Correlating Spam Activity with IP Address Characteristics

Chris Wilcox, Christos Papadopoulos
Department of Computer Science
Colorado State University
Fort Collins, Colorado

John Heidemann
University of Southern California
Information Sciences Institute
Los Angeles, California

***Abstract*—It is well known that spam bots mostly utilize compromised machines with certain address characteristics, such as dynamically allocated addresses, machines in specific geographic areas and IP ranges from AS' with more tolerant spam policies. Such machines tend to be less diligently administered and may exhibit less stability, more volatility, and shorter uptimes. However, few studies have attempted to quantify how such spam bot address characteristics compare with non-spamming hosts. Quantifying these characteristics may help provide important information for comprehensive spam mitigation.**

We use two large datasets, namely a commercial blacklist and an Internet-wide address visibility study to quantify address characteristics of spam and non-spam networks. We find that spam networks exhibit significantly less availability and uptime, and higher volatility than non-spam networks. In addition, we conduct a collateral damage study of a common practice where an ISP blocks the entire /24 prefix if spammers are detected in that range. We find that such a policy blacklists a significant portion of legitimate mail servers belonging to the same prefix.

I. Introduction

Spam email is a persistent problem that negatively affects Internet users and administrators. The importance of spam mitigation has grown with the volume of spam emails, now estimated at around 90% of Internet email traffic [1]. Existing methods for controlling spam include a) detecting and filtering spam email *on-the-fly* by analyzing email contents and header data, and b) maintaining a blacklist of spamming host IP addresses in order to reject undesirable email connections. The identification of spamming hosts and classification of spam emails are therefore key to reducing spam.

Studying network traffic in order to identify spamming hosts poses many challenges. A combination of legal issues, privacy concerns, and data encryption prevent examination of payload data. Researchers are left to infer malicious host behavior by studying packet headers and other artifacts of network flows. These include source and destination addresses, however Internet addressing lacks a mechanism to enforce host accountability. Many IP addresses are dynamic, so legitimate and malicious hosts can bind to different addresses over time. IP addresses can be spoofed or hijacked, making it difficult to reliably identify misbehaving hosts. Network address translation and network firewalls further complicate the task.

Despite these problems, IP ownership and visibility are an important data source that can be studied. There is a growing body of literature that suggests there are consistent and quantifiable differences between the IP address characteristics of spamming and non-spamming hosts [2], [3], [4], [5], [6]. Such differences can help identify and mitigate spam, for example by using them as predictors for blacklisting IP addresses.

In this paper we correlate the results of an Internet-wide study of IP characteristics [7], [8] against a large commercial spam blacklist. In contrast to previous research, the combination of the address study and blacklist allows us to quantify the differences between the IP characteristics of spamming and non-spamming hosts and address prefixes. We pose the following questions:

1) Do differences exist in IP availability, volatility, and uptime between spammers and non-spammers? (We will precisely define these metrics shortly.)
2) Are there differences in the domain names associated with spamming and non-spamming IP addresses?
3) What is effectiveness of blocking entire /24 address prefixes based on the IP addresses of spammers?

Our results show major differences between the IP address characteristics of spamming and non-spamming prefixes. We also measure a high level of variance between domain names used by spamming and non-spamming hosts. Both of these disparities may be valuable in the identification of spamming hosts. Finally, we find that blacklisting the addresses in a /24 IP prefix can produce significant collateral damage.

II. Data Sources

Our study uses data from the Internet IP visibility study in conjunction with a commercial IP blacklist from eSoft.com.

A. Visibility Study

The address visibility study [7], [8] examines the population of visible hosts on the Internet. The study uses censuses and surveys implemented using an active probing technique based on the ICMP protocol. The sampling methodology includes both randomly and uniformly sampled addresses, biased towards previously responding addresses. Active probing is used to provide a systematic traversal of addresses on a consistent schedule. Each survey samples around 1% of the allocated space, or ∼24,000 /24 prefixes. All hosts in each /24 prefix are probed approximately every 11 minutes for the survey period which is 1–2 weeks.

