Understanding the Network-Level Behavior of Spammers

Anirudh Ramachandran and Nick Feamster
College of Computing, Georgia Tech
{avr, feamster}@cc.gatech.edu

ABSTRACT

This paper studies the *network-level* behavior of spammers, including: IP address ranges that send the most spam, common spamming modes (*e.g.*, BGP route hijacking, bots), how persistent across time each spamming host is, and characteristics of spamming botnets. We try to answer these questions by analyzing a 17-month trace of over 10 million spam messages collected at an Internet "spam sinkhole", and by correlating this data with the results of IP-based blacklist lookups, passive TCP fingerprinting information, routing information, and botnet "command and control" traces.

We find that most spam is being sent from a few regions of IP address space, and that spammers appear to be using transient "bots" that send only a few pieces of email over very short periods of time. Finally, a small, yet non-negligible, amount of spam is received from IP addresses that correspond to short-lived BGP routes, typically for hijacked prefixes. These trends suggest that developing algorithms to identify botnet membership, filtering email messages based on *network-level* properties (which are less variable than email content), and improving the security of the Internet routing infrastructure, may prove to be extremely effective for combating spam.

Categories and Subject Descriptors

C.2.0 [**Computer Communication Networks**]: Security and protection; C.2.3 [**Computer Communication Networks**]: Network operations – network management

General Terms

Design, Management, Reliability, Security

Keywords

spam, botnet, BGP, network management, security

1. Introduction

This paper presents a study of the network-level characteristics of unsolicited commercial email ("spam"). Much attention has been devoted to studying the content of spam, but comparatively little attention has been paid to spam's *network-level properties*. Conventional wisdom often asserts that most of today's spam comes from botnets, and that a large fraction of spam comes from Asia; a few studies have attempted to quantify some of these characteristics [5]. Unfortunately, little is known about how much spam comes from botnets versus other techniques (*e.g.*, short-lived route announcements, open relays, etc.), the geographic and topological distribution of where most spam originates (in terms of Internet Service Providers, countries, and IP address space), the extent to which different spammers use the same network resources, the stationarity of these properties over time, and so forth. A primary goal of this paper is to shed some light on these relatively unstudied questions.

Beyond merely exposing spammers' behavior, gathering information about the network-level behavior of spam could be a major asset for designing spam filters that are based on spammers' network-level behavior (presuming that the network-level characteristics of spam are sufficiently different than those of legitimate mail, a question we explore further in Section 4). Whereas spammers have the flexibility to alter the content of emails—both per-recipient and over time as users update spam filters—they have far less flexibility when it comes to altering the network-level properties of the spam they send. It is far easier for a spammer to alter the content of email messages to evade spam filters than it is for that spammer to change the ISP, IP address space, or botnet from which spam is sent.

Towards the goal of developing techniques that will help in the design of more robust network-level spam filters, this paper characterizes the network-level behavior of spammers as observed at a large spam sinkhole domain, which stores complete logs of all spam received from August 2004 through December 2005. We perform a *joint analysis* of the data collected at this sinkhole with an archive of BGP route advertisements as heard from the receiving network, traces from the "command and control" of a Bobax botnet, and traces of legitimate email from the mail server logs of a large email service provider. Although many aspects of mail headers can be forged, we base our analysis strictly on properties of the sender that are difficult to forge (*e.g.*, IP addresses that made connections to our mail servers, passive TCP fingerprints, corresponding route announcements, etc.).

We draw the following surprising conclusions from our study:

- *The vast majority of received spam arrives from a few concentrated portions of IP address space (Section 4).* Spam filtering techniques currently make no assumptions about the distribution of spam across IP address space. In a related area, many worm propagation models assume a uniform distribution of vulnerable hosts across IP address space (*e.g.*, [29]). In contrast, we find that the vast majority of spamming hosts—and, perhaps not coincidentally, most Bobax-infected hosts—lie within a small number of IP address space regions. Unfortunately, with a few exceptions (*e.g.*, 60.* – 70.*), most legitimate email comes from the same regions of IP address space, which suggests that, in general, effective filtering based on network-level properties may require determining second-order characteristics (*e.g.*, botnet membership).

SIGCOMM'06, September 11-16, 2006, Pisa, Italy.

- *Most received spam is sent from Windows hosts, each of which sends a relatively small volume of spam to our domain (Section 5).* Most bots send a relatively small volume of spam to our sinkhole (*i.e.*, less than 100 pieces of spam over 17 months), and about three-quarters of them are only active for a single time period of less than two minutes (65% of them send all spam in a "single shot").
- *A small set of spammers continually use short-lived route announcements to remain untraceable (Section 6).* A small portion of spam is sent by sophisticated spammers, who briefly advertise IP prefixes, establish a connection to the victim's mail relay, and withdraw the route to that IP address space after spam is sent. Anecdotal evidence has suggested that spammers might be exploiting the routing infrastructure to remain untraceable [1, 30]; this paper quantifies and documents this activity for the first time. To our surprise, we discovered a new class of attack, where spammers attempt to evade detection by hijacking *large* IP address blocks (*e.g.*, /8s) and sending spam from widely dispersed "dark" (*i.e.*, unused or unallocated) IP addresses within this space.

Beyond these findings, this paper's joint analysis of several datasets provides a unique window into the network-level characteristics of spam. To our knowledge, this paper presents the first study that examines the interplay between spam, botnets, and the Internet routing infrastructure.

We acknowledge that our spam corpus represents only a single vantage point, and, as such, drawing general conclusions about Internet-wide spam is not possible. Our goal is not to present conclusive figures about Internet-wide characteristics of spam. Indeed, the data we have collected is a small, localized sample of all spam traffic, and our statistics may not be reflective of Internet-wide characteristics. However, the spam we have collected represents an interesting dataset as it reflects the *complete set of spam emails received by a single Internet domain*. This dataset exposes spamming as a typical network operator for some Internet domain might also witness it. This unique view can help us better understand whether the features of spam that any single network operator observes could be useful in developing more effective filtering techniques.

With these goals in mind and an understanding of the context of our data, we offer the following additional observations on the implications of our results for the design of more effective techniques for spam mitigation, which we revisit in more detail in Section 7. First, the ability to trace the identities of spammers hinges on securing the routing infrastructure. Second, the distribution of spam and botnet activity across IP space suggests that, for some IP address ranges and networks, spam filters might monitor *network-wide* spam arrival patterns and attribute higher levels of suspicion to spam originating from networks with higher spam activity. Given the highly variable nature of the content of spam messages, incorporating general network-level properties of spam into filters may ultimately provide significant gains over more traditional methods (*e.g.*, content-based filtering), both through increased robustness and the ability to stop spam closer to its source.

The rest of this paper is organized as follows. Section 2 provides background on spamming and an overview of previous related work. In Section 3, we describe our data collection techniques and the datasets we used in our analysis. In Section 4, we study the distribution of spammers, spamming botnets, and legitimate mail senders across IP address space. Section 5 presents our findings regarding the relationship between the spam received at our sinkholes and known spamming bots. Section 6 examines the extent to which spammers use IP addresses that are generally unreachable (*e.g.*, using short-lived BGP route announcements) to send spam untraceably. Based on our findings, Section 7 offers positive recommendations for designing more effective mitigation techniques. We conclude in Section 8.

2. Background and Related Work

This section provides an overview of techniques both for sending and for mitigating spam and discusses related work in these areas.

2.1 Spam: Methods and Mitigation

In this section, we offer background on the main techniques used by spammers to send email, as well as some of the more commonly used mitigation techniques.

2.1.1 Spamming methods

Spammers use various techniques to send large volumes of mail while attempting to remain untraceable. We describe several of these techniques, beginning with "conventional" methods and progressing to more intricate techniques.

Direct spamming. Spammers may purchase upstream connectivity from "spam-friendly ISPs", which turn a blind eye to the activity. Occasionally, spammers buy connectivity and send spam from ISPs that do not condone this activity and are forced to change ISPs. Ordinarily, changing from one ISP to another would require a spammer to renumber the IP addresses of their mail relays. To remain untraceable and avoid renumbering headaches, spammers sometimes obtain a pool of dispensable dialup IP addresses, send outgoing traffic from a high-bandwidth connection the IP address spoofed to appear as if it came from the dialup connection, and proxy the reverse traffic through the dialup connection back to the spamming hosts [25].

Open relays and proxies. Open relays are mail servers that allow unauthenticated Internet hosts to connect and relay email through them. Originally intended for user convenience (*e.g.*, to let users send mail from a particular relay while they are traveling or otherwise in a different network), open relays have been exploited by spammers due to the anonymity and amplification offered by the extra level of indirection. It appears that the widespread deployment and use of blacklisting techniques have all but extinguished the use of open relays and proxies to send spam [21, 26].

