

An Empirical Study of Behavioral Characteristics of Spammers: Findings and Implications*

Zhenhai Duan,[†] Kartik Gopalan, Xin Yuan

Abstract

In this paper we present a detailed study of the behavioral characteristics of spammers based on a two-month email trace collected at a large US university campus network. We analyze the behavioral characteristics of spammers that are critical to spam control, including the distributions of message senders, spam and non-spam messages by spam ratios; the statistics of spam messages from different spammers; the spam arrival patterns across the IP address space; and the active duration of spammers, among others. In addition, we also formally confirm an informal observation that spammers may hijack network prefixes in sending spam messages, by correlating the arrivals of spam messages with the BGP route updates of the corresponding networks. In this paper we present the detailed results of the measurement study; in addition, we also discuss the implications of the findings for the (content-independent) anti-spam efforts.

Keywords: Spam, Spammer Behavior, Network Reachability of Spammers, Content-Independent Spam Filtering

1 Introduction

In recent years, content-independent (CI) anti-spam techniques that aim to determine the likelihood of an incoming message to be spam based on the properties of the message sender, instead of the message content, have attracted a lot of attention [4, 15, 18, 21]. However, in order for the CI anti-spam schemes to be effective, we must have a clear understanding of the behavioral characteristics of spammers that can distinguish spammers from the senders of legitimate messages. Behavioral characteristics of spammers such as the distributions of spam and non-spam messages by spam ratios, the statistics of spam messages from different spammers, the spam arrival patterns across the IP address space, the number of mail servers in different (spam) networks, and the active duration of spammers, can significantly affect both the feasibility and effectiveness of the CI anti-spam mechanisms. Moreover, a clear understanding of the behavioral characteristics of spammers can also facilitate the design of new anti-spam mechanisms and new email delivery architectures that are inherently spam-resistant.

In this paper we perform a detailed study of the behavioral characteristics of spammers at both the mail server and the network levels by analyzing a two-month trace of more than 25 million emails received at a large US university campus network, of which more than 18 million are spam. We also correlate the

*A preliminary version of this paper appeared in the *Proceedings of IEEE ICC 2007* with the title “Behavioral Characteristics of Spammers and Their Network Reachability Properties”.

Zhenhai Duan and Xin Yuan are with the Department of Computer Science, Florida State University, Tallahassee, FL 32306 USA (email: {duan,xyuan}@cs.fsu.edu).

Kartik Gopalan is with the Department of Computer Science, State University of New York at Binghamton, Binghamton, NY 13850 USA (email: kartik@cs.binghamton.edu)

[†]Corresponding author. Phone: 1-850-645-1651. Fax: 1-850-644-0058. Email: duan@cs.fsu.edu.

arrivals of spam with Border Gateway Protocol (BGP) route updates to investigate the network reachability properties of spammers [16]. Our study confirms the informal observation that the spam arrivals from some spammers are often closely correlated in time with the BGP announcement of the corresponding network prefixes [19]. These network prefixes are short-lived in that they are withdrawn quickly after the spamming activity is over. This sophisticated spamming technique can make it hard to track and identify the responsible spammers. In this paper we formally study the prevalence of such behavior.

We use the following terms in the exposition of our findings. The *spam ratio* of a message sender is the fraction of the messages sent by the sender that is spam. A *spam only mail server* sends only spam messages, and a *non-spam only mail server* sends only legitimate messages. A sender mail server sending both spam and legitimate messages is referred to as a *mixed mail server*. The term *spam mail servers* refers to the set of both spam only and mixed mail servers. A spam mail server sends at least one spam message in the trace. The term *non-spam mail servers* refers to the set of non-spam only and mixed mail servers. A non-spam mail server sends at least one legitimate message in the trace. Sender networks are classified similarly. The major findings from our study are as follows:

- *Most mail servers send either mostly spam or mostly non-spam messages (Section 4.1).* For example, about 5% of mail servers have a spam ratio of at most 1%, and about 90% of mail servers have a spam ratio of at least 99%. Only less than 5% of mail servers have a spam ratio between 1% and 99%. *The vast majority of spam messages come from mail servers with a high spam ratio, and a large portion of non-spam messages come from mail servers with a low spam ratio (Section 4.2).* For example, more than 91% of spam messages come from mail servers with a spam ratio of at least 90%, and about 76% of non-spam messages come from mail servers with a spam ratio of at most 10%.
- *The majority of spammers send only a small number of spam messages (Section 4.3).* For example, 93% of spam only mail servers and 58% of spam only networks send no more than 10 messages each during the two-month trace collection period. In contrast, about 0.04% of spam only mail servers send more than 1,000 messages each and are responsible for 16% of all spam messages. About 0.5% of spam only networks send more than 1,000 messages each and are responsible for 2% of all spam messages. *A large portion of spammers send spam only within a short period of time (Section 4.6).* For example, 81% of spam only mail servers and 27% of spam only networks send spam only within one day out of the two-month email collection period.
- *The vast majority of both spam messages and spam only mail servers are from mixed networks (Sections 3 and 4.3).* For example, about 91.7% of spam messages and 91% of spam only mail servers are from mixed networks. Moreover, only 6.5% of mixed networks send more than 1,000 messages each but are responsible for 75% of all spam messages.
- *The majority of both spam messages and spam mail servers are from a few concentrated regions of the IP address space (Sections 4.4 and 4.5).* For example, 68% of spam messages and 74% of spam mail servers are from top 20 “/8” IP address spaces. The top “/8” address spaces of spam messages and spam mail servers largely overlap with each other. In addition, *spam networks tend to have more mail servers than non-spam only networks.* For example, less than 1% of non-spam only networks have more than 10 mail servers. In contrast, about 14% of spam only networks have more than 10 mail servers. Alarming, about 10% of mixed networks have more than 100 mail servers, and about 1% have more than 1,000 mail servers. It is likely that a large portion of mail servers in the mixed networks are infected machines (popularly called *bots*).
- *Network prefixes for a small portion of spam only networks are only visible within a short period of time, and the short life span of these network prefixes coincides with the delivery of spam from the corresponding networks. (Section 5).* For example, during the two-month trace collection period, the

network prefixes of about 6% of spam only networks are visible for no longer than one week. In contrast, only about 2% of non-spam only networks and 2% of mixed networks have a life span less than one week.

In the remainder of the paper, we will present the details of the findings and we will also discuss the implications of the findings for the design and development of (content-independent) anti-spam schemes. In Section 2 we describe the collection of the email and BGP traces, analysis methodology, and the terminology used in the paper. We present an overview of the email trace in Section 3. We study the behavioral characteristics of spammers and their network reachability properties in Sections 4 and 5, respectively. We discuss the implications of the findings for spam control in Section 6. We describe the related work in Section 7 and conclude the paper in Section 8.

2 Preliminaries

2.1 Data Sources

The email trace was collected at a mail relay server deployed in the Florida State University (FSU) campus network between 8/25/2005 and 10/24/2005 (excluding 9/11/2005). During the course of the email trace collection, the mail server relayed messages destined for 53 sub-domains in the FSU campus network. The mail relay server ran SpamAssassin [14] to detect spam messages. The email trace contains the following information for each incoming message: the local arrival time, the IP address of the sender mail relay, and whether or not the message is spam. Due to user privacy concerns, we did not have access to the contents of any emails.

In order to study the network reachability properties of spammers, we collected the BGP updates from one peering point at the University of Oregon Route Views Project [9] over the same time period of our email trace collection. We also collected one BGP RIB table *rib(0)* from the same peering point at the beginning of our email trace collection. The BGP routing table and BGP updates were stamped with the GMT time [9]. We converted the local arrival time of incoming email messages to the GMT time so as to correlate the timestamps of the spam arrivals and the BGP route updates for the corresponding network prefixes. Ideally, both the email and BGP traces should have been collected at the same site so as to have a more precise view of the network prefixes of the message sending machines in the email trace. However, due to logistical constraints, we were unable to collect BGP traces on the FSU campus network and instead used the BGP traces from the Route Views Project. We further discuss the impact of using two separate locations for the email and BGP trace collection in the next section when we detail our analysis methodology.

2.2 Analysis Methodology

An incoming email message is classified as either *spam* or *non-spam* by SpamAssassin [14] deployed in the FSU mail relay server. SpamAssassin has a small rate of both false-positives and false-negatives. In the absence of access to the contents of the emails, it is difficult to identify these cases. We ignore the message classification errors of SpamAssassin and consider the message classification results of SpamAssassin as the ground truth. To ease exposition, we refer to the set of all incoming messages as the *aggregate* emails including both spam and non-spam. We consider each distinct IP address of the sender mail relays observed in the trace as a separate sender *mail server*. In reality, multiple IP addresses may be associated with a single mail server. We ignore this in our study. A mail server is classified as either *non-spam only*, *spam only*, or *mixed*, depending on if spam messages are received from the server. A mail server is classified as a non-spam (spam) only server if we *only* receive non-spam (spam) messages from the server. If we receive both spam and non-spam messages from a mail server, we classify it as a mixed server. To ease exposition, we

refer to the set of spam only and mixed servers as *spam* mail servers, which sent at least one spam message; we also refer to the set of non-spam only and mixed servers as *non-spam* mail servers, which sent at least one non-spam message.

