Mitigating Attacks on Open Functionality in SMS-Capable Cellular Networks

***Abstract*— The evolution of phone networks from isolated voice carriers to Internet-enabled multipurpose data and voice networks has introduced exceptionally dangerous vulnerabilities. For example, a recent report showed that a carefully crafted DOS attack on text-messaging could incapacitate all cell phone communications in Manhattan with little more than a cable modem. This attack highlights a critical vulnerability of the evolving phone network infrastructure: cellular network control channels are exposed to adversaries in the phone network and the wider Internet. In this paper, we consider novel countermeasures to attacks on the control channel interfaces of the cellular networks. We adapt existing network admission control control techniques and develop novel channel allocation strategies in addressing these threats. The costs and tradeoffs associated with the mitigation strategies are analytically quantified. We further introduce X, a extensive GSM simulator that characterizes the queuing and air interface behavior in cell phone to base-station communication. Our analysis and simulation shows that we can sustain legitimate communications in the presence of highly targeted and intense attacks. (rewrite this part numbers?) We conclude by considering how these techniques can be applied not only to preventing SMS traffic misuse, but to the range of media forms emerging in current and next generation networks.**

I. Introduction

Cellular networks are an increasingly essential means of communication. In addition to traditional voice telephony, these systems now offer a wide variety of data and text/short messaging services (SMS). As a means of increasing the adoption of such services, cellular providers have increasingly created gateways between their own networks and the Internet. This heightened usability and utility are responsible for soaring usage statistics. In the United States alone, some five billion text messages are sent each month [21]. Indeed, for significant numbers of users, text messaging has become as or more popular a means of communication than traditional voice telephony [3].

While a great deal of beneficial new functionality is now possible, the interconnection of these systems inherently exposes cellular networks to many of the problems prevalent in the Internet. Because these systems were designed to operate in the absence of influence from external networks, many of these exploits violate core assumptions upon which these systems were built. Enck, et al. [13] present one such example. Given a list of phone numbers for a metropolitan area, an adversary can use Internet gateways to inject a relatively small number of text messages per second into a cellular network. In so doing, the attacker is able to deny voice service to the targeted area. In events such as September 11th, when the use of communication systems is absolutely critical, the cost of such a vulnerability becomes decidedly human.

This paper creates and develops a number of traffic engineering techniques and evaluates their ability to mitigate or eliminate the damage caused by targeted SMS attacks. Because this work addresses issues caused by the interconnection of the Internet and telecommunications networks, we seek to solve these problems through a combination of techniques from both domains. Our work begins by challenging the effectiveness of so-called "edge solutions" including per-user rate limitation and spam filtering. We then apply well-known queueing techniques including variants of Weighted Fair Queueing (WFQ) and Weighted Random Early Detection (WRED), which are well tested for addressing traffic overload in the Internet. Our work then focuses on the alleviation of congestion by reapportioning the wireless medium through novel methods including Strict Resource Provisioning (SRP), Dynamic Resource Provisioning (DRP) and Direct Channel Allocation (DCA). We finish by exploring the effects of combining multiple countermeasures.

At the current time, an adversary is able to deny voice service to cities the size of Washington D.C and Manhattan with the bandwidth available to a cable modem. Through the application of the above techniques, we allow cellular networks to operate safely even when the signaling links delivering voice and SMS traffic reach maximum capacity. More importantly, the implementation of these mechanisms allows providers to securely offer Internet-coupled services without necessitating a significant and expensive re-engineering of their networks.

In this paper, we make the following contributions:

- **Simulator Design and Development:** Using publically available GSM standards, we have designed and implemented a tool to simulate the GSM air interface.
- **Network/Attack Characterization:** Through mathematical modeling and simulation, we create detailed characterizations of system behavior for networks experiencing targeted SMS attacks. Previous work in this area was limited to a more coarse-grained description of messaging volume.
- **Mechanism Development and Evaluation:** Using variations on well established and novel new approaches, we characterize the ability of a number of traffic engineering techniques from the Internet and telecommunications domains to mitigate such attacks. These mechanisms range in complexity and effectiveness and offer a range of solutions to service providers.

The remainder of this paper is organized as follows: Section II discusses pertinent related work; Section III provides an overview of cellular signaling networks and characterizes targeted SMS attacks; Section IV offers a number of solutions and mathematical measures of their ability to mitigate these attacks; Section V details simulations of the above solutions;

Section VI offers concluding remarks and discusses future work.

II. Related Work

Physical disconnection from external networks has long been one of the most effective means of providing security for communication systems. From small internal corporate networks to global telecommunications systems, "air-gap" separation has simplified the job of protecting networks from the majority of potential adversaries. Accordingly, the focus of security in these networks has typically centered around fraudulent access and billing. The authentication control messages between SS7 core network components, for example, was not available before 2002 [36]. The changing needs of users, however, have forced the gradual erosion of such well defined borders. Whether due to new access patterns (e.g. wireless access points, traveling users, etc) or the advent of new services (e.g. data networking in cellular telecommunications), many of the systems that once relied upon isolation as a major portion of their defenses are no longer able to do so. Security measures addressing new classes of threats are therefore essentially.

Telecommunications networks are not the only systems to suffer from vulnerabilities related to expanded connectivity. Systems including Bank of America's ATMs and 911 emergency services for Bellevue, Washington were both made inaccessible by the Slammer worm [24]. Although neither system was the target of this attack, simply being connected to the Internet made them experience significant collateral damage. Systems less directly connected to the Internet have also been subject to attack. Byers, et al. [9] demonstrated one such attack using simple automated scripts and webforms. Immense volumes of junk postal mail could then be used to launch denial of service (DoS) attacks on individuals.

The typical targets of DoS attacks, however, are more traditional online resources service. In 2000, for example, users were unable to reach Amazon, eBay and Yahoo! as their servers were bombarded with over a gigabit per second of traffic [31]. Since that time, sites ranging from software vendors [15] and news services [33] to online casinos [7] have all fallen victim to such attacks. While significant research has been dedicated to categorizing [23], mitigating [34], [17] and eliminating [38] such attacks, no solutions have seen widespread implementation. Because of the various transformations of data transiting between the Internet and telecommunications networks, the direct application of the above techniques would be ineffective.

Whether accidental or the result of malicious behavior, denial of service incidents have been studied and documented in telecommunications networks. The National Communications System published a study on the effects of text messages during emergency situations. Given realistic scenarios for usage, this technical bulletin argued that SMS resources needed to be increased 100-fold in order to operate under such conditions [27]. Operators have also reported problems with connectivity during holidays due to increased volumes of SMS traffic [22]. Enck, et al. [13] demonstrated that an adversary would be able to cause the same congestion in targeted metropolitan areas by injecting a relatively small amount of traffic. While a number of solutions were proposed in that work, none have yet been measured and compared.

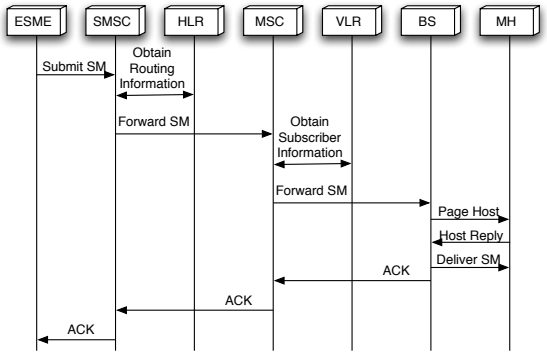

Fig. 1. A high level description of SMS delivery in an SS7 network.

III. System/Attack Characterization

A. Message Delivery Overview

In the following subsection, we provide a high-level, simplified tutorial on text message delivery in cellular networks.