Botnets. Conventional wisdom suggests that the majority of spam on the Internet today is sent by botnets—collections of machines acting under one centralized controller [3, 4, 31]. The `W32/Bobax` ("Bobax") worm (of which there are many variants) exploits the DCOM and LSASS vulnerabilities on Windows systems [18], allows infected hosts to be used as a mail relay, and attempts to spread itself to other machines affected by the above vulnerabilities, as well as over email. This paper studies the network-level properties of spam sent by Bobax drones. Agobot and SDBot are two other bots purported to send spam [12].

BGP spectrum agility. This study has discovered a new type of cloaking mechanism—BGP "spectrum agility"—whereby spammers briefly announce (often hijacked) IP address space from which they send spam and the routes to that IP address space once the spam has been sent. Although we have observed this behavior informally several years ago [6] and subsequent anecdotal evidence has suggested that spammers may use this technique [1], our study thoroughly documents this activity, and further finds that spammers may be using spectrum agility to complement spamming by other methods.

2.1.2 Mitigation techniques

Techniques for mitigating spam are as varied as techniques to send spam, and most existing techniques have significant draw-

backs. One of the most widely used anti-spam techniques is *filtering*, which typically classifies email based on its *content*; content-based filtering uses features of the contents of an email's headers or body to determine whether it is likely to be spam. Content-based filters, such as those incorporated by popular spam filters like SpamAssassin [27], successfully reduce the amount of spam that actually reaches a user's inbox. On the other hand, content-based filtering has drawbacks. Users and system administrators must continually update their filtering rules and use large corpuses of spam for training; in response, spammers devise new ways of altering the contents of an email to circumvent these filters. The cost of evading content-based filters for spammers is negligible, since spammers can easily alter content to attempt to evade these filters.

In addition to performing content-based checks, many mail filters, including SpamAssassin, also perform lookups to determine whether the sending IP address is in a "blacklist". Blacklists of known spammers, open relays and open proxies remain one of today's predominant spam filtering techniques. There are more than 30 widely used blacklists in use today; each of these lists is separately maintained, and insertion into these lists is based on many different types of observations (*e.g.*, operating an open relay, sending mail to a spam trap, etc.). The results in this paper—in particular, that IP address space is often "stolen" to send spam and that many bot IP addresses are short-lived—indicate that this long-standing method for filtering spam could become much less effective as spammers adopt these more sophisticated techniques.

2.2 Related Work

In this section, we first review previous work that has studied various spamming and spam-mitigation techniques, as well as the behavior of various worms and botnets. We then briefly discuss previous studies of unorthodox routing announcements. Previous work has studied each of these phenomena to some degree in isolation, but this study is the first to perform a joint analysis of spamming behavior, botnet characteristics, and Internet routing to better understand the characteristics and network-level behavior of spammers.

2.2.1 Spam and botnets

Previous studies have investigated the behavior and properties of worms, botnets, and other spam sources. Casado *et al.* used passive measurements of packet traces captured from about 2,500 spam sources to estimate the bottleneck bandwidths of roughly 25,000 TCP flows from spam sources and found peaks at common bandwidths (*e.g.*, modem speeds) [2]. Kumar *et al.* deconstructed the source code of the "Witty" worm to estimate various properties about Internet hosts (*e.g.*, host uptime) as well as about the propagation of the worm itself (*e.g.*, who infected whom) [14]. In contrast, our work explores the behavior of spammers in depth, although we also peripherally study malware whose exclusive purpose is to send spam (*i.e.*, the "Bobax" drone).

Several previous and ongoing projects are studying spammers' attempts to harvest email addresses for the purposes of spamming. For instance, Project Honeypot sinks email traffic for unused MX records and hands out "trap" email addresses to investigate harvesting behavior and to help identify spammers [23]. A previous study has used the data from Project Honeypot to analyze the methods employed by spammers; monitor the time it takes from when an email address is harvested to the time when that address first receives spam; the countries where most harvesting infrastructure is located; and the persistence (across time) of various harvesters [22]. We present preliminary results from a similar study in a technical report version of this paper [24].

In Section 5, we correlate spam arrivals with traces of hosts known to be infected with malware. Moore *et al.* found that the majority of hosts—and more than 80% of the hosts in Asia—did not patch the relevant vulnerability until well after actual outbreak [19], which makes it more reasonable to assume that IP addresses of Bobax drones remain infected for the duration of our spam trace.

2.2.2 Mitigation

A recent presentation from the SpamAssassin project discusses several techniques that the SpamAssassin spam filtering tool has incorporated to detect forged `X-Mailer` headers, weak "hash-busting" schemes, etc. [17]. Although their work also involves reverse engineering, the project focuses on analyzing mail *contents* to reverse-engineer spamming tools and techniques (with the goal of using this analysis to incorporate better content-filtering rules into SpamAssassin). Though our paper also studies such properties of spam, our analysis hinges on network-level properties—for instance, the IP address of the last remote mail relay (which previous work has also observed as one of the few parts of the SMTP header that cannot be forged [10])—rather than the artifacts of spamming software that appear in email content.

Jung *et al.* performed a study of DNS blacklist (DNSBL) traffic and the effectiveness of blacklists [13] and observed that 80% of the IP addresses that were sending spam were listed in DNSBLs two months after the collection of the traffic trace. Our study also measures the effectiveness of DNSBLs albeit in real time—we examine whether a host IP is listed in a set of DNSBLs *at the time the host spammed our domain*. While we also find that about 80% of the received spam was listed in at least one of eight blacklists, hosts that employ spamming techniques such as BGP spectrum agility tend to be listed in far fewer blacklists. We also find that even the most aggressive blacklist has a false negative rate of about 50%.

2.2.3 Unorthodox route announcements

Feamster *et al.* studied route advertisements for "bogon" IP address space (*i.e.*, private address space or unassigned addresses) [8]. However, since bogus or reserved address ranges are well-known, transit ISPs often filter them, resulting in little or no spam from such ranges. Cursory studies have suggested that spammers advertise routes to hijacked IP prefixes for short amounts of time to send spam [6, 28, 30]. In Section 6, we quantify the extent to which the sending of spam coincides with short-lived BGP route announcements for IP prefixes containing the mail relays that send spam.

3. Data Collection

This section describes the datasets that we use in our analysis. Our primary dataset consists of the actual spam email messages collected at a large spam sinkhole. To study the specific characteristics of certain subsets of spammers, we augment this dataset with three other data sources. First, to compare the network-level characteristics of spam received at our sinkhole with similar characteristics of *legitimate* email traffic, we obtain a corpus of email logs from a large email provider who automatically rejects email likely to be spam (thus allowing us to distinguish legitimate mail from spam). Second, we intercept the "command and control" traffic from a Bobax botnet at a sinkhole to identify IP addresses that were infected with the Bobax worm (and, hence, are likely members of botnets that are used for the sole purpose of sending spam). Third, we collect BGP routing data at the upstream border router *of the same network where we are receiving spam* and monitor the routing activity for the IP prefixes corresponding to the IP addresses from which spam was sent.

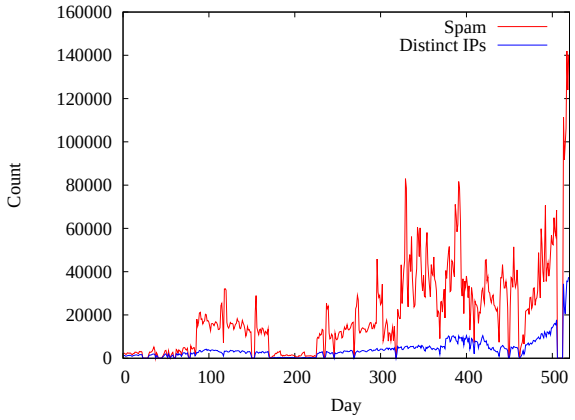

Figure 1: The amount of spam received per day at our sinkhole from August 2004 through December 2005.

3.1 Spam Email Traces

To obtain a sample of spam, we registered a domain with *no legitimate email addresses* and established a DNS Mail Exchange (MX) record for it. Hence, all mail received by this server is spam. The "sinkhole" has been capturing spam since August 5, 2004. Figure 1 shows the amount of spam that this sinkhole received per day through January 6, 2006 (the period of time over which we conduct our analysis). Although the total amount of spam received on any given day is rather erratic, the data indicates two unsettling trends. First, the amount of spam that the sinkhole is receiving generally appears to be increasing. Second, and perhaps more troubling, the number of distinct IP addresses from which we see spam on any given day also appears to be on the rise.

