UNDERSTANDING AND DEFENDING AGAINST INTERNET INFRASTRUCTURES SUPPORTING CYBERCRIME OPERATIONS

A Thesis
Presented to
The Academic Faculty

by

Maria Konte

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computer Science

Georgia Institute of Technology
December 2015

UNDERSTANDING AND DEFENDING AGAINST INTERNET INFRASTRUCTURES SUPPORTING CYBERCRIME OPERATIONS

Approved by:

Professor Nick Feamster, Advisor
Department of Computer Science
Princeton University

Professor Roberto Perdisci
Department of Computer Science
University of Georgia

Professor Wenke Lee
School of Computer Science
Georgia Institute of Technology

Professor Ellen Zegura
School of Computer Science
Georgia Institute of Technology

Professor Manos Antonakakis
Department of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: 13 November 2015

To those who help others.

ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. Nick Feamster, for all his support, enthusiasm, and optimism, which have been an invaluable source of inspiration for me. Nick has been a great mentor, who has had an incredible impact on my path. Also, I would like to thank Prof. Roberto Perdisci, for all his support, boundless positive energy, and persistence. I feel extremely fortunate, that I have had the chance to work with him.

I would like to thank the rest of the members of the thesis committee; Prof. Wenke Lee, Prof. Ellen Zegura and Prof. Manos Antonakakis, for their constructive feedback and comments.

Also, I would like to thank all my lab mates at the Networks Research Lab, and especially Samantha Lo, Bilal Anwer, Yogesh Mundada and Karim Habak for their invaluable support and positive energy.

On a personal level, I would like to thank my family; Efi and Vangelis Kontes for their encouragement. Especially, I would like to thank my brother, Nikos Kontes, who I have always been admiring and looking up to. Also, I would like to thank my partner Konstantinos Dovrolis for his boundless support on so many levels, but most importantly for his encouragement through this journey, and true dreaming. I would like to thank my daughter Anafi, for always being there, since we started the PhD program together - literally. But most importantly, I would like to thank her for just being who she is; I never thought I would meet a person like her. Finally, I would like to thank my friend Ronda Shiff, for her true and brilliant presence, always next to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Today's cybercriminals must carefully manage their network resources to evade detection and maintain profitable businesses. For example, a rogue online enterprise has to have multiple technical and business components in place, to provide the necessary infrastructure to keep the business available. Often, cybercriminals in their effort to protect and maintain their valuable network resources (infrastructures), they manipulate two fundamental Internet protocols; the Domain Name System (DNS) and the Border Gateway Protocol (BGP).

A popular countermeasure against cybercriminal infrastructures are Autonomous Systems (AS) reputation systems. Past research efforts have developed several AS reputation systems that monitor the traffic for illicit activities. Unfortunately, these systems have severe limitations; (1) they cannot distinguish between malicious and legitimate but abused ASes, and thus it is not clear how to use them in practice, (2) require direct observation of malicious activity, from many different vantage points and for an extended period of time, thus delaying detection.

This dissertation presents empirical studies and a system that help to counteract cybecriminal infrastructures. First, we perform empirical studies that help to advance our understanding, about how these infrastructures operate. We study two representative types of infrastructures: (1) fast-flux service networks which are infrastructures based on DNS manipulation, (b) malicious ASes (hubs of cybercriminal activities) which are infrastructures that are primarily based on BGP manipulation. Second, we build on our observations from these studies, and we design and implement, *ASwatch*; an AS reputation system that, unlike existing approaches, monitors exclusively the routing level behavior of ASes, to expose malicious ASes sooner. We build *ASwatch* based on the intuition that, in an attempt to evade possible detection and remediation efforts, malicious ASes exhibit agile routing behavior (e.g. short-lived routes, aggressive re-wiring). We evaluate *ASwatch* on known malicious ASes, and we compare its performance to a state of the art AS reputation system.

