A Detection Mechanism for SMS Flooding Attacks in Cellular Networks

Eun Kyoung Kim, Patrick McDaniel, and Thomas La Porta

Dept. of Computer Science and Engineering
Pennsylvania State University, PA, USA
ekkim@cse.psu.edu

Abstract. In recent years, cellular networks have been reported to be susceptible targets for Distributed Denial of Service (DDoS) attacks due to their limited resources. One potential powerful DDoS attack in cellular networks is a SMS flooding attack. Previous research has demonstrated that SMS-capable cellular networks are vulnerable to a SMS flooding attack in which a sufficient rate of SMS messages is sent to saturate the control channels in target areas. We propose a novel detection algorithm which identifies a SMS flooding attack based on the reply rate to messages sent by a handset. We further propose a mitigation technique to reduce the blocking rate caused by the attack. Our simulation results show that the false positive and false negative rates of our detection algorithm are low even when the attack traffic is blended with flash crowd traffic and/or the attack traffic mimics flash crowd traffic, and that the blocking rate is successfully reduced through the mitigation technique.

Keywords: SMS flooding attack, DDoS attack, flash crowd, anomaly detection, modeling, cellular network.

1 Introduction

Text messages continue to grow as the most popular data service of cellular networks. The total number of text messages sent globally has tripled over the past three years to reach 6.1 trillion in 2010. In other words, people around the world are sending 200,000 text messages every second [1]. In the U.S. 66% of mobile subscribers use text messaging service and over 600 text messages on average are sent or received monthly by a subscriber [2].

With this growing popularity of text messages, the reliability of Short Message Service (SMS) is becoming increasingly important. However, previous studies have shown that the control channels in the cellular networks may be a bottleneck for both SMS and voice services due to their limited capacity and shared nature. The stand alone dedicated control channel (SDCCH) is the most vulnerable since it is used for call setup and location updates as well as SMS [3,4,5]. An abnormal increase of SMS traffic may result in high occupancy of the SDCCH and high blocking rate of text messages and voice calls threatening the reliability of cellular networks.

A.D. Keromytis and R. Di Pietro (Eds.): SecureComm 2012, LNICST 106, pp. 76–93, 2013.

There are two kinds of the events that may cause a sudden increase in the SMS traffic volume in the cellular network: flash crowds and DDoS attacks. Flash crowds are an unusual burst of legitimate traffic produced by an increased number of users; these are frequently observed during special occasions [6,7]. For example, the volume of text messages sent on the New Year's Eve increases each year [8] and the resulting congestion causes lost and delayed text messages [9].

DDoS attacks through SMS are another cause of abnormal increase of SMS traffic. Typical SMS attacks aim to degrade target networks by depleting the control channel resources with a flood of SMS messages. In previous research, the feasibility of a SMS flooding attack was proved [10], and mitigation techniques were proposed [11]. However, they do not address how to detect SMS flooding attacks.

In this paper, we propose a novel anomaly detection mechanism that identifies malicious SMS flooding traffic causing intentional congestion in cellular networks. The difficulty is that the attack traffic mimics flash crowd traffic to circumvent detection. As the traffic behavior in flash crowds and flooding attacks are very similar, we need to find some features that can be used to distinguish them to reduce the false positive rate of our detection algorithm. Due to the lack of attack traffic traces, we analyze normal SMS traffic to infer the difference between flash crowds and flooding attacks. We find through the analysis that a mobile user replies to a message from a close friend with high probability, and is unlikely to answer a message from an unknown number. Therefore, we infer that if the reply rate for a handset which sends messages into a congested network is lower than a threshold, it is likely to be a malicious handset attempting to deplete the control channels.

We also develop a mitigation technique which classifies SMS traffic as normal, suspicious, or malicious and separates the traffic into three distinct queues with decreasing priorities to reduce the blocking caused by attack traffic and allow for fast identification of malicious handsets. The blocking of the normal handsets' traffic is efficiently diminished since a higher priority for obtaining the limited control channels is given to the normal handsets rather than the suspicious and malicious handsets.

Our simulation results show that our baseline algorithm performs the detection of unmixed flooding traffic with a very low false positive rate. The detection of attacks occurring in a flash crowd event and/or mimicking flash crowds is much more challenging. The mitigation technique, however, reduces the blocking rate of the messages from normal handsets successfully.

We compare our results to those of SMS-Watchdog, the most similar algorithm to ours in the literature, and show that we are more effective at distinguishing between attack traffic and naturally occurring flash crowd traffic.

The remainder of the paper is organized as follows. In Section 2, we discuss related works. The characteristics of the SMS network architecture and the different types of SMS traffic are introduced in Section 3. Our detection algorithm follows in Section 4. We evaluate our detection algorithm and mitigation technique through simulation in Section 5 and conclude in Section 6.

2 Related Work

The increasing popularity of short messages in cellular networks has led to much research on SMS capacity. In [3], it was shown that severe congestion may occur when the SDCCH channels are exhausted as they are shared by SMS, call setup and location updates. Agarwal et al. [5] conducted capacity analysis using a queueing model to show that the SDCCHs can be a bottleneck which increases the blocking probability of SMS as well as voice calls during elevated message loads. The possibility of an attack exploiting the limited and shared property of the SDCCHs was addressed in [4].

Enck et al. [10] demonstrated that SMS-capable cellular networks are vulnerable to a SMS flooding attack where a sufficient rate of SMS messages is sent from the Internet to local cell phones in order to saturate the SDCCH capacity. Furthermore, Traynor et al. [11] evaluated the performance of this attack by modeling and simulation and proposed mitigation techniques. However, they do not address how to detect flooding attacks. We propose a detection algorithm to identify SMS flooding attacks and a mitigation technique to lower the blocking rates at the control channels.