We consider each distinct *network prefix* announced by BGP updates as a separate *network* [3]. We perform the longest prefix match to determine the network to which a mail server belongs. Consider an email message arriving at time t . We recursively apply the BGP updates up to time t to the original BGP RIB table $rib(0)$ to obtain the BGP RIB table $rib(t)$ at time t . We then perform the longest prefix match against $rib(t)$ to identify the network to which the sender mail server belongs. Because of propagation delays and policy issues, the two vantage points where we collected the email trace and BGP updates may have different views on the network reachability. Consequently, we may not identify the network a mail server belongs to for all mail servers in this way. If we cannot identify the network the mail server belongs to in this manner, we consider the mail server belongs to the network with the longest matched prefix that we have seen up to time t (note that this prefix must have been withdrawn before time t). After this stage, we successfully identified the networks for 2, 460, 502 mail servers out of the total 2, 461, 114 mail servers we observed. For the remaining 612 mail servers we determine the networks they belong to by matching them to the longest network prefix that we observe during the complete course of the BGP update collection, or by manually querying a *whois* server maintained by the Merit Networks [8].

Similarly, a network is classified as either *non-spam only*, *spam only*, or *mixed*, depending upon whether we receive spam messages from any mail servers belonging to the network. We refer to the set of spam only networks and mixed networks as the *spam networks*, and the set of non-spam only networks and mixed networks as the *non-spam networks*.

2.3 Terminology

The *spam ratio* of a message sender (either a mail server or a network) is the fraction of messages sent by the sender that is spam. Let t_1 and t_n be the times when we receive the first and the last messages from a mail server, respectively, then $t_n - t_1$ is referred to as the *active duration of the mail server*. The active duration of a network is similarly defined.

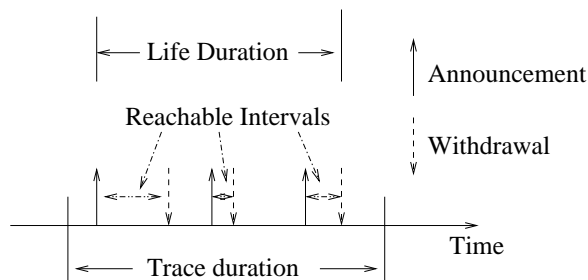


Figure 1: An example illustration of reachable intervals and life duration.

Now we define the notation of reachable intervals and life duration for a network prefix (see Figure 1 for an example illustration). Informally, a reachable interval of a network is a time interval in which the prefix is continuously announced, and the life duration of a network is the time interval within which we observe the BGP announcements of the corresponding network prefix. Let t_a be the time when we receive the first BGP announcement of a network prefix (following a BGP withdrawal of the prefix), and let t_w be the time when we receive the first withdrawal following the BGP announcement, then we refer to $t_w - t_a$ as a *reachable interval of the network*.

Consider an arbitrary network. If its corresponding prefix appears in the original BGP RIB table $rib(0)$, we let t_b be the time when $rib(0)$ was collected. Otherwise, let t_b be the time when a BGP announcement

of the prefix is first observed. Let t_e be the last BGP withdrawal message of the prefix without any BGP announcement messages of the same prefix following the withdrawal. If no such BGP withdrawal message exists, let t_e be the end of the time period of the BGP update collection. Then $t_e - t_b$ is referred to as the *life duration of the network* over the course of the two-month BGP trace collection.

Figure 1 shows the reachable intervals and life duration of an example network. The example network has three reachable intervals. In the example we have assumed that the prefix of the network does not appear in $rib(0)$. Therefore the life duration of the network is the time interval between the very first BGP announcement and the very last BGP withdrawal of the corresponding prefix during the course of the trace collection.

Ideally, any message arrival must occur within a reachable interval of the network to which the mail servers belong. However, due to separate locations at which the email and BGP traces are recorded, the two vantage points may have shifted views of the BGP update and spam arrivals. As a result, an email message may arrive between two neighboring reachable intervals. For such messages, we consider their arrival to be within the closest reachable interval of the corresponding networks.

3 Overview of the Email Trace

Table 1: Summary of the email trace (CV = Coefficient of Variation).

| Measure | Non-spam | Spam | Aggregate |
|-----------------------------|--|----------------|---------------|
| Period | 8/25/2005 – 10/24/2005 (excluding 9/11/2005) | | |
| Total # of emails | 6,712,392 | 18,537,364 | 25,249,756 |
| Total # of mail servers | 236,360 | 2,340,011 | 2,461,114 |
| Total # of networks | 39,158 | 61,888 | 68,732 |
| Avg # msgs/day (CV) | 110,039 (0.4) | 303,891 (0.17) | 413,930 (0.2) |
| Avg # mail servers/day (CV) | 14,191 (0.34) | 75,168 (0.13) | 86,664 (0.14) |
| Avg # networks/day (CV) | 5,730 (0.31) | 16,342 (0.1) | 19,340 (0.12) |

The email trace was collected between 8/25/2005 and 10/24/2005 (excluding 9/11/2005). The trace contains more than 25 M emails, of which more than 18 M, or about 73% of the messages, are spam (see Table 1). During the course of the trace collection, we observe more than 2 M mail servers, of which more than 95% send at least one spam message, and only about 10% of which sent at least one non-spam message. The messages come from 68,732 networks, of which more than 90% send at least one spam message.

In Table 2 we categorize email senders in more detail. As we can see from the table, the vast majority (more than 90%) of mail servers are spam only servers. They are responsible for 56.26% of all email messages and 76.6% of all spam messages. Only less than 5% of mail servers are non-spam only servers, and they are responsible for about 5% of all email messages (which are non-spam). About 5% of mail servers are mixed ones, and they are responsible for more than 38% of all email messages and 23.4% of all spam messages.

At the network level (Table 2), about 43% of networks we observe in the trace are spam only networks. However, only a small percentage of all email messages (6%) and spam messages (8%) come from such spam only networks. Moreover, only about 8% of mail servers belong to such networks. About 10% of networks are non-spam only ones; they are responsible for 0.56% of all messages and 0.39% of mail servers we observe. About 47% of networks are mixed ones, sending both spam and non-spam messages. They are responsible for a large portion of all messages (93.36%) and mail servers (91.55%) we observe. They also send a large fraction of spam (91.7% of all spam messages) and host a high percentage of spam only mail servers (91.1% of all spam only mail servers). These observations indicate that spammers most likely use

Table 2: Distribution of email senders.

| Group | Mail servers | | Networks | | |
|---------------|--------------|----------------------|----------|----------------------|------------------------------------|
| | % | % of Msg (% of spam) | % | % of Msg (% of spam) | % of Mail servers (% of spam only) |
| Non-spam only | 4.9 | 5.02 | 9.96 | 0.56 | 0.39 |
| Spam only | 90.4 | 56.26 (76.6) | 43.03 | 6.08 (8.3) | 8.06 (8.9) |
| Mixed | 4.7 | 38.71 (23.4) | 47.01 | 93.36 (91.7) | 91.55 (91.1) |

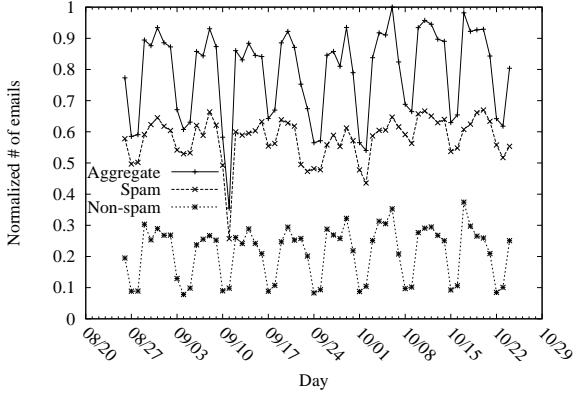


Figure 2: Normalized daily # of emails (max # of daily aggregate emails = 518, 145).

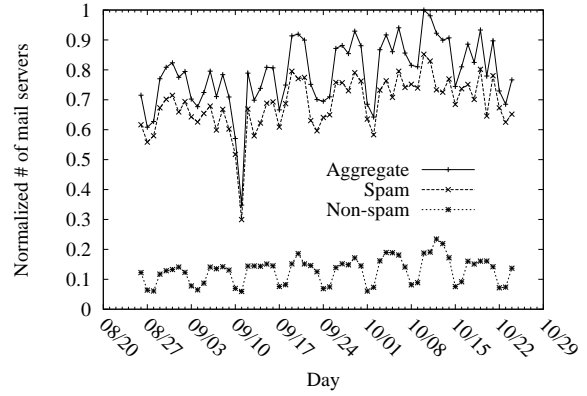


Figure 3: Normalized daily # of mail servers (max # of mail servers = 109, 933).

