

Abnormally Malicious Autonomous Systems and their Internet Connectivity

Craig A. Shue, Andrew J. Kalafut, and Minaxi Gupta

Abstract—While many attacks are distributed across botnets, investigators and network operators have recently identified malicious networks through high profile autonomous system (AS) de-peering and network shut-downs. In this paper, we explore whether some ASes indeed are safe havens for malicious activity. We look for ISPs and ASes that exhibit disproportionately high malicious behavior using ten popular blacklists, plus local spam data, and extensive DNS resolutions based on the contents of the blacklists. We find that some ASes have over 80% of their routable IP address space blacklisted. Yet others account for large fractions of blacklisted IP addresses. Several ASes regularly peer with ASes associated with significant malicious activity. We also find that malicious ASes as a whole differ from benign ones in other properties not obviously related to their malicious activities, such as more frequent connectivity changes with their BGP peers. Overall, we conclude that examining malicious activity at AS granularity can unearth networks with lax security or those that harbor cybercrime.

Index Terms—Autonomous Systems, Security.

I. INTRODUCTION

The Internet is plagued by malicious activity, from spam and phishing to malware and denial-of-service (DoS) attacks. Much of it thrives on armies of compromised hosts, or *botnets*, which are scattered throughout the Internet. However, malicious activity is not necessarily evenly distributed across the Internet: some networks may employ lax security, resulting in large populations of compromised machines, while others may tightly secure their network and not have any malicious activity. Further, some networks may exist solely to engage in malicious activity. Several recent ISP enforcement actions, such as the Atrivo and McColo autonomous system (AS) de-peering [1], [2] and the FTC closure of Pricewert networks [3], highlight that there are networks that exist simply to launch attacks. In this paper, we examine whether we can find malicious networks in a systematic manner using existing blacklists.

A systematic detection of disproportionately malicious networks can be used to build metrics which may be used to determine if a network is harboring a significant amount of malicious activity. Such metrics may offer several practical

benefits. First, ISPs could use them to build identification of malicious networks into their peering agreements. As an example, provider ISPs may use the metrics to require their customers to limit the amount of malicious activity in their networks to avoid harboring criminals. ISPs could also use the metrics to determine the effectiveness of their efforts to combat abuse and compare themselves with other networks. Also, when receiving traffic, a destination network could prioritize traffic based on the cleanliness of ASes, which the metrics can help estimate. This would allow a network under attack to prioritize traffic that is less likely to be associated with attackers. Finally, such metrics could also aid spam filtering programs in their scoring of email messages.

To determine which ASes are malicious, we use ten of the most commonly-used blacklists for spam, phishing, malware and botnet activities for a period of a month, in addition to URLs from spam collected at our department's email server. These blacklists either contain host names or IP addresses to be blacklisted. For host name-based blacklists, we first determine the IP addresses for each blocked host using real-time DNS queries. This gives us IP addresses of all blacklisted hosts in our blacklists. We then use BGP routing tables to group these IP addresses into their originating ASes. Upon grouping these addresses by AS, we compare ASes by the percent of infected machines and the rate at which they are cleaned up. Using data from the RouteViews Project [4], we examine other characteristics of the malicious ASes, such as whether their connectivity to other ASes changes more often than those without malicious activity. The key findings of our study are:

- A large fraction of routable space is malicious for some ASes: Four ISPs, 2 from Ukraine, one from Iran, and one from Belarus, have over 80% of their routable IP addresses blacklisted. This raises concerns regarding the purpose of such ISPs.
- Some ASes account for significantly large fractions of blacklists: Four ASes, three of which are US-based hosting providers and one large broadband service provider in Turkey, account for over 6% of at least one of the blacklists we tested.
- Some providers regularly peer with malicious ASes: We find 22 provider ISPs with 100% of their customer ASes engaged in significant malicious activity.
- Malicious ASes differ from benign ones in other ways: They are more likely to become completely unreachable than those which have less malicious activity, and they are likely to have more peers. However, the duration of unreachability is short for these ASes, which may have

Manuscript received June 11, 2010; revised January 31, 2011; accepted May 16, 2011; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor Z. Mao.

A. J. Kalafut is with the School of Computing and Information Systems, Grand Valley State University, Allendale, MI 49401 USA (e-mail: kalafuta@gvsu.edu).

C. A. Shue is with the Cyberspace Sciences and Information Intelligence Research Group, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA (e-mail: cshue@ornl.gov).

M. Gupta is with the School of Informatics and Computing, Indiana University, Bloomington, IN 47401 USA (e-mail: minaxi@cs.indiana.edu).

implications for orchestrated de-peering attempts.

Overall, these results confirm that examining malicious activity at the AS granularity can help find networks that are disproportionately bad, providing a metric for focusing network clean-up efforts.

The remainder of this paper is organized as follows. In Section II, we describe our data collection and data sets. In Section III, we examine the degree to which ASes are malicious. Section IV examines the characteristics of normal and malicious ASes, including BGP behavior and AS size. We describe related work in Section V and discuss the limitations and alternatives to using blacklists for this analysis in Section VI. We conclude in Section VII.

II. DATA COLLECTION

To create a comprehensive evaluation of an AS, we use a diverse set of data sources. Each of our data sources list machines reported as engaging in some form of malicious activity. Our data sets have a few limitations. We discuss them in Section VI.

A. Data Sets

For each set, data was collected from June 1, 2009 to June 30, 2009 unless otherwise indicated. We summarize the data sources in Table I, and describe them below.

1) *Phishing Sites*: Phishing sites attempt to collect sensitive data, such as login credentials, credit card numbers, account numbers, and social security numbers, from users by impersonating legitimate organizations or brands. The Anti-Phishing Working Group (APWG) [5] and PhishTank [6] have among the largest data feeds listing such phishing sites. We use their feeds, referred to subsequently as APWG and PhishTank data sets, respectively. Both of these feeds contain URLs of phishing sites, along with other meta-data. On an hourly basis, we extract host names from URLs currently in the feed, and perform DNS resolutions in each host name to get lists of IP addresses associated with these feeds. The PhishTank data set had a two-day outage on June 20 and June 21 causing us to only have 28 days of data.

2) *Spam/Scam Sites*: Similar to their phishing site brethren, scam sites are sites that are advertised in unsolicited messages. These spam-advertised sites may actually be phishing sites, be running some other type of scam, or provide actual legitimate products or services. We use lists of scam sites from two of the major collectors of such information, Support Intelligence [7] and SURBL [8].

We receive the feed from Support Intelligence every six hours. This feed contains URLs from spam as well as associated IP addresses. We use the IP addresses as our SI-Feed data set. Not every URL in this feed has an associated IP address, and for some that do, when we resolve the associated host names we get different addresses. Therefore, we use our own resolutions of these as another data set, SI-DNS.

SURBL also collects domain names from URLs contained in spam. Although they typically only allow users to perform lookups on the domain names in their list, we have also arranged to receive the associated IP addresses from them.

These IP addresses are those associated with the domain itself, and with the domain with `www` prepended. We receive this feed once per day, and refer to it as SURBL.

