A Macroscopic Study of Network Security Threats at the Organizational Level

by

Jing Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science and Engineering)
in the University of Michigan
2015

Doctoral Committee:

Associate Professor Michael Donald Bailey, UIUC, Co-Chair
Professor Mingyan Liu, Co-Chair
Assistant Professor J. Alex Halderman
Professor Farnam Jahanian, Carnegie Mellon University
Professor Brian D. Noble
Professor Nicholas A. Valentino

To my mother Min, my father Jianwei, and my husband Hao.

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my advisors Michael Bailey and Mingyan Liu. During my doctoral studies at the University of Michigan, they have been close advisors to me on personal, academic, and professional levels. I cannot thank Bailey enough for acting as a great mentor and friend and for providing invaluable direction for my research and life. Moreover, Mingyan inspired me with lots of valuable ideas and gave me great support especially after Bailey moved to UIUC.

I want to thank the rest of my doctoral committee, including Alex Halderman, Farnam Jahanian, Brian Nobel, and Nicholas Valentino — all of whom have provided guidance and feedback for my dissertation. I'd like to especially thank Farnam, who guided me through my first steps as a graduate student and continued providing that guidance regarding the direction of my research even after he moved to other positions.

I also want to express my gratitude to many people that made this dissertation possible. I'd like to thank the IT and security staff at the University, including: Paul, Will, Matt, who have been great resources in problem discussions, providing access to data, and operational advice. I am very fortunate to have worked with many experienced researchers outside of the University, who have provided data access and valuable comments throughout my doctoral study: Robin Berthier and William Sanders at University of Illinois at Urbana-Champaign, Manish Karir and Michael Kallitsis with Merit Network, Marc Eisenbarth at Arbor Networks, Paul Royal at Gatech, and Elie Bursztein at Google.

I'd like to thank all of the software faculty that provide a friendly and helpful environment for the doctoral program. And I'd like to thank my friends and colleagues at the University that have not only provided support, but given me an enjoyable graduate school experience: Yunjing Xu, Jakub Czyz, Kee Shen Quah, Zakir Durumeric, Yang Liu, Shirley

Zhe Chen, Zhen Qi, Feng Qian, Xiaoen Ju, and Huan Feng.

Finally, and most importantly, I wish to thank my beloved family. This dissertation is dedicated to my mother, Min, my father, Jianwei, and my husband, Hao. I'd like to thank them for their constant support and encouragement throughout my life.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ABSTRACT

A Macroscopic Study of Network Security Threats at the Organizational Level

by

Jing Zhang

Co-Chairs: Michael Donald Bailey and Mingyan Liu

Defenders of today's network are confronted with a large number of malicious activities such as spam, malware, and denial-of-service attacks. Although many studies have been performed on how to mitigate security threats, the interaction between attackers and defenders is like a game of Whac-a-Mole, in which the security community is chasing after attackers and malicious hosts rather than helping defenders to build systematic defensive solutions. As a complement to these studies that focus on attackers or end hosts, this thesis studies security threats from the perspective of the organization, the central authority that manages and defends a group of end hosts. This perspective provides a balanced position to understand security problems and to deploy and evaluate defensive solutions.

This thesis explores how a macroscopic view of network security from an organization's perspective can be formed to help effectively measure, understand, and mitigate security threats. To realize this goal, we bring together a broad collection of reputation blacklists that cover malicious sources involved in Spam, Phishing/Malware, and active scanning. We first measure the properties of the malicious sources identified by these blacklists and their impact on an organization. We reveal that the malicious sources have a surprisingly high impact on an organization's traffic — about 17% of the organization's traffic is from or to malicious sources. We then aggregate the malicious sources to Internet organizations and characterize the maliciousness of organizations and their evolution over a period of two and half years. We find that the maliciousness of organizations varies greatly: while more than half of the organizations remain “clean”, some organizations have a disproportionally large fraction of their IP addresses involved in malicious activities. We also find that both the average magnitude and the dynamic of maliciousness have increased significantly over the past two and half years. Next, we aim to understand the cause of different maliciousness levels in different organizations. By systematically examining the relationship between eight security mismanagement symptoms and the maliciousness of organizations, we find a strong positive correlation between mismanagement and maliciousness of organizations. Lastly, motivated by the observation that there are organizations that have a significant fraction of their IP addresses involved in malicious activities, we evaluate the tradeoff of one type of mitigation solution at the organization level — network takedowns. Based on a broad set of cost and benefit metrics and tradeoff analysis, we identify hundreds of Internet organizations for whom this analysis shows significant favorable returns, with minor costs when they are shuttered.

Thesis Statement: By forming a macroscopic view of security postures of organizations, it is possible to develop security solutions that measure, understand, and mitigate network malicious activities.

CHAPTER 1

Introduction

Network security problems have drawn great attentions for decades. Although security communities have sought various solutions to minimize the impact of malicious activities, both the scale and sophistication of attacks have increased dramatically in recent years. Every day there are thousands of large distributed denial-of-service attacks [17], tens of thousands of new drive-by download pages [105], hundreds of thousands of new malware samples [26], and billions of spam emails [31]. In no uncertain terms, defenders of today's networks are overwhelmed. Therefore, it is vital to the Internet ecosystem that we develop effective security mechanisms that can measure, understand, detect, and mitigate the malicious threats.

1.1 Perspectives of Network Security Studies

While focusing on different participants in the network security ecosystem, the study of a security threat can be conducted from different perspectives: *the perspective of the attacker*, *the perspective of the organization*, and *the perspective of the end host*. To illustrate these different perspectives, we use the DNS amplification attack as an example. In DNS amplification attacks, attackers utilize open DNS resolvers to flood target hosts with a large number of DNS responses. These attacks are innately dependent on both widely-distributed misconfigured open DNS resolvers and the ability of attackers to forge request packets [52, 20]. Using this example, we can see the approaches researching the attack from each perspectives. The research that understands the attack mechanism and measures the global impact of the attack can be classified as studies from the attacker's perspective [119, 19, 41]. On the other hand, the DDos detection systems deployed at an organization's network edges are a form of detection solution from the organization's perspective [13, 18]. Lastly, the best security practices on open resolver configurations provides mitigation solutions from the end host's perspective [76].

From the example we can see that these three perspectives are equally important in studying security threats. This thesis focuses on the organizational perspective because it provides a balanced tradeoff among breadth, depth, and actionablity. Understanding the attacker's perspective is critical because knowing your enemy is imperative to establishing any effective defensive strategy. The study from the attacker's perspective aims to answer questions such as Who are the attackers? Why are they attacking? and How do they attack? The attacker's perspective typically offers a wide breadth of visibility with an Internet-wide view of security threats. However, it lacks the depth of visibility because it cannot observe the detailed information from end hosts. Another drawback of the attacker's perspective is the lack of actionability, because the observations, although very useful to understand the threats, are hardly translated into solutions that can be deployed on the Internet.

In contrast, the perspective of the end host provides opportunities for deploying security solutions, but has a very limited view on the threats. An end host, which can be a computer or a user, is the basic participant in the security ecosystem. Studies from the perspective of the end host aim to understand the vulnerabilities posed by the end hosts, how those vulnerabilities can be explored, and how to fix them. These studies may be able to inspect deep activities and develop detection and mitigation solutions for a single end host, but they do not provide the global visibility that the attacker's perspective can.

While the attacker's perspective and the end host's perspective present clear tradeoffs in depth, breadth and actionablity, the organization, being between of the Internet and individual end users, provides a balanced vantage point to understand security problems and to deploy and evaluate defensive solutions. Organizations are the central authorities that manage and defend groups of end hosts. Because organizations aggregates a large number of end hosts, the visibility at this organizational level is broader and more persistent than the visibility individual end hosts have. Organizations administer networks and therefore manage and deploy network policies such as operating systems and patches, firewall policies, training of IT personnel, and even end-user education.

Besides the balanced position for studying security threats, it is important to study organizational security as we have seen an increase of targeted attacks, especially those targeting organizations [50]. For example, the recent data breaches, such as those at Target [96], JP Morgan [62], and Home Depot [112], resulted in a huge economic loss and social impact. The JP Morgan Chase attack was believed to have been one of the largest in history, affecting nearly 76 million households [62]. Organizations that poorly-manage security policies are low-hanging fruits for attackers and pose great risk to not only the organizations themselves but also to their customers and to the Internet and society as a whole. While traditional security studies are chasing after attackers and malicious hosts, because these are constantly changing, the fight is like a game of Whac-a-Mole, and the overall network security ecosystem has not been improved. Therefore, it is critical to help organizations to understand their security problems and build systematically defensive solutions.

1.2 Overview of Thesis

This thesis focuses on studying security threats from the organizational perspective. It aims to explore how a macroscopic view of network security from an organization's perspective can be formed to help effectively measure, understand, and mitigate security threats. The study of organizational security can be conducted both internally and externally. While the internal study has deep understanding of the organization, it is hard to scale to multiple organizations because of the heterogeneous environment and confidentially of operational information. To achieve our goal of a macroscopic view of organizational security, we choose the external study as it is possible to get Internet-wide external observations with scanning [80] and other data collection methods. However, our studies may lose the depth in understanding the security postures of individual organizations.

Specifically, we bring together a broad collection of reputation blacklists that cover malicious sources involved in Spam, Phishing/Malware, and active scanning on the Internet. We first analyze the properties of these IP-based reputation lists and their impact on an organization. Then we formed a measure of *organization maliciousness* by aggregating the malicious IP addresses to an organization level. Equipped with the defined organization maliciousness, this thesis demonstrates how this macroscopic view of maliciousness can be useful for security studies, including measuring the evolution of maliciousness and defenses, understanding the causes of maliciousness, and mitigating security threats at the organizational level. While this thesis advocates for the exploration of organization-focused security solutions, it is not aimed at discounting existing host-centric and attacker-centric studies. Rather, these three perspectives are equally important and complement each other.

1.3 Main Contributions

The contributions of this thesis are the following:

- ***Measuring* the characteristics of IP-based reputation blacklists and their impact on an organization's traffic.** Reputation blacklists are a form of reactive policy enforcement in which malicious hosts are actively detected and recorded. Organizations usually use them as real-time feeds so that connections to these hosts can be rejected or carefully inspected. In this thesis, we collected nine to twelve IP-based reputation blacklists that cover Spam, Phishing/Malware, and active scanning sources on the Internet. We find that while the blacklists are stable in size, the IP addresses within the blacklists are highly dynamic, growing between 150% to 500% over a one-week period. And blacklisted IP addresses share affinity for specific geographic distributions (e.g., RIPE and APNIC dominate spam; ARIN and RIPE dominate phishing/malware). Rather than solely focusing on the lists themselves, we analyze the impact of these blacklists on Merit Networks [36] to gain better insight into the interplay between malicious sources and organizations. By tainting traffic whose source or destination IP address is listed, we find a surprisingly high proportion — up to 17% of the collected network traffic at Merit Networks [36] — is tainted by at least one of the blacklists.

- ***Measuring* the longitudinal evolution of maliciousness and defenses at the organizational level.** Equipped with the IP-based blacklists, we first calculate *organization maliciousness*, which is defined as the fraction of the organization's IP addresses that are blacklisted. The results show that organizations' malicious levels vary greatly: while more than half of the organizations remain "clean", some organizations have a disproportionally high fraction of IP space that is identified as "malicious" by the blacklists. By observing the organization maliciousness for two and half years, we then characterize its evolution. The evolution of the organization maliciousness is determined both by external factors, such as the rise of attack activities or botnet takedowns, and by complex incentives and defensive efforts done by the organization. Therefore, characterizing the evolution of organization maliciousness is a fundamental step toward understanding the evolution of Internet crimes and defenses and facilitating preventive solutions. Using the time series of organization maliciousness, we extract metrics that capture the magnitude and dynamic properties. We show that the magnitude of spam and active scanning activities has increased, while phishing-related malicious sources remained unchanged over the past two and half years. The dynamic of scanning activities per AS is higher than that of Spam and phishing-related activities, and has increased significantly over the same time period. When using dynamics as a proxy to obtain some measure of organizations' responses to or defenses against security threats, it indicates that organizations have become faster in responding to security incidents and threats. However, the effectiveness of such defenses is limited and not long-lasting, as the overall magnitude of maliciousness continues increasing.

- ***Understanding* the relationship between mismanagement and maliciousness of organizations.** The large variance in maliciousness levels for different organizations inspires us to explore the causes for such a difference. Anecdotal evidence suggests that mismanaged networks are often taken advantage of for launching external attacks, posing a risk not only to themselves, but to the Internet as a whole (such as the DNS amplification attack [20, 119]). This thesis complements the existing evidence of individual incidents with a macroscopic, systematic study on the relationship between security mismanagement and organization maliciousness. For the purpose of this study, we define mismanagement as the failure to adopt commonly accepted guidelines or policies when administrating and operating networks. By leveraging Internet-scale measurements of eight varied mismanagement symptoms, we show that misconfigured systems and servers are pervasive, including over 27 million open recursive DNS resolvers, 22 thousand open SMTP relays, and 227 thousand DNS resolvers that do not utilize source port randomization, etc. Then, we analyze their relationship to the maliciousness of the network. We find strong positive correlations between organizations' mismanagement and maliciousness, even when controlled by selected social and economic factors.

- ***Mitigating* network security threats with network takedowns.** The finding of "bad" organizations, who consistently have a very large fraction of their IP addresses involved in malicious activities, motivates us to study a particular mitigation solution at the organization level — network takedowns. In response to the growing pressure of malicious activities, the network operator community has, in limited cases, resorted to vigilante justice — shutting off Internet access to networks who are egregiously involved in acts of malice or misbehavior, such as the Russian Business Network [58], Atrivo [15], and McColo [6]. The hotly-contested debate around such actions involves not only legal and ethical issues, but also raises concerns about the ad hoc nature of the decision making. In this thesis, we investigate a principled approach to analyzing network takedowns that explores both the technical justification and feasibility of such takedowns, as well as allows for comparison between various takedown options. We select and apply a broad collection of cost and benefits metrics, including Internet routing, naming, and transit. With the measured costs and benefits associated with certain takedown candidates, a Pareto efficiency is then applied to find the optimal operating points that maximize security benefits while minimizing costs. We apply this methodology to an investigation of the Internet over a period of one year, and we identify hundreds of networks for whom this analysis shows significant favorable returns with minor costs when the networks are shuttered. While providing strong justification for takedowns in individual cases, our analysis shows the existing vigilante network takedown approach will likely only provide modest improvements to the overall health of the global Internet.

1.4 Structure of Thesis

The rest of the thesis is organized into six parts. We characterize the IP-based blacklists and their impact on traffic in Section 2. Our study of organization maliciousness and its longitudinal evolution is presented in Section 3. In Section 4, we explore the relationship between organization maliciousness and mismanagement. Then in Section 5, we evaluate the tradeoffs of network takedowns. In Section 6, we review existing works related to this thesis. Finally, we conclude this thesis and discuss some directions for future work in Section 7.

CHAPTER 2

Characterization of IP-based Reputation Blacklists and Their Impact on An Organization's Traffic

Reputation information is a useful resource for organizations to evaluate and design their security policies. In this thesis, we leverage various IP-based reputation blacklists to characterize the maliciousness of Internet organizations. This chapter measures the properties of these reputation blacklists and their impact on the organization's traffic.

IP-based reputation blacklists are a form of coarse-grained, reputation-based, dynamic policy enforcement in which real-time feeds of malicious hosts are sent to networks so that connections to these hosts may be rejected. We collect reputation blacklists covering three broad categories of attacks: Spam, Phishing/Malware website, and active scanning. Together, these represent the most prevalent attacks in today's Internet [115].

In this chapter, we first analyze the size, timing, and geographic properties of those reputation blacklists. Our findings include:

- While stable in size, the IP addresses in blacklists are highly dynamic, growing between 150% to 500% over a one-week period.
- Classes of reputation blacklists show significant internal entry overlap, but little similarity is seen between classes.
- Reputation blacklists in the same classes share an affinity for specific geographic distributions (e.g., RIPE and APNIC dominate Spam lists; ARIN and RIPE dominate phishing/malware lists).

Then, instead of focusing solely on the blacklists themselves, we analyze their impact on Merit Network [36], a large Internet Service Provider (ISP). By examining what network traffic is tainted by these blacklists, we gain better insight into the utility of these mechanisms and their impact on Internet organizations. Surprisingly, we find a very high proportion, up to 17%, of the collected network traffic is tainted by at least one of our reputation blacklists. In addition, our work indicates that an organizational view of network threats can differ from the global perspective — Merit network only saw traffic to a small portion, between 3% and 51%, of IP addresses within the blacklists.

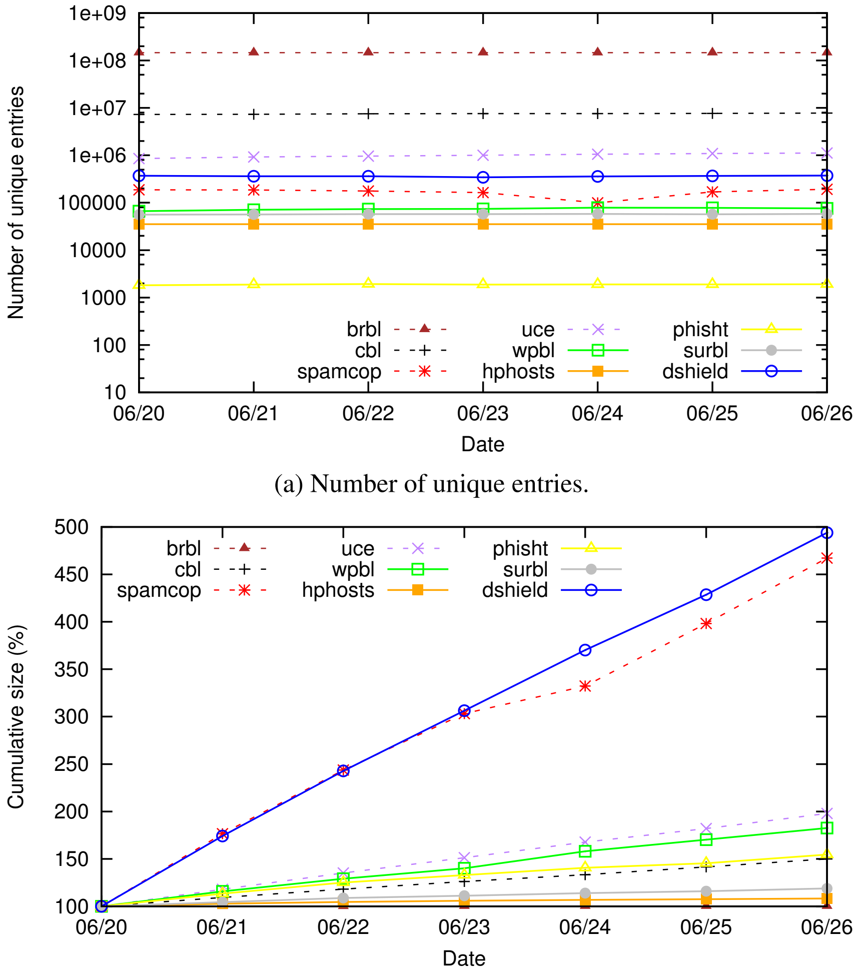
2.1 Data Set

Attack Category	**Blacklists**
Spam	BRBL[10], CBL[12] , SBL[53], SpamCop[47], WPBL[57], UCEPROTECT[55]
Phishing/Malware Website	SURBL[49], PhishTank[40], hpHosts[24]
Active Scanning	Darknet scanners list, Dshield[21], OpenBL[39]

Table 2.1: Blacklists data sources and attack categories.

IP-based reputation blacklists are lists managed by various organizations that contain IP addresses believed to have originated some malicious behavior. Blacklists generally focus on some specific suspicious behavior. Merit collects 12 commonly used blacklists on a daily basis, which are typically fetched directly from the publisher via rsync or wget. Table 2.1 organizes these lists into three broad categories based on the malicious behavior being monitored: spam (unsolicited bulk e-mail), phishing/malware website, or active scanning. Together, these represent the most prevalent attacks in today's Internet [115].

This set of IP-based reputation blacklists is the main data source used in this thesis. In the following chapters, 9-12 blacklists are used depending on the data availability when the work was done.



(a) Number of unique entries.

(b) Relative cumulative size (%).

Figure 2.1: Daily size and cumulative size of blacklists. While stable in size, the IPs in the blacklists are highly dynamic, growing between 150% to 500% over a one week period.

2.2 Characterize IP-based Reputation Blacklists

We first characterize nine IP-based reputation blacklist to answer three questions: how dynamic are the lists? What is the geographical distribution of the malicious IP addresses? And what is the relationship between different lists?

2.2.1 Timing

We examined the stability of each blacklist with respect to *the daily number of unique IP addresses*. As shown in Figure 2.1a, the size varied across blacklists with BRBL being

	Spam					Phishing/Malware			Active
	BRBL	**CBL**	**Spamcop**	**UCE**	**WPBL**	**hpHosts**	**Phisht**	**SURBL**	**Dshield**
AFRINIC	3.02	7.70	5.89	6.37	4.19	0.20	0.58	0.04	2.19
APNIC	25.20	47.14	51.94	48.45	51.27	8.45	11.56	5.58	36.19
ARIN	6.23	1.05	2.53	1.84	6.17	53.32	43.93	54.70	13.54
LACNIC	17.11	16.19	12.15	15.89	10.59	1.66	5.32	1.44	8.54
RIPENCC	48.44	27.93	27.50	27.44	27.77	36.37	38.6	38.24	39.53

Table 2.2: Geographic distribution of IPs for each blacklist (%). Reputation blacklists in the same classes share affinity for specific geographic distributions: RIPE and APNIC dominate SPAM; ARIN and RIPE dominate phishing and malware.

much larger than the others, but the size of blacklists was consistent over the week measured. In order to understand the churn of unique IP addresses, we calculated the relative size of cumulative entries in Figure 2.1b. Spamcop and Dshield updated their entries aggressively, with nearly 500% turnover in one week, while BRBL, hpHosts, and SURBL were relatively static during the week, with less than 110% turnover.

2.2.2 Regional Characteristics

We mapped the blacklisted IP addresses to their registries by using the IP to ASN mapping services provided by Team Cymru [51]. Table 2.2 demonstrates that a given class of blacklists has consistent geographical properties. SPAM- and Active-attack-related lists have more entries in the APNIC (Asia/Pacific) and RIPENCC (Europe) regions, while ARIN (North America) and RIPENCC are the most common regions in Phishing/Malware blacklists. Even though monitoring position and listing methodologies are different for each blacklist, they share consistent views of the regional distribution of malicious activity.

2.2.3 Overlap

We examined to what extent blacklists overlap with other; we expected that overlap within the same category of blacklists would be significantly larger than the overlap among different classes. Our results in Table 2.3 match our expectation: BRBL and CBL, the two largest SPAM blacklists, cover about 90% of other SPAM-related lists, and the intersection within hpHosts, PhishTank, and SURBL is also large. Meanwhile, the overlaps between different classes are trivial.

	Spam					**Phishing/Malware**			**Active**
	BRBL	**CBL**	**Spamcop**	**UCE**	**WPBL**	**hpHosts**	**Phisht**	**SURBL**	**Dshield**
BRBL	100.0	75.2	94.6	89.8	93.8	5.3	10.0	30.7	33.2
CBL	3.9	100.0	98.1	91.7	70.2	0.5	0.7	6.2	9.3
Spamcop	0.1	2.3	100.0	12.6	21.5	0.1	0.1	0.8	1.2
UCE	0.6	12.1	69.4	100.0	50.6	0.3	1.5	1.2	4.8
WPBL	0.0	0.7	8.8	3.7	100.0	0.0	0.2	0.9	0.4
hpHosts	0.0	0.0	0.0	0.0	0.0	100.0	33.7	7.3	0.0
Phisht	0.0	0.0	0.0	0.0	0.0	1.8	100.0	1.7	0.0
SURBL	0.0	0.0	0.3	0.1	0.7	11.8	52.8	100.0	0.1
Dshield	0.1	0.4	2.4	1.8	2.2	0.4	0.7	0.3	100.0

Table 2.3: The average % (of column) overlap between blacklists (row, column). Classes of blacklists show significant internal entry overlap, but little similarity is seen between classes.

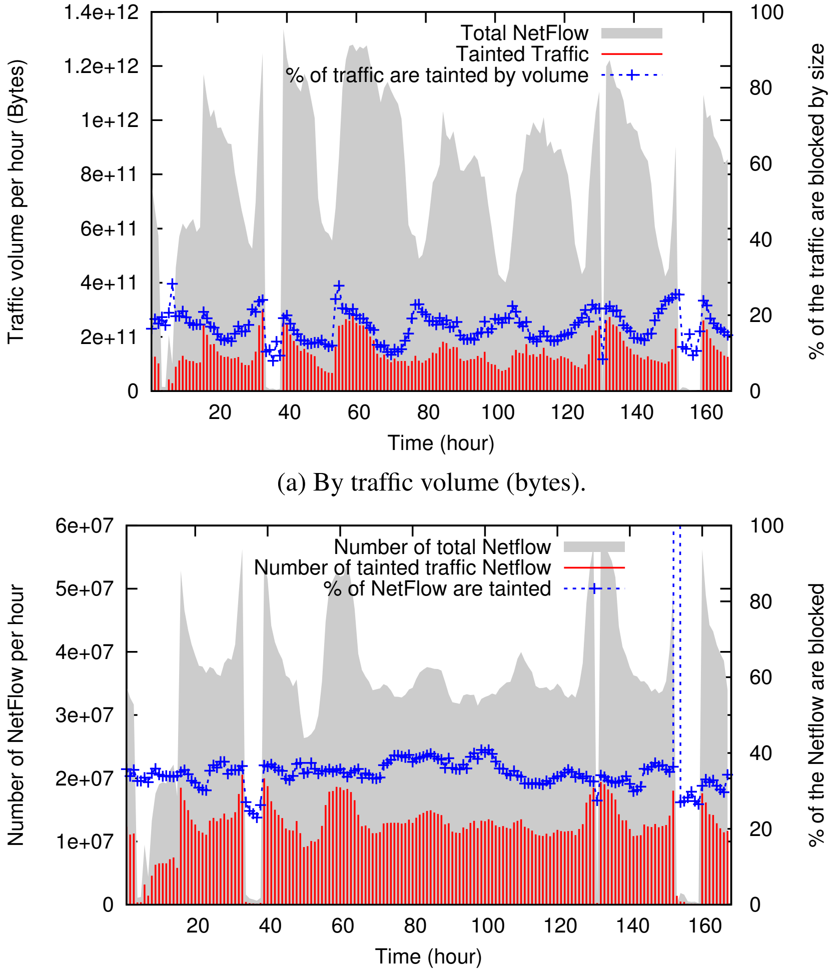
2.3 Impact of IP-based Reputation

One of the key questions we considered in our study was, *what fraction of traffic carries a negative reputation?*

To answer the question, we collected records of the traffic at Merit. Merit is a large regional ISP, which provides high-performance computer networking and related services to educational, government, healthcare, and nonprofitable organizations located primarily in Michigan. This network experiences a load which varies daily from a low of four Gbps to a high of eight Gbps. Though Merit has over 100 customers, the top five make up more than half of the total traffic, and HTTP accounts for more than half of the traffic volume. Our traffic data was collected via NetFlow [32] with a sampling ratio of 1:1. The traffic was monitored at all peering edges of the network for a period of one week, starting on June 20, 2012. During this period, we experienced several collection failures, each lasting from one to seven hours, for a total of 17 hours lost. The collected NetFlow represents 118.4TB of traffic with 5.7 billion flows and 175 billion packets.

We taint the NetFlow if one or both of the collected NetFlow's source and destination IPs are listed by any blacklist. While we expected that perhaps as much as 10% of network traffic might be potentially malicious [29], we found that tainted traffic accounted for an average of 16.9% of the total traffic volume over the week. When measured by flow count, the proportion is even larger, with 39.9% of the flows being tainted (Figure 2.2b).

This is, of course, a very liberal approach to tainted traffic analysis: tainting all the traffic of a host by all the entries in all the blacklists. We conjecture that there may be



(a) By traffic volume (bytes).

(b) By number of flows.

Figure 2.2: Total traffic v.s. tainted traffic. Tainted traffic accounted for an average of 16.9% of the total traffic volume or 39.9% of the flows over the week.