(compromised) machines within established networks to send spam instead of building their own networks. As a consequence, filtering spam at the native network level (as specified by network prefixes) would most likely also penalize legitimate email users.

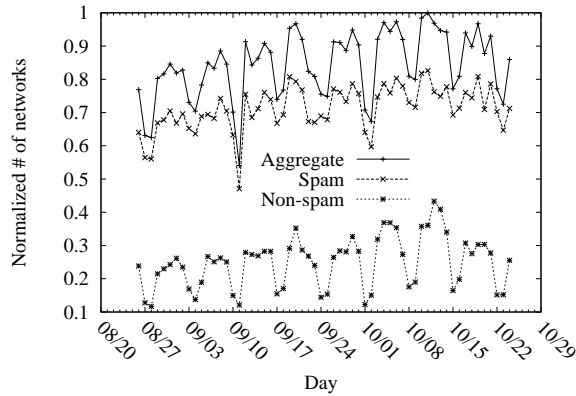


Figure 4: Normalized daily # of networks (max # of networks = 22, 968).

Figure 2 shows the daily arrivals of spam, non-spam, and aggregate emails (normalized against the maximum daily aggregate email arrivals 518, 145). As we can see from the figure, the arrivals of non-spam messages show a clear weekly pattern. This weekly arrival pattern is less evident for spam messages. Table 1 presents the average number of daily email arrivals and the coefficient of variation (CV) in parentheses. The smaller value of CV for the spam messages again indicates that spam messages arrive in a more constant manner compared to non-spam messages. Figures 3 and 4 show the daily numbers of mail servers and networks observed in the trace, respectively. We can see again a clear weekly pattern in the number of non-spam mail servers and networks observed. Note also that the daily numbers of spam mail servers and networks appear to be increasing. The notable deep dips in the three figures correspond to the date of

9/11/2005 (GMT time). This is caused by the fact that we do not have email trace on 9/11/2005 (local FSU time).

4 Behavioral Characteristics of Spammers

In this section we present a detailed study on the behavioral characteristics of spammers. In particular, we study the distribution of mail senders by spam ratios, the distribution of spam and non-spam messages by spam ratio of senders, the distributions of spam messages from different spammers, the spam arrival patterns across the IP address space, the number of mail servers in different spam networks, and the active duration of spammers, among others. We also briefly discuss the important implications of the findings for the anti-spam efforts, wherever it is proper. We discuss the implications of the findings in greater details in Section 6.

4.1 Distribution of Mail Senders by Spam Ratios

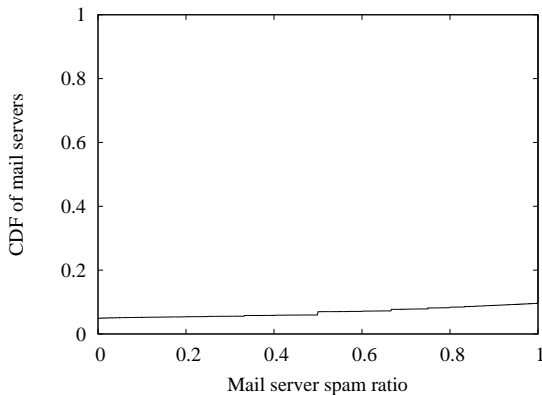


Figure 5: Distribution of mail server spam ratios.

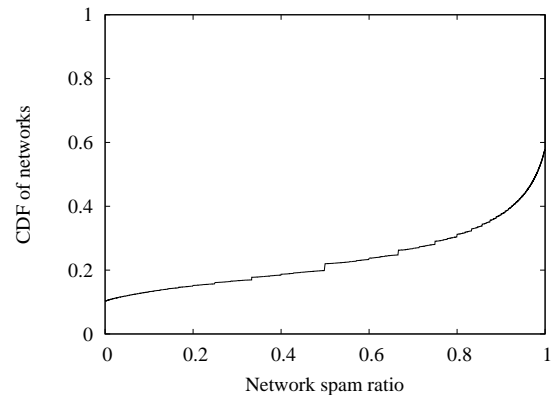


Figure 6: Distribution of network spam ratios.

In this section we study the spam ratio of mail senders and examine if spam ratio can be used as an effective indicator to distinguish legitimate senders from spam senders. Figure 5 shows the cumulative distribution function (CDF) of mail servers by spam ratios. From the figure we can see that most mail servers send either mostly spam messages or mostly non-spam messages. In particular, more than 90% of mail servers have a spam ratio that is at least 99%. About 5% of mail servers have a spam ratio that is at most 1%. It is critical to note that only less than 5% of mail servers have a spam ratio that is between 1% and 99%. Therefore, we could identify most legitimate mail servers and spam mail servers by examining the spam ratios associated with the mail servers. The mixed mail servers with a spam ratio between 1% and 99% are only a small portion of the total number of mail servers.

Figure 6 shows the corresponding distribution of networks by spam ratios. This figure shows a similar trend as in Figure 5, albeit with a quite different distribution for different spam ratios. In particular, only about 56% of networks have a spam ratio that is at least 99%, and more than 10% of networks have spam ratio that is at most 1%. A large fraction (about 43%) of networks have a spam ratio that is between 1% and 99%. Compared to the observations at the mail server level, it is less effective to identify legitimate networks and spam networks by examining the spam ratios associated with the networks. (However, we are still able to identify the majority of networks that are legitimate networks and spam networks.)

4.2 Distribution of Messages by Sender Spam Ratios

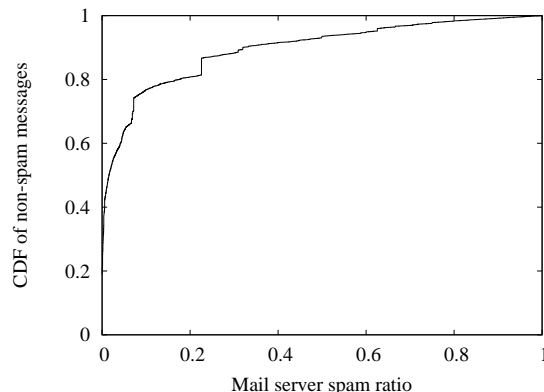
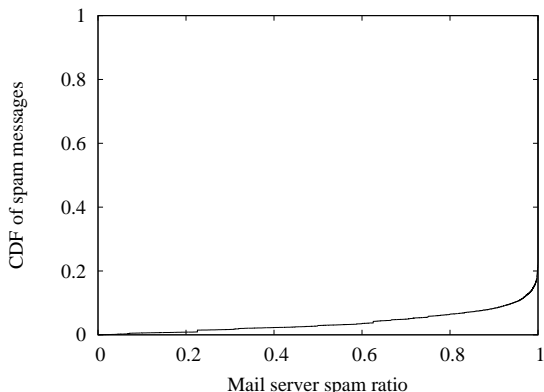


Figure 7: CDF of spam by mail server spam ratios. Figure 8: CDF of non-spam messages by mail server spam ratios.

In this section we study how spam and non-spam messages are contributed from senders with different spam ratios. Figure 7 shows the distribution of spam messages as a function of mail server spam ratios. As we can see from the figure, the vast majority of spam messages come from mail servers with a high spam ratio. In particular, 91% of spam messages come from mail servers with a spam ratio of at least 90%, and about 76% of spam messages come from mail servers with a 100% spam ratio (that is, they are spam only mail servers). The contribution of mail servers with a low spam ratio to the total spam is very small. For example, mail servers with a spam ratio of no higher than 50% only contribute less than 3% of total spam messages. This observation indicates that we can filter a large fraction of spam messages by blocking messages from the mail servers with a high spam ratio.

Figure 8 shows the distribution of non-spam messages as a function of the mail server spam ratios. As we can see from the figure, the distribution of non-spam messages is more diffused compared to the distribution of spam messages in Figure 7. In particular, mail servers with a spam ratio no more than 10% contribute about 76% of non-spam messages, and mail servers with a zero spam ratio (that is, they are non-spam only mail servers) contribute only about 19% of non-spam messages. A large portion (20%) of non-spam messages are contributed by mixed mail servers with a spam ratio between 10% and 90%. These observations indicate that spam ratio is a less effective metric to determine non-spam messages than to determine spam messages.

Figures 9 and 10 show the distributions of spam messages and non-spam messages as a function of spam ratios at the network level, respectively. In general we observe the similar trends at the network level as that at the mail server level (Figures 7 and 8), and we can draw a similar conclusion in controlling spam based on spam ratios of networks, as we have discussed above.