Finally, we harvest URLs from spam sent to the Computer Science Department at Indiana University (IU) and use it to create the Local Spam data set. This is a daily feed. We extract host names from this feed and perform DNS resolutions to obtain corresponding IP addresses.

3) *Spam Senders*: Mail server can use IP blacklisting to prevent compromised machines from sending mail directly. Spamhaus runs the most widely-used blacklist in this context, the SBL [9]. We obtain a copy of this blacklist every hour, and extract IP addresses to create the Spamhaus SBL data set. Data collection for the Spamhaus SBL data set started a day later than the others, beginning on June 2, 2009.

4) *Exploited Hosts*: Spamhaus also maintains a second blacklist, known as the XBL [10]. This list contains prefixes (often individual IP addresses) of hosts infected with exploits often used to send spam. This includes open proxies, computers infected with viruses that are known to send spam, and other exploits. This data is updated every half hour, and is labeled Spamhaus XBL. Data collection for this data set started a day later than the others, beginning on June 2, 2009.

5) *Malware Downloads*: Malicious software, or *malware*, including viruses, worms, and trojans, have harmful effects on the computers they infect. Three of our data sets list Web sites which host malware downloads. The Clean-MX Viruswatch mailing list [11], eSoft [12], and Malware Patrol [13], all independently collect URLs which host malware. The Viruswatch mailing list periodically sends out emails indicating newly discovered URLs with viruses. We receive mails from eSoft with new URLs containing malware, along with a malware sample, as they are discovered. We download new URLs from Malware Patrol every hour. In each case, we extract host names, and perform DNS resolutions to obtain the set of IP addresses we use. We label these data sets CleanMX, eSoft, and Malware Patrol, respectively.

6) *Bot Command and Control*: Botnets consist of groups of compromised machines used for malicious purposes on the Internet. Miscreants often use them for sending spam and for hosting phishing and scam sites. Bots must get their instructions from their bot masters, often through command and control servers. The ShadowServer Foundation [14] provides lists of botnet command and control servers along with their IP addresses. We have an hourly access to this data, referred to as Bot C&C subsequently.

B. Data Set Comparisons

Due to differing goals, methodologies, and data sources, each data set we use can be expected to contain IP addresses not seen in other data sets. By examining the overlap of IP addresses from different data sets, we can see how often IP addresses are used for multiple different malicious purposes. In Table II, we show the number of data sets containing each IP address. The Spamhaus XBL is roughly three orders of magnitude larger than any other data set, so the vast majority of IP addresses appear only in that single data set. It is further

TABLE I
OVERVIEW OF DATA SETS

Label	Description	Duration (in days)	Unique IP Addresses	Unique ASes	Median IPs Per AS	Std. Dev. IPs per AS
APWG	Phishing URLs from the Anti-Phishing Working Group	30	9,560	1,803	2	18.0
Bot C&C	Botnet command and control IPs from the ShadowServer Foundation	30	1,986	611	1	11.4
CleanMX	Malware serving sites from the CleanMX VirusWatch mailing list	30	2,974	687	1	12.0
eSoft	Malware serving sites from eSoft, Inc.	30	8,000	1,196	2	27.2
Local Spam	URLs from spam messages received by the IU CS Department	30	5,495	1,024	1	16.5
Malware Patrol	MalwarePatrol's block list for malware-serving sites	30	871	368	1	5.3
PhishTank	Phishing URLs from PhishTank	28	7,143	1,580	1	14.2
Spamhaus SBL	Verified spam sources from Spamhaus.org Block List	29	6,422	2,005	1	8.9
Spamhaus XBL	Hijacked machines from Spamhaus.org Exploit Block List	29	29,585,604	13,580	9	31,568.1
SI-Feed	URLs and IP addresses from spam emails from Support Intelligence	30	7,591	1,420	1	20.2
SI-DNS	IP addresses from DNS resolutions on the SI-Feed data set	30	4,448	911	1	11.8
SURBL	Host names appearing in spam messages from SURBL	30	29,324	2,739	2	47.2

unsurprising that some IP addresses appear in two or three data sets since some of our data sets track the same information. We see that some IP addresses appeared in multiple data sets, with 8 IP addresses appearing in 9 of our data sets and another 7 appearing in 8 sets. This indicates that malicious machines are occasionally used for many forms of malicious activity; however, a large majority appear not to be.

TABLE II
DEGREE TO WHICH AN IP ADDRESS APPEARS IN MULTIPLE BLACKLISTS

Number of Blacklists with Given IP Address	Number of IP Addresses
1	29,631,573
2	9,566
3	3,650
4	1,290
5	320
6	112
7	29
8	7
9	8

Now, we look at similarity between any two data sets. We calculate the Jaccard similarity coefficient between the sets of IP addresses in each. Let IP_{S_i} be the set of IP addresses in data set S_i . Then the Jaccard similarity of two data sets is given by $J(IP_{S_i}, IP_{S_j}) = \frac{|IP_{S_i} \cap IP_{S_j}|}{|IP_{S_i} \cup IP_{S_j}|}$. Results for all data sets except for Spamhaus XBL are shown in Table III. We ignore the XBL because its size is orders of magnitude bigger than any other data set, hence the Jaccard coefficients involving it would be extremely small. As expected, we see the highest similarity between the two phishing data sets, and the two derived from Support Intelligence data. Notably, the Bot C&C data set shares at most 4 IP addresses with any other data set, while most others, even measuring different types of bad behavior, have greater similarity to each other. Although one of the most similar, the two phishing data sets still only share 24% of their combined IP addresses with each other. The malware data sets have even less similarity in IP addresses. This analysis exposes some of the practical limitations of using blacklists: some malicious behavior is reported and captured by some blacklists, while other behavior goes unreported. By using many different blacklist providers, we have a better view of malicious activity than would otherwise be possible.

Next, we map these IP addresses to their autonomous systems and repeat a similar calculation for the overlap between

TABLE III
JACCARD SIMILARITY BETWEEN IP ADDRESSES IN EACH DATA SET

	Bot C&C	CleanMX	eSoft	Local Spam	Malware Patrol	Phishtank	Spamhaus SBL	SI-Feed	SI-DNS	SURBL
APWG	0	.06	.05	.02	.01	.24	.01	.04	.03	.10
Bot C&C	0	0	0	0	0	0	0	0	0	0
CleanMX		0	.07	.01	.06	.07	0	.01	.01	.02
eSoft			0	.01	.01	.05	0	.02	.01	.02
Local Spam				0	.01	.02	.01	.06	.09	.05
Malware Patrol					0	.01	0	.01	.01	.01
Phishtank						0	.01	.02	.02	.05
Spamhaus SBL							0	.01	.01	.01
SI-Feed								0	.49	.06
SI-DNS									0	.06

the ASes represented by the IP addresses contained in each data set. In order to map IP addresses to an AS, we used a June 15, 2009 BGP routing table from the RouteViews Project [4]. We chose this date because it is in the middle of our data collection and is expected to give us the best estimate of the routing information from that duration. We loaded each advertised BGP prefix and originating AS from the RouteViews data into a trie data structure, commonly used by routers in deciding the next interface to use to forward packets, and performed longest prefix matches on each IP address to determine the AS associated with the address.

