provide a reasonable approximation.

IV. Numerical Results

A. Parameters for the Model

Table I shows the numerical values which have been chosen for each parameter of the mathematical model during the evaluation. For calculating the energy consumption of the radio subsystem of UEs, the values of the power levels Π_X were taken from [14], and the power consumed during promotion from I to H was also measured to be $\Pi_{IH} = 1.21W$. We assumed that the same power is consumed by the radio subsystem when it is transitioning back from L to I. For all the other transitions between power states, we used the approximation $\Pi_{XY} = 0.5(\Pi_X + \Pi_Y)$.

Note that we chose Ω to follow an exponential distribution with mean $60s$ but any other distribution could be used to obtain numerical results. In the sequel, we illustrate the impact of signaling storms first on the eNodeB and then on the UEs, because it is through the increase in the control and data plane loads of the eNodeB that misbehaving UEs can have adverse effects on the normal UEs.

B. Effects of Signaling Storms at eNodeB

Fig. 2(a) shows that signaling load at the eNodeB increases rapidly when the proportion Z of misbehaving UEs increases in the population of UEs connected to the eNodeB. A larger proportion of misbehaving UEs means that the eNodeB will receive a larger number of signaling messages to process. The maximum signaling load occurs when Z is around 0.28, where the eNodeB reaches its maximum processing capacity. The increase in signaling rate also increases the signaling queueing delay at the eNodeB, which results in the time T_{IH} needed for a single UE to promote from I to H (i.e. access latency of the UE) to increase as shown in Fig. 2(b).

Table I
General parameters for LTE network

Parameter	Value
Total no. of UEs, β	175
Total no. of eNode-Bs	1
Rate of CQI messages at eNodeB, λ_c	$100s^{-1}$
Processing time of CQI at eNodeB, δ_c	$2ms$
Inter-arrival time distribution of normal traffic bursts	Exponential
Mean inter-arrival time of normal traffic bursts, $\mathbb{E}[\Omega]$	$60s$
Total bandwidth at eNodeB, B_e	150Mbps
Average size of normal traffic requests, D_N	1.88MB
Average size of misbehaving traffic requests, D_A	10KB
Timeout for demotion from h, T_h	$100ms$
Timeout for demotion from *short DRX* state, T_S	$200ms$
Timeout for demotion from *long DRX* state, T_L	$10s$
Duration of a short *DRX* cycle, t_s	$20ms$
Duration of a long *DRX* cycle, t_l	$40ms$
Duration of "on" period during a *DRX* cycle, t^{on}	$1ms$
Proportion of user uplink traffic, U	0.3
No. of signaling messages exchanged during $I \to H$ promotion, r_{IH}	16
Average processing time of each $I \to H$ promotion message at eNodeB, δ_{IH}	$3ms$
No. of signaling messages exchanged during $L \to I$ demotion, r_{LI}	5
Average processing time of each $L \to I$ demotion messages at eNodeB, δ_{LI}	$3ms$
Power consumed in state I, Π_I	0.031W
Power consumed in state L, Π_L	1.076W
Power consumed in state S, Π_S	1.091W
Power consumed in state h, Π_h	1.29W
Power consumed when transmitting at 1Mbps, π_H^{tx}	0.438W/Mbps
Power consumed when receiving at 1Mbps, π_H^{rx}	0.052W/Mbps

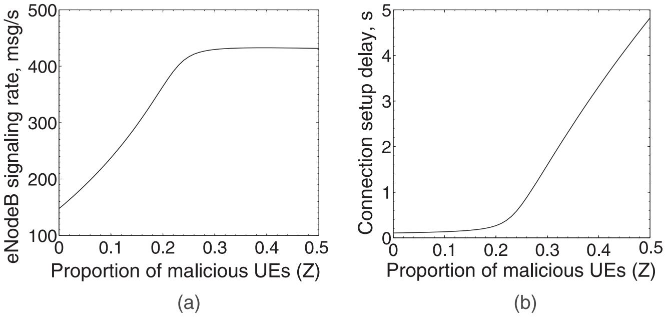


Figure 2. (a) Signaling load Γ_e and (b) promotion delay T_{IH} from I to H at the eNodeB, when the proportion Z of misbehaving UEs increases.

Fig. 3 shows that the number of concurrent malicious bandwidth requests G_A being served at the eNodeB grows rapidly when Z is initially increased before growing at a lower rate when congestion in the control plane of the eNodeB becomes more severe, effectively reducing the rate λ_a at which misbehaving data connections can be made. The number of concurrent normal bandwidth requests G_N also