In addition to simply collecting spam traces, the sinkhole runs Mail Avenger [16], a customizable Simple Mail Transfer Protocol (SMTP) server that allows us to take specific actions upon receiving email from a mail relay (*e.g.*, running traceroute to the mail relay sending the mail, performing DNSBL lookups for the relay's IP address, performing a passive TCP fingerprint of the relay). We have configured Mail Avenger to (1) accept all mail, regardless of the username for which the mail was destined and (2) gather network-level properties about the mail relay from which spam is received. In particular, the mail server collects the following information about the mail relay *when the spam is received*:

- the IP address of the relay that established the SMTP connection to the sinkhole
- a traceroute to that IP address, to help us estimate the network location of the mail relay
- a passive "`p0f`" TCP fingerprint, based on properties of the TCP stack, to allow us to determine the operating system of the mail relay
- the result of DNS blacklist (DNSBL) lookups for that mail relay at eight different DNSBLs.

Note that, unlike many features of the SMTP header, these features are not easily forged.

3.2 Legitimate Email Traces

One of the motivations for our study was to determine whether the network-level characteristics of spam differ markedly from those of legitimate email. To perform this comparison, we obtained a corpus of mail logs from a large email provider that runs a Postfix mail server. Because this provider manages millions of mailboxes, it performs extensive spam filtering at its incoming SMTP servers. Accordingly, the logs for this mail server record, for each SMTP connection attempt, the time at which the connection attempt was made, the IP address of the connecting host, whether the mail was accepted or rejected, and, if the email was rejected, the reason for rejection. Using these logs, we can estimate the network-level properties of email that this domain deems to be legitimate. We performed our analysis over approximately 700,000 pieces of legitimate mail, as received at this provider's mail server on June 13, 2006. Although the corpus of legitimate mail is from a different domain than our sinkhole, both the spam sinkhole and the domain for legitimate email constitute large, domain-wide data sources for spam and legitimate mail, respectively, and are representative samples of spam and legitimate email that could be expected at any Internet domain.

3.3 Botnet Command and Control Data

To identify a set of hosts that are sending email from botnets, we used a trace of hosts infected by the `W32/Bobax` ("Bobax") worm from April 28-29, 2005. This trace was captured by hijacking the authoritative DNS server for the domain running the command and control of the botnet and redirecting it to a machine at a large campus network. This method was only possible because (1) the Bobax drones contacted a centralized controller using a domain name, and (2) the researchers who obtained the trace were able to obtain the trust of the network operators hosting the authoritative DNS for that domain name. This technique directs control of the botnet to the honeypot, which effectively disables it for spamming for this period. On the upside, because all Bobax drones now attempt to contact our command-and-control sinkhole rather than the intended command-and-control host, we can collect a packet trace to determine the members of the botnet.

To obtain a sample of spamming behavior from known botnets, we correlate Bobax botnet membership from the 1.5-day trace of Bobax drones with the IP addresses from which we receive spam in the sinkhole trace. This technique, of course, is not perfect: over the course of our spam trace, hosts may be patched. Although we cannot precisely determine the extent to which the transience of bots affects our analysis, previous work suggests that, even for highly publicized worms, the rate at which vulnerable hosts are patched is slow enough to expect that many of these infected hosts remain unpatched [19]. We also acknowledge another shortcoming of our approach: if hosts use dynamic addressing, different hosts (some of which may be Bobax-infected and some of which may not be) may use one of the IP addresses observed in the Bobax trace. However, we believe that the resulting inaccuracies are small: We observe a significantly higher percentage of Windows hosts in the subset of spam messages sent by IP addresses in our Bobax trace than in the complete spam dataset, which indirectly suggests that the hosts with IP addresses from the Bobax trace were indeed part of a spamming botnet when they spammed our sinkhole.

3.4 BGP Routing Measurements

In this paper, we study whether an IP address of the mail relay from which we receive spam is *reachable* and how long it remains reachable. We are particularly interested in cases where a route for an IP address is reachable for only a short period of time, coinciding with time at which spam was sent. To measure network-layer reachability from the network where spam was received, we co-located a "BGP monitor" in the same network as our spam sinkhole, similar to that in our previous work [7]. The monitor receives BGP updates from the border router, and our analysis includes a BGP update stream that overlaps with our spam trace. Since the moni-

tor has an internal BGP session to the network's border router, it will see only those BGP updates that cause a change in the border router's choice of *best* route to a prefix. Despite not observing all BGP updates, the monitor receives enough information to allow us to study the properties of *short-lived BGP route announcements*: the monitor will have *no* route to the prefix at all if the prefix is unreachable.

4. Network-level Characteristics of Spammers

In this section, we study some first-order network-level characteristics of spam sources. We survey the portions of IP address space from which our sinkhole received spam and the ASes that sent spam to the sinkhole. We also observe the persistence of these characteristics over time. To determine whether these network level characteristics could be suitable for filtering spam, we compare the network-level characteristics of spam to the same characteristics for *legitimate* email, as received at a large domain that manages approximately 40 million mailboxes.

We find that the distribution of spam across IP address space is (1) nearly identical to the legitimate mail distributions (with a few exceptions), and (2) quite persistent over time. Still, the distribution of spam senders across IP address space is far from uniform, and spam arrival by *IP address range* is much more pronounced, persistent, and concentrated than similar characteristics by IP address. Additionally, we find that a large fraction of spam is received from just a handful of ASes: nearly 12% of all received spam originates from mail relays in just two ASes (from Korea and China, respectively), and the top 20 ASes are responsible for sending nearly 37% of all spam. This distribution (as well as the main perpetrators) is also persistent over time. This heavily skewed distribution suggests that spam filtering efforts might better focus on identifying high-volume, persistent groups of spammers (*e.g.*, by AS number), rather than on blacklisting individual IP addresses, many of which are transient.

4.1 Distribution Across Networks

To determine the address space from which spam was arriving ("prevalence") and whether the distribution across IP addresses changes over time ("persistence"), we tabulated the spam in our trace by IP address space. We find that spam arrivals across IP space are far from uniform.

Finding 4.1 (Distribution across IP address space) *The majority of spam is sent from a relatively small fraction of IP address space.*

Figure 2 shows the number of spam email messages received over the course of the entire trace, as a function of IP address space. Several ranges of IP address space originate large amount of email traffic (both spam and legitimate), including space allocated to cable modem providers (*e.g.*, 24.*) and the address space allocated to the Asia Pacific Network Information Center (APNIC) regional Internet registry (*e.g.*, 61.*). Although most IP address ranges that originate a significant amount of spam also originate a lot of legitimate mail traffic, a few IP address ranges have significantly more spam than legitimate mail (*e.g.*, 80.*–90.*), and vice versa (*e.g.*, 60.*–70.*). This characteristic suggests that it may be possible to use IP address ranges to distinguish spam from legitimate email.

We repeated the analysis of the network-level characteristics of spam per day across months, per month across years, and so forth. We also compared the distribution of spam collected at our sinkhole to the distribution of *rejected* SMTP connections at the domain where we performed our analysis of legitimate email and found

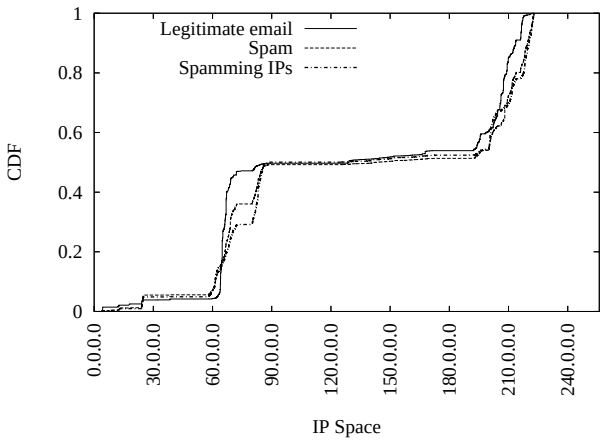

Figure 2: Fraction of spam email messages and comparison with legitimate email received (as a function of IP address space); also, fraction of client IP addresses that sent spam, binned by /24.

