How an SMS-Based Malware Infection Will Get Throttled by the Wireless Link

Ilona Murynets
AT&T Services Inc.
Security Research Center
New York, NY 10007
Email: ilona@att.com

Roger Piqueras Jover
AT&T Services Inc.
Security Research Center
New York, NY 10007
Email: roger.jover@att.com

Abstract—As smart phones increase in popularity, they become an attractive target for attackers and spammers. This paper presents a new simulation model that evaluates the effects of an SMS-based malware infection in GSM and UMTS networks. It is the first known work that accounts for the wireless link of the network in modeling of malware propagation. The paper demonstrates propagation of the SMS-transmitted malware in a densely populated metropolitan area. It shows that spreading rate of cellular malware is tightly bounded by the actual network architecture and strongly diverges from the pattern seen in regular wired Internet-connected networks.

I. INTRODUCTION

The increasing popularity and computational resources of smart phones create a whole new set of network threats and open the doors to cyber attacks. Mobile phones are also the latest platform for malware spreading in data networks [18]. Either motivated by theft of service, spam propagation, elaborated distributed denial-of-service attacks or simple vandalism, both malicious applications and attacks to the mobility network have been reported in the media over the last year [5], [6], [4]. There has also been reports of attackers exploiting system vulnerabilities to jam and remotely control iPhones [8].

These problems have been analyzed extensively in the literature. Current studies on SMS(Short Message Service)-based malware propagation scenarios [9], [14] analyze the problem from different perspectives. On one hand, propagation through direct contact (phone to phone spreading on wireless local or personal area network) and Bluetooth outbreaks are investigated. In this case, the spreading of such malware is shown to be mild and easily controllable by lowering the susceptibility of the population to the infection. On the other hand, SMS-based infections spreading on cellular voice and data networks are analyzed. These studies conclude that such propagation can potentially saturate transmission channels and prevent users from legitimate communications. No published study considers yet the actual wireless hop of the network in the propagation modeling, though. Traynor et al. [13] stress that this first link in the access network is precisely the main bottleneck when it comes to massive short message distribution. Therefore, known models for SMS-based malware propagation might provide misleading results.

In this paper, we focus on the specific case of a malware infection spreading throughout the mobility network on both GSM and UMTS. We compare it against the propagation on a traditional wired Internet-connected network [26], usually described with models developed for human epidemics [2]. The new model presented in this paper provides more accurate results by including the essential wireless link component into the equation. Malware is assumed to propagate as a result of an infection on a smart phone running on a specific operating system. Upon reception of a short message from a known contact, a user is tricked into downloading a malicious application appearing to be harmless. This results into a slow initial infection rate that increases exponentially. However, the infection will be unable to propagate beyond a certain point due to the network’s limited resources.

The remainder of the paper is organized as follows: Section II discusses related work; Section III presents a general overview of the SMS network architecture as well as specific details on SMS transmission over the air interface; Section IV describes the network model used to analyze the propagation of an SMS-based infection. Details on the simulation methodology and procedure are discussed as well. Section V presents examples of an infection spreading on both GSM and UMTS networks, Finally, Section VI discusses the concluding remarks.

II. RELATED WORK

Mobile malware propagation has been and continues to be a very popular topic both in academic literature and media. Different malware-related problems with mobile phone applications, especially on Android platforms, have made it to the headlines [5], [6], [4]. It has also been observed that SMS is one of the preferred means for malware propagation and can result in an infection that allows attackers to take over a phone and remotely control it [8]. Fleizach et al. study in [9] the effects of a mass spreading of malware by means of Multimedia Messaging Services (MMS) and discuss the the slower propagation of an infection spreading in a mobility network with respect to a wired network. However, while a deep analysis is presented, the authors consider that a bottleneck of malicious SMS dissemination is located at the Short Messaging Service Center (SMSC). This node is
assumed to have a capacity for 100 multimedia messages per second for a total population of 5,621,336 users. This seems unrealistic for the current estimates of 187.7 billion text messages being sent monthly [7].