	Spam					**Phishing/Malware**			**Active**
	BRBL	CBL	Spamcop	UCE	WPBL	hpHosts	Phisht	SURBL	Dshield
Touched entries	4,142,394	577,583	44,383	134,024	16,288	13,989	983	14,043	105,918
% of the list	2.8%	7.7%	29.3%	39.5%	51.2%	25.2%	24.4%	13.9%	22.1%

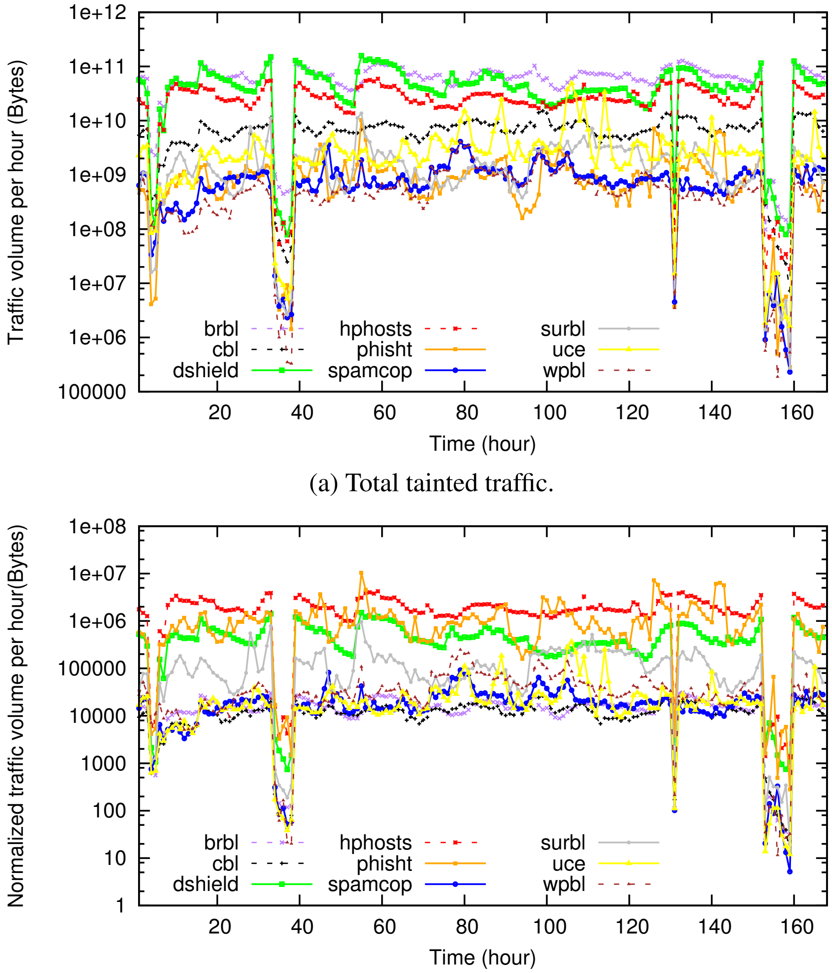
Table 2.4: Blacklist entries touched by our network traffic.

several sources of overestimation: (i) some blacklists are intended to taint only one kind of application traffic instead of an entire host, (ii) the blacklists may contain false positives, (iii) some IP addresses are shared via mechanisms like Network Address Translation (NAT) and therefore some traffic was tainted due to “guilt by association”. To provide a tighter lower bound, we applied the blacklists solely to the type of traffic they pertain to (e.g., SPAM blacklists are only applied to SMTP traffic). The results show that 10.5% of total traffic was tainted by this more conservative approach. Further we observed that several list entries were for well known services on the network, such as Amazon Web Services, Facebook, and CDNs. Although previous work has shown that the cloud services have been used for malicious activities [122], we nevertheless conservatively whitelisted these service providers. As a result, the volume of tainted traffic was reduced to 7.5% of total traffic. Therefore, we believe a realistic value for tainted traffic is likely to lie within the range of 7.5% to 17% of the total traffic by bytes.

Next, we investigated *the potential impact of global reputation blacklists when applied locally.* Prior work in this area has suggested that there might be some entries in global blacklists that are never used by an organization [123], and our results validated this argument. In Table 2.4, we show the average number of daily entries touched for each blacklist. Only a small fraction of entries were touched by our network traffic. For our ISP, only small portions of blacklists are relevant, even though these portions may change over time.

Finally, we examined *whether lists, or a class of lists, have the greatest impact on our traffic.* The traffic volume tainted by each blacklist is shown in Figure 2.3a. There is a clear variance among tainted traffic volumes, ranging from more than ten GB per hour by Dshield, BRBL, and hpHosts to about tens of MB per hour by Spamcop, PhishTank, and SURBL.

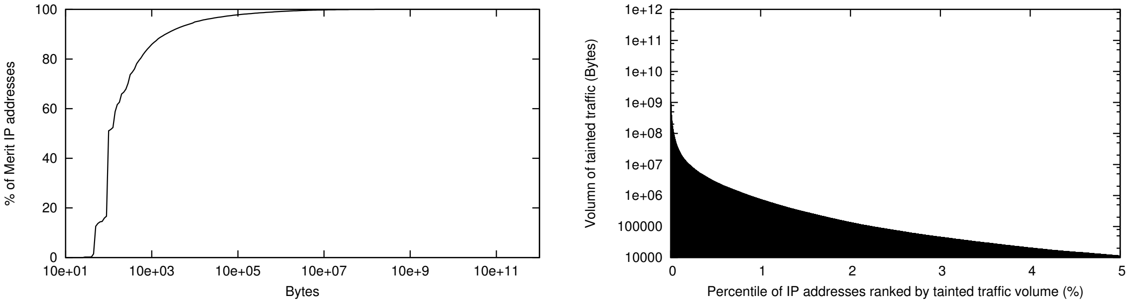
Since the number of entries in each blacklist differs, we then normalized the volume of tainted traffic per hour (i.e. $\frac{Volume\ of\ tainted\ traffic\ by\ the\ blacklist}{Number\ of\ touched\ entries\ in\ the\ blacklist}$) in Figure 2.3b. Interestingly,



(a) Total tainted traffic.

(b) Normalized tainted traffic volume.

Figure 2.3: Tainted traffic per blacklist. Tainted traffic volumes various, ranging from more than ten GB per hour by Dshield, BRBL, and hpHosts to about tens of MB per hour by Spamcop, PhishTank, and SURBL. After normalization, the contribution of entries in the SPAM-related blacklists is about two orders of magnitude lower.



(a) CDF of traffic volume per IP.

(b) Tainted traffic volume of top 5% of IPs.

Figure 2.4: Tainted traffic to/from external IP addresses. The top 50 external IP addresses contributed about 40% of total tainted traffic.

we show that each entry in hpHosts, PhishTank, and Dshield taints about one MB of traffic per hour on average; but, the contribution of entries in the SPAM-related blacklists is about two orders of magnitude lower.

2.4 Impact of Heavy Hitting IPs

In this section, we investigate whether any specific IPs are responsible for skewing the traffic distribution. Toward this end, we divided the traffic into two categories: those IP addresses belonging to Merit (internal IP addresses) and those not belonging to Merit (external IP addresses).

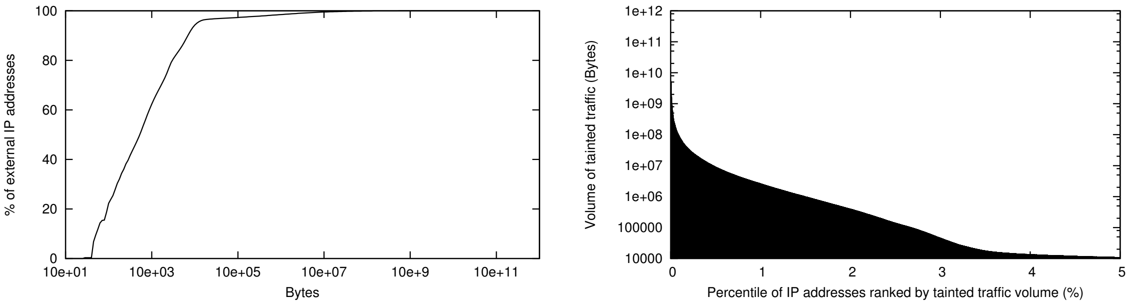
2.4.1 External IP Addresses

Of the 11,016,520 unique external IP addresses in the tainted traffic, 99.5% of them had less than 10 MB of tainted traffic each (as shown in Figure 2.4a). However, the top contributors had more than 100 GB of tainted traffic associated with each of them (Figure 2.4b). In fact, the top 50 external IP addresses contributed about 40% of total tainted traffic. In the following analysis, we try to define *what these hitters are* and *what comprises their traffic.*

External Heavy Hitters Among the top 50 external IP addresses, 39 are listed in at least one blacklist. It is surprising to see that 27 of those are hosting service providers or

Ports	*80*	*443*	*1935*	*1256*	*1509*	*1046*	*1077*	*1224*	*1121*	*1065*
% of volume	60.65	35.31	3.48	1.12	1.06	1.03	0.71	0.66	0.64	0.58

Table 2.5: Distribution over TCP/UDP ports for top blacklisted external IPs.



(a) CDF of traffic volume per IP.

(b) Tainted traffic volume of top 5% IPs.

Figure 2.5: Tainted traffic to/from internal IP addresses. The top 50 internal IP addresses contributed 38% of the total tainted traffic.

caching servers, including Amazon Web Services hosts (10 IPs listed on hpHosts, Phisht, SURBL, or Dshield), Facebook content distribution network (CDN) servers (six IPs listed on Dshield), Pandora media servers (six IPs listed on Dshield), EDGECAST Network hosts (three IPs listed on hpHosts, Phisht, or Dshield), and BOXNET servers (two IPs listed on BRBL). These hosts are owned by popular service providers and their traffic is dominated by HTTP, as shown in Table 2.5.

External Heavy Hitters Not on a reputation blacklist The remaining 11 external IP addresses in the top 50 are IP addresses communicating with tainted Merit hosts, who send large volumes of traffic. Of these external destinations, 10 are owned by Netflix and one belongs to Yahoo!. 99% of the tainted traffic within these 11 IP addresses was over HTTP.

2.4.2 Internal IP Addresses

Analysis of the 2,515,080 Internal IP addresses observed in the tainted traffic also showed the existence of heavy internal hitters (as shown in Figure 2.5). In this case, the top 50 internal IP addresses contributed 38% of the total tainted traffic.

Organization	CDN	EDU				LIB	MED
	Akamai	University	College	Intermediate	Regional		
Num of IPs	9	6	4	1	1	4	4
Total	9	12				4	4

Table 2.6: Organization of blacklisted internal IP addresses.

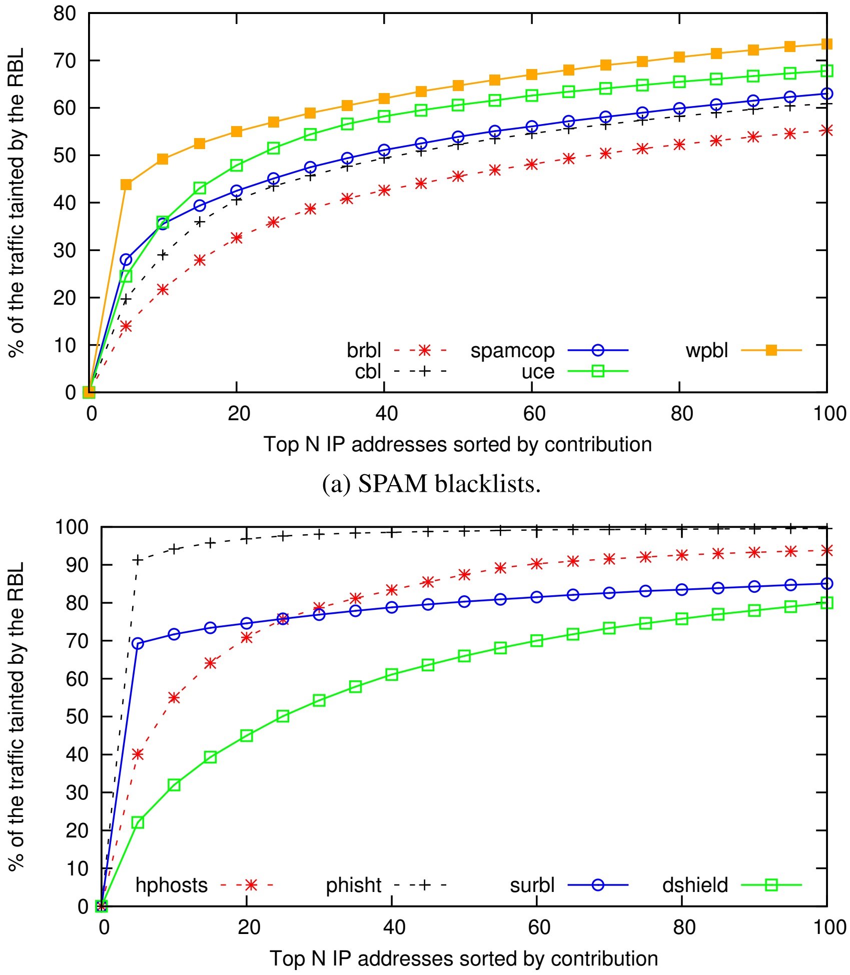
Internal Heavy Hitters Our results showed that there are only 35 IP addresses in the top 50 listed by the Reputation Blacklists, and of the 35 IP addresses, only 29 were resolvable to host names. When categorized by owner (as shown in Table 2.6), we see that nine of these blacklisted IP addresses are owned by Akamai [8], a provider of content delivery network (CDN) and shared hosting services; others are hosts registered by educational institutions, library network providers, and medical centers. Interestingly, there are two Virtual Private Network servers, a mail server, and one web site server from educational institutions.

Internal Heavy Hitters Not on a reputation blacklist We found the top three internal heavy hitters, which accounted for 12% of total tainted traffic, are not themselves on a blacklist, and 81.6% of their traffic is HTTPS traffic. Furthermore, by inspecting the blacklisted hosts they communicated with, we noticed that about 80% of their tainted traffic is to/from Amazon Web Services (AWS) IP addresses that are blacklisted.

2.4.3 Heavy Hitter Distribution

Heavy hitters constitute a significant portion of tainted traffic. *How are these heavy hitters distributed across reputation blacklists?*

To understand the heavy hitters in each reputation blacklist, we defined the contribution of $entry_i$ in $LIST_j$ as $\frac{V_{entry_i}}{V_{LIST_j}}$, where V_{entry_i} is the volume of traffic tainted by $entry_i$ and V_{LIST_j} is the total volume of traffic tainted by $LIST_j$. We then sorted the entries by their contribution in decreasing order for each RBL, and then derived the cumulative contribution of the top N entries (Figure 2.6). The top entries contribute greatly to the blacklists — the traffic tainted by the top 50 entries accounted for more than half of the total tainted traffic of each. In the case of Phishing/Malware blacklists, the top 50 entries contributed even more (80%) of the tainted traffic (as shown in Figure 2.6b). Once again, we find a small amount of entries dominating the tainted traffic.



(a) SPAM blacklists.

(b) Phishing/Malware and Scan blacklists.

Figure 2.6: Cumulative contributions of the top *N* entries per blacklist.

BRBL	**CBL**	**Spamcop**	**UCE**	**WPBL**
80 (59.62)	*80* (34.01)	*80* (26.394)	*3389* (27.03)	*25* (26.71)
443 (22.30)	*443* (21.26)	*44794* (16.51)	*53* (14.16)	*80* (23.30)
1935 (2.22)	*4444* (11.78)	*4025* (16.16)	*25345* (12.80)	*44794* (19.30)
3578 (1.26)	*25* (6.67)	*25* (11.14)	*80* (12.54)	*4025* (18.89)
17391 (1.21)	*3389* (4.96)	*37101* (7.60)	*25* (8.18)	*1080* (9.73)

(a) SPAM.

hpHosts	**Phisht**	**SURBL**
80 (84.99)	*80* (65.05)	*443* (52.30)
443 (15.00)	*443* (32.32)	*80* (44.84)
1256 (1.95)	*49729* (2.96)	*25* (1.85)
1121 (1.10)	*42652* (1.80)	*1288* (1.51)
1605 (1.01)	*52951* (1.48)	*1032* (1.12)

(b) Phishing/Malware.

Dshield
80 (60.75)
443 (32.26)
1935 (3.55)
993 (1.68)
1509 (1.16)

(c) Scan.

Table 2.7: Top TCP/UDP ports for traffic tainted by top 50 contributors per blacklist.

	Spam					**Phishing/Malware**			**Scan**
	BRBL	CBL	Spamcop	UCE	WPBL	hpHosts	Phisht	SURBL	Dshield
CDN	2	0	0	0	0	35	3	1	26
HOST	0	0	1	0	2	3	19	17	12
TOR	1	11	0	0	0	1	0	0	0
MAIL	0	0	0	3	5	0	1	0	1
VPN	3	0	0	1	0	0	0	0	0
Total	**10**	**13**	**1**	**4**	**7**	**39**	**23**	**18**	**39**

Table 2.8: Service hosts in top 50 contributors for each blacklist.

Next, we characterized the tainted traffic by the top 50 contributors for each blacklist (Table 2.7). Though not dominating, SMTP (port 25) traffic occupied a large proportion of the tainted traffic for each of the SPAM related blacklists (except BRBL). This matches our expectation that SPAM related IP addresses send email more aggressively than other hosts. In the other blacklist, we see a higher proportion of Web related traffic. This could be associated with either Phishing and Malware distribution activities or other, potentially benign, traffic from these hosts.

Finally, we looked at the network and domain information of the top contributers (shown in Table 2.8). We found that 60 of these IP addresses are used by content delivery networks and 51 of them are owned by hosting companies. Four VPN servers are listed in BRBL and UCEProtector, while 11 Tor nodes are shown in CBL. Nine different mail servers (some of them belonging to LinkedIn) are also in the top 50 entries of some reputation blacklists. These entries form a sizable fraction of network traffic. This holds especially true for the Phishing/Malware and Active reputation blacklists, whose tainted traffic included from 29% to 68% of these heavy hitters.

2.5 Summary

In this chapter, we first characterized nine reputation blacklists. The blacklists are highly dynamic, growing between 150% to 500% over a one-week period. While there is a significant overlap among blacklists within the same class, little similarity is seen between classes. And geographically, RIPE and APNIC dominate Spam blacklists while ARIN and RIPE dominate phishing/Malware blacklists.

Then we analyzed the impacts of reputation blacklists on traffic from a live operational Internet organization. We demonstrated that up to 17% of the traffic could be considered tainted, as it flowed to or from addresses that were on various blacklists. The surprisingly high fraction of traffic that is tainted indicates that malicious sources identified by the reputation blacklist have great impact on organizations.

CHAPTER 3

Measuring the Longitudinal Evolution of Maliciousness at the Organization-Level

While security ecosystem is complex and can be studied from different perspectives, our study in this chapter is motivated by the desire to better understand security ecosystem within the context of the malicious sources at the organizational level.

An organization's maliciousness and its longitudinal evolution is determined both by external factors such as the raise of attack activities or botnet takedowns and by complex incentives and defensive efforts of the organization. For individual organizations, understanding the threats can facilitate risk assessment and adaptive defensive policies. For example, the activities from an organization with consistently high level of maliciousness should be closely monitored while the activities from a "clean" organization can be whitelisted to utilize limited resources. In addition, by analyzing the trend of its own maliciousness, security administrators can systematically evaluate the effective of current or new defensive policies. From a global perspective, characterizing organizations' malicious activity is a fundamental step toward understanding the evolution of Internet crime and facilitating preventive solutions. What are the maliciousness level of organizations? How malicious activities in the Internet has evolved over time in terms of magnitude and dynamics? Are there differences for different types of attacks? Are there geographical and topological difference in the evolution? Those are questions that we seek to answer in this chapter.

This chapter makes two main contributions. First, we form an organization-level view of malicious sources based on a collection of 11 IP-based reputation blacklists. Traditional reputation is usually generated for individual IP-address or individual host. However, the highly dynamic nature of IP addresses [122] and the large number of IP addresses on host reputation lists can significantly diminish the accuracy and utility of evaluating network security threats. By contrast, one may expect the network of an organization to exhibit more stable and thus more predictable behavior over time, due to the fact that the factors influencing an organization's security posture generally vary on a slower time scale. These factors include various network policy related issues such as operating systems and patch levels, firewall policies, the expertise and training of IT personnel, and even user awareness levels. The notion of organization maliciousness is consistent with the observation of clusters of malicious activity in specific networks [107, 70, 117, 73]. Specifically, we aggregate malicious IP addresses by their autonomous systems or BGP routing prefixes, which are coarse-grained proxies of Internet organizations. We then define the maliciousness of an organization as the fraction of its IP addresses that appear on the blacklists. Then we compare the organization-level maliciousness to IP-based blacklisting and show that aggregation can achieve greater persistence and predictability of entities with high level of maliciousness.

Secondly, we characterize the longitudinal evolution of organization maliciousness in a period of two and half year. We collect the reputation blacklists and form the organization maliciousness for the period of time. Using the resulting time series of organization maliciousness, we then extract metrics capturing the magnitude and dynamic properties. In subsequent analysis we first examine the first order statistics, i.e., how the average magnitude of the maliciousness time series has evolved over time, resulting in a sequence of long-term trends. We then examine the second order statistics (dynamic), i.e., how the frequency of short-term changes in the maliciousness time series has evolved over time. The short-term change in maliciousness may be taken as a proxy for gleaning at how fast and effective an organization responds to or defends against malicious activities. Thus this latter exercise allows us to indirectly measure how organization's response has evolved over time. We find that in the past two and half years, the magnitude of spam and active scanning activities have increased while phishing/malicious websites related malicious sources remained unchanged. The dynamic in scanning activities per AS is higher than that of Spam and phishing/malicious websites related activities, and has increased significantly over the same time period. Given the increasing trend of dynamics of maliciousness, organizations have become faster in responding to security incidents and threats. However the effectiveness of such defenses is limited and not long-lasting as the overall magnitude of maliciousness continues increasing.

3.1 Constructing Organization Maliciousness

3.1.1 Abstraction for Organizations

Ideally, the best aggregation can capture the management boundaries and cluster IP addresses into groups in which IPs in the same group are share similar characters and are managed with the same policies. However, given the confidentiality of operation information, such management boundary is hardly known. To address this challenge, various proxies, including autonomous systems, BGP routing prefixes, DNS administrative domains, and /24 prefix, are used. The more coarser the aggregation, the less dynamic the measure of maliciousness. But one drawbacks of coarse-grain aggregation is that it can also negatively impact individual hosts that become a part of the aggregate. The choice of best aggregation level is out of the scope of this thesis. In this thesis, we mainly adopt the aggregation level of Autonomous System (AS). But our method is applicable to any arbitrary choice of aggregation level.

3.1.2 Construction

We define the maliciousness of an organization as the fraction of its announced IP addresses that appeared in the union of collected blacklists. Specifically, we construct organization maliciousness in three steps.

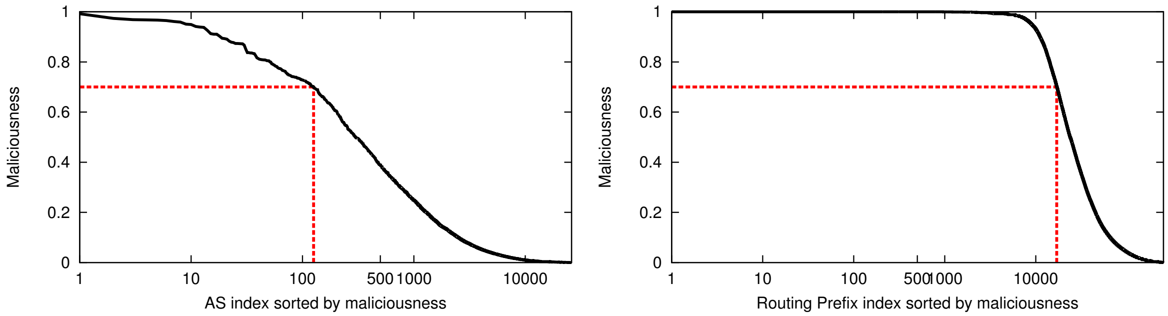
Blacklist Consolidation In this first step, various blacklists are combined into a single list. There are different ways in which blacklists can be combined. The most intuitive method — union, minimizes false negatives. On the other hand, union method results in the highest false positives because even a single listing on the most inaccurate list will result in that IP address being included and given equal weight with union method.

In addition to the Union method, we also explored other combination methods such as Voting (IP in at least seven of 11 lists) and Intersection (IP in at least one list of each attack category). However, these techniques resulted in huge number of false negative as the lists use disparate techniques to capture malicious IP addresses, and have varying levels of confidence in their listing methodology. As a result, these more complicated methods offered little intuitive or technical rationale and therefore we reverted to a simple Union list for the purposes of this study.

As mentioned in Section 2.1, our collection of reputation blacklists cover three category of attacks: spam, phishing/malware, and active scanning. In addition to the *overall union list* across all three types of attacks, we also perform individual analysis on each type of attack. Namely, we union the blacklists under each attack category and get three by attack category union lists, which we refer as *Spam-related union list*, *Phishing-related union list*, and *Scan-related union list* in the following analysis.

Aggregation As discussed above, we choose BGP routing prefix and autonomous system as illustration of aggregation in our study. Equipped with the union list, we use routing information to map IP addresses to BGP prefix and then to autonomous systems. Specifically, we used the BGP tables collected by all vantage points of Route Views [104] and RIPE [42] project. We first match IP addresses to BGP prefixes via longest prefix matching, and then map the BGP prefixes to their origin autonomous systems. While our reputation blacklists are refreshed on a daily basis, we collected the BGP table everyday at midnight and perform the mapping process.

Computing Organization Reputation In the final step, we quantify organization maliciousness of an organization as the fraction of its IP addresses that are blacklisted on the union list. We use an upper bound estimation of the size — all the IP addresses in advertised prefix and the sum of the IP addresses of all its advertised prefixes for an AS. As there



(a) Maliciousness of Autonomous Systems

(b) Maliciousness of routing prefixes

Figure 3.1: An illustration of organization maliciousness at AS and routing prefix levels. Some disreputable organizations have disproportional large fraction of IP addresses blacklisted: More than 70% of their IP addresses are blacklisted for the top 126 ASes and for the top 17k prefixes.

might be unused IP addresses in each netblock, the number of advised IP addresses is the upper bound estimation of the total size of an Internet organization. We note that dynamic IP allocation could cause our measure to overestimate the malicious level of the network as a compromised machine may use different IP addresses. For the reminder of this thesis, we refer the fraction as *organization maliciousness*.

3.1.3 Impact of Abstraction on Persistency and Predictability

As we discussed above, the abstraction of organization would have impacts on the resulted maliciousness. In this section, we examine the difference in the content of the blacklisted IPs in the combined list and the worse prefixes/ASes and how this leads to difference in their security benefits. For simplicity, blocking individual IPs listed on the Union list will be referred to as IP-based blocking, while blocking out entire addresses within a blacklisted AS or prefix will be referred to as AS-based blocking and prefix-based blocking, respectively, or aggregated blocking collectively.

Based on the method described in section 3.1.2, we compute organization maliciousness at both abstractions on October 16, 2012 based on the overall union lists. There are about 10 million malicious IP addresses in the overall union list on that day. While aggregating to Internet organizations, 26,051 autonomous systems (59.3% of all ASes seen in routing

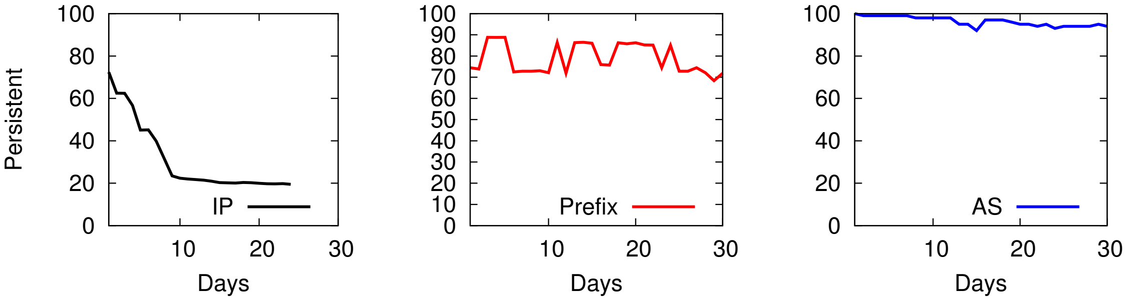


Figure 3.2: Persistence of malicious IP addresses and the disreputable prefixes and ASes. The coarser the aggregation level, the more persistent are the worst entities: only 20% of the IP addresses remain on the list during the one-month period; more than 90% of the disreputable ASes persist in this group, while about 75% prefixes among the disreputable prefixes persist during the same period.

tables on that day) and 253,297 BGP prefixes (45.8% of total routed prefixes) had at least one IP addresses blacklisted in the Union list. Figure 3.1 shows the distribution of organizations' maliciousness under our definition. On the x-axis, we ranked autonomous systems or routing prefixes by their maliciousness and on the y-axis, we show their maliciousness (i.e. the fraction of their advised IP addresses that are listed in the Union list).

It can be seen that there are a set of organizations that have very high level of maliciousness. More than 70% of their IP addresses are blacklisted for the 100 worst ASes. The result is more pronounced at a finer level of aggregation: nearly 100% for the worst 9,000 prefixes, and 70% for the worst 15,000 prefixes. These observations confirm the existence of bad organizations that have disproportional large fraction of IP addresses blacklisted.

