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 (in time) each spamming host is, botnet spamming characteristics, and techniques for harvesting email addresses. This paper studies these questions by analyzing an 18-month trace of over 10 million spam messages collected at one Internet "spam sinkhole", and by correlating these messages with the results of IP-based blacklist lookups, passive TCP fingerprinting information, routing information, and botnet "command and control" traces.

We find that a small, yet non-negligible, amount of spam is received from IP addresses that correspond to short-lived BGP routes, typically for hijacked addresses. Most spam was received from a few regions of IP address space. Spammers appear to make use of transient "bots" that send only a few pieces of email over the course of a few minutes at most. These patterns suggest that developing algorithms to identify botnet membership, filtering email messages based on *network-level* properties (which are less variable than an email's contents), and improving the security of the Internet routing infrastructure, may be prove extremely effective for combating spam.

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 contents of spam, but comparatively little attention has been focused on spam's *network-level properties*. Conventional wisdom often asserts that most of today's spam comes from botnets, and a large fraction of spam comes from Asia, and a few studies have attempted to quantify some of these characteristics [11].

Unfortunately, little is known about what quantity of spam comes from botnets vs. 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 simply exposing spammers' behavior, gathering information about the network-level behavior of spam could well prove to be a huge asset for designing spam filters that are based on spammers' network-level behavior. 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. Specifically, our insight is that 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 spam sinkholes for two domains. The trace at one domain contains all spam received at the domain since August 2004 and serves as our primary dataset. The trace at the second domain contains all spam received at a newly registered domain since November 2005; while its spam volume to date is modest, the fact that we can observe spam arrival at this domain from "time zero" has allowed us to better understand harvesting techniques.

We perform a *joint analysis* of the data collected at these sinkholes, together with packet traces, an archive of BGP route advertisements as heard from the receiving network, traceroutes from the receiving mail relay to the spammer's mail relay at the time the relay sent the mail, traces from the botnet "command and control" of the Bobax worm, and traces of legitimate email from the border router of a large campus network. Although many aspects of mail headers can be forged, we base our analysis strictly on properties of the sender that cannot be forged (*e.g.*, the IP addresses that made connections to our mail servers, passive TCP fingerprints, packet traces of those connections, 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).* Many models of worm propagation assume a uniform distribution of vulnerable hosts across IP address space (*e.g.*, [25]), and spam filtering techniques currently make no assumptions about the distribution of spam across IP address space. In fact, 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 (predominantly 61.* – 80.* and 200.* – 215.*).

- *Most received spam is sent from transient Windows hosts, each of which sends a relatively small volume of spam (Section 5).* Most bots send a relatively small volume of spam (*i.e.*, less than 100 pieces of spam over 18 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 remain untraceable (Section 6).* A small portion of spam is sent by sophisticated spammers, who briefly advertise IP prefix space, establish an SMTP connection to the victim's mail relay, and withdraw the route to that IP address space after the client sends spam. Anecdotal evidence has suggested that spammers exploit the routing infrastructure to remain untraceable [1, 26]; 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" IP addresses within this space.
- *Harvesting entities and spamming appear to be conducted from distinct infrastructure, if not totally separate organizations (Section 7).* This finding also suggests that filtering spam by observing entities that first perform harvesting is not likely to be successful—in fact, these "crawlers" never appear to send spam.

We readily 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. The spam we have collected still represents an interesting dataset since it reflects the *complete set of spam received by two Internet domains*. This dataset exposes spamming as a typical network operator for a single Internet domain might also witness it. This unique vantage point can help us better understand whether the features of spam that any single network operator observes can be useful in developing more effective filtering techniques.

Beyond this practical utility, 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.

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 8. First, the ability to trace the identities of spammers hinges on securing the routing infrastructure. Second, the uneven distribution of spam (and botnet activity) across IP space—and the differences in this distribution from legitimate email—suggests that spam filters and intrusion detection systems might monitor *network-wide* spam arrival patterns for changes in these distributions to detect anomalies such as a surge in spam activity. This characteristic also suggests that individual spam filters might be able to attribute higher levels of suspicion to spam originating from IP address space with higher spam activity. Given the transient nature of most spamming hosts, incorporating general network-level properties of spammers may ultimately provide significant gains over more traditional filtering methods (*e.g.*, content-based filtering).

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 properties of the infrastructure used by spammers, in particular the relationship between the spam received at our sinkholes and known spamming bots. Section 6 examines the extent to which spammers use transient IP addresses—specifically, short-lived BGP route announcements—to send spam untraceably. In Section 7, we describe preliminary case studies of harvesting we have observed at a newly created spam sinkhole. Based on our findings, Section 8 offers positive recommendations for designing more effective mitigation techniques. We conclude in Section 9.

2. Background and Related Work

In this section, we provide an overview of known spamming techniques. Although many of these spamming techniques have been acknowledged anecdotally, several of them (*e.g.*, does spam actually arrive from short-lived BGP route announcements?) have not been confirmed or quantified prior to this study.