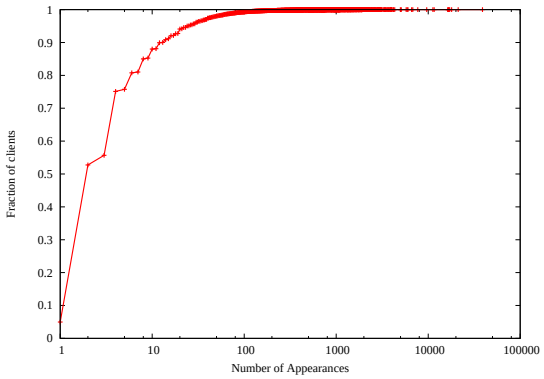

Figure 3: The number of distinct times that each client IP sent mail to our sinkhole (regardless of the number emails sent in each batch).

that the distribution of these connections across IP address space is similar to that shown in Figure 2. All of these distributions have remained roughly constant over time (*i.e.*, the results look similar to those shown in Figure 2). In contrast, individual IP addresses are far more transient. Figure 3 shows that even though a few IP addresses sent more than 10,000 emails, about 85% of client IP addresses sent less than 10 emails to the sinkhole, indicating that targeting an individual IP address might not help mitigate spam without sharing information across domains. This finding has an important implication for spam filter design: Though the individual IP addresses from which spam is received changes from day-to-day, the fact that spam continually comes from the same IP address *space* suggests that incorporating these more persistent features may be more effective, particularly in portions of the IP address space that send either mostly spam or mostly legitimate email.

In many cases, IP address ranges are not adequate for distinguishing spam from legitimate email. To determine whether other network-level properties, such as the AS from which the email was sent, could serve as better classifiers, we examined the distribution of spam across ASes and compared this feature to the distribution of legitimate email across ASes.

Finding 4.2 (Distribution across ASes) *More than 10% of spam received at our sinkhole originated from mail relays in two ASes,*

AS Number	*# Spam*	*AS Name*	*Primary Country*
766	580559	Korean Internet Exchange	Korea
4134	560765	China Telecom	China
1239	437660	Sprint	United States
4837	236434	China Network Communications	China
9318	225830	Hanaro Telecom	Japan
32311	198185	JKS Media, LLC	United States
5617	181270	Polish Telecom	Poland
6478	152671	AT&T WorldNet Services	United States
19262	142237	Verizon Global Networks	United States
8075	107056	Microsoft	United States
7132	99585	SBC Internet Services	United States
6517	94600	Yipes Communications, Inc.	United States
31797	89698	GalaxyVisions	United States
12322	87340	PROXAD AS for Proxad ISP	France
3356	87042	Level 3 Communications, LLC	United States
22909	86150	Comcast Cable Corporation	United States
8151	81721	UniNet S.A. de C.V.	Mexico
3320	79987	Deutsche Telekom AG	Germany
7018	74320	AT&T WorldNet Services	United States
4814	74266	China Telecom	China

Table 1: Amount of spam received from mail relays in the top 20 ASes. 11 of the top 20 networks from which we received spam are primarily based in the United States.

AS Number	*# Email*	*AS Name*	*Primary Country*
15169	49500	Google Inc.	United States
5731	38238	AT&T WorldNet Services	United States
26101	30406	Yahoo	United States
3561	22730	Savvis	United States
4355	17381	Earthlink, Inc	United States
8560	16666	Schlund Partner AG	Germany
8075	14699	Microsoft Corp	United States
14779	13115	Inktomi Corporation	United States
6541	12493	GTE.net LLC	United States
14780	11597	Inktomi Corporation	United States

Table 2: Top 10 ASes (by email volume) in our legitimate email trace.

and 36% of all received spam originated from only 20 ASes. With a few exceptions, the ASes containing hosts responsible for sending large quantities of spam differ from those sending large quantities of legitimate email.

The concentration of spammers in a small collection of offending ASes—and the fact that this collection of ASes differs from the ASes responsible for sending legitimate email (with the exception of ASes 5731 and 8075)—suggests that spam filters should attribute more suspicion to email coming from ASes where spam commonly originates. This observation begs the question about why Figure 2 does not show similar differences. Indeed, the spamming behavior of specific IP address ranges deserves further study, since Figure 2 really only exposes macro-level behavior of IP address ranges (*i.e.*, differences for small IP address ranges may not be visible in the figure). We are studying the behavior of fine-grained address ranges in ongoing work.

Recent reports have claimed that most spam originates in the United States [5]. On the other hand, Figure 2 suggests that many spamming hosts reside in IP address space that is allocated to the Asia-Pacific region (*e.g.*, 61.0.0.0/8). To perform a rough estimate of the amount of spam originating from each country, we associated the ASes from which we received spam to the countries where those ASes were based.[1] Table 1 also shows the distribution of hosts that

[1] Although some ASes span multiple countries, typically even large transit providers have different AS numbers for backbone networks in different countries. In any case, we use the *primary* country where the AS is based.

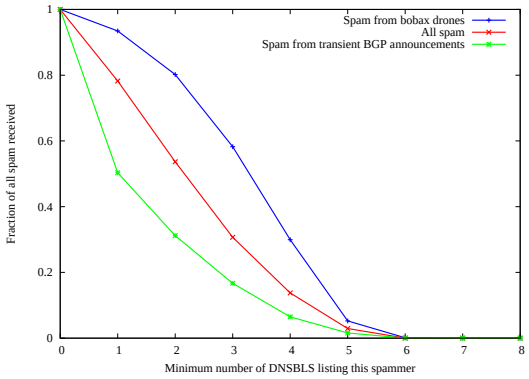

Figure 4: The fraction of spam emails that were listed in a certain number of blacklists or more, *at the time each mail was received*.

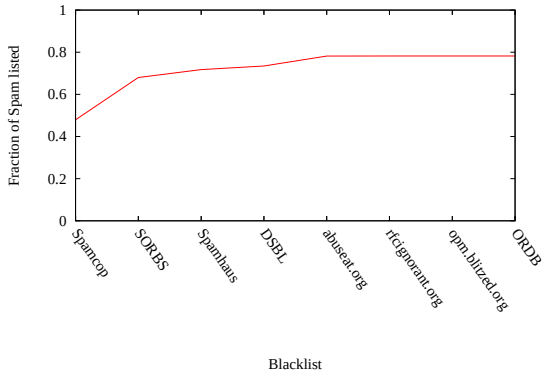

Figure 5: The cumulative fraction of spam emails that were listed in each blacklist *at the time each mail was received*, sorted from most aggressive to least aggressive blacklist.

sent spam to the sinkhole by country, for the top 20 ASes from which we received spam.

Finding 4.3 (Distribution by country) *Although the top two ASes from which we received spam were from Asia, 11 of the top 20 ASes from which we received spam were from the United States and comprised over 40% of all spam from the top 20 ASes.*

We mapped the most prolific IP address (*i.e.*, the top 11.6% of IP addresses, responsible for 65% of all spam received at the sinkhole) to their respective countries. Our analysis indicates that nearly three times as much spam in our trace originates from ISPs based in the United States than from either of the next two most prolific countries (Korea and China, respectively). This conclusion does differ from other reports, which also indicate that most spam comes from the U.S., but to a much lesser degree. The distribution of spam by country, when compared to the statistics for legitimate email (Table 2), also suggests that, in some cases, assigning a higher level of suspicion according to an email's *country* of origin may be an effective filtering technique for some networks.

4.2 The Effectiveness of Blacklists

Given the transience of each IP address sending spam to our sinkhole (*i.e.*, the results shown in Figure 3), we suspected that filtering based on IP address, a method commonly employed by DNSBLs, would be relatively ineffective. To test this hypothesis, we used the

results from real-time DNSBL lookups performed by Mail Avenger to 8 different blacklists *at the time the mail was received* .

Figure 4 indicates that IP-based blacklisting is still working reasonably well *if many blacklists are consulted simultaneously*: Although 20% of spam came from IP addresses that were not listed in *any* blacklist, (as shown by the middle line "All spam", where about 80% spam was listed in at least one blacklist), more than 50% of such spam was listed in two or more blacklists, and 80% was listed in two or more blacklists.

More troubling, however, is that the spam that we received from spammers using "BGP spectrum agility" techniques (as described in Section 2) are not blacklisted nearly as much: half of these IP addresses do not appear in *any* blacklist, and only about 30% of these IP addresses appear in more than one blacklist.

Finding 4.4 (Effectiveness of blacklists) *Nearly 80% of all spam was received from mail relays that appear in at least one of eight blacklists. A relatively higher fraction of Bobax drones were blacklisted, but relatively fewer IP addresses sending spam from short-lived BGP routes were blacklisted—only half of these mail relays appeared in any blacklist.*

Although this finding appears to suggest that DNSBLs are effective at identifying most types of spam based on IP address, the reality is actually not as bright as it appears. First, this result is based on an aggressive approach that sends queries to *eight* blacklists; Figure 5 shows the cumulative fraction of spam listed in each blacklist, from most aggressive DNSBL to least aggressive and shows that even the most aggressive blacklist, Spamcop, only lists about half of all spam received. Second, many of the more aggressive blacklists are known to have a significant number of false positives. Finally, even aggressive mechanisms, such as querying eight different blacklists, are fairly ineffective at identifying IP addresses using more sophisticated cloaking techniques (*e.g.*, the BGP spectrum agility technique, which we discuss in more detail in Section 6).