CHAPTER I

INTRODUCTION

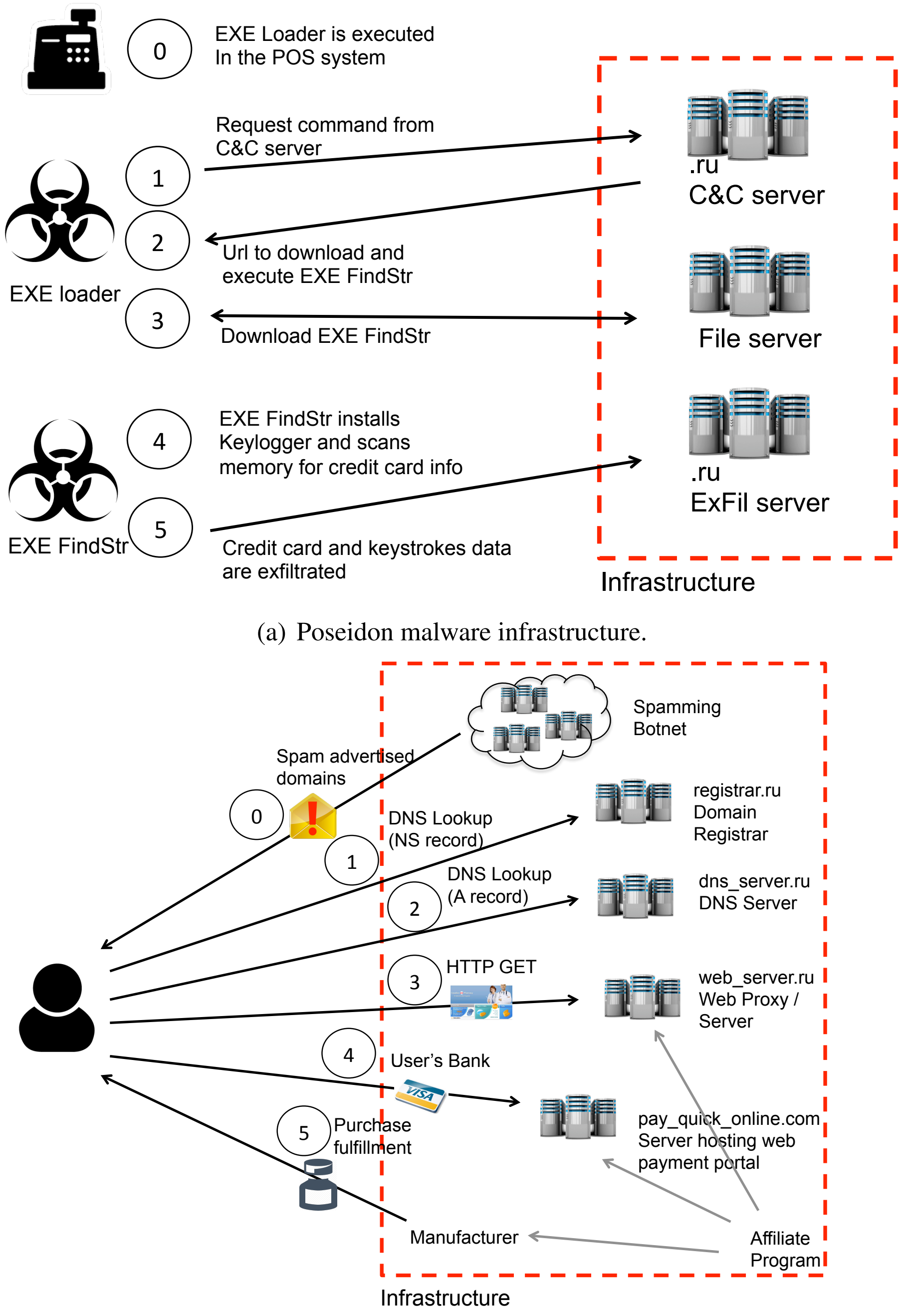
1.1 Introduction

Today's cybercriminals must carefully manage their network resources to evade detection and maintain profitable businesses. For example, a rogue online enterprise has to have multiple technical and business components in place, to provide the necessary infrastructure to keep the business available and profitable. Some elements of the technical components, may include botnet services, domain registration services, name servers, and hosting or proxy services. Cybercriminals take measures to protect these resources; for example, botmasters protect their botnet command-and-control (C&C) servers from take-downs, spammers secure spam delivery by rotating IP addresses to evade trivial blacklisting, and illicit business operators protect scam hosting services by setting up proxies, etc. We refer to the distribution, maintenance and protection of these resources as *operations*. We refer to the network resources that support these operations as *infrastructure*.