2.1 Spam: Methods and Mitigation

In this section, we provide 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: Old and New

Spammers use various techniques to send large volumes of mail while remaining as untraceable as possible, including:

Botnets. Conventional wisdom suggests that the majority spam on the Internet today is sent by botnets—collections of machines acting under one centralized controller [27, 4, 5]. The `W32/Bobax` ("Bobax") worm (of which there are many variants), exploits the DCOM and LSASS vulnerabilities [17], allows the 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. Agobot and SDBot are two other bots purported to send spam [10].

Direct spamming. Spammers often 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. To avoid renumbering problems in these cases, spammers sometimes obtain a pool of dialup IP addresses, send outgoing traffic from the high-bandwidth connection, and proxy the reverse traffic through the dialup connection back to the spamming hosts [22].

BGP spectrum agility. This paper exposes a new type of cloaking mechanism—BGP "spectrum agility"—whereby spammers briefly announce (often stolen) IP address space from which they send spam and withdraw the routes to that IP address space once the spam is sent, in order to remain untraceable. Although anecdotal evidence has suggested that spammers use may use this technique [1], our study finds that spammers may be using spectrum agility to complement spamming by other methods. This paper documents several interesting cases of this activity.

Open relays and proxies. Some SMTP servers will allow any client to connect to it for the purposes of sending email. Originally intended for convenience purposes (*e.g.*, to let users send mail from a particular SMTP server while traveling or otherwise in a different network), open relays were readily exploited by spammers because the layer of indirection allowed them to remain untraceable. It would appear that the widespread deployment and use of blacklisting techniques have all but extinguished the use of open relays to send spam.

2.1.2 Mitigation Techniques

Techniques for stemming the tide of spam are as varied as the techniques to send spam. 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 an the contents of an email message's headers or body to determine whether an email is likely to be spam. Content-based filters, such as those incorporated by popular spam filters such as SpamAssassin [23], have been quite successful to date at reducing 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 continue to come up with 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 email contents to attempt to evade these filters. In contrast, altering the network characteristics of where spam is being sent from, and how it is being sent, is more costly. For all the work that has focused on developing filters based on email contents, scant attention has been devoted to the *network-level properties* associated with spamming behavior.

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, open proxies exist today and remain one of the 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 ranges 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 is likely to become much less effective over time.

2.2 Related Work

In this section, we review previous work in three areas: spam, worms and botnets, and unorthodox interdomain routing announcements. While previous work has studied each of these phenomena to some degree in isolation, we believe that 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 Previous Studies of Spamming Behavior

A recent presentation from the SpamAssassin project discusses several techniques that the SpamAssassin spam filtering tool has incorporated to to detect forged `X-Mailer` headers, weak "hashbusting" schemes, etc. [16]. Although this work also involves reverse engineering, the project focuses on analyzing mail *contents* to reverse-engineering spamming tools and techniques (with the goal of using this analysis to incorporate better content-filtering rules into SpamAssassin). In this paper, we also study properties of spamming behavior, but we focus on network-level properties, rather than artifacts of spamming software that appear in email content. In particular, we focus on properties of the spam, such as the IP address of the last relay from which the mail was sent before the local domain, which previous work has also observed is one of the few parts of the SMTP header that cannot be forged [8].

Previous studies have studied 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) [3]. Although we have not yet estimated bandwidths of spammers that send spam to our sinkhole, studying the passive port 25 packet trace that we have also captured at our sinkhole is part of our future work. 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) [13]. 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).

Jung *et al.* previously performed a study of DNS blacklist (DNSBL) traffic and the use of blacklists [12] and observed that 80% of of the IP addresses that were sending spam were listed in DNSBLs two months after the collection of the traffic trace. Our study also studies the effectiveness of DNSBLs but examines whether a client is listed in the DNSBL *at the time the corresponding piece of mail was received*, and with a different dataset. While we also find that about 80% of the received spam was listed in at least one of eight blacklists, hosts that employ certain spamming techniques such as BGP spectrum agility tend to be listed in far fewer blacklists. We also find that most spam comes from only a handful of address ranges; thus, blacklisting on ranges, rather than individual IP addresses, may also help improve the effectiveness of blacklists.

Several previous and ongoing studies are studying spammers' attempts to harvest email addresses for the purposes of spamming. Project Honeypot also sinks email traffic for unused MX records and hand out "trap" email addresses to harvesting behavior and help identify spammers [21]. A previous study has used the data from Project Honeypot to analyze the methods spammers use to 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 [20]. We present some preliminary results from a similar study in Section 7.

In this paper, we correlate spam arrivals with traces of hosts known to be infected with malware. Moore *et al.* used "backscatter" traces to a /8 network to study the spread of the CodeRed word in July 2001 [18]. Although we do not study the spread of malware in this paper, their paper's findings that the majority of hosts—and more than 80% of the hosts in Asia—did not patch the relevant vulnerability well after actual outbreak make it more reasonable to assume that IP addresses of positively identified Bobax drones remain infected across the course of our spam trace.