4.3 Number of Messages From Email Senders

Figure 11 shows the CDF of the number of messages from the observed mail servers. For the mixed servers, we show both the number of all messages with label *Mixed (all emails)* and the number of spam messages with label *Mixed (spam)*. Note first that about 50% of spam only mail servers send only a single message, and about 93% send no more than 10 messages over the two-month period. Second, about 28% of mixed mail servers send only a single *spam* message, and about 75% send no more than 10 *spam* messages. Combined, about 92% of all spam mail servers send no more than 10 spam messages and are responsible for about 26% of all spam messages we observe. It can be challenging, if not impossible, for reactive IP-address-based

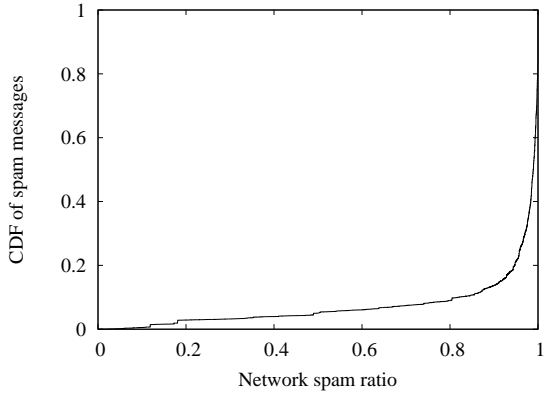


Figure 9: CDF of spam by network spam ratios.

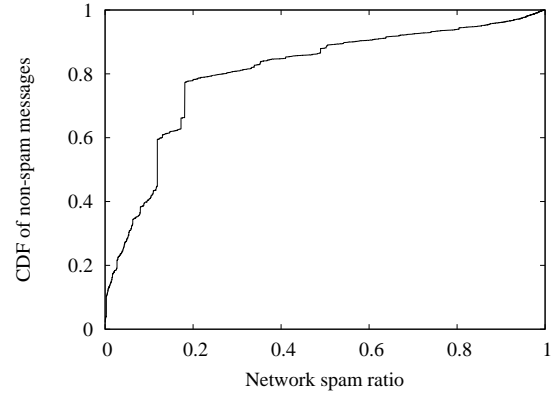


Figure 10: CDF of non-spam messages by network spam ratios.

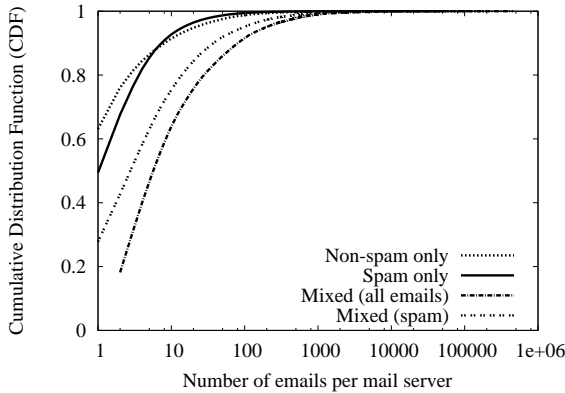


Figure 11: Number of emails from each mail server.

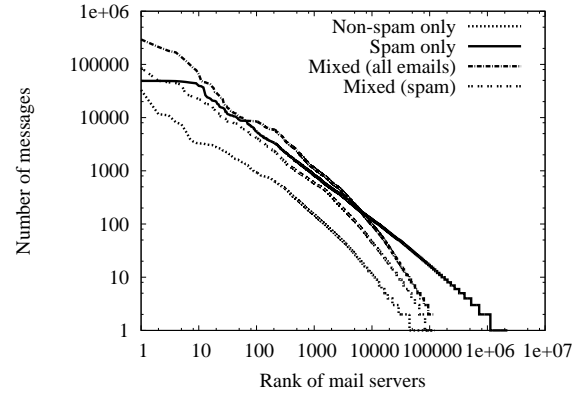


Figure 12: Rank of mail servers based on # of messages.

anti-spam schemes such as RBL [11] to combat such spam mail servers, given the small amount of spam sent by the majority of these mail servers.

From the figure we can see that non-spam mail servers also show a trend similar to the spam mail servers. For example, about 63% of non-spam only mail servers send only one message, and about 91% send no more than 10 messages. For mixed mail servers, the proportion of mail servers for sending one non-spam message and no more than 10 non-spam messages are 52% and 86% (not shown), respectively. Combined, about 89% of all non-spam mail servers send no more than 10 non-spam messages and are responsible for 6% of all non-spam messages we observe.

To have a better view of the tails of the distributions of email arrivals, in Figure 12 we rank the mail servers based on the number of messages they send. From the figure we can see that only a small number of mail servers generate a relatively large number of messages for all types of mail servers. In particular, only 814, or about 0.04%, of spam only mail servers send more than 1,000 messages each, and they are responsible for 16% of all spam messages we observe. Out of 115,257 mixed mail servers, only 482 servers send more than 1,000 spam messages each, and they are responsible for 10% of all spam messages we observe. These observations suggest that filtering aggressive spam mail servers can considerably reduce the amount of spam on the Internet. However, we must note that it is generally infeasible to filter mixed mail servers that send both spam and legitimate emails.

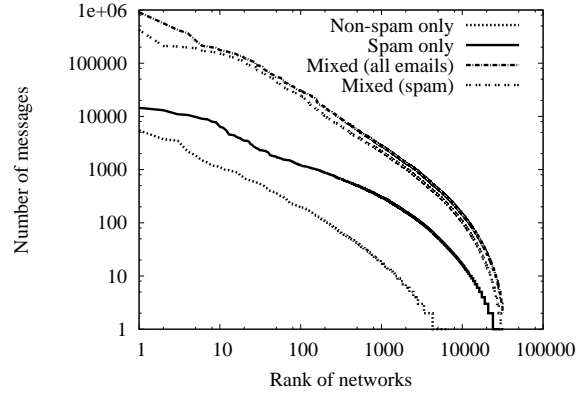
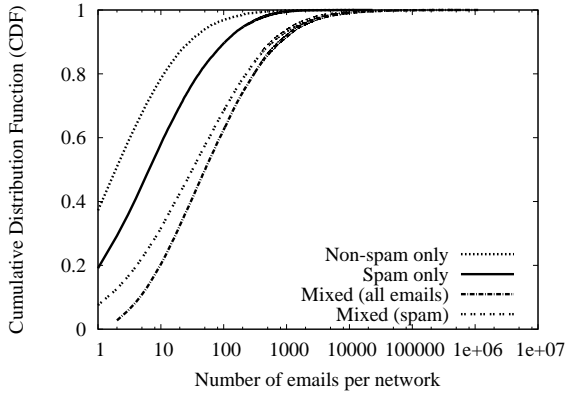


Figure 13: Number of emails from each network. Figure 14: Rank of networks based on # of messages.

Figures 13 and 14 present the CDF of the number of messages from the observed networks and the ranks of the networks according to the number of the messages they send, respectively. In principle, these two figures show the similar trends as Figures 11 and 12, respectively. In particular, about 58% of spam only networks send no more than 10 messages each and are only responsible for 0.3% of all spam messages we observe. About 32% of mixed networks send no more than 10 spam messages each and are responsible for 0.2% of all spam messages we observe. Combined, about 44% of spam networks send no more than 10 spam messages each and are responsible for 0.5% of all spam messages we observe.

Out of 29,574 spam only networks, 158, or 0.5%, send more than 1,000 messages each and are responsible for 2% of all spam messages we observe. 2,103 out of 32,314, or 6.5% of mixed networks send more than 1,000 spam messages each and are responsible for 75% of all spam messages we observe. These results show that the majority of spam messages are sent from mixed networks that generate both spam and non-spam messages. This presents significant challenges in filtering spam at the native network prefix level as announced by the BGP updates.

4.4 IP Address Origins of Spam Messages

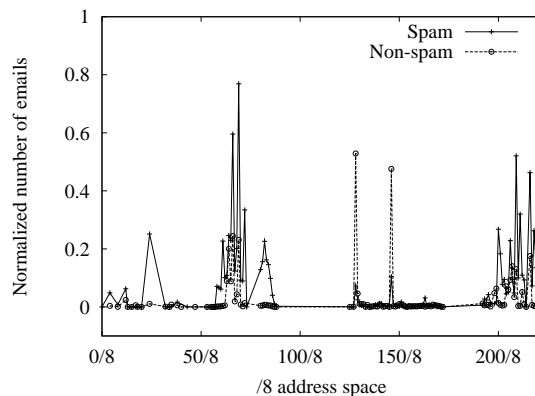


Figure 15: Normalized number of messages from each “/8” address space (max # of aggregate emails = 2,077,655).

In order to study the origins of spam messages across the IP address space, we classify messages into each “/8” address space depending on the mail servers that delivered the messages. In Figure 15, we plot

the number of spam and non-spam messages that originate from each “/8” address space, normalized by the maximum number of aggregate emails (2,077,655) generated by one of the “/8” address spaces. Note first that spam messages originate from a few concentrated “/8” address spaces. For example, the top 20 “/8” address spaces originate 68% of all spam messages we observe, and the top 40 originate more than 91% of all spam messages. The top “20” “/8” address spaces are, 69/8, 66/8, 209/8, 216/8, 72/8, 211/8, 200/8, 218/8, 24/8, 64/8, 65/8, 206/8, 61/8, 82/8, 222/8, 68/8, 221/8, 201/8, 220/8, and 83/8, ordered according to the number of spam messages they originate. Note that the address space of 24/8 is the “cable block,” used by the companies that provide Internet access via cable systems [5]. This “/8” address space sends about 2.8% all spam messages we observe. We do not observe spam messages from any of the “/8” address spaces reserved by the Internet Assigned Number Authority [5]. This is possibly because ISPs filter non-routable network prefixes on the Internet including private RFC 1918 address blocks and unassigned address prefixes.