For example, Figure 1 illustrates the infrastructure that supports the operations of: (1) an illicit pharmaceutical business (Figure 1(b)) and (2) a malware which attacks point of sales (PoS) systems (Figure 1(a)). In the case of the illicit pharmaceutical business, the infrastructure involves a spamming botnet, servers that provide domain registration, name service, web hosting and payment. In the case of the PoS malware, named Poseidon, the malware targets PoS systems, and infects machines to access the memory for credit card information, which exfiltrates at a later time. The infrastructure includes a command and control (C&C) server, file servers and exflitration servers.

Often, cyber-criminals in their effort to maintain and protect their infrastructure, they manipulate two fundamental internet protocols that are critical for connectivity, namely the Domain Name System (DNS) and the Border Gateway Protocol (BGP).

Examples of how cybercriminals manipulate DNS to maintain and protect their infrastructures, include DNS based botnets, and scam hosting using fast flux service networks (FFSN). Figure 2(a) shows an example infrastructure of a FFSN. Somewhat similar to a technique used by content distribution networks (CDNs) such as Akamai, a fast-flux domain is served by many distributed machines and short time-to-live (TTL) values are used to quickly change a mapping between a domain and an IP address. However, the hosts involved for serving a fast-flux domain are botnet zombie drones and instead of hosting



(a) Poseidon malware infrastructure.

(b) Illicit online business infrastructure.

Figure 1: Example cyber-criminal infrastructures; 1(a) illustrates the poseidon malware infrastructure as described by Cisco security blog post [18], 1(b) illustrates the infrastructure involved to support an illicit pharmaceutical business, as described by [62].

actual content, these zombies often act as front-end proxies that relay messages between a client and a mothership node. Consequently, using this fast flux technique, cybercriminals can easily throw in and out a large number of compromised hosts as needed while effectively hiding their mothership node.