2.2.2 *Unorthodox route announcements*

Anecdotal evidence and cursory studies have suggested that spammers advertise routes to IP prefixes for short amounts of time to send spam while remaining undetectable [1, 24, 26]. This paper is the first to quantitatively confirm this suspicion. Feamster *et al.* performed an empirical study on route advertisements in bogus address spaces (*i.e.*, private address space or unassigned addresses) [6]. In Section 6, we document cases where the sending of spam coincides with short-lived BGP route announcements for IP prefixes containing the mail relays that send spam. To our knowledge, this paper is the first to quantify the extent to which spam originates from mail relays that are only reachable for short periods of time.

3. Data Collection

This section describes the datasets that we use in our analysis. Our primary dataset is are the actual spam email messages collected at two sinkhole domains. To study the specific characteristics of certain subsets of spammers, we augment this dataset with two additional datasets: First, 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. We also intercept the "command and control" traffic from the Bobax worm at a sinkhole at a large campus network 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).

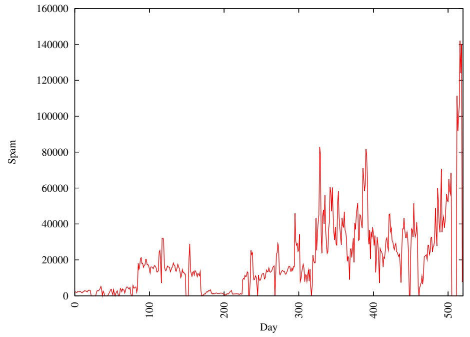

Figure 1: The amount of spam received per day at our sinkhole from August 2004 through December 2005.

3.1 Spam Email Traces

To obtain a sample of spam, we registered a domain and established a corresponding DNS Mail Exchange (MX) record with *no legitimate email addresses*. 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 the sinkhole has 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 (not shown in the graph) is also increasing.

We established a second sinkhole in November 2005 to measure the process by which spammers "ramp up" in sending spam to a domain (*e.g.*, the process by which email addresses are harvested and exchanged, methods that spammers use for harvesting, etc.). We registered the second domain in November 2005, linked to a web site for that domain from our personal web pages, and seeded Google's web crawler with the domain. On this page, we set up links to a "contact" web page that randomly generates a list of email addresses at that domain. Joining these randomly generated email addresses, as they may become seen at a later date in our spam logs, allows us to track the "life" of an email address from when it was harvested to when we receive spam from that address (and from whom we receive spam).

In addition to simply collecting spam traces, these spam sinkholes run MailAvenger [15], a customizable SMTP server that allows us to take specific actions upon the receipt of 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). These sinkholes are hosted by domains that resolve to mail exchangers that that run the MailAvenger SMTP server, which we have configured 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, these mail servers collect the following information about the mail relay *at the same time that the spam itself was 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 BGP Routing Measurements

To gain a view of network-layer reachability from the network where spam was received, we co-located a "BGP monitor" in the same network with our spam sinkholes. The monitor receives BGP updates from the border router, and our analysis includes a BGP update stream that overlaps with almost all of our spam trace, ending on December 28, 2005. Because the monitor has an internal BGP session to the campus network's border router, it will not see *all* BGP messages heard by the border router. Rather, it will see only BGP messages 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*. In this study, we are primarily concerned with 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. Even though our BGP monitor receives only the best route for each IP prefix, we can nevertheless determine whether a prefix is reachable by virtue of the fact that the monitor will have *no* route to the prefix at all if the prefix is unreachable.

3.3 Botnet Command and Control Data

To gain a definitive accounting of hosts that are sending email from botnets, we use 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 DNS hijacking technique directs control of the botnet to the honeypot, which effectively disables it for spamming for this period (*i.e.*, the 1.5-day period in April 2005). On the upside, since all infected drones now attempt to contact the honeypot, rather than the intended command-and-control host, we can take a packet trace to obtain a reasonable estimate for the size of the botnet and 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 18 months, hosts may be patched, in cases of dynamic addressing, multiple different hosts (some of which may be Bobax-infected and some of which may not be) may use one of the IP addresses logged from the Bobax trace. 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 is slow enough that we can expect that many of these infected hosts remain unpatched [18].

4. Network-level Characteristics of Spammers

In this section, we study some "traditional" network-level characteristics of spammers. We survey the portions of IP address space from which our sinkhole received spam and the persistence of this distribution over time. While we do not present specific results to this effect, we find that these distributions are quite persistent over time. The distribution of spam senders across IP address space is far from uniform, and it differs significantly from the distribution of IP addresses of senders of legitimate email in certain parts of the address space. Further, spam arrival by *IP prefix* is much more pronounced, persistent, and concentrated than similar characteristics by IP address. Finally, 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.

