Behavioral Characteristics of Spammers and Their Network Reachability Properties

Zhenhai Duan
Florida State University
duan@cs.fsu.edu

Kartik Gopalan
Binghamton University
kartik@cs.binghamton.edu

Xin Yuan
Florida State University
xyuan@cs.fsu.edu

***Abstract*—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 messages, we characterize the spammer behavior at both the mail server and the network levels. We also correlate the arrivals of spam with the BGP route updates to study the network reachability properties of spammers. Among others, our significant findings are: (a) the majority of spammers (93% of spam only mail servers and 58% of spam only networks) send only a small number of spam messages (no more than 10); (b) the vast majority of both spam messages (91.7%) and spam only mail servers (91%) are from mixed networks that send both spam and non-spam messages; (c) the majority of both spam messages (68%) and spam mail servers (74%) are from a few regions of the IP address space (top 20 "/8" address spaces); (d) a large portion of spammers (81% of spam only mail servers and 27% of spam only networks) send spam only within a short period of time (no longer than one day out of the two months); and (e) network prefixes for a non-negligible portion of spam only networks (6%) are only visible for a short period of time (within 7 days), coinciding with the spam arrivals from these networks. In this paper, in addition to presenting the detailed results of the measurement study, we also discuss the implications of the findings for the current anti-spam efforts, and more importantly, for the design of future email delivery architectures.**

***Index Terms*—Spam, Spammer Behavior, Network Reachability of Spammers**

I. Introduction

The majority of earlier studies on the email spam have focused on the contents of email messages so as to distinguish spam messages from legitimate ones [1], [2], [3], [4]. However, there is a growing realization in the networking community that effective anti-spam techniques can be developed only with a clear understanding of the spammer behavior at various levels, in particular, the network-level behavior. Behavioral characteristics of spammers such as 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 the effectiveness (or even the feasibility) of many anti-spam mechanisms [5], [6], [7], [8]. 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 arrivals of spam with BGP route updates to investigate the network reachability properties of spammers [9]. Our study confirms the informal observation [10] that the spam arrivals from some spammers are often closely correlated in time with the BGP announcement of the corresponding network prefixes [10]. These network prefixes are short-lived in that they are withdrawn quickly after the spamming activity is over. This 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. 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:

- *The majority of spammers send only a small number of spam messages (Section V-A).* 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.
- *The vast majority of both spam messages and spam only mail servers are from mixed networks (Sections IV and V-A).* 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 V-B and V-C). 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. Alarmingly, 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*).

- *A large portion of spammers send spam only within a short period of time (Section V-D).* 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.
- *Network prefixes for a non-negligible portion of spam only networks are only visible within a short period of time (Section VI).* 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. The short life span of these network prefixes coincides with the delivery of spam from the corresponding networks. In contrast, only about 2% of non-spam only networks and 2% of mixed networks have a life span less than one week.

These findings have profound implications for the current anti-spam efforts and the design of future email delivery architectures that are inherently spam-resistant. **(1)** The fact that the majority of spammers only send a small number of spam messages and are only active for a short period of time suggests that the effectiveness of IP-address-based spam filters may be limited in combating such spammers. **(2)** 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 based purely on the network prefix information. **(3)** Given that a large portion of spam messages are sent from infected machines, sender authentication schemes such as sender policy framework [6] are in urgent need and can be very effective. 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. **(4)** The findings that the majority of spammers are only active for a short time period, and more alarmingly, some sophisticated spammers are utilizing short-lived network prefixes to hide their identities, 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 the Differentiated Mail Transfer Protocol (DMTP) [11] and IM2000 [12], have made progress in incorporating these design lessons. A recent, independent and parallel work [13] also studied some aspects of the network-level behavior of spammers. We discuss the similarities and differences between the two studies in Section III.