Previous research conducted by Yan et al. [12] proposed a SMS-related attack detection scheme named SMS-Watchdog that detects abnormal activities of SMS users by checking deviations from their normal social behaviors. Their approach is applicable to SMS flooding attacks because the attacker's behavior may be changed from the behavioral profile trained before the attack starts.

However, SMS-Watchdog gives false alarms when a flash crowd event occurs because the behavioral characteristics of normal SMS users are changed during flash crowd events. On the contrary, our algorithm can distinguish flooding attacks and flash crowds reducing the false alarms.

As DoS attacks and flash crowds are the two major concerns threatening the reliability and stability of the Internet, many studies on how to discriminate them have been conducted in the IP networks [6,13,14,15]. However, the direct application of these solutions is unsuitable because the IP flow and text messages of flash crowds have different properties. For example, the flash crowd traffic in IP networks is destined to a small number of servers while the messages exchanged in flash crowd events are scattered over many users in cellular networks.

We characterize flash crowd traffic and attack traffic based on the analysis of normal SMS traffic. [7] provides us with the statistics of flash crowd traffic in cellular networks. We obtain real SMS traces of three service providers from [16] and analyze them to infer the difference between a recipient's behavior to normal messages and attack messages.

3 SMS Communication Characterization

3.1 Network Characterization

The basic network structure of SMS is depicted in Fig. 1. A mobile handset B can receive a text message from one of two sources - another mobile handset A or

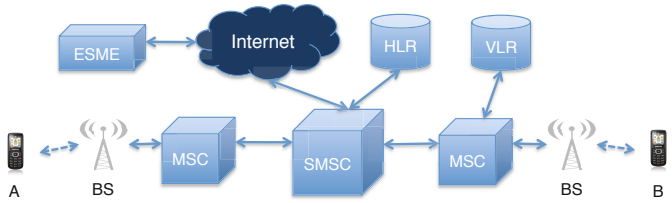


Fig. 1. SMS network structure

External Short Messaging Entity (ESME). An ESME is typically a non-mobile entity that submits messages to, or receives messages from a Short Messaging Service Center (SMSC) via the Internet. A text message from either is delivered to a recipient through a SMSC, a Mobile Switching Center (MSC), and a Base Station (BS).

A SMSC is responsible for storing and forwarding short messages to a terminating MSC. It obtains routing information from the Home Location Register (HLR) to locate the proper MSC. The MSC performs the switching functions of the system and delivers SMS messages to the specific mobile subscriber by retrieving the subscriber's location from the Visitor Location Register (VLR). A MSC can store the messages in a queue for a short time during which it retransmits the messages if acknowledgements are not received within a specific time. If a message is not successfully delivered to the mobile station after the maximum number of retransmission attempts, the MSC sends an error message to the SMSC [17,18].

Between a BS and a mobile handset, a message is delivered via the air interface using control channels. First, a BS transmits a paging request with an identifier on the Paging Channel (PCH). When a mobile handset hears its identifier, it responds to the BS on the Random Access Channel (RACH). Then, the BS assigns a SDCCH to the handset. The handset authenticates with the BS and receives the text message via the SDCCH. As a SDCCH is used for call setup and location updates in addition to SMS transfer, it may be flooded by an increase of SMS requests blocking both voice and SMS communication.

3.2 Normal Traffic Characterization

There have been prior efforts on characterizing normal SMS traffic patterns. Some researchers analyze SMS traces collected from a nationwide cellular carrier with more than 20 million subscribers over a period of three weeks [19,7]. They present thread-level characteristics in addition to the SMS message-level characteristics, where a thread is defined as messages exchanged between the same two users within a predefined timeout period. According to their analysis with 10 minutes as a timeout, the number of messages in each thread, or the thread length, is 4.9 on average and the average thread duration is 8 minutes. That implies that the average interval between receiving and responding to a message is 2 minutes.

However, not all the recipients make a reply to a message that they have received. In our own analysis on the SMS trace data provided in [16], 22% of the messages are "single" messages which are not followed by another message, where we only consider the messages from the handsets which generate at least ten messages per a day on average. Thus, the average length of a thread including "single" messages is $1 * 0.22 + 4.9 * 0.78 \approx 4$.

Another measurement study on SMS traffic, logged in records from three different companies over a one month period, examines the distribution of the intervals between messages belonging to one thread [16]. The empirical results show that the inter-arrival time and the waiting time of normal messages have power-law distribution within a thread duration and a new thread is initiated by an exponential distribution. The arrival rates of calls and SMS messages in a single sector per second and the service rates of calls and SMS messages at SDCCH are also known as shown in Table 1.

Table 1. System Variables and Parameters

λ_{call}	Arrival rate of voice calls	0.25 calls/sector/sec [11]
λ_{SMS}	Arrival rate of SMS msgs	0.7 msgs/sector/sec [11]
$\mu^{-1}_{SDCCH,call}$	Service rate of voice calls at SDCCH	1.5 sec [20]
$\mu^{-1}_{SDCCH,SMS}$	Service rate of SMS msgs at SDCCH	4 sec [4]

Flash crowd traffic shows different characteristics from regular SMS traffic. The traffic looks anomalous because cellular networks suffer a sudden increase of SMS traffic in a flash crowd event. For example, the volume of messages exchanged during the New Year's Eve in India reaches almost eightfold the normal traffic level [7]. Such an increase in traffic is affected more by an increase in the number of SMS users sending and receiving messages rather than an increment of messages per user. Therefore, the SMS communication in a flash crowd is different from a regular SMS communication in that the increased volume of traffic is caused by an increased number of users without a change in the number of messages sent by a user.