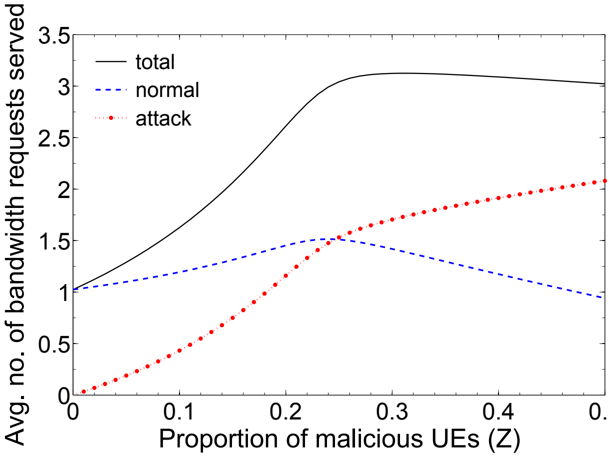


Figure 3. Average number of normal G_N, misbehaving G_A and total bandwidth requests served at the eNodeB, when the proportion Z of misbehaving UEs increases.

initially increases with Z due to the bandwidth congestion at the eNodeB created by the presence of an increasing number of misbehaving UEs making malicious bandwidth requests. The bandwidth congestion at the eNodeB results in the normal UEs spending more time τ_H^N to complete a normal bandwidth request as shown in Fig. 4. When Z exceeds 24%, we notice that signaling congestion causes G_N to drop with further increase in Z, because normal UEs find it more difficult to access the network in order to communicate. Thus in this scenario performance is dominated by bandwidth congestion when $Z < 24\%$, and it is dominated by signaling congestion when $Z > 24\%$; otherwise G_N would have continued to grow if the control plane was not overloaded.

C. Effects of Signaling Storms on Normal and Misbehaving UEs

In Fig. 4, we plot the average response time R from (24) along with its two delay components: connection setup delay and communication time. We see that the average communication time τ_H^N initially increases with Z during the data-plane congestion phase of the eNodeB, before starting to drop and level off as congestion in the control plane becomes more severe and fewer users are able to access the network, as shown previously in Fig. 3. However, although the attack rate slows down during signaling overload, the average connection setup time increases rapidly, causing the UEs to experience severe signaling delay when transitioning from I to H, and also higher response times.

Finally, Fig. 5(a) shows that the energy consumed by the radio subsystem of a normal UE to complete one data session increases rapidly with increasing Z; therefore, signaling storms do not only have a negative impact on the access time of normal users but also on their energy consumption. Note that from Fig. 1, there are four variables which affect energy consumption and change with

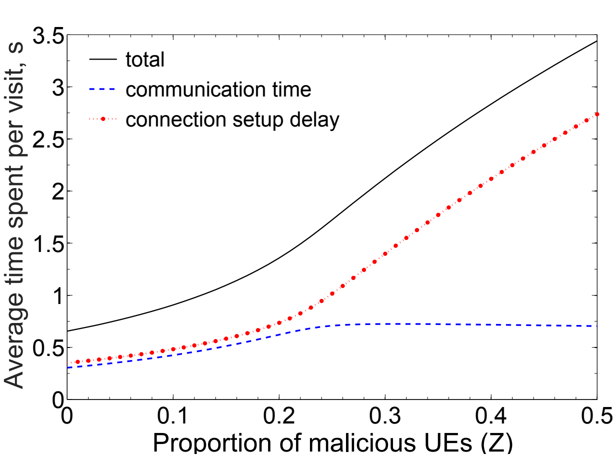
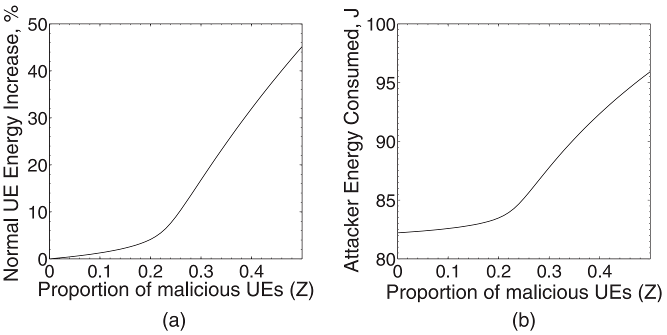


Figure 4. The average time time spent: communicating a traffic burst T_C, per visit to H and transitioning from another state X to H for a normal UE when the proportion Z of misbehaving UEs increases.



(a)

(b)

Figure 5. The percentage increase in energy consumption of radio subsystem of a normal UE for one data session when Z increases (left), and the corresponding energy consumed by a misbehaving UE (right). The energy consumed by a normal UE in the absence of attack is $17.9J$.

Z: τ_H, τ_h—the times spent in H, h, and T_{IH}, T_{LI}—the connection setup and release times. Energy increases more quickly during the later signaling congestion phase, since more energy is needed to perform the transitions between I and C. Fig. 5(b) shows that the energy consumed by the radio subsystem of a misbehaving UE during one normal data session increases with Z, because similar to the normal data sessions, the misbehaving UEs suffer from increased bandwidth and signaling congestion, and so they require more energy in order to setup and release connections.