The remainder of the paper is structured as follows. In Section II we describe the collection of the email and BGP traces, analysis methodology, and the terminology used in the paper. Subsequently, we discuss related work in Section III. We present an overview of the email trace in Section IV. We study the behavioral characteristics of spammers and their network reachability properties in Sections V and VI, respectively. We summarize the findings and the implications in Section VII.

II. Preliminaries

A. 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 [4] 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. Specifically, we did not have access to the contents of any emails, due to privacy issues.

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 [14] 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 [14]. 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. However, due to logistical constraints, we were unable to do so and instead used the BGP trace 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.

B. Analysis Methodology

An incoming email message is classified as either *spam* or *non-spam* by SpamAssassin [4] deployed in the FSU mail relay server. (SpamAssasin 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.) 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* [15], [9], [16]. 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 [17].

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*.

C. Terminology

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.

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*.

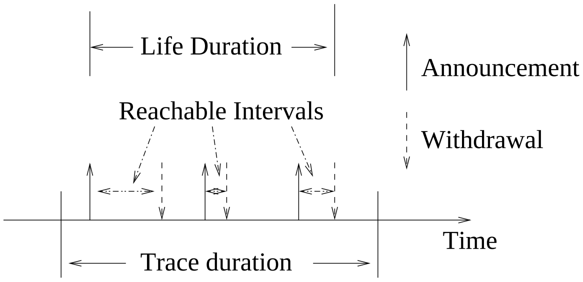

Fig. 1. An example illustration of reachable intervals and life duration.

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.

III. Related Work

A recent, independent and parallel work [13] 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, [13] based its study mainly on spam traces collected at two spam sinkholes, which presumably contain only spam messages. In contrast, our email trace contains both spam and legitimate messages, which presents us with the opportunity to compare the behavior of spammers and legitimate email users. In particular, because our trace has both spam and non-spam emails, even though we obtain some observations similar to those in [13], the conclusions may differ. For example, like in [13], we also

TABLE I
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)

observe that a significant portion of spam messages come from a small region in the IP address space. However, our conclusion is that it can be challenging to filter spam based on the network prefix information because the vast majority of spam messages are from mixed networks that also send legitimate emails.

Second, [13] 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. In other words, while the study [13] focused on the network prefixes with persistent short *reachable intervals*, we focus on the network prefixes with short *life duration* (see Figure 1 in Section II). In particular, we study the spamming pattern of the networks whose prefixes are visible only within one week out of the two-month trace collection period.

Last, while the study in [13] 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. These two studies confirm and complement each others' findings from different vantage points.

Another recent work [18] 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.

Given the importance of controlling spam on the Internet, many anti-spam schemes have been proposed including numerous email spam filters [2], [8], sender authentication schemes [5], [6], [7], and sender-discouragement mechanisms [19], [20], [21], [22]. A Differentiated Mail Transfer Protocol (DMTP) was recently proposed in [11], which advocates a receiver-driven email delivery architecture to hold spammers accountable for spamming, to throttle the spam delivery rate of spammers, and to remove spammers' flexibility in frequently changing their locations and/or Internet Service Providers. In this paper we discuss the implications of our findings for the current anti-spam efforts and the design of future email delivery architectures that can proactively resist spam.

IV. OVERVIEW OF THE EMAIL TRACE

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%, are spam (see Table I). 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. The messages come from $68,732$ networks, of which more than 90% send at least one spam message.

In Table II 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 II), about 43% of networks we observe in the trace are spam only networks. However, only a small percent 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 (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.

TABLE II
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)

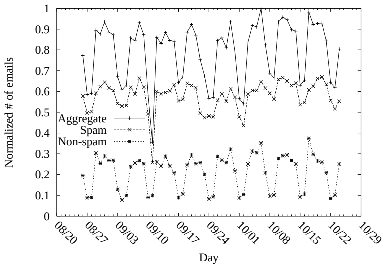

Fig. 2. Normalized daily # of emails (max # of daily aggregate emails = $518,145$).