We also observe that 60% of handsets participating in a flash crowd do not send any messages in three days before the event [7]. These new participants have a higher probability to be mistakenly classified as malicious as they have weak prior relationship with legitimate recipients.

Even though flash crowd traffic may slow down the message delivery or even cause some messages to be discarded because of congestion [9], it should be serviced as legitimate because it naturally occurs from normal handsets. Therefore, we develop an anomaly detection algorithm to distinguish malicious attack traffic from flash crowd traffic even when they are intermingled and malicious attack traffic mimics flash crowd traffic to avoid the detection.

3.3 Attack Traffic Characterization

Previous studies on SMS network capacity have proven that the SDCCH can be a bottleneck in cellular networks due to its limited capacity and shared characteristics [3,5]. This makes a SMS flooding attack feasible because an attacker can paralyze cellular communications in a certain area by overloading SDCCHs in that area. Such an attack will be performed by sending enough messages to potential target lists which can be created by several efficient methods [10].

We assume that an attacker has the capability to compromise a number of handsets so that they can send attack traffic under the control of the attacker without any awareness of the owners. Even though we only consider a mobile-to-mobile attack in this paper, a SMS flooding attack using bulk messaging services can be detected if we cast each ESME of bulk messaging providers as an attacking mobile handset in the algorithm.

Furthermore, we assume that the attacker is intelligent enough to mimic the behavior of normal users in a flash crowd. The attacker can compromise a large number of handsets and make them generate bogus messages with seemingly normal rates so that the aggregated traffic saturates the SDCCHs in a target area. The attacker can even launch the attack purposely during a flash crowd event to reduce the probability of being detected.

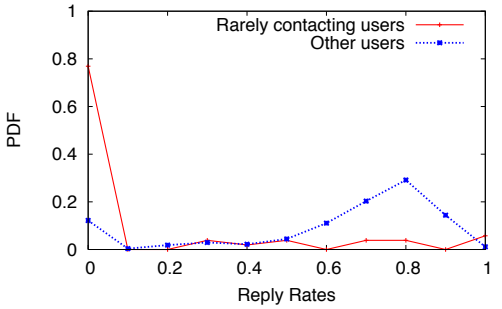


Fig. 2. Probability distribution function of reply rates

Consequently, a flooding attack and a flash crowd cannot be easily distinguished by the traffic characteristics determined by senders' behavior such as the total volume of generated messages, message generation rate per handset and contents of the messages. However, we infer that the recipients' behaviors for the messages sent by a normal user in a flash crowd event and an attacker disguised as a normal user are distinguishable.

We suppose that a user who sends out only one message to a recipient during over a one month period represents an unknown or unfamiliar sender to the recipient. Through the analysis of the SMS trace given in [16], we find that the reply rates for the unfamiliar senders and the other normal users have distinguishable distributions as seen in Fig. 2 with a 15% and 62% average value, respectively.

Since an attacker is an unknown originator from a recipient's point of view, it is expected to have a similar distribution of reply rates to that for an unfamiliar sender. Thus, we can distinguish an attack handset from a normal one based on the reply rates regardless of whether the attacker mimics a normal user's message sending characteristics.

4 Detection Algorithm

4.1 Attack Model

The purpose of a SMS flooding attack is to paralyze the cellular network in a specific area by overloading the SDCCHs. In this paper, the target of the attack is a sector served by the BS of a MSC. The handsets serviced in the target area are called local handsets and the handsets outside the targeted area are remote handsets. The incoming attack occurs from remote handsets while the outgoing attack is performed by the local handsets. Because handsets are authenticated, while they can be infected with a virus that causes them to launch an attack, their addresses cannot be spoofed.

The attack can be classified as a *mixed attack* or *unmixed attack* according to whether it occurs in concurrence with a flash crowd or not. An attacker may launch a mixed attack to accelerate the attack's efficiency and avoid detection by having the attack traffic intermingled with flash crowd traffic. Our base algorithm aims at successfully identifying the messages sent from malicious handsets among the intermingled traffic even under a mixed attack and keeping the false positive rate low by adaptively changing the expected reply rate for the benign messages during the congestion.

From the perspective of the intensity of the attack traffic from a single handset, we can classify the attack as *high-intensity* or *low-intensity*. The attacker can choose low-intensity with a large number of compromised handsets mimicking a flash crowd; however, a high intensity attack with a small number of compromised handsets is easier to carry out. Detection of a low intensity attack takes more time as the interval between attack messages sent by a handset and the number of attackers are larger. However, the blocking rate for the normal handsets decreases efficiently through our mitigation technique even when the false negative rate is not low.

Consequently, we carry out a performance evaluation for four types of attack - 1) *unmixed attack with high intensity*, 2) *unmixed attack with low intensity*, 3) *mixed attack with high intensity*, and 4) *mixed attack with low intensity*. Intuitively, the detection of the mixed attack with low intensity is the most challenging while the unmixed attack traffic with high intensity is detected with the shortest delay.