5. Spam from Botnets

In this section, we amass circumstantial evidence that suggests that a majority of spam originates from bots. Although, given our limited datasets, we cannot determine a precise fraction of the total amount of spam that is coming from bots, we use our trace of "Bobax" command and control data to study the patterns of spam that are being sent from hosts that are known to be bots. First, we study the activity profile of drones from the "Bobax" botnet and find that the IP address space where we observe worm activity bears close similarity to the IP address space where we observed spamming activity (Finding 4.1). Second, we observe that about 70% of all remote hosts spamming our sinkhole—and 95% of hosts for which we could attribute some operating system—appear to be running Windows; additionally, these hosts each send relatively low volumes of spam to the sinkhole, regardless of their persistence.

5.1 Bobax Topology

We studied the prevalence of spamming hosts versus the prevalence of known Bobax drones to better understand how the distribution of IP addresses of Bobax-infected hosts compared to the IP distribution of spammers in general. Figure 6 shows the results of this analysis; the distribution of *all* Bobax-infected hosts is quite similar to that of the distribution of all spammers (Figure 2).

Finding 5.1 (Bobax vs. spammer distribution) *Spamming hosts and Bobax drones have similar distributions across IP address space, which indirectly suggests that much of the spam received at the sinkhole may be due to botnets such as Bobax.*

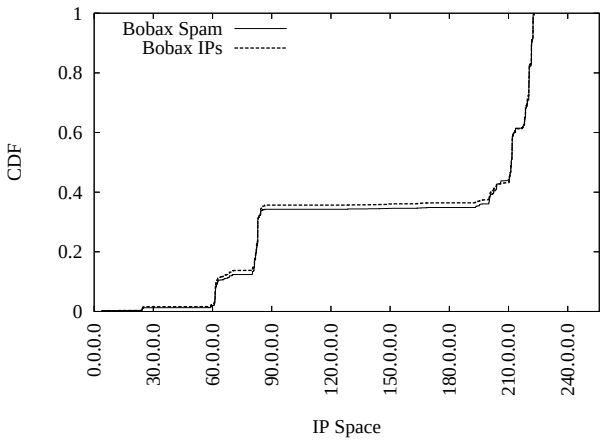

Figure 6: The number of *all* Bobax drones, and the amount of spam received from those drones at the sinkhole, as a function of IP address space. On the x-axis, IP address space is binned by /24.

This similarity provides evidence of correlation, not causality, but the fact that the distribution of IP addresses from which spam is received more closely resembles botnet activity than the spread of IP addresses of legitimate email suggests that a significant amount of spam activity may be due to botnet activity.

Although the range 60.* – 67.* has a significant fraction of spamming IP addresses (Figure 2), we see relatively less spam from Bobax drones from this space, which led us to suspect that spammers may be using techniques other than botnets for sending spam from many of the hosts in this range. Indeed, in Section 6, we present findings that suggest that one or more sophisticated groups of spammers appear to be sending spam from a large number of machines (or, perhaps, a smaller number of machines with changing IP addresses), numbered from portions of unused IP space within this range that are unroutable except for when they are sending spam.

5.2 Operating Systems of Spamming Hosts

In this section, we investigate the prevalence of each operating system among the spam we received, as well as the total amount of spam we received from hosts of each type. For this purpose, we used the passive OS fingerprinting tool, `p0f`, which is incorporated into Mail Avenger; thus, we can attribute an operating system to each remote host that sends us spam. Using this technique, we were able to identify the operating system for about 75% of all hosts from which we received spam. Table 3 shows the results of this study. Roughly 70% of the hosts from which we receive spam, and 95% of these hosts to which we could attribute an operating system, run Windows; this fraction is consistent with the fact that roughly 95% of all hosts on the Internet run Windows [20].

More striking is that, while only about 4% of the hosts from which we receive spam are from hosts are running operating systems other than Windows, this small set of hosts appears to be responsible for at least 8% of the spam we receive. The fraction, while not overwhelmingly large, is notable because of the conventional wisdom that most spam today originates from compromised Windows machines that are serving as botnet drones.

Finding 5.2 (Prevalence of spam relays by OS type) *About 4% of the hosts sending spam to the sinkhole are not Windows hosts but our sinkhole receives about 8% of all spam from these hosts.*

Operating System	*Clients*	*Total Spam*
Windows	854404 (70%)	5863112 (58%)
- Windows 2000 or XP	604252 (49%)	4060290 (40.2%)
- Windows 98	13727 (1.1%)	54856 (0.54%)
- Windows 95	559 (<0.1%)	2797 (<0.1%)
- Windows (other/unconfirmed)	235866 (19%)	1745169 (17.2%)
Linux	28132 (2.3%)	557377 (5.5%)
FreeBSD	6584 (0.5%)	152456 (1.5%)
MacOS	2944 (0.2%)	46151 (0.4%)
Solaris	1275 (< 0.1%)	18084 (0.2%)
OpenBSD	797 (< 0.1%)	21496 (0.2%)
Cisco IOS	736 (< 0.1%)	5949 (<0.1%)
NetBSD	44 (< 0.1%)	327 (<0.1%)
HP-UX	31 (< 0.1%)	120 (<0.1%)
Tru64	26 (< 0.1%)	143 (<0.1%)
AIX	23 (< 0.1%)	366 (<0.1%)
OpenVMS	18 (< 0.1%)	62 (<0.1%)
IRIX	7 (< 0.1%)	62 (<0.1%)
Other/Unidentified	128580 (10.4%)	1212722 (12%)
No Fingerprint	204802 (16.7%)	2225410 (22%)
Total	1228403	10103837

Table 3: The operating system of each unique sender of received spam, as determined by passive OS fingerprinting.

A significant fraction of the spamming infrastructure is apparently still Unix-based.[2]

5.3 Spamming Bot Activity Profile

The results in Section 5.2 indicate that an overwhelming fraction of spam is sent from Windows hosts. Because a very large fraction of spam comes from Windows hosts, our hypothesis is that many of these machines are infected hosts that are bots. In this section, we investigate the characteristics of spamming hosts that are known to be Bobax drones. Specifically, we seek to answer the following three questions:

1. **Intersection:** *How many of the known Bobax drones send spam to our sinkhole?*
2. **Persistence:** *For how long does any particular Bobax drone send spam?*[3]
3. **Volume:** *How much of the spam from Bobax drones originates from hosts that are only active for a short period of time?*

The rest of this section explores these three questions. Although our trace sees spam from only a small fraction of all Bobax-infected drones, this sample nevertheless can offer insight into the behavior of spamming bots.

5.3.1 Intersection and prevalence

To satisfy our curiosity (and to compare with other claims about the amount of spam coming from botnets [3]), we wanted to determine the total fraction of received spam that originated from botnets versus other mechanisms. The circumstantial evidence in Sections 5.1 and 5.2 suggests that the fraction of spam that originates from botnets is quite high. Unfortunately, there are no techniques for isolating botnets from mail logs alone; we can only determine whether a particular piece of spam originated from a botnet based on whether the IP address of the relay sending the spam appears in our trace of machines known to be infected with Bobax.

Even this information is not sufficient to answer questions about the amount of spam coming from botnets, since machines other than Bobax-infected hosts may be enlisted in spamming botnets. Indeed, good answers to this question depend on both additional vantage points (*i.e.*, sinkhole domains) and better botnet detection heuristics and algorithms. Not only will more vantage points and better detection algorithms aid analysis, but they may also prove useful for massively collaborative spam filtering—identification of botnet membership, for example, could prove a very effective feature for identifying spammers.

At our spam sinkhole, we receive spam from only 4,693 of the 117,268 Bobax-infected hosts in our command-and-control trace. This small (though certainly non-negligible) view into the Bobax botnet emphasizes the need for observing spamming behavior at multiple domains to observe more significant spamming patterns of a botnet. Nevertheless, this set of hosts that appear both in our spam logs and in the Bobax trace can provide useful insight into the spamming behavior and network-level properties of *individual* bots; it also appears to be a reasonable cross-section of all spamming bots (Figure 6 indicates that the IP distribution of bots from which our sinkhole receives spam is quite similar to the distribution of all spamming hosts across IP address space as shown in Figure 2).