The probing methodology sends an ICMP echo request to each address and waits for a timeout period of ∼5s. Possible replies include echo reply, destination unreachable, administratively prohibited, and no reply. *Echo reply* is a positive reply which indicates the presence of a host. *Destination unreachable* is a negative reply indicating the host is unavailable or the address is unused. *Administratively prohibited* indicates some level of administrative control. This response often sent by an intervening network firewall. *No reply* or timeout can be caused by a number of factors: a lost probe, an unoccupied address, a host is temporarily down, or a firewall or router declined to generate a reply.

Several IP address characteristics can be computed based on probe responses, including availability, volatility, and median uptime statistics for each IP address. These have already been shown to distinguish dynamic addresses from static [8]. We combine the statistics to produce an average over an entire IP prefix. The method for calculating statistics is detailed in Section III. We combine address surveys with hostname data which we acquire via reverse-DNS lookup.

For this paper we computed statistics based on two surveys. The it24w survey was done Feb. 3-12, 2009, finding 2.05 million visible hosts over 20,121 /24 prefixes [9]. The it28w survey was done Sep. 14-28, 2009, finding 1.95 million visible hosts over 20,002 /24 prefixes [10]. The results were so similar that we have published numbers only from the later survey.

B. Spam Blacklist

Our blacklist comes from eSoft [11], which maintains a reputation-based filter that stores the spam score and IP address. Information used to compile the blacklist comes from more than 12,000 eSoft customers around the world. The spam score varies from approximately -60 to 70, where >30 is high spam, ≤30 and >10 is medium spam, and ≤10 is low spam. The score is based on a variety of factors:

1) Sender Address Verification: Ensures that the sender is a valid email recipient.
2) Sender Policy Framework: Verifies the sending email server is authorized to send from domain.
3) Heuristic Analysis: Looks for special characters, unusual capitalization and other signatures of spam.
4) Reputation Filtering: Takes into account the spam ratings from the eSoft customer base.
5) Bayesian Filtering: Learns "on the fly" from customer decisions about what is and isn't spam.
6) Historical Averaging: Looks at addresses over time to find repeat offenders.

Colorado State University receives a blacklist from eSoft every 30 minutes, which provides us with a historical database of spammers which spans the last 18 months. eSoft also provides raw data files at the same interval, which contain an accumulated spam score and the associated number of spam email deliveries. The calculation of the spam score is proprietary, but spammers are added to the blacklist based on a threshold of spam email deliveries over a specified period.

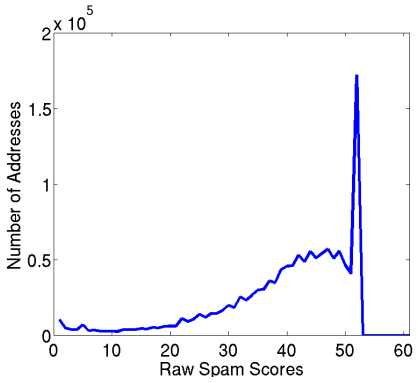


Fig. 1. Spam Score Distribution

While the actual blacklist distributed by eSoft is a subset of the raw list, the blacklist contains additional post-processing of spam scores, and *whitelist* entries. We have chosen to use the raw data for this study because of the direct and easy to understand relation to spam activity. The distribution of spam scores from the raw data is shown in Figure 1. The majority of the raw data distribution has spam scores in the high spam range. All blacklists are subject to a small percentage of false positives, but we limit the effect of these by discarding low and medium spamming addresses.

III. Methodology

Our basic methodology is to evaluate correlations between address visibility surveys and eSoft spam reports. We therefore review how we interpret the metrics used for evaluating IP visibilty from [8], the eSoft spam reports, and the specific datasets we use. Our study makes use of the datasets which were collected by the IP visibility study.

Address Metrics: Prior evaluation of IP visibility has used three metrics: address availability (A), volatility (V), and median uptime (U), described below:

1) *Availability* is the number of positive replies divided by the total number of probes. The minimum availability (0) means no reply was ever received, and the maximum (1) means a host replied to every probe.
2) *Volatility* is computed by dividing the number of up periods by the number of probes, then multiplying by two. The minimum volatility (0) represents a host which is consistently down or down during the entire survey, and the maximum (1) is when a host only replies every other time it is probed.
3) *Uptime* is the median number of seconds of all up periods. The maximum uptime is the period of the survey, which is ∼1.2 million seconds (226 hours).