Aggregation can dampen the dynamic nature of individual IP addresses and result in more persistent measures of malicious entities. We first compare the *persistence* of malicious IPs and the disreputable ASes/prefixes under our definitions. We take the ASes/prefixes with more than 70% of their IP addresses listed as the sets of disreputable ASes/prefixes. As a result, there are 126 ASes and 17k prefixes in the disreputable AS and disreputable prefix set, respectively. Figure 3.2 shows the malicious IPs (or disreputable ASes/prefixes) on the first day remaining on the Union list on day *x* as a function of *x*. As expected, these disreputable ASes/prefixes are much more persistent than malicious IP addresses, and the coarser the level, the more persistent are the worst entities: only 20% of the IP addresses

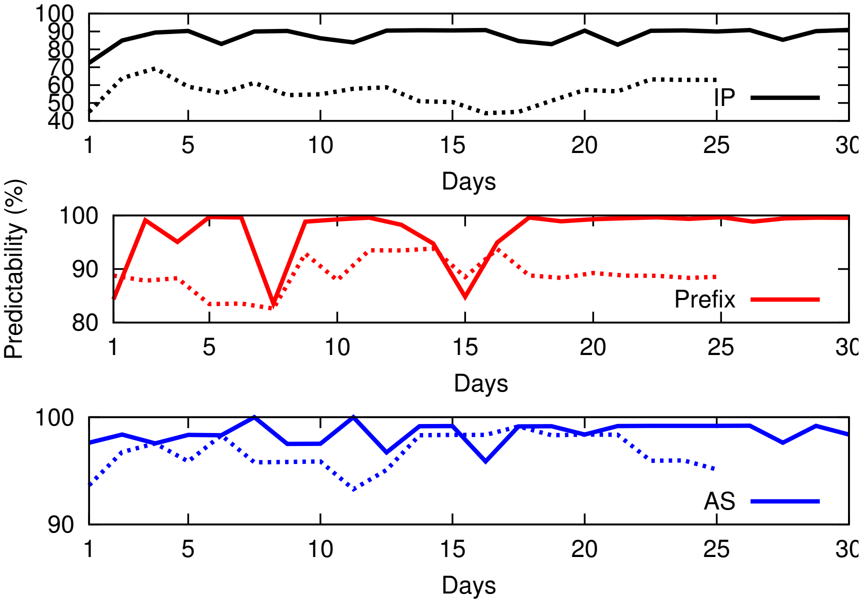


Figure 3.3: The predictability of IP-based blocking and aggregated blocking. The solid line shows the predictability within a time lag of one day, and the dash line five days.

remain on the list during the one-month period; more than 90% of the disreputable ASes persist in this group, while about 75% prefixes among the 17k disreputable prefixes persist during the same period.

The difference in persistency is then reflected in their ability to predict malicious sources. As already pointed out, the IP-based lists are often inaccurate and time-delayed. In another word, the lists published on a given day capture the malicious sources that are active over the past few days, but miss the malicious sources that currently or will be involved in malicious activities. This results in both false positives (some malicious IPs on the list have since been cleaned up) and false negatives (new malicious IPs that have not been captured to show up on the list). Our conjecture is that if aggregation offers more steady measures of malicious entities, then the worst ASes/prefixes do not change as fast as the malicious IPs. Thus aggregated blocking would cancel out some of the false negatives and even predict future malicious sources. We refer this benefit as *predictability*.

We compare the predictability of IP-based reputation and aggregated reputation under the assumption that the lists are a time-delayed version of ground truth. More specifically, the actually malicious sources on a certain day is reflected on the lists published x days later. Then the predictability is the fraction of actually malicious sources (i.e., reflected by the lists after a time lag) that can be captured by what is known on a given day (i.e., the lists on that day). We do not yet have a valid method of verifying precisely how long is this time lag. So the results here use a one-day and five-day time lag as possible examples.

We show the predictability of IP-based and aggregated-based blocking in Figure 3.3. It can be seen that IP-based blocking has 90% predictability with a one-day time lag. However, the it drops quickly to around 50% if the truth is only known after five days (i.e., a five-day time lag). By contrast, aggregated blocking provides more predictive power. Both prefix and AS level aggregation provides 99% predictability with a one-day time lag. More importantly, the predictability remains at high levels over the longer time lag (about 95% for AS-based and 90% for prefix-based blocking with a five-day time lag), and the higher the aggregation level, the higher the predictability.

However, AS/prefix-based blocking results in other types of errors even if the Union list on a given day represents the truth: false positives due to the collateral damage inflicted on those innocent IPs deemed guilty by association, and false negatives because this process ignores those malicious IPs falling outside the worst set of ASes or prefixes.

In the next sections, we characterize the longitudinal evolution of maliciousness at the autonomous system level in terms of magnitude and dynamic.

3.2 Evolution on the Magnitude of Maliciousness

This section aims to answer the following questions. Is the overall malicious level of organizations increasing or decreasing over this period of time? Is there any difference for different types of attacks? What are the regional differences? Do edge networks and service provides have the same trend?

3.2.1 Characterizing Magnitude of Maliciousness

We first describe our methodology in extracting a metric to describe the magnitude of maliciousness. For an organization denoted by Ni, we quantify its daily maliciousness on

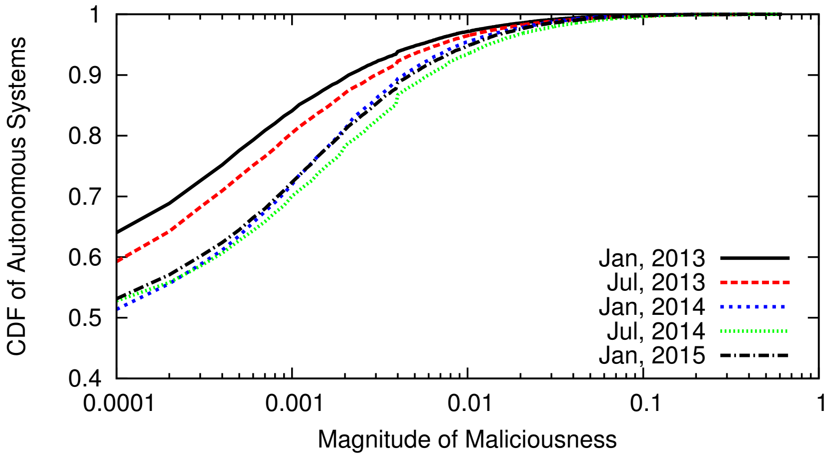


Figure 3.4: An illustration of organization distribution over Spam malicious magnitude.

day t (1 t 31) of any month k (mi,k(t)) as the fraction of its advertised IP addresses that are blacklisted on a union list (i.e. Spam-related union list, Phishing-related union list, or Scan-related union list) on that day. We then combine the daily maliciousness in month k into a time series mi,k = [mi,k (1), mi,k (2),], and the concatenating each month we obtain the complete time series mi,kk=1,2, for organization Ni. This aggregation is done separately for each of the three union lists for the three types of attacks.

We define the metric *malicious magnitude* based on the constructed organization maliciousness time series. The malicious magnitude of an organization is defined as the average of its daily maliciousness in a certain month. Specifically, the magnitude of organization N_i in month k is given by: $\bar{m}_{i,k} = \frac{\sum_{t=1}^{T} m_{i,k}(t)}{T}$, with T being the number of days in month k. We use the average over a month instead of daily maliciousness to reduce possible noises introduced in the data collection.

With the magnitude metric, we show the distribution of ASes based on their malicious magnitude of Spam activities as an illustration. Figure 3.4 shows the CDF of autonomous system over their malicious magnitude in Spam. Overall, we see more than half of autonomous systems are very “clean” with less than 0.01% of their IP addresses listed in spam activities (i.e. magnitude is less than 0.0001). At the same time, about 10% of the organizations are highly malicious — on average more than 1% of their IP addresses were sending Spam (i.e. magnitude is large than 0.01). We also notice that there are tens of

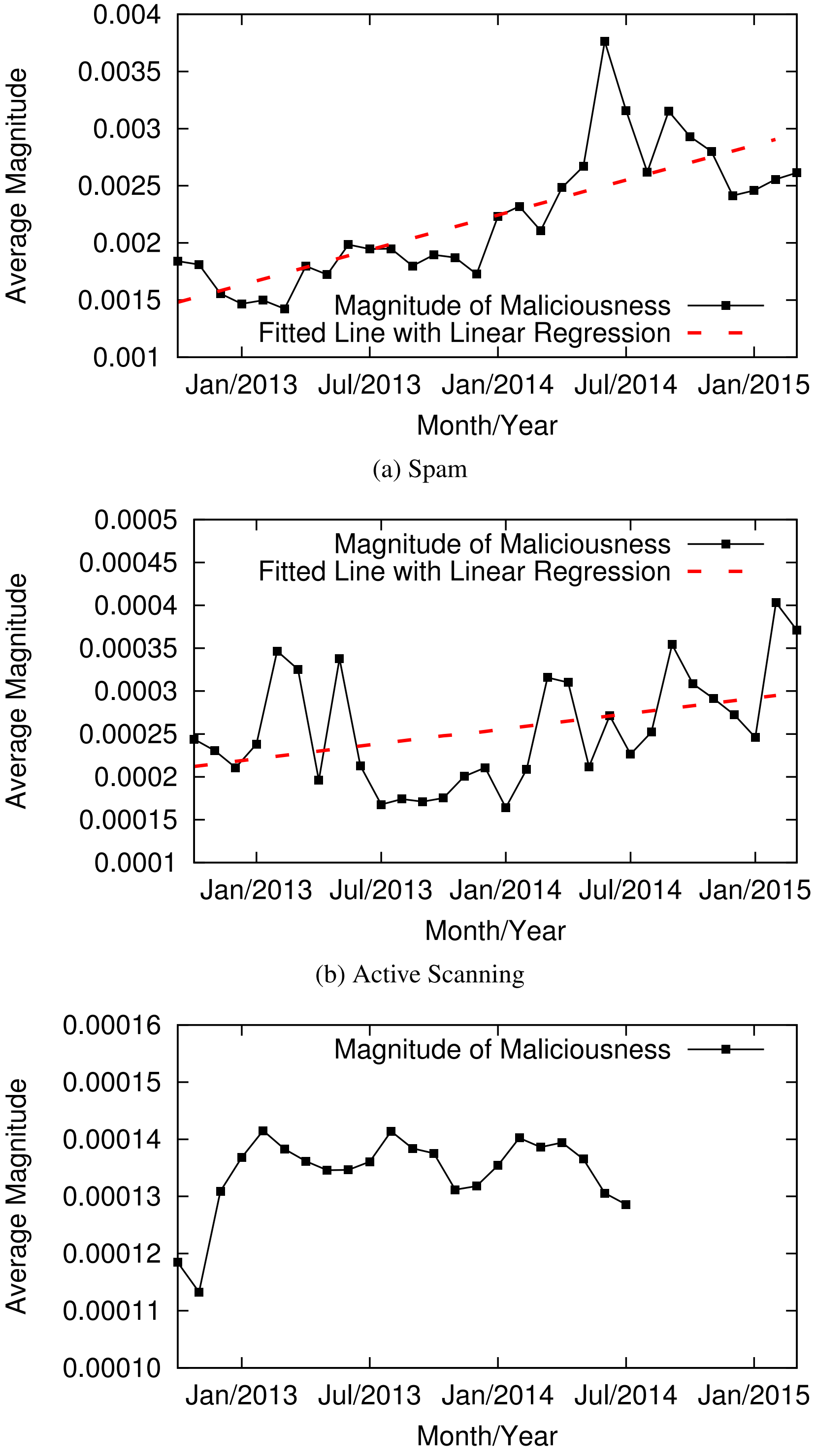
ASes with an average of more than 20% of their IP addresses blacklisted, indicating the existence of “bad” organizations that are responsible for an disproportionate fraction of malicious sources. The distribution is similar for active scanning and phishing/malicious webpage related maliciousness, but at a lesser scale.

3.2.2 Trend in Average Malicious Magnitude

When comparing the distribution of different month in Figure 3.4, one can see that the curve was shifting to a higher magnitude. At the beginning of 2013, about 65% of the autonomous systems were with a magnitude less than 0.0001; but the fraction reduced to 52% in 2014 and 2015. And the ASes whose malicious magnitude was higher than 0.01 have increased from 2.8% to 6.5% in the two years. This observation inspires us to systematically look at the trends of malicious magnitude.

Overall Trend To better capture the overall trend of malicious magnitude, we calculate the malicious magnitude of all autonomous systems in each month and plot the average over time in Figure 3.5. It can be seen that while fluctuating, the magnitude of both Spam and active scanning has been increasing since the end of 2012. By applying linear regression, we obtain the following statistically significant models ($p - value < 0.05$): $\bar{m}_k = 5.1 \cdot 10^{-5} \cdot k + 1.4 \cdot 10^{-3}$ for Spam and $\bar{m}_k = 3.0 \cdot 10^{-6} \cdot k + 2.1 \cdot 10^{-4}$ for active scanning, where $\bar{m}_k$ is the average magnitude over all autonomous systems and k is the time (month).For Phishing/malicious websites, the magnitude experienced an increase at the end of 2012, but has been staying stable since January 2013. And a statistical test also confirms that the magnitude does not depend on the time.

Regional Differences We then examine the regional difference in the trend of malicious magnitude. We calculate the average magnitude of organizations respectively within five different regions and show the results in Figure 3.6. For Spam, AFRINIC has the highest average malicious magnitude while autonomous systems in ARIN have a very small fraction of their IP addresses associated with Spam. Over this period, except for AFRINIC, the other four regions share similar increasing trends. The magnitude of AFRINIC decreased

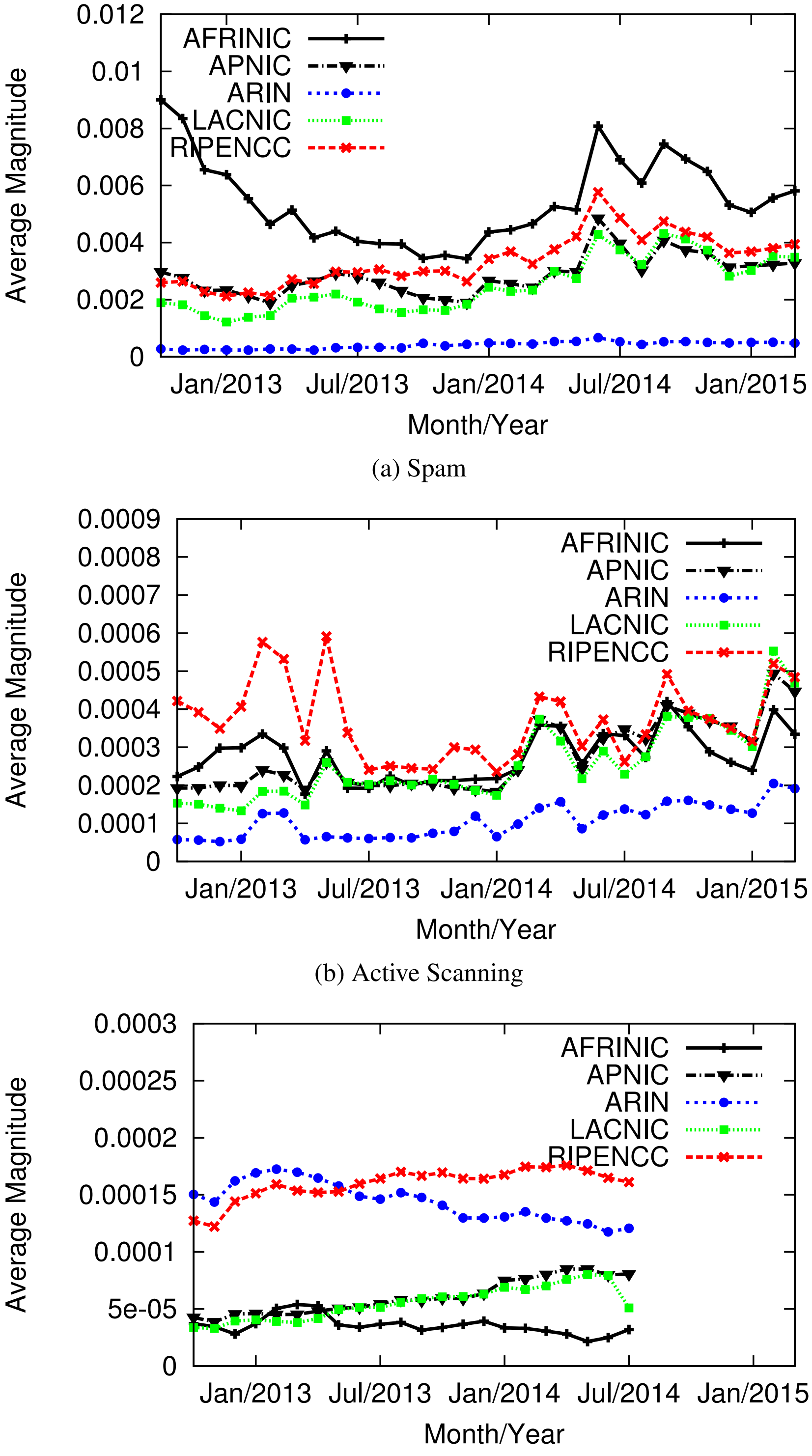


(a) Spam

(b) Active Scanning

(c) Phishing/Malicious websites

Figure 3.5: Evolution of Magnitude of Maliciousness.



(a) Spam

(b) Active Scanning

(c) Phishing/Malicious websites

Figure 3.6: Regional difference in the magnitude of maliciousness.

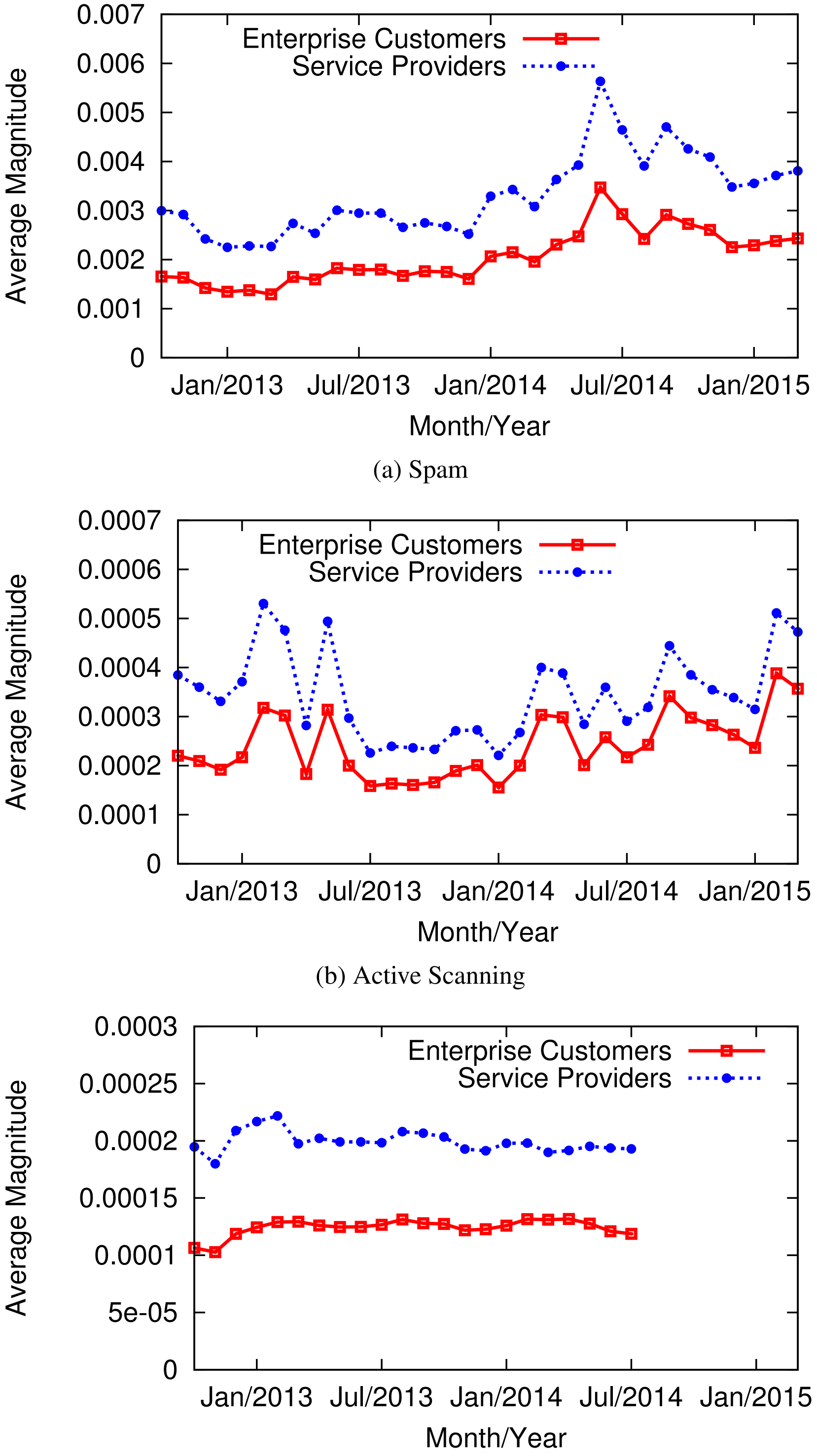
significantly in the year of 2013 and stayed close to the overall malicious magnitude. However, the difference have been increasing again since July 2014.

For Scan activity, ARIN also has the lowest malicious magnitude among the five regions. RIPENCC had a relatively high magnitude (about two times higher than that of the others) at the beginning of our study. But we observe a sharp decrease in the middle of 2013 and since then the magnitude and trend are very similar for the four regions other than ARIN.

Interestingly, AFRINIC, while having the highest malicious magnitude in Spam sources, has the lowest magnitude in phishing/malicious websites related maliciousness. One possible explanation is that end hosts in AFRINIC are poorly protected so that they are easily to be compromised and used as Spam sources. However, for phishing/malicious websites, the attackers tend to utilize networks who have better infrastructures (i.e., high network speed, stable services, etc). This is consistent with the observation that phishing/malicious websites are concentrated in ARIN and RIPENCC, in which web infrastructure and hosting providers are highly concentrated. For long term trend, we do observe a decrease in the magnitude of phishing in ARIN while an increase in that of RIPENCC, APNIC, and LACNIC.

Topological Differences We next examine the question whether organizations in different topological categories have different malicious magnitude. We break the autonomous systems into enterprise customers and service providers by their topological position. The rationale is that these two types of ASes play very different roles on the Internet and thus may be managed in different ways, or one type of ASes may be more attractive to the attackers than the other. For this purpose we use the definition and categorization techniques proposed by Dhamdhere et al [79], with the modification that we combine transit service providers and content/accessing/hosting service providers into a single service provider category in our study.

The results are shown in Figure 3.7 with the average malicious magnitude of enterprise customers and service providers. It is interesting to see that for all three types of attacks, the magnitude of the two types of ASes trend in synchrony; this indicates that attacks or



(a) Spam

(b) Active Scanning

(c) Phishing/Malicious websites

Figure 3.7: Topological difference on the magnitude of maliciousness.

malicious campaigns do not appear to distinguish network types when choosing potential host targets. On the other hand, the magnitude of service providers is consistently (about two times) higher than that of enterprise customers; this might have been caused by the complexity of managing large and diverse networks.

3.3 Dynamics in Maliciousness

In this section we examine the frequency of short-term changes in the malicious magnitude. This short-term change, or dynamics in maliciousness may be taken as a proxy for how fast and effective an organization responds to or defends against malicious activities. This therefore allows us to indirectly measure how organization response has evolved over time. Below we first define how we quantify such dynamics and measure their change for individual organizations. We then break down organizations into different types based on both magnitude and dynamics and analyze the evolution.

3.3.1 Characterizing Dynamics

In order to measure the frequency in change, we will first need to define what constitutes a "change" in the malicious magnitude. Furthermore, such a definition needs to be consistent with our intention of using it as a proxy for measuring organization responses. To do so, we will first quantize the magnitude time series of a particular organization in any month into three regions: "good", "normal" and "bad", on a scale relative to that organization's average within that month.

Specifically, a point $\mathbf{m_{i,k}}(t)$ at t belongs to the "normal" region if $m_{i,k}(t) \in [(1-\delta)\bar{m}_{i,k}, (1+\delta)\bar{m}_{i,k}]$, the "good" region if $m_{i,k}(t) < (1-\delta)\bar{m}_{i,k}$, and the "bad" region if $m_{i,k}(t) > (1+\delta)\bar{m}_{i,k}$, where $0 < \delta < 1$ is a constant and $\bar{m}_{i,k}$ is the average over month k for organization N_i. The parameter choice for δ is obviously not unique and will lead to different quantities in our analysis; however, such a choice should have limited impact on our study of trending over time. We have used the value 0.2 in this study.

This quantization of the magnitude values is motivated by clear phase transitions in the

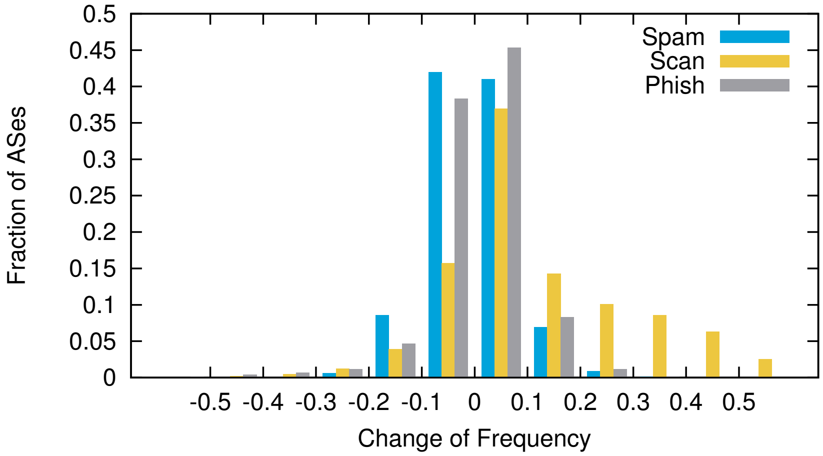


Figure 3.8: Histogram of the change in dynamic.

maliciousness time series; it is plausible that these significant phase transitions correspond to events associated with attacks and responses. For example a "good" region/period may correspond to events that have had certain IP addresses removed from the blacklists due to potential security enhancement efforts. Similarly, "bad" region/periods may correspond to security breaches and host infections. We examine the frequency of regions as they transition into good, bad and normal separately and do not notice any significant differences in the transition into one state or another. We therefore take any such transition, i.e., the time series moving from one region to another, as a change in the magnitude. This is also referred to as the short-term change in the magnitude which is distinguished from the long-term trend observed in the previous section. Our second step consists of calculating the frequency at which the magnitude changes during each month.

3.3.2 Change in Dynamics

We first examine how the malicious dynamics have evolved over the period of our study. We calculate the average frequency for each organization over the first and last six months in this period (Oct 2012 - Mar 2013 and Oct 2014 - Mar 2015, respectively). The difference between these two frequencies is shown in Figure 3.8 as a histogram that presents the fraction of organizations with a certain level of difference. A negative difference indicates that an organization's maliciousness change less frequently at the end of the study period compared to two years ago.

For Spam related maliciousness, we see that most organizations (83%) have a small change falling within [-0.1, 0.1]. The numbers of organizations with positive changes (becoming more dynamic) and those with negative changes (less dynamic) are roughly the same. On the other hand, for Active Scan the distribution is very different: we see a clear trend in organizations becoming more dynamic. A total of 78% of the organizations have an increase in their frequency. The heavy tail on the positive side also indicate that some organizations have significant increases in their dynamics. We observe a similar result for phishing/malicious webpage related maliciousness, but to a less degree: about 55% of the autonomous systems have an increase in their malicious dynamics, among which about 45% is a small increase (less than 0.1).

3.3.3 Dynamics v.s. Magnitude

In the above analysis, we have analyzed the trend in magnitude and dynamics separately, in this section we further examine the relationship between magnitude and dynamics as well as its trending over time.