The authors in [9] also propose a model that considers mobile terminals connected to the network by means of links that assume signaling and control channel effects not significant when compared to packetized bandwidth limits. Under this assumption, the link between the Base Station and the RNC is identified as being the bottleneck of the process. However, in this paper we show that the wireless hop presents the main bottleneck in the SMS-based malware propagation due to the high cost of establishing a connection.

SMS-related network threats have also been investigated in the academia. Most of these studies focus on the bottleneck that results from the delivery of short messages on channels shared with essential system control messages. A general overview of the transmission of SMS messages and its related cellular network vulnerabilities are discussed in [11]. This work is extended in [10], where the authors present solutions to this problem based on queue management and resource provisioning. The same authors analyze in [12] the causality of several Denial of Service (DoS) attacks targeting cellular networks. They show that the recent physical connection with external packet-based networks and the Internet has brought a whole new spectrum of vulnerabilities and threats to cellular networks. In parallel, other works present phone-targeted threats. The authors in [15] analyze the exhaustion of mobile phone batteries by constantly sending messages that are unnoticed to the user. A similar threat aiming to DoS a given phone by means of unnoticed SMSs is presented in [16].

Malware spreading on the Internet has also been a recurrent topic in the literature. From initial studies on self-propagating code and how to contain such infections [29] to recent analysis of modern worm propagation [27]. All these studies have in common that the malicious software always uses very popular communication pathways to propagate, such as email [30], social networks [28] and, as studied in this paper, text messages.

In summary, extensive work has been presented on malware propagation and SMS spreading on mobility networks. However, this is the first known study that models the wireless link. This new model demonstrates that the bottleneck of the process is indeed at this link of the access network. A smartphone infection propagating by means of text messages is shown to potentially saturate the wireless access link.

III. SMS OVER GSM AND UMTS

This section presents a description of the SMS service over GSM and UMTS networks, from the network architecture supporting the service to the actual process of delivery over the wireless interface. It is precisely at this point where the spreading of network messages face a severe bottleneck.

A. SMS Network Architecture

The Short Messaging Service Center (SMSC) is the central node that handles all SMS traffic in a network, independently from its origin, prior to routing and final delivery. The large amount of traffic generated by text messages, with reports of billions of SMSs flying through the airwaves daily [1], often results in multiple SMSCs per provider being supported to enhance the capacity. In this work we assume that a given provider is equipped with enough capacity at the SMSC level, because the operator can always add more capacity at this stage.

![Fig. 1: SMS Network Architecture](image-url)

The network architecture involved in the routing of a text message is shown in Figure 1. The SMSC, upon reception of a short message, examines its contents and, when necessary, formats it to the appropriate SMS format. Mobile-terminated messages are queued to be forwarded according to their destination. A copy of each message is stored until the reception of a delivery acknowledgement from the destination point. In the case of mobile-originated short messages, SMSs are stored in the phone’s memory until they can be sent over the wireless interface to the SMSC. In the case of a prolonged period of time with no network coverage, the message delivery fails and it has to be resent manually by the user.

The SMSC needs to know the location of the recipient in order to appropriately route the message. A query to a data base node known as Home Location Registry (HLR) provides the address of the Mobile Station Controller (MSC) to which the phone is attached and the message is forwarded. Once at the MSC, the location in terms of Base Station (BS) can be obtained from a query to a local Visitor Location Registry (VLR) data base or by means of a paging process. Finally, once at the BS, the message is delivered to the target mobile phone over the air interface and a confirmation message is sent back to the SMSC.

Note that the overall network architecture and procedure for SMS routing is very similar in both GSM and UMTS. In the case of UMTS, the node in charge of controlling multiple Base Stations is known as Radio Network Controller (RNC).
B. SMS Wireless Delivery on GSM

The radio traffic on GSM access networks is divided into two segments. With a FDMA/TDMA multiple access architecture, these networks establish radio resources for both control and data traffic. User communications (i.e., voice traffic) are delivered over a set of Traffic Channels (TCH) while the network’s signaling, both common to each user within a given cell or specific to one single phone, is transmitted over a set of Control Channels.