1) Message Insertion: An Internet-originated SMS message can be generated by any one of a number of *External Short Messaging Entities* (ESMEs). ESMEs include devices and interfaces ranging from email and web-based messaging portals to service provider websites and voice mail services and can be attached to telecommunications networks either by dedicated connection or the Internet. When a message is injected into the network, it is delivered to the *Short Messaging Service Center* (SMSC). These servers are responsible for the execution of a "store-and-forward" protocol that eventually delivers text messages to their intended destination.

When a message is received from an ESME, it is examined by an SMSC. The contents and destination information from the message are then copied into a properly formatted packet. At this point, messages originating in the Internet and those created in the network itself become indistinguishable. Formatted text messages are then placed in an egress queue in the SMSC and await service.

2) Message Routing: Before an SMSC can forward a text message to a targeted mobile device, it must first determine the location of that device. To accomplish this, the SMSC queries a database known *Home Location Register* (HLR). The HLR is responsible for storing subscriber data including availability, billing information, available services and current location. With the help of other elements in the network, the HLR determines the routing information for the targeted device. If the desired phone is not available, the SMSC stores the message until a later time for subsequent retransmission. Otherwise, the SMSC receives the address of the *Mobile Switching Center* (MSC) providing it service. Through its attached *base stations* (BS), the MSC wirelessly delivers the text message. Figure 1 illustrates the path described above.

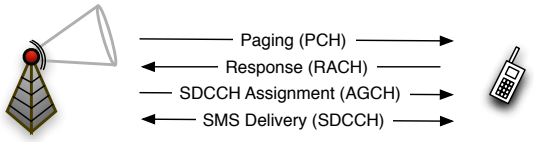

Fig. 2. An overview of SMS message delivery on the wireless or air interface. Incoming voice calls would follow a similar procedure except that they would receive a TCH after using the SDCCH.

3) Wireless Delivery: The air interface, or radio portion of the network, is traditionally divided into two main logical categories - the *Control Channels* (CCHs) and *Traffic Channels* (TCH). TCHs carry voice traffic after call setup has occurred. CCHs, which provide information about the network and assist in call setup/SMS delivery, are subclassified further. In order to alert a targeted device that a call or text message is available, a message is broadcast of the *Paging Channel* (PCH). Note that multiple base stations broadcast this page in an attempt to quickly determine the sector in which the targeted recipient is located. Upon hearing its temporary identifier on the PCH, available devices inform the network of their readiness to accept incoming communications using the slotted aloha-based *Random Access Channel* (RACH). A device is then assigned a *Standalone Dedicated Control Channel* (SDCCH) by listening to the *Access Grant Channel* (AGCH). If a text message is available, the base station authenticates the device, enables encryption, supplys a new temporary identifier (to preserve future anonymity) and then delivers the contents of the message over the assigned SDCCH. If instead a call is incoming for the device, the SDCCH is used to authenticate the device and negoriate a TCH for voice communications.

Figure 2 offers an overview of the wireless portion of message delivery.

B. System Vulnerability

All large scale attacks, whether targeting the digital or physical domain, evolve in the following phases: *recognition* (identification of a vulnerability), *reconnaisance* (characterization of the conditions necessary to attack the vulnerability), *exploit* (attacking the vulnerability) and *recovery* (cleanup and foresensics). We therefore approach targeted SMS attacks in the same fashion.

The vulnerability in GSM cellular networks that allows for targeted text message DoS attacks to occur is the result of bandwidth allocation on the air interface. Under normal operating conditions, the small ratio of bandwidth allocated to control versus traffic data is sufficient to deliver all messages with a low probability of blocking. However, because text messages use the same control channels as voice calls for delivery (SDCCHs), contention for resources occurs when SMS traffic is elevated. Given a sufficient number of SMS messages, each of which require on average four seconds for delivery [27], arriving voice calls will be blocked for lack of available resources.

Sending text messages to every possible phone number is not an effective means of attacking a network. The haphazard submission of messages is in fact more likely to overwhelm gateways between the Internet and telecommunications networks than to disrupt cellular service. An adversary must efficiently blanket only the targeted area with messages so as to reduce the probability of less effective collateral damage. The information to achieve such a goal, however, is readily available. Using tools including NPA-NXX Area Code Databases, Internet search engines and even feedback from service provider websites, an attacker can easily construct a "hit-list" of potential targets. Armed with this information, an adversary can then begin exploiting the bandwidth vulnerability.

The exploit itself involves saturating base station towers to their SDCCH capacity for some period of time. In so doing, the majority of attempts to establish voice calls are blocked. For all of Manhattan, a perfectly executed attack (against 12 SDCCCHs) would require the injection of only 165 messages per second. Because downtime in telecommunications networks has historically proven expensive [11], we more fully characterize these attacks such that effective solutions can be developed.

C. Attack Characterization

In order to judge the efficacy of any countermeasure against targeted SMS attacks, it is necessary to fully characterize such an event. We seek to understand the observed conditions and the subtle interplay of network components given a wide range of inputs. For example, because text messages injected as part of an attack potentially deviate from the traditionally assumed Poisson interarrival behavior, we look at attacks exhibiting a number of different flow characteristics. To achieve these ends, we have developed a detailed GSM simulator. The design considerations and verification of its accuracy are discussed in the Appendix.

A cellular deployment similar to that found in Manhattan [27], in which each of the 55 sectors in the city has 12 SDCCHs, is used in the base scenario [1]. In our simulations, call and SMS requests arrive throughout the city with a Poisson distribution with an average rate of $\lambda_{call} = 50K\ call/hour$ and $\lambda_{SMS} = 138.6K\ msg/hour$ (where λ represents an arrival rate). Voice calls occupy TCHs for an average 120 seconds and are exponentially distributed around this mean. Text messages and voice calls use SDCCHs for 4 [27] and 1.5 [28] seconds, respectively. Such values are well within standard operating conditions [20], [25], [21]. An area is observed for a total of 60 minutes, in which the middle 30 minutes are exposed to a targeted SMS attack during which SMS arrival rates are increased by approximately 4-13 times their normal rates ($\lambda_{SMS} = 165\ msg/sec$ (3 messages/second/sector) to $\lambda_{SMS} = 495\ msg/sec$ (9 messages/second/sector))[2] All results are the average of 1000 runs, each using randomly generated traffic patterns consistent with the above parameters.

[1]In reality, only the highest capacity sectors would be so overprovisioned [27], making this a conservative estimate for every sector in a city.

[2]Because DoS attacks on the Internet frequently exhibit increases of more than 1000 times normal traffic rates [], such an increase is relatively insignificant.

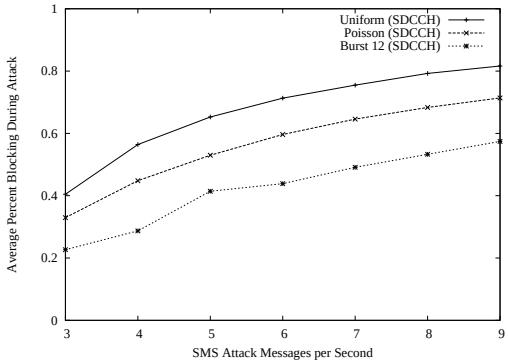

Fig. 3. The blocking probability for traffic exhibiting uniform and Poisson interarrival characteristics over varying attack strengths. Note that 3 messages/second/sector corresponds to an attack of 165 messages/second on Manhattan.