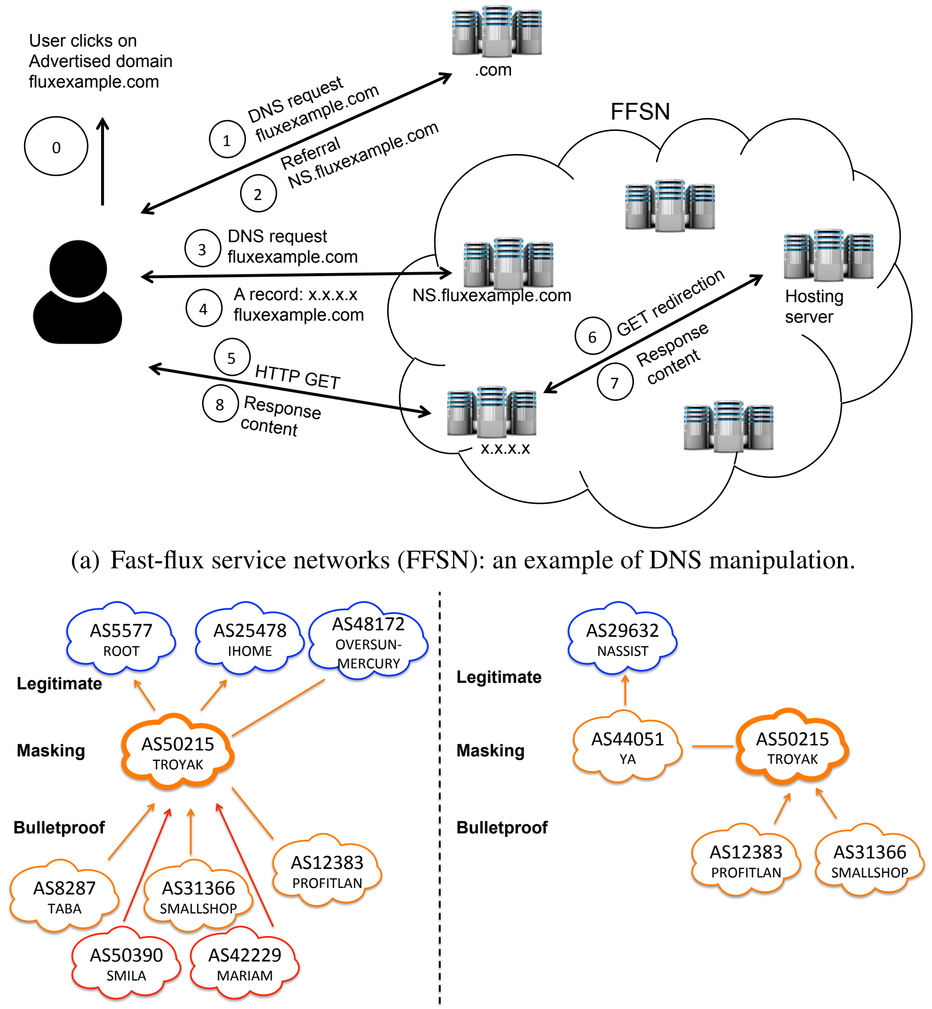
Examples of how cybercriminals manipulate BGP to maintain and protect their infrastructures, include short-lived prefix announcements to perform an illicit operation (*e.g.*large volumes of spam, prefix hijacking), hosting their services in a dedicated AS (called bulletproof AS) which is often connected to a masking upstream provider. Figure 2(b) shows an example of the latter case. We show the upstream and downstream connectivity of a crime-friendly AS, Troyak, around the time it was reported in a blog. Before the report, it was connected with ASes Root, Ihome, and Oversun-Mercury. After the blog report, AS Troyak lost *all* of its upstream providers and relied on a peering relationship with AS Ya for connectivity. After the report, AS Troyak and its customers went offline. We refer to an AS as *malicious*, if it is managed and operated by cybercriminals, and if its main purpose is to support illicit network activities (*e.g.*, phishing, malware distribution, botnets). In contrast, we refer to an AS as *legitimate*, if its main purpose is to provide legitimate Internet services. In some cases, a legitimate AS's IP address space may be abused by cybercriminals to host malicious activities (*e.g.*, sending spam, hosting a botnet command-and-control server). Such abuse is distinct from those cases where cybercriminals operate and manage the AS.

1.1.1 Challenges of defenses against cyber-criminal infrastructures

We describe the challenges that researchers face, to design defenses against cybercriminal infrastructures. These challenges motivate the measurement studies that we perform, and the detection system that we develop in this dissertation.

Distinguishing between malicious and abused ASes.

A countermeasure against cybercriminal infrastructures are AS reputation systems. The community has developed several AS reputation systems that monitor *data-plane* traffic for illicit activities. Existing AS reputation systems typically monitor network traffic from different vantage points to detect the presence of either malware-infected machines that contact their C&C servers, send spam, host phishing or scam websites, or perform other illicit activities. These systems establish AS reputation by measuring the "density" of malicious network activities hosted within an AS. For instance, FIRE [101] tracks the number of botnet C&C and drive-by malware download servers within an AS. ASes that host a large concentration of malware-related servers are then assigned a low reputation. Similarly, Hostexploit [35] and BGP Ranking [12] compute the reputation of an AS based on data collected from sources such as DShield [26] and a variety of IP and domain name



(a) Fast-flux service networks (FFSN): an example of DNS manipulation.

(b) Malicious ASes rewire and hide behind masking upstream ASes: an example of BGP manipulation

Figure 2: Cyber-criminals manipulate critical Internet protocols to maintain and protect their infrastructures. Examples of DNS 2(a) and BGP 2(b) manipulation.

blacklists.