With this mapped data, we then calculate the Jaccard similarities of the ASes in the data sets. Let AS_{S_i} be the set of ASes represented in data set S_i . The Jaccard similarity of the two data sets at the AS granularity is then given by $J(AS_{S_i}, AS_{S_j}) = \frac{|AS_{S_i} \cap AS_{S_j}|}{|AS_{S_i} \cup AS_{S_j}|}$. Results for this calculation are shown in Table IV. Between all pairs of data sets, there is much more similarity with regards to ASes than there was in terms of IP addresses. While the same IP address is not often used for multiple different malicious activities, multiple IP addresses in an AS appear to be used this way more often. Regardless of the type of malicious activity an AS was seen engaged in, the presence of an AS in multiple blacklists could be used as a characteristic to help determine if other later suspicious activities are truly malicious.

III. DEGREE OF AUTONOMOUS SYSTEM MALICIOUSNESS

Using the AS information corresponding to each malicious IP, we examined the extent of AS maliciousness from two

TABLE IV
JACCARD SIMILARITY BETWEEN ASes IN EACH DATA SET

	Bot C&C	CleanMX	eSoft	Local Spam	Malware Patrol	Phishtank	Spamhaus SBL	SI-Feed	SI-DNS	SURBL
APWG	.17	.26	.34	.25	.14	.49	.26	.31	.24	.43
Bot C&C		.18	.17	.15	.16	.18	.14	.16	.15	.14
CleanMX			.35	.25	.27	.17	.22	.22	.20	.20
eSoft				.21	.20	.33	.23	.30	.25	.30
Local Spam					.17	.22	.20	.31	.33	.25
Malware Patrol						.16	.12	.15	.17	.12
Phishtank							.26	.29	.23	.38
Spamhaus SBL								.27	.20	.29
SI-Feed									.58	.33
SI-DNS										.26

perspectives: the percentage of the AS found to be blacklisted and the percentage of the blacklist each AS constitutes. We now describe both approaches and their results in detail. We then examine the temporal behavior of listed machines and the peering relationships of malicious networks.

A. Examination of ASes by Fraction of Advertised IP Space

Given the number of malicious IP addresses associated with an AS, the most straight-forward approach to evaluating the ASes for maliciousness would be to simply order the ASes by the number of malicious IP addresses. However, such an analysis would penalize the larger ASes: they simply have more addresses so they have more hosts that could be compromised and blacklisted. Accordingly we must consider the overall size of the AS as a factor when looking for ASes that are disproportionately bad.

There are no direct sources that help estimate the size of an AS. Even the *whois* database, which contains contact information about ASes in addition to detailed information about domain names and IP addresses, does not contain information about which AS owns which IP prefix. However, the prefixes advertised by an AS can be used to determine the maximum number of IP addresses associated with the AS. While ASes often have unused IP addresses in each of their prefixes, and it is difficult to determine just how many addresses are unused, this allows us to obtain a rough approximation for the AS size, which may be considered an upper bound. We again extracted the prefix and originating AS information from the June 15, 2009 BGP RouteViews routing table. We loaded this information into a trie data structure as before. For each prefix associated with an originating AS, this allowed us to determine the number of IP addresses associated with the prefix. In the process, we were careful to exclude any sub-prefixes associated with other ASes. Such a sub-prefix may exist, for example, if an ISP leases part of its address space to a customer with their own AS. After adding together the address space from each of the prefixes for each AS, we had the total number of IP addresses advertised by each AS.

Next, we determine the rough percentage of each AS that appears in each of our data sets. In Figure 1, we show the percentage of badness for each AS present in our data sets, excluding the Spamhaus XBL data set. We separated out the Spamhaus XBL due to its much larger size which made

the other results difficult to read. This Figure shows several interesting results. First, a total of 31,263 ASes were advertised in our BGP routing data and 46.8% of these had at least one malicious IP in them. While a majority of them have little to no malicious activity, a small number of ASes have as much as 0.5-10% of their IP addresses engaged in maliciousness. In fact, in the SI-Feed data set, one AS had 9.25% of its addresses in the data set. No other AS had 5% or more of its addresses in any of these data sets.

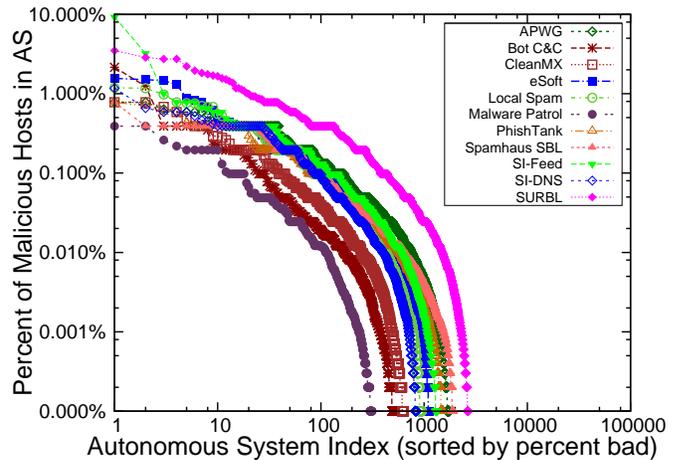


Fig. 1. Percentage of badness for each AS. The AS indices on the x-axis are independent across data sets with different ASes exhibiting the highest percentage of maliciousness in each data set.

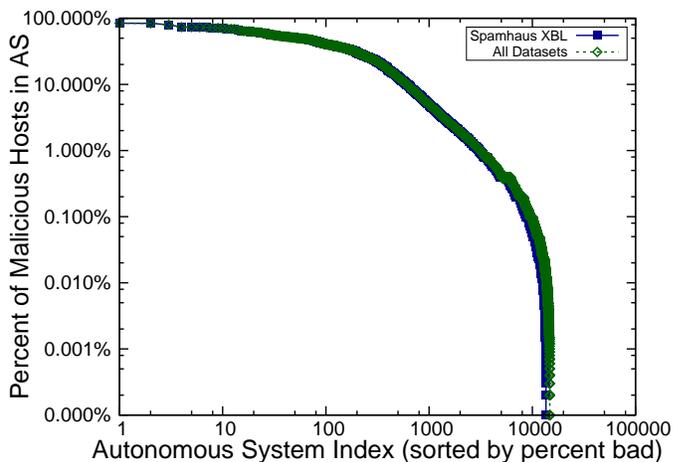


Fig. 2. Percentage of badness for each AS in the Spamhaus XBL blacklist and across all blacklists combined.

In Figure 2, we show the same results for the Spamhaus XBL data set and the combination of each data set. We note that the two lines are very similar and almost completely overlap because of the size of the Spamhaus XBL data set. We found 58 ASes with over 100,000 compromised machines in this data set. Additionally, 255 ASes had between 10,000 and 100,000 machines blacklisted. When looking at the percentage of each AS's advertised address space marked as malicious, we found that *four ISPs, two from Ukraine, one from Iran, and one from Belarus, had at least 80% of their advertised*

IP space blacklisted. Another 49 in the Spamhaus XBL data set had 50-80% of their addresses listed. Further, 556 ASes had at least 10% but less than 50% of their addresses listed. These ASes have a disproportionately high degree of reported malicious behavior, which may be caused by reporting bias, lax security at the AS, or intentional support of cybercrime. In any of these cases, these ASes may warrant greater attention.