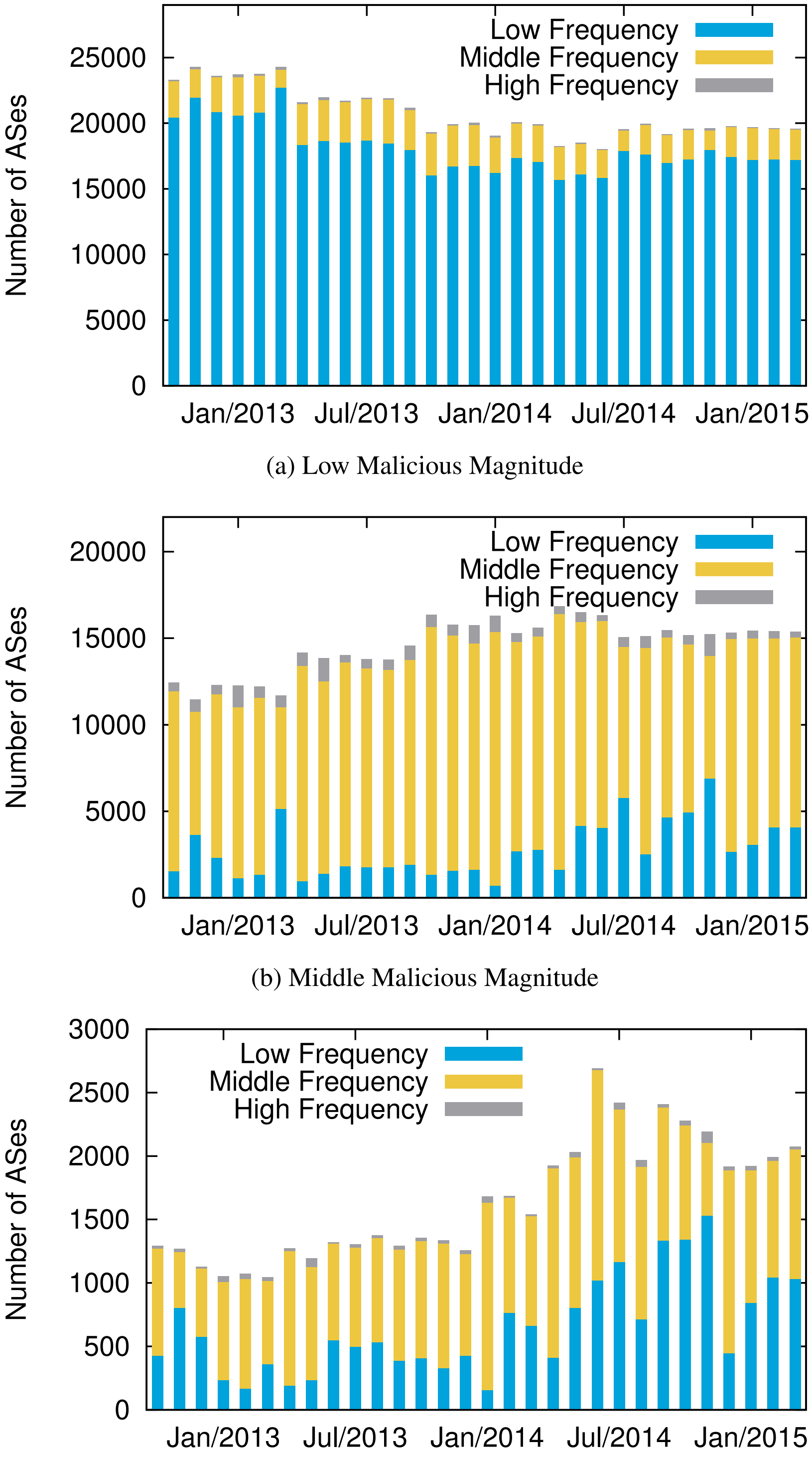
We first classify organizations into nine groups (clusters) based on their malicious magnitude and dynamic in each month. For simplicity of presentation we use four thresholds for the grouping: α_H indicating high magnitude, α_L indicating low magnitude, β_H for high dynamics and β_L for low dynamics. We then classify organizations, for instance, into High ($\bar{m}_{i,k} \geq \alpha_H$), Medium ($\bar{m}_{i,k} \in [\alpha_L, \alpha_H]$), and Low Magnitude ($\bar{m}_{i,k} \leq \alpha_L$) groups and similarly three groups of High, Medium and Low Frequency. In total, we have nine groups (3×3) when combining these two metrics. These thresholds in principle should be chosen by experimenting on their effectiveness in separation. For instance, one could employ clustering methods (see e.g., spectral clustering [103]) to obtain precise quantities. In this study we have used the following thresholds. For Spam, we set $\alpha_H = 0.01$ and $\alpha_L = 0.0001$, while for active scanning and phishing/malicious websites, the value are adjusted given the lower number of malicious sources ($\alpha_H = 0.001$ and $\alpha_L = 0.00001$). For frequency, since the dynamics (state changes) are w.r.t. to each organization's own magnitude, we use the unified threshold $\beta_H = 0.3$ and $\beta_L = 0.05$ for all three types of attacks.

We show the classification results for Spam, active scanning, and phishing/malicious websites in Figure 3.9, 3.10, and 3.11 respectively. We then highlight our key observations on the relationship between magnitude and dynamics as well as the evolution of network types.

First, the dynamics of maliciousness varied with the malicious magnitude. The organizations with low malicious magnitude also have low dynamics (Figure 3.9a, 3.10a, and 3.11a). This confirm the existence of "clean" organizations that consistently have small number of IP addresses that involve in malicious activities. For Spam, we also notice that around 40% of organizations in high malicious magnitude groups also have low dynamics, which indicating the existence of autonomous systems that are consistently "bad". And in general, the autonomous systems with middle malicious magnitude have the highest dynamics.

Second, we observed very different distribution over dynamics for different types of attacks. For active scanning, although the organizations with low magnitude have low dynamics, the organizations with medium and high magnitude almost completely fall into the medium and high frequency groups (Figure 3.10b and 3.10c). This observation indicates that the IP addresses performing active scanning change more frequently. One possible explanation is that large-scale scanning is relatively easily detected using network intrusion systems and are usually met with quick responses from network operators. We also see that organizations have the lowest dynamics in Phishing/malicious websites among all three types of attacks: more than 99% of low malicious organizations, and more than 54% of medium and high malicious organizations fall into the low dynamic groups. While previous studies have shown that most phishing campaigns only lasts for very short periods of time (usually a couple of hours) [110], one possible explanation for the static nature of phishing-related maliciousness is that the attackers tend to find new hosts which are located topologically close to the old ones.

Third, the number of autonomous systems with medium and high magnitude continues to increase throughout the period. For example, the number of organizations with high magnitude in Spam increased from around 1k in 2013 to more than 2k in late 2014 and

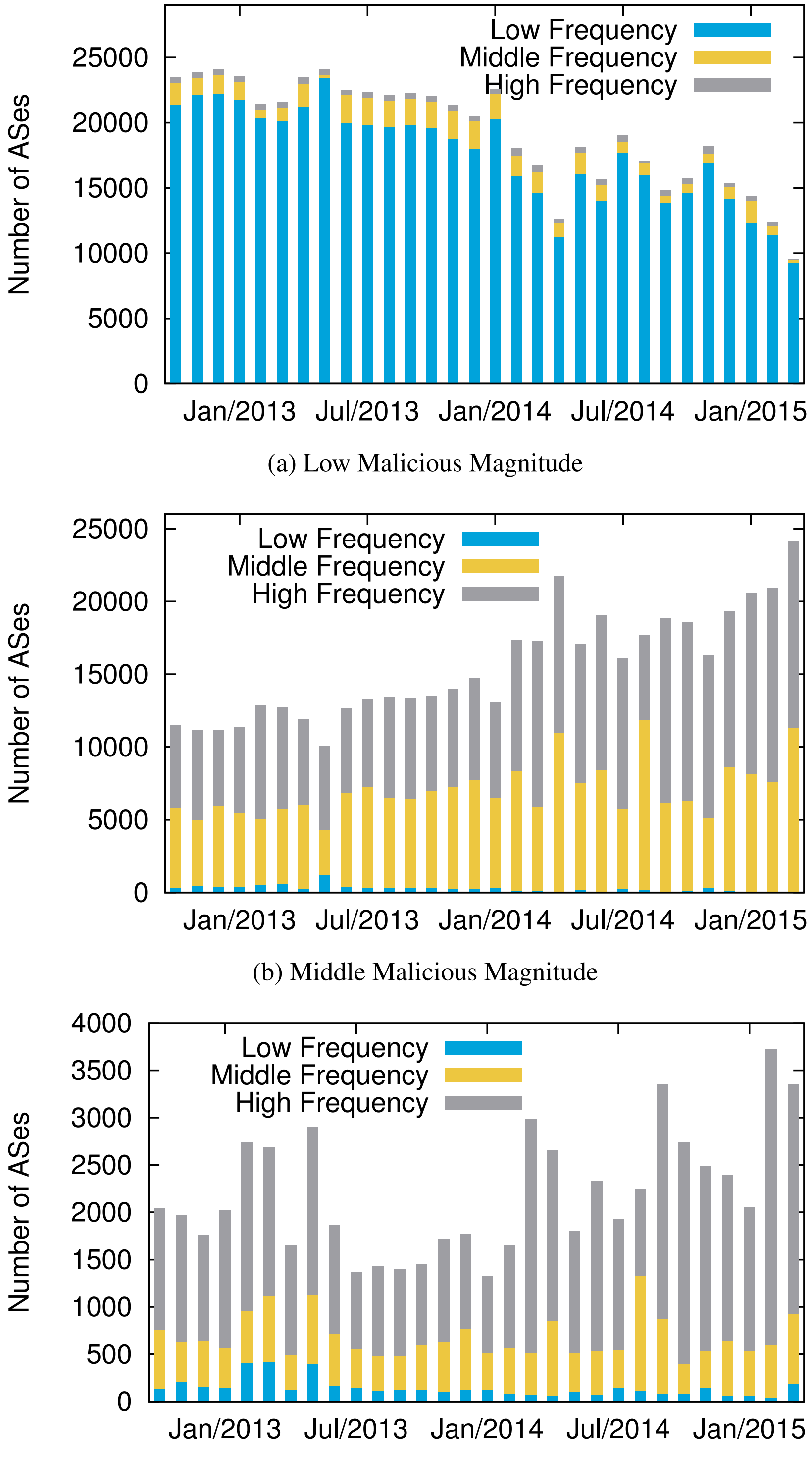


(a) Low Malicious Magnitude

(b) Middle Malicious Magnitude

(c) High Malicious Magnitude

Figure 3.9: Distribution of ASes based on magnitude and frequency categories in Spam activities.

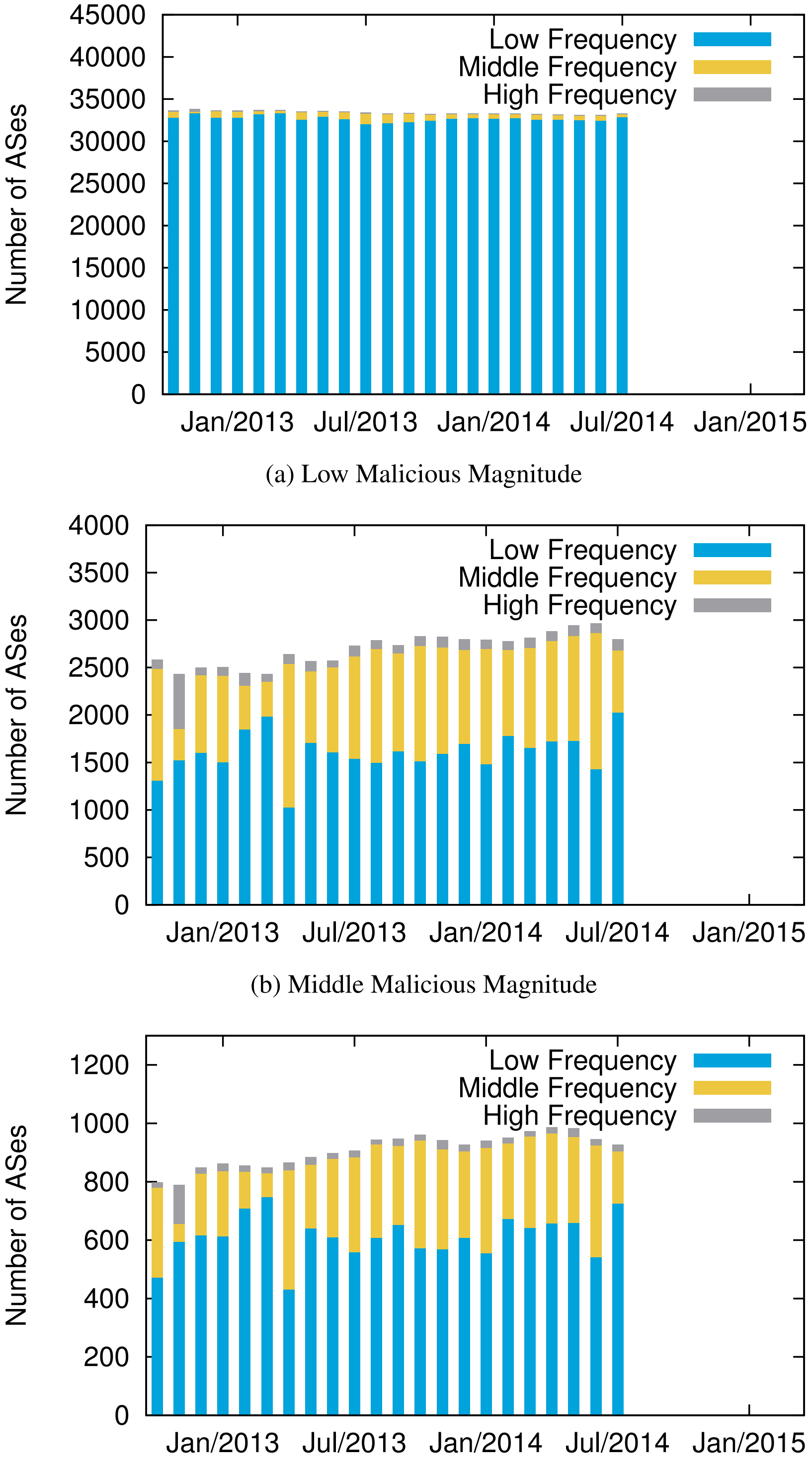


(a) Low Malicious Magnitude

(b) Middle Malicious Magnitude

(c) High Malicious Magnitude

Figure 3.10: Distribution of ASes based on magnitude and frequency categories in active scanning activities.



(a) Low Malicious Magnitude

(b) Middle Malicious Magnitude

(c) High Malicious Magnitude

Figure 3.11: Distribution of ASes based on magnitude and frequency categories in phishing/malicious websites.

2015 (Figure 3.9c). And the number of middle magnitude in SCAN has increased from 11k to more than 10k during the period of our study (Figure 3.10b). This is in consistent to the increasing of overall malicious magnitude.

Forth, while we see a clear increasing trend into higher magnitude, the dynamics is also increasing for active scanning. Organizations with high frequency has increased from 20% in 2013 to about 35% in 2015 and organizations with medium frequency also increased from 19% to 26%. However, for Spam, we observe an increasing in the number of organizations with high magnitude but low frequency. This indicates that there are more stable "bad" organizations for which we can target our efforts.

3.4 Summary

In this chapter, we study the longitudinal evolution of Internet threats landscape from the perspective of organization maliciousness. Using Autonomous Systems and routing prefixes as proxies of Internet Organizations, we formed the maliciousness at a organization level and characterize the malicious behavior of an organization with both the level of maliciousness (magnitude) and how frequently the maliciousness changes (dynamics). With our defined maliciousness, we quantify the security benefits of aggregated blocking compared to IP-based blocking. Our results show that aggregated maliciousness of organizations, both at autonomous system and prefix levels, can dampen the dynamic nature of IP addresses, thus resulting in more consistent and predictive measure of malicious sources. Then we analyze the longitudinal evolution of organization maliciousness. We found that both the magnitude and dynamics of maliciousness have been increasing during the two and half years. While taking the dynamics as a proxy of responsiveness, organizations have become faster in responding to malicious activities.

CHAPTER 4

Understanding the Relationship between Organization Maliciousness and Security Mismanagement

In chapter 3, we find that the maliciousness of organizations various greatly. This chapter aims to understand the causes for such a variances in organization maliciousness. Specifically, we explore the question that whether security mismanagement is one the cause for organization maliciousness.

Misconfigured networks have long been attractive resources for hackers [23], and anecdotal evidence suggests that mismanaged networks are often taken advantage of for launching external attacks, posing a risk not only to themselves, but to the Internet as a whole. One example of this can be seen in DNS amplification attacks in which attackers utilize open DNS resolvers to flood target hosts with a large number of DNS responses. These amplification attacks have long been observed in the wild and continue to occur with increasing scale and impact [119, 20]. These attacks are innately dependent on both widely-distributed misconfigured open DNS resolvers and the ability of attackers to forge request packets. In spite of calls by the Internet security community to address both of these issues by following standard deployment practices [76, 82], serious attacks continue to occur [52, 41]. As a result, these events are frequently described in terms of economic externalities: "a situation where a party could efficiently prevent harm to others—that is, a dollars worth of harm could be prevented by spending less than a dollar on prevention—but the harm is not prevented because the party has little or no incentive to prevent harm to strangers [81]."

In this chapter, we complement these anecdotes of individual incidents with a macroscopic, systematic study of one such externality—security mismanagement. For the purpose of this study, we define mismanagement as the failure to adopt commonly accepted guidelines or policies when administrating and operating organizations' networks. We explore the relationship of such misconfiguration with our defined organization maliciousness. We seek to understand the relationship between different types of network mismanagement and global Internet security.

Rather than focusing on how individual vulnerabilities influence the likelihood of a host becoming compromised (e.g., CVE-2008-4250 resulting in Conficker infections), we instead investigate how symptoms of network mismanagement, such as the presence of open recursive resolvers or instances of observed BGP misconfiguration, relate to organizational reputation built from externally observed malicious behavior, such as malware hosting and SPAM. While these features are merely proxies, ultimately, we hope to answer the question of what relationships exist between poor network management and maliciousness of the organization.

Our results show a statistically significant, strong positive correlation (0.64 correlation coefficient <0.01 p-value) between mismanagement and maliciousness of organizations. However, a correlation test is not sufficient as there might be latent variables that causes both maliciousness and mismanagement. We then assume organization management and maliciousness can both be impacted by common social and economic factors such as country GDP and number of customers in routing topology. By applying multivariate regressions, we find that the relationship still holds while controlling for these social and economic considerations.

4.1 Symptoms of Mismanagement

There are many symptoms that externally reflect poor network management. We analyze eight of these symptoms, which we list in Table 4.1. While these symptoms do not necessarily comprehensively describe all manners in which a network could be mismanaged, we choose to focus on these particular symptoms because they are well-documented

Symptoms	Best Current Practices	Functions	Attacks	Dataset
Open Recursive Resolvers	BCP 140/RFC 5358	Naming	DNS Amplification	Global
DNS Source Port Randomization	RFC 5452	Naming	DNS Cache Poisoning	Global
Consistent A and PTR records	RFC 1912	Naming	-	Partial
BGP Misconfiguration	RFC 1918, RFC 6598	Routing	-	Global
Egress Filtering	BCP 38/RFC 2827	Transit	-	Partial
Untrusted HTTPS Certificates	RFC 5246, RFC 2459	Web	Man-in-the-middle	Global
Open SMTP Mail Relays	RFC 2505	Mail	SPAM	Global
Publicly Available IPMI cards	Manufacturer's Guideline	Server	Compromising Hosts	Global

Table 4.1: Summary of mismanagement metrics and the third-party, public data sources used for validation

in published Request for Comments (RFCs) and Best Current Practices (BCPs) [68], and are part of the security community's best practices. We attempt to focus on characteristics that are symptomatic of overall network mismanagement rather than on specific vulnerabilities that could be used for mounting an attack. This is intended to reduce any bias between mismangement symtoms and maliciousness metrics we consider later in this work (e.g., CVE-2008-4250 resulting in Conficker infections).

We choose a range of symptoms ranging from BGP routing stability to the management and patching of SMTP, HTTPS, and DNS servers. This range of symptoms has several merits. First, it provides a global perspective of an organization's network management. For example, different teams potentially manage different services and by analyzing a range of different symptoms, we focus on the overall organizational network mismanagement rather than a single misconfigured service. Second, the analysis of multiple symptoms allows us to analyze the relationships between different symptoms. Although care was taken in choosing these symptoms, we make no claim that they are complete or without bias. We discuss potential drawbacks of these symptoms invidually in the follwing sections and reflect on them at the end of this section.

In the following subsections, we discuss each of the mismanagement symptoms with respect to their security implications, associated best practices, and our data collection methodology.

4.1.1 Open Recursive Resolvers

Open DNS resolvers respond to recursive queries for any domain and pose a direct threat to the Internet due to their role in DNS amplification attacks. In an amplification attack, an attacker sends simple DNS queries to an open resolver with a spoofed source IP address. While the DNS lookup request itself is small, the response to the victim is much larger and, as a result, the responses overwhelm the victim. BCP 140 [76] provides several recommendations for how to configure open resolvers to mitigate these threats. Ultimately, recursive lookups should be disabled unless specifically required and, when enabled, limited to intended customers.

In order to analyze the misconfiguration of open resolvers, we utilize data provided by the Open Resolver Project [38], which conducts active scans of the public IPv4 address space by sending a DNS query to every public address on port 53 and capturing the responses. The project has been performing these scans weekly since April, 2013, and has identified more than 30 million open resolvers. Detailed data collection methodology and preliminary results can be found in their recent presentation at NANOG [61].

We specifically consider the scan from June 2, 2013, which found 34.2 millions open resolvers in total. We consider the hosts that support open recursive queries as misconfigured, given their potential risk to the Internet and their failure to implement even the simplest best practices. Ultimately, we find 27.1 million open recursive resolvers on the Internet.

4.1.2 DNS Source Port Randomization

DNS cache poisoning is a well-known attack in which an attacker injects bogus DNS entries into a recursive name server's local cache. Traditionally, DNS resolvers used a randomized query ID in order to prevent cache poisoning attacks. However, in 2008, Dan Kaminsky presented a new subdomain DNS cache poisoning attack that has two new advantages [37]. First, it extends the window of attack because there is no valid reply from the authoritative name server with which to compete. Second, the multiple identical queries allow attackers to brute-force the 16-bit transaction ID that was previously relied upon for preventing these types of attacks.

Current best practices (RFC 5452 [87]) recommend randomizing the source port when performing DNS lookups in order to prevent these brute force attacks. In this configuration, a DNS server will use a large range of source ports instead of a single preset port, which significantly increases the search space for an attacker. For example, if a DNS server utilizes 2,000 source ports, the search space would increase from 64,000 to more than 100 million possibilities. Most popular DNS packages have already issued patches that implement source port randomization [1, 2].

In order to determine whether networks have patched their DNS resolvers with source port randomization, we analyze the set of DNS queries made against VeriSign's [56] .com and .net TLD name servers on February 26, 2013. In total, we observed approximately 5 billion queries from 4.7 million DNS resolvers.

In this experiment, we track the source ports utilized to make DNS queries against these TLD servers and infer that resolvers that only utilize the default source port without implementing source port randomization are misconfigured. We find that 226,976 resolvers, which account for 4.8% of total resolvers seen in the data, do not utilize source port randomization.

4.1.3 Consistent A and PTR records

DNS supports several types of records, of which Address (A) and Pointer (PTR) records are two of the most common. An A record is used to map a hostname to an IP address. A PTR record resolves an IP address to a canonical name.

One merit of PTR records is that they facilitate the validation of connecting clients and are widely used for detecting and blocking malicious IP addresses. For example, SMTP servers often discard messages from IP addresses without a matching PTR or MX record. The DNS operational and configuration guidelines (RFC1912 [65]) dictate that every A record should have a corresponding PTR record[25].

In our study, we utilize two datasets in order to estimate the global status of DNS records: the .com and .net second level domains stored in the VeriSign zone files and the domains in the Alexa Top 1 Million popular websites [9].

In order to determine which A records have associated PTR records, we perform a DNS query for each domain in our two datasets, finding 116 million A records. We then perform a reverse DNS lookup of the IP addresses appearing on these 116 million A records. We find that 27.4 million A records, which account for 23.4% of A records we queried, do not have a matching PTR record.

We note that our dataset is biased toward domains within North America and Europe. However, given that .com and .net domains account for more than half of all domains on the Internet [59] and that Alexa includes the most popular sites in the world, we believe our results still provide insights into the status of DNS records management.

4.1.4 BGP Misconfiguration

Publicly routed networks utilize Border Gateway Protocol (BGP) in order to exchange advertised routes. A router can announce a new route for a prefix or withdraw a route when it is no longer available. Routers are expected to not send updates unless there are topological changes that cause its advertised routes to change. However, misconfiguration and human error can result in unnecessary updates, which can potentially lead to both security vulnerabilities (e.g., Bogons [108, 121]) and downtime (e.g., AS7007 incidents [3]).

Mahajan et al. note that 90% of short-lived announcements (less than 24 hours) are caused by misconfiguration [101]. This is because policy changes typically operate on human time-scales, while changes due to misconfiguration typically last for a much shorter time.

In order to measure BGP misconfigurations, we use this simple heuristic in coordination with BGP updates from 12 BGP listeners in the Route Views project [104]. In our experiment, we track the time period for every new route announcement during the first two weeks of June, 2013 and infer that routes that last less than a day were likely caused by misconfiguration. We detect 42.4 million short-lived routes, which account for 7.8% of announced routes during the period of two weeks. We note that the Mahajan methodology is dated, and a fruitful area of future work would be to validate this methodology in the context of current routing practice (i.e., the current practice of fine-graned routing announcements).

4.1.5 Egress Filtering

Attackers often spoof source IP addresses to achieve anonymity or as part of DDoS attacks [66, 60]. In order to counter these attacks, it has been a best practice since 2000, to perform egress filtering as documented in BCP 38 [82].

In order to measure which networks have implemented egress filtering, we consider data from the Spoofer Project [48], which utilizes approximately 18,000 active clients to send probes to test for the presence of egress filtering. We specifically analyze data from April 29, 2013 and check in which netblocks an arbitrary routable source IP address can be spoofed. Because spoofed IP addresses are primarily used by attackers, we consider netblocks that do not implement address filtering to be misconfigured. The dataset from April 29th contained results for 7,861 netblocks, of which 35.6% have not implemented egress filtering. Unfortunately, the status of the remaining 195,000 netblocks is unknown.

4.1.6 Untrusted HTTPS Certificates

HTTPS sites present X.509 certificates as part of the TLS handshake in order to prove their identity to clients. When properly configured, these certificates are signed by a browser-trusted certificate authority.

Now that browser-trusted certificates are available for free from several major providers, the best practice is for public websites to use browser-trusted certificates. As such, we consider the presence of untrusted certificates as a potential symptom of misconfiguration. However, a large number of sites utilize self-signed certificates or certificates that have not been validated by a trusted authority.

In order to understand the state of HTTPS certificate utilization, we consider a scan of the HTTPS ecosystem that was completed as part of the ZMap network scanner project [80]. In this scan, Durumeric et al. performed a TCP SYN scan on port 443 of the public IPv4 address space on March 22, 2013 using the ZMap network scanner. It then performed a follow-up TLS handshake with hosts that responded on port 443, and collected and parsed the presented certificate chains using libevent and OpenSSL.

Using this dataset, we consider whether presented certificates are rooted in a browser-trusted certificate authority or are not browser trusted (i.e. self-signed or signed by an unknown certificate authority). We found 33 million hosts with port 443 open, 21.4 million hosts who successfully completed a TLS handshake, and 8.4 million distinct X.509 certificates. Among these certificates, only 3.2 million (38%) were browser-trusted, and only 10.3 million (48%) of the hosts presented browser-trusted certificates.

4.1.7 SMTP server relaying

Open mail relays are SMTP servers that do not perform any filtering on message source or destination and will relay e-mail messages to any destination. These servers are frequently abused by spammers in order to avoid detection or to offload traffic onto third parties. Given their consistent abuse, the Internet community strongly recommends against their use (RFC 2505 [99], RFC 5321[93]).

In order to investigate the prevalence of open mail relays, we performed a TCP SYN scan of the IPv4 address space for port 25 using ZMap on July 23, 2013 and attempted the initial steps of an SMTP handshake in order to determine whether the server would reject the sender or receiver. After determining whether the server would accept the message, we terminated the connection without sending any mail.

Our scan identified 10.7 million servers with port 25 open of which 7.0 million identified themselves as SMTP servers. Of the responsive SMTP servers, 6.2 million explicitly rejected our sender, 433,482 terminated the connection or timed out, and 22,284 SMTP servers accepted the message, identifying them as open mail relays.

4.1.8 Publicly Available Out-of-Band Management Cards

Out-of-band management cards that allow remote control of power, boot media, and in some cases, remote KVM capabilities, are now commonplace on servers. Most of these management cards are implementations of the Intelligent Platform Management Interface (IPMI) industry standard, but come under a variety of names, including Dell's Remote Access Card (iDRAC), HP Integrated Lights Out (iLO), and Super Micro's Base Management Card (BMC).

While these interfaces are a valuable tool for systems administrators, they also pose a severe security risk if publicly available on the Internet [67]. These devices have recently been found to be riddled with vulnerabilities, and manufacturers explicitly recommend that the devices be isolated on a private management network and not be made available on the public Internet [86, 67, 118]. As such, we consider any publicly-available management card to be a misconfiguration.

In order to measure the public availability of these IPMI cards, we consider the TLS certificate data set collected by Durumeric et al. by searching for known default certificates presented by IPMI cards manufactured by Dell, HP, and Super Micro. In this dataset, we found IPMI cards hosted on 98,274 IP addresses.