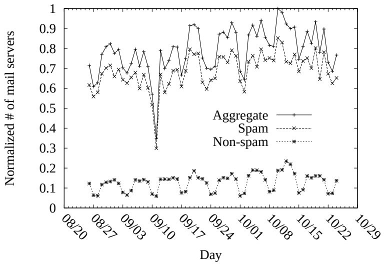

Fig. 3. Normalized daily # of mail servers (max # of mail servers = $109,933$).

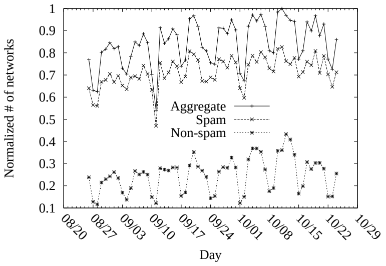

Fig. 4. Normalized daily # of networks (max # of networks = $22,968$).

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 I 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).

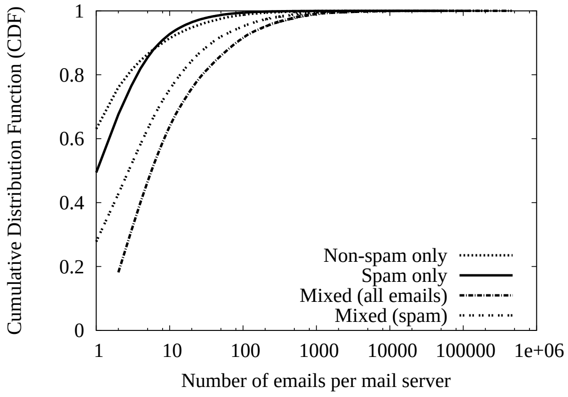

Fig. 5. Number of emails from each mail server.

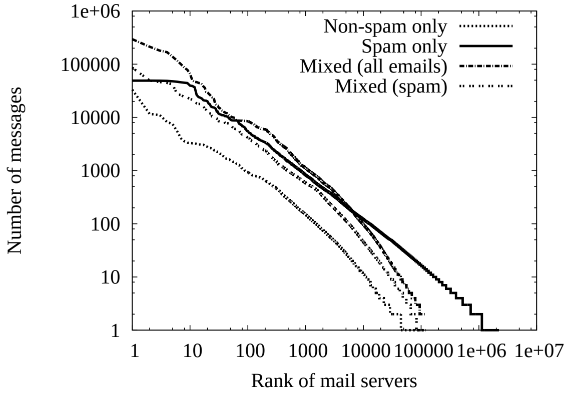

Fig. 6. Rank of mail servers based on # of messages.

V. BEHAVIORAL CHARACTERISTICS OF SPAMMERS

In this section we present a detailed study on the behavioral characteristics of spammers. In particular, we study 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 discuss the important implications of the findings for the current anti-spam efforts and the design of future email delivery architectures.

A. Number of Messages From Email Senders

Figure 5 shows the CDF of the number of messages from the observed mail servers. 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 IP-address-based anti-spam schemes such as RBL [8] 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.

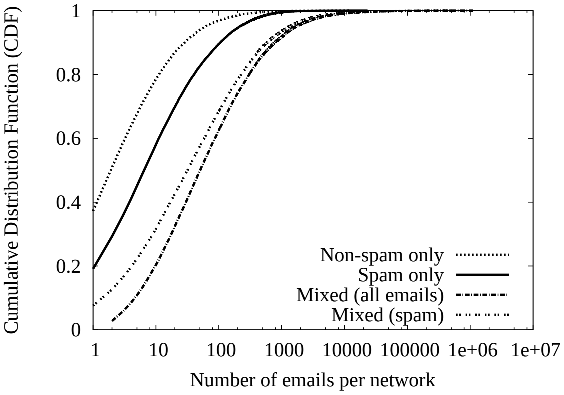

Fig. 7. Number of emails from each network.