These *network-level characteristics* of spam, which the rest of this section surveys in greater detail, suggests that spam filters that focus on the relatively small fraction of /24 prefixes where spam arrives continually would complement techniques that blacklist based only on individual IP addresses. This heavily skewed distribution, both in IP space and by AS number, suggests that spam filtering efforts might better focus their energy on identifying high-volume, persistent groups of spammers, rather than on blacklisting individual IP addresses, many of which are transient. As we will see in Section 5, this conclusion is even stronger when we restrict our analysis to the set of spamming hosts that are known to be botnets.

4.1 Distribution Across Networks

The fact that the vast majority of spam originates from a relatively small portion of the IP address space that differs from the distribution of legitimate email suggests that it may be possible to design spam filters that target small portions

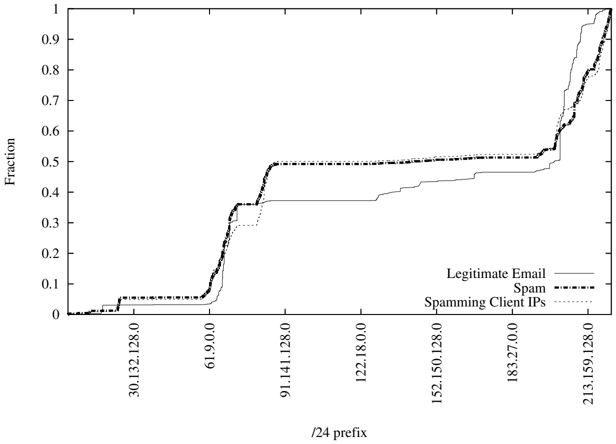

Figure 2: Fraction of spam email messages and legitimate email addresses received as a function by IP address space; also, fraction of client IP addresses that sent spam, binned by /24.

of the IP address space. This distinction also suggests that it may be possible for a network operator to automatically detect a sudden influx of spam by tracking the changes in distribution of IP address space for incoming mail.

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

Finding 4.1 (IP Address Space Distribution) *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. The cumulative graph clearly shows several "knees", the most distinctive of which are in the IP address spaces for cable modems (*e.g.*, 24.*) and in the address space allocated to the Asia Pacific Network Information Center (APNIC) regional Internet registry (*e.g.*, 61.*).

We repeated this study per day across months, per month across years, and so forth. Surprisingly, this distribution has remained roughly constant over time. This finding offers two implications for the design of spam filters. First, although the individual IP addresses from which spam is received may change from day-to-day, the fact that spam continually comes from the same IP address *space* suggests that spam filters should incorporate this feature when assessing whether a piece of email is in fact spam.

Despite the massive spread of Internet hosts across IP space, Figure 2 suggests that, in fact, most spam is coming from a relatively concentrated portion of the address space. We compared this distribution to that of IP addresses of *all* mail relays that sent mail to a large campus network and found that, while the distributions are largely similar, significantly more spam than legitimate email comes from the range from 70.* – 80.*; this characteristic is notable because several of these blocks (*i.e.*, 77/8, 78/8, and 79/8) are reserved by the Internet Assigned Numbers Authority (IANA). These differences in distribution suggest that spam filters could assign a higher level of suspicion to email sent from relays in this address space (particularly the reserved space).

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.

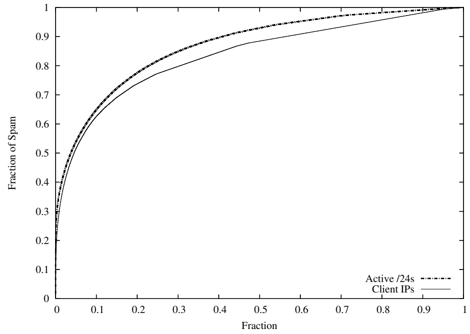

Figure 3: The distribution of spam messages across the /24 has any hosts that send spam all IP addresses that send spam and all "active" /24s (*i.e.*, those that send at least one piece of spam).

Figure 3 shows that roughly half of the received spam arrives from less than 3% of /24s that receive any spam at all (only about 486,614 /24s receive any spam at all); half of the spam comes from only about 0.01% of all /24s. Figure 4 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 will typically not be fruitful in mitigating spam without sharing information across domains. The concentration of spammers in relatively concentrated regions of IP address space and the relative transience of individual IP addresses suggests that network operators (and spam filters) should attribute a higher level of suspicion to spam coming from IP address space where spam commonly originates.

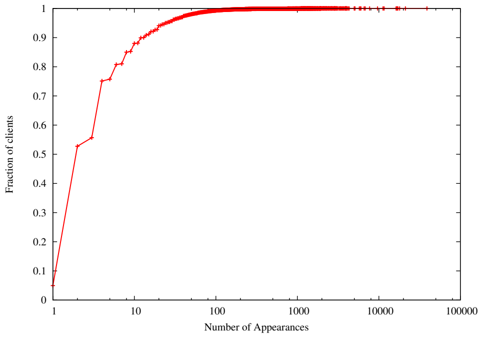

Figure 4: The number of distinct times that each client IP sent mail to our sinkhole (regardless of the number emails sent in each batch).