4.2 Algorithm for Identifying Attackers

We deploy a detector on each MSC to detect anomalies in air interfaces under the coverage of a MSC. Because we make use of the reply rate of mobile handsets

Alg. 1 : Monitor Threads

```
1: for each message M observed in W do
2:   if M is an outgoing message from L to R then
3:     if T = (L, R) exists then
4:       Increase L_s by 1
5:     else if T = (R, L) exists then
6:       Increase R_r by 1
7:     else
8:       Create T = (L, R)
9:       Increase L_s by 1
10:    end if
11:  end if
12:  if M is an incoming message from R to L then
13:    if T = (R, L) exists then
14:      if M is delivered to L then
15:        Increase R_s by 1
16:      end if
17:    else if T = (L, R) exists then
18:      Increase L_r by 1
19:    else
20:      if M is delivered to L then
21:        Create T = (R, L)
22:        Increase R_s by 1
23:      end if
24:    end if
25:  end if
26: end for
```

Table 2. Variables for Alg. 1

M	SMS message collected during W
T	Message thread represented by a pair of $(sender, receiver)$
L/R	Local/remote handset
L_s/R_s	The number of sent messages from L/R
L_r/R_r	The number of replies to L/R

for distinguishing benign and malicious traffic, the detector gathers (sender ID, recipient ID, timestamp) information of both outgoing and incoming messages. At the end of every time window W with duration ω, the detector looks into all the information collected during the time window and creates message threads and updates the number of sent messages and the corresponding replies for a handset as shown in Alg. 1 with the variables in Table 2.

Note that for incoming messages, only a message delivered to the destination successfully can increase R_s. Otherwise, the reply rates will be underestimated.

A detector contains an analyzer which identifies the attackers sending overloading messages. Upon the detection of congestion, the analyzer is activated and

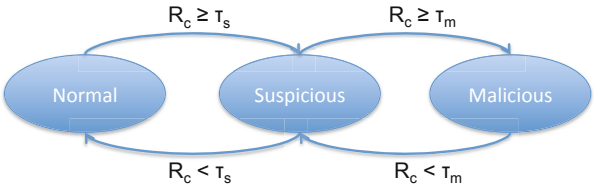


Fig. 3. SMS state transition diagram

calculates the reply rate threshold under normal network conditions, θ, considering the upper bound on the expected false positive rate for the threshold. Let X be a random variable of reply rates with finite expected value μ and non-zero variance σ^2. Given the arbitrary distribution of reply rates in the normal traffic observed before the congestion, the false positive rate is bounded to $\frac{1}{1+k^2}$ if we choose $\mu - k\sigma$ as the threshold, and consider a reply rate of less than that value as an anomaly, which is supported by the one-sided *Chebyshev's inequality*:

$$Pr(X \leq \mu - k\sigma) \leq \frac{1}{1+k^2}.$$

Accordingly, by setting $k = 1$ and $\theta = \mu - \sigma$, the false positive rate does not exceed 0.5. To limit the upper bound of false positive rate to 0.2, we choose $k = 2$ and $\theta = \mu - 2\sigma$. Even though a larger k and smaller θ guarantee a lower expected false positive rate, this might cause a lower detection probability and longer delay in return. The uncertainty of this effect is caused by the fact that we do not know distribution of the reply rates for malicious handsets a priori, although we expect them to be lower than that for normal handsets. In order to remove this uncertainty and resolve the trade-off, we choose $k = 1$ and $\theta = \mu - \sigma$ to increase the detection probability and reduce the detection delay, and use a scoring-based technique [21] to reduce the false positives by confirming the anomalies as attacks only when the anomaly score exceeds a threshold.

Note that θ is based on the measured distribution of reply rates during normal network conditions before the congestion occurs. The reply rate from the handsets in the area under attack will be decreased from its normal rate because the congestion in the attack area. This congestion will cause messages to be blocked and thus, send no reply. Therefore, the reply rate threshold for successfully delivered messages from a remote handset, τ_r, should be dynamically changed to reflect the blocking rate caused by congestion. As the ratio of the number of unblocked replied messages to that of replies in uncongested network is $(1 - B_{avg})$, we set τ_r to $\theta * (1 - B_{avg})$. We use B_{avg}, the moving average of blocking rate for ω_B to smooth the change of the blocking rate.

The anomaly score representing the degree to which a handset is considered an anomaly or attacker is initially set to 0. The score increases if the current reply rate is lower than θ for a normal message under current network conditions and decreases otherwise. When the anomaly score for a handset reaches a threshold designated for suspicious handsets, τ_s, the analyzer marks the handset as suspicious. If the score keeps increasing to a threshold for malicious handsets, τ_m, the handset is deemed malicious. As the analysis progresses, the score may

Alg. 2 : Identify Attackers

```
calculate θ
τ_r = θ * (1 − B_avg)
R_c = 0
for each handset R in T = (R, L) do
    if R send or receive a message then
        R_rr = R_r / R_s
        if R_rr < τ_r then
            R_c + +
        else
            R_c = max(R_c − −, 0)
        end if
        if R_c ≥ τ_m then
            Mark R as malicious
        else if R_c ≥ τ_s then
            Mark R as suspicious
        else
            Mark R as normal
        end if
    end if
end for
```

Table 3. Variables for Alg. 2

R_{rr}	Reply rate for R
R_c	Anomaly score representing the likelihood that R is an attacker
B_{avg}	Moving average of blocking rates for duration ω_B
θ	Reply rate threshold for normal handsets in normal network condition
τ_r	Reply rate threshold for normal local/remote handsets in congested network
τ_s	Anomaly score threshold for suspicious handsets
τ_m	Anomaly score threshold for malicious handsets

go lower and higher than the each threshold causing the change of the status of a handset as shown in Fig. 3.