A mobile phone is by default in a battery-saving sleep mode, while keeping track of the Paging Channel (PCH), where alerts of incoming transmissions are received. Upon reception of a paging message identified with its Temporary Mobile Subscriber ID (TMSI), the device attempts to contact the BS over the Random Access Channel (GSM-RACH) to inform of its availability to receive incoming calls or SMSs (the device’s SMS buffer could be full) and establish a connection with the network. After this response, the BTS instructs the user to communicate using the Standalone Dedicated Control Channel (SDCCH). Through this channel authentication and encryption are established and the SMS is finally delivered.

The SDCCH channel is a highly bandwidth-limited GSM control channel used for short transactions including registration, initial call set up and SMS delivery. Generated as an aggregation of four logically consecutive time slots within the GSM multi-frame, this channel has an effective bandwidth of 782 bps [20]. Authentication, encryption enabling and final SMS delivery hold this channel busy for about 4 to 5 seconds [21]. This represents the main bottleneck of the SMS delivery process, given that, independently of the core network architecture, SMSs cannot be delivered at a higher rate than approximately 900 messages per hour per SDCCH channel. Nevertheless, in real deployments the number of SDCCHs is often equal to twice the number of carriers. Most cells have a total of 8 SDCCHs, so the total capacity per cell would be of about 8 times that of a single channel [21].

C. SMS Wireless Delivery on UMTS

Unlike in GSM networks, there is no SDCCH in UMTS. The exchange of control messages, authentication, encryption and connection establishment between a mobile terminal and the serving Base Station take place over the UMTS Random Access Channel (RACH) [25]. This channel is used for SMS transmission and delivery. The authors of [17] conclude that the Random Access Channel, despite its significantly higher bandwidth, as compared to the GSM SDCCH, is still a network bottleneck.

The Physical Random Access Channel (PRACH) carries the RACH transport channel, used by the mobile terminals to communicate with the Base Station in the uplink. The transmission over this channel is organized in 20ms frames that are, in turn, subdivided in 15 slots [25].

The Random Access Channel is shared by all the users within a given cell, so collisions might occur. In order to prevent and minimize collisions, a random access method based on preamble messages is established. The transmission of each data/control packet has two steps: the transmission of a set of preambles and the transmission of the actual packet [22]. Transactions on the RACH are, in turn, associated with a parallel physical channel in the downlink, the Acquisition Indication Channel (AICH), used by the BS to send responses to the User Equipment (UE) and organized as well in frames with 15 slots.

In the first step, whenever a UE has something to transmit, it randomly selects a slot from the 15 available. A short packet, known as preamble, is sent on the selected slot with a specific initial value of transmission power, derived from an open loop power control scheme. This initiates what is known as Preamble Cycle. The transmission of a preamble is often bounded to a persistence probability. This is, with probability $p$ the preamble is sent and, with probability $(1-p)$, transmission is postponed to the next frame. This preamble contains a randomly selected signature and acts as a petition to obtain access to the spectrum. Once a preamble has been sent, the UE listens to the corresponding slot in the AICH for a response from the BS.

Once at this point, three possible replies are to be expected [23]. Upon reception of an acknowledgement message, the UE is given access to transmit a packet in the following frame(s). If a negative acknowledgement message is received, the UE defers further attempts for a random number of frames (random back off time) and starts a new preamble cycle. In the case of not receiving an acknowledgement, a new preamble is sent with a pre-defined increase in the transmission power. This is known as Power Ramping. If a maximum number of retransmissions is reached or the transmission power is increased above a maximum value, the preamble cycle ends. The UE will wait for a randomly generated backoff time and will start a new preamble cycle unless it has reached the maximum number of allowed preamble cycles. Note that the value of $p$, distribution of the random back off time, initial power level, power increase step, maximum transmission power and the maximum number of preamble cycles are parameters broadcasted by the network and known by all the phones.