B. Examination of ASes by Proportion of Data Set

While examining the percentage of an AS that is blacklisted can highlight ASes with disproportionately high concentrations of blacklisted hosts, it requires large data sets of malicious hosts. While the Spamhaus XBL data set shows this clearly, other data sets are not large enough to distinguish atypically malicious networks. However, rather than consider the AS to be malicious based on the percentage of its blacklisted address space, we can instead examine the percentage of a data set for which an AS accounts. This can be used to highlight ASes with a large number of blacklisted hosts that might not otherwise stand-out due to the size of some blacklists. We note that not all highlighted ASes are equal: in smaller data sets, an AS may be highlighted because of chance with small numbers. However, this approach may find concentrations of malicious activity, even using smaller data sets.

In Table V, we show for each data set the number of ASes containing at least 0.25% of the IP addresses in the data set. However, we wanted to avoid penalizing large ASes that advertise large address spaces and do not necessarily account for a disproportionate amount of maliciousness in that data set. Toward that goal, we perform the following calculations. Let AS_{S_i} be the set of ASes represented in data set S_i , and IP_{S_i} be the set of IP addresses in the data set. For each AS $a_j \in AS_{S_i}$, let IP_{a_j} be the set of IP addresses in the AS (without regards to whether the IP addresses themselves are in the data set). Then the IP addresses we count as malicious are those which satisfy the following two inequalities.

$$\frac{|IP_{a_j} \cap IP_{S_i}|}{|IP_{S_i}|} > .0025$$

$$\frac{|IP_{a_j}|}{\sum_{a_k \in AS_{S_i}} |IP_{a_k}|} * 10 < \frac{|IP_{a_j} \cap IP_{S_i}|}{|IP_{S_i}|}$$

The first of these inequalities simply captures ASes containing at least 0.25% of the IP addresses in the data set. The second ignores ASes where the proportion of the address space advertised by all ASes belonging to the data set advertised by the AS in question is greater than a factor of ten less than its proportion of the IP addresses. For example, if an AS contained exactly 0.25% of the IP addresses in the data set, we would list it if it accounted for less than 0.025% of the address space of all ASes in the data set, but ignore it otherwise.

We can see that some ASes have a high concentrations of malicious activity. Focusing on the top few rows of Table V, we note that several ASes account for more than 6% of blacklisted IP addresses in various data sets. *For example, in*

the Bot C&C data set, we see that one AS contains 9.11% of the IP addresses in the data set, yet its advertised address space represents only 0.002% of the address space advertised by all ASes in the data set. The next AS in this list, with 8.66% of the listed IP addresses represents only 0.006% of the advertised addresses in the listed ASes. These two ASes are a large broadband ISP from Turkey and a hosting service provider from the US. Incidentally, the US-based hosting provider also accounts for 7-8% of all blacklisted IP addresses. Further, in Spamhaus XBL and SI-Feed data sets, we find two more US-based hosting providers that account for over 6-8% of these blacklists.

Overall, our results show that a small number of ASes have a disproportionately high fraction of reported malicious hosts. These ASes could warrant more attention, such as the investigations of Atrivo or McColo [1], [2]. We believe that legitimate ISPs with disproportionately high malicious activity need to provide tighter account controls, particularly in the case of hosting providers, or seek opportunities to provide anti-virus or firewalling services to prevent malicious activity.

C. ASes with Unruly Children

Our data establishes that malicious activity is often disproportionately clustered by AS. We now look at whether ASes with disproportionate malicious activity are tightly clustered. We begin by labeling as malicious any AS with at least 1% of its IP addresses appearing in any blacklist, as described in Section III-A. We then examine each of the BGP updates for June 2009 to find provider-customer (or parent-child) relationships. Given two adjacent ASes, we infer which one is the parent by examining the degrees of the two ASes, similar to the algorithm described by Gao [15]. We consider the AS with largest degree to be the provider.

For each provider AS, we consider the extent to which its customer ASes have been found to be malicious. In the second column of Table VI, we show the number of provider ASes with at least three children that have the indicated percentage of its children as malicious. *We see 22 ASes with 100% of their customers classified as malicious. A total of 194 providers have at least 50% malicious customer ASes.* In comparison, random links between ASes would result in an AS having an average of less than 10% of its links to a malicious AS, since only 3,082 of the 32,193 ASes were labeled malicious.

We repeated this analysis using the definition of maliciousness from Section III-B: the AS must have at least 0.25% of the malicious IP addresses in a data set. We show these results in the third column of Table VI. *Five providers have at least 50% of their customer ASes labeled as malicious.*

This analysis shows that there are dense clusters of malicious activity in the Internet. Accordingly, efforts to systematically reduce malicious activity, via regulation or other means, could have a substantial impact by targeting a small number of networks.

IV. AUTONOMOUS SYSTEM CHARACTERISTICS

Having examined the degree of AS malicious behavior, we now search for other characteristics that differ between malicious and benign ASes. Specifically, we compare ASes where

TABLE V
NUMBER OF ASes IN EACH DATA SET CONTAINING THE GIVEN PERCENTAGE OF ALL IP ADDRESSES IN THE DATA SET.

Percent of IPs in data set	All Sets	APWG	Bot C&C	CleanMX	eSoft	Local Spam	Malware Patrol	PhishTank	Spamhaus SBL	Spamhaus XBL	SI-Feed	SI-DNS	SURBL
$\geq 10\%$													
[9%, 10%)			1										
[8%, 9%)			1										
[7%, 8%)	1									1			
[6%, 7%)											1		
[5%, 6%)					1			1					
[4%, 5%)	1	1	1	2						1	1		1
[3%, 4%)	1					3	1			1		1	2
[2%, 3%)	3	2	2	2	3	2	1	1		3	1	2	
[1%, 2%)	7	5	5	3	7	11	6	3		7	5	10	8
[0.50%, 1%)	16	12	10	16	6	19	16	11		16	20	19	14
[0.25%, 0.50%)	19	20	26	27	25	20	18	18	18	18	27	33	38

TABLE VI
PERCENTAGE OF MALICIOUS CUSTOMER ASes FOR PROVIDERS WITH MORE THAN THREE CUSTOMERS.

Percent of Malicious Customer ASes	Number of Provider ASes	
	Fraction of Advertised IP Space	Proportion of Data Set
100%	22	
[90%, 100%)	2	
[80%, 90%)	8	
[70%, 80%)	17	
[60%, 70%)	72	3
[50%, 60%)	73	2
[40%, 50%)	78	5
[30%, 40%)	202	24
[20%, 30%)	239	45
[10%, 20%)	204	78

we have not observed any malicious IP addresses (good ASes), ASes where we have seen at least one malicious IP address, ASes which have at least 1% of their IP addresses in one of our malicious data sets, and ASes representing at least 0.25% of a blacklist as described in Sections III-A and III-B. For ASes falling in these categories, we compare BGP behavior, AS size, and their connectivity. ASes can be disproportionately malicious for several reasons, such as malicious intent by the operator of the AS, or just lax administration practices. Therefore, we do not expect all malicious ASes to have the same properties as each other or for there to be no overlap with good ASes. However, we do hope to see trends in the characteristics of malicious ASes.