The algorithm for an incoming attack is summarized in Alg. 2 and the variables used in Alg. 2 are presented in Table 3.

4.3 Mitigation Technique

We devise a 3-queue mitigation mechanism in which each kind of traffic classified by the detector - normal, suspicious, and malicious traffic - is served by one of three different queues with different weights. Weighted Fair Queueing [22] is used for scheduling messages in the queues.

Normal traffic is processed with a weight of 2 while suspicious traffic has a weight of 1. The malicious traffic is placed in the lowest priority queue and is only served when the two higher priority queues are empty. The blocking

rate for the messages from normal handsets is efficiently reduced by prioritizing the process of the normal messages while reducing the number of requests for the wireless control channels by delaying or refusing service for suspicious or malicious handsets.

5 Simulation Results

In this section we evaluate our algorithm. We also compare it with the most similar related work, called SMS-Watchdog, and show that we achieve better results for the challenging circumstances.

5.1 Simulation Settings

To evaluate the performance of our algorithms, we implement a simulator based on the characteristics of SMS communication and the proposed algorithms Alg. 1 and 2. We explain the settings in our simulation.

Network Settings. Assuming a SMS network with the network components in Fig. 1, our detector modules are deployed at the MSCs because they can provide all the information - (1) the blocking rate of SDCCHs in each sector, and (2) (sender ID, recipient ID, timestamp) of messages needed to detect the attacks targeting the SDCCHs in sectors controlled by the MSCs.

The message queues for mitigation techniques are implemented in the BS. The forwarded messages from the MSC have indicators to which queue they belong. If the corresponding queue is full, the MSC retries the delivery. The maximum number of attempts is set to 2. After that, an error message returns to the SMSC and the message is deleted from the MSC.

Parameter Settings. In Alg. 1, the interval of analysis on message threads, ω, needs to be set considering the tradeoff between timely detection and computational overhead. In our simulation, we set $\omega = 10$ seconds because it is short enough to capture each message of one thread in each time window and long enough not to overload the detector. We set the value of blocking rate acceptable in cellular networks, β, to 1%. If the average blocking rate for ω is greater than β, an analyzer is activated to identify the attackers.

In Alg. 2, the duration for the calculation of the moving average of blocking rates, ω_B, is set to 120 seconds. Since the average waiting time for a reply is 120 seconds, we expect the previous message to have been transmitted 120 seconds prior to the message just received. Therefore, the average blocking rate for the last 120 seconds affects the reply rate of the message.

Traffic Settings. We simulate 24 hours of SMS communication. Local and remote handsets constantly transmit regular SMS traffic during the simulation. The regular messages are submitted by 4800 handsets at 0.7 msgs/sector/sec rate according to the normal traffic characteristics.

Attack traffic is emitted for one hour from 23 to 24 hours. The reply rates for normal handsets observed for 23 hours before the attack prevent the detector from misclassifying the normal handsets as malicious handsets due to the transient low reply rates during the congestion. The longer training period builds a stronger "send-reply" relationship among normal users making the discrimination between the normal and malicious messages easier.

The aggregated volume of the attack traffic is 8 times more than the value of regular traffic. For the mixed attack, flash crowd traffic fourfold the normal traffic is generated in addition to the attack traffic.

5.2 Performance Evaluation

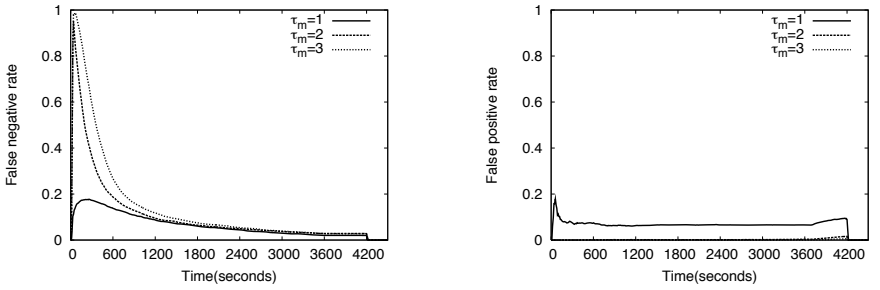
We evaluate our algorithm using several fundamental metrics: false positive rate (FPR), false negative rate (FNR), and blocking rate. The false positive rate is the fraction of benign handsets that are misjudged as malicious over all benign handsets. The false negative rate is the fraction of malicious handsets that are mistakenly judged as benign over all malicious handsets. The blocking rate is the portion of messages which are blocked due to insufficient channel resources.

We show the performance of our baseline detection mechanism which identifies the malicious handsets but does not resolve the congestion, and the performance of the detector with a mitigation technique which reduces the blocking in the air interface by placing the identified attackers in a low priority queue. The algorithm performs significantly better with mitigation in places because malicious handsets are removed from the traffic flow making it easier to detect remaining malicious handsets.

Without Mitigation Techniques. We first examine the FNR and FPR of the baseline algorithm for unmixed incoming traffic with high intensity with τ_m= 1, 2, and 3. The results are presented in Fig. 4a and 4b respectively for the time elapsed after the start of the attack. The FNR decreases more quickly for a smaller τ_m because the attackers' score, R_c, exceeds the threshold, τ_m, in a shorter time. When $\tau_m = 1$, however, the resulting FPR is over 5% on average whereas for $\tau_m = 2$ and 3, FPR is reliably low throughout the attack period. This is because the attack likelihood score for a normal handset which has not exchanged messages with recipients before the attack turns to 1 when the detector sees the first incoming message and exceeds the threshold in the case that $\tau_m = 1$.