The origins of non-spam messages are also highly concentrated. For example, the top 20 and 40 “/8” address spaces send 88% and 94% of all the non-spam messages, respectively. Moreover, the top 20 “/8” address spaces of spam messages and the top 20 “/8” address spaces of non-spam messages share 8 common “/8” address spaces, namely 64/8-69/8 (excluding 67/8), 206/8, 209/8, and 216/8. It is worth noting that internal servers at Florida State University contribute to the two notable spikes (corresponding to 128/8 and 146/8) of the non-spam message curve. These two “/8” address spaces are responsible for about 31% of all non-spam messages we observe.

4.5 Number of Mail Servers and Their Origins

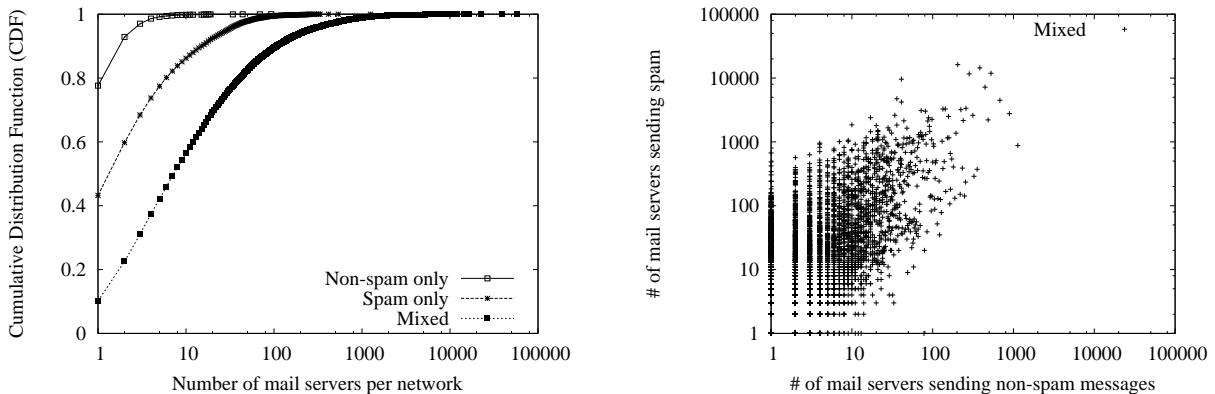


Figure 16: Number of mail servers of each network. Figure 17: Spam vs. non-spam mail servers (mixed networks).

Figure 16 presents the CDF of the number of mail servers in the individual networks we observe. We only observe a single mail server for about 78% of non-spam only networks, and no more than 10 mail servers for over 99% of non-spam only networks. The maximum number of mail servers we observe in a non-spam only network is 93. In contrast, we generally observe more mail servers in spam only networks. For example, only about 43% of spam only networks have one mail servers, and about 86% no more than 10 mail servers. The maximum number of mail servers we observe in a spam only network is 1,249.

Note from the figure that a larger portion of mixed networks have a large number of mail servers compared to both spam only and non-spam only networks. This can be understood by noting that mixed networks normally have both legitimate and spam mail servers. In particular, only about 10% of mixed networks have a single mail server (which send both spam and non-spam messages), and about 56% no more than 10 mail servers. More than 10% mixed networks have more than 100 mail servers, and more than 1% have more

than 1,000 mail servers. The maximum number of mail servers we observe in a mixed network is 57,106.

Next we examine if the network size in terms of number of mail servers affects the spamming behavior of the corresponding networks. Figure 17 shows the correlation between the number of non-spam mail servers and spam mail servers in the mixed networks. In the figure, each point represents a network, and the corresponding x-axis value shows the number of non-spam mail servers and y-axis value the number of spam mail servers. From the figure we can see that the spamming behavior in terms of number of mail servers sending spam messages appears to be independent of the network size. Almost regardless of the network size, there is a large number of mail servers sending spam for most of the network domains. In addition, the majority (68%) of mixed networks have more spam mail servers than non-spam mail servers.

A large portion of mail servers in the mixed networks are likely to be infected machines. It again suggests that filtering spam at the native network level as announced by BGP updates may not be feasible. It also indicates that sender authentication schemes such as sender policy framework [20] are in urgent need and can be effective in combating such spam. Note that although spammers can easily turn an infected machine into a spam mail server, it is much harder for them to fake it as a *legitimate* mail server.

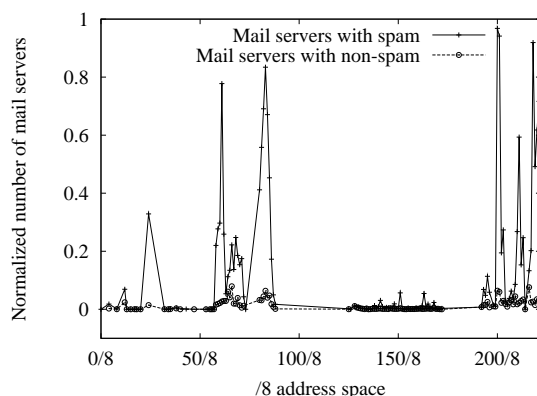


Figure 18: Normalized # of mail servers of each “/8” address space (max # of mail servers = 147,130).

In Figure 18 we classify each mail server into the corresponding “/8” address space and show the number of the mail servers in each address space (normalized by the maximum number of mail servers (147,130) observed in one of the “/8” address spaces). Similar to our observation in Figure 15 for the origins of spam messages, the spam mail servers are also from a few concentrated “/8” address spaces. In particular, the top 20 “/8” address spaces are responsible for about 74% of all the spam mail servers, and the top 40 are responsible for about 94% of all spam mail servers. In contrast, the non-spam mail servers are less concentrated in the IP address space. For example, the top 20 “/8” address spaces are only responsible for about 57% of all the non-spam mail servers, and the top 40 are only responsible for about 84% of all non-spam mail servers.

The top 40 “/8” address spaces of spam mail servers and the top 40 “/8” address spaces of spam messages largely overlap; they share 34 common “/8” address spaces. Spam only mail servers often employ neighboring IP addresses or IP addresses within close proximity, such as “128.121.31.104-128.121.31.114” and “128.121.31.143-128.121.31.152”. These observations suggest that spam filters should take into account the following factors in identifying spam messages or spam mail servers: the number of mail servers deployed in a network and the IP address pattern of these mail servers.

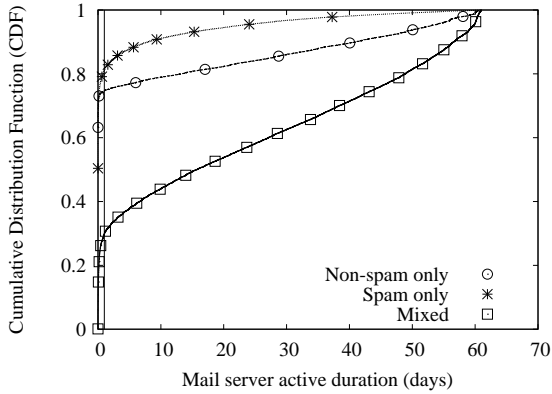


Figure 19: Mail server active duration (days).

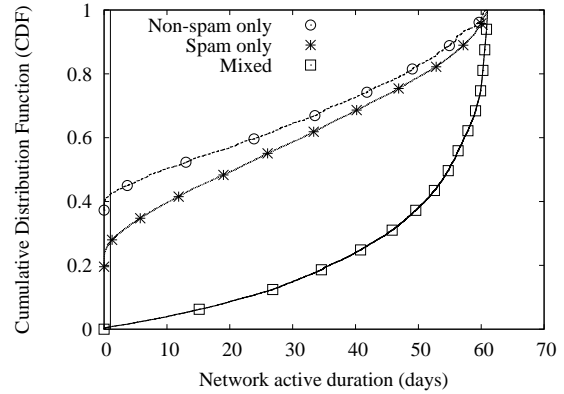


Figure 20: Network active duration (days).

4.6 Active Duration of Spammers

Recall that the active duration of a sender is defined as the time interval between the first message and the last message we observe from the sender. From Figure 19 we see that about 81% of all spam only mail servers send spam messages only within one day (the vertical line). This short active duration of the vast majority of spam only mail servers again makes it challenging for IP-address-based spam filters such as RBL to work effectively, if feasible at all. It is critical to note that a large majority (75%) of non-spam only mail servers are also active only within one day. Therefore, the length of active duration of mail servers is not a reliable indicator for distinguishing spam mail servers from non-spam mail servers. In comparison, only about 30% of mixed mail servers are active only within one day. This can be caused by the fact that both spammers and legitimate email users send messages from such mixed mail servers.