Unfortunately, these existing AS reputation systems have a number of limitations: (1) They cannot distinguish between *malicious* and *legitimate but abused* ASes. Legitimate ASes often unwillingly host malicious network activities (*e.g.*, C&C servers, phishing sites) simply because the machines that they host are abused. For example, AS 26496 (GoDaddy) and AS 15169 (Google) repeatedly appeared for years among the ASes with lowest reputation, as reported by Hostexploit. Although these ASes are legitimate and typically respond to abuse complaints with corrective actions, they may simply be unable to keep pace with the level of abuse within their network. On the other hand, *malicious* ASes are typically unresponsive to security complaints and subject to law-enforcement takedown. (2) Because of the inability to distinguish between *malicious* and *legitimate but abused* ASes, it is not clear how to use the existing AS rankings to defend against *malicious* ASes. (3) Existing AS reputation systems require direct observation of malicious activity from many different vantage points and for an extended period of time, thus delaying detection.

Limitations of defenses against DNS-based infrastructures.

In the case of infrastructures that are based on DNS manipulation, most defense approaches have naturally focused on exposing the associated domain names. Unfortunately, DNS-based infrastructures are highly agile, and they often exhibit high ”churn” rates. For example a botnet may iterate over thousands of domains names and IP addresses over the course of a few days [88]. Domain names are a cheaper resource in comparison to IP addresses [71, 72], and thus we expect IP addresses to be ”consumed” at a lower rate than the associated domain names. Thus, focusing primarily on the domain names, even though it is effective for early exposure of malicious domains names, it maybe not enough to keep up with the hosting IPs of DNS-based infrastructures.

1.1.2 Understanding malicious infrastructures and designing control-plane based defenses.

Thesis Statement: This dissertation counteracts cybercriminal infrastructures by targeting malicious ASes, which are hubs of cybercriminal operations. We demonstrate that it is possible to achieve more accurate and earlier detection of malicious ASes, by monitoring the control-plane AS behavior. We perform empirical studies that help to advance our understanding of how cyber-criminals manipulate two critical Internet protocols - DNS and BGP - to maintain and protect their infrastructures. We then build on this understanding, to design and implement an AS reputation system, to expose malicious rather than abused ASes.

This dissertation makes the following contributions in defense of the thesis statement:

- We present an empirical study of the dynamics of fast-flux service networks - a representative DNS based infrastructure - as they are used to host point-of-sale sites for email scam campaigns. We actively monitor the DNS records for URLs for scam campaigns received at a large spam sinkhole over a one-month period to study the rates of change in fast-flux networks, the locations in the DNS hierarchy that change, and the extent to which the fast-flux network infrastructure is shared across different campaigns.
- We present the first systematic study of the re-wiring dynamics of malicious ASes.

We track the ASes that were listed by Hostexploit over a period of two years and compared their AS-level re-wiring dynamics with non-reported ASes. Using a publicly available dataset of Customer-Provider (CP) relations in the Internets AS graph, we study how interconnection between autonomous systems evolves, both for ASes that provide connectivity for attackers and ASes that were not reported as malicious.

- We present a fundamentally different approach to establishing AS reputation. We design and implement a system, *ASwatch*, that aims to identify malicious ASes using exclusively *control-plane* data (*i.e.*, the BGP routing control messages exchanged between ASes using BGP). Unlike existing *data-plane* based reputation systems, *ASwatch* explicitly aims to identify *malicious* ASes, rather than assigning low reputation to legitimate ASes that have unfortunately been abused.

We now elaborate on the contributions:

Understanding the dynamics of FFSN, a DNS based infrastructure. We study the roles of fast-flux nodes in hosting different parts of the infrastructure (e.g., authoritative name server, Web server, or spammer) and how these roles evolve over time. We study: (a) the rates at which fast-flux networks redirect clients to different authoritative name servers, or to different Web sites entirely. (b) The extent to which individual fast-flux networks "recruit" new IP addresses and how the rate of growth varies across different scam campaigns. (c) the location of change; the extent to which fast-flux networks change the Web servers to which clients are redirected. (d) the use and sharing of infrastructure. We study the geographical and topological locations of fast-flux hosts (both authoritative nameservers and Web servers), as well as how fast-flux infrastructure is shared over time, across scam campaigns, and between spamming and hosting infrastructure.