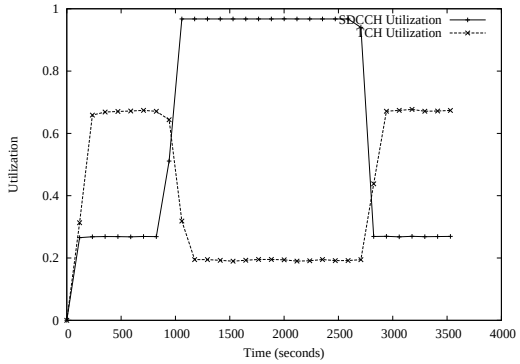

Fig. 4. The utilization of SDCCHs and TCHs for an attack exhibiting a Poisson interarrival at a rate of 495 messages/second.

Figure 3 shows the blocking probability for a number of traffic patterns and network condition. The most effective attack, which replicates the attack proposed in Enck, et al [13], sends a burst of 12 SMS messages in sequential frames once every four seconds. Whereas telecommunications networks are traditionally designed to experience blocking probabilities of less than 1% [29], [30], [6], this attack is able to prevent approximately 90.14% of all calls from being completed.

Because variability within the network is possible, we examined a number of attack flow types for which the perfect alignment of messages is virtually unachievable. For example, instead of holding an SDCCH for a constant period of time, incoming SMS and voice calls occupy their SDCCH for an average of 4 and 1.5 seconds, respectively, and are exponentially distributed around this mean. Figure 3 shows the probability of blocking for a sector under SMS attacks exhibiting uniform, Poisson and bursty interarrival characteristics. Notice that, due to the addition of variability, bursty attacks are the least successful of the three. This is because the next burst of incoming messages almost certainly experiences blocking on approximately half of the SDCCHs. Accordingly, some portion of SDCCHs are almost always available to legitimate traffic. The attack in which SMS messages are delivered at a uniform rate may also be difficult to achieve due to variability. In order to perform a more accurate study of these attacks, we therefore assume Poisson interarrival behavior for the rest of this research.

In our remaining experiments, we will use an attack of 495 messages/second, which is equal to 9 message/second/sector and yields a blocking probability of 71.34%. Figure **??** offers additional characterization of channel utilization. Notice that this rate is not significantly larger than that suggested in Enck, et al [13] and would only occupy less than 22% of the bandwidth of a single 56Kb SS7 signaling link.

IV. Traffic Management Techniques

Voice communications have traditionally received priority in telecommunications networks. Because voice has been the dominant means by which people interact via these networks, providers allow for the degredation of other services in order to achieve high availability for the voice services on these networks. There are, however, an increasing set of scenarios in which the priority of services begins to change.

On September 11th, 2001, service providers experienced significant surges in usage. Verizon Wireless reported the number of calls made increased by more than 100% above average levels. Cingular Wireless experienced an increase of over 1000% for calls bound for the greater Washington D.C area [27]. Although telecommunications networks are designed to operate in the presence of elevated traffic levels, these spikes were significantly above the capacity of even the best provisioned systems. In spite of the increased call volume, SMS messages were still received in even the most inundated areas because the control channels used for their delivery remained uncongested. In both emergency and day-to-day situations, the utility of text messaging has increased to the same level as voice communications for significant portions of the population.

For this reason, attractive mitigation solutions must not only protect voice services from the direct SMS attack, but also allow SMS service to continue. In particular, differentiated service for SMS delivery based upon the source of the SMS traffic is desireable. For instance, authenticated messages originated by emergency responders should be given higher priority than messages submitted by unauthenticated sources.

There are three traditional approaches to combating congestion. Typically, the most effective is to limit the rate of the traffic source, in this case the interfaces on which messages are submitted. Because this is not always effective, it is important for elements to protect the network by perhaps shedding traffic or using scheduling mechanisms. Finally, resources may be reallocated to alleviate the network bottleneck. We examine these solutions below.

A. Current Solutions

Cellular providers have introduced a number of mitigation solutions into phone networks to combat the SMS-based DOS

attacks. These solutions focus on *rate limiting* the source of the messages and are ineffective against all but the least sophisticated adversary. To illustrate, the primary countermeasure discovered by the authors of the original study was a per-source volume restriction at the SMS gateway [13]. Such restrictions would, for example, allow only 50 messages from a single IP address. The ability to spoof IP addresses, the existence of botnets, and wide availability of IP addresses renders this solution impotent. Another popular deployed solution filters SMS traffic based on the textual content. Similar to SPAM filtering, this approach is effective in eliminating undesirable traffic only if the content is predictable. However, an adversary can bypass this countermeasure by generating legitimate looking SMS traffic from randomly generated simple texts, e.g. "*I will meet you at Trader Joe's at 5:00pm. -Alice*"

Note that these and the overwhelming majority of other solutions deployed in response to the SMS vulnerability can be classified as *edge solutions*. Ineffective by construction, such solutions try to regulate the traffic flowing from the Internet into the provider network at its edge. Provider networks cover huge geographic areas and consist of hundreds of thousands of network elements. Any compromised element can be a conduit for malicious traffic. Moreover, if left unregulated, the connections between provider networks can also be exploited to inject SMS traffic.

Rate limitation is largely unattractive even within the core network. The distributed nature of Short Messaging Service Centers (SMSCs), through which all text messages flow, makes it difficult to coordinate real-time filtering in response to targeted attacks.

Therefore, for the purposes of this discussion, we assume that an adversary is able to successfully submit a large number of text messages into a cellular network. The defenses below are dedicated to protecting the resource that is being exploited in the SMS attack – the bandwidth constrained SDCCHs. Note that the Internet faces a similar conundrum: once dominant perimeter defenses are failing in the face of dissolving network borders, e.g., as caused by wireless connectivity and larger and more geographically distributed networks [19]. As is true in the Internet, we must look to other methods to protect telecommunications networks.

B. Queue Management Techniques

1) Weighted Fair Queueing: Because we cannot rely on rate limitation at the source of messages, we now explore network-based solutions. Fair Queueing [26] is a scheduling algorithm which separates flows into individual queues and then apportions bandwidth equally between them. Designed to emulate bit-wise interleaving, Fair Queueing services queues in a round-robin fashion. Packets are transmitted when their calculated interleaved finishing time is the shortest. Building priority into such a system is a simple task of assigning weights to flows. Known as *Weighted Fair Queueing* (WFQ) [10], this technique can be used to give incoming voice calls priority over SMS.

We apply WFQ to the service queues of the SDDCH. We create two waiting queues, one for voice requests and one for SMS requests, respectively. The size of the call queue is 6 and the size of the SMS queue is 12. We give a weight of two to the call queue.

We provide a simplfied analysis to characterize the performance of WFQ in this scenario. To determine the relative blocking probability and utilization of the voice and SMS flows, we begin by assuming the conditions set forth in Section III-C. WFQ can be approximated as a general processer sharing system (GPS) [35]. The average service rate of such systems is the weighted average of the service rates of all classes of service requests. In our case we have two types of request: voice requests with $\mu_{voice}^{-1} = 1.5$ seconds and $\lambda_{voice} = 0.2525$/second, and SMS requests with $\mu_{SMS}^{-1} = 4$ seconds and $\lambda_{SMS} = 9.7$/second. Therefore, for our system, $\mu^{-1} = 3.94/second$.

Although our system has multiple servers (SDDCHs), and is thus an M/M/m system, because it is operating at high loads during an attack, it may be approximated by an M/M/1 system with its $\mu = m\mu^{'}$, where $\mu^{'}$ is the service rate calculated above. Using these values, and accounting for the weighting of 2:1 for servicing call requests, the call request utilization $\rho_{call-queue} = 0.04$, and the expected queue occupancy is about 1%. Because the $\rho_{SMS-queue}$ is much greater than 1, its queue utilization is approximately 100%. When combined, the total queue occupancy is approximately 67%. These numbers indicate that the WFQ-based approach would sufficiently protect voice calls from targeted SMS attacks. Section V offers additional insight through simulation.