Figure 20 shows the active duration of sender networks. Again, there is a large portion of spam only and non-spam only networks that are only active within a short period of time. In particular, the proportions of spam only networks and non-spam only networks that are active only within one day are 27%, 42%, respectively. The mixed networks in general have longer active duration. For example, more than 85% of mixed networks are active for more than 30 days, and only about 0.8% are active only within one day. Again, this can be caused by the fact that both spammers and legitimate email users send messages from such mixed networks.

4.7 Network Prefix Length and Network Types

In this section we investigate how spam senders are distributed in terms of the types of the networks. In particular, we study the network prefix length of various spam and non-spam network domains, and where spam and non-spam senders are located in the Internet.

Figure 21 shows the prefix length of the networks observed. The dominant prefix lengths for non-spam only networks are 24 or 16, representing about 65% of all non-spam only networks. For spam only networks, the general trend is that the longer the prefix, the larger portion of networks have this prefix. For example, the top 6 prefix lengths (24 - 19) represent about 93% of all spam only networks. In particular, about 23% of all spam only networks have a network prefix length of 24. Similarly, about 23% of mixed networks have a network prefix length of 24.

Next we investigate the locations of the spam and non-spam senders. In particular, we would like to know if they are more likely located at the core of the Internet or at the edge of the Internet. For this purpose, we classify a network into either a *stub* network or a *transit* network. Informally, a stub network is on the edge of the Internet, while a transit network is on the core of the Internet. A network is considered

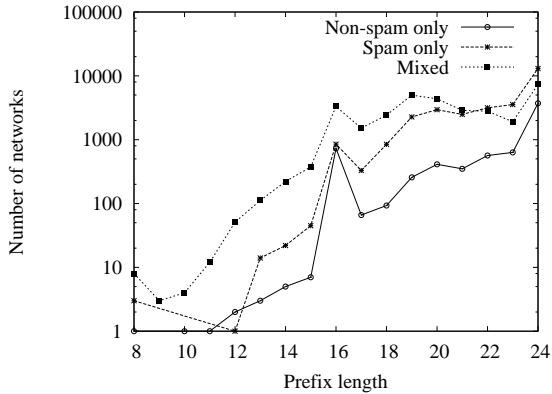


Figure 21: Network prefix length.

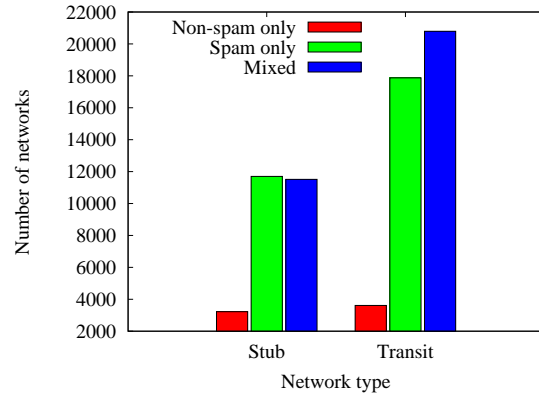


Figure 22: Network types.

as a stub network if its origin ASes only appear as the first (rightmost) AS in the AS paths of the BGP announcements of the prefix [16]. Otherwise, it is considered as a transit network. Figure 22 shows the network types (stub or transit). The proportion of stub networks in the non-spam only networks is 47%, and for spam only networks and mixed networks, the proportions are 40% and 36%, respectively. The proportion of stub networks in the non-spam only networks is only slightly higher than the spam only and mixed networks. (But note that only about 10% networks observed are non-spam only network.) From this observation we can see that both stub and transit networks are likely to host spam and non-spam senders, although non-spam only senders have a higher probability to be located in stub networks.

5 Network Reachability Properties of Spammers

An important objective of this section is to verify an informal observation by Paul Vixie that the spam arrivals from some spammers are often closely correlated in time with the BGP announcement of the corresponding network prefixes [19]. These network prefixes are short-lived in that they are withdrawn after the spamming activity is finished. This technique makes it hard to identify the spammers that are responsible for spamming. In this section we formally confirm this behavior and investigate the prevalence of this behavior. Our major finding is that the network prefixes of a non-negligible portion of spam only networks are only visible within a short period of time during the complete course of the email trace. In particular, about 6% of all spam only networks have life duration of no longer than one week out of the two-month trace collection period. The life duration of these prefixes coincides with the arrivals of spam from the networks. Note that this observation is different from the one reported in [10], which focused on network prefixes with *persistent short reachable intervals* (i.e., persistent short-lived BGP announcements [10]), which may have a life duration spanning the complete course of the email trace.

In Section 5.1, we first study how network prefixes with *short life duration* are used by spammers and investigate the prevalence of this spamming technique. In Sections 5.2 and 5.3, we next investigate patterns in spam messages from network prefixes with *persistent short reachable intervals* (i.e., persistent short-lived BGP announcements) as first reported in [10] and compare their observations with ours. We discuss the two different spamming patterns at the end of this section.

5.1 Network Life Duration

In the following we study the life duration of networks (Section 2). Figure 23 shows the CDF of the life duration of all networks observed. From the figure we can first see that most of networks are reachable

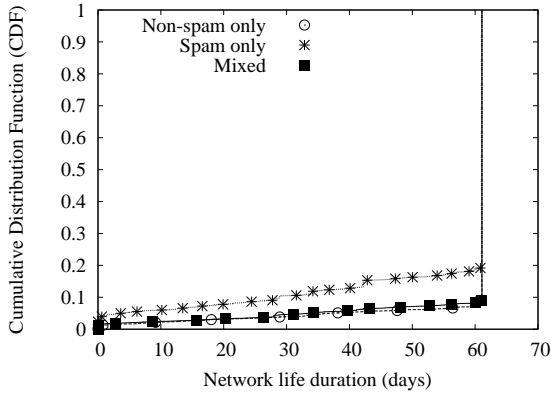


Figure 23: Network life duration.

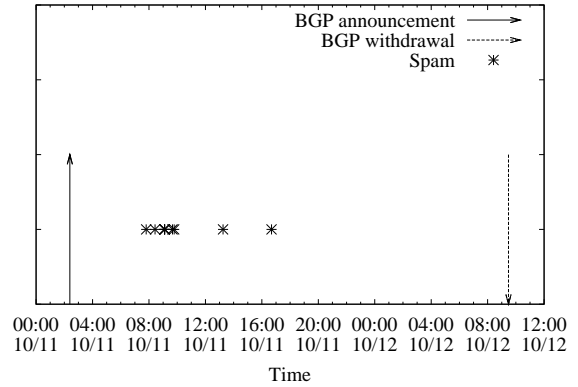


Figure 24: Spam network with short life duration (222.46.32.0/20).

during the whole time period of the email trace collection. We can also see from the figure that a larger portion of spam only networks have a shorter life duration compared to non-spam only and mixed networks. In particular, more than 4% of spam only networks have life duration less than one day. In contrast, only about 1% of non-spam only networks and 1.6% of mixed networks have this short life duration. (Note also that only about 10% of all networks observed are non-spam only networks, see Table 2.) Moreover, about 6% of spam only networks have life duration no longer than one week, and the corresponding proportion of non-spam only networks and mixed networks are both about 2%.

The short life duration of spam only networks coincides with the spamming activity of the spammers. Figure 24 shows a typical example of the correlation between the BGP updates and spam arrivals from the network 222.46.32.0/20. The network prefix was announced around 2:00AM on 10/11/2005 (FSU local time) and we never saw the BGP updates of the prefix before. Spam messages arrived around 7:00AM on that day and continued till around 5:00PM on the same day. The prefix was subsequently withdrawn around 9:00AM next day (10/12/2005) and we never saw the BGP updates of the prefix thereafter. It is worth noting that we observe the spamming activity of the spammer only from a single vantage point (FSU). It is highly possible that during the life duration of the network, the spammer may have also sent spam messages to other Internet users.

There is no clear understanding of how long the spammers using the technique of short-lived networks keep their network prefixes announced. If we take one week (one day) as the threshold, then from Figure 23 we can see that about 6% (4%) of network-level spammers use this technique for spamming.

5.2 Network Reachable Intervals

As defined in Section 2, the reachable interval of a network is the time between the first BGP announcement of the prefix and the first BGP withdrawal of the same prefix after the announcement. In other words, a reachable interval of a network is a time interval that the corresponding prefix is continuously visible.

Figure 25 shows the CDF of the reachable intervals (the vertical dash line corresponds to the time of one day). It is interesting to note that a higher percentage of reachable intervals of non-spam only networks are short compared to the reachable intervals of both spam only networks and mixed networks. This can be caused by a few factors, for example, a few non-spam only networks are extremely unstable and originate a large number of short reachable intervals. This figure indicates that in general we cannot distinguish spam networks from non-spam networks simply based on the stability of the BGP routes of the networks.