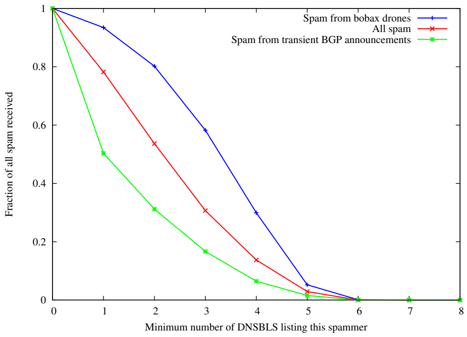

Figure 5: The fraction of mails that were listed in a certain number of blacklists or more, *at the time each mail was received*.

Finding 4.2 (Distribution of spammers across ASes) *More than 10% of spam received at our sinkhole originated from mail relays in two ASes, and 36% of all received spam originated from only 20 ASes.*

Recent claims have suggested that most spam in fact originates in the United States [11]. On the other hand, Figure 2 suggests that a significant number of spamming hosts reside in an 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. Table 1 shows also the distribution of hosts that sent spam to the sinkhole by country, for the top 20 ASes from which we received spam.

Finding 4.3 (Distribution of spammers 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 nearly 40% of all spam from the top 20 ASes.*

Furthermore, our estimates over 65% of the corpus suggest that nearly three times as much spam in our trace originates from ISPs based in the US 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 the most spam comes from the U.S., but to a much lesser degree. The fact that most spam comes from a large number of United States-based providers that also provide service for many legitimate customers (*e.g.*, Comcast, Level3, etc.) suggests that filtering spam based on the AS of the mail relay is not likely to be effective.

4.2 The Effectiveness of Blacklists

Our observations that most spam comes from a small portion of the address space led us to wonder whether filtering techniques that used network level properties other than a mail relay's IP address might improve the effectiveness of blacklist-based filtering strategies. Indeed, we also wondered how effective DNSBL filtering based on IP address would be *at all*, given that, as shown in Figure 4, most mail relays never send spam at more than two distinct instances in time. To test this hypothesis, we used the results from real-time DNSBL lookups performed by MailAvenger at the time the mail was received at 8 different blacklists.

Figure 5 indicates that, in fact, IP-based blacklisting is still working reasonably well: while 20% of spammers' IP addresses were not listed in *any* blacklist (as shown by the middle line "all spam", where about 80% of IP addresses were listed in at least one blacklist), more than 50% of all spam was listed in two or more blacklists, and 80% of spam from known botnets were from IP addresses that were 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) *About 80% of all spam was received from mail relays that appear in at least one blacklist. 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.*

We discuss BGP spectrum agility in more detail in Section 6, but the general ineffectiveness of blacklists for detecting IP addresses from this space suggests that this technique is quite effective and may gain prominence, and possibly used in conjunction with botnets (which appears to be the predominant spamming technique, as we discuss in Section 5).

5. Spam from Botnets

In this section, we amass circumstantial evidence which 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 perform a joint analysis with 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" worm and find that the IP address space where we observe worm activity bears close similarity to the IP address

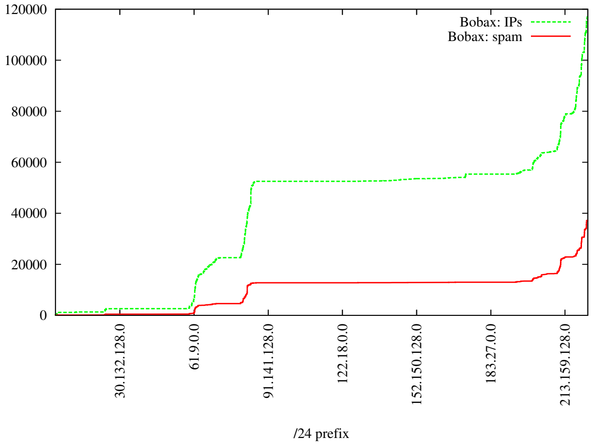

Figure 6: The number of *all* Bobax drones, and the amount of spam received from those drones at the sinkhole, as a function of IP address space. On the x-axis, IP address space is binned by /24.

space where we observed spamming activity (Finding 4.1). Second, we observe that about 95% of the spam received at our sinkhole appears to be sent by Windows hosts that each send relatively low volumes of spam.

5.1 Bobax Topology

We studied prevalence of spammers versus the prevalence of known Bobax drones to gain a better understanding of how the distribution of IP addresses of Bobax-infected hosts compared to our observations of IP distribution of spammers in general. Figure 6 shows the results of this analysis, which shows, surprisingly, that the distribution of *all* Bobax-infected hosts is quite similar to that of the distribution of all spammers (Figure 2).

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

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 61.* – 74.* has a sizable number of Bobax-infected hosts, we see relatively less spam from the them in this space. One possible explanation for this is that spammers may be using other techniques besides 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 sizable number of machines (or, perhaps, a smaller number of machines with changing IP addresses), numbered from portions of unused IP space (within this IP address range) that are typically unroutable, except for the times when they are sending spam.