5.3.2 Persistence

Figure 7 shows the persistence of each Bobax-infected IP address that sent spam to the sinkhole. The figure indicates that the majority of botnets make only a single appearance in our trace; these "single shot" bots account for roughly 25% of all spam that is known to be coming from Bobax drones.

Finding 5.3 (Single-shot bots) *More than 65% of IP addresses of hosts known to be infected with Bobax send spam only once, and nearly 75% of these addresses send spam to our sinkholed domain for less than two minutes, although many of them send several emails during their brief appearance.*

Of the spam received from Bobax-infected hosts, about 25% originated from hosts that only sent mail from IP addresses that only appeared once. The persistence of Bobax-infected hosts appears to be mildly bimodal: although roughly 75% of Bobax drones persist for less than two minutes, the remainder persist for a day or longer, about 50 persist for about six months, and 10 persist for entire length of the trace. Although these short-lived bots do not yet send the majority of spam coming from botnets, this "single shot" technique may become more prominent over time as network-level filtering techniques improve and spammers employ more sophisticated evasion techniques.

Because most bot IP addresses are short-lived, we hypothesized that IP-based blacklists (*e.g.*, DNSBL filtering) would be somewhat ineffective for blocking spam. To our surprise, Figure 4 shows that the botnet hosts from which we received spam were actually *more* likely to be listed than the typical spamming mail relay (although, as we describe in Section 4.2, the technique appears to be somewhat ineffective in general). Intuitively, this result is justifiable, because other domains likely received spam from drones with the same IP addresses. This result also demonstrates the benefits of collaborative spam filtering, which facilitates the identification of spammers

[2] Alternatively, this spam might be sent from Windows machines whose stacks have been modified to emulate those of other operating systems. Although we doubt that this is likely, since most spam filters today do not employ `p0f` checks, we acknowledge that it may become more common in the future, especially as spammers incorporate these techniques.

[3] Previous work has noted that the "DHCP effect" can create errors in estimation for both persistence and prevalence (*e.g.*, a single host could dynamically be assigned different IP addresses over time) [19]. Although the DHCP effect can introduce problems for estimating the total population of a group of spammers, it is not as problematic for the questions we study in this paper. Since one of our objectives is to study the effectiveness of IP-based filtering (rather than, say, count the total number of hosts), we are interested more in measuring the persistence of *IP addresses*, not hosts.

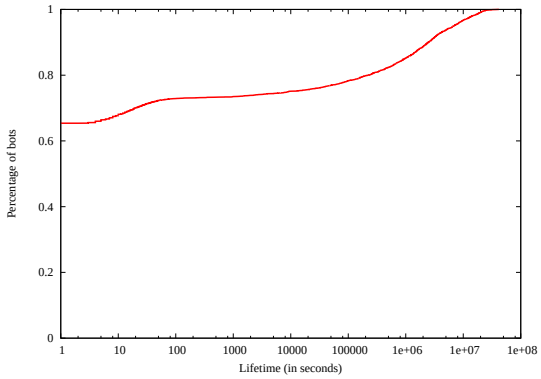

Figure 7: Bobax drone persistence.

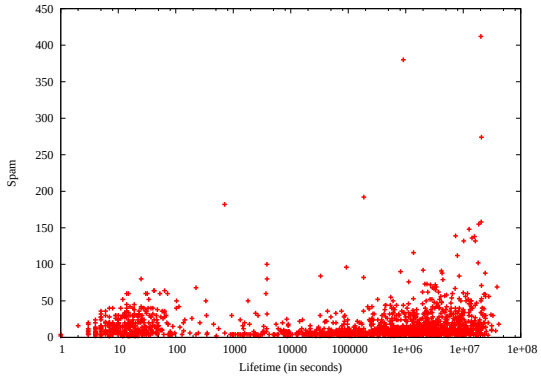

Figure 8: Number of spam email messages received vs. bobax drone persistence.

that send only a single piece of spam but send spam to multiple domains.

5.3.3 *Volume and Rate*

Figure 8 shows the amount of spam sent for each Bobax drone, plotted against the persistence of each drone. This graph shows that most Bobax drones do not send a large amount of spam, *regardless of how long the drone was active*. Indeed, nearly all of the Bobax drones observed in our trace send fewer than 100 pieces of spam over the entire period of the trace. This finding suggests that spammers have the ability to send spam from a large number of hosts, each of which is typically used for a short period of time and nearly always used to send only a relatively small amount of spam. Thus, not only are IP-based filtering schemes likely to be ineffective, but *volume-based* detection schemes for spamming botnets may also be ineffective.

Finding 5.4 (Spam arrives from bots at very low rates)
Regardless of persistence, 99% of bots sent fewer than 100 pieces of spam to our domain over the entire trace.

Most persistent bots sent fewer than 100 pieces of spam to our sinkhole, indicating that typical rates of spam from Bobax drones, *for spam received by a single domain*, are less than a single piece of spam per bot per day.

6. Spam from Transient BGP Announcements

Many spam filtering techniques leverage the ability to positively identify a spammer by its IP address. For example, DNS blacklists catalog the IP addresses of likely spammers so that spam filters may later send queries to determine whether an email was sent by a likely spammer. Of course, this technique implicitly assumes a connection between an IP address and the physical infrastructure that a spammer uses to distribute email. In this section, we study the extent to which spammers use such transient identities by examining spam received by the sinkhole domain that coincides with short-lived BGP route announcements.

Informal anecdotes have claimed that some spammers briefly advertise portions of IP address space, send spam from mail relays with IP addresses in that space, and subsequently withdraw the routes for that space after the relays have sent spam [1, 28, 30]. This practice makes it difficult for end users and system administrators to track spam sources because the network from which a piece of spam was sent is likely to be unreachable at the time a user lodges a complaint. Although it is technically possible to log BGP routing announcements and mine them to perform post-mortem analysis, the relative difficulty of doing so (especially since most network operators do not monitor interdomain routes in real time) essentially makes these spammers untraceable.

Little is known about (1) whether the technique is used much in practice (and how widespread it is), (2) what IP space spammers tend to use to mount these types of attacks and (3) the announcement patterns of these attacks. This study seeks to answer two sets of questions about the use of short-lived BGP routing announcements for sending spam:

- *Prevalence across ASes and persistence across time.* How many ASes use short-lived BGP routing announcements to send spam? Which ASes are the most guilty, in terms of number of pieces of spam sent, and in terms of persistence across time?
- *Length of short-lived BGP announcements.* How long do short-lived BGP announcements last (*i.e.*, long enough for an operator to catch)?

As we will see, sending spam from IP address space corresponding to short-lived route announcements is *not*, by any means, the dominant technique that spam is sent today (when this technique is actively being used, it accounts for no more than 10% of all spam we receive, and it generally accounts for much less). Nevertheless, because our domain only observes spamming behavior from a single vantage point, this technique may be more common than we are observing. Additionally, because this technique is not well defended against today, and because it is complementary to other spamming techniques (*e.g.*, it could conceivably be used to cloak botnets), we believe that this behavior is worth attention, particularly since some of the techniques we observe (*i.e.*, hijacking large prefixes) represents a significant departure from conventional wisdom on prefix hijacking.

6.1 BGP Spectrum Agility

Figure 9 shows an example of `61.0.0.0/8` being announced by AS 4678 for a brief period of time on September 30, 2005, during which spam was also sent from IP addresses contained within this prefix.

To investigate further the extent to which this technique is used in practice, we performed a joint analysis of BGP routing data (described in Section 3.4) and the spam received at our sinkhole, which is co-located with the BGP monitor. Given the sophistication required to send spam under the protection of short-lived routing announcements (especially compared with the relative simplicity of purchasing access to a botnet), we doubted that it was particularly prevalent. To our surprise, a small number of parties appear to be

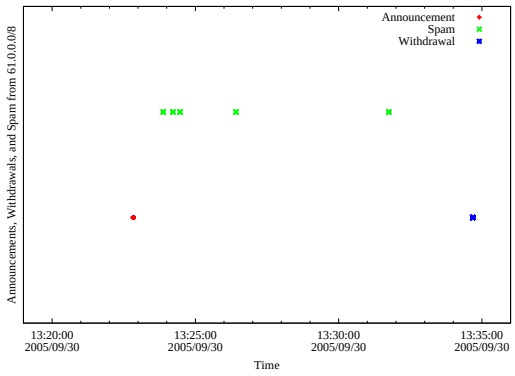

Figure 9: Observation of a short-lived BGP route announcement for `61.0.0.0/8`, spam arriving from mail relays in that prefix, and the subsequent withdrawal of that prefix.