In order to receive an SMS, the destination terminal must, upon reception of the paging message, connect to the network through the RACH. On the other hand, the RACH is the channel through which a mobile-originated SMS is sent into the network and towards the SMSC. Therefore, both for message transmission and reception, access to the UMTS Random Access Channel has to be obtained.

IV. Simulation Model

The transmission of short messages on GSM is modeled assuming cells equipped with 8 SDCCH channels. Therefore, as described in section III-B, approximately 8 SMSs can be transmitted every 5 seconds in each cell. In the case of UMTS, cells are assumed to be equipped with one RACH channel. The UMTS Random Access Procedure is implemented in each cell by means of a custom Matlab script to model the SMS delivery on 3G networks. The RACH access parameters (probability $p$,
initial transmitted power, power step, backoff time distribution, etc) belong to UMTS’ Access Service Class number 4 (ASC4)[24]. Mobile terminals are distributed uniformly within each cell and both thermal noise and interference from other mobiles are modeled.

We assume that the malware propagates via SMS messages that contain a link to download a malicious application. This link may be disguised as a game or any other appealing application. This application obtains access to the phone’s contact book and determines which targets to send an SMS in order to spread the infection. Let the probability that a susceptible user follows the link and downloads the application be equal to 0.5. Once the user installs the application, the phone becomes infected and sends the same message to \( k \) randomly chosen contacts from the phone book every \( \tau \) minutes (\( \tau \) is exponentially distributed with mean \( \bar{\tau} = 40 \) minutes).

We demonstrate the propagation of the SMS-transmitted malware in a densely populated metropolitan area, equivalent to Washington D.C (68.2 mi\(^2\)), served by 120 cells [3]. The mobility network consists of 600000 mobile devices uniformly distributed over the area [21], with 30% of the phones susceptible to the infection (i.e. with the same Operating System). We assume that the network structure obeys the scale-free distribution of the Barabasi and Albert model [19]. Therefore, the size of the mobile devices’ contact book follows a power law distribution. We also assume that most of the contacts for a given user belong to the neighboring cells (i.e. the area around the subscriber’s home) as well as several other cells (areas where the subscriber works, lived before or where family and relatives reside). Several other contacts may be randomly distributed over the entire area to model other random connections (i.e. classmates, friends, etc.).

Note that we do not consider any legitimate background traffic to be loading the control channels (SDCCH and RACH). This implies that the same malware infection in a real system would spread even more slowly than the presented by the model in this paper. The results from the simulations characterize a best case scenario and can be considered a lower bound of the negative impact such an infection could potentially cause to the network. Finally, unlike in [9], we do not consider the capacity effects of the links between the different nodes described in Figure 1 beyond the actual wireless channel. This is because, as demonstrated in Section V, the transmission of short messages starts saturating the wireless link in GSM with a load of just about \( 8.3 \times 10^{-5} \) messages per second per user. Therefore, the transmission of short messages is stopped way before the load in any wired link reaches a potential saturation level. In this way, the clogging never reaches higher layers and only occurs locally at the wireless link between the phone and the BS.

V. RESULTS

In this section we present the results of the simulations of SMS-based malware propagating both on GSM and UMTS networks. These results are compared against the case of an infection spreading on an equivalent network without capacity constraints, modeling an Internet-connected wired network. Each run models 5 hours of malware spreading and we assume there are 3 infected phones at the beginning of each simulation. Time is represented in slots of 5 seconds and, at each slot, the system computes the number of new messages generated, queued messages at the SMSC and phone memory (see Section III-A) and SMSs delivered over the air interface. This is done
based on the capacity of the GSM system of about 8 short messages every 5 seconds, as described in Section III-B, and the capacity of the UMTS RACH Channel (Section III-C).