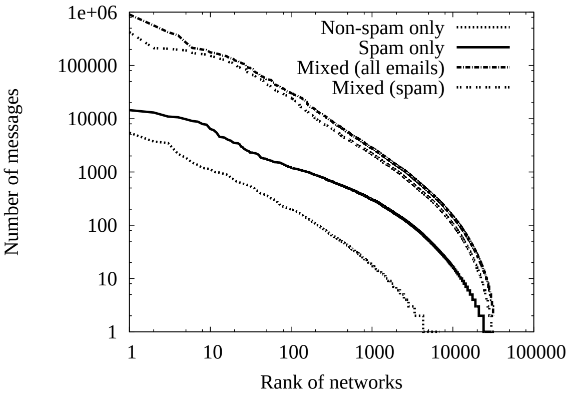

Fig. 8. Rank of networks based on # of messages.

To have a better view of the tails of the distributions of email arrivals, in Figure 6 we rank the mail servers according to 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.

Figures 7 and 8 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 5 and 6, 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.

B. IP Address Origins of Spam Messages

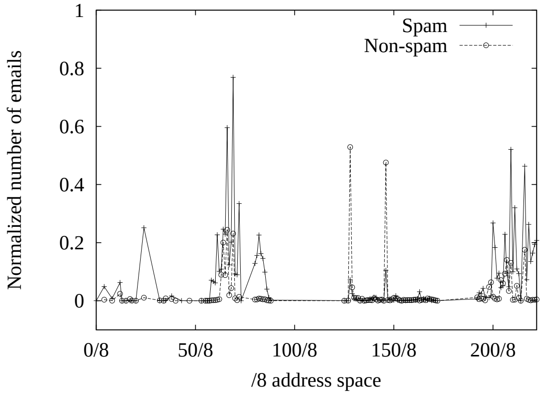

Fig. 9. Normalized number of messages from each “/8” address space (max # of aggregate emails = 2,077,655).

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 9, 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 [23]. 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 [23]. This is possibly because ISPs filter non-routable network prefixes on the Internet including private RFC 1918 address blocks and unassigned address prefixes [24].

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.

C. Number of Mail Servers and Their Origins

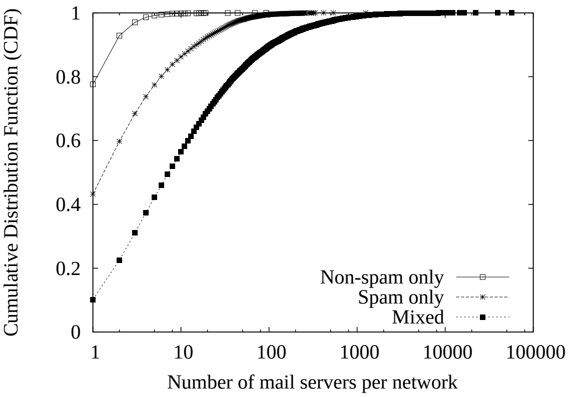

Fig. 10. Number of mail servers of each network.

Figure 10 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 $57,106$.

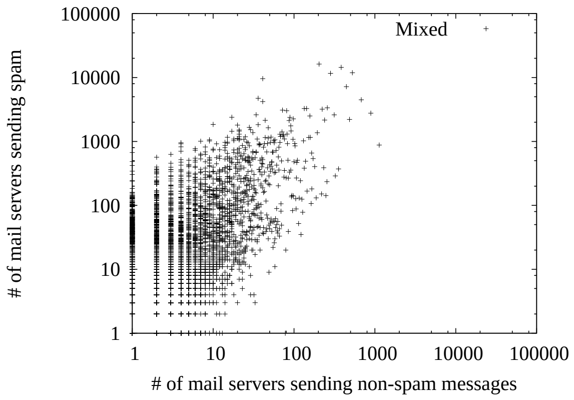

Fig. 11. Spam vs. non-spam mail servers (mixed networks).

Figure 11 plots 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 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 [6] 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.