Spam Metrics: We use the eSoft raw data as ground truth when determining whether a host is a spammer. The differentiation between spammers and non-spammers is based on the spam score from the eSoft raw data, described in the previous section. We additionally filter the eSoft raw data to discard hosts with a spam score less then 20.0, to ensure that only highly active spamming hosts are compared.

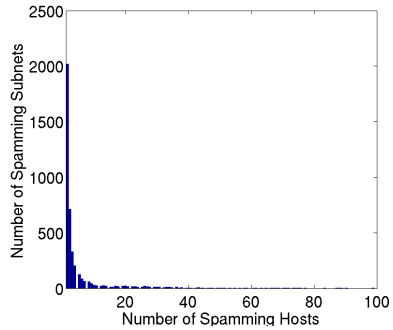
Datasets: The results in this paper are derived from the it28w survey. We coalesce address characteristics for a /24 prefix by computing the mean availability, volatility, and uptime of all of the hosts it contains. We coalesce the eSoft raw data from the dates coinciding with the survey in a similar manner, from 1.43 million addresses to 261,943 /24 prefixes. The spam score for a prefix is the mean of its host scores. After coalescing, we intersect the survey prefixes with the eSoft prefixes. Survey prefixes containing hosts from the eSoft data set are labeled as spamming, those without are labeled non-spamming. Those from eSoft that do not intersect the survey are discarded.

IV. Results

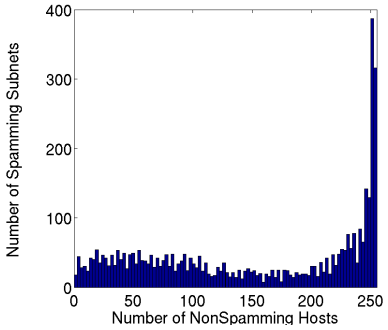
We begin by considering how many spamming hosts are in each /24 prefix. The result of the intersection between the it28w survey and the eSoft raw data for the same period is 4,126 spamming and 15,876 non-spamming prefixes. The spamming prefixes are ∼20% of the prefixes from the survey, representing 646K hosts that answer a reverse-dns query. The non-spamming prefixes are ∼80% the of the survey, with around 2,252K hosts. We use these two sets of prefixes for the analysis that follows.

A. IP Address Intersection

Figure 2(a) and Figure 2(b) show the distribution of of spammers and non-spammers in the intersected /24 prefixes. We observe that many spamming prefixes contain only a single spamming host, and very few prefixes have more than 10. This mirrors the distribution of spammers in all of the eSoft raw data, which is not shown. The sparse distribution of spammers suggests the possibility of significant collateral damage if the entire prefix is blocked, we return to this question in IV-D. Non-spamming prefixes have a more even distribution of hosts, although there is a cluster of highly populated prefixes starting around 240 hosts.



(a) Spamming hosts per /24 prefix (4,126)



(b) Non-Spamming hosts per /24 prefix (15,876)

Fig. 2. Number of hosts per /24 prefix

B. IP Address Characteristics

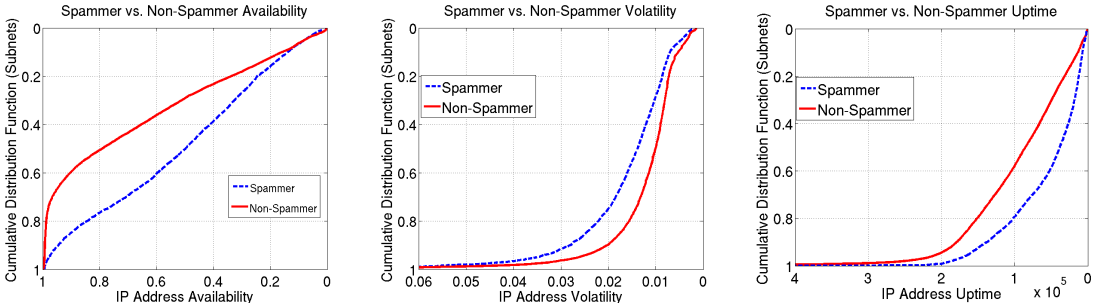
We correlate ping-observable characteristics of prefixes and spam origination, with the goal of understanding whether address surveys can predict spamming. These results corroborate prior observations that dynamic addresses originate more spam, and confirm that address surveys provide an alternative method for identifying spammers.