Operating System	*Clients*	*Total Spam*
Windows	854404	5863112
- Windows 2000 or XP	604252	4060290
- Windows 98	13727	54856
- Windows 95	559	2797
- Windows (other/unconfirmed)	235866	1745169
Linux	28132	557377
FreeBSD	6584	152456
MacOS	2944	46151
Solaris	1275	18084
OpenBSD	797	21496
Cisco IOS	736	5949
NetBSD	44	327
HP-UX	31	120
Tru64	26	143
AIX	23	366
OpenVMS	18	62
IRIX	7	62
Other/Unidentified	128580	1212722
No Fingerprint	204802	2225410
Total	1228403	10103837

Table 2: The operating system of each unique sender of received spam, as determined by passive OS fingerprinting.

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 MailAvenger; using this technique, we can associate each piece of spam with an operating system. Using this technique, we were able to identify the operating system for about 75% of all hosts from which we received spam. Table 2 shows the results of this study. Approximately half of the hosts from which we receive spam run Windows; this fraction is surprisingly small, given that roughly 95% of all hosts on the Internet run Windows [19].

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.*

A significant fraction of the spamming infrastructure is apparently still Unix-based.[1] Over time, this fraction may in fact *increase*, both as spammers develop different, more sophisticated cloaking techniques.

[1] 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.

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 disproportionately large fraction of spam comes from Windows hosts, our hypothesis is that many of these machines are infected hosts that are bots. (To test this hypothesis, we intend to check the distribution of legitimate email by operating system type, but we have not yet done so.) 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?*[2]
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 personal curiosity (and to compare with other claims about the amount of spam coming from botnets [4]), we wanted to determine the total fraction of received spam that originated from botnets versus other mechanisms. The circumstantial evidence we have amassed in Sections 5.1 and 5.2 suggest that the fraction of spam that originates from botnets is quite high. Unfortunately, we have not yet developed a technique 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, as well as 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 bots across IP space).

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 become more sophisticated.

Based on the short lifespans of the majority of bots, we hypothesized that IP-based blacklists (*e.g.*, DNSBL filtering) are unlikely to be effective in blocking spam from, at least the 65% of bots that send spam to our sinkholed domain only once. This hypothesis turns out to be generally incorrect. As Figure 5 shows, the botnet hosts from which we received spam were actually *more* likely to be listed in more DNSBLs than the typically spamming mail relay. Intuitively, this can be justified, since other domains likely received spam from the same drones, even the ones from which our domain only received a single piece of spam, but this result also demonstrates the benefits of collaborative spam filtering (of which DNSBLs are the primary example): they can facilitate identification of spammers that send only a single piece of spam to a domain when those spammers recur across domains.

5.3.3 *Volume*

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

[2] 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) [18]. 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.

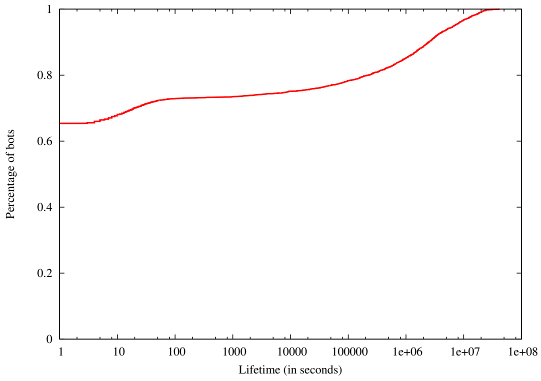

Figure 7: Bobax drone persistence.

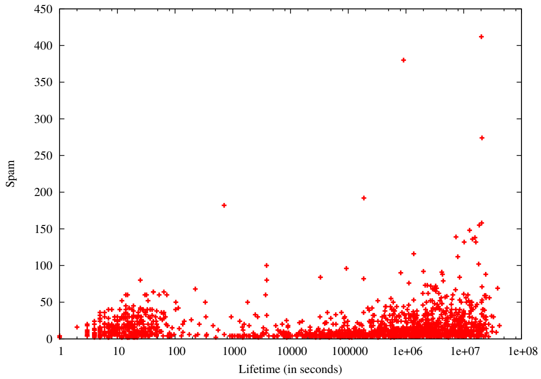

Figure 8: Number of spam email messages received vs. bobax drone persistence.

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 have typically sent no more than 100 pieces of spam, 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 the extent to which the sinkhole domain receives mail that coincides with short-lived BGP route announcements.

Anecdotal evidence has previously suggested 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 route announcements for that IP address space after the relays have sent spam [1, 24, 26]. This practice make 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. Because this IP address space is unreachable the vast majority of the time, it is unlikely that the IP address that sent the spam will even be reachable at the time when a network operator is investigating the incident.

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 (it accounts for no more than 10% of all spam we receive, and probably 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 certainly worth attention, particularly since hiacking large prefixes is a practice that 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. (This particular announcement appears to be particularly interesting; we will return to this example shortly.)