In Figure 12 we classify each mail servers 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 9 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

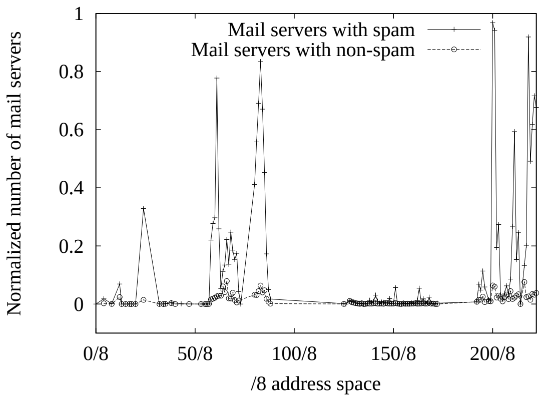

Fig. 12. Normalized # of mail servers of each "/8" address space (max # of mail servers = 147,130).

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.

D. Active Duration of Spammers

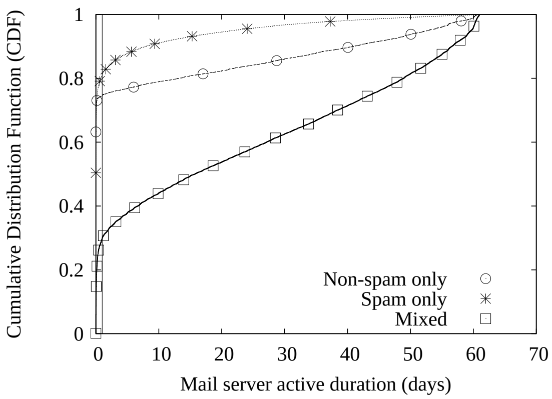

Fig. 13. Mail server active duration (days).

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 13 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.

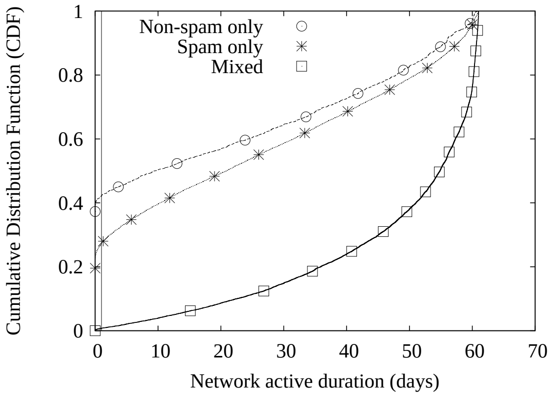

Fig. 14. Network active duration (days).

Figure 14 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.

E. Network Prefix Length and Network Types

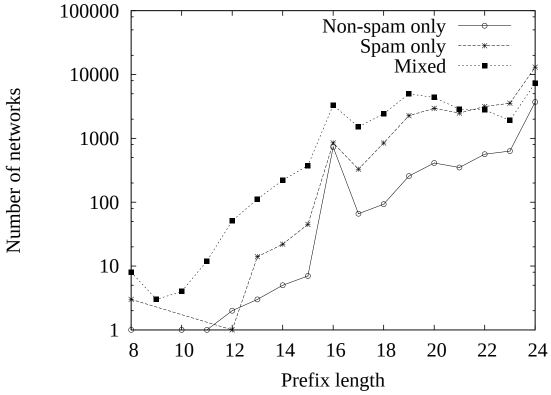

Fig. 15. Network prefix length.

Figure 15 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.

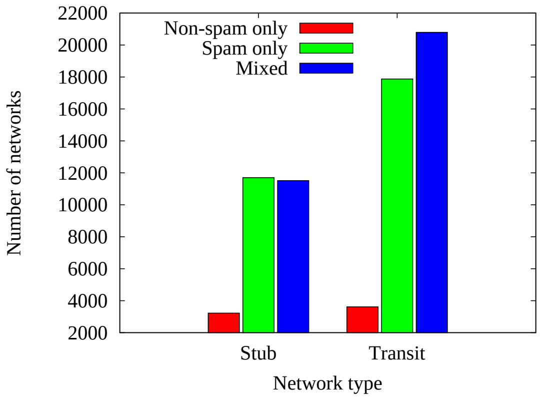

Fig. 16. Network types.