Figure 3(a) shows the cumulative distribution function (CDF) for IP address availability. All graphs in this section are plotted with an inverted CDF to make the differences in address characteristics more visually comprehensible. The non-spamming prefixes are clearly more available than the spamming prefixes. 72% of non-spammers and 50% of spammers have >0.5 availability, and 50% of non-spammers and 24% of spammers have >0.8 availability. This corresponds to previous results [12], and provides additional evidence that spammers use dynamic address spaces more heavily. One interesting deviation is the crossover apparent when availability is <0.1. We theorize that spammers are less likely to make use of these addresses because of low availability, which makes them less vulnerable and less desirable for malicious use.

Figure 3(b) shows the CDF for IP address volatility. The graph is plotted so that a higher line represents more volatility. We see that spamming prefixes are more volatile than non-spamming prefixes. From the graph, 90% of non-spammers and 75% of spammers have <0.02 volatility, and 50% of non-spammers and 28% of spammers have <0.01 volatility.

Figure 3(c) shows the CDF for IP address uptime. The graph is plotted so that a higher line represents higher median uptime. It is apparent that non-spamming prefixes have a significantly longer median uptime than the spamming prefixes. 70% of non-spammers and 42% of spammers have >50000s (14 hours) uptime, and 44% of non-spammers and 22% of spammers have >100000s (28 hours) uptime. We conclude that spamming hosts tend to have lower uptimes than non-spammers; another aspect of dynamic address usage. Shorter uptimes imply that addresses are often reused, suggesting that blacklists based on individual addresses may cause some collateral damage by filtering hosts that are reusing addresses previously occupied by spammers.

Searching for more evidence of the relationship between spamming and IP visibility, we plot the CDF of availability for several different levels of spam scores. Figure 4 shows that higher spam scores correlate closely with lower availability. The lines in the graph follow the eSoft definitions of low spammers (spam score <10) and high spammers (spam score >30), and adds two categories in between. From the graph we see that 83% of low spammers have >0.9 availability, for high



(a) IP Availability (b) IP Volatility (c) IP Uptime

Fig. 3. IP Address Characteristics

spammers this number is only 14%. Note the clear separation between the different levels of spam scores.

To summarize the IP visibility data, we find that non-spamming prefixes exhibit more availability, less volatility, and more uptime than spamming prefixes. Availability decreases as spam score increases, in what appears to be a direct correlation. Thus it appears from our data that more stable networks tend to have fewer spammers, suggesting that availability is a possible indicator of spam origination.

C. Domain Names

Domain names often contain words that describe usage, allocation, ownership, or access type. For example 'server' may indicate a mail or web server, 'static' may indicate a static address allocation, and 'cable' may indicate a cable customer. We use reverse-DNS lookups on our set of intersected prefixes to supply domain names which we analyze according to the categories from [8], as shown in Table I.

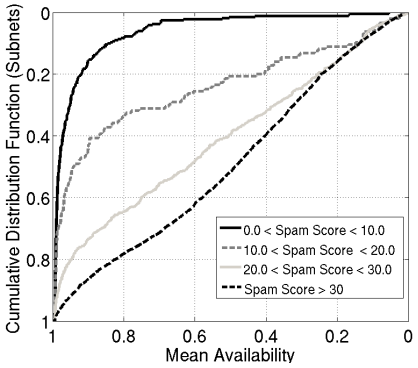


Fig. 4. IP Availability by Spam Score

TABLE I
CATEGORIES OF DOMAIN NAMES

GROUP	CATEGORY	KEYWORDS
ALLOCATION	STATIC	STATIC
	DYNAMIC	DYNAMIC, DYN
	DHCP	DHCP
	POOL	POOL, POND
	PPP	PPP
ACCESS	DIAL	DIAL, MODEM
	DSL	DSL
	CABLE	CABLE
	WIRELESS	WIRELESS, WIFI
	DED	DED, DEDICATED
CONSUMER	BIZ	BIZ, BUSINESS
	RES	RES, RESIDENT
SERVER	SERVER	SERVER, SRV, SVR, MAIL, SMTP, WWW, NS, FTP
	ROUTER	ROUTER, RTR, RT, GATEWAY, GW
	CLIENT	CLIENT

We extend the result to show a comparison of domain names based on spammer versus non-spammer classification, shown in Figure 5. A number of disparities are apparent. Spammers prevail by a factor of 2X in dynamic domains, and non-spammers prevail by more than 3X in static domains. On top of IP address characteristics, this provides independent corroboration that spammers are more commonly associated with dynamic addresses [8]. Non-spammers are more common by a factor of 1.5X in server domains, and they completely dominate router domains. This is unsurprising since server domains are commonly associated with static addresses. Spammers are more common in the dsl, pool, and dial categories. We suspect based on the category name that these addresses are commonly used by ISPs to dynamically allocate addresses to customers. The reverse is true in cable domains, this is a topic for further investigation.