4.1.9 Summary and Limitations of Symptoms

In this work, we choose to focus on eight symptoms that we believe expose mismanaged networks and, for the most part, are not vulnerabilities that will directly influence the the blacklists we consider later in this work. We further focus on symptoms that have clear and accepted best practices, which have been documented by the security community.

We note that these symptoms are not the only externally visible metrics for network mismanagement—there most likely exist networks that contain other mismanaged services, which may correlate to the symptoms we present. Additionally, we acknowledge that biases may exist between the symptoms that we select that cannot be discerned without operational details of an organization (e.g., open recursive DNS resovers and open SMTP relays).

Regardless, we observe pervasive failures in implementing common security practices in the symptoms that we do consider, several of which can, by themselves, result in easily exploitable vulnerabilities. Specifically, we find that there exist (1) 27 million open recursive resolvers, (2) 226,976 DNS resolvers that have not been patched to use source port randomization, (3) 27.4 million A records that do not have matching PTR records, (4) 42.4 million short-lived BGP routes, (5) 35.6% of the tested netblocks that have not implemented egress filtering, (6) 10.2 million HTTPS servers using untrusted certificates, (7) 22,284 SMTP servers that allow open mail relays, and (8) 98,274 public accessible IPMI cards.

4.2 Mismanagement Symptoms at Autonomous System Level

In this section, we analyze the previously discussed symptoms at the AS level in order to determine the global misconfiguration of different networks and to measure the relationships between different types of misconfiguration.

4.2.1 Abstracting Networks

While it would be ideal to measure the correlation between mismanagement and maliciousness at the management boundaries, there exist no easily visible or authoritative network boundaries from an external perspective—it is often difficult or impossible to detect what sociopolitical organizations own or manage network blocks or specific hosts within a network block.

Several methodologies have emerged for aggregating networks ranging from AS-level aggregation, to BGP routed prefix [97], to aggregating hosts by adminstrative domains defined by authoritative name server [120, 106, 107]. We choose to aggregate hosts at the AS level because several of our metrics are only available at this granularity and because as we move forward, we ultimately hope to send information to owning organizations.

We make no claim that this choice of administrative boundary is ideal. For example, several uniquely managed organizations make exist within a single AS (e.g., customers of a large provider). Strictly speaking, we do not show in all cases a correlation between mismanaged organizations and malicious networks, but rather between mismanaged ASes and ASes that have been the source of malicious traffic.

4.2.2 Distribution of Misconfigured Systems

We hypothesize that the security postures of networks will differ based on the varied effort placed in management and security. To validate this hypothesis, we consider the distribution of each of type of misconfiguration based on host IP addresses in each AS.

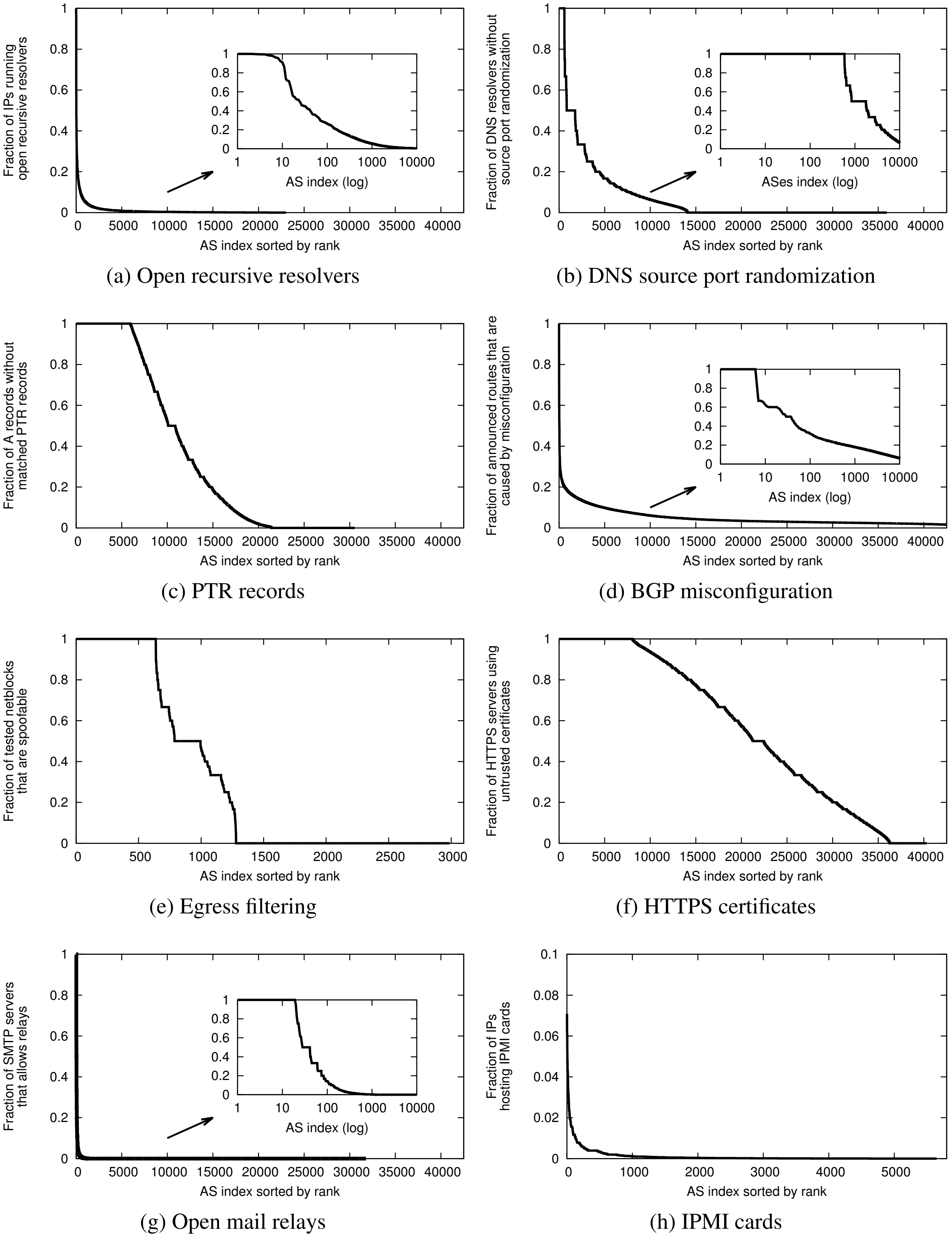
We rank networks by the normalized number of misconfigured systems, and show the breakdown of vulnerabilities in Figure 4.1. In line with our hypothesis, mismanagement is different between different networks—symptoms of misconfiguration are typically concentrated in a small number of networks.

In the remainder of this section, we discuss how we normalized each metric and the results of aggregating specific vulnerabilities by AS.

Open recursive resolvers We normalize the number of open recursive resolvers by total number of IP addresses announced by the AS. In Figure 4.1a, we show the normalized number of open recursive resolvers (i.e., fraction of IP addresses that are running open recursive resolvers) for each AS, ranked by a decreasing order. We find that in the top 10 most misconfigured ASes, close to 100% of the ASes' advertised addresses are running misconfigured open resolvers. While we do not know for sure why this is occurring, we suspect that these networks are centrally managed and hosts are similarly configured. Beyond these several cases, 477 ASes (1.2%) have more than 10% of IPs running misconfigured open recursive resolvers. The long-tail distribution shows that approximately 95% of all ASes are well-managed, with a small number of no open recursive resolvers.

DNS source port randomization We normalize the number of DNS resolvers without source port randomization by the total number of unique resolvers in the AS. The results are shown in Figure 4.1b. There are 14,102 ASes (33%) with at least one misconfigured DNS server. Among these, the top 584 most misconfigured ASes have 100% of their resolvers misconfigured, and more than 50% of the resolvers do not implement source port randomization in the top 1,762 ASes.

Consistent A and PTR records We define the normalized number of unmatched PTR records as the fraction of the AS' A records that do not have a corresponding PTR record. We show the results of this normalization in Figure 4.1c. At least one A record is mismatched in 21,418 ASes (49%). A large number of ASes have a disproportionally higher fraction of their A records mismatched: none of the A records in the top 5,929 ASes have



(a) Open recursive resolvers

(b) DNS source port randomization

(c) PTR records

(d) BGP misconfiguration

(e) Egress filtering

(f) HTTPS certificates

(g) Open mail relays

(h) IPMI cards

Figure 4.1: Normalized distribution of misconfigured systems in autonomous systems. All of the symptoms show that there are a few ASes that have a disproportional number of misconfigured systems.

corresponding PTR records and more than half of the A records are mismatched in the top 10,863 ASes.

BGP misconfiguration In order to normalize BGP misconfigurations, we consider the fraction of routing announcements originating from an AS that is misconfigured. Results are shown in Figure 4.1d. Unlike the previously discussed metrics, we do not find clearly divided groups of ASes. Instead, we find many ASes that announce a similar number of short-lived routes. Only 37 ASes have more than half of their updated routes as short-lived, and only a few ASes have less than 5% of their updates that are caused by misconfiguration. We suspect that this is because the causes of BGP misconfiguration are numerous and complex [101].

Egress Filtering Ideally, the number of netblocks without egress filtering would be normalized by the total number of netblocks in an AS. However, our dataset only includes information for a fraction of the netblocks in 2,987 ASes. Therefore, we estimate the normalized number by calculating the fraction of known netblocks that are spoofable in these 2,987 ASes. As shown in Figure 4.1e, approximately half of these ASes do not have any netblocks that allow address spoofing, while all of the tested netblocks in the top 638 ASes do not implement egress filtering and are spoofable.

We note that this particular metric may not accurately represent the distribution of networks without egress filtering. First, we can only estimate the deployment of source address validation in 6% of all ASes. Secondly, the results may be biased given that the tested netblocks in a particular AS may not accurately represent the behavior of the entire AS. However, even with these limitations, we believe that the existence of egress filtering is a symptom worth considering when discussing mismanaged networks due to the potential abuse for attacks.

Untrusted HTTPS certificates We normalize the servers that present untrusted certificates with the total number of HTTPS servers seen in each AS. The results are plotted in Figure 4.1f. While there is less risk associated with using self-signed certificates, we find that a large number of ASes contain servers with a self-signed certificate. Specifically, more than 36,000 ASes (82%) have at least one mismanaged HTTPS server. In 8,042 ASes, all hosts serving HTTPS on port 443 use a self-signed or an otherwise untrusted certificate.

Open SMTP mail relays We normalize open mail relays with the total number of SMTP servers in each AS, and we show the per-AS normalized number of open mail relays in Figure 4.1g. In comparison to other mismanagement symptoms, we find that mail servers are relatively well maintained. Only 1,328 ASes hosted open mail relays and only 135 ASes contained more than 10% of mail servers that are misconfigured.

Publicly available IPMI devices We find relatively few publicly available IPMI cards in comparison to the previously listed metrics; in total we find IPMI cards in 5,648 ASes. Normalized by the total number of IP addresses of the ASes, the number is tiny (Figure 4.1h). But, a few ASes are relatively poorly managed—2% of IP addresses have been detected with IPMI cards in the top 44 ASes.

4.2.3 Correlations between Symptoms

We next explore the question of what relationship, if any, exists between the different mismanagement symptoms within an AS. To quantify the relationship between two symptoms, we use Spearman's rank correlation test, which measures the statistical dependence between two ranked variables. We use rank-based correlation rather than value-based tests because of the differences in scale between ASes and the varying implications of each mismanagement symptom. Further, rank-based correlation is a nonparametric measure that does not require data from a normal distribution.

The result of Spearman's test is a value between -1 and 1, where a negative value indicates a negative relationship between two metrics and positive value indicates a positive relationship. For any nonzero value, we perform a hypothesis test with a 95% confidence level in order to determine whether the observed correlation is significant (i.e., if $p - value < 0.05$). For a significant nonzero correlation coefficient, the larger the absolute value, the stronger the relationship. According to Cohen's guidelines [71], values with absolute correlation coefficients from 0.1 to 0.3 can be considered weakly correlated, 0.3 to

	Open resolver	**Port random.**	**PTR record**	**BGP misconfig.**	**Egress filter.**	**HTTPS cert.**	**SMTP relay**	**IPMI cards**
Open resolver	-	**0.35** (< 0.01)	0.09 (< 0.01)	**0.17** (< 0.01)	0.09 (< 0.01)	**0.46** (< 0.01)	**0.14** (< 0.01)	**0.26** (< 0.01)
Port random.	**0.35** (< 0.01)	-	**0.14** (< 0.01)	0.07 (< 0.01)	0.04 ($= 0.02$)	**0.23** (< 0.01)	**0.16** (< 0.01)	**0.26** (< 0.01)
PTR record	**0.10** (< 0.01)	**0.15** (< 0.01)	-	0.03 (< 0.01)	0.01 ($= 0.46$)	0.00 ($= 0.37$)	**0.11** (< 0.01)	**0.15** (< 0.01)
BGP misconfig.	**0.17** (< 0.01)	0.07 (< 0.01)	0.03 (< 0.01)	-	0.04 ($= 0.04$)	**0.16** (< 0.01)	0.02 (< 0.01)	0.03 (< 0.01)
Egress Filter.	0.09 (< 0.01)	0.04 ($= 0.02$)	0.01 ($= 0.46$)	0.04 ($= 0.04$)	-	-0.02 ($= 0.32$)	**0.14** (< 0.01)	**0.10** (< 0.01)
HTTPS cert.	**0.46** (< 0.01)	**0.23** (< 0.01)	0.00 ($= 0.37$)	**0.16** (< 0.01)	-0.02 ($= 0.32$)	-	0.06 (< 0.01)	**0.15** (< 0.01)
SMTP relay	**0.14** (< 0.01)	**0.16** (< 0.01)	**0.10** (< 0.01)	0.02 (< 0.01)	**0.14** (< 0.01)	0.06 (< 0.01)	-	**0.26** (< 0.01)
IPMI cards	**0.26** (< 0.01)	**0.26** (< 0.01)	**0.15** (< 0.01)	0.03 (< 0.01)	**0.10** (< 0.01)	**0.15** (< 0.01)	**0.26** (< 0.01)	-

Table 4.2: Correlation coefficients and p-values between different mismanagement symptoms. There are significant correlations between different symptoms. (RED: Moderate correlation; BLUE: Weak correlation.)

0.5 moderately correlated, and 0.5 to 1.0 to be strongly correlated.

The pair-wise correlation coefficients and p-values are shown in Table 4.2. We find a statistically significant correlation between 25 of the 28 comparisons at a 95% confidence level. Of these, two of the pairs are moderately correlated, 14 pairs are weakly correlated, and the remaining correlations are trivial. Of the symptoms, we find the strongest correlation within vulnerability-related symptoms: open DNS resolvers, failure to implement source port randomization, and using untrusted HTTPS certificates.

Missing PTR records and BGP misconfiguration have the weakest correlation to other metrics. In the case of the PTR records, this may be caused by the biased dataset as discussed in Section 4.1.3. For BGP misconfiguration, we expect to see little correlation with other metrics due the complexity and potential inaccuracy of measurements (see Section 5.2.2.3).

We expected to find the lack of egress filtering significantly correlated with other symptoms, which we do not observe. However, we note that the relatively size sample size of this metric has skewed its results. Specifically, the measured ASes in our egress filtering dataset are biased toward fewer misconfigured systems as indicated by other metrics. As such, we do not draw any conclusions based on this metric.

One plausible explanation for the correlation between these technically disparate mismanagement metrics is that they are likely impacted by the organizational culture of security management. In other words, while we expect that disparate systems are managed by different groups within an organization, we suspect that members in an organization are influenced by its culture, including its hiring process, daily operating procedures, and general awareness of security vulnerabilities.

4.3 Unified Network Mismanagement Metric

We next analyze the mismanagement of networks as a whole using the eight metrics we previously described. We first combine the individual symptoms into an overall mismanagement metric. Our rationale is that while each symptom may be an inaccurate measure of the AS' mismanagement, the combination of disparate metrics provides a more holistic view. Using this global metric, we consider different attributes of ASes including their geographic region and topological role.

4.3.1 Combining Symptoms

We combine different symptoms into a single metric using Borda's method [77], which is a linear combination algorithm for aggregating ranked results. This provides us with an overall score for each AS that is equivalent to an unweighted average of the AS' rank in each individual symptom. We exclude our metrics on ingress filtering and PTR records given that they only represent a small number of ASes. We rank ASes by their overall mismanagement scores from the worst to best managed.

4.3.2 Geographical Distribution

We first consider the geographic distribution of mismanagement by mapping ASes to their geographical regions using the WHOIS services provided by Team Cymru [4]. To compare mismanagement of ASes, we group ASes into five groups based on their rank percentile in the overall mismanagement metric. We show the distribution of ASes in these five groups in Figure 4.2.

We find that networks allocated by ARIN are relatively well-managed, and that ASes in AFRINIC and LACNIC have a disproportionally large number of poorly-managed ASes. Approximately 15% of their ASes fall into the 5th percentile of mismanaged ASes, and 60% fall into the 25th percentile of mismanaged ASes.

One possible explanation for the regional differences is that less developed areas may devote less resources to network management. In addition, with different network opera-

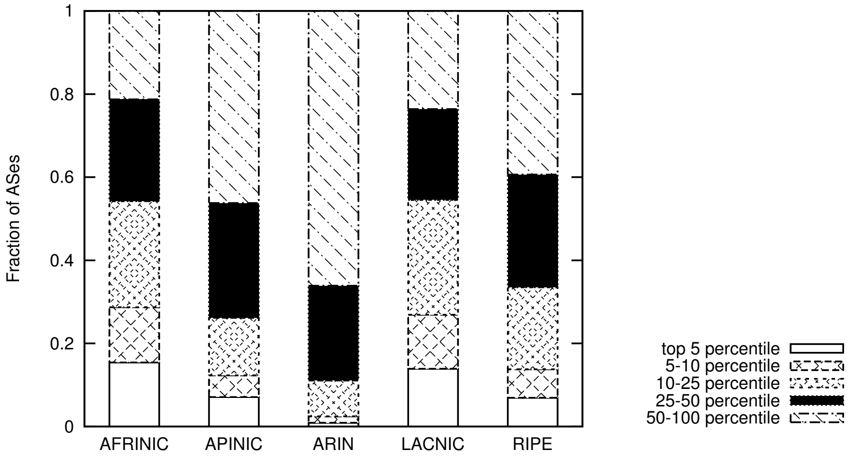


Figure 4.2: Regional differences in mismanagement. Networks assigned by ARIN are relatively well managed, while a larger fraction of networks under AFRINIC and LACNIC are poorly managed.

tor groups being geographically based, the exposure to management regulations and best practices could potentially vary between geographic regions.

4.3.3 Topological Roles

Next, we examine whether the topological role of an AS influences its management. One might expect that mismanagement would increase with AS size and breadth of services due to the complexity of managing large and diverse networks. We use the categorization of ASes' topological roles and techniques proposed by Dhamdhere et al. [79] in order to estimate AS size and functionality. In this categorization, ASes are separated into four topological roles: Enterprise Customers (ECs), Small Transit Providers (STPs), Large Transit Providers (LTPs), and Content/Access/Hosting Providers (CAHPs).

In order to determine the topological position of an AS, we attempt first to infer business relationships (provider-customer, peers, and sibling) between ASes using Gao's algorithm against the global BGP routing table [**?**]. We then build a decision tree based on the number of customers and peers of each AS and train against a set of 150 known ASes. We then utilize this decision tree on the remaining set of ASes in order to determine their topological roles. With this methodology, we find 39,653 ECs, 3,667 STPs, 30 LTPs, and 1,599 CAHPs

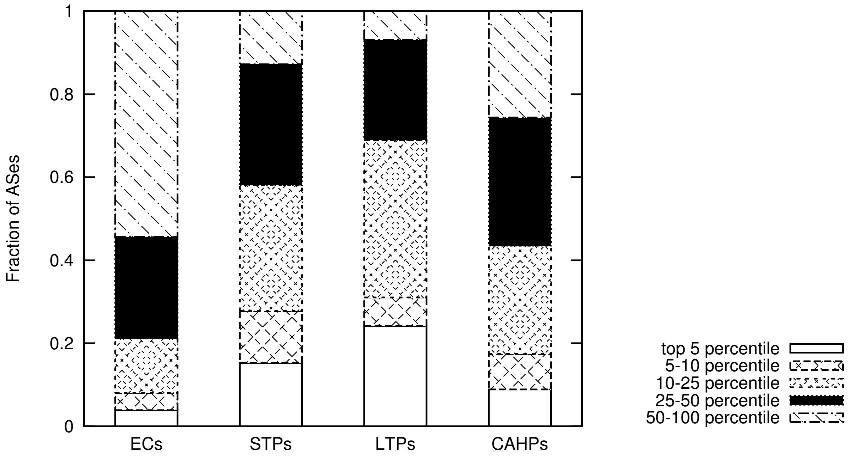


Figure 4.3: Topological role differences in mismanagement. In general, Transit providers (LTPs and STPs) are the the most mismanaged type of organization, followed by content, access and hosting providers (CAHPs).

in June 2013.

Figure 4.3 shows the distribution of ASes in each topological role. As expected, we find that *Enterprise Customers are the best managed while Transit Providers are poorest managed followed by Content/Access/Hosting Providers*. Except the complexity of managing large networks, it is also plausible that these large providers receive lower scores because they contain a large number of small customers that are not large enough or staffed to receive an AS. In either case, service providers need to be held responsible for maintaining their networks or alerting their customers to these vulnerabilities.

4.4 Mismanagement and Maliciousness

In this section, we explore whether there is a relationship between the eight mismanagement symptoms we measured and our definition of organization maliciousness.

4.4.1 Maliciousness of Autonomous Systems

As defined in Section 3, we form the maliciousness of autonomous systems with 12 blacklists which contains approximately 160 million unique IP addresses. We note that

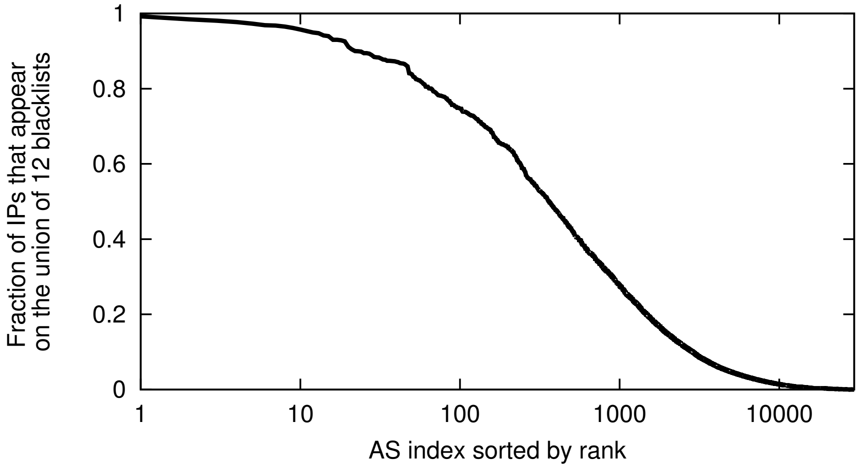


Figure 4.4: Maliciousness of autonomous systems. Similar to the distribution of misconfigured systems, a few ASes have a disproportionally large number of malicious IP addresses.

while we attempt to choose mismanagement symptoms that appear to be unrelated to the blacklists in question, there is potential for bias between some mismanagement symptoms and these blacklists. We specifically acknowledge that there is likely a bias between the SPAM blacklists we use and the existence of open SMTP relays on a network. However, we ultimately find only a weak positive correlation between the two, less than the correlation with many of the other mismanagement symptoms we investigated.

As a result, we find that 29,518 ASes (67%) have at least one blacklisted IP address. Figure 4.4 depicts the maliciousness of ASes sorted in descending order. Similar to the distribution of misconfigured systems, the maliciousness of ASes varies greatly: the top 350 ASes have more than 50% of their IP addresses blacklisted, while the bottom ASes have a negligible number of blacklisted IPs.

4.4.2 Are Mismanaged Networks more Malicious?

We hypothesize that there is a positive correlation between mismanagement and maliciousness for two reasons. First, well-managed networks will expose fewer attack vectors, which will ultimately lead to fewer infected hosts and will prevent attackers from using well-managed networks as launch points for attacks. Second, well-managed networks are more likely to adopt other reactive approaches (e.g., anomaly detection, filtering/blocking)

Metric	**Coefficient**	**P-value**	**Interpretation**
Open recursive DNS resolvers	0.59	< 0.01	strong positive
DNS source port randomization	0.45	< 0.01	moderate positive
Consistent A and PTR records	0.20	< 0.01	weak positive
BGP misconfiguration	0.19	< 0.01	weak positive
Lack of Egress filtering	0.04	< 0.01	no correlation
Untrusted HTTPS certificates	0.44	< 0.01	moderate positive
Open SMTP mail relays	0.23	< 0.01	weak positive
Mismanaged IPMI cards	0.22	< 0.01	weak positive
Overall	0.64	< 0.01	strong positive

Table 4.3: Correlation coefficients and p-values between mismanagement and maliciousness. There is a statistically significant correlation between our mismanagement symptoms and maliciousness.

to mitigate the impact of compromise. Therefore, if compromise were to occur, hosts would not remain online long enough to be found in our scans or to be placed on a global blacklist.

In order to determine the relationship between mismanagement and maliciousness, we examine the correlation between the two metrics. We first calculate Spearman's correlation between each individual mismanagement symptom and maliciousness. All of the symptoms we examine have a statistically significant positive relationship with networks' apparent maliciousness at a 95% confidence level. We present the results in Table 4.3. In particular, the vulnerability-related symptoms (open DNS resolvers, DNS source port randomization, and HTTPS server certificates) have a stronger correlation to maliciousness than the other symptoms. One possible explanation for this is that the mismanaged open resolvers, DNS servers, and HTTPS servers, once compromised, can be turned into malicious sources directly. Other mismanagement symptoms (e.g., BGP misconfiguration or mismatched DNS records), which pose risk to the Internet but cannot be turned into malicious sources, have a weak correlation to maliciousness. And we find that the correlation between anti-spoofing and maliciousness is negligible, which we believe is due to biased datasets (i.e., our egress filtering dataset only covers 4% of total networks).

And we find that our aggregated mismanagement metric has the strongest correlation with maliciousness. Given that our overall mismanagement metric is an approximation of the true management posture of a network, this observation shows that researchers need to consider a more holistic view of network health, rather than only consider specific vul-

Variable	Coefficient	t-value	p-value
(Intercept)	3.089e+04	273.648	<2e-16
GDP	-1.732e-10	-16.043	<2e-16
GDP per capita	-2.175e-01	-58.963	<2e-16
# of Customers	1.107e+01	8.553	<2e-16
# of Peers	2.102e+01	8.238	<2e-16

Table 4.4: Multivariate regression with rank of mismanagement as the dependent variable and the four social and economic factors as independent variables. All the factors significantly influence the organization's rank of mismanagement: the higher the GDP/GDP per capita, the better the management; the more the customers and peers, the worse the management.

nerabilities or symptoms. However, it is possible that the strongest correlation is caused by statistical errors—the sample randomness could be reduced by combining several samples togerther. We do not explore the statistical details here, which is one limitation of our current work.

Correlation does not imply any cause-effect relationship; there could very well be a third variable that impacts both mismanagement and maliciousness [113]. For example, as we discussed in Section 4.3, mismanagement differs between geographical and topological locations, which indicates that external social and economic factors influence mismanagement. Therefore, we need to further examine whether mismanagement causes maliciousness when controlling for social and economic factors.

We assume that the aforementioned differences in management within different geographic regions (Section 4.3.2) are caused by the differing economic development in these regions. Networks in a developed region might invest more in management and security than in less-developed countries. For each country, we use gross domestic product (GDP) and GDP per capita as economic indicators for the ASes located in the country. In addition, we look at the business relationship between ASes to infer their social and financial status. Specifically, we use two variables: number of customers and number of peers in Internet routing. The number of routing customers and peers are calculated with Gao's algorithm as discussed in 4.3.3.

The multivariate regression results show that all the four social and economic variables are significant influencing factors for both mismanagement and maliciousness of organiza-

Variable	Coefficient	t-value	p-value
(Intercept)	1.656e+04	191.644	<2e-16
GDP	-7.103e-11	-8.592	<2e-16
GDP per capita	-1.833e-01	-64.897	<2e-16
# of Customers	6.352e+00	6.408	1.49e-10
# of Peers	1.105e+01	5.659	1.53e-08

Table 4.5: Multivariate regression with rank of maliciousness as the dependent variable and the four social and economical factors as independent variables. The results are similar to those of mismanagement in Table 4.4.

tions. In Table 4.4 and Table 4.5, we show the detailed regression results. It can be seen that all the four factors statistically influence the rank of an organization's maliciousness and mismanagement at a 95% confidence level. The higher the GDP and GDP per capita, the better the management status of the organization. From the coefficients, we can see that a 10 billion dollar increase in the country's GDP would result in a decrease of 0.71 in the maliciousness rank and a decrease of 1.7 in the mismanagement rank when controlled by other factors. A 10 dollar increase in the GDP per capita would cause the maliciousness rank to decease by 1.8 and the mismanagement rank to decease by 2.1 when controlled by other factors. Similarly, the more the customers and peers, the worse the management. With each additional customer the organization has, its maliciousness rank would increase by 6.4 and its mismanagement rank would increase by 11; and if the organization have one more peer, its maliciousness rank will increase by 11 and its mismanagement rank will increase by 21.