2) Weighted Random Early Detection: Active queue management has received a great deal of attention as a congestion avoidance mechanism in the Internet domain. *Random Early Detection* (RED) [14], [8], one of the better known techniques from this field, is a particularly effective means of coping with potentially damaging quantites of text messages. While traditionally used to address TCP congestion, RED helps to prevent queue lockout and was therefore investigated. RED drops packets arriving to a queue with a probability that is a function of the weighted queue occupancy average. Packets arriving to a queue capacity below a threshold, t_{min}, are never dropped. Packets arriving to a queue capacity above some value t_{max} are always dropped. Between t_{min} and t_{max}, packets are dropped with a linearly increasing probability. This probability, p_{drop}, is calculated as follows[3]:

$$p_{drop} = p_{drop-max} * (Q_{avg} - t_{min})/(t_{max} - t_{min}) \quad (1)$$

The advantages to this approach are twofold: first, lockout becomes more difficult as packets are purposefully dropped with greater frequency; secondly, because the capacity of busy queues stays closer to a moving average and not capacity, space typically exists to accomodate sudden bursts of traffic. However, one of the chief difficulties with traditional RED is that it eliminates the ability of a provider to offer quality of service (QoS) guarantees. Because all traffic entering a queue is dropped with equal probability, ensuring that the most time sensative messages arrive quickly becomes difficult. *Weighted*

[3]Some variants of RED additionally incorporate a *count* variable. Equation 1 is the simplest version of RED defined by RFC 2309 [8].

Random Early Detection (WRED) solves this problem by basing the probability a given incoming messages is dropped on an attribute such as its contents, source or destination. Arriving messages not meeting some priority are therefore subject to increased probability of drop. The dropping probability for each class of message is tuned by setting t_{min} and t_{max} for each class.

We consider the use of authentication as a means of creating messaging priority classes. For example, during a crisis, messages injected to a network from the Internet by an authenticated municipality or from emergency personel could receive priority over all other SMS messages. A number of municipalities already use such systems for emergency [32] and traffic updates [37]. Messages from authenticated users within the network itself receive secondary priority. Unauthenticated messages originating from the Internet are delivered with the lowest priority. Such a system would allow the informative messages (i.e. evacuation plans, additional warnings, etc) to be quickly distributed amongst the population. The remaining messages would then be delivered at ratios corresponding to their priority level. We assume that packet priority marking occurs at the SMSCs such that additional computational burden is not placed on base stations.

Here we show how using WRED, we can provide differentiated service to different classed of SMS traffic using the attack scenario described in Section III-C. In this example we assume messages arrive with the following distribution: 10% priority 1, 10% priority 2, and 80% priority 3. To accomodate sudden burts of high priority traffic, we choose an SMS queue size of 12. Because we desire low latency delivery of high priority messages, we target an average queue occupancy $Q_{ave} = 3$.

To meet this objective, we must set t_{min} and t_{max}. For M/M/n systems with a finite queue of size m, the number of messages in the queue, N_Q, is

$$N_Q = P_Q \frac{\rho}{1-\rho} \tag{2}$$

where

$$P_Q = \frac{p_0 (m\rho)^m}{m!(1-\rho)} \tag{3}$$

where

$$p_0 = \left[\sum_{n=0}^{m-1} \frac{(m\rho)^n}{n!} + \frac{(m\rho)^m}{m!(1-\rho)} \right]^{-1} \tag{4}$$

Setting $N_Q = 3$, we derived a target load $\rho_{target} = 0.855$. ρ_{target} is the utilization desired at the queue. Thus, the packet dropping caused by WRED must reduce the actual utilization, ρ_{actual}, caused by the heavy offered load during an attack, to be reduced to ρ_{target}. Therefore

$$\rho_{target} = \rho_{actual}(1 - P_{drop}) \tag{5}$$

where P_{drop} is the overall dropping probability of WRED. For an attack with average arrival rate of $9.7\ msg/sec$ ($\lambda = 9.7$), $\rho_{actual} = 3.23333$. Solving for P_{drop},

$$P_{drop} = 1 - \frac{\rho_{target}}{\rho_{actual}} = 0.735567 \tag{6}$$

P_{drop} can be calculated from the dropping probabilties of the individaul classes of messages by

$$P_{drop} = \frac{P_{drop \cdot 1} \cdot \lambda_1 + P_{drop \cdot 2} \cdot \lambda_2 + P_{drop \cdot 3} \cdot \lambda_3}{\lambda_{total}} \tag{7}$$

Because we desire to deliver all messages of priroty 1 and 2, we set $P_{drop \cdot 1} = P_{drop \cdot 2} = 0$. Using equation 7, we find $P_{drop \cdot 3} = 0.746339$. This value can then be used in conjunction with Equation 1 to determine t_{min} and t_{max}.

The desired average queue occupancy, Q_{ave}, is 3. From equation 1, t_{min} must be an integer less than the average queue occupancy. This leaves three possible values for t_{min}: 0, 1, and 2. The best fit is found when $t_{min} = 0$ and $t_{max} = 4$, resulting in 75% dropping of priority 3 traffic.

Using this method it is possible to set thresholds to meet delivery targets. Of course, depending on the intensity of an attack, it may not be possible to meet desired targets according to equation 7, i.e., it may not be possible to limit blocking to only low priorty traffic. While the method outlined here provides just an approximate solution, given the quantization error in setting t_{min} and t_{max} (they must be integers), we beleive the method is sufficient. We provide more insight into the performance of WRED in Section V.

C. Air Interface Provisioning

The difficulty with the above methods is that they do not deal with the system bottleneck directly; rather, they sacrifice the quality of service for one type of flow over another. Attempts to reallocate the bandwidth available to message delivery would therefore have a greater impact in combating targeted SMS attacks. We therefore investigate a variety of techniques that modify the way in which the air interface is used.

To analyze these techniques we resort to simple Erlang-B queuing analysis. We prevent a brief background here. For more details see [35]. In a system with N servers, and an offered load in Erlangs of A, the probability that an arriving request is blocked because all servers are occupied is given by:

$$P_B = \frac{\frac{A^N}{N!}}{\sum_{l=0}^{l=N-1} \frac{Al}{l!}} \tag{8}$$

The load in Erlangs is the same as the utilization, ρ, in a queuing system; it is simply the offered load multiplied by the service time of the resource. The expected occupancy of the servers is given by:

$$E(n) = \rho(1 - P_B) \tag{9}$$

In our system, the SDCCHs are the servers.

1) Strict Resource Provisioning: Under normal conditions, the resources for service setup and delivery are overprovisioned. At a rate of 50,000 calls per hour in our baseline scenario, for example, the calculated average utilization of SDCCHs per sector is approximately 2%. Given this observation, if a subset of the total SDCCHs can be made available only to voice calls, blocking due to targeted SMS attacks can

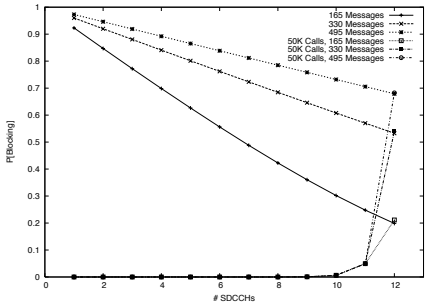

Fig. 5. The probability that incoming calls and SMS messages are blocked in a system implementing SRP. The allocated number of SDCCHs (x-axis) is listed in terms of those to which SMS delivery is restricted.