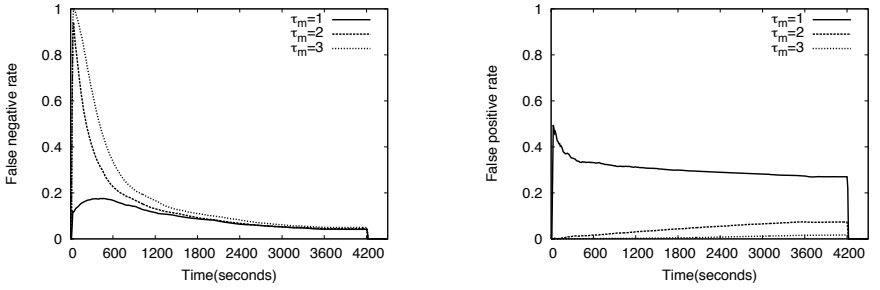
Our algorithm operates even in more challenging situations. When the attacker generates a high intensity attack traffic in the middle of a flash crowd event, it is difficult to distinguish malicious traffic from benign traffic because more than a half of the benign handsets in flash crowds have not participated in conversational message threads prior to the event. However, even with the mixed traffic, our baseline algorithm identifies the attackers based on the difference between reply rates of malicious and benign messages.

Fig. 5a and 5b show more clearly that the FPR increases as τ_m decreases and the FNR increases as τ_m increases. When $\tau_m = 1$ or 2, the FPR increases to



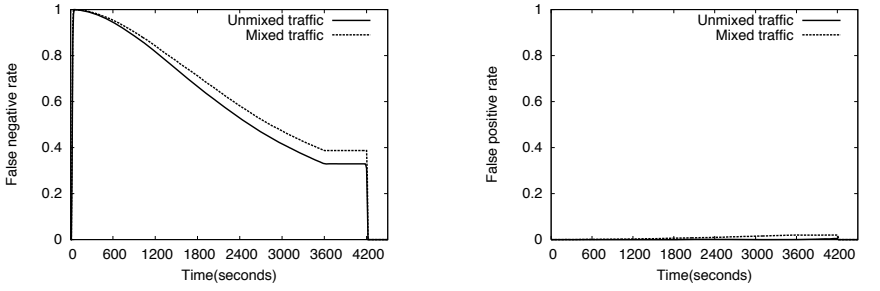
(a) (b)

Fig. 4. (a) FNR and (b) FPR of unmixed attack traffic with high intensity without a mitigation scheme



(a) (b)

Fig. 5. (a) FNR and (b) FPR of mixed attack traffic with high intensity without a mitigation scheme



(a) (b)

Fig. 6. (a) FNR and (b) FPR of two kinds of attack traffic with low intensity without a mitigation scheme

a high value even though the FNR decreases quickly. The dramatic increase of FPR for $\tau_m = 1$ is caused by the sudden increase of new SMS users in flash crowds. Thus, we set τ_m to 3 considering trade-off between the FPR and the FNR.

With $\tau_m = 3$, the performance of mixed and unmixed attacks with low-intensity is shown in Fig. 6. The detection of the attacks is carried out slower than the high-intensity attacks, but the false positive rates are very close to those of the corresponding high-intensity attacks. Our observation is that the intensity of attack messages initiated from a handset determines how fast the attackers can be detected and the ratio of messages from new active normal users during the attack determines the accuracy of the detection. The strength of our detection algorithm is low false positive rates even in the extreme case of the mixed traffic with low intensity even though the detection of the attacking handsets is inherently slow due to the low arrival rate of the attack messages.

With Mitigation Techniques. Our detection algorithm identifies the attacking handsets but cannot resolve the blocking caused by the attack messages. We devise a 3-queue mitigation mechanism which places the three kinds of traffic - normal, suspicious, and malicious traffic - classified by the detection algorithm into the corresponding queues and schedules each messages using Weighted Fair Queueing [22]. By providing normal messages with more wireless channel resources, the blocking rate for normal messages is efficiently reduced.

For message classification, we need to determine the proper value of τ_s for $\tau_m = 3$. The attack likelihood scores for all incoming messages after the detection starts are initially 0. A handset which has not established message threads with a recipient before the attack is likely to have 0 as a reply rate and 1 as the attack likelihood score at the first classification process. So, if we set τ_s to 1, the handset is classified as suspicious. With τ_s set to 2, the handset is still regarded as normal.

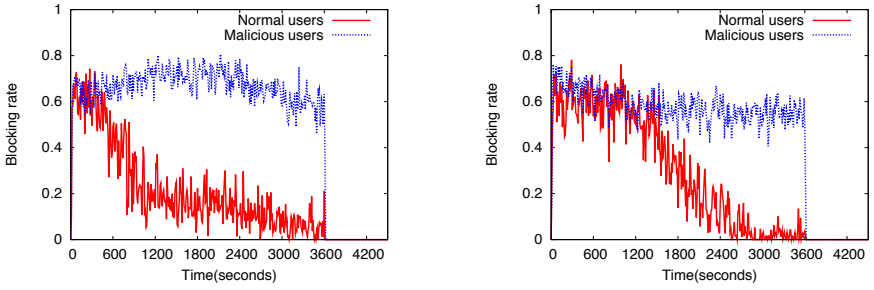


Fig. 7. Blocking rates for mixed attack traffic with low intensity when a mitigation technique is applied with (a) τ_s=1 and (b) τ_s=2

We determine the proper value for τ_s taking into account blocking rate as the blocking rate is the ultimate measure of the performance of the mitigation system. We show in Fig. 7 the blocking rate for the mixed attack with low intensity when the mitigation technique is applied. The blocking is mitigated most efficiently with $\tau_s = 1$, from 60% to 20% in approximately 20 minutes. Therefore, we set τ_s and τ_m to 1 and 3, respectively.