The reachable intervals of spam networks are often in concert to the arrivals of spam messages, as

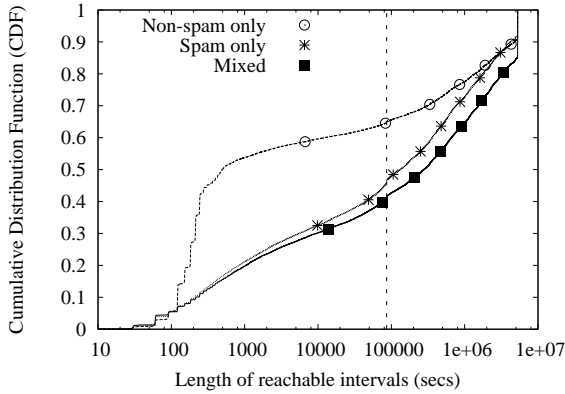


Figure 25: Network reachable intervals.

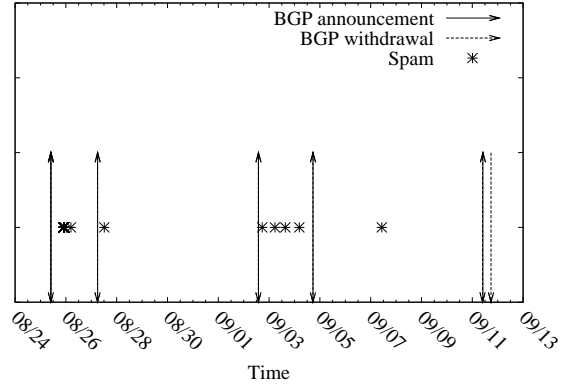


Figure 26: Spamming correlated with BGP updates (62.215.96.0/19).

shown in Figure 26. In the figure, spam arrivals from mail servers in the network 62.215.96.0/19 follow the announcement of the prefix for some of the reachable intervals. However, in general, it is hard to infer if this is done intentionally. This can occur simply because any messages (spam or not) must arrive during some reachable intervals of the corresponding prefixes.

5.3 Active Network Reachable Intervals

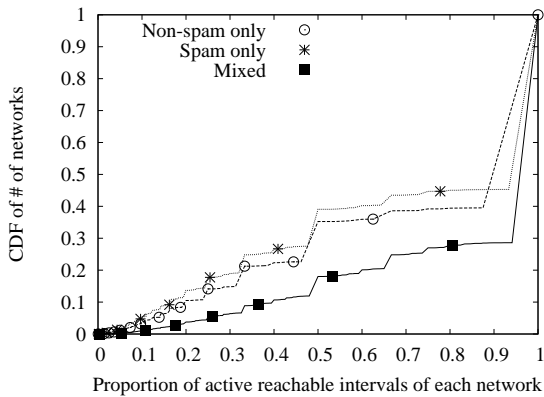


Figure 27: Proportions of intervals with emails.

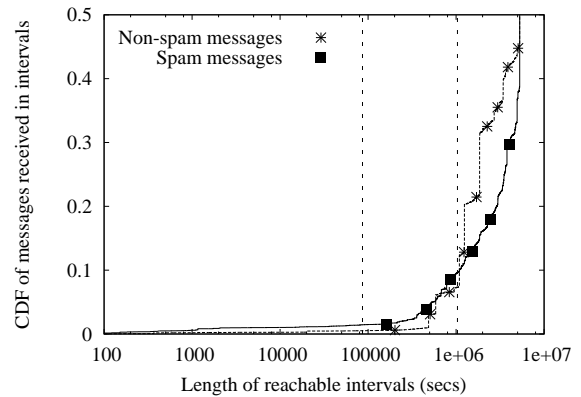


Figure 28: Email arrivals in different network reachable intervals.

A network reachable interval is active when it coincides with the arrivals of messages. We define the proportion of active reachable intervals of a network as the ratio of active reachable intervals to all the reachable intervals of the prefix. Figure 27 shows the CDF of the proportion of active reachable intervals of the networks. From the figure we can see that the reachable intervals of mixed networks are better utilized than spam only networks and non-spam only networks—a higher portion of reachable intervals of mixed networks coincide with the arrivals of (spam or non-spam) messages. For example, more than 80% mixed networks send messages in more than 50% of their reachable intervals. Only about 65% non-spam only networks and 60% spam only networks send messages in more than 50% of their reachable intervals. This can be explained by the fact that the vast majority, i.e., more than 93% of all messages are sent from mixed networks (see Table 2). The lower proportion of active reachable intervals of spam only networks again

suggests that the coincidence of spam arrivals and reachable intervals of most spam only networks may not be intentional. Otherwise, we would expect a higher proportion of active reachable intervals from these networks.

In order to study the arrival patterns of spam messages and non-spam messages, in Figure 28, we plot the CDF of the number of messages arriving in different reachable intervals. The two dash lines in the figure correspond to one day and 12 days. From the figure we see that a higher portion of spam messages arrive within shorter reachable intervals (shorter than 12 days) than non-spam messages. In particular, about 1.5% of spam messages and 0.5% of non-spam messages arrive within reachable intervals shorter than a day. However, when we consider the reachable intervals longer than about 12 days (the right dash line), a higher portion of non-spam messages arrive within shorter reachable intervals than spam messages.

In summary, when considering *life duration* of network prefixes, we found that network prefixes for a non-negligible portion of spam only networks are only visible within a short life duration (Figure 23). For example, during the two-month trace collection period, the network prefixes of about 6% of spam-only networks have life duration less than one week. This technique makes it hard to identify the spammers that are responsible for spamming.

On the other hand, when considering *reachable intervals* of network prefixes, we found that, in general, it is difficult to conclusively infer whether or not spammers intentionally use short reachable intervals to hide their spamming activity. In particular, a higher percentage of reachable intervals of non-spam only networks are actually shorter compared to the reachable intervals of spam only networks (Figure 25). Additionally, spam only and non spam only networks have similarly low proportion of *active* reachable intervals (Figure 27).

These collective observations from our email and BGP traces suggest the following: spammers are more likely to rely upon the technique of using network prefixes with *short life duration* rather than network prefixes with *persistent short reachable intervals* (or persistent short-lived BGP announcements [10]). Note however, that these observations are specific to our two month long email and BGP traces, and it would be worthwhile to independently confirm these findings from other sources.

6 Implications and Discussion

In this section we discuss the implications of our findings for the (content-independent) anti-spam efforts. In particular, we examine the implications of the findings on email sender authentication, sender reputation based spam filtering techniques, and new email delivery architectures.

6.1 Implications on Email Sender Authentication

Nowadays, the majority of spam messages are originated from compromised machines (or spamming bots), and our study confirmed that a large number of observed “mail servers” are compromised machines. Note that most these compromised machines are likely to be end-user machines, instead of real server machines, given that server machines are normally better protected. This observation indicates that it is critical for a recipient network to know if a machine in a remote network is a legitimate mail server when an incoming message delivery request comes. There have been a number of research and development efforts following this direction. The most notable approach is the formal sender authentication protocols such as the Sender Policy Framework (SPF) [20]. Our study shows that such sender authentication systems will play a critical role in blocking spam messages. Note that although spammers can easily turn an infected end-user machine into a spam mail server, it is much harder for them to fake it as a *legitimate* mail server for a particular network domain.

Another somewhat ad-hoc approach in identifying end-user IP addresses that should not behave as a

mail server is the Policy Block List (PBL) maintained by the Spamhaus project [15]. PBL is a database of end-user IP addresses that are maintained by both the participating network domains and the Spamhaus team. PBL contains various (end-user) IP addresses that should not initiate a message delivery request to a remote network domain, for example, dynamic IP addresses. However, PBL is far from complete based on our recent experience with the database. A possible reason is that the voluntary participation of network domains is still quite limited. The third approach to distinguish mail servers from end-user machines is to extract and model the distinct features of the two classes of machines, as we have discussed in [13]. Currently we are developing and evaluating a machine learning based approach on classifying legitimate mail servers and end-user machines. We will report our results in a separate upcoming paper. It is important to note that all the approaches that we have discussed in this subsection will help to proactively block spam messages delivered from illegal “mail servers”, based on the nature of the sending machines. They do not rely on the spamming history of the sending machines. Filtering mechanisms based on the spamming history of sending machines will be discussed in the next subsection.

6.2 Implications on Sender Reputation based Anti-Spam Schemes

In this section we discuss the implications of our findings on the sender reputation based anti-spam schemes, that is, the schemes that block spam messages based on the spamming history of the sender machines. The implications of our findings on these schemes are multi-fold. First, our study shows that the majority of sending machines send either mostly spam messages or mostly non-spam messages, which indicates that spam ratio can be an effective and reliable metric to distinguish legitimate mail servers from spam sending machines. Moreover, blocking mail servers with a high spam ratio can effectively filter the vast majority of spam messages (with a small false positive rate of blocking some non-spam messages, depending on the specified spam ratio threshold).