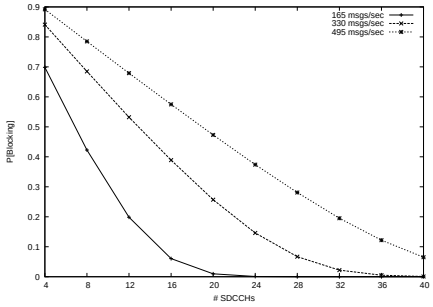

Fig. 6. The probability of an incoming call/message blocking in a sector for a varying number of SDCCHs

be significantly mitigated. Our first air interface provisioning technique, *Strict Resource Provisoning* (SRP), attempts to address this contention by allowing text messages to occupy only a subset of the total number of SDCCHs in a sector. Requests for incoming voice calls can compete for the entire set of SDCCHs, including the subset used for SMS. In order to determine appropriate parameters for systems using SRP, we apply equations 8 and 9.

To illustrate the effectiveness of SRP, we consider a system with no queue. Figure 5 shows the blocking probabilites for a system using SRP when we vary the number of SDCCHs that will accept SMS requests from 0 (none) to 12 (all). Because incoming text messages only compete with voice calls for a subset of the resources, any resulting call blocking is strictly a function of the size of the subset of voice-only SDCCHs. The attacks of intensity 165, 330 and 495 messages per second have virtually no impact on voice calls until the full complement of SDCCHs are made available to all traffic. In fact, it is not until 10 SDCCHs are made available to SMS traffic that the blocking probability for incoming voice calls reaches 1%.

By limiting the number of SDCCHs that will serve SMS requests, the blocking for SMS is increased. When only six SDCCHs are available to text messages, blocking probabilities for SMS are as high as 84%. Because significant numbers of people rely upon text messaging as their primary means of communication, such parameters should be carefully tuned. We will discuss the impact of additional factors after examining the results of simulation in Section V.

2) Dynamic Resource Provisioning: While SRP reprovisions capacity on existing SDCCHs, other over-provisioned resources in the sector could be manipulated to alleviate SDCCH congestion. For example, at a rate of 50,000 calls per hour, each sector uses an average of 67% of its TCHs. If a small number of unused TCHs could be repurposed as SDCCHs, additional bandwidth could be provided to mitigate such attacks.

Our second air interface technique, *Dynamic Resource Provisioning* attempts to mitigate targeted text messaging attacks by reclaiming a number of TCHs (up to some limit) for use as SDCCHs. This approach is highly practical for a number of reasons. First, increasing the bandwidth (762 bits/second) of individual SDCCHs is difficult without making significant changes to the either the radio encoding or the architecture of the air interface itself. Because major changes to the network are extremely expensive and typically occur over the course of many years, such fixes are not appropriate in the short term. Secondly, dynamically reclaiming channels allows the network to adjust itself to current conditions. During busy hours such as morning and evening commutes, for example, channels temporarily used as SDCCHs can be returned to the pool of TCHs to accomodate elevated voice traffic needs. Lastly, because SDCCHs are assigned via the AGCH, allocating incoming requests to seemingly random timeslots requires almost no changes to handset software.

Figure 6 demonstrates the blocking probability for incoming calls and text messages in a sector using DRP to add a variable number of SDCCHs. Again, no queue was used. The ability of an attacker to block all channels is significantly reduced as the number of SDCCHs increases. Attackers are therefore forced to increase the intensity of their attack in order to maintain its potency. For attacks at a rate of 165 messages/second, doubling the number of available SDCCHs reduces the calculated blocking caused by an attack by two orders of magnitude. The blocking probability caused by attacks at higher rates, in which the number of Erlangs is greater than the number of SDCCHs, decreases in roughly a linear relationship to the number of SDCCHs added.

One potential drawback with DRP is that by substracting TCHs from the system, it is possible to increase call blocking because of TCH exhaustion. In fact, the reclamation of TCHs for use as SDCCHs increases the blocking probability for voice calls from 0.2% in the base scenario (45 TCHs, 12 SDCCHs) to 1.5% where 40 SDCCHs are available (a reduction to 38 TCHs). Section V offers additional insight into the tradeoffs inherent to this scheme.

3) Direct Channel Allocation: The ideal means of eliminating the competition for resources between call setup and SMS delivery would be through the separation of shared mechanisms. Specifically, delivering text messages and incoming call requests over mutually exclusive sets of channels would prevent these flows from intefering with each other. The challenge of implementing such a mechanism is to do so without requiring significant restructuring of the network architecture. As

previously mentioned, such fundamental changes in network operation are typically too expensive and time consuming to be considered in the short term. While the SRP technique provides a rudimentary separation, it is possible to further separation of these two types of traffic.

As mentioned in the previous section, DRP is easily implementable because the AGCH specifies the location of the SDCCH allocated for a specific session. After call requests finish using their assigned SDCCH, they are instructed to listen to a specific TCH. Because the use of a TCH is the eventual goal of incoming voice calls, it is therefore possible to shortcut the use of SDCCHs for call setup. Incoming calls could therefore be directed to a TCH, leaving SDCCHs exclusively for the delivery of SMS messages. This technique, which we refer to as *Direct Channel Allocation* (DCA), removes the shared SDCCH channels as the system bottleneck.

Calculating blocking probabilites for a system implementing DCA is a simple matter of analyzing SDCCH and TCH blocking for the two independent flows. For 165 messages/second, text messages have a calculated blocking probability of approximately 20%. This value increases to 68% as the attack intensity increases to 495 messages/second. Voice calls, at an average rate of 50,000/hour, have a blocking probability of 0.2%. Note that because the shared bottleneck has been removed, it becomes extremely difficult for targeted text messsaging attacks to have any effect on voice communications. In Section V, we will highlight these new potential points of contention.

V. Experimental Results

In order to characterize each of our proposed mitigation techinques, we simulate attacks against networks with the same parameters used in Section III. Attacks exhibit Poisson interarrival characteristics and arrive at an average rate of 9 messages/second/sector. This is equivalent to an attack on Manhattan with a rate of 495 messages/second. These messages were marked as follows: 10% emergency, 10% users within the network and 80% originating from the Internet as part of an attack. Blocking on the RACH, the parameters of which were set using optimal settings [] was not a factor in these experiments.

A. Queue Management Strategies

1) Weighted Fair Queueing: In order to implement queue management techniques, buffers were added to the system before messages are assigned to SDCCHs. If all SDCCHs are occupied, newly arriving voice and SMS messages are placed into their own queues. While a number of different buffer sizes were considered and examined, the following queue management experiments occur in the presence of two buffers - one of size 12 for SMS and a second of size 6 for voice. Matching the number of SDCCHs, queues of this size offer the a good tradeoff between message delay and protection from message overload. Note that buffer size alone is not sufficient to protect against congestion [26], [18]. Queues are work conserving and are served in a round robin fashion, with voice calls receiving a preferential weighting of 2 to 1 over text messages.

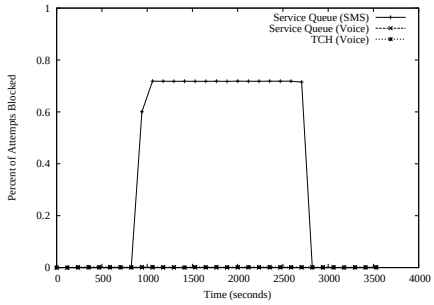

Fig. 7. The simulated blocking probability for a sector implementing WFQ. Notice that voice calls are unaffected by the attack, whereas the majority of text messages are dropped.

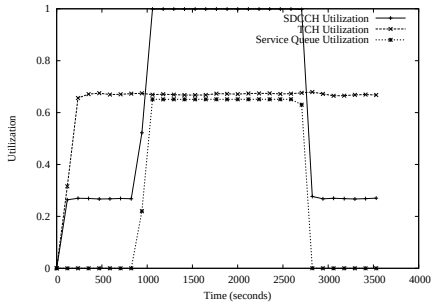

Fig. 8. The simulated utilization for a sector implementing WFQ. Notice that TCH utilization remains constant throughout the attack.