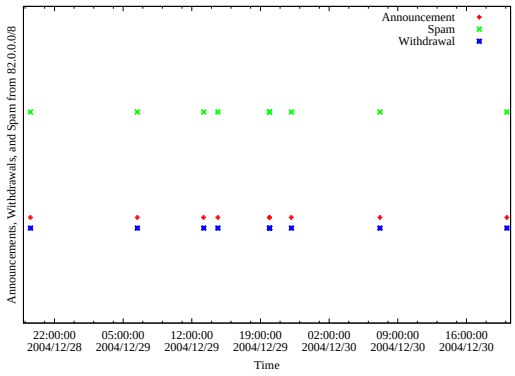

Figure 10: Observation of a short-lived BGP route announcement for `82.0.0.0/8`, spam arriving from mail relays in that prefix, and the subsequent withdrawal of that prefix.

using this technique to send spam quite regularly. In fact, looking in further detail at the several (prefix, AS) combinations, we observed the following remarkable patterns:

- AS 21562, an Internet service provider (ISP) in Indianapolis, Indiana (according to `ra.net` and `arin.net`), originated routing announcements for `66.0.0.0/8`.
- AS 8717, an ISP in Sofia, Bulgaria, originated announcements for `82.0.0.0/8`.
- In a third, less persistent case, AS 4678, an ISP in Japan, Canon Network Communications (according to `apnic.net`), originated routing announcements for `61.0.0.0/8`.

We were surprised that three of the most persistent prefixes involved in short-lived BGP routing announcements were so large. Although some short-lived routing announcements may be misconfigurations [15], the fact that these routing announcements continually appear, that they are for large address blocks, and that they typically coincide with spam arrivals (as shown in Figure 9) raised our suspicion about the veracity of these announcements. Indeed, not only are these route announcements short-lived and hijacked, but they are also for large address blocks. Although the use of large address blocks might initially seem surprising, the distribution of the IP addresses of hosts sending spam using this technique suggests the following theory.

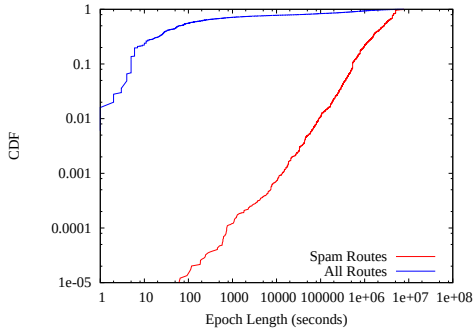

Figure 11: CDF of the length of each short-lived BGP episode, from September 2005–December 2005.

Finding 6.1 (Spectrum Agility) *A small, but persistent, group of spammers appear to send spam by (1) advertising (in fact, hijacking) large blocks of IP address space* (i.e., */8s*)*, (2) sending spam from IP addresses that are scattered throughout that space, and (3) withdrawing the route for the IP address space shortly after the spam is sent.*

We have called this technique "spectrum agility" because it allows a spammer the flexibility to use a wide variety of IP addresses within a very large block from which to send spam. The large IP address block allows the mail relays to "hop" between a large number of IP addresses, thereby evading IP-based filtering techniques like DNSBLs. Judging from Figure 4 and our analysis in Section 4.2, the technique seems to be rather effective. As an added benefit, route announcements for shorter IP prefixes (*i.e.*, larger blocks of IP addresses) are less likely to be blocked by ISPs' route filters than route announcements or hijacks for longer prefixes.

Upon further inspection, we also discovered the following interesting features: (1) the IP addresses of the mail relays sending this spam are widely distributed across the IP address space; (2) the IP addresses from which we see spam in this address space typically appear only once; (3) on February 6, 2006, attempts to contact the mail relays that we observed using this technique revealed that that roughly 60-80% of these hosts were not reachable by `traceroute`; (4) many of the IP addresses of these mail relays were located in allocated, albeit unannounced and unused IP address space; and (5) many of the AS paths for these announcements contained reserved (*i.e.*, to-date unallocated AS numbers), suggesting a possible attempt to further hamper traceability by forging elements of the AS path. We are at a loss to explain certain aspects of this behavior, such as why some of the machines appear to have IP addresses from allocated space, when it would be simpler to "step around" the allocated prefix blocks, but, needless to say, the spammers using this technique appear to be very sophisticated.

Whether spammers are increasingly using this technique is inconclusive. Still, many of the ASes that send the most spam with this technique also appear to be relative newcomers. Variants of this type of technique may be used in the future to make it more difficult to track and blacklist spamming hosts, particularly since the technique allows a spammer to relatively undetectably commandeer a very large number of IP addresses.

6.2 Prevalence of BGP Spectrum Agility

Because of the volume of data and the relatively high cost of performing longest-prefix match queries, we performed a more extensive analysis on a subset of our trace, from September 2005 till December 2005, to detect the fraction of spam coming from

short-lived announcements and to determine a reasonable threshold for studying short-lived announcements across the entire trace. Figure 11 shows that, for all of the IP addresses for which we received spam over the course of these four months, almost 99% of the corresponding BGP routing announcements were announced continuously for at least a day. In other words, most of the received spam corresponded to routing advertisements that were *not* short-lived. On the other hand, this technique appears to be used intermittently, and during time periods when this activity was more prevalent, as much as 10% of all received spam coincides with routing announcements that lasted less than a day.

Finding 6.2 (Prevalence: Spam from Short-Lived Routes) *Only about 1% of spam was received from route that persisted for less than a single day, although during intervals when this technique was used more commonly, as much as 10% of all spam coincided with routes that lasted less than a day.*

Unfortunately for traditional filtering techniques, the spammers who are the most persistent across time are, for the most part, *not* the spammers who send the most spam using this technique. Indeed, only two ASes—AS 4788 (Telekom Malaysia) and AS 4678 (Canon Network Communications, in Japan)—appear among both the top-10 most persistent and most voluminous spammers using short-lived BGP routing announcements.

6.3 How Much Spam from Spectrum Agility?

A comparatively small fraction of spam originates from IP addresses that correspond to short-lived BGP route announcements (*i.e.*, routing announcements that persist for less than a day) that coincide with spam arrival. The total amount of spam received as a result of this technique seems to pale in comparison to other techniques: no more than 10% of all spam—and more likely as little as 1%—appears to be sent using this technique. Although this technique is not apparent for most of the spam we receive (after all, a botnet makes traceability difficult enough), the few groups of spammers that employ this technique typically use it quite regularly. We also observed that many of the ASes using this technique for the longest period of time do *not*, in fact, rely on this technique for sending most of their spam. Even the most prolific spamming AS in this group, Malaysia Telekom, appears to send only about 15% of their spam in this fashion.

Finding 6.3 (Persistence vs. Volume) *The ASes from where spammers most continually use short-lived route announcements to send spam are not the same ASes from which the most spam originates via this technique.*

Many ASes that advertise short-lived BGP routing announcements and send large volumes of spam from these routes do not appear to be hijacking IP prefixes. In the case where spam volume is high, these short-lived routing announcements may simply coincide with spam being sent via another means (*e.g.*, from a botnet). The ASes that persistently advertise short prefixes, however, appear to be doing so intentionally.

7. Lessons for Better Spam Mitigation

Existing spam mitigation techniques have focused on either throttling senders (*e.g.*, recent attention has focused on cost-based schemes [9, 11]) or having receivers filter spam according to the *content* of a message. The results of this paper, however, highlight several important lessons that strongly indicate that devoting more attention to the network-level properties of spammers that may be a useful addition to today's spam mitigation techniques. Using network-level information to help mitigate spam not only provides a veritable font of new features for spam filters, but network-level properties have two important properties that could potentially lead to more robust filtering.

1. Network-level properties are *less malleable* than those based on an email's contents.
2. Network-level properties may be *observable in the middle of the network*, or closer to the source of the spam, which may allow spam to be quarantined or disposed of before it ever reaches a destination mail server.

From our findings, we derive the following lessons regarding the network-level behavior of spammers that could help in designing better mitigation techniques.

Lesson 1 *Spam filtering requires a better notion of host identity.*

We observed a significant amount of spam from "one-shot" bots and spammers using spectrum agility. Short-lived bots, short-lived BGP route hijacks, and dynamic addressing effects foil the common practice of using a host's IP address as its identity.

Lesson 2 *Detection techniques based on aggregate behavior are more likely to expose nefarious behavior than techniques based on observations of a single IP address.*

Although comprehensive IP-based blacklisting is somewhat effective, blacklisting techniques may also benefit by exploiting other network-level properties such as IP address *ranges*, some of which (*e.g.*, 80.*–90.*) send mostly spam.