In order to determine whether a correlation exists between mismanagement and maliciousness when controlling for these factors, we test their relationship with a multivariate regression. We first quantify the relationship between maliciousness and mismanagement with a univariate regression. Then we add the four social and economical factors as independent variables to test whether this relationship holds when controlled by these factors.

In the univarite regression between maliciousness and mismanagement, we find the following regression model:

$$rank\ of\ maliciousness = 0.451 * rank\ of\ mismanagement - 131.9$$

Variable	Coefficient	t-value	p-value
(Intercept)	4.581e+03	36.445	<2e-16
Mismanagement	3.879e-01	118.336	<2e-16
GDP	-3.841e-12	-0.537	0.5911
GDP per capita	-9.891e-02	-38.988	<2e-16
# of Customers	2.057e+00	2.405	0.0162
# of Peers	2.902e+00	1.722	0.0851

Table 4.6: Multivariate regression with rank of maliciousness as the dependent variable, and mismanagement and the four social and economic factors as independent variables.

Variable	Coefficient	t-value	p-value
(Intercept)	4.576e+03	36.413	<2e-16
Mismanagement	3.883e-01	118.956	<2e-16
GDP per capita	-9.958e-02	-47.837	<2e-16
# of Customers	2.324e+00	2.764	0.00572

Table 4.7: Final multivariate regression model for maliciousness and mismanagement when controlled by social and economic factors. Mismanagement is a significant influencing factor for organization maliciousness when controlled by these latent variables.

in which the rank of mismanagement is a significant factor with $p-value < 2e-16$ for rank of maliciousness. This indicates that when the mismanagement rank increases by one, the organization's maliciousness rank would increase by 0.45. When adding all four factors and applying a multivariate regression, we get a model as shown in Table 4.6. This shows that the GDP and number of peers are not statistically significant independent variables in this model (with $p-value > 0.05$). Therefore, we remove them to get the final model, as shown in Table 4.7. In the final model, although when controlled by these social and economic factors, the increase of mismanagement rank has a lower impact on the maliciousness, but mismanagement is still a significant influencing factor for maliciousness. When controlled by these factors, every increase in the mismanagement rank causes an increase of 0.39 in the organization's maliciousness rank. Therefore, we conclude that organization maliciousness is correlated with its mismanagement when controlled with our selected social and economic factors.

We note that there might be missed variables that may diminish the relationship between mismanagement and maliciousness. For example, the legal regulations for security managements in different countries or the education level of population would directly

Metric	Coefficient	P-value	Interpretation
Open recursive DNS resolvers	0.54	< 0.01	strong positive
DNS source port randomization	0.24	< 0.01	weak positive
Untrusted HTTPS certificates	0.39	< 0.01	moderate positive

Table 4.8: Aggregation at BGP prefix level: Correlation coefficient and p-value between mismanagement and maliciousness.

impact the organizations management and behavior patterns. In the future, more social, behavioral and economic factors could be added into the analysis to verify this relationship.

4.4.3 Impact of Aggregation Type on Maliciousness Correlations

In order to explore our choice of using autonomous system as the proxy of Internet organizations instead of at a more granular level, such as by routed block or authoritative name server, we consider the correlation between three of the mismanagement symptoms and our global maliciousness metric at the routed block granularity. We find very slight differences between the level of correlation (e.g., the correlation between DNS port randomization moves from a moderate positive correlation to a weak positive correlation while other correlations remain unchanged). Ultimately, we find that there continues to be strong positive correlations for all of the mismanagement symptoms. We show the exact values in Table 4.8.

4.5 Limitations

There are several limitations in the data that we collect in this work. First, we utilize a large number of external data sources that were collected using disparate collection methodologies, from multiple networks with differing coverage, and during multiple time frames. While utilizing these datasets reduces the impact of active scanning on destination networks, these discrepancies could potentially impact the correlations we present. While we are not aware of any impact, there is future work to collect more consistent datasets.

As discussed throughout the paper, we aggregate networks at an AS level. Additional insight may be gained by grouping hosts at a more granular level or with the addition of organizational data. As well, there is future work required to determine whether hidden biases exist between the symptoms we select or between the symptoms and the mismanagement metrics we utilize, particularly between mail server mismanagement and SPAM metrics.

4.6 Summary

There is a widely held, anecdotal belief that mismanaged networks not only pose a risk to themselves, but to the Internet as a whole. In this paper, we systematically examine the relationship between mismanagement and maliciousness by analyzing eight Internet-scale mismanagement metrics and twelve commonly used global blacklists. Through this analysis, we find that different symptoms of mismanagement are highly correlated among themselves, and we ultimately find the relationship between mismanagement and maliciousness holds while controlling for social and economic considerations.

While security has primarily been reactionary, the understanding of the relationship between mismanagement and maliciousness is the first step in developing proactive security systems. We encourage the security community to switch some of their attention from studying attacks to researching defensive mechanisms and incentivizing organizations to implement even the simplest security best practices. Ultimately, we hope that networks can be secured proactively from such research instead of primarily reactively.

CHAPTER 5

Exploring the Tradeoffs of Network Takedowns

We observe the existence of "bad" organizations that consistently have a high fraction of their IP addresses involved in malicious activities. In response to these rogue Internet organizations, network operator community have turned to a form of vigilante justice we call *network takedowns*. Network takedowns is one of the mitigation solutions at the organizational level, which aims to eliminate rogue Internet organizations that provide bulletproof hosting for malicious activities [11] by pressuring upstream provider to de-peer from or refuse to transit a target network's traffic. Since 2007, more than ten disreputable ISPs have been successfully taken down [11], including Russian Business Network [58], Atrivo [15], and McColo [6].

At first blush, these high profile actions appear to have merit—the takedown of McColo, for example, contributed to a two-third reduction in global spam traffic [45]. But recently, the community has begun to question both the legal and ethical basis for such takedowns as well as, perhaps more importantly, their ad hoc nature:

> "Some have suggested that ISPs and Internet backbone providers should not be allowed to serve as judge, jury and executioner of problematic customers...[This] stance was echoed by Marcus Sachs, director of the SANS Internet Storm Center. 'There are others out there who need to be cut off but we've got to find a better way to do it than by creating the virtual equivalent of a lynch mob.'" [30]

While the thorny legal and ethical issues offer opportunities to test recent ethical assessments of Information Technology Research, such as those espoused in the Menlo Report [64], we are most intrigued by the ad hoc nature of the evaluations used by the community to assess a potential takedown target. In this thesis, therefore, we propose a more principled quantitative approach to measure the costs and benefits for such takedown decisions. Specifically, we propose sets of cost and benefit metrics, ranging from security, Internet naming, routing and transit. The cost metrics include the collateral damages to both legitimate services and innocent networks. The benefit metrics cover not only the security gain, but also the improvement in control plane stability and network workloads. To measure the cost and benefit metrics, we assemble seven Internet-scale datasets covering network attacks, routing, and naming, as well as the traffic collected by a regional Internet Service Provider. We further apply Pareto efficiency on the individual measure of costs and benefits to find the optimal operating points of takedowns that optimizes benefits while minimizing costs.

The framework can be used for evaluating any future takedown actions. To illustrate our framework, we apply the cost and benefit analysis on a set of candidate networks for takedowns. Specifically we identify 672 candidate autonomous systems (ASes), which had a significant fraction of their IP addresses engaged in acts of malice based on the 11 reputation blacklists collected during a period of one year. While applying our metrics, we have the following findings:

- These 672 disreputable autonomous systems contributed 20% of all malicious IP addresses on our collected blacklists, and 7% of sites hosting Malware detected by Google Safe Browsing.

- Compare to other networks, disreputable ASes play less critical roles in the Internet naming and routing infrastructures (e.g., hosting statistically significant fewer name servers, or having smaller degree in routing topology). Therefore, the takedown of them only results in minor costs to innocent users and services.

- The disreputable organizations not only perform maliciously in network security, but also have undesirable behaviors in other Internet functions. They contribute significantly more to the BGP instability and DNS workloads than other ASes. Therefore, the takedowns have moderate benefits on other Internet core functions.

- Applying Pareto efficiency on the costs and benefits of takedowns, we identified sets of optimal operating points with favorable cost-benefit tradeoffs. Conservatively, taking down 14 ASes (8 enterprise customers and 6 service providers) would eliminate 1% of malicious IP addresses (about 96K malicious IPs) in our collected blacklists with very minor costs. At a more aggressive operating point, taking down 153 enterprise customers and 24 service providers would remove 7% of malicious sources at the cost of 0.4% ASes becoming unreachable.

Although the cost-benefit tradeoff is impressive, when we explore the question *what are the global impacts of large numbers of these takedowns?*, the case for such community policing is not as clear. By taking down the 14 candidate ASes, we only eliminate 1% of malicious IP addresses in our view of the Internet. Even by taking down more than six hundreds of disreputable autonomous system, only 20% of malicious IP addresses would be removed. When taken along side the dismaying anecdotes decrying the success of previous takedowns (e.g., "By the second half of March, seven-day average spam volume was at the same volume we saw prior to the blocking of the McColo ISP in November 2008 [44]."), our observations about the modest global impact of network takedowns highlight the fact there are likely to be no quick and dirty solutions to the pervasive security problems we face. Instead such minor victories " will be short but sweet [45]."

5.1 Tradeoff Analysis Framework

In this section we describe a tradeoff analysis framework aimed at understanding the costs and benefits associated with a given takedown decision. Specifically, we want to answer two questions: *What are the costs and benefits of takedowns and how to measure them?* and *What is the optimal operating points for takedowns where certain security benefit is achieved with minimal costs?* To answer the first question, we select various metrics ranging from network security, Internet naming, routing, and transit and use normalized global impacts to measure them. For the second question, we use Pareto efficiency to find the optimal operating points which maximize security benefits while minimizing the costs. The framework takes as input a given set of candidate networks for takedown, which may

	Metrics	**Internet Functions**
Security Benefits	Reduced Malicious IP addresses	-
	Reduced Malware sites	-
Costs	Invalid Domains	Naming
	Unreachable ASes	Routing
	Loss of untainted traffic	Transit
Other Benefits	Reduced TLD queries	Naming
	Reduced BGP Instability	Routing

Table 5.1: Summary of cost and benefit metrics.

be arbitrarily defined. Therefore, our methodology can be used to evaluate any future takedown actions. Although the Pareto solution provides guidelines for takedown decisions, we believe there should be some operation guidelines to facilitate any such decision, which is out of the scope of this work.

5.1.1 Network Abstraction for Takedowns

There are different network abstractions, such as autonomous systems [6, 58, 15], netblocks [54], or groups of servers [33], at which takedown actions can be conducted. In this thesis, we assume that the takedowns will be conducted at the level of autonomous systems. This is in accordance with most of historical takedown actions and has three merits. First, autonomous systems are administrative boundaries that can reflect the networks' policies. So it is a natural choice for capturing rogue networks [117]. Secondly, autonomous systems are the entities of routing policy. Therefore, it is the level where de-peer by upstream providers is feasible [30, 35]. Thirdly, AS level provides a useful benchmark for tradeoff analysis. It represents the worst case scenario which has the largest costs. A finer abstraction, for example, of the prefix or administrative domains levels, provides opportunities for finer-grained policies and more favorable tradeoffs. In practice, the choice of best network abstraction should be determined on the basis of operation, which is out of the scope of this work.

Security	Malicious IP addresses from a collection of 11 blacklists Sites hosting Malware detected by Google Safe Browsing
Naming	Zone snapshot of second-level domains in .com and .net .com and .net TLD queries seen by Verisign TLD servers
Routing	Routing tables collected by Route Views and RIPE projects BGP updates seen by Route Views and RIPE
Traffic	NetFlows collected at a large regional ISP

Table 5.2: Summary of datasets used in measuring costs and benefits.

5.1.2 Cost and Benefit Metrics

To evaluate the costs and benefits associated with takedowns, we propose three sets of metrics as summarized in Table 5.1. These metrics have a broad coverage ranging from network security, Internet naming, routing and transit. And to make fair comparisons between different metrics possible, we use normalized global scale measures for each metric. As listed in Table 5.2, we collect seven global-scale datasets to measure the costs and benefits. Although these metrics have been carefully selected, we make no claim that they are complete or without bias. Next we will discuss each of these metrics with respect to the reason of their inclusion and measurement methodology.

5.1.2.1 Security Benefit Metrics.

Our first set of metrics aims to measure the global impact of takedowns on malicious sources. The first question we want to answer is *what percentage of malicious IP addresses would be removed after the takedowns?* To answer this question, we collect 11 IP address-based reputation blacklists (as described in Section 2.1) and calculate the reduction in the number of unique malicious IP addresses in these blacklists to represent the security benefit of takedowns. This is an unweighted measure of the improvement in security because we treat each malicious IP equally.

To cross-validate the security benefits, we use an empirical metric — the reduction in Malware sites detected by Google Safe Browsing [22]. Other security metrics, for example, the reduction in spam traffic or phishing websites, could also be added to our framework. However, we do not explore them in the present study.

5.1.2.2 Cost Metrics

The costs of network takedown are mainly the collateral damages to legitimate services and innocent users. We proposed cost metrics that cover three core Internet functions — Internet naming, routing, and transit.

For Internet naming, the most important collateral damage from takedowns is the disruption of legitimate services. Given the design of the DNS system, domains are only reachable in practice if they have at least one reachable authoritative name server. The takedown of ASes would make domains whose authoritative name servers are located in blocked ASes unreachable to the rest of the Internet. Therefore, we use the metric of *the percentage of domains becoming invalid after takedowns* as a proxy for this cost. We collect Verisign zone information snapshot that contains the authoritative name server information for all second-level domains under .com and .net. and quantify this cost by examining the effect on accessibility of all .com and .net domains, which account for half of all registered domain names on the Internet [59].

From a routing perspective, the takedowns might not only knock off the targeted networks, but also render their customers disconnected from the Internet. Therefore, we measure the cost as *the percentage of untargeted ASes that become unreachable after takedowns.* We measure this cost based on the routing tables collected by Route Views [104] and RIPE [42].

Similarly, for network transit, the cost is the loss of legitimate traffic. Generally speaking, it is very difficult, if at all possible, to measure the volume of purely malicious/legitimate traffic. Instead, we provide a rough estimate of the malicious traffic utilizing the tainted traffic approach [126], where tainted traffic is defined as traffic whose source or destination IP address is shown to be used for malicious activities (e.g., appears on a blacklist), while untainted traffic is the complement set. The untainted traffic represents a lower bound of legitimate traffic. We use the *the fraction of untainted traffic that is lost following a takedown* as the cost metric in transit. To measure the transit cost, we collected Netflow traffic from a regional ISP that provides high-performance computer networking and related services to educational, government, healthcare, and non-profit organizations.

5.1.2.3 Other Benefit Metrics

Besides security benefits, we select two metrics which measure the possible improvements in the Internet core functions. Although these benefits are not the primary goal of takedowns, they are interesting and welcome if present.

For naming infrastructure, we propose a benefit metric of the reduced DNS queries. As the naming systems are intensively used for botnet communication, we expect there would be a reduction in the queries after taking down disreputable ASes. This benefit is measured as *the fraction of global top-level domain queries that are reduced after takedowns*. We measure the metric of lighted name server loads with a dataset that consists of DNS queries captured at four geographically distributed VeriSign .com and .net TLD server clusters (Dulles, VA; New York, NY; San Francisco, CA; and Amsterdam, NL).

We also propose to use the BGP instability metrics to measure the potential benefit in routing. BGP routing updates happen for a variety of reasons, such as device failures, reconfiguration, and policy changes. Each change causes overhead due to issues such as convergence delay and network congestion. The stability of BGP routing tables has drawn considerable attention [98] as it is critical to the routing operation. We evaluate *what fraction of BGP update events can be reduced after takedowns* to measure the benefit metric of reduced BGP instability.We use routing updates collected from Route Views [104] and RIPE [42] to measure this benefit.

5.1.3 Pareto Efficiency Analysis of Costs and Benefits

Based on the costs and benefits data we measured, we further answer the question of *how to make takedown decisions to achieve optimized benefits while minimizing costs?* We apply Pareto efficiency [78], which compare the costs of different solutions for achieving certain benefit and then selects the most effective solutions that cannot be outperformed by other solutions. In our study, denote the cost for naming, routing and transit as c_1, c_2, c_3 and the security benefits of takedowns as b. All these parameters are functions of N which is the number of top candidates for takedowns. Our goal is to find Pareto optimal operating points of N such that no other operating point would gain the same benefit with lower costs.

A common way of approximating the Pareto optimal solutions is to solve a series of optimization problems formed by a linear combination of the cost and benefit functions. We first specify a set of numbers $\{\omega_1, \omega_2, \omega_3\}$ on the hyperplane $\{\omega_1, \omega_2, \omega_3 : \omega_1, \omega_2, \omega_3 \leq 1, \omega_1 + \omega_2 + \omega_3 \leq 1.\}$. Then we formulate the following optimization problem to find an optimal N

$$\min_N \quad \sum_{i=1}^{3} \omega_i \cdot c_i - (1 - \sum_{i=1}^{3} \omega_3) \cdot b$$

Ideally we should formulate and search such N over all possible points on the hyperplane, which is impossible as the above set is infinite. The way we simplify the search is to uniformly quantize the $[0,1]$ interval and select ω_is only on the grids, for instance take ω_i as $[0, 0.1, 0.2, ..., 1]$ with an interval being 0.1. After solving and getting the optimal N for each $\{\omega_1, \omega_2, \omega_3\}$ we obtain a set of N which approximates the Pareto optimal solution set.

The Pareto efficiency provides a guideline for making takedown decisions. However, in real operations, a careful cost-benefit analysis [14], in which the costs and benefits are compared after being unified into a single measure (e.g., money, infected population, etc.), is needed to evaluate the soundness of a decision. Such conversion and analysis is a highly non-trivial process as it depends on operational considerations, which is beyond our measurement capabilities. This remains an interesting direction of future work.

5.2 Application

In this section, we will illustrate the application of our tradeoff analysis framework. We first identify a set of disreputable Autonomous Systems, which are candidates for takedown, based the 11 blacklists collected over the period of one year starting from March 2014. Then we measure each cost and benefit metrics of taking down the top N candidates with our Internet-scale datasets. In the end, we apply our Pareto efficiency analysis to provide guidelines for the choice of N.

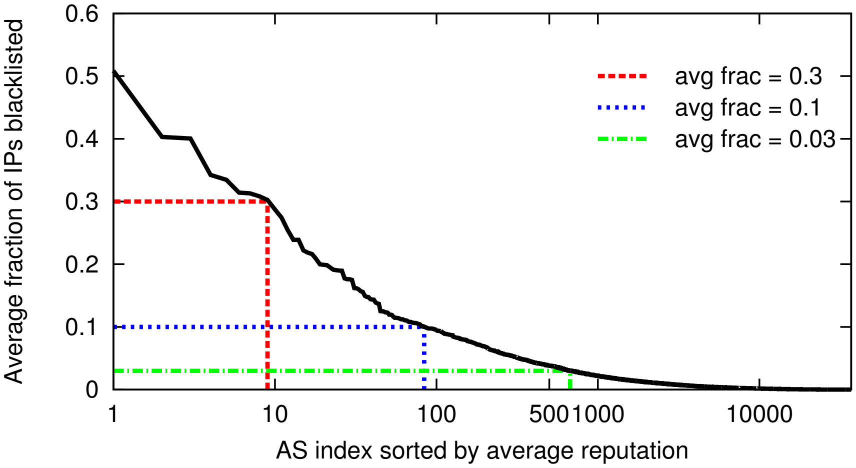


Figure 5.1: Average Maliciousness of Autonomous Systems over a period of one year. There are ASes with very high level of maliciousness—the top nice disreputable ASes have an average of more than 30% IP addresses are malicious during the year; more than 10% of IP addresses are blacklists for the top 84 disreputable ASes; and more than 3% IP addresses for the top 672 disreputable ASes.

5.2.1 Identifying Disreputable Organizations

The evidences that motivated historical takedown are external security reports showing that major portions of the network were being used for malicious activities [30, 35]. In consistence to the operational practices, we identify potentially takedown organizations based on their maliciousness. Specifically, we collect 11 IP address-based reputation blacklists over a period of one year starting on March, 2014 and calculate the organization maliciousness as described in Section 3.1.2. To identify disreputable organizations as candidates for takedown, we calculate the average maliciousness of each autonomous system over the period of one year.

We ranked autonomous systems by their average maliciousness over the year and show the average in Figure 5.1. Overall, 33,717 autonomous systems had at least one IP addresses blacklisted during the year. Similar to our observations in previous chapters, the average maliciousness varies greatly for different ASes: most autonomous systems had only a very small fraction of their IP addresses involved in malicious activities; but on the other hand, we do observed the existence of disreputable organizations that have a large fraction (as high as 51%) of their IP addresses involved in malicious activities. The top nice disreputable ASes have an average of more than 30% IP addresses are malicious during the year. More than 10% of IP addresses are blacklists for the top 84 disreputable ASes, and more than 3% IP addresses for the top 672 disreputable ASes. These observations confirm the existence of disreputable organizations that are subjected to be taken down. For the following cost and benefit analysis, we will focus on the top 672 ASes that had an average of more than 3% of their IP addresses blacklisted during the year.

We note that our method, which gives equal weight for different malicious sources, can be improved in multiple ways. For example, different malicious types can be given different weights: the IP address that hosts a C&C server could be weighted more than a spamming IP address. As we mentioned that our framework can take an arbitrary set of candidates as input, it can be used to assess the candidates selected by other consolidation methods in the future.

In our following analysis, we break these disreputable ASes into enterprise customers (ECs) and service providers (SPs) by their topological position. The rational is that the two types of ASes play very different roles on the Internet, thus making the impact of takedown varied. For example, a transit service provider is in a more critical position in the Internet hierarchy and will have a higher impact on routing than a stub AS. For our categorization, we use the definition and categorization techniques proposed by Dhamdhere et al [79]. While 539 disreputable ASes are enterprise customers, 133 of them are service providers including both transit service providers and content/access/hosting providers.

5.2.2 Measuring the Costs and Benefits

In this section, we apply our framework on the selected disreputable autonomous systems and measure the costs and benefits of taking down the top *N* disreputable ASes.

5.2.2.1 Improvements of Security

5.2.2.1.1 Reduced Blacklisted IPs We first look at what fraction of malicious IP addresses in our union list can be culled if one takes down the top *N* disreputable ASes. As we evaluate the benefit within our collection of blacklists that are also used for identifying

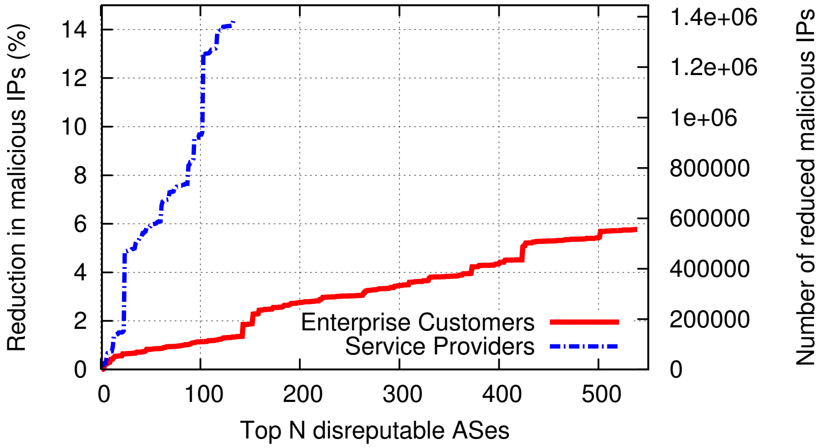


Figure 5.2: Benefit on Security — Reduced malicious IP addresses. There is a moderate improvement in security. About 20.2% of blacklisted IPs would be eliminated after taking down the 672 disreputable ASes.

disreputable ASes, our results represent the upper bound of the security gains.

We collected the 11 reputation blacklist for three months starting from March 2015. On average, there are 9.6 million malicious IP addresses on these lists on each day. We calculate the average fraction of malicious IP addresses that would be eliminated after taking down the top *N* disreputable service providers/enterprise customers as a function of *N* and show the results in Figure 5.2. It can be seen that the security improvement increase linearly with the increase of *N*. In total, the 672 disreputable ASes contribute about 20.2% blacklisted IP addresses. Taking down the 133 disreputable service providers would eliminate 14.4% of total malicious IP addresses, while only 5.8% of total malicious IP addresses would be removed after the takedown of 539 disreputable enterprise customers. The limited impact of taking down enterprise customers is cause by the small size of these ASes. Although they have a very large fraction of their IP addresses blacklisted, the absolute number of malicious IP addresses in them only accounts for a small fraction of total malicious IP addresses.

We also note that more than 80% of the blacklisted IPs used for identifying disreputable organizations fall into the category of spam. Therefore, the identified disreputable organizations might be biased toward the malicious activities of sending spam and the most profound security gain would be the reduction in spam. By breaking down the reduction

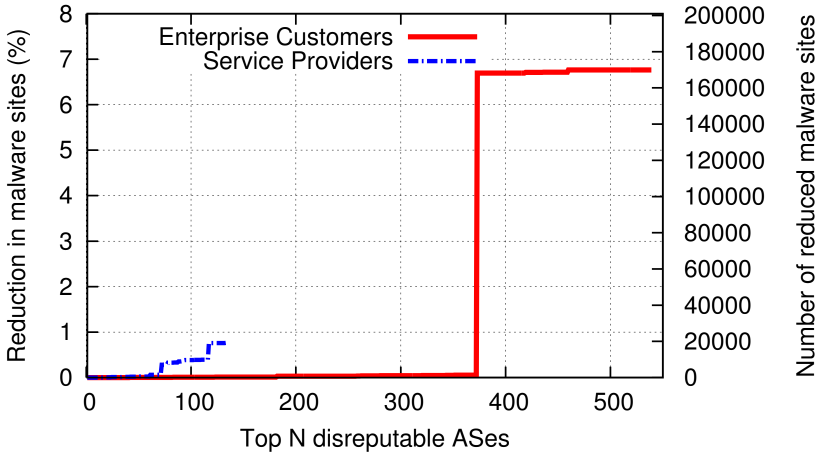


Figure 5.3: Benefit on Security — Reduced Malware site detected by Google Safe Browsing. About 7.6% of Malware sites would be removed after taking down the 672 disreputable ASes.

of malicious IP addresses into three different attack types, we see that taking down the top 672 disreputable AS would eliminate about 5% malicious IPs performing active scanning and 27% malicious IPs that hosting Phishing or Malware site. Therefore, although our selection methods are biased towards Spam sources, there are also moderate security benefits for other malicious activities.

5.2.2.1.2 Reduced Malware Sites The second metric we used to measure the security benefit is the reduction in malware sites detected by Google Safe Browsing [22] . We use the data collected on March 20, 2015 which includes all the sites hosting Malware in the past year. In total, there are 2.5 million malware sites in the dataset.

In Figure 5.3, we show the fraction of malware site that would be removed if one takes down the top *N* disreputable ASes. Only 0.8% of the detected malware sites are hosted in the 133 disreputable service providers, while about 6.8% are hosted in the 539 disreputable enterprise customers. However, the large spike indicates that malware sites are concentrated in certain networks while there are only several malware sites in other disreputable ASes. This concentration may be caused by the data collection method. As Google Safe Browsing does not scan the entire IP address space and the scan coverage varies from less than 1% to more than 99% for different autonomous systems. Therefore, our results could be affected

	Mean (disreputable)	Mean (others)	p-value
ECs	29.0	36.6	0.14
SPs	63.3	285.0	5.41e-09

Table 5.3: T-tests on the difference in number of authoritative name servers between disreputable ASes and other ASes. Disreputable service providers *host significantly fewer name servers* than other ASes at a 95% confidence level.

by the biased data sample.

5.2.2.2 Impact on Naming

5.2.2.2.1 Cost—Invalid Domains To quantify the cost in naming, we use the Verisign zone information snapshot taken on March 01, 2015. In total, there are 314 million second-level domains with 2.1 million resolvable name servers. For each domain, we map its authoritative name servers into ASes. In total, we find 25,063 ASes that host authoritative servers for at least one domain.