A. BGP Behavior

In order to examine BGP behavior, we begin with the earliest BGP routing table available from the RouteViews project for June 1, 2009. We then replay in order all of the BGP updates for the month of June, examining how routes change in the updates.

We begin by examining routing changes that result in any AS which originates a prefix becoming completely unreachable. We consider an AS to become unreachable when all of the routes to all of the prefixes originated by that AS have been withdrawn according to all of the routers that peer with RouteViews. In total, 5,069 ASes become unreachable at some point in the month. This is 15.7% of the 32,193 total ASes we ever see originating a route.

In our data sets of malicious activity, we observed IP addresses from 14,807 ASes. Of these, 2,319 become unreachable at some point. This is the same percentage, 15.7%, that became unreachable when examining all ASes. It appears that the chances of becoming completely disconnected or unreachable is not affected by small degrees of maliciousness. However, looking at just those ASes where 1% of their IP addresses have been marked as bad, we see that 24.4% become unreachable. *ASes with the most malicious activity appear to be disconnected more often than others.* However, among the ASes which make up at least 0.25% of the malicious IP addresses in their data sets, only 8 (3.0%) ever become unreachable.

Many of the ASes which become unreachable do not stay that way for long. We now look at if how long they are unreachable is dependent on the degree of maliciousness of the AS. Figure 3 shows the duration of time ASes in each category become unreachable, except for those making up at least 0.25% of malicious IP addresses in a data set, which we exclude from this figure due to the low number of data points. Some become unreachable multiple times for short durations; however, the time plotted in this figure represents the aggregate for each AS. Timestamps on the BGP updates are at a resolution of one second, so when an AS becomes unreachable for less than one second, we count it as becoming unreachable but do not add time for this period.

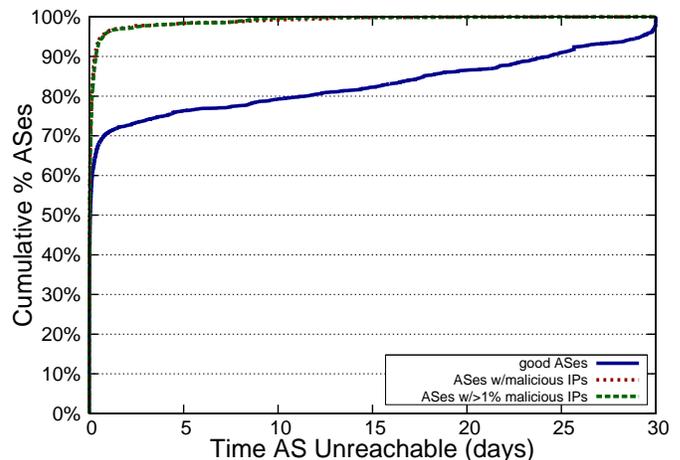


Fig. 3. Unreachability duration for good and bad ASes which become unreachable during our data period.

We see a significant difference among our categories here. 96% of malicious ASes are disconnected for less than a single day, with a similar number for ASes with 1% bad IP addresses. Correspondingly, only 71% for the ASes not identified as malicious become disconnected for less than a day. On the high end, while 45.7% of ASes which become unreachable have malicious behaviors, just 1% of those unreachable for more than 2 weeks have malicious behaviors. *When malicious ASes become unreachable, they do not tend to stay that way for long.* If these disconnections are intentional de-peering, the approach is not effective at isolating the AS for long.

The results for the length of time an AS becomes unreachable were opposite of what we initially expected. To examine routing behavior in further detail, we now consider all connectivity changes to ASes which originate a route (gaining or losing a peer), not just those which change its overall reachability. Of all ASes originating a prefix, 17,286 (53.7%) have some change during our data period. For malicious ASes, this is 8,695 (58.7%), and for those with at least 1% malicious IP addresses, this is 2,036 (66.1%). For those making up at least .25% of one of our data sets, this is 166 (60.9%). Malicious behavior in an AS is clearly associated with routing instability; however, this could be the result of other factors and not simply the malicious activity.

The presence of connectivity changes may be due to problems with the other peer involved in the connection. This is less likely to be the explanation for such changes if an AS had such changes in its relationships with more than a single peer. Figure 4 shows the number of peers involved in connectivity changes with each AS that had such changes. Among good ASes, only 36% with changes had connectivity changes with multiple peers. However, among bad ASes, this is much higher: 50% had a change in relation to more than one peer. This was similar for those with more than 1% bad IP addresses, but was worse for those ASes making up at least 0.25% of their data set. For these, 70% changed in relation to more than a single peer.

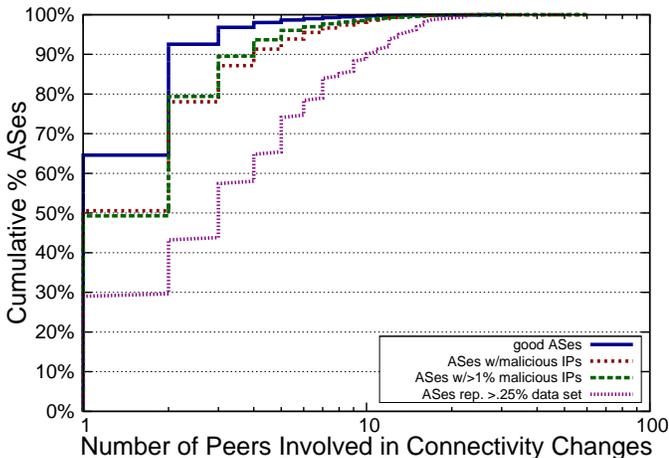


Fig. 4. Number of peers involved in connectivity changes for each origin AS with such changes in our data period.

Similarly, Figure 5 shows the total number of connectivity changes. Among good ASes, 75% of those with changes had

10 or fewer total changes, while this was only 62% for bad ASes and 45% for bad ASes representing 0.25% of their data set. *Overall, among those with changes, ASes harboring malicious behavior have a greater number of connectivity changes than good ASes, and these changes involve more of their peers.*

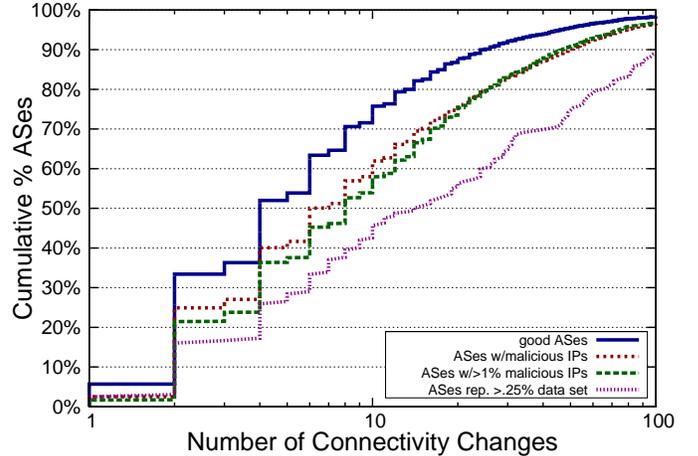


Fig. 5. Number of connectivity changes for each origin AS with such changes in our data period.