Figure 7 illustrates the resulting blocking for a sector implementing WFQ. The preferential treatment of voice traffic eliminates the blocking previously seen in an unprotected system. Incoming text messages, however, continue to experience roughly the same probability (71.8%) of blocking observed by all traffic in the base attack scenario. As is shown in Figure 8, the queue itself does nothing to prevent congestion. Total queue utilization is 65.1%. As two-thirds of the queue space is available to text messaging, this represents a near total average occupancy of the SMS queue and a virtually unused voice traffic queue. Such an observation confirms our analytical results.

The advantage to implementing the WFQ mechanism is not only its relative simplicity, but also its effectiveness in preventing degredation of voice services during targeted SMS attacks. Unfortunately, the granularity for prioritizing text messages is insufficient to provide adequate service to those users relying upon text messaging as their dominant means of communication. We discuss means of adding such granularity through the use of WRED.

2) Weighted Random Early Detection: While WFQ could be expanded to provide prioritization for flows with different origins through the use of multiple queues, the increased complexity of managing such a system as the number of

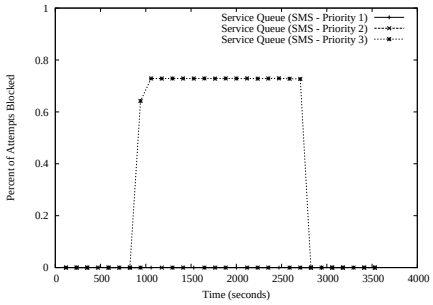

Fig. 9. The simulated blocking probability for a sector implementing WRED. Unlike WFQ, only Internet-originated text messages are dropped at an elevated frequency.

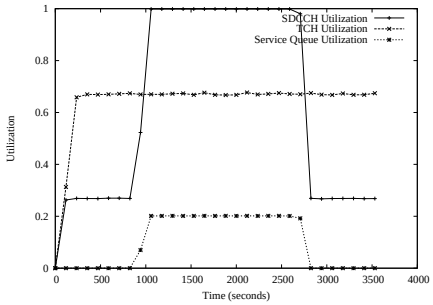

Fig. 10. The simulated utilization for a sector implementing WRED. Notice that the queue occupancy stays low due to the decreased priority of Internet-originated messages.

flows grows quickly becomes unmanageable. The use of a prioritized dropping policy allows a system to offer similar prioritization while maintaining only a single queue. In our implementation of WRED, we assume that SMS traffic is marked upstream as having either high ($t_{high,max} = t_{high,min} = 12$), medium ($t_{med,max} = 10, t_{med,min} = 6$) or low ($t_{low,max} = 2, t_{low,min} = 1$ priority. These priorities correspond directly to emergency priority users, network customers and Internet-originated messages, respectively. Dropping decisions are made in an event-driven fashion [8] with a $p_{drop-max}$ of 1 for all flows and a weight of 0.8 on the most recent sample data use for determining the average queue length. Like the previous queue management technique, a queue of size 12 is allocated for both voice and text messages.

Figure 9 gives the blocking probabilities for each of the three priorities of text messages. Because voice calls never block in these simulations, they are ommitted from this graph. Both high and medium priority flows experienced a blocking probability of zero throughout all of the simulations. The blocking of Internet-originated messages averages 72.8%, approximately the same blocking probability experienced my all incoming messages in the base attack scenarios. Service queue utilization, shown in Figure 10, corresponds with WAIT FOR WILL'S DATA.

Systems implementing WRED not only match the elimination of voice call blocking seen through the use of WFQ, but also offer significantly improved performance in terms of message delivery. Implementing this solution, however, faces its own challenges. The verification of high priority messages, for example, would require the use of additional infrastructure. High priority messages originating outside the network, such as emergency messages distributed by a city, may require the use of a dedicated line and/or the use of public keys for verification. Because of historical difficulties effectively achieving the latter [12], implementing such a system may prove difficult. Even with such protections, this mechanism fails to protect the system against insider attacks. If the machine responsible for sending high priority messages into the network or user phones are compromised by malware, systems implementing WRED lose their messaging performance improvements over the WFQ solution. Note that networks not bounding priority to specific geographic regions cab potentially be attack through any compromised high priority device.

B. Air Interface Strategies

1) Strict Resource Provisioning: Before characterizing the SRP technique, careful consideration was given to the selection of operating parameters. Because many MSCs are capable of processing up to 500,000 calls per hour, we engineer our solution to be robust to large spikes in traffic. We therefore allow SMS traffic to use 6 of the 12 total SDCCHs, which yields a blocking probability of 1% when voice traffic requests reach 250,000 per hour. Note that calls would experience an average blocking probability of 70.6% due to a lack of TCHs with requests at this intensity. Because these networks are designed to operate dependably during elevated traffic conditions, we believe that the above settings are realistic.

The blocking probabilites for SMS and voice flows in a sector implementing SRP are shown in Figure 11. Because SRP prevents text messages from competing for all possible SDCCHs, voice calls experience no blocking on the SDCCHs throughout the duration of the attack. Text messages, however, are blocked at a rate of 82.8%. Channel utilization, illustrated in Figure 14, gives additional insight into network conditions. Because calling behavior remains the same during the attack, the resources allocated by the network are more than sufficient to provide voice service to users. By design, SDCCH utilization plateaus well below full capacity. Whereas the SDCCHs used by text messages have an average utilization of 96.9%, the SDCCHs used by incoming voice calls average a utilization of 6.3%. This under-use of resources, represents a potential loss of utility as the majority of text messages (legitimate or otherwise) go undelivered.

The difficulty with this solution becomes correct parameter setting. While theoretical results indicated that allocating up to 10 SDCCHs only increased call blocking to 1%, voice traffic volumes fluctuate throughout the day. Provisioning resources in a static fashion must account for worst-case scenarios and therefore leads to conservative parameter settings. While protecting the network from an attack, such a mechanism may actually hinder the efficiency of normal operation. When traffic

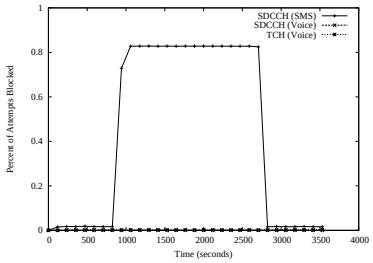

Fig. 11. SRP Blocking

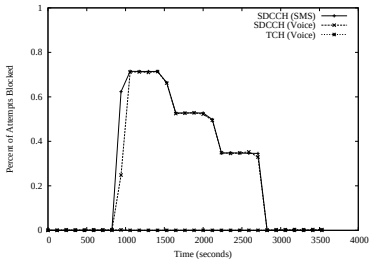

Fig. 12. DRP Blocking

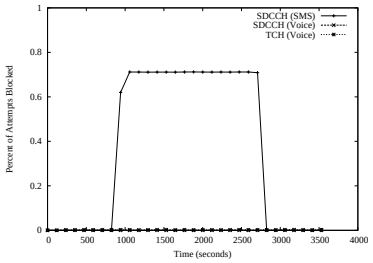

Fig. 13. DCA Blocking

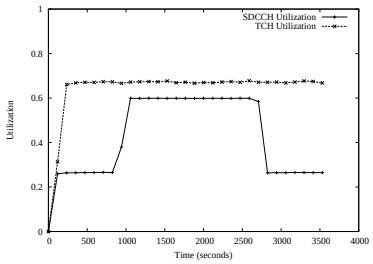

Fig. 14. SRP Utilization

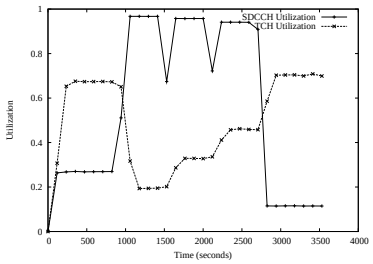

Fig. 15. DRP Utilization

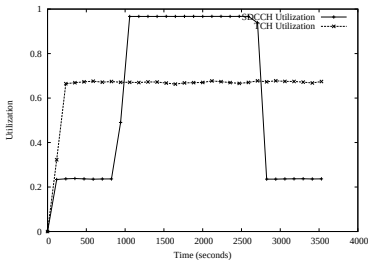

Fig. 16. DCA Utilization

channels are naturally saturated, as may be common during an emergency, such hard limits actually prevent users from communicating. Determining the correct balance between insulation from attacks and resource utilization becomes non-trivial. Accordingly, we look to our other techniques for more complete solutions.