Next we examine the types of the networks observed. 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 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 16 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 slightly higher than the spam only and mixed networks. (But note that only about 10% networks observed are non-spam only network.)

VI. 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 [10]. 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 [13], which focused on network prefixes with *persistent short reachable intervals* (i.e., persistent short-lived BGP announcements [13]), which may have a life duration spanning the complete course of the email trace.

In Section VI-A, we first study how network prefixes with *short life duration* are used by spammers and investigate the prevalence of this spamming technique. In Sections VI-B and VI-C, 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 [13] and compare their observations with ours. We discuss the two different spamming patterns at the end of this section.

A. Network Life Duration

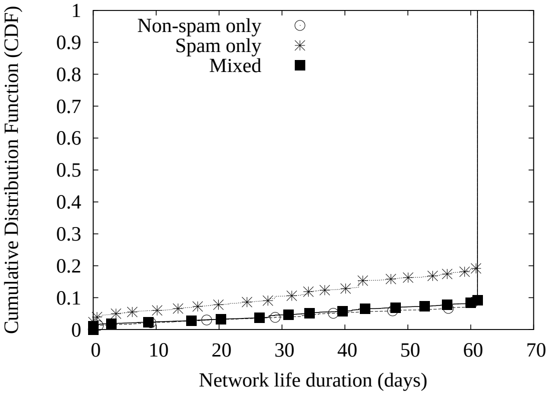

Fig. 17. Network life duration.

Recall from Section II that the life duration of a network is defined as the time interval between the first BGP announcement of the prefix (or the beginning of the two-month trace collection period if the prefix is in *rib(0)*) and the last BGP withdrawal of the prefix (or the end of the two-month period if the last BGP update of the prefix is an announcement in the period). Figure 17 shows the CDF of the life duration of all networks observed. From the figure we can see 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 II.) 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 18 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

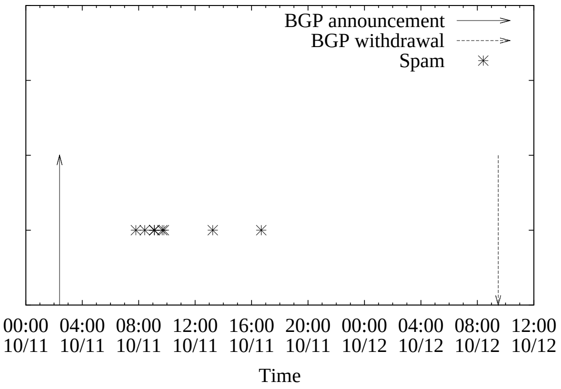

Fig. 18. Spam network with short life duration (222.46.32.0/20).

(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 17 we can see that about 6% (4%) of network-level spammers use this technique for spamming.

B. Network Reachable Intervals

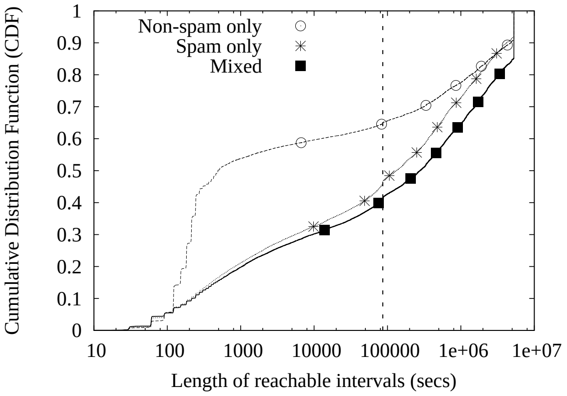

Fig. 19. Network reachable intervals.