Given this data, we first examine *whether disreputable ASes plays less or more critical roles than others in the naming infrastucture.* Specifically, we compare the number of unique name servers hosted by disreputable ASes and the rest. The average number of hosted name servers of ASes in different groups are shown in Table 5.3. Both disreputable service providers and enterprise customers host less name servers than reputable ones. To further validate *whether the difference is statistically significant*, we use a *t-test*, which can examine the impact of the value of a binary variable on the value of other variables. To meet the normality assumption of t-test, we begin with a log-transformation and then apply a t-test to the transformed data. The null hypothesis H_o here is that there is no difference between disreputable ASes and others in terms of the number of hosted name servers. With a 95% confidence level, we can reject H_o if p-value$<$ 0.05. The results of t-tests are shown in Table 5.3. We can reject the null hypothesis for service providers and conclude the differences are statistically significant at a 95% confidence level.

This observation indicates that disreputable service providers are playing less critical roles in naming infrastructure, thus resulting in less negative impact of takedown actions. To quantify the cost on domain accessibility, we look at *the percentage of domains which*

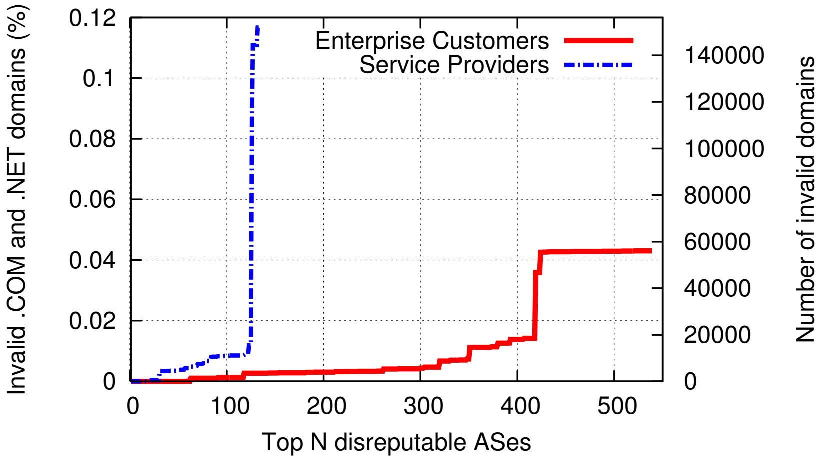


Figure 5.4: Cost on Naming: Unresolvable .COM and .NET domains after taking down the top *N* disreputable ASes. The cost is minor.

become unresolvable after taking down the top *N* disreputable ASes. We note that domains may have multiple authoritative name servers. A domain becomes unresolvable only when all of its authoritative name servers are unreachable. In our analysis, we do not distinguish malicious domains. This results in an overestimation of cost, since we take the removal of malicious domains as part of costs.

We show the results in in Figure 5.4 and it can be see that the cost on domain availability is negligible. Only about 0.003% domains would be affected after taking down the top 200 disreputable enterprise customers; and taking down all the 539 disreputable enterprise customers would affect 0.04% of domains. And as expected, taking down service provides has much larger negative effects. About 0.12% of domains would be unresolvable after taking down the 133 disreputable service providers.

5.2.2.2.2 Benefit — Reduced TLD Loads We then measure the benefit metric of lighted name server loads with a dataset that consists of DNS queries captured by VeriSign TLD server clusters on Dec 23, 2013. The data contains nearly 4.1 billion queries from 3.6 million unique resolvers.

We first map the IP addresses of resolvers into ASes and calculate the number of TLD queries per AS. Again, we first compare the TLD querying behavior of disreputable ASes and others. We perform the t-tests as before to see *whether there are statistically significant*

	Mean (disreputable)	Mean (others)	p-value
ECs	102,729	26,508	$< 2.2\text{e-}16$
SPs	474,560	382,528	6.7e-13

Table 5.4: T-tests on the difference in number of TLD queries between disreputable ASes and other ASes. Both disreputable ASes enterprise customers and service providers *sent significantly more queries* than other ASes at a 95% confidence level.

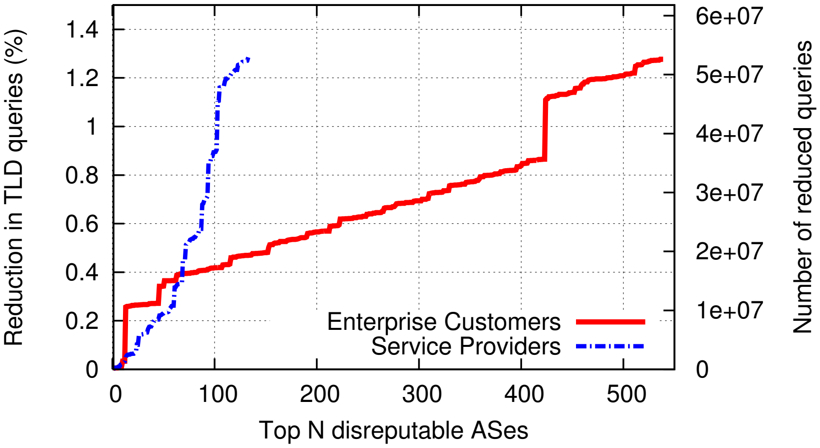


Figure 5.5: Benefit on Naming: TLD queries reduced after taking down the top *N* disreputable ASes. There would be a moderate reduction in TLD queries.

differences in the number of queries sent by disreputable ASes and the rest. The results are shown in Table 5.4. From the mean number of queries per AS, we can see that disreputable ASes sent more queries on average than other ASes in the same category. Especially, the disreputable enterprise customers sent about four times more queries than the other ECs. And the t-tests confirm that the differences are statistically significant at a 95% confidence level (i.e., p-value< 0.05).

Although we do not know why the disreputable ASes contribute more TLD loads, taking them down would definitely have benefits. We show the reduction on the number of queries after taking down the top *N* disreputable ASes in Figure 5.5. The improvements are moderate. Taking down the 133 disreputable service provides would eliminate 1.3% of global TLD queries, while taking down all the disreputable enterprise customers would result in a 1.3% reduction.

Summary of Impact on Naming: The disreputable ASes plays an less critical role in

	Mean (disreputable)	Mean (others)	p-value
ECs	2.1	2.1	0.63
SPs	26.4	48.8	1.9e-13

Table 5.5: T-tests on the centrality (degree) in AS-graph between disreputable ASes and others. The disreputable service providers occupied significantly less central positions in routing topology than other providers at a 95% confidence level.

the naming infrastructure than the rest of ASes, as they on average host less authorization name servers. Taking down the disreputable ASes would have minor cost on domains accessibility ($\sim 0.1\%$). Meanwhile, these disreputable ASes contribute more TLD queries than others. Thus taking down them would have positive effects of reducing TLD loads by more than 2.5%.

5.2.2.3 Impact on Internet Routing

5.2.2.3.1 Cost—Loss of Reachability Next, we measure the cost of takedowns in Internet routing reachability. Before we evaluate the actual cost, we first compare the disreputable ASes to reputable ones with respect their roles in the routing topology. We calculate the node degree of each AS in the AS-graph constructed from the routing tables in March 1, 2015. The higher the degree, the more critical role played by the AS in routing topology.

We show the mean degree of ASes of different groups in Table 5.5. It can be seen that disreputable service providers have smaller degree than other ASes in the same category. And the t-test results indicate that the differences are statistically significant at a 95% confident level. Since enterprise customers are the edge networks, there would not be differences for disreputable ones from others.

We measure this impact by examining the ASes other than disreputable ones that become unreachable by any monitor from Route Views or RIPE after removing all paths involving the top *N* disreputable ASes. We only use the explicitly announced routes in BGP table snapshots without inferring any paths in this exercise. It has been shown that the Route Views and RIPE data fails to capture the complete Internet topology. A significant fraction of peering and backup links are missed [72]. Therefore, our approach, again, represents the worse case because the loss in reachability is likely smaller in reality than what

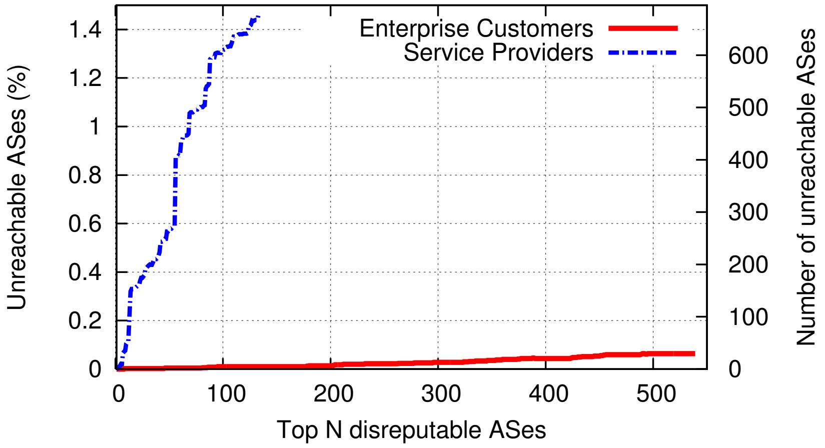


Figure 5.6: Cost on Routing: ASes that becomes unreachable after taking down the top *N* disreputable ASes. The cost is minor for taking down enterprise customers.

we see here.

As one might have expected, Figure 5.6 shows that taking down service providers has much higher cost on reachability than enterprise customers. Since some of them provide transit services, taking them down would make their customers unreachable from other parts of the Internet. On the other hand, taking down edge networks has little impact, even if hundreds of them are taken down. Overall, about 1.5% of ASes would become unreachable after taking down the disreputable ASes.

5.2.2.3.2 Benefit—Reduced BGP Instability We use the BGP updates seen by Route Views and RIPE on March 01, 2015 to measure the benefit of improved BGP stability. Rather than simply counting the number of updates for a given destination AS, which may lead to counting a single routing change multiple times, we use the method proposed by Rexford et al. [109], which is to collapse updates that occur close in time (within 45 seconds) and that are associated with the same destination prefix into a single update event. In total, we observe 144 million update events.

We first seek to answer the question of *whether disreputable ASes generate more or fewer update events than others.* We show the mean number of update events for each group in Table 5.6. It can be seen that the disreputable ASes announced more BGP update events than other ASes. And the t-tests confirm that the differences are statistically significant at a

	Mean (disreputable)	Mean (others)	p-value
ECs	5,756	1,543	9.3e-06
SPs	33,797	11,598	0.0005

Table 5.6: T-tests on the difference in number of update events between disreputable ASes and others. The disreputable ASes announced significantly more BGP update events than other ASes at a 95% confidence level.

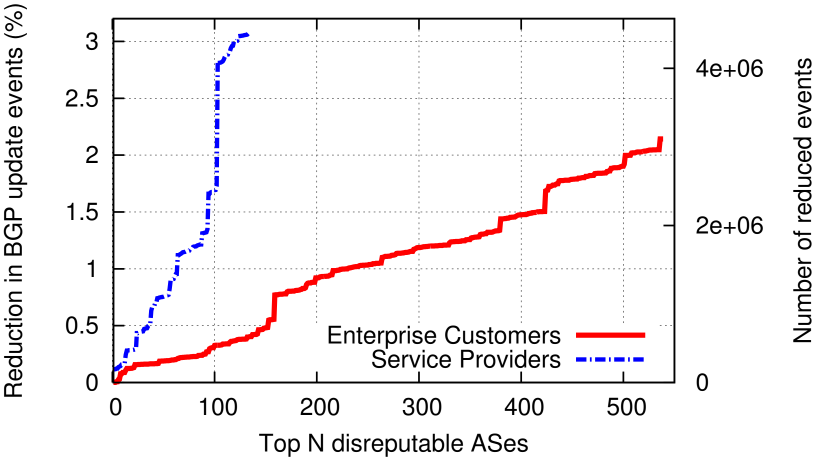


Figure 5.7: Benefit on Routing: BGP update events that would be eliminated by taking down the top *N* disreputable ASes. There is a moderate improvement in BGP instability.

95% confidence interval.

Given the finding that disreputable ASes generate more update events than others, we expect to see an improvement in stability with network takedown. Therefore, we examine *what fraction of BGP instability is mitigated if we take down the top N disreputable ASes.* In Figure 5.7 we show the fraction of update events that can be eliminated by taking down the top *N* disreputable ASes. About 3.1% of update events would be eliminated after taking down the 133 disreputable service providers. In addition, taking down the 539 disreputable enterprise customers would also remove about 2.1% of update events. To conclude, we see a moderate improvement in BGP instability by taking down the disreputable ASes.

Summary of Impact on Routing: The disreputable ASes play less central position in the routing topology. Thus, taking down enterprise customers would only result in minor cost on reachability while taking down service providers would result in moderate cost on reachability. Meanwhile, these disreputable ASes announce more BGP updates, which

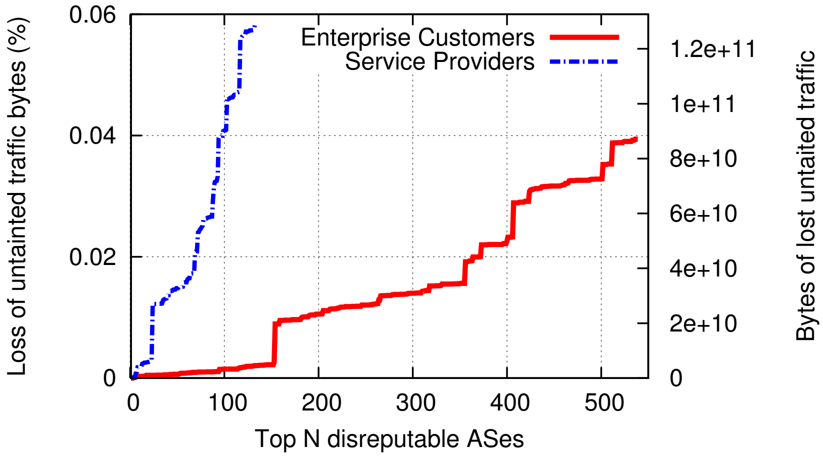


Figure 5.8: Cost on Transit: Loss of untainted traffic after taking down the top *N* disreputable ASes. The cost is minor.

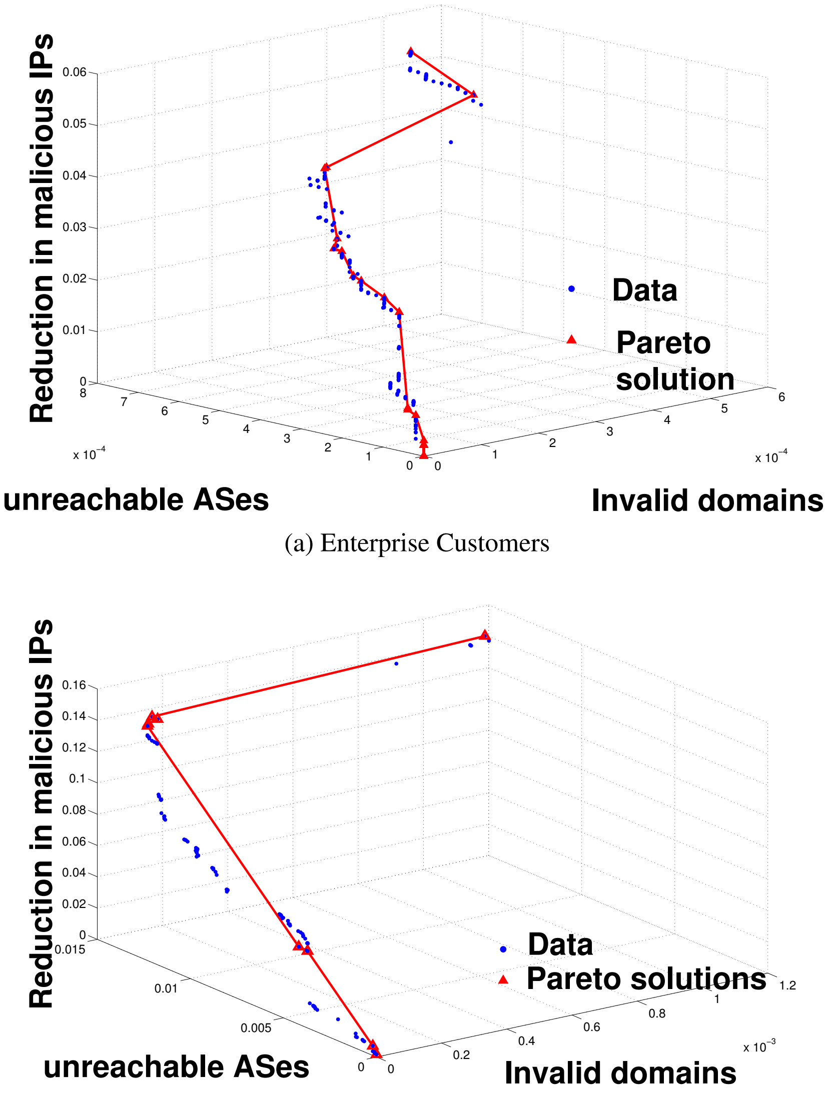
is an undesirable property. The takedown would moderately improves BGP stability by eliminating about 5% update events .

5.2.2.4 Impact on Transit

5.2.2.4.1 Cost—Loss of Untainted Traffic To measure the costs in transit, we use a dataset that consists of NetFlow traffic records collected from a regional ISP for a period of one week starting on March 17, 2015. Figure 5.8 shows the loss of untainted traffic after taking down the top *N* disreputable service providers/enterprise customers. It can be seen that the regional ISP would lose about 0.06% and 0.04% untainted traffic when taking down the top 133 disreputable service providers or the top 539 disreputable enterprise customers respectively. In total, less than 0.1% of untainted traffic would lost when taking down all the disreputable ASes. Therefore, the cost is minor.

5.2.3 Pareto Efficiency Analysis

Throughout the preceding analysis, we have deliberately avoided the question of how many disreputable ASes should be taken down (i.e., the choice of *N*). In this section, we use Pareto efficiency analysis to give recommendations on the choice of *N*. Here, we want to justify the takedowns from quantitative perspective and provide recommendations.



(a) Enterprise Customers

(b) Service Providers

Figure 5.9: An illustration of Pareto solution for the choice of N. Red triangles are Pareto solutions.

	Type		**ECs**		**SPs**	
	(*N*)	**(8)**	**(58)**	**(153)**	**(6)**	**(24)**
Security	*Reduced malicious IPs*	0.30%	0.86%	2.28%	0.62%	4.88%
	Reduced Malware sites	0.000%	0.003%	0.012%	0.000%	0.011%
Costs	*Invalid domains*	0.000002%	0.000027%	0.002779%	0.000003%	0.000320%
	unreachable ASes	0.000%	0.004%	0.010%	0.020%	0.376%
	Loss of untainted traffic	0.0001%	0.0004%	0.0013%	0.0004%	0.0062%
Other benefits	*Reduced TLD queries*	0.01%	0.37%	0.50%	0.01%	0.10%
	Reduced BGP instability	0.08%	0.20%	0.55%	0.14%	0.43%

Table 5.7: Costs and benefits with recommended choice of *N* based on Pareto efficiency.

The goal of Pareto efficiency is to find the optimal choices of *N* to maximize security benefits while minimizing collateral damage. To illustrate the analysis, we visualize the Pareto solutions with the costs in naming and routing, alongside with the reduction of malicious IP addresses as the benefit in Figure 5.9. The plots are generated for enterprise customers and service providers separately. In each Figure, the small nodes correspond to a (c_1, c_2, b) combination for a certain *N* while the line with triangles represents the Pareto solutions. The derived Pareto solutions capture a small set of *N*s that achieved the optimal operation points while eliminating the sub-optimal ones. In addition, the line between Pareto solution shows the trends of cost-benefit slope. With the choice of a small *N*, the benefit increases with limited increase of costs. However, with the increase of *N*, the costs increase much faster for the same increase in benefit. In addition, taking down enterprise customers is more effective than taking down service providers, as it can achieve the same benefit with a lower cost.

Based on the Pareto efficiency on all the three costs in naming, routing, and transit, we recommend a set of *N*s ($N = 8, 58, 153$ for enterprise customers and $N = 6, 24$ for service providers). The recommended choice of *N* is a Pareto optimal solution and has a favorable cost-benefit slope. To illustrate the global impact of such takedowns, we summarize all the costs and benefits we measured with the recommended *N*s in Table 5.7. Conservatively, if one takes down eight enterprise customers and six service providers, although the cost is very minor, only less than 1% of global malicious IP addresses would be eliminated. On the other hand, at a more aggressive operating points, taking down 153 enterprise customers and 24 service providers would remove 7% malicious sources at the cost of 0.4% ASes being unreachable.

There is a limitation of our current analysis—we do not unify different costs and benefits in the same terms, which would normalize the impact of different metrics and support more rational operational decisions. For example, one might expect that the monetary impact of an invalid domain would be much less than the loss caused by one inaccessible autonomous system. And similarly, the security benefit of eliminating control and command domain are expected to be higher than the benefit of taking down a spam machine. Therefore, it is an important future direction to transform the metrics into unified measures, such as money or affected populations, to facilitate operational decisions.

5.3 Implications

In the previous sections, we discussed how to move from ad hoc evaluations of network takedowns to a more principled approach for performing tradeoff analysis. Such analysis showed that there are 14 networks for which the security benefits of their takedown far outpace their costs. In evaluating the global impact of these take downs, however, we found that taking them down has little down impact (1%) on the global malicious sources. Even by being more aggressive, taking down all the 672 disreputable ASes, only contributes to 20% reduction in globally misbehaving hosts. In this section we briefly discuss what we believe to be the implications of this result.

Historically, high-profile takedowns resulted in significant drops in malicious activities. For example, the takedown of McColo in 2008 resulted in a two-thirds reduction in global spam traffic. Unfortunately, it only took four months for the spam level to return to its prior level after the takedown of McColo [43], as the attackers quickly restored their command-and-controls in other networks. What is worse, the trend of distributed cybercrime infrastructures significantly shortened recovery time. For instance, it only took three days for the spam level to be restored after the takedown of Real Host in August 2009. Unless the cybercrime infrastructures were comprehensively rooted out, any takedowns can only be able to temporarily disturb the malicious activities.

The limited security benefits we revealed are consistent with the trend of more distributed and resilient cybercrime infrastructures. The changes are, in fact, catalyzed by takedown actions, and attackers have made efforts to a void putting all of their eggs into one basket [7]. With more distributed infrastructures, the botnets would be better able to resist network takedowns. In addition, they act more stealthily to avoid exposing new ISPs as targets for takedowns [43].

Both the distributed nature and temporary effects of network takedowns indicate that they are not sufficient to solve security problems: "There's no point to aggressively pursuing malware networks if we don't choke up on their collective ability to regain Internet access, even/especially if the entire point of that access is to re-aim their networks at a different set of servers. This fight is already far too much like a game of Whack-a-Mole — there's no need to turbocharge the competition [34]."

So where do we go from here? One direction is to gain a deeper understanding on the interplay between attackers and takedown policies. When a takedown happens, the attackers will make efforts to restore their infrastructures. We expect the cost of rebuilding a rogue ISP that has been taken down would be high, as it likely involves efforts to allocate IP addresses and to connect to upstream providers. Therefore, one possible solution is to automate the whack-a-mole process or at least reduce the time it takes to detect and take down new rogue networks as they emerge. This would not only reduce the malicious activities, but also increase the cost and hassle of creating future rouge ISPs. And then the lack of bulletproof hosting services would further increase the cost to attackers, thus improving the security ecosystem in the long term.

Another direction is to understand the social, behavioral, and economic factors that influence the attackers [91, 92]. The models derived from this understanding have lead to more targeted and structural actions than the whack-a-mole nature of takedowns. For example, consider the spam-level reductions from takedowns which seem to always return to previous levels. It's only with the removal of the SpamIT affiliate program that we see sustained reductions in spam volume [46].

Another area of potential advancement is dealing with the factors that create these bad networks in the first place. From an economic point of view, security has long been recognized as a public good [102], for which the investment by one agent has positive effects on others. However, the existence of positive effects can create free-riders—networks that do not make any investment to improve Internet security, but instead relies on the contribution of others. However, if everyone is selfish, everyone is worse off than if they had all agreed on collective efforts. This is called the tragedy of commons, where “a party could efficiently prevent harm to others — that is, a dollar worth of harm could be prevented by spending less than a dollar on prevention — but the harm is not prevented because the party has little or no incentive to prevent harm to strangers [81].” Legal, policy, or market based solutions (e.g., cyber insurance) are needed to help incentivize good behavior.

5.4 Summary

While more than ten high-profile network takedowns have occurred over the last several years, the process for evaluating who deserves to be removed remains largely ad hoc. In this chapter we proposed a tradeoff analysis framework including examples of applicable cost and benefit metrics. Utilizing this framework, we show that there are autnomous systems for which the takedown would provide moderate benefits with little costs. While this is true for tens of individual organizations, we find that the global security gains of all such takedowns, even in aggregate, are limited. The whack-a-mole takedown actions increase the cost of attacks, but we argue there is no easy solution for maintaining network health. Rather we call for more work that addresses the social, behavioral, and economic factors behind both the attacker ecosystems and those of the network defenders. Such work can assist in creating appropriate disincentives for malicious behavior and incentivize health behavior in networks and, in the long term, succeed better than our current game of whack-a-mole.

CHAPTER 6

Related Work

In this section, we review the related work regarding the study of organizational security. From organizations' perspective, the study are commonly understanding the security problems faced by organizations and developing defensive mechanisms. We classify these studies by their purpose as measuring, detecting, mitigating, or understanding the threats.

6.1 Measuring security threats at the organizational level

There are numerous works on measuring the security postures and cyber incidents of organizations. For example, The annual Data Breach Investigations Report from Verizon lists the data breach incidents in organizations and measures the evolution of cyberthreats from the organization's perspective [16]. Among the measurement studies, the works focusing on the malicious activities of organizations and the mismanagement of organizations are most related to the studies in this thesis.

There are many evidences of the concentration of malicious activities in the "bad" organizations. For example, Ramachandran et al. studied the network-level distribution of Spam found that about 36% of the observed spam are originated from only 20 autonomous systems [107]. Similarly, as reported by Koddkdis et al, 90% of total observed spam volume came from 10% of spamming autonomous systems [94]. The "bad" organization have also been identified for other malicious resources. For example, Collins et al. studied the spatial uncleanness of malicious sources of botnets. They reveal that the compromised hosts are disproportionally concentrated in some organizations [73]. Stone-Gross et al. found that drive-by-downloads and phishing websites are also concentrated in rouge autonomous systems [117]. Shue et al. [111] use a union approach to combine and aggregate IP-based reputation lists into reputation of autonomous systems and examine the Internet connectivity properties of the malicious ASes. They found that malicious ASes have more frequent changes with their BGP peers.

For mismanagement, there have been numerous studies on the adoption and efficacy of various best practices that we build upon. For example, in 2009, Beverly et al. [66] performed an active measurement experiment from 12,000 clients in order to study the deployment of egress filtering. Their team showed that 31% of the clients are able to spoof any arbitrary routable source address and that 77% can forge an address within their /24 subnetwork. The results are consistent with our findings, and indicate a lack of anti-spoofing deployment improvement within the past 5 years. In 2002, Mahajan et al. showed that 90% of short-lived routes are caused by misconfiguration and that 0.2%-1% of the global routing table consists of misconfigured routes [101]. The study also found that these misconfigured routes had a variety of causes, including human error, configuration errors, and software bugs. In our study, we use their heuristics to define updates caused by BGP misconfiguration.

6.2 Detecting rogue organizations and Detecting threats at the organization-level

The detection studies from organization's perspective can be categorized into two categories. The first category is to detect rogue organizations, which provide bulletproof hosting service for malicious activities. Stone-Gross et al. developed FIRE, a project that aims to detect rogue autonomous systems based on malicious activities such as drive- by-downloads and phishing [117]. And a follow-up study found that it is also feasible to detect rogue organizations based on their control-panel activities such as routing rewire activities and IP space fragmentation and churn [95].

The second category of work is to detect cyber attacks based on the aggregated observation at the organization-level. For example, various network intrusion detection system (NIDS) are deployed on an organization's network to detect distributed denial-of-service attacks [13, 18] and botnet activities [75, 85]. And our previous work successfully detected compromised user accounts in two universities by looking at the account activities observed by the organizations.

6.3 Mitigating solutions at the organization level

While network takedown discussed in this thesis is a form of mitigation solution that the security community takes against rouge organization, there are mitigation solutions that organizations can deployed to reduce the risk or impact of security threats.

From security management's perspective, there exist a large number of best practices for specific services and for organizationally managing security, including ISO 17799 [28], the Information Security Forum [27], and Network Protection Practices [74]. As defenders in the security ecosystem, organizations can contribute to mitigating security problems by following these best security practices. On the other hand, active security systems, such as Network access control (NAC) and Firewalls are commonly deployed at organization network edges to mitigate cyberthreats. Lastly, organization can train its IT personal and do security education on users to proactively prevent threats from happening.