The fist systematic study to understand the control-plane behavior of malicious ASes. We track reported ASes and non-reported ASes, with the goal of improving our understanding of how malicious networks exploit interconnection through different upstream ASes to cover their traces. Rather than attempting to detect any individual type of attack (e.g., spam, denial of service), we characterize the re-wiring activity of malicious networks that are primarily responsible for attacker activities. This preliminary study motivated us to design *ASwatch*.

Designing control-plane features to capture the behavior of malicious ASes. The main intuition behind *ASwatch* is that malicious ASes may manipulate the Internet routing system, in ways that legitimate ASes do not, in an attempt to evade current detection and remediation efforts. For example, malicious ASes "rewire" with one another, forming groups of ASes, often for a relatively short period of time [49]. Only one AS from the group connects to a legitimate upstream provider, to ensure connectivity and protection for the group. To capture this intuition, we derive a collection of *control-plane features* that is evident solely from BGP traffic observed via Routeviews [86]. We identify three families of features that aim to capture different aspects of the "agile" control plane behavior typical of malicious ASes. (1) *AS rewiring* captures aggressive changes in AS connectivity; (2) *BGP routing dynamics* capture routing behavior that may reflect criminal illicit operations; and (3) *Fragmentation and churn of the advertised IP address space* capture the partition and rotation of the advertised IP address space.

An AS reputation system to expose malicious ASes, rather than abused ASes. We present *ASwatch*, an AS reputation system that aims to identify malicious ASes by monitoring their *control plane behavior*. We evaluate *ASwatch* on real cases of malicious ASes. We collect *ground truth* information about numerous malicious and legitimate ASes, and we show that *ASwatch* can achieve high true positive rates with reasonably low false positives. We evaluate our statistical features and find that the rewiring features are the most important. We compare the performance of *ASwatch* with BGP Ranking, a state-of-the-art AS reputation system that relies on data-plane information. Our analysis over nearly three years shows that *ASwatch* detects about 72% of the malicious ASes that were observable over this time period, whereas BGP Ranking detects only about 34%.

Practical help for network operators to defend against malicious traffic: Our work is motivated by the practical help that an AS reputation system, which accurately identifies malicious ASes, may offer: (1) Network administrators may handle traffic appropriately from ASes that are likely operated by cyber criminals. (2) Upstream providers may use reliable AS reputation in the peering decision process (e.g. charge higher a low reputation customer, or even de-peer early). (3) Law enforcement practitioners may prioritize their investigations and start early monitoring on ASes, which will likely need remediation steps.

1.1.3 Outline

This thesis is organized as follows: Chapter 2 discusses related work. Chapters 3 and 4 present measurement studies of cyber-criminal infrastructures that rely on DNS and BGP manipulation, respectively. Chapter 5 presents *ASwatch*, an AS reputation system that can help to defend against malicious ASes. Chapter 6 discusses general lessons we learned, and future work.

1.1.4 Bibliographic Notes

The FFSN measurement study appeared as a paper at PAM 2009 [51], and received the best dataset award. A longer version of this paper is available as a technical report [50]. The measurement study of rewiring behavior of malicious ASes appeared as a poster at SIGCOMM 2011 [48], and as a paper at PAM 2012 [49]. ASwatch was presented at NANOG62 Research Track in 2014 [67], and appeared as a paper at SIGCOMM 2015 [52].

CHAPTER II

RELATED WORK

In this chapter, we review measurement studies of FFSN, defenses against DNS-based infrastructures, studies of the structure of online crime, studies of "unclean" ASes and existing AS reputation systems, as well as applications of machine learning and signal processing to detect BGP anomalies. Also, we describe how our work relates to these studies.

2.1 Measurement studies of FFSN

The operation of fast-flux service networks was first described in detail by the Honeynet Project [102]. By closely monitoring the behavior of fast-flux agents executed in test environments, the report showed two different types of fast-flux service networks—single-flux and double-flux. Their findings provide insights on the changing nature of fast-flux service networks and lead us to design the multi-level measurement method that form the bases of our study of dynamics of scam infrastructure.