As defined in Section II, 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 19 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.

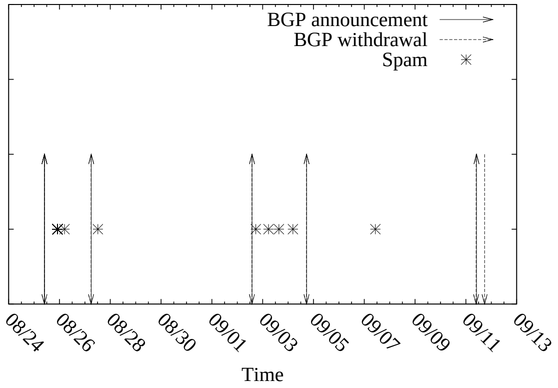

Fig. 20. Spamming correlated with BGP updates (62.215.96.0/19).

The reachable intervals of spam networks are often in concert to the arrivals of spam messages, as shown in Figure 20. 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.

C. Active 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 21 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 II). 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 22, we plot the CDF of the number of messages arriving in different reachable intervals.

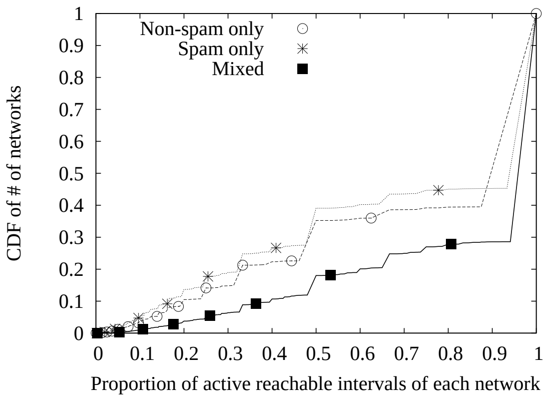

Fig. 21. Proportions of intervals with emails.

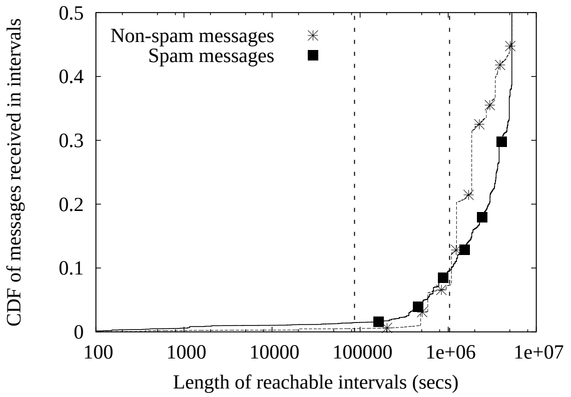

Fig. 22. Email arrivals in different network reachable intervals.

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.

D. Discussions

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 17). 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 19). Additionally, spam only and non spam only networks have similarly low proportion of *active* reachable intervals (Figure 21).

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 [13]). 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.

VII. Summary and Implications for Future Email Architectures

In this paper we studied the behavioral characteristics of spammers at both the mail server and network levels. We also investigated the network reachability properties of spammers as indicated by their BGP routing dynamics. We found that **(a)** the majority of spammers send only a small number of spam messages; **(b)** the vast majority of both spam messages and spam only mail servers are from mixed networks that send both spam and non-spam messages; **(c)** the majority of both spam messages and spam mail servers are from a few regions of the IP address space; **(d)** a large portion of spammers send spam only within a short period of time; and **(e)** network prefixes for a non-negligible portion of spam only networks are only visible for a short period of time, coinciding with the spam arrivals from these networks. The collective observations from our email and BGP traces suggest that 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 [13]).