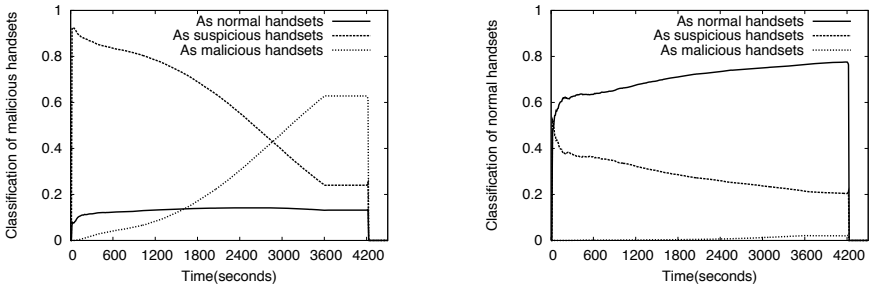
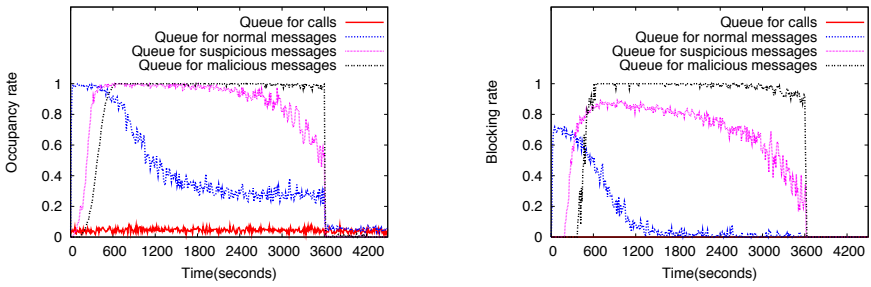


Fig. 8. Classification of (a) malicious and (b) normal handset for mixed attack traffic with low intensity with a mitigation scheme

Fig. 8a and 8b gives us insight into how the detector with the 3-queue scheme operates to classify the message-sending handsets in case of mixed attack with low intensity. Most of the malicious handsets are considered as suspicious at first as we see in Fig. 8a. But, they are subsequently classified as malicious, and so the ratio of malicious handsets classified as suspicious starts to decrease and the ratio of malicious handsets classified correctly increases. Fig. 8b shows that the normal handsets occupy both the normal queue and the suspicious queue. This is because normal handsets from flash crowds are likely to be classified as suspicious due to the absence of previous message threads while the normal handsets which have sent messages and received replies during the prior normal network situation are likely to be classified as normal.

The normal handsets in normal queue are served with the highest priority without much competition with the malicious handsets. Moreover, the competition in the suspicious queue between the normal handsets and malicious handsets is resolved as more malicious handsets are classified as malicious. Therefore, the blocking rate for normal handsets decreases efficiently while the messages from malicious handsets are suspended in lower priority queues or discarded after the maximum number of retransmissions.

Fig. 9a presents the occupancy in each of the three queues of 3-queue scheme. This results from the classification performed at the detector. The occupancy at the normal queue is almost 1 at the start of the attack. As the classification of suspicious handsets occurs, the occupancy of the suspicious queue increases and the occupancy of the normal queue decreases. Then, the handsets in the suspicious queue are judged as normal or malicious by the detector, and the occupancy of the suspicious queue decreases. As more handsets are classified as malicious, the occupancy of the normal and suspicious queues decreases because the messages in these queues are served quickly. The blocking rate in each of the queues of the 3-queue scheme is shown in Fig. 9b. The blocking rate of a queue goes up when the occupancy of the queue is high and falls if the queue has space for new messages.



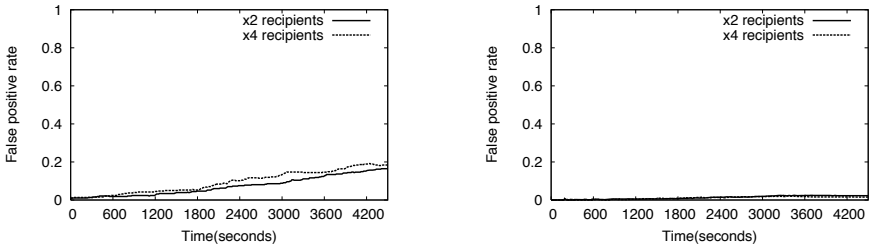
(a) (b)

Fig. 9. (a) Occupancy and (b) blocking rate for mixed attack traffic with low intensity with a mitigation scheme

5.3 Comparison with SMS-Watchdog

The work most similar to ours is SMS-Watchdog [12]. In this work, SMS-based blending attacks are detected using each user's regular social behaviors. For example, anomalies are detected by checking if the number of recipients in a window of messages from a sender deviate significantly from the average number of unique recipients in training messages.

Because a blending attack has similar characteristics to flash crowds in terms of the increased number of recipients per sender, the SMS-Watchdog algorithms is not effective at distinguishing an attack from a flash crowd. Fig. 10a shows that the false positive rate of SMS-Watchdog's R detection scheme for a flash crowd increases to 17% and 20% for a twofold and fourfold increase in the number of recipients per sender, respectively, in a case in which the number of messages and senders increases up to 4 times more than that under regular conditions. On the contrary, our scheme has false positive rate of less than 2% as shown in Fig. 10b, which means we correctly classify flash crowd traffic.