Holz *et al.* [33] analyzed fast flux domains using periodic DNS lookups and presented the characteristics focusing on the diversity of A records and the network locations (AS numbers) that these A records reside at. They also showed various analysis results including the percentage of scam campaigns leveraging fast-flux service networks and the rate at which new machines are added to fast flux domains for a few selected ones. In addition to these measurements, we measured the change of fast flux domains at *multiple* levels of the DNS hierarchy (A records, NS records, and IPs of NS records) and found many different structures of fast-flux service networks, some of which are previously unknown. We also present the geographic and topological distribution of flux hosts, the prevalence resource sharing found across different scam campaigns, and the relationship between fast-flux agents and various blacklists.

Previous work observed fast-flux domains via DNS measurement [111]. From the analysis of passively collected DNS responses at a university gateway, Zdrnja *et al.* observed an instance of fast-flux domain that was short lived (only for three days) and resolved to 80 different IP addresses [111]. In comparison, our study *actively* probes a large number of suspicious DNS domains to profile different types of fast-flux networks over a longer period of time.

Our primary data is drawn from emails collected at a spam trap. Spam traps provide a window to glimpse into the underlying network operations of online scammers who use bulk email for solicitation. Because of relatively easy deployment and data processing, many previous studies used email collected at spam traps for measuring the effectiveness of DNS-based blacklists [44], studying the network-level behavior of spammers [82], characterizing the scam hosting infrastructure [6], and studying the dynamism of the IP addresses of spammers [109]. Others have used passive DNS monitoring to study the dynamics of botnets [22, 79], which are now believed to host fast-flux networks.

Content-based scam campaign clustering is a commonly used technique for analyzing spammer behavior. Anderson *et al.* used image fingerprints to group similar Web pages [6]. Holz *et al.* used strings found from HTML documents for grouping [33]. Pathak *et al.* propose that spammers could be clustered into campaigns by looking for relationships across SMTP connections [76]. In comparison, we used image comparison in addition to string matches of the names of embedded files of each URL. Although the similarity metric employed in each work is slightly different, we believe that clustering results would not bear too much difference.

2.2 Defenses against FFSN and DNS-based cyber-crminal infrastructures

FFSN detection systems: Earlier approaches, [33, 73, 75] studied fast-flux domains advertised through spam email messages and attempted to detect them as follows: collected spam emails, extracted the urls advertised through these messages, and extracted features features from the urls, then performed active probing (repeated DNS queries to collect the resolved IP addresses) and classified each individual domain as fast-flux.

In contrast, Fluxbuster [77] was based on passive analysis of recursive DNS traces collected from large ISPs, to classify entire groups of domains into fast-flux or non-fast-flux. More specifically, they deployed a sensor in front of the recursive DNS server of large ISPs, to passively monitor the DNS queries and responses from the clients, and applied prefiltering rules to selectively store information about potential fast-flux domains, into a central DNS data collector. Finally, they clustered relative domains together, they extracted features from these groups (*e.g.*total number of resolved IPs, diversity of BGP prefixes), and classified each group of domains as likely fast-flux or not.

Botnet takedowns: Nadji *et al* [71] proposed a methodology to perform effective botnet takedowns. The authors enumerate the C&C infrastructure (IP addresses that host C&C

domains from the same botnet). They perform the enumeration using: 1) passive DNS data to identifying historic relationships between domain names and hosts, and 2) interrogating malware samples. In earlier work, Nadji *et al* [72] studied cyber-criminal infrastructures using graphs; the authors considered as nodes the IP addresses of hosts that perform malicious activities (spam, scam hosting, C&C domain) and draw an edge between two nodes if they have found historic relationship based on passive DNS data. Then, they applied community detection algorithms to identify parts of the graph which have are densely connected, and hence likely to belong to the same cyber-criminal network. They also studied the importance of different parts of the networks, with respect to volumes of DNS queries performed by the victim clients.