V. Conclusions and Future Work

This paper demonstrates that signaling storms do not only affect the waiting time that normal users experience before accessing the network but also significantly increase the energy consumption of the radio subsystem and the time it takes for a normal UE to communicate the same amount of data after gaining access to the network. These important results were obtained through three coupled mathematical models which capture the adverse effects of signaling storms on the control and data plane of LTE mobile networks including UEs. Future generations of mobile networks, such as the upcoming 5G, need to take into account that small and frequent data transmissions are going to be a more prevalent activity in mobile networks due to the increasing use of mobile networks for M2M and Internet of Things communication and hence, future mobile networks should be designed to be resilient to such network activity. Future work will involve extending the analysis to include changes in user behavior due to degraded quality of service. We will also investigate the impact of signaling storms on the energy consumption of the core components of LTE networks, and how they can be modified with minimum cost to detect and mitigate signaling storms.

VI. Acknowledgments

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

References

[1] C. Gabriel, "DoCoMo demands Google's help with signalling storm," Rethink Wireless, http://www.rethink-wireless.com/2012/01/30/docomo-demands-googles-signalling-storm.htm, Jun. 2012.

[2] O. H. Abdelrahman, E. Gelenbe, G. Gorbil, and B. Oklander, "Mobile network anomaly detection and mitigation: The NEMESYS approach," in *Proc. 28th Int'l Symp. on Computer and Information Sciences (ISCIS)*, ser. LNEE, E. Gelenbe and R. Lent, Eds. Springer, Oct. 2013, vol. 264, pp. 429–438.

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

[4] S. Corner, "Angry Birds + Android + ads = network overload," IT Wire, http://www.itwire.com/business-it-news/networking/47823, Jun. 2011.

[5] 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 Annual Network and Distributed System Security Symp. (NDSS)*, San Diego, CA, Feb. 2012.

[6] 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.

[7] P. P. 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.

[8] J. Puttonen, E. Virtej, I. Keskitalo, and E. Malkamaki, "On LTE performance trade-off between connected and idle states with always-on type applications," in *Proc. 23rd IEEE Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sydney, Australia, Sep. 2012, pp. 981–985.

[9] O. H. Abdelrahman and E. Gelenbe, "Signalling storms in 3G mobile networks," in *Proc. IEEE Int'l Conference on Communications (ICC)*, Sydney, Australia, Jun. 2014, pp. 1017–1022.

[10] G. Gorbil, O. H. Abdelrahman, and E. Gelenbe, "Storms in mobile networks," in *Proc. 10th ACM Symp. on QoS and Security for Wireless and Mobile Networks (Q2SWinet)*, Montreal, Canada, Sep. 2014, pp. 119–126.

[11] G. Gorbil, O. H. Abdelrahman, M. Pavloski, and E. Gelenbe, "Modeling and analysis of RRC-based signalling storms in 3G networks," *IEEE Trans. Emerg. Topics Comput.*, 2015.

[12] L. Zhou, H. Xu, H. Tian, Y. Gao, L. Du, and L. Chen, "Performance analysis of power saving mechanism with adjustable DRX cycles in 3GPP LTE," in *Proc. 68th IEEE Vehicular Technology Conf. (VTC Fall)*, Calgary, BC, Sep. 2008, pp. 1–5.

[13] C. S. Bontu and E. Illidge, "DRX mechanism for power saving in LTE," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 48–55, Jun. 2009.

[14] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *Proc. 10th ACM Int. Conf. on Mobile Systems, Applications, and Services (MobiSys)*, Lake District, UK, Jun. 2012, pp. 225–238.

[15] A. Castiglione, F. Palmieri, U. Fiore, A. Castiglione, and A. D. Santis, "Modeling energy-efficient secure communications in multi-mode wireless mobile devices," *J. Comput. System Sci.*, 2014, in press.

[16] E. Gelenbe and O. H. Abdelrahman, "Time-outs and counters against storms," 2014, submitted for publication.

[17] E. Gelenbe, G. Gorbil, D. Tzovaras, S. Liebergeld, D. Garcia, M. Baltatu, and G. Lyberopoulos, "Security for smart mobile networks: The NEMESYS approach," in *Proc. IEEE Global High Tech Congress on Electronics (GHTCE)*, Shenzhen, Nov. 2013, pp. 63–69.

[18] ETSI 3GPP, "LTE; evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification (3GPP TS 36.331 version 12.3.0 release 12)," Sep. 2014.

[19] S. Deng and H. Balakrishnan, "Traffic-aware techniques to reduce 3G/LTE wireless energy consumption," in *Proc. 8th ACM Int'l Conf. on Emerging Networking Experiments and Technologies (CoNEXT)*, Nice, France, 2012, pp. 181–192.

[20] ETSI 3GPP, "LTE; evolved universal terrestrial radio access (E-UTRA); physical layer procedures (3GPP TS 36.213 version 8.8.0 release 8)," Oct. 2010.

[21] F. P. Kelly, *Reversibility and Stochastic Networks*. Wiley, 1979.

[22] J. Cohen, "The multiple phase service network with generalized processor sharing," *Acta Informatica*, vol. 12, no. 3, pp. 245–284, 1979.