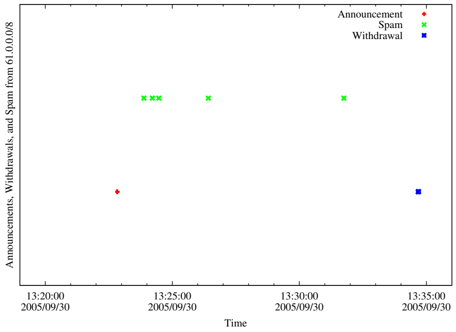

Figure 9: Observation of a short-lived BGP route announcement for `61.0.0.0/8`, spam arriving from mail relays in that prefix, and the subsequent withdrawal of that prefix.

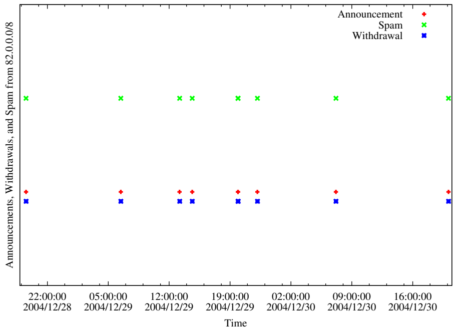

Figure 10: Observation of a short-lived BGP route announcement for `82.0.0.0/8`, spam arriving from mail relays in that prefix, and the subsequent withdrawal of that prefix.

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 3.2) 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, there are a small number of parties who appear to be 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 involved such large portions of IP address space. Although some short-lived routing announcements may be misconfigurations [14], the fact that these routing announcements continually appear, they are for large address blocks, and 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. While the use of large address blocks might initially seem surprising, the dispersity of IP addresses of the clients sending spam corresponding to the short-lived analysis has suggests the following alternate theory.

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, thus evading filters in two ways. First, route announcements for shorter IP prefixes are less likely to be blocked by route filters. Second, the larger 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 5 and our analysis in Section 4.2, the technique seems to be rather effective.

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 somewhat inconclusive. Still, many of the ASes that send the most spam with this technique also appear to be relative newcomers, and it is our belief that variants of this type of technique may 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.

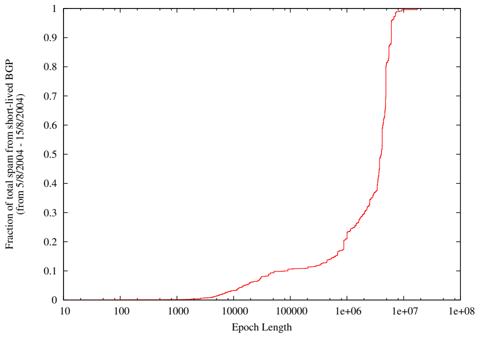

Figure 11: CDF of the length of each short-lived BGP episode, *in seconds*, for ten days in August 2005.

6.2 Prevalence of 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 August 5, 2005 to August 15, 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 ten days of our trace, 90% 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, as much as 10% of all received spam may appear to coincide with this type of behavior.

Finding 6.2 (Prevalence: Spam from Short-Lived Routes) *Approximately 10% of spam was received from routing announcements that lasted less than a single 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 as one of 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 received 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 spammers that do use this technique typically use it quite regularly. We also observed that many of the ASes where this technique has been witnessed 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 that coincide with spam do not appear to be hijacking IP prefixes to do so. 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. A Preliminary Survey of Harvesting

To better understand the harvesting techniques used by spammers, we established a new domain and pointed its DNS mail exchanger (MX) record to our second spam sinkhole, as mentioned in Section 3.1. Establishing this domain has allowed us to observe the relationship between harvesting to actual spam arrival, similar to that which being performed in other studies [21]. After registering the MX record, we built a web site for that domain, with a "contacts" list that consists of randomly generated, non-existent email addresses at that domain. Since these email addresses are random combinations of letters, it is very unlikely that email sent to those addresses are the result of a dictionary attack. We also log a list of the email addresses that are fed to clients accessing the contact page, together with other information such as time of crawling, client IP, HTTP User Agent, etc.). By combining these logs with the "To:" addresses to which we receive spam, we were able to identify some of the techniques spammer use to harvest email addresses and send spam.

The domain was registered on November 19, 2005, and the SMTP server (MailAvenger [15]) was set up on December 6, 2005. The setup is similar to our primary sinkhole's configuration: email to any username is accepted and logged.

Though our first two pieces of spam appeared within 5 days, it appeared to be a random attack: an analysis of their headers and our logs showed no evidence of email being received to "fed" addresses. Our first real evidence of active harvestation of email addresses, a *Phishing* [2] attack appeared over the course of a day starting on January 20th, 2006, from two Windows machines. Three days after the first attack, one of the machines spammed our domain again, under the guise of a different organization.