D. Collateral Damage

The prior sections have shown that both address and hostname characteristics confirm that spam originates from dynamic addresses. We use these results to consider a new

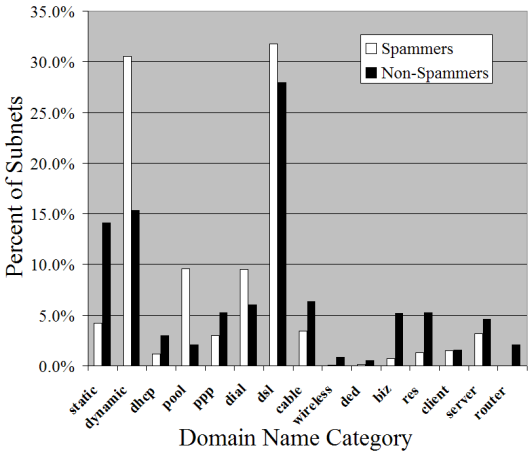


Fig. 5. IP Domain Name Comparison

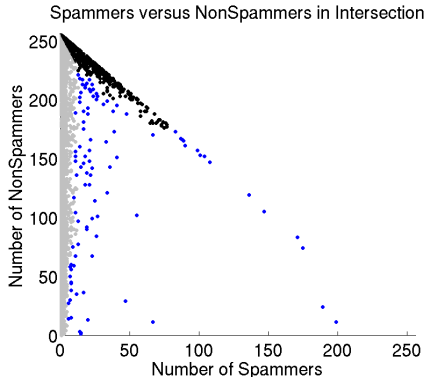


Fig. 6. Number of Spammers versus Non-Spammers

question: Is blacklisting an entire /24 prefix based on the presence of one or more spamming hosts an effective policy? While many blacklists enumerate individual IP addresses, blocking entire /24 prefixes are also common [13]. We are concerned about reducing spamming, but also about the blocking of legitimate outgoing email. We define collateral damage [13] as the number of legitimate mail servers which would be incorrectly filtered when an entire /24 prefix is blacklisted. First we identify all survey prefixes which have spamming hosts. If these prefixes also contain non-spamming hosts, then they are subject to collateral damage. Figure 6 compares the number of spammers versus non-spammers in the set of intersected prefixes.

Except for outliers, the graph shows that many of the prefixes seem to cluster along the left-axis (grey) or the top diagonal (black). The diagonal is present because the sum of spammers and non-spammers is never more than the size of a prefix (255). The left-axis cluster shows prefixes with a reasonably uniform number of non-spammers and a small number of spammers. The diagonal cluster shows a large number of spammers residing in highly populated prefixes. These clusters may reflect two different situations. The diagonal cluster shows heavily compromised prefixes, which we believe may have negligent administration or a collaborating provider. The other cluster represents a limited number of compromised hosts in an otherwise normal prefix, we believe these may be caused by bots. The latter are prone to collateral damage, since they contain a high number of non-spamming hosts and a low number of spamming hosts.

Anti-spammers typically assume there is no collateral damage in blacklisting a /24 prefix, because many ISPs forward legitimate mail through the ISPs mail server, rather than allowing hosts to send mail directly. We are only able to quantify whether a blacklisted prefix contains mail sources, by studying their hostnames and DNS mail forwarding records. For this study we extract the reverse hostname and MX record for each address in the prefix, using the Linux `host` and `dig` commands. Finally we intersect the mail server IP addresses against the survey dataset to see if any reside in the blacklisted prefixes. Table II shows the progression of our data analysis.