Lesson 3 *Securing the Internet routing infrastructure is a necessary step for bolstering identity and traceability of email senders.*

Although BGP spectrum agility is by no means responsible for most received spam, several characteristics make the technique extremely troubling. Most notably, the technique can be combined with other spamming techniques (possibly even spamming with botnets) to give spammers more agility in evading IP-based blacklists. Indeed, our analysis of DNSBLs indicates that spammers may already be doing this. A routing infrastructure that instead provided protection against route hijacking (specifically, unauthorized announcement of IP address blocks) would make BGP spectrum agility attacks more difficult to mount.

Lesson 4 *Some network-level properties of spam can be incorporated relatively easily into spam filters and may be quite effective at detecting spam that is missed by other techniques.*

Although the BGP spectrum agility attack is particularly wily—and effective against DNSBLs—incorporating additional network-level features into spam filtering software such as "recently announced BGP announcement" should prove remarkably effective at quenching this attack.

Given the benefits that network-wide analysis could provide for stemming spam, we imagine that the ability to witness the network-level behavior of spammers *across* multiple distinct domains could also expose patterns that are not evident from a single domain. One organization might be able amass such a dataset either by sinkholing a large number of domains; for example, Project Honeypot [23] solicits donations of MX records for registered domains that do not receive email (though its corpus is still significantly smaller than

ours). As we have discovered thus far from our initial experiences establishing new sinkholes, attracting spam to a new domain takes some effort (we found some amusement in the difficulty of attracting spam when we actually wanted to receive it). In addition to using sinkholes, network operators might share network-level statistics of received email from *real* network domains to pre-emptively detect and filter spamming hosts.

8. Conclusion

This paper has studied the network-level behavior of spammers using a joint analysis of a unique combination of datasets—a 17-month-long trace of all spam sent to a single domain with real-time traceroutes, passive TCP fingerprints, and DNSBL lookup results; BGP routing announcements for the network where the sinkholes are located; command and control traces from the Bobax spamming botnet; and mail logs from a large commercial email provider.

This analysis allowed us to study some new and interesting questions that should guide the design of better spam filters in the future, based on the lessons in Section 7. We studied network-level behavior of spammers and compared these characteristics to those of legitimate email, noting some differences that could help identify spammers by IP address space or AS. We also used "ground truth" Bobax drones to better understand the characteristics of spamming botnets, and we found that most of these drones do not appear to revisit the same domain twice. While this property does not appear to hamper the use of blacklists for identifying bots (emphasizing the benefits of collaborative spam filtering), we found that blacklists were remarkably ineffective at detecting spamming relays that sent spam from IP addresses scattered throughout a briefly announced (and typically hijacked) IP address block—a new technique we call "BGP spectrum agility". This technique is lethal because it makes traceability and blacklisting significantly more difficult. Spam filters that incorporate *network-level* behavior could not only mitigate this class of attack and many others, but they could also prove to be more resistant to evasion than content-based filters.

Acknowledgments

We thank David Mazières, David Dagon, and Suresh Ramasubramanian, whose traces made this study possible. We are also grateful to David Mazières for supporting Mail Avenger, to Hari Balakrishnan for inspirational discussions and the use of physical resources at MIT, and to David Andersen, Randy Bush, Wenke Lee, Vern Paxson, Michael Walfish, the anonymous reviewers, and our shepherd, Adrian Perrig, for helpful feedback. Finally, we thank the Datapository project and Emulab for providing resources that we used in our analysis.

REFERENCES

[1] D. Bank and R. Richmond. Where the Dangers Are. *The Wall Street Journal*, July 2005. http://online.wsj.com/public/article/SB112128442038984802-w4qR772hjUeqGT2W0FIcA3_FNjE_20060717.html.

[2] M. Casado, T. Garfinkel, W. Cui, V. Paxson, and S. Savage. Opportunistic measurement: Extracting insight from spurious traffic. In *Proc. 4th ACM Workshop on Hot Topics in Networks (Hotnets-IV)*, College Park, MD, Nov. 2005.

[3] CNN Technology News. Expert: Botnets No. 1 emerging Internet threat. http://www.cnn.com/2006/TECH/internet/01/31/furst/, Jan. 2006.

[4] Description of coordinated spamming, Feb. 2005. http://www.waltdnes.org/spam.

[5] J. Evers. Most spam still coming from the U.S. http://news.com.com/Most+spam+still+coming+from+the+U.S./2100-1029_3-6030758.html, Jan. 2006.

[6] N. Feamster. Open problems in BGP anomaly detection. In *CAIDA Workshop on Internet Signal Processing*, San Diego, CA, Nov. 2004.

[7] N. Feamster, D. Andersen, H. Balakrishnan, and M. F. Kaashoek. Measuring the Effects of Internet Path Faults on Reactive Routing. In *Proc. ACM SIGMETRICS*, pages 126–137, San Diego, CA, June 2003.

[8] N. Feamster, J. Jung, and H. Balakrishnan. An Empirical Study of "Bogon" Route Advertisements. *ACM Computer Communications Review*, 35(1):63–70, Nov. 2004.

[9] Goodmail Systems, 2006. http://www.goodmailsystems.com/.

[10] J. Goodman. IP Addresses in Email Clients. In *First Conference on Email and Anti-Spam*, Mountain View, CA, July 2004.

[11] S. Hansell. Postage is due for companies sending email, February 5, 2006. http://www.nytimes.com/2006/02/05/technology/05AOL.html.

[12] Honeynet Project. Know Your Enemy: Tracking Botnets. http://www.honeynet.org/papers/bots/botnet-commands.html, 2006.

[13] J. Jung and E. Sit. An Empirical Study of Spam Traffic and the Use of DNS Black Lists. In *Proc. ACM SIGCOMM Internet Measurement Conference*, pages 370–375, Taormina, Sicily, Italy, Oct. 2004.

[14] A. Kumar, V. Paxson, and N. Weaver. Exploiting Underlying Structure for Detailed Reconstruction of an Internet-scale Event. In *Proc. ACM SIGCOMM Internet Measurement Conference*, Berkeley, CA, Oct. 2005.

[15] R. Mahajan, D. Wetherall, and T. Anderson. Understanding BGP Misconfiguration. In *Proc. ACM SIGCOMM*, pages 3–17, Pittsburgh, PA, Aug. 2002.

[16] MailAvenger, 2005. http://www.mailavenger.org/.

[17] J. Mason. Spam Forensics: Reverse-Engineering Spammer Tactics. http://spamassassin.apache.org/presentations/2004-09-Toorcon/html/, Sept. 2004.

[18] Microsoft security bulletin ms04-011. http://www.microsoft.com/technet/security/bulletin/ms04-011.mspx, Apr. 2004.

[19] D. Moore, C. Shannon, and J. Brown. Code-red: A case study on the spread and victims of an internet worm. In *Proc. ACM SIGCOMM Internet Measurement Workshop*, Marseille, France, Nov. 2002.

[20] Operating System Market Shares. http://marketshare.hitslink.com/report.aspx?qprid=2, Jan. 2006.

[21] The Open Relay Database, 2006. http://ordb.org/.

[22] M. Prince, B. Dahl, L. Holloway, A. Keller, and E. Langheinrich. Understanding How Spammers Steal Your E-Mail Address: An Analysis of the First Six Months of Data from Project Honey Pot. In *Second Conference on Email and Anti-Spam*, Stanford, CA, July 2005.

[23] Project Honey Pot. http://www.projecthoneypot.org/.

[24] A. Ramachandran and N. Feamster. Understanding the Network-Level Behavior of Spammers. Technical Report GT-CSS-2006-001, Georgia Tech, Feb. 2006.

[25] S. Ramasubramanian. Port 25 filters - how many here deploy them bidirectionally? http://www.merit.edu/mail.archives/nanog/2005-01/msg00127.html, Jan. 2005.

[26] The Spam and Open Relay Blocking System (SORBS), 2006. http://www.sorbs.net/.

[27] SpamAssassin, 2005. http://www.spamassassin.org/.

[28] Spammer-X. *Inside the Spam Cartel*. Syngress, Nov 2004.

[29] S. Staniford, V. Paxson, and N. Weaver. How to 0wn the Internet in Your Spare Time. In *Proc. 11th USENIX Security Symposium*, San Francisco, CA, Aug. 2002.

[30] J. Todd. AS number inconsistencies, July 2002. http://www.merit.edu/mail.archives/nanog/2002-07/msg00259.html.

[31] ZDNet Security News. Most spam genrated by botnets, expert says. http://news.zdnet.co.uk/internet/security/0,39020375,39167561,00.htm, Sept. 2004.