B. AS Sizes

We now investigate whether bad ASes have differing sizes than good ones, to see if either larger ASes or smaller ASes have a greater tendency towards malicious behaviors. For each AS, we use the BGP routing table from June 15 to determine the size of the AS based on the size of the prefixes they advertise. Results are plotted for our four categories in Figure 6.

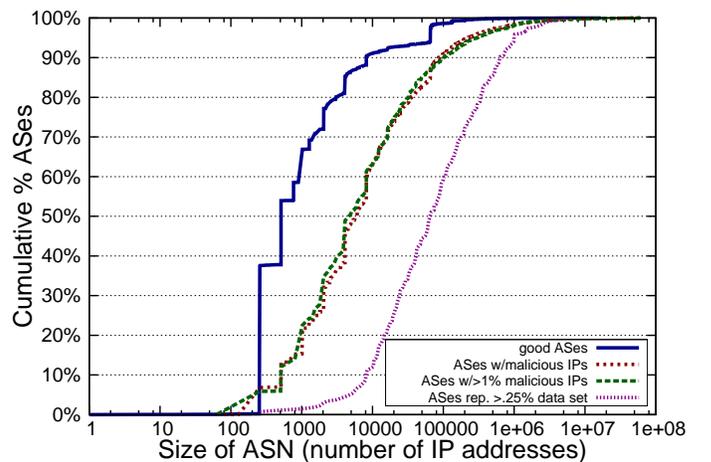


Fig. 6. Sizes of ASes containing or not containing malicious IP addresses in our blacklists.

We see significant differences between sizes of good ASes and those containing malicious IP addresses. While the median size for a good AS is 512 IP addresses, the median for ASes with any malicious IP addresses at all and those with more than

1% of their IP addresses malicious is an order of magnitude larger, and the median for those that represented more than 0.25% of a data set is yet another order of magnitude larger. Similarly, while 67% of ASes without malicious IP addresses have 1024 or fewer IP addresses, this is only 22% for those containing malicious IP addresses, and 1.5% for those that made up at least 0.25% of a data set.

This result is somewhat expected. The more addresses in an AS, the more likely at least one will be compromised. However, the plot for those with more than 1% of their addresses marked as malicious closely follows the plot for those with any malicious addresses at all. This is unexpected because larger ASes would need more total IP addresses to be malicious to end up in this category. *Overall, it appears that larger ASes are more likely to contain malicious addresses.*

C. Degree of AS Peering

We now look at the degree of each AS, which is the number of other ASes with which it directly connects. In Figure 7, we show the degrees of ASes containing or not containing blacklisted IPs. We see that *ASes with malicious IP addresses are more likely to have a higher degree.* Both have a median degree between 1 and 2 indicating that a large portion of both are stub ASes. However, 99% of good ASes have a degree of 10 or less, while 91% for ASes with at least one malicious host have that degree. Further, only 65% for ASes with at least 0.25% of the malicious IP addresses in a data set have a degree of 10 or less. In general, ASes harboring malicious traffic appear to have good connectivity, which may affect efforts to isolate these systems.

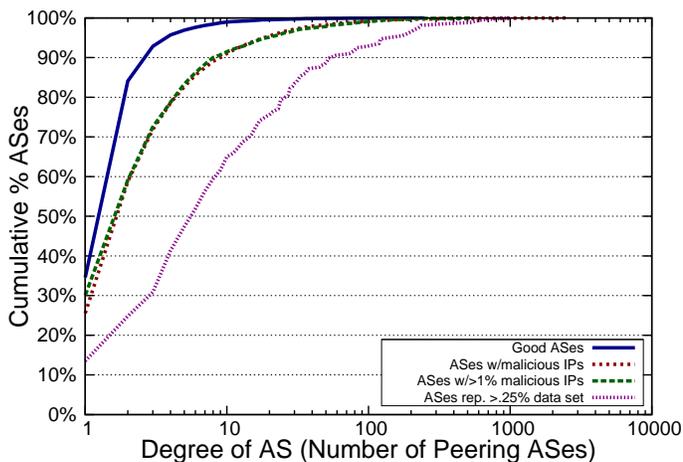


Fig. 7. Degrees of ASes containing or not containing malicious IP addresses in our blacklists.

V. RELATED WORK

This work focuses on finding the ASes harboring malicious activity on the Internet, and investigating the behaviors of and connections between those ASes. Accordingly, the related work falls into two broad categories, work which examines AS topology, and work which attempts to characterize the location of malicious behaviors.

A. AS Topology

Numerous studies have focused on accurately determining types of AS relationships, including those by Di Battista *et al.* [16], Dimitropoulos *et al.* [17], Gao [15], and Subramanian *et al.* [18]. Where we deal with connections between ASes, we are most concerned just with if a malicious AS is related to other malicious ones. Therefore to infer the type of relationship, we use a simple algorithm similar to the one Gao describes as her basic algorithm.

Rexford *et al.* [19] examined BGP routing stability for the ASes of popular destinations on the Internet. They found that most BGP instability was from unpopular destinations and that popular destinations had more stable routes. Work by Feldmann *et al.* [20] identifies ASes which cause routing changes. In our work, we find that ASes containing malicious IP addresses have disproportionately high routing changes.

Other work deals with malicious activities in the BGP system itself. One such activity is prefix hijacking, where a prefix is stolen by advertising a false route. Several papers, including work by Zhang *et al.* [21] and many others, seek to identify hijacking attempts. Zhang *et al.* also have work on defending against hijacking attempts [22]. Several other systems, such as SBGP [23] and soBGP [24], also prevent prefix hijacking. While we recognize the importance of mitigating such malicious behaviors, we focus on malicious hosts and their originating AS and AS behaviors, not attacks against BGP itself.

B. Locating malicious behaviors

Our work is directly motivated by the disconnections of ASes belonging to the ISP Atrivo [1] and the Web hosting provider McColo [2], [25] in 2008 by their upstream providers, as well as the 2009 FTC-ordered shutdown of Pricewert [3]. In all three cases, the networks were accused of large amounts of botnet activity, malware hosting, and spamming. While these three attracted enough attention for high profile action against them and coverage by much of the technology news media, we wanted to see to what extent malicious activity clustered together in other ASes which have not received so much attention or drastic action.

Some previous works attempt to locate malicious behavior at granularities other than ASes. In their study of spyware, Moshchuk *et al.* [26] find that certain categories of Web sites contain more spyware than others. Similarly, work by Provos *et al.* [27] finds that 67% of malware download sites in drive-by downloads are hosted in a single country: China. While there is insight to be gained by examination at these other granularities, we focus solely on the AS location of malicious behavior in the paper.

Other work touches on AS locations of malicious behaviors on the Internet. In a paper on spammers' behaviors, Ramachandran *et al.* [28] find that a small number of ASes are responsible for sending a large amount of spam, with 36% of all spam coming from just 20 ASes. Konte *et al.* [29] examined scam hosting infrastructure. Among their findings was that for the spam campaigns they examined there was almost no overlap in the ASes of the spamming machines and

the ASes where the scam Web sites were hosted. However, none of these papers has the AS locations of the behavior as their main focus as we do, and none go on to examine BGP behaviors of those ASes identified.