The total number of transmitted messages per time in the whole network and within one randomly chosen cell are plotted. These are shown for three cases: a traditional Internet-connected network (Figure 2 (a) and (b)), a GSM network (Figure 2 (c) and (d)) and an UMTS network (Figure 2 (e) and (f)). As expected, the spreading starts at a very low scale but soon reaches large values due to the exponentially increasing number of infected phones.

The results show that the spreading rate in mobility networks is much lower than in a regular wired topology. The number of transmitted messages is bound to the capacity of the control channels used to transport SMSs over the wireless link. Only a portion of the total generated short messages is transmitted successfully, remaining the rest queued for a later delivery. In the case of GSM, the 8 SDCCH channels in each cell will start clogging when the rate of SMS exceeds 8 transmitted messages every 5 seconds. Note that we consider an infinite queue size and unlimited memory in the mobile devices.

The case of UMTS is especially interesting. Unlike in GSM, the capacity of the channel used to transport text messages is not constant. A large increase in the load of messages leads to a higher collision rate that, in turn, decreases the overall capacity of the channel (see Section III-C). Note that, if the load steadily increases, the UMTS Random Access Channel could reach a saturation stage in which its capacity would be very close to zero. Figure 3 shows the number of successfully transmitted messages as a function of the offered SMS load to the RACH channel. Thus, in the case of malware spreading over SMS in UMTS, RACH collisions result in less messages being successfully transmitted as the infection spreads beyond a certain point. Therefore, the rate at which new phones become infected decreases. The overall consequence is the observation of a slight decrease of the number of sent messages towards the end of the simulation in Figure 2 (e) and (f).

Figure 4 depicts the number of queued messages as a function of time in the entire network and a detailed view of the first two hours in the case of GSM (Figure 4 (a) and (b)) and in the case of UMTS (Figure 4 (c) and (d)). Note that, given the infinite size of the buffers in the simulations, the quantity of stored messages increases to very large values. Independently of that, the network can be considered to be undergoing a saturation stage for as long as the queues are full. SMS traffic will be queued resulting in a later delivery. Messages start being queued earlier in a GSM network due to the lower capacity of the SDCCH. However, despite starting later, it can be observed that the total amount of queued messages in UMTS increases at a much faster rate.

Zooming into the figures (Figure 4 (b) and (d)) one can observe how the GSM network starts saturating after about 2.7 hours (2000 slots of 5 seconds). It is very important to note that, when this happens, the total SMS load in the network is of about 250 messages every 5 seconds, which is equivalent to just $8.3 \times 10^{-5}$ messages per second per user. This could
hide this behavior in a real network as it would be masked by regular SMS traffic.

Finally, Figure 5 shows the propagation of the infection on the three analyzed networks (wired Internet-connected network, GSM and UMTS). The malware spreads up to an eventual point where the 600,000 users in the system become infected. However, this occurs only in the case of a network with no wireless link. The inherent capacity bounds for short text messages in GSM and UMTS hold the spreading rate at a much lower rate, infecting just about 45,000 and 200,000 users respectively. Therefore, the infection will require a much longer time to fully propagate. In parallel, the network will be in a saturation state that will last for as long as the malware is active.

VI. CONCLUSION

This work argues that the wireless hop of a cellular mobility network is the bottleneck of the short message delivery process. Therefore, the architecture of a cellular network will constrain the spreading rate of an SMS-based infection. This paper demonstrates this effect by means of simulations. It also presents a new model that provides a more accurate analysis of malware propagation in GSM and UMTS networks. Unlike in previous works analyzing SMS-based malware infection and spreading, we show that this process will not generate a massive outbreak with a spreading rate equivalent to the one of a wired network.

Simulations of an SMS-based malware propagation are presented for a scenario with a size, capacity, population and density equivalent to the city of Washington DC. The results show that the total number of infected phones after 5 hours of infection spreading in GMS and UMTS networks are 10 and 3 times lower than that predicted by other models for wired Internet-connected networks.

ACKNOWLEDGMENT

The authors would like to thank Dr. Gustavo De Los Reyes, AT&T Executive Director of Technology Security, for his comments and valuable suggestions.

REFERENCES