TABLE II
COLLATERAL DAMAGE STUDY

DESCRIPTION	DOMAINS	HOSTS	PREFIXES
INTERSECTED PREFIXES		646,040	4,126
DOMAIN QUERY TIMEOUT		12,899	
DOMAIN QUERY INVALID		175,535	
DOMAIN QUERY VALID		457,606	
UNIQUE DOMAIN NAMES	4,044		
NUMBER MAIL SERVERS		6,718	
UNIQUE MAIL SERVERS		3,872	2,154
COLLATERAL DAMAGE		1,377	365

We start with 646,040 addresses that reside in the 4,126 spamming prefixes in our intersection set. We subtract addresses that timeout or fail to return a valid domain name. From the remainder, our programs identify a set of unique domain names, and the addresses of the corresponding mail servers. Intersecting these addresses with our spamming prefixes, we find collateral damage of 1,377 mail servers and 365 prefixes, which is ∼8.8% of all spamming prefixes. We conclude that prefix blocking incurs a high rate of collateral damage, suggesting the need for finer-grain filtering.

V. ROBUSTNESS OF RESULTS

When studying the full Internet, *all* data sources concerning spam and address usage are imperfect. Since the conclusions of this paper depend on those datasets, we need to examine the potential sources of error in our address usage, domain name, and spam score data. Ping-based address probes are known to under-represent the number of responsive addresses [7] by about one-third, primarily due to firewalls. This suggests that the number of non-spammers is likely to be low, thus our conclusions may under-estimate the amount of collateral damage. Additional evaluation of the ping survey methodology is found in [8].

Accuracy of the spam data from eSoft is another potential source of error. Spam blacklists vary greatly between vendors, due to differences in the data and methods. Evaluation of this work against an alternative spam service is therefore a candidate for future work. The spam scores used in this study are specific to eSoft, as of yet there is no industry standard defined for scoring spamming activity. The eSoft methodology does ensure that spammers on the blacklist have recently sent spam emails, which increases our confidence that the survey and blacklist data coincide during the dates of the survey.

Finally our methodology for determining collateral damage can underestimate or overestimate legitimate mail servers. Organizations can locate sending servers at different addresses from their receiving servers. Mail servers that only receive email should not be counted as collateral damage. Mail servers that only send email are not required to maintain an MX record, so our study may miss them. Finally we consider only the presence of mail servers, not the email message volume. In spite of these challenges, we believe that our methods for estimating address usage and differentiating spammers represent a reasonable evaluation based on current methods.

VI. Related Work

There is a large body of work characterizing the behavior of spammers and spam botnets. Below we outline the work most relevant to ours, specifically research that explores the address characteristics of hosts sending spam.

Several papers have examined the relation between host behavior and IP address ownership. Xie at al. investigated the dynamic nature of IP addresses using Hotmail traces [12]. They established that the majority of mail servers using dynamic addresses sent spam only, accounting for almost half of the spam emails received by Hotmail. Kokkodis and Faloutsos identified a previously unidentified IP space with high spamming activity, and show that spamming is becoming more equally distributed in the IP space [14]. Chen, Ji and Barford studied the distribution and longevity of IP addresses used for malicious behavior [5]. They found that 80% of malicious sources attack from the same 20% of IP space. A large percentage of malicious addresses appear only a small number of times in their trace, implying that malicious hosts use IP addresses in a transient and therefore hard to filter fashion. We confirm and extend the results in these papers by quantifying the relationship between IP address characteristics and spamming behavior.

Ramachandran and Feamster used network-level behavior to study spamming hosts [2]. They correlated spam email against DNS blacklist lookups, BGP routing information, botnet IRC traces, and TCP flow behavior. The authors found that the majority of spam comes from a few regions of the IP address space and a non-negligible amount of spam is sent from hijacked prefixes. The authors concluded that network-level properties could be useful for spam mitigation. Follow-up work by Ramachandran and Feamster proposed the creation of blacklists based solely on the network-level behavior of spamming hosts [3]. We extend the work of these authors by adding a control group, thus allowing a direct comparison of spamming and non-spamming hosts.

Duan, Gopalan, and Yuan [4] correlated the IP addresses of mail senders in a large email trace containing both legitimate and spam email, with concurrent BGP announcements and withdrawals. They analyzed network behavior such as the IP address distribution and duration and the network prefix length, and were able to distinguish between spamming and non-spamming hosts. Further results included a quantification of the number of spamming and non-spamming hosts within the same network prefix. Overall this paper confirms the results from [2] and [4]. Our work differs in that we use active probing to characterize IP addresses, and we analyze prefixes at /24 instead of BGP prefix granularity. We extend the analysis of spammer behavior by quantifying the intersection between malicious and legitimate email hosts within /24 prefixes.