The most closely related work, that by Stone-Gross *et al.* [30], analyzes ISP networks based on their degree of involvement in malicious activity. The authors use one of our blacklists, from PhishTank, as well as other data sources, including botnet communication and drive-by-download activity, for their analysis. Our work augments this work by analyzing an additional data sources of malicious activity and examining the BGP behavior of the ASes reported as malicious.

VI. DISCUSSION

In this Section, we describe the inherent limitations of using blacklists for research of similar nature to ours and describe alternative approaches that could be used. We then discuss how this research could be used in practice.

A. Limitations of Blacklisting

In this work, we used blacklists as ground truth for whether a machine was engaged in malicious activity or not. However, this is not the purpose for which these blacklists were created: the blacklists are designed to allow other organizations to prevent possibly malicious traffic from arriving at their infrastructure. This is different from our goal of providing a real-time feed of malicious activity that can be used to assess an AS. As a result, there are a number of factors that can affect these results:

- **Blacklist Administration:** Some blacklists use an “automatic addition/manual deletion” system, automatically blacklisting any reported IP addresses, with network administrators being required to manually remove entries from the blacklists. These administrators may not be aware of the blacklisting or simply choose not to remove their machines from the list even after they are reported. In practice, these lists may see many additions with few removals, causing entries from long-ago attacks to remain in the lists even after the termination of the malicious activity and penalizing clean machines.
- **Reporting Bias:** Some destinations may report attacks to blacklists, causing the attacking systems to be listed. Other destinations may choose not to report attacks. Accordingly, the reporting may be biased, causing systems and networks attacking the destinations that report to be identified while others may attack freely without being listed. This may introduce biases in the conclusions drawn through our research.
- **DNS Resolutions:** Some blacklists provide a list of IP addresses while others provide host names. To obtain a consistent data set, these host names must be resolved to their IP addresses. However, the original reporter of the IP address may perform the DNS resolution and obtain different IP addresses from what is resolved later by researchers. This can be due to temporal changes in addresses, such as those present in fast flux [31], [32], or responses tailored to the resolver’s location. Even with

regular resolutions from a large number of vantage points, it is impossible to be certain that the same IP address was examined that was reported as malicious.

- **DHCP/NAT Effects:** Some networks may have a single machine that is compromised and engaging in attacks on the Internet, but due to DHCP leasing, it may have a larger number of IP addresses from which it has attacked. In our analysis, this machine may be counted multiple times, causing the associated AS to appear more malicious. Likewise, NAT allows multiple machines to share the same address. Accordingly, multiple malicious machines would appear to be only a single compromised machine if they were translated to the same address. Further, a mixture of malicious and benign machines may be labeled as malicious and multiple legitimate machines would be recorded only as a single legitimate machine. Differences in deployment strategies can greatly amplify or dampen these effects: a short DHCP lease time with a mandatory IP change at lease expiration would have much greater impact than a DHCP environment with long lease times and little IP address churn. Accordingly some networks may be disproportionately penalized.
- **Hijacked IP Space:** Attackers have used IP prefix hijacking to acquire control over legitimate address space, use the addresses to launch attacks, then release the route [33]. If these attacks were reported, the AS whose prefix was hijacked would be penalized, rather than the actual perpetrators.
- **Toxic Assets:** With a decreasing availability of IPv4 address space, IP ranges that were previously assigned but are now vacant are being reused, causing innocent parties to obtain addresses that have been reported as malicious. Such innocent parties may be incorrectly identified as malicious through our research.

B. Alternative Data

With the limitations of using blacklists for assessing AS maliciousness, we explore alternative data that could be analyzed. This attack history data must be collected by the destination or by network routing infrastructure.

Some destinations have banded together to share their attack information to quickly detect industry-wide attacks. A number of organizations have established Information Sharing and Analysis Centers (ISACs) to share information about attacks they encounter [34]. Some of these ISACs operate darknets, or data collection operations on unused IP address space, to detect attacks being randomly targeted and operate honeypots to learn more about attacks. These organizations may have live data feeds that can be used to detect systems engaged in scanning or attacks. However, this data will have reporting biases: 1) these institutions regard security as a critical need and do not represent the Internet as a whole and 2) only some attacks may appear in the data, depending on the collection systems they use. Accordingly, attackers that avoid these systems, or attackers that use techniques that are not easily automatically detected, would not appear in these data sets.

Network and host-based intrusion detection services may collect and aggregate data on attacks and provide them to the

security service vendors to analyze. These vendors can analyze sources of the attack traffic and perform analysis similar to our own. Unfortunately, these vendors must typically carefully guard this information as it may expose sensitive information about their client organizations. This limits the ability of third-party researchers to analyze the data.

Other data can be captured by leveraging and observing the infrastructure used by attackers. In the work by Stone-Gross *et al.* [30], the authors observed botnet command and control channels, allowing them to see members of the botnets in real-time. Such work can facilitate observations about infection duration and cleansing efforts, which is challenging with blacklists. However, attackers regularly change and evolve their command and control infrastructure, requiring researchers to continually infiltrate new attacker infrastructure. Further, this analysis introduces bias: researchers can only observe botnet-based attacks and only attacks in which they have identified the command and control infrastructure. Accordingly, researchers risk missing smaller botnets which may be more region-focused.

While there are a few possible alternative routes for obtaining attack data, there is no perfect data set for analysis. Without a system for widespread, systematic reporting of attacks, it will be impossible to obtain unbiased data to analyze for attacks. However, such a system could be created by regulators to evaluate ISP networks, allowing complaints to be filed. Industry self-regulation strategies, such as those used in organizations like the Better Business Bureau, could also create a system for reporting. However, while creating such a production system, researchers would have to create appropriate trust models to ensure that attackers could not simply pollute the system with false entries.

C. Applications of Our Research

Comparing ASes and their degree of maliciousness can be used in several applications, including public policy, peering preference, and destination prioritization.

Governments have increasingly recognized that critical national assets are exposed to the Internet and that cyber attacks can have profound implications on national operation. Due to the distribution of compromised machines, these nations must address attacks coming from within their borders. Accordingly, regulators may seek to curtail computer attacks; however, mechanisms to evaluate ASes, regulators would be unable to establish baselines for compliance and what constitutes responsible network management. Our approach can provide these metrics.

Alternatively, ISPs may choose to self-regulate to ward off government intervention. To influence others to adopt better security practices, peers may place requirements for controlling the spread of malicious machines in their peering agreements in exchange for lower peering costs. Larger ISPs could pressure their customers to practice better security.

Finally, destination networks can leverage information on AS maliciousness to determine how to prioritize traffic. In the case of bandwidth contention, a destination may prefer traffic from an AS with low malicious activity over a highly

malicious AS since doing so would be more likely to service a legitimate user. Maliciousness scores could also be used in spam filtering; however, this cannot be a sole discriminator since there may be legitimate machines in many highly malicious networks.