6.4 Understanding security threats at the organizational level

As a complement to these studies that focus on technical solutions, some researches have paid attention to understand the social, behavioral, and economic factors that influence the security investment and policies in organizations. For example, the Cyberthreat Defense Report [5] conducts surveys among IT security practitioners and aims to understanding their security concerns, practices, and investment strategies and facilitating comparison and experience sharing among different organizations. Our previous internal study in two universities aims to understand the effectiveness of security practice such as vulnerability scanning, password policy, and user educations [124]. It shows that all these security practice can significantly reduce the user susceptibility toward identity theft. Jiang et al. studies the selfish behavior in Internet security investment while treating security as public goods [88]. It shows that the improvement of technology alone may not improve the security ecosystem if there is a lack of incentives in security investment. Gill et al. studied the low deployment rate of Secure BGP and argue that it is caused by the lack of incentives as the Secure BGP is a public good [83]. Then they proposed a strategy that adds economic incentives to drive the deployment.

CHAPTER 7

Conclusion and Future Work

This thesis has demonstrated how a macroscopic view of network security at the organizational level can be used to measure, understand, and mitigate security threats. While malicious activities have evolved considerably over the past decade, Internet organizations are chasing after the attackers. The central premise of this thesis is to help organizations systematically understand their security problems and to then develop defensive strategies. In the four primary chapters, the research presented in this thesis demonstrates how to leverage Internet-wide observations to study security problems at the organizational level, including: measuring the impact of malicious sources on organizations, characterizing the longitudinal evolution of security threats, understanding the root cause of maliciousness, and then mitigating threats via network takedowns. In the following sections, we summarize our contributions and key insights from this thesis, and discuss future work.

7.1 Summary of Contributions

This thesis begins with describing the collecting and characterizing of a set of reputation blacklists. Reputation blacklists are widely used by organizations to make security policies, such as blocking and filtering. Existing work has studied how these lists can be created [63], evaluated for their effectiveness [90, 114], and then further explored for the properties of the networks that make them effective [120, 123, 111]. For this thesis, tens of IP-based, widely-used, commercial reputation blacklists, which cover Spam, Phishing/Malware, and active scanning, were collected on a daily basis and analyzed. By analyzing the lists, we find that the size of these blacklists are relatively stable, but that some of them have a high turnover rate. The blacklists in the same attack category have significant internal entry overlap, but little similarity is seen across the different attack categories.

Rather than focusing solely on the lists themselves, this thesis analyzes their impact on an Internet organization — Merit Networks, a large Internet Service Provider. By examining which part of network traffic is tainted by these blacklists, we gain better insight into the utility of these mechanisms and the nature of malicious traffic on our networks. We find that a surprisingly high proportion, up to 17%, of the collected network traffic is tainted by at least one of the blacklists. In addition, our network only saw traffic to a small portion (between 3% and 51%) of IP addresses within the blacklists, indicating there is a difference between the global view and the organizational view of the security threats.

Equipped with these IP-based reputation blacklists, we then formed a measure of organization maliciousness which is as the fraction of the organization's IP addresses that are blacklisted. In general, one can expect hosts in a network to be governed by similar network policies, such as DHCP pool blocks, workstations with the same operating systems or patch levels, firewall policies, password strength checks, and even user-awareness levels. This observation hints at the potential benefit of aggregating IP-based reputation lists. By aggregating individual IP addresses, one can form an equivalent measure of reputation, but perhaps a measure that is more stable; we call this *organization maliciousness*. Specifically, we quantify the security benefits of aggregation, compared to blocking, using the unaggregated list of malicious IP addresses. Our results show that the aggregated reputation can achieve greater persistency and predictability of entities with poor reputations, thus being more effective in filtering out malicious IP addresses.

Then, this thesis studies the longitudinal evolution of the Internet threats landscape with respect to the organization maliciousness. We characterize the malicious behavior of a network by both the level of maliciousness (magnitude) and by how frequently the maliciousness changes (dynamics). We find that the distribution of organizations' malicious magnitude varies greatly and that there are "bad" organizations that consistently have a very high magnitude of maliciousness. We also show that over a period of two and half years, both the magnitude and dynamics of maliciousness have been increasing. While taking the dynamics as a proxy of responsiveness, we find that organizations have become faster in responding to malicious activities. However, the impact of the response is not as lasting because the average magnitude of maliciousness has increased significantly.

Inspired by the observation that maliciousness varies greatly across different organizations, this thesis then explores the cause for the maliciousness of organizations. To do so, we systematically examine the relationship between security mismanagement and maliciousness by analyzing eight Internet-scale mismanagement metrics. Through this analysis, we find that different symptoms of mismanagement are highly correlated to organization maliciousness, and we ultimately find a causal relationship between mismanagement and maliciousness, while controlling for social and economic considerations.

Lastly, after measuring and understanding organization maliciousness, this thesis discusses the investigations of one organization-level mitigation solution — network takedowns of organizations, specifically ones that consistently have a very large fraction of their IP addresses blacklisted. This thesis proposes a cost-and-benefit tradeoff analysis framework that includes examples of applicable cost and benefit metrics. Utilizing this framework, we show that there are networks for which the takedown would provide moderate benefits with little cost. While this is true for tens of individual networks, we find that the global security gains of all such takedowns, even in aggregate, are limited. The Whack-a-Mole takedown approach increases the cost of attacks, but we argue there is no easy solution for maintaining network health.

7.2 Insights and Future Work

In this section, we review several of the key insights learned from the research presented in this thesis and discuss potential future research directions.

7.2.1 Granularity of organizations

As discussed in Section 3.1.1, external studies of an organization's security face the challenge of sharing confidential operational information, in particular the organization's management boundaries. To deal with this, various proxies of organizational boundaries have been proposed in previous works. For example, autonomous systems [107], routing prefixes [120], and DNS administrative domains [106] are used to study the organizational-level behavior of spamming. However, most of the current work utilizes existing natural boundaries of other Internet core functions as proxies of organizations. For example, autonomous systems and routing prefixes are boundaries in Internet routing, while DNS administrative domains are boundaries in Internet naming. But the security management boundaries could be different from other management domains. Therefore, it is an important future direction to identify the right level of granularity based on security postures. For example, a hierarchical clustering [69] can be used to divide the address space into netblocks based on the similarity of security behavior, such as malicious behavior and management status.

Technically, the approaches proposed in this thesis can be extended to any arbitrary level of granularity. For example, as demonstrated in Sections 3.1.3 and 4.4.3, our methods are applicable to routing prefixes. But the result of the policies might differ when being applied on different level of granularity. For example, as shown in Section 3.1.3, the persistency and predictability of autonomous systems is better than that of routing prefixes, indicating that autonomous systems are the more stable and consistent entity for profiling and predicting organization behaviors. However, when we looked at actions, such as network takedowns, those at the autonomous systems AS level represented the worst-case scenario with the largest collateral damage across nearly all cost metrics. The question of what is the best granularity for different policies is out of the scope of this thesis, but remains an interesting direction to explore in the future.

7.2.2 External and internal study of organizational security

Studies from the organization's perspective can be conducted internally or externally. External studies, such as our research presented in this thesis, that rely on external observations, could have Internet-wide visibility with Internet scanning tools. But because of the confidentiality of operational information within organizations, it is hard to get detailed security management information to answer questions such as: Where is the management boundary of organizations? What is the organization's security management status? and How many security incidents happen every year? Instead, we need to use various proxies to infer the internal management status.

In contrast, an internal study is effective in better understanding security problems and developing solutions within the organization. However, it is hard to scale such internal studies to reflect global security postures or to achieve unified solutions for different organizations for two reasons. First, the security solution for one organization might be ineffective for others because of the heterogeneous nature of the Internet. Internet organizations vary in their functions, infrastructure, and user groups, which results in different security problems and solutions. Therefore, it might require non-trivial efforts to customize security solutions for different organizations. Second, cross-organization cooperation and information sharing is also hard to achieve. Without the cooperation among organizations, the security community cannot have broad visibility of the security threats.

Our current study is based solely on the external observation of an organization's malicious activities. In the future, we can conduct more studies such as these within organizations that help University of Michigan to detect emerging threats [125] and understand the effectiveness of current preventive solutions [124]. Another future direction is to cross-validate our observations, such as fluctuation of maliciousness, with the defensive action taken by the organization and the process of attack campaigns. Such correlation or causality could help us to better understand the utility of defensive policies and to facilitate more effective future actions. In addition, these studies can be extended to more organizations in the future.

7.2.3 Proactive and reactive defense

Traditional security is of a reactive nature. For example, most reputation blacklists are generated based on the observation of malicious activities. However, reactive mechanisms are ineffective due to the latency between exploit and detection. In most cases, by the time an attack is detected, it is already too late as the damage has occurred. In contrast, an effective proactive defense can substantially reduce the potential cost incurred by an incident. In this thesis, we discuss three mechanisms of proactive defense and elaborating on those possible directions for future work.

The first, proactive reputation shows the management status of organizations, and it can help the security community to target efforts to risky networks. Equipped with proactive reputation, we can knowledgeably answer questions when making policies. Is there a point at which a network becomes too dangerous to be allowed to remain connected to the public Internet? Is it appropriate to proactively blacklist open mail relays in SPAM filters or to drop DNS responses originating from known open recursive resolvers?

The second mechanism is to explore whether we can build systems to predict future attacks. Inline with the research presented in this thesis, we have successfully developed a method to predict future security incidents in organizations [100]. This work demonstrates that mismanagement and the historical maliciousness are effective predictors of future incidents. In addition, there are existing works that effectively predict future malicious sources [123] and compromised web servers [116]. In the future, we would like to work more to refine the prediction system to be more accurate and to cover even more security threats.

Lastly, proactive defense suggests more security considerations for computer systems. As shown in this thesis, there are a large number of misconfigured systems that fail to implement even the simplest patches and then ultimately pose a threat to the Internet as a whole. Proactive defense calls for better default configuration and automatic patch mechanisms for computer and network systems. Developing intelligent management tools for network operators can also be helpful to prevent the exploit of misconfigured systems. These are all directions for future studies.

7.2.4 Incentive to secure networks

This thesis identifies a large number of organizations that have a disproportional high number of malicious sources. An area of potential advancement is dealing with the factors that create these bad networks in the first place. From an economic point of view, security has long been recognized as a public good [102], for which the investment by one agent has positive effects on others. However, the existence of positive effects can create free-riders—networks that do not make any investment to improve Internet security, but instead rely on the contribution of others. However, if everyone is selfish, everyone is worse off than if they had all agreed on collective efforts. As pointed out by Felten [81], "A party could efficiently prevent harm to others — that is, a dollar's worth of harm could be prevented by spending less than a dollar on prevention — but the harm is not prevented because the party has little or no incentive to prevent harm to strangers." For example, the case of egress filtering poses a risk to the rest of the Internet, but poses little internal threat. As a result, organizations have little incentive to fix these services.

Recent work has shown that providing social or financial incentives may be more effective than developing new technical solutions for improving overall security [89]. As such, if we are able to develop strategies in which edge networks are incentivized to better manage their systems, we may be to able increase the stability of the Internet as a whole. In one example, Gill et al. propose a strategy for increasing BGP security in which network operators assign a higher priority to routes that adopt appropriate security measures. This increased traffic translates to increased revenue, serving as a financial incentive for securing networks [84]. However, the strategy is limited to BGP security. We show that one future research direction is to develop legal, policy, or market-based solutions (e.g., cyber insurance) to encourage better organizational network management.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] https://kb.isc.org/article/AA-00924/0.

[2] http://technet.microsoft.com/en-us/library/dd197515(v=ws.10).aspx.

[3] http://www.merit.edu/mail.archives/nanog/1997-04/msg00444.html.

[4] http://www.team-cymru.org/Services/ip-to-asn.html.

[5] 2015 cyberthreat defense report. http://www.cyber-edge.com/2015-cdr/.

[6] A Closer Look at McColo. http://voices.washingtonpost.com/securityfix/2008/11/the_badness_that_was_mccolo.html.

[7] A year later: A look back at McColo. http://voices.washingtonpost.com/securityfix/2009/11/a_year_later_a_look_back_at_mc.html.

[8] Akamai. www.akamai.com/.

[9] Alexa - Top Sites. http://www.alexa.com/topsites.

[10] Barracuda Reputation Blocklist. http://www.barracudacentral.org/.

[11] Bulletproof hosting. http://en.wikipedia.org/wiki/Bulletproof_hosting.

[12] CBL: Composite Blocking List. http://cbl.abuseat.org/.

[13] Cloudflare advanced ddos protection. https://www.cloudflare.com/ddos.

[14] Cost-benefit analysis.

[15] Cybercrime's U.S. Home. http://www.spamhaus.org/news/article/636.

[16] Data breach investigations report. http://www.verizonenterprise.com/DBIR/.

[17] DDoS & Security Reports: DDoS Protection. http://www.arbornetworks.com/asert/ddos-protection/.

[18] Ddos attacks & protection - arbor networks. http://www.arbornetworks.com/ddos-attacks.

[19] Deep Inside a DNS Amplification DDoS Attack. http://blog.cloudflare.com/deep-inside-a-dns-amplification-ddos-attack.

[20] DoS Attack against DNS? http://seclists.org/nanog/2006/Jan/294.

[21] DShield. http://www.dshield.org/.

[22] Google Transparency Report: Malware Distribution by Autonomous Systems. https://www.google.com/transparencyreport/safebrowsing/?hl=en.

[23] Hackers focus on misconfigured networks. http://forums.cnet.com/7726-6132_102-3366976.html.

[24] hpHosts for your pretection. http://hosts-file.net/.

[25] Importance of PTR records for reliable mail delivery. http://www.mxpolice.com/email-security/importance-of-ptr-records-for-reliable-mail-delivery/.

[26] Infographic: The State of Malware 2013. http://www.mcafee.com/us/security-awareness/articles/state-of-malware-2013.aspx.

[27] Information security forum. https://www.securityforum.org/.

[28] International standard ISO/IEC 17799:2000 code of practice for information security management. http://17799.denialinfo.com/whatisiso17799.htm.

[29] Internet has a garbage problem, researcher says. http://www.pcworld.com/article/144006/article.html.

[30] Internet Shuns U.S. Based ISP Amid Fraud, Abuse Allegations. http://voices.washingtonpost.com/securityfix/2008/09/internet_shuns_us_based_isp_am.html.

[31] Internet threats trend report. http://www.commtouch.com/uploads/2013/04/Commtouch-Internet-Threats-Trend-Report-2013-April.pdf.

[32] Introduction to Cisco IOS NetFlow. http://www.cisco.com/en/US/products/ps6601/prod_white_papers_list.html.

[33] Level 3 tries to waylay hackers. http://www.wsj.com/articles/level-3-tries-to-waylay-hackers-1432891803?mod=WSJ_article_EditorsPicks_2.

[34] McColo reconnect highlights network security gap. http://arstechnica.com/security/2008/11/mccolo-reconnect-highlights-network-security-gap/.

[35] Mccolo takedown: Internet self policing or vigilantism. http://www.computerworld.com.au/article/269267/mccolo_takedown_internet_self_policing_vigilantism/.

[36] Merit Network INC. http://www.merit.edu/.

[37] Multiple DNS implementations vulnerable to cache poisoning. http://www.kb.cert.org/vuls/id/800113.

[38] Open Resolver Project. http://openresolverproject.org/.

[39] OpenBL. http://www.openbl.org/.

[40] PhishTank. http://www.phishtank.com/.

[41] Putting the Spamhaus DDoS attack into perspective. http://www.arbornetworks.com/corporate/corporate/blog/4813-putting-the-spamhouse-ddos-attack-in-perspective.

[42] RIPE Routing Information Service (RIS) Raw data Project. http://www.ripe.net/data-tools/stats/ris/ris-raw-data.

[43] Spam back up to pre-McColo levels. http://www.securityfocus.com/brief/938.

[44] Spam data and trends: Q1 2009.

[45] Spam Volumes Drop by Two-Thirds After Firm Goes Offline. http://voices.washingtonpost.com/securityfix/2008/11/spam_volumes_drop_by_23_after.html.

[46] Spam Volumes: Past & Present, Global & Local.

[47] SpamCop Blocking List. http://www.spamcop.net/.

[48] Spoofer project. http://spoofer.cmand.org/index.php.

[49] SURBL: URL REPUTATION DATA. http://www.surbl.org/.

[50] Targeted attack campaigns and trends: 2014 annual report. http://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/targeted-attack-campaigns-and-trends-2014-annual-report.

[51] Team Cymru comumity Services. http://www.team-cymru.org/Services/ip-to-asn.html.

[52] The DDoS that knocked Spamhaus offline. http://blog.cloudflare.com/the-ddos-that-knocked-spamhaus-offline-and-ho.

[53] The SPAMHAUS project: SBL, XBL, PBL, ZEN Lists. http://www.spamhaus.org/.

[54] Threat spotlight: Sshpsychos. http://blogs.cisco.com/security/talos/sshpsychos.

[55] UCEPROTECTOR Network. http://www.uceprotect.net/.

[56] Verisign. Inc. www.verisigninc.com.

[57] WPBL: Weighted Private Block List. http://www.wpbl.info/.

[58] Russian Business Network study. http://www.bizeul.org/files/RBN_study.pdf, 2007.

[59] Verisign's future looks stable with .com and.net registries in the bag. http://www.forbes.com/sites/greatspeculations/2012/08/20/verisigns-future-looks-stable-with-com-/and-net-registries-in-the-bag/, 2012.

[60] DDoS strike on Spamhaus highlights need to close DNS open resolvers. http://www.techrepublic.com/blog/security/ddos-strike-on-spamhaus-highlights-need-to-close-dns-open-resolvers/9296, 2013.

[61] Open Resolver Project — Results from 3 months of active scans. http://www.nanog.org/sites/default/files/tue.lightning3.open_resolver.mauch_.pdf, 2013.

[62] Tanya Agrawal, David Henry, and Jim Finkle. Jpmorgan hack exposed data of 83 million, among biggest breaches in history. http://www.reuters.com/article/2014/10/03/us-jpmorgan-cybersecurity-idUSKCN0HR23T20141003, October 2014.

[63] Manos Antonakakis, Roberto Perdisci, David Dagon, Wenke Lee, and Nick Feamster. Building a Dynamic Reputation System for DNS. In *USENIX Security Symposium*, pages 273–290, 2010.

[64] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The menlo report. *IEEE Security and Privacy*, 10:71–75, 2012.

[65] D. Barr. Common DNS operational and configuration errors. RFC 1912, 1996.

[66] Robert Beverly, Arthur Berger, Young Hyun, and k claffy. Understanding the efficacy of deployed Internet source address validation filtering. In *Proceedings of IMC '09*, 2009.

[67] Anthony J. Bonkoski, Russ Bielawski, and J. Alex Halderman. Illuminating the security issues surrounding lights-out server management. *Proceedings of the 7th USENIX Workshop on Offensive Technologies*, August 2013.

[68] S. Bradner. The Internet Standards Process – Revision 3. RFC 2026 / BCP 9, 1996.

[69] Xue Cai and John S. Heidemann. Understanding block-level address usage in the visible internet. In *SIGCOMM*, 2010.

[70] Zesheng Chen, Chuanyi Ji, and Paul Barford. Spatial-temporal characteristics of internet malicious sources. In *INFOCOM*. IEEE, 2008.

[71] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic, 1988.

[72] R. Cohen and D. Raz. The internet dark matter - on the missing links in the as connectivity map. In *Proceedings of INFOCOM '06*, 2006.

[73] M. Patrick Collins, Timothy J. Shimeall, Sidney Faber, Jeff Janies, Rhiannon Weaver, Markus De Shon, and Joseph Kadane. Using Uncleanliness to Predict Future Botnet Addresses. In *Proceedings of IMC '07*, 2007.

[74] Reliability Communication Security and Interoperability Council. Internet Service Provider (ISP) Network Protection Practices. http://transition.fcc.gov/pshs/docs/csric/CSRIC_WG8_FINAL_REPORT_ISP_NETWORK_PROTECTION_20101213.pdf, 2010.

[75] Evan Cooke, Farnam Jahanian, and Danny McPherson. The zombie roundup: Understanding, detecting, and disrupting botnets. In *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop*, SRUTI'05, 2005.

[76] J. Damas and F. Neves. Preventing use of recursive nameservers in reflector attacks. RFC 5358 / BCP 140, 2008.

[77] Jean C. de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1784.

[78] R. De Neufville. *Applied Systems Analysis: Engineering Planning and Technology Management*. McGraw-Hill, 1990.

[79] Amogh Dhamdhere and Constantine Dovrolis. Ten years in the evolution of the internet ecosystem. In *Proceedings of IMC'08*, pages 183–196, New York, NY, USA, 2008.

[80] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. ZMap: Fast Internet-wide scanning and its security applications. In *Proceedings of the 22nd USENIX Security Symposium*, 2013.

[81] Ed Felten. Security Lessons from the Big DDoS Attacks. https://freedom-to-tinker.com/blog/felten/security-lessons-from-the-big-ddos-attacks/.

[82] P. Ferguson and D. Senie. Network ingress filtering: Defeating denial of service attacks wich employ IP source address spoofing. RFC 2827 / BCP 38, 2000.

[83] Phillipa Gill, Michael Schapira, and Sharon Goldberg. Let the market drive deployment: A strategy for transitioning to bgp security. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11, 2011.

[84] Phillipa Gill, Michael Schapira, and Sharon Goldberg. Let the market drive deployment: a strategy for transitioning to bgp security. In *Proceedings of SIGCOMM '11*, 2011.

[85] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. Botminer: Clustering analysis of network traffic for protocol- and structure-independent botnet detection. In *Proceedings of the 17th Conference on Security Symposium*, SS'08, 2008.

[86] Hewlett-Packard. *HP Integrated Lights-Out security*, 7 edition, December 2010. `http://bizsupport2.austin.hp.com/bc/docs/support/SupportManual/c00212796/c00212796.pdf`.

[87] A. Hubert and R. Van Mook. Measures for making DNS more resilient against forged answers. RFC 5452, 2009.

[88] L. Jiang, V. Anantharam, and J. Walrand. How bad are selfish investments in network security? *Networking, IEEE/ACM Transactions on*, 19(2):549–560, April 2011.

[89] Libin Jiang, Venkat Anantharam, and Jean Walrand. How bad are selfish investments in network security? *IEEE/ACM Trans. Netw.*, 19(2), 2011.

[90] Jaeyeon Jung and Emil Sit. An empirical study of spam traffic and the use of DNS black lists. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 370–375, New York, NY, USA, 2004. ACM.

[91] Chris Kanich, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM Conference on Computer and Communications Security (CCS '08)*, 2008.

[92] Chris Kanich, Nicholas Weaver, Damon McCoy, Tristan Halvorson, Christian Kreibich, Kirill Levchenko, Vern Paxson, Geoffrey M. Voelker, and Stefan Savage. Show me the money: Characterizing spam-advertised revenue. In *USENIX Security Symposium*, 2011.

[93] J. Klensin. Simple mail transfer protocol. RFC 5321, 2008.

[94] Marios Kokkodis, Michalis Faloutsos, and Athina Markopoulou. Network-level characteristics of spamming: An empirical analysis. In *ICNP*, 2011.

[95] Maria Konte, Roberto Perdisci, and Nick Feamster. Aswatch: An as reputation system to expose bulletproof hosting ases. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, SIGCOMM '15, 2015.

[96] Brian Krebs. The target breach, by the numbers. `http://krebsonsecurity.com/2014/05/the-target-breach-by-the-numbers/`, May 2014.

[97] Balachander Krishnamurthy and Jia Wang. On network-aware clustering of web clients. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '00, 2000.

[98] Craig Labovitz, G. Robert Malan, and Farnam Jahanian. Internet routing instability. *IEEE/ACM Trans. Netw.*, 6(5):515–528, 1998.

[99] G. Lindberg. Anti-Spam recommendations for SMTP MTAs. BCP 30/RFC 2505, 1999.

[100] Yang Liu, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, and Mingyan Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 1009–1024, Washington, D.C., August 2015. USENIX Association.

[101] Ratul Mahajan, David Wetherall, and Tom Anderson. Understanding BGP misconfiguration. In *Proceedings of SIGCOMM '02*, 2002.

[102] Deirdre Mulligan and Fred Schneider. Doctrine for cybersecurity. *The Journal of the American Academy of Arts & Sciences*, 140(4), 2011.

[103] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On Spectral Clustering: Analysis and an algorithm. In *NIPS*, pages 849–856, 2001.

[104] University of Oregon. Route views project. `http://www.routeviews.org/`.

[105] Niels Provos, Dean McNamee, Panayiotis Mavrommatis, Ke Wang, and Nagendra Modadugu. The ghost in the browser analysis of web-based malware. In *Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets*, HotBots'07, pages 4–4, Berkeley, CA, USA, 2007. USENIX Association.

[106] Zhiyun Qian, Zhuoqing Morley Mao, Yinglian Xie, and Fang Yu. On network-level clusters for spam detection. In *Proceedings of NDSS'10*, 2010.

[107] Anirudh Ramachandran and Nick Feamster. Understanding the network-level behavior of spammers. In *Proceedings of SIGCOMM '06*, pages 291–302, 2006.

[108] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. Address allocation for private internets. BCP 5/RFC 1918, 1996.

[109] Jennifer Rexford, Jia Wang, Zhen Xiao, and Yin Zhang. Bgp routing stability of popular destinations. In *Proceedings of IMW '02*, 2002.

[110] Steve Sheng, Lorrie F. Cranor, Jason Hong, Brad Wardman, Gary Warner, and Chengshan Zhang. An Empirical Analysis of Phishing Blacklists. In *Sixth Conference on Email and Anti-Spam*, 2009.

[111] Craig A. Shue, Andrew J. Kalafut, and Minaxi Gupta. Abnormally malicious autonomous systems and their internet connectivity. *IEEE/ACM Trans. Netw.*, pages 220–230, 2012.

[112] Robin Sidel. Home depot's 56 million card breach bigger than target's. `http://www.wsj.com/articles/home-depot-breach-bigger-than-targets-1411073571`, September 2014.

[113] Herbert A. Simon. Spurious Correlation: A Causal Interpretation. *Journal of the American Statistical Association*, (267):467–479, 1954.

[114] Sushant Sinha, Michael Bailey, and Farnam Jahanian. Shades of Grey: On the Effectiveness of Reputation-based ”blacklists”. In *Proceedings of MALWARE '08*, pages 57–64, October 2008.

[115] SOPHOS. Security Threat Report 2012. `http://www.sophos.com/medialibrary/PDFs/other/SophosSecurityThreatReport2012.pdf`.

[116] K. Soska and N. Christin. Automatically Detecting Vulnerable Websites Before They Turn Malicious. In *Proceedings of the 23rd USENIX Security Symposium*, San Diego, CA, August 2014.

[117] Brett Stone-Gross, Christopher Kruegel, Kevin Almeroth, Andreas Moser, and Engin Kirda. FIRE: FInding Rogue nEtworks. In *Proceedings of the 2009 Annual Computer Security Applications Conference (ACSAC '09)*, 2009.

[118] Johannes Ullrich. IPMI: Hacking servers that are turned “off”. ISC Diary blog, June 2012. `https://isc.sans.edu/diary/IPMI%3Aminimal+Hacking+servers+that+are+turned+%22off%22/13399`.

[119] Randal Vaughn and Gadi Evron. DNS Amplification Attacks. `http://www.isotf.org/news/DNS-Amplification-Attacks.pdf`, 2006.

[120] Shobha Venkataraman, Subhabrata Sen, Oliver Spatscheck, Patrick Haffner, and Dawn Song. Exploiting network structure for proactive spam mitigation. In *Proceedings of Usenix Security 2007*, 2007.

[121] J. Weil, V. Kuarsingh, C. Donley, C. Liljenstolpe, and M. Azinger. IANA-Reserved IPv4 Prefix for Shared Address Space. BCP 153/RFC 6598, 2012.

[122] Yinglian Xie, Fang Yu, Kannan Achan, Eliot Gillum, Moises Goldszmidt, and Ted Wobber. How dynamic are ip addresses? In *Proceedings of SIGCOMM '07*, pages 301–312, 2007.

[123] Jian Zhang, Phillip Porras, and Johannes Ullrich. Highly Predictive Blacklisting. In *Usenix Security 2008*, 2008.

[124] Jing Zhang, Robin Berthier, Will Rhee, Michael Bailey, Partha Pal, Farnam Jahanian, and William Sanders. Learning from Early Attempts to Measure Information Security Performance. In *Proceeding of the 5th Workshop on Cyber Security Experimentation and Test (CSET '12)*, Bellevue, WA, USA, August 2012.

[125] Jing Zhang, Robin Berthier, Will Rhee, Michael Bdailey, Partha Pal, Farnam Jahanian, and William Sanders. Safeguarding Academic Accounts and Resources with the University Credential Abuse Auditing System. In *Proceedings of the 42th Annual IEEE International Conference on Dependable Systems and Networks (DSN '12)*, Boston, MA, USA, June 2012.

[126] Jing Zhang, Ari Chivukula, Michael Bailey, Manish Karir, and Mingyan Liu. Characterization of blacklists and tainted network traffi. In *Proceedings of PAM'13*, 2013.