2) Dynamic Resource Provisioning: Although it is possible to reclaim any number of TCHs for use as SDCCHs under the DRP mechanism, we limited the candidate number of channels for this conversion two. In these experiments, a single TCH was repurposed into 8 SDCCHs every 10 minutes during the attack. This separation was designed to allow the network to return to steady state between channel allocations. While converting only two channels is not enough to completely eliminate attacks at high intensities, our goal is to understand the behavior of this mechanism.

The blocking probabilities for SMS and voice flows in a sector implementing the DRP technique are illustrated in Figure 12. As TCHs are converted for use as SDCCHs, the blocking probabilites for both incoming SMS and voice requests fall from 71.19% to 52.55% and eventually 34.88%. This represents a total reduction of the blocking probability by approximately half. The reduced number of available TCHs results in no additional blocking for voice calls. Figure 15 illustrates a gradual return towards pre-attack TCH utilization levels as additional SDCCHs are allocated. The effects of the reprovisioning are also obvious for SDCCH utilization. The downward spikes represent the sudden influx of additional, temporarily unused channels. While SDCCH utilization quickly returns to nearly identical levels after each reallocation, more voice calls are able to be completed due to a decrease probability of the attack holding all SDCCHs at any given time.

As was a problem for SRP, determining the correct parameters for DRP is a difficult undertaking. The selection of two TCHs for conversion to SDCCHs illustrates the utility of this mechanism, but is not sufficient for real settings. To reduce the blocking probability on SDCCHs below the values observed for TCHs, a total of 48 SDCCHs would have to be made available. This leaves 39 TCHs, with a call blocking rate of 2.1%, for use by voice calls. Elevations in the volume of voice calls would likely require the release of some number of reclaimed TCHs to be repurposed to their original use. The decision to convert channels is also non-trivial. Whereas the decision to reallocate channels at specific times was decided statically, dynamically determining these parameters would prove significantly more challenging. Basing reclaimation decisions on small observation windows, while offering greater responsiveness, may result in decreased resource use due to thrashing. If the observation window becomes too large, an attack may end before appropriate action can be taken. As was observed for SRP, the static allocation of additional SDCCHs faces similar inflexibility problems. Low resource utilization under normal operating conditions again represent a potential loss of opportunity and revenue.

3) Direct Channel Allocation: To simulate the DCA mechanism, incoming voice calls skip directly from the RACH to the next available TCH. An average of 1.5 additional seconds was added to each incoming call to replicate the processing formerly occurring on an SDCCH. As is shown in Figure 13, voice calls arriving in a sector implementing the DCA scheme experience no additional blocking during a targeted SMS attack. The decoupling of these mechanisms limits similar denial of service attacks to the RACH, which has exhibited no call dropping throughout the entirity of our experiments. Figure 16 confirms the results in the previous

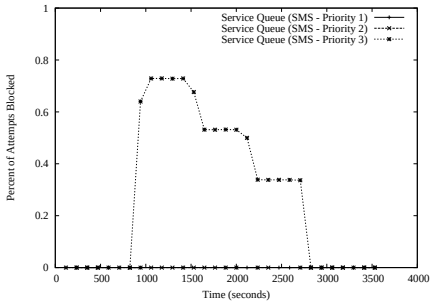

Fig. 17. Blocking probabilities for a system implementing both WRED and DRP.

figure by showing the constant TCH utilization throughout the duration of the attack. No additional assistance is provided to assist in the delivery of text messages under DCA.

While removing the bottleneck on the shared path of SMS delivery and voice call setup, DCA potentially introduces new vulnerabilities into the network. The advantage of using SDCCHs to perform authentication before calls is that TCHs are never allocated to anyone but authorized users. Under the DCA model, however, valuable traffic channels can be occupied before users are ever verified. Using a single phone planted in a targeted area, an attacker could simply respond to all paging messages and then ignore all future communications from the network. Because there are legitimate reasons including low signal to wait tens of seconds for a phone to reply, an attacker could force the network to open and maintain state for multiple connections that would eventually go unused. Note that because paging for individual phones occurs over multiple sectors, a single rogue phone could quickly create a black-hole effect. Such an attack is very similar to a combination of ARP spoofing [] and the classic 'syn' attack observed throughout the Internet. While seemingly the most complete, the potential for additional damage made possible because of the DCA approach should be carefully considered.

C. Combined Approaches

There is no "silver-bullet" for maintaining a high quality of service for both text messaging and voice calls during a targeted SMS attack. As the above techniques demonstrate, each potential solution has its own weaknesses. Combining of such solutions, however, offers techniques robust to a wider array of threats. We examine two instances for which the fusion of mechanisms provides additional protections.

1) DRP and WRED: While directly addressing the bandwidth issue that makes targeted SMS attacks possible, the DRP technique lacks granularity to separate incoming voice and SMS requests. WRED, on the other hand, provides such traffic classification but is unable to react to attacks originating from trusted sources. Figure 17 shows the result of the combination of the two techniques.

2) DRP and SRP: **DRP and SRP -** Gives DRP granularity and allocates extra channels to SRP.

Please add your thoughts!

VI. Conclusion

This is a great place for a conclusion!

References

[1] OPNET Network Simulator. `http://www.opnet.com/`.

[2] The Network Simulator – ns-2. `http://www.isi.edu/nsnam/ns/`.

[3] Young 'prefer texting to calls'. `http://news.bbc.co.uk/2/hi/business/2985072.stm`, June 2003.

[4] 3rd Generation Partnership Project. Physical layer on the radio path; General description. Technical Report 3GPP TS 04.18 v8.26.0.

[5] 3rd Generation Partnership Project. Physical layer on the radio path; General description. Technical Report 3GPP TS 05.01 v8.9.0.

[6] A. Acampora and M. Naghshineh. Control and Quality-of-Service Provisioning in High- Speed Microcellular Networks. *IEEE Personal Communications*, 1(2):36–43, 1994.

[7] S. Berinato. Online Extortion – How a Bookmaker and a Whiz Kid Took On an Extortionist and Won. *CSO Online*, May 2005.

[8] B. Branden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski, and L. Zhang. RFC 2309 - Recommendations on Queue Management and congestion Avoidance in the Internet. rfc2309.txt, 1998.

[9] S. Byers, A. Rubin, and D. Kormann. Defending Against an Internet-based Attack on the Physical World. *ACM Transactions on Internet Technology (TOIT)*, 4(3):239–254, August 2004.

[10] A. Demers, S. Keshav, and S. Shenker. Analysis and Simulation of a Fair Queueing Algorithm. In *Proceedings of ACM SIGCOMM*, pages 3–12, 1989.

[11] L. Dryburgh and J. Hewett. Signaling System No. 7: The Role of SS7. `http://www.ciscopress.com/articles/article.asp?p=330805\&rl=1`, 2004.

[12] C. M. Ellison and B. Schneier. Ten Risks of PKI: What You're Not Being Told About Public-Key Infrastructure. *Computer Security Journal*, 16(1):1–7, 1999.