An analysis of the attack unearthed a number of interesting features. First, all email addresses to which we received spam from these two machines were harvested in a single attempt on January 16, 2006. The IP address which harvested the spam was logged as 69.192.210.155, which is IP space belonging to Rogers Cable, but the IP space of the machines that sent us spam (65.220.17.5 and 65.220.15.30) belongs to UUNET Technologies Inc. Though the email addresses that we the harvester were in no particular order, the spamming organization appears to have *sorted* the list alphabetically and delegated approximately half the set to each machine. We also found that both machines were active at the same time and sent spam at approximately 15-minute intervals, which indicates some level caution on the spammer's part to avoid triggering network alerts. Unsurprisingly, many of the mail headers were also forged. For instance, the *X-Mailer* headers, which usually identifies the Mail User Agent (MUA), were consistently forged: the same machine had different X-Mailer strings ("AOL 7.0 for Windows US sub 118", "Microsoft Outlook Express 5.50.4133.2400", "Microsoft Outlook IMO, Build 9.0.2416 (9.0.2910.0)" etc.) for consecutive emails it sent.

Although this spam dataset is relatively small compared to our primary sinkhole, the short interval between harvesting and actual spamming is worrisome. Moreover, *all* the addresses that were harvested were spammed within a short period. The sophistication of spammers (trying to aviod detection by not flooding the domain with emails, pre-processing and balancing the "load" between available machines, tampering with message headers, etc.) and the apparent cooperation between different groups (harvesting from one IP block, spam from another) are all cause for concern.

8. 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 [7, 9]) 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 far 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 five main insights regarding the network-level behavior of spammers that could help in our design of better mitigation techniques.

Lesson 1 *Effective spam filtering requires a better notion of end-host identity.*

We observed a non-trivial amount of spam coming from "one-shot" botnets. The notion of a using an IP address to pin down a spammer is now meaningless. Short-lived botnets and short-lived BGP routing announcements (with spectrum agility) make the notion of pinning an identity to an IP address (or even IP prefix space) effectively impossible.

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

Although comprehensive IP-based blacklisting is reasonably effective (indeed, for 80% of received spam, the IP address of the sending relay was blacklisted at the time the mail was received), blacklisting techniques may also benefit by exploiting other network-level properties such as IP address *ranges*, some of which (*e.g.*, 70.* – 80.*, particularly the reserved blocks within this range) send mostly spam.

Lesson 3 *The distribution of spammers (and received) spam across IP address space is highly skewed, despite the fact that any given IP address sends a very small amount of spam.*

70% of spam is received from only 20% of all IP address space. This uneven distribution suggests that spam filters that take into account suspicious regions of IP address space (rather than simply blacklisting individual IP addresses) may be a more efficient way of identifying spammers.

Lesson 4 *Trends indicate that securing the Internet routing infrastructure is a necessary step for bolstering identity and traceability of email senders.*

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. Our study suggests that while this spamming technique 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.

Lesson 5 *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 of exploring 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* domains could also provide significant benefits by exposing patterns that are not evident from the trace of a single domain alone. One organization might be able amass such a dataset either by sinkholing a large number of domains; Project Honeypot [21], in fact solicits donations of MX records (though its corpus is still significantly smaller than ours)—*i.e.*, for registered domains that do not receive email. As we have discovered thus far from our own experience, attracting spam to a new domain takes some effort (we found some amusement in the difficulty of attracting spam that we actually wanted). Additionally, in addition to using sinkholes, network operators might share network-level statistics of received email from *real* network domains to detect anomalous behavior and, possibly pre-empt spam.

9. Conclusion

This paper has studied the network-level behavior of spammers using a joint analysis of a unique combination of datasets—an 18-month-long trace of all spam sent to a single domain with real-time traceroutes, passive TCP fingerprints, DNSBL lookup results, and traceroutes; a similar, shorter trace for a domain with a Web server that generates random email addresses and tracks who harvests them; BGP routing announcements for the network where the sinkholes are located; command and control traces from the Bobax spamming botnet; and port 25 packet traces for legitimate mail for a large campus network.

This comprehensive joint 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 8. We studied "traditional" network-level behavior (*e.g.*, where in IP space we are receiving spam from) of spammers and compared these characteristics to those of legitimate email, noting some significant differences that could help identify spammers by IP space. We also used "ground truth" Bobax drones to better understand the characteristics of spamming botnets, finding 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 Botnet drones (emphasizing the benefits of collaborative spam filtering), we also find that blacklists were remarkably ineffective at detecting spamming relays that sent spam hosts scattered throughout a briefly announced (and typically hijacked) IP address block—a new technique we call "BGP spectrum agility". Although 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 attack and many others, but could also prove to be more resistant to evasion than content-based filters.

Acknowledgments

We thank David Mazières, David Dagon, and Russ Clark, whose traces made this study possible, and Randy Bush for feedback. We are also grateful to David Mazières for supporting MailAvenger, and Hari Balakrishnan for many inspirational discussions and the use of physical resources at MIT.

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] S. Bellovin. Inside risks: Spamming, phishing, authentication, and privacy. *Communications of the ACM*, 47, 2004.

[3] 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.

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

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

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

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

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

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

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

[11] Joris 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.

[12] 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.

[13] 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.

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

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

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

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

[18] 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.

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

[20] 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.

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

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

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

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

[25] 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.

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

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