Our findings have important implications for the current anti-spam efforts as we have previously discussed in the paper. More importantly, they also shed light on the design of future email delivery architectures that can inherently resist spam. In the current SMTP-based email delivery architecture [25], it is hard to hold spammers accountable for spamming; spammers can vanish (go offline) immediately after pushing a deluge of spam to receivers. This is confirmed by our findings that a large portion of spammers send spam only within a short period of time, and more alarmingly, some sophisticated spammers utilize short-lived networks for spamming. Our findings suggest that in order to effectively control spam, we must hold spammers accountable, force them to stay online for longer periods of time while throttling their spamming rates, and limit the spammers' flexibility in frequently changing their locations and/or Internet Service Providers. Design of new pull-based email delivery architectures, such as DMTP [11] and IM2000 [12] have made progress in incorporating these lessons.

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 and Kartik Gopalan were 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] X. Carreras and L. Márquez, "Boosting trees for anti-spam email filtering," in *Proceedings of 4th International Conference on Recent Advances in Natural Language Processing*, Tzigov Chark, BG, 2001.

[2] P. Graham, "A plan for spam," Jan. 2003, http://www.paulgraham.com/spam.html.

[3] T. Z. Jr., "The fight against v1@gra (and other spam)," *The New York Times*, May 2006, http://www.nytimes.com/2006/05/21/business/yourmoney/21spam.html.

[4] SpamAssassin, "The apache spamassassin project," http://spamassassin.apache.org/.

[5] M. Delany, "Domain-based email authentication using public-keys avertised in the DNS (domainkeys)," Internet Draft, Aug. 2004, work in Progress.

[6] M. Lentczner and M. W. Wong, "Sender policy framework (spf): Authorizing use of domains in MAIL FROM," Internet Draft, Oct. 2004, work in Progress.

[7] J. Lyon and M. Wong, "Sender ID: Authenticating e-mail," Internet Draft, Aug. 2004, work in Progress.

[8] RBL, "Real-time spam black lists (RBL)," http://www.email-policy.com/Spam-black-lists.htm.

[9] Y. Rekhter and T. Li, "A border gateway protocol 4 (BGP-4)," RFC 1771, Mar. 1995.

[10] P. Vixie, "Private communication," USENIX/SRUTI, Jul. 2005.

[11] Z. Duan, Y. Dong, and K. Gopalan, "DMTP: Controlling spam through message delivery differentiation," *Computer Networks Journal (Elsevier)*, Jul. 2007.

[12] D. Bernstein, "Internet Mail 2000 (IM2000)," http://cr.yp.to/im2000.html.

[13] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," in *Proc. ACM SIGCOMM*, Sep. 2006.

[14] U. of Oregon, "Route Views project," http://www.routeviews.org/.

[15] S. Halabi and D. McPherson, *Internet Routing Architectures*, 2nd ed. Cisco Press, 2000.

[16] J. Stewart, *BGP4: Inter-Domain Routing In the Internet*. Addison-Wesley, 1999.

[17] Merit Network Inc., "Merit network routing assets database," whois.radb.net.

[18] L. Gomes, C. Cazita, J. Almeida, V. Almeida, and W. Meira, "Characterizing a spam traffic," in *Proceedings of IMC'04*, Oct. 2004.

[19] J. Goodman and R. Rounthwaite, "Stopping outgoing spam," in *Proc. of EC'04*, 2004.

[20] E. Harris, "The next step in the spam control war: Greylisting," White Paper, Aug. 2003.

[21] A. Juels and J. Brainard, "Client puzzles: A cryptographic defense against connection depletion attacks," in *Proceedings of NDSS-1999*, Feb. 1999.

[22] K. Li, C. Pu, and M. Ahamad, "Resisting spam delivery by TCP damping," in *Proceedings of First Conference on Email and Anti-Spam (CEAS)*, Jul. 2004.

[23] Internet Assigned Number Authority, "IP address services," http://www.iana.org/ipaddress/ip-addresses.htm.

[24] Team Cymru, "The team cymru bogon route server project," http://www.cymru.com/BGP/bogon-rs.html.

[25] J. Klensin, "Simple Mail Transfer Protocol," RFC 2821, Apr. 2001.