VII. CONCLUSION

In this study, we examined whether some networks are safe harbors for malicious activity. We found that several ASes have high concentrations of malicious IP addresses while others represent disproportionately higher malicious activity than their equivalently sized peers. This shows that while botnets are commonly being used to launch attacks, malicious hosts may still be clumped by network providers. In spite of these results, traffic cannot simply be declared malicious based solely on its originating AS even for ASes with the high degree of maliciousness, as this would have extensive collateral damage, penalizing legitimate traffic as well. However, identifying if traffic is coming from ASes known to be malicious can be used as one component to help make such a decision.

Our analysis can be used to help increase ISP accountability and can become a mechanism to combat malicious activity. By providing a comparison with equivalently-sized networks, we can highlight the ASes most in need of attention and which would only offer diminishing returns. This information can also be used in peering agreements to place pressure on ISPs to respond to malicious activity.

ACKNOWLEDGMENTS

We would like to thank the RouteViews project for their extensive publicly available BGP data. We would also like to thank the APWG, eSoft, ShadowServer, Spamhaus, Support Intelligence, SURBL, and Rob Henderson of the IU Computer Science department for access to their lists of malicious IP addresses and URLs, as well as Clean-MX, Malware Patrol, and Phishtank, whose publicly available lists we also used.

This submission was sponsored by a contractor of the United States Government under contract DE-AC05-00OR22725 with the United States Department of Energy and by the National Science Foundation (NSF) under Grant No. CNS-0831988. The United States Government retains, and the publisher, by accepting this submission for publication, acknowledges that the United States Government retains, a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this submission, or allow others to do so, for United States Government purposes. Also, any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] J. Hruska, "Bad seed ISP Atrivo cut off from rest of the Internet," 2008, <http://arstechnica.com/security/news/2008/09/bad-seed-isp-atrivo-cut-off-from-rest-of-the-internet.ars>.
- [2] B. Krebs, "Major source of online scams and spams knocked offline," 2008, http://voices.washingtonpost.com/securityfix/2008/11/major_source_of_online_scams_a.html.

- [3] J. Cheng, "FTC forces hive of scum and villainy ISP offline," 2009, <http://arstechnica.com/tech-policy/news/2009/06/ftc-forces-hive-of-scum-and-villainy-isp-offline.ars>.
- [4] U. of Oregon Advanced Network Technology Center, "Route Views project," <http://www.routeviews.org/>.
- [5] APWG, "Anti-phishing working group," <http://www.antiphishing.org/>.
- [6] OpenDNS, "PhishTank," <http://www.phishtank.com/>.
- [7] Support Intelligence, LLC, <http://www.support-intelligence.com/>.
- [8] SURBL, <http://www.surbl.org/>.
- [9] Spamhaus Project, "Spamhaus block list (SBL)," <http://www.spamhaus.org/sbl/index.lasso>.
- [10] —, "Exploits block list (XBL)," <http://www.spamhaus.org/xbl/index.lasso>.
- [11] NETpilot GmbH, "Viruswatch mailing list," <http://lists.clean-mx.com/cgi-bin/mailman/listinfo/viruswatch>.
- [12] eSoft Inc., <http://www.esoft.com/>.
- [13] Malware Patrol, "Malware block list," <http://www.malwarepatrol.net/lists.shtml>.
- [14] ShadowServer Foundation, <http://www.shadowserver.org/wiki/>.
- [15] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Transactions Of Networking*, vol. 9, no. 6, pp. 733–745, Dec. 2001.
- [16] G. D. Battista, M. Patrignani, and M. Pizzonia, "Computing the types of the relationships between autonomous systems," in *IEEE Conference on Computer Communications (INFOCOM)*, 2003.
- [17] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, kc claffy, and G. Riley, "AS relationships: Inference and validation," *ACM SIGCOMM Computer Communications Review (CCR)*, vol. 37, no. 1, pp. 29–40, Jan. 2007.
- [18] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *IEEE Conference on Computer Communications (INFOCOM)*, 2000.
- [19] J. Rexford, J. Wang, Z. Xiao, and Y. Zhang, "BGP routing stability of popular destinations," in *ACM SIGCOMM Internet Measurement Workshop (IMW)*, 2002.
- [20] A. Feldmann, O. Maennel, Z. M. Mao, A. Berger, and B. Maggs, "Locating Internet routing instabilities," in *ACM SIGCOMM*, 2004.
- [21] Z. Zhang, Y. Zhang, Y. C. Hu, and Z. M. Mao, "iSPY: detecting IP prefix hijacking on my own," in *ACM SIGCOMM*, 2008.
- [22] —, "Practical defenses against BGP prefix hijacking," in *Conference on Future Networking Technologies (CoNext)*, 2007.
- [23] S. Kent, C. Lynn, and K. Seo, "Secure border gateway protocol (S-BGP)," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 4, pp. 582–592, 2000.
- [24] R. White, "Securing BGP through secure origin BGP (soBGP)," *The Internet Protocol Journal*, vol. 6, no. 3, pp. 15–22, 2003.
- [25] J. Hruska, "Spam sees big nosedive as rogue ISP McColo knocked offline," 2008, <http://arstechnica.com/security/news/2008/11/spam-sees-big-nosedive-as-rogue-isp-mccolo-knocked-offline.ars>.
- [26] A. Moushchuk, T. Bragin, S. Gribble, and H. Levy, "A crawler-based study of spyware on the web," in *Network and Distributed System Security Symposium (NDSS)*, 2006.
- [27] N. Provos, P. Mavrommatis, M. Rajab, and F. Monrose, "All your iFRAMEs point to us," in *USENIX Security Symposium*, 2008.
- [28] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," in *ACM SIGCOMM*, 2006.
- [29] M. Konte, N. Feamster, and J. Jung, "Dynamics of online scam hosting infrastructure," in *Passive and Active Measurement Conference*, 2009.
- [30] B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda, "FIRE: Finding rogue networks," in *Annual Computer Security Applications Conference (ACSAC)*, 2009.
- [31] The HoneyNet Project, "Know your enemy: Fast-flux service networks," Whitepaper, Jul. 2007, <http://www.honeynet.org/papers/ff>.
- [32] D. K. McGrath, A. Kalafut, and M. Gupta, "Phishing infrastructure fluxes all the way," *IEEE Security and Privacy Magazine Special Issue on Securing DNS*, September/October 2009.
- [33] X. Hu and Z. M. Mao, "Accurate real-time identification of IP prefix hijacking," in *IEEE Symposium on Security and Privacy*, 2007, pp. 3 – 17.
- [34] (2010) National council of ISACS. [Online]. Available: <http://www.isaccouncil.org/>



Craig A. Shue received the Ph.D in computer science from Indiana University, Bloomington, in 2009.

He is currently a Cyber Security Research Scientist with the Cyberspace Sciences and Information Intelligence Research Group, Oak Ridge National Laboratory, Oak Ridge, TN. His research interests are in online deception, measurements, and Web security.



Andrew J. Kalafut received the Ph.D in computer science from Indiana University, Bloomington, in 2010.

He is currently an Assistant Professor with the School of Computing and Information Systems, Grand Valley State University, Allendale, MI. His research interests are in network measurement and security.



Minaxi Gupta received the Ph.D. degree in computer science from the Georgia Institute of Technology, Atlanta, in 2004.

She is an Associate Professor with the School of Informatics and Computing, Indiana University, Bloomington. Her research interests are in computer networks and security.