VII. Future Research

There are three immediate directions for future work. We plan to extend the collateral damage work by evaluating a large e-mail trace in order to 1) identify mail servers missed by our DNS queries, and 2) quantify how many legitimate emails would be incorrectly filtered. We also will examine the outliers in our intersection set to understand why some prefixes have such a high percentage of spamming hosts. In addition, we plan to investigate whether newly allocated IP addresses are targeted by spammers.

VIII. Conclusions

In this paper, we combine two datasets, an Internet-wide IP address characteristics study and a commercial spam blacklist, to quantify differences between spamming and non-spaming prefixes. We conclude that the combination allows differences to clearly emerge and be quantified. Finally, we carry out a collateral damage study of a practice where ISPs block entire /24 prefixes when spam activity is detected. We show that this practice can disrupt legitimate email activity.

Acknowledgments

We acknowledge access to the IP visibility survey data sets and other help from Xue Cai and Yuri Pradkin at USC/ISI. Additional help was provided by Dr. Dan Massey and Steve DiBenedetto at CSU.

This material is based on work partially supported by the United States Department of Homeland Security contract number NBCHC080035 ("LANDER-2007"), and by the National Science Foundation (NSF) under grant number CNS-0823774, "NeTS-NBD: Maltraffic Analysis and Detection in Challenging and Aggregate Traffic (MADCAT)", and by ARO under grant number W911NF-06-1-0094, "Spatio-Temporal Nonlinear Filtering with Applications to Information Assurance and Counter Terrorism". All conclusions of this work are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] [Online]. Available: www.pcworld.com/article/id,165533/article.html

[2] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 291–302, 2006.

[3] A. Ramachandran, N. Feamster, and S. Vempala, "Filtering spam with behavioral blacklisting," in *Proceedings of CCS 2007*. ACM, 2007, pp. 342–351.

[4] Z. Duan, K. Gopalan, and X. Yuan, "Behavioral characteristics of spammers and their network reachability properties," in *Proceedings of ICC 2007*, 2007, pp. 164–171.

[5] Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of internet malicious sources," in *Proceedings of INFOCOM 2008*. IEEE, 2008, pp. 2306–2314.

[6] N. Syed, S. Hao, N. Feamster, A. Gray, and S. Krasser, "Detecting spammers with snare: Spatio-temporal network-level automatic reputation engine," Georgia Institute of Technology, Tech. Rep. GT-CSE-08-02, May 2008.

[7] J. Heidemann, Y. Pradkin, R. Govindan, C. Papadopoulos, G. Bartlett, and J. Bannister, "Census and survey of the visible internet," in *Proceedings of IMC 2009*. ACM, 2008, pp. 169–182.

[8] X. Cai and J. Heidemann, "Understanding address usage in the visible internet," USC/Information Sciences Institute, Tech. Rep. ISI-TR-2009-656, February 2009.

[9] USC/LANDER project, "Internet addresses survey dataset, PREDICT ID USC-LANDER/internet_address_survey_reprobing_it24w-20090203," web page http://www.isi.edu/ant/lander, 2009.

[10] ——, "Internet addresses survey dataset, PREDICT ID USC-LANDER/internet_address_survey_reprobing_it28w-20090914," web page http://www.isi.edu/ant/lander, 2009.

[11] ESoft, Inc., Web page http://www.esoft.com, January 2009.

[12] Y. Xie, F. Yu, K. Achan, E. Gillum, M. Goldszmidt, and T. Wobber, "How dynamic are IP addresses?" in *Proceedings of SIGCOMM 2007*. ACM, 2007, pp. 301–312.

[13] A. Church, "DNS blacklists considered harmful," Internet Draft, Work in Progress, August 2005. [Online]. Available: http://ietfreport.isoc.org/all-ids/draft-church-dnsbl-harmful-01.txt

[14] M. Kokkodis and M. Faloutsos, "Spamming botnets: Are we losing the war?" in *Proceedings of CEAS 2009*, 2009.