[13] W. Enck, P. Traynor, T. F. La Porta, and P. McDaniel. Exploiting Open Functionality in SMS-Capable Cellular Networks. In *Proceedings of the ACM Conference on Computer and Communication Security (CCS)*, November 2005.

[14] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1993.

[15] C. Haney. NAI is latest DoS victim. *IDG News Service*, February 5 2001.

[16] J. Hedden. Math::Random::MT::Auto - Auto-seeded Mersenne Twister PRNGs. `http://search.cpan.org/~jdhedden/Math-Random-MT-Auto-5.01/lib/Math/Rand%om/MT/Auto.pm`. Version 5.01.

[17] J. Ioannidis and S. Bellovin. Implementing Pushback: Router-Based Defense Against DDoS Attacks. In *Proceedings of Network and Distributed System Security Symposium (NDSS)*, February 2002.

[18] R. Jain. Myths about congestion management in high speed networks. *Internetworking: Research and Experience*, 3:101–113, 1992.

[19] G. Kunene. Perimeter Security Ain't What It Used to Be, Experts Say. *DevX.com*, 2004.

[20] Lucent Technologies. 5ESS(R) 2000 - Switch Mobile Switching Centre (MSC) for Service Providers. `http://www.lucent.com/products/solution/0,,CTID+2019-STID+10048-SOID+82%4-LOCL+1,00.html`, 2006.

[21] K. Maney. Surge in text messaging makes cell operators :-). July 27 2005.

[22] Mike Grenville. Operators: Celebration Messages Overload SMS Network. `http://www.160characters.org/news.php?action=view&nid=819`, November 2003.

[23] J. Mirkovic and P. Reiher. A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2):39–53, 2004.

[24] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the Slammer Worm. *IEEE Security and Privacy*, 1(4), July 2003.

[25] Motorola Corporation. Motorola GSM Solutions. `www.motorola.com/networkoperators/pdfs/GSM-Solutions.pdf`, 2006.

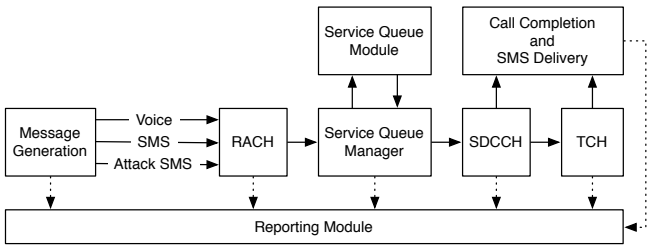

Fig. 18. Simulator Architecture

[26] J. B. Nagle. On Packet Switches with Infinite Storage. *IEEE Transactions on Communications*, COM-35(4), April 1987.
[27] National Communications System. SMS over SS7. Technical Report Technical Information Bulletin 03-2 (NCS TIB 03-2), December 2003.
[28] Nyquetek, Inc. Wireless Priority Service for National Security. `http://wireless.fcc.gov/releases/da051650PublicUse.pdf`, 2002.
[29] R. Ramjee, R. Nagarajan, and D. F. Towsley. On optimal call admission control in cellular networks. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, pages 43–50, 1996.
[30] R. F. Rey, editor. *Engineering and Operations in the Bell System*. Bell Telephone Laboratories, INC, second edition, 1984.
[31] M. Richtel. Yahoo Attributes a Lengthy Service Failure to an Attack. *The New York Times*, February 8 2000.
[32] Roam Secure. 17 Counties & Cities in Washington, DC Region deploy Roam Secure Alert Network. `http://www.roamsecure.net/story.php?news_id=52`, September 2005.
[33] P. Roberts. Al-Jazeera Sites Hit With Denial-of-Service Attacks. *PCWorld Magazine*, March 26 2003.
[34] S. Savage, D. Wetherall, A. Karlin, and T. Anderson. Practical network support for IP traceback. In *Proceedings of ACM SIGCOMM*, pages 295–306, October 2000.
[35] M. Schwartz. . In *Telecommunication Networks - Protocols, Modeling and Analysis*. Addison-Wesley Publishing Company, 1987.
[36] G. Shannon. Security Vulnerabilities in Protocols. In *Proceedings of ITU-T Workshop on Security*, May 13-14, 2002.
[37] Tamara Neale. VDOT LAUNCHES NEW 511 EMAIL ALERT SERVICE. `http://www.virginiadot.org/infoservice/news/newsrelease.asp?ID=CO-511-0%6`, February 2006.
[38] B. Waters, A. Juels, J. Halderman, and E. Felten. New client puzzle outsourcing techniques for DoS resistance. In *Proceedings of ACM CCS'04*, pages 246–256, 2004.

APPENDIX

While analytics easily characterize simplistic network loads, it is difficult to analyze multi-faceted input flows. To more fully characterize targeted SMS attacks and mitigation techniques, we developed an extensible, detailed GSM simulator. A number of simulation environments, including OPNET [1] and NS2 [2], were considered. The GSM air interface modules for both OPNET and NS2 focus on mobility management and did not contain support for SMS. More importantly, neither designs were amenable to modeling potential countermeasures. Commercial GSM simulators were also considered, but were highly complex, specialized, and cost prohibitive. Our goals became two-fold: to create an open source, freely available GSM simulator that would allow us to model the SMS and other attacks and potential countermeasures.

The core simulator is over 5,000 lines of C with supplement scripts written in various languages. In total, the project contains over 8,000 lines of code. All code was written with flexibility of configuration and extensibility in mind. Figure 18 depicts our simulator architecture. Solid lines indicate voice and SMS message flow; dashed lines indicate reporting of events, including creation, stage entrance, blocking, and completion. Simulation begins in the *Message Generation stage*, where messages are randomly created using a Mersenne Twister Pseudo Random Number Generator [16]. Message generation can follow Poisson, uniform, or bursty arrival patterns. After creation, messages proceed to the *RACH stage*, which strictly follows 3GPP TS 04.18 [4] and is tunable using the *max_retrans* and *tx_integer* variables. The *Service Queue Manager stage* assigns messages to an SDCCH. If desired, a pluggable *Service Queue Module* can be defined using standard interface callback functions. If possible, the Service Queue Manager assigns a message to an SDCCH. Rather than simulating exact communication and compensating for retransmittion, messages are held in the *SDCCH stage* for an exponential mean service time corresponding to message type. For accuracy, each SDCCH services messages by decreasing counters only during frames defined in 3GPP TS 05.01 [5]. When counters reach zero, SMS messages complete and voice messages attempt to aquire a TCH. Like the SDCCH stage, the *TCH stage* uses an exponential mean hold time to simulate channel occupancy. TCHs service messages during every frame, and when the hold time counter reachs zero, the call is complete, and the TCH is released.

The accuracy of the simulator was verified by running simple base scenarios and comparing blocking and utilization to values obtained using equations 8 and 9. Base scenarios similar to those in the paper were created: 12 SDCCHs, 45 TCHs, voice and SMS service times with exponential means of 1.5 and 4.0 seconds respectively, voice holding time with exponential mean of 120 seconds, voice call loads representing 25K, 50K, 75K, and 100K calls per hour to Manhattan, and SMS messages loads of 1, 2, and 3 messages per second to a sector. Voice and SMS loads were simulated separately and the average blocking and utilization of 1,000 runs was compiled. Both blocking and utilization are measured on a scale from 0 to 1.0. Statistical analysis showed that 95% of all samples fell within 0.006 of the mean, with many close to 0.001. Compared to calculated blocking and utilization, all simulated means deviated less than 20%, with all utilization means less than 2%. All but one simulated blocking mean was withing 10% calculated values. The outlier was resulted from simulating a very small blocking value and was off by -5×10^{-4}, which therefore is acceptable.