(a) (b)

Fig. 10. False positive rates of (a) R-type detection in SMS-Watchdog scheme and (b) our scheme for a flash crowd

6 Conclusion

We propose a novel detection algorithm which identifies a SMS flooding attack regardless of whether the attack traffic is mixed with legitimate flash crowd traffic and/or the attack traffic is mimicking flash crowd traffic. To distinguish malicious handsets, we consider the reply rate to messages sent by a handset. If the reply rate of a certain handset is lower than that expected for a normal handset, the handset is likely to be an attacker. We show that our baseline algorithm performs the detection of unmixed traffic with a very low false positive rate. The detection of attackers mimicking benign users during a flash crowd event takes longer, but the false positive rate is still low.

We propose a 3-queue mitigation scheme to reduce the congestion on the wireless control channels. The mitigation scheme employs three queues with different priorities to serve normal, suspicious, and malicious traffic differentially. We show that the blocking rate of normal handsets is efficiently diminished by prioritizing normal messages.

Acknowledgment. This work was supported by the US National Science Foundation (NSF) (CNS-0905447 and CNS-0643907).

References

1. ITU, The world in 2010: Ict facts and figures, ITU. Tech. Rep. (2010), www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf
2. Nielsen: State of the media (January 2011), http://blog.nielsen.com/nielsenwire/wp-content/uploads/2011/01/nielsen-media-fact-sheet-jan-11.pdf
3. Kyriazakos, S., Karetsos, G., Kechagias, C., Karabalis, C., Vlahodimitropoulos, A.: Signalling channel modelling for congestion management in cellular networks. In: IEEE VTS 54th Vehicular Technology Conference on VTC 2001 Fall, vol. 4, pp. 2712–2715. IEEE (2002)
4. Sms over ss7, National Communications System. Tech. Rep. (2003)
5. Agarwal, N., Chandran-Wadia, L., Apte, V.: Capacity analysis of the GSM short message service. In: National Conference on Communications (2004)
6. Jung, J., Krishnamurthy, B., Rabinovich, M.: Flash crowds and denial of service attacks: Characterization and implications for CDNs and web sites. In: Proceedings of the 11th International Conference on World Wide Web, pp. 293–304. ACM (2002)
7. Meng, X., Zerfos, P., Samanta, V., Wong, S., Lu, S.: Analysis of the reliability of a nationwide short message service. In: 26th IEEE International Conference on Computer Communications, INFOCOM 2007, pp. 1811–1819. IEEE (2007)
8. Cellular News. An estimated 43 billion text messages were sent globally on new years eve. (January 2008), http://www.cellular-news.com/story/28496.php
9. Cellular News. Congestion causes text message slowdown (January 2008), http://www.cellular-news.com/story/28391.php
10. Enck, W., Traynor, P., McDaniel, P., La Porta, T.: Exploiting open functionality in SMS-capable cellular networks. In: Proceedings of the 12th ACM Conference on Computer and Communications Security, pp. 393–404. ACM (2005)

11. Traynor, P., Enck, W., McDaniel, P., La Porta, T.: Mitigating attacks on open functionality in SMS-capable cellular networks. IEEE/ACM Transactions on Networking 17(1), 40–53 (2009)
12. Yan, G., Eidenbenz, S., Galli, E.: SMS-Watchdog: Profiling Social Behaviors of SMS Users for Anomaly Detection. In: Balzarotti, D. (ed.) RAID 2009. LNCS, vol. 5758, pp. 202–223. Springer, Heidelberg (2009)
13. Le, Q., Zhanikeev, M., Tanaka, Y.: Methods of Distinguishing Flash Crowds from Spoofed DoS Attacks. In: 3rd EuroNGI Conference on Next Generation Internet Networks, pp. 167–173. IEEE (2007)
14. Marnerides, A., Pezaros, D., Hutchison, D.: Flash crowd detection within the realms of an internet service provider (isp). In: The 9th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (2008)
15. Li, K., Zhou, W., Li, P., Hai, J., Liu, J.: Distinguishing DDoS attacks from flash crowds using probability metrics. In: 2009 Third International Conference on Network and System Security, pp. 9–17. IEEE (2009)
16. Wu, Y., Zhou, C., Xiao, J., Kurths, J., Schellnhuber, H.: Evidence for a bimodal distribution in human communication. Proceedings of the National Academy of Sciences (2010)
17. 3GPP, TS 23.040, Technical Realization of the Short Message Service (SMS); Release 9. v9.3.0 (2010)
18. 3GPP, TS 24.011, Point-to-Point (PP) Short Message Service (SMS) support on mobile radio interface; Release 9. v9.0.1 (2010)
19. Zerfos, P., Meng, X., Wong, S., Samanta, V., Lu, S.: A study of the short message service of a nationwide cellular network. In: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, pp. 263–268. ACM (2006)
20. Fonash, P., McGregor, P.: National Security/Emergency Preparedness Wireless Priority Service. In: Proc. 8th Int'l. Conf. Intelligence in Next Generation Networks
21. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys (CSUR) 41(3), 15 (2009)
22. Demers, A., Keshav, S., Shenker, S.: Analysis and simulation of a fair queueing algorithm. In: Symposium Proceedings on Communications Architectures & Protocols, pp. 1–12. ACM (1989)