On the other hand, our study also shows the majority of spammers only send a small number of spam messages and are only active for a short period of time. This suggests that most network domains with a limited view of the behavior of the sending machines may not be able to collect enough spamming history and respond quickly enough to flag most of the spamming machines. Before such network domains can identify these spamming machines, they may have finished the corresponding spam campaign. This possibly helps explain why sender reputation based systems have been mostly deployed and effective at large email service providers such as Gmail [17], but not much so at other network domains. Large email service providers are at a strategic position to have a greater view of the spamming behavior of sending machines, given the large volume of email messages they receive each day. One potential approach to overcoming this problem is through network collaboration as used in anti-spam schemes such as CARE [21], where multiple network domains exchange the sender spamming history. This helps provide a greater view of the sender behavior and helps to quickly flag a spammer.

Another approach to addressing this problem is to aggregate the spamming history at the sending network domain level instead of considering only at the individual mail server level. Our study shows that although many individual spamming machines only send a small number of spam messages and are short-lived, the corresponding networks are normally much longer lived and send much higher volume of spam messages, which provides a wider time window and more evidence for the receiving network domains to respond. However, extra care must be given to this approach when it is adopted. First, given that the vast majority of spam messages and spam mail servers are from mixed networks that send both spam and legitimate messages, it can be challenging to filter spam purely based on the network prefix information of the corresponding sender network domain. Second, our study shows that a large percentage of network domains are mixed networks. For example, about 43% of networks have a spam ratio that is between 1% and 99%. We must take a very conservative spam ratio threshold in identifying spamming network domains. Another potential use of the network domain spam ratio is to feed this information as one factor into a comprehensive

(content-dependent) spam filter to determine if an incoming message is likely to be spam (instead of directly blocking messages based on the information of the sending network domain) [4].

6.3 Implications on New Email Delivery Architectures

Our findings on the behavioral characteristics of spammers and their network reachability properties also have important implications for the design of new email delivery architectures. One of our key findings is that the majority of spammers are only active for a short period of time. In the current email system, spammers can push out a deluge of spam within a very short period of time, making spamming profitable because of the economy of scale. Moreover, the current email system makes it hard to hold spammers accountable for spamming. Spammers can vanish (go offline) immediately after pushing spam to receivers. This makes it quick and easy for spammers to hide their identities and provides spammers with the flexibility to frequently change their locations and/or Internet service providers. In addition, our study also shows that sophisticated techniques such as network prefix hijacking have been employed by spammers, which can further help spammers to hide their identities.

All these observations suggest that in future spam-resistant email delivery architectures, it is important to force spammers to stay online for longer periods of time while throttling their spam delivery rates and to remove spammers' flexibility in frequently changing their locations and/or Internet Service Providers. New email delivery architectures such as Differentiated Mail Transfer Protocol (DMTP) [1] have made progress in incorporating these design lessons. Our confirmation that sophisticated techniques such as network prefix hijacking have been used in spamming also highlight the urgency in improving the overall security of the Internet architecture, in particular, the inter-domain routing system, which is beyond the scope of the current paper and we do not further elaborate (but see [6, 12] and references therein).

7 Related Work

The work [2] studied the characteristics of spam traffic aiming to identify the features that can distinguish spam from legitimate messages. They found that key email workload aspects including the email arrival process, email size distribution, and distributions of popularity and temporal locality of email recipients can distinguish spam from legitimate messages. They also discussed the inherently different natures of spammers and legitimate email users that contribute to the distinct features of spam traffic. However, this work did not study the behavioral characteristics of spammers at the mail server level or at the network levels as we do in this paper.

As part of the effort to investigate the feasibility of mitigating the spam problem by prioritizing the delivery of legitimate messages, the authors of [18] studied the effectiveness of network-aware clustering as an aggregation scheme for exploiting the history and structure of sending machines. Our study confirmed some of the findings reported in their paper. Using a 4-year spam archive, the authors of [7] studied the spamming activity across the IP address space. They found that spamming bots are more spread across the IP address space in recent years. We provided a more comprehensive study of the behavioral characteristics of spammers and we discussed the implications of the findings in a broader term compared to these two studies.

An independent and parallel work [10] also studied some aspects of the network-level behavior of spammers. While their work shares some of our objectives, there are a number of significant differences. First, while the study in [10] only focused on the network-level behavior of spammers, we study the behavioral characteristics of spammers at both the network level and the mail server level. Second, [10] identified a spamming pattern where spam arrivals coincided with *persistent short-lived BGP route announcements* of the corresponding network prefixes. The BGP announcements of these prefixes may span the complete

course of the spam trace collection. In contrast, we reveal another important spamming pattern where *the prefixes of spam networks are short-lived*; such network prefixes are visible only *briefly during the complete course* of our data collection period, coinciding with the arrivals of spam from the corresponding networks. These two studies confirm and complement each others' findings from different vantage points.

8 Conclusion

In this paper we studied the behavioral characteristics of spammers at both the mail server and network levels using a two month email trace collected on the FSU campus network. We focused on the behavioral characteristics that have important implications for spam control, including the distributions of mail servers, spam and non-spam messages by spam ratios; the statistics of spam messages from different spammers; the spam arrival patterns across the IP address space; and the active duration of spammers, among others. In addition, we also investigated the network reachability properties of spammers as indicated by their BGP routing dynamics, and formally confirmed an informal observation that spammers may hijack network prefixes in sending spam messages. In addition to presenting the detailed findings of the empirical study, we also discussed the implications of the findings for anti-spam efforts, including the implications on sender authentication schemes, sender reputation based anti-spam schemes, and the design of new email delivery architectures. We are currently developing effective content-independent spam filtering schemes based on the insights obtained from this empirical study of spammer behavioral characteristics, and will report the results in an upcoming paper.

Acknowledgments

We thank Paul Vixie for the informative discussion that motivated this study. We thank Mary Stephenson and Arthur Houle for helping collect the email trace and the University of Oregon Route View Project for making the BGP trace publicly available. Zhenhai Duan was supported in part by NSF grants CCF-0541096 and CNS-1041677. Kartik Gopalan was supported in part by NSF Grant CCF-0541096. Xin Yuan was supported in part by NSF Grants ANI-0106706, CCR-0208892, CCF-0342540, and CCF-0541096. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of National Science Foundation.

References

- [1] Z. Duan, Y. Dong, and K. Gopalan. DMTP: Controlling spam through message delivery differentiation. *Computer Networks (Elsevier)*, 51:2616–2630, July 2007.
- [2] L. Gomes, C. Cazita, J. Almeida, V. Almeida, and W. Meira. Characterizing a spam traffic. In *Proceedings of IMC'04*, pages 356–369, Taormina, Sicily, Italy, Oct. 2004.
- [3] S. Halabi and D. McPherson. *Internet Routing Architectures*. Cisco Press, 2 edition, 2000.
- [4] S. Hao, N. A. Syed, N. Feamster, E. G. Gray, and S. Krasser. Detecting Spammers with SNARE: Spatio-temporal Network-level Automatic Reputation Engine. In *USENIX Security Symposium*, pages 101–118, Montreal, Canada, Aug. 2009.
- [5] IANA. Internet assigned numbers authority. <http://www.iana.net/>.
- [6] S. Kent, C. Lynn, and K. Seo. Secure border gateway protocol (S-BGP). *IEEE Journal on Selected Areas in Communications*, 18(4):582–592, 2000.

- [7] M. Kokkodis and M. Faloutsos. Spamming botnets: Are we losing the war? In *Proceedings of 6th Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, July 2009.
- [8] Merit Network Inc. Merit network routing assets database. whois.radb.net.
- [9] U. of Oregon. Route Views project. <http://www.routeviews.org/>.
- [10] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of ACM SIGCOMM*, pages 291–302, Pisa, Italy, Sept. 2006.
- [11] RBL. Real-time spam black lists (RBL). <http://www.email-policy.com/Spam-black-lists.htm>.
- [12] F. Sanchez and Z. Duan. Region-based bgp announcement filtering for improved bgp security. In *Proceedings of 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pages 89–100, Beijing, China, Apr. 2010.
- [13] F. Sanchez and Z. Duan. Understanding forgery properties of spam delivery paths. In *Proceedings of 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS)*, Redmond, WA, July 2010.
- [14] SpamAssassin. The Apache SpamAssassin project. <http://spamassassin.apache.org/>.
- [15] Spamhaus. The spamhaus project. <http://www.spamhaus.org/>.
- [16] J. Stewart. *BGP4: Inter-Domain Routing In the Internet*. Addison-Wesley, 1999.
- [17] B. Taylor. Sender reputation in a large webmail service. In *Proceedings of Third Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, July 2006.
- [18] S. Venkataraman, S. Sen, O. Spatscheck, P. Haffner, and D. Song. Exploiting network structure for proactive spam mitigation. In *USENIX Security Symposium*, pages 149–166, Boston, MA, Aug. 2007.
- [19] P. Vixie. Private communication. USENIX/SRUTI, July 2005.
- [20] M. Wong and W. Schlitt. Sender policy framework (spf): Authorizing use of domains in e-mail, version 1. RFC 4408, Apr. 2006.
- [21] M. Xie and H. Wang. A collaboration-based autonomous reputation system for email services. In *Proceedings of IEEE INFOCOM*, pages 992–1000, San Diego, CA, Mar. 2010.