DNS reputation systems: Antonakakis *et al* [8–10] and Bilge *et al* [14] developed DNS reputation systems. These systems build models of known legitimate domains and malicious domains, to compute a reputation score for unknown domains. They primarily rely on the following types of features: network-based features (*e.g.*total number of IPs historically as sociated with a domain), zone-based features (*e.g.*distinct TLDs, frequency of different characters in the domain string), and evidence-based features (*e.g.*distinct malware samples that contacted the domain or the hosting IP).

2.3 Measurement studies and defenses against malicious networks

Studies of "unclean" ASes. Previous studies have attempted to identify "unclean" ASes, which are ASes with a high concentration of low reputation IP addresses. In contrast, we attempt to understand the behavior of ASes that are *controlled and managed by attackers*, rather than ASes which are heavily abused. Collins [20] first introduced the term network uncleanliness as an indicator of the tendency for hosts in a network to become compromised. They gathered IP addresses from datasets of botnets, scan, phishing, and spam attacks to study spatial and temporal properties of network uncleanliness; this work found that compromised hosts tend to cluster within unclean networks. Kalafut *et al.* [45] collected data from popular blacklists, spam data, and DNS domain resolutions. They found that a small fraction of ASes have over 80% of their routable IP address space blacklisted. Konte *et al.* [49] studied ASes that are reported by Hostexploit and how they changed their upstream connectivity. Johnson *et al.* introduced metrics for measuring ISP badness [42]. Moura *et al.* studied Internet bad neighborhoods aggregation. Earlier papers have looked into IP addresses that host scam websites or part of spamming botnets are organized intro infrastructures [17,30,110]. Finally, Ramachandran *et al.* found that most spam originates from a relatively small number of ASes, and also quantified the extent to which spammers use short-lived BGP announcements to send spam [80, 81]. These studies suggest that it is possible to develop an AS reputation system based on analysis of control-plane features, which is the focus of our work.

AS reputation systems . The state of the art in AS reputation systems is to use features that are derived from data-plane information, such as statistics of attack traffic. Current systems correlate data from multiple sources such as spam, malware, malicious URLs, spam bots, botnet C&C servers, phishing servers, exploit servers, cyber-warfare provided by other organizations or companies. Then, then rank ASes based on the concentration of low reputations IP addresses. Organizations, such as Hostexploit [90], Sitevet [90], and BGP Ranking [12] rate each AS with an index based on the activity of the AS weighted by the size of its allocated address space. FIRE [101] examines datasets of IRC-based botnets, HTdetection-based botnets, drive-by-download and phishing hosts and scores ASes based on the longevity of the malicious services they host and the concentration of bad IP addresses that are actively involved. ASMATRA [107] attempts to detecting ASes that provide upstream connectivity for malicious ASes, without being malicious themselves.

Zhang *et al.* [113] find that there is a correlation between networks that are mismanaged and networks that are responsible for malicious activities. The authors use a mismanagement metric to indicate which ASes may be likely to exhibit malicious behaviors (e.g. spam, malware infections), which does not necessarily indicate if an AS is actually operated by cyber-criminals or not. In contrast, we focus on detecting ASes that are operated by attackers, rather than ASes that are mismanagement and likely abused. Also, [113] examined short-lived BGP announcements as an indication of BGP misconfigurations. Even though we also examine the duration of prefix announcements, this is only one of the features we use to capture control plane behavior. Our analysis shows that this feature alone is not enough to distinguish between legitimate and malicious ASes.

Roveta *et al.* [87] developed BURN, a visualization tool, that displays ASes with malicious activity, with the purpose to identify misbehaving networks. In contrast to these reputation systems that rely on data-plane observations of malicious activity from privileged vantage points, *ASwatch* establishes AS reputation using control-plane (*i.e.*, routing) features that can be observed without privileged vantage points and often before an attack.

Machine learning and signal processing approaches. These approaches detect BGP anomalies (*e.g.*, burstiness), with the goal to help system administrators diagnose problematic network behaviors, but they do not provide a connection between BGP anomalies and criminal activity. In contrast to these approaches, *ASwatch* attempts to capture suspicious control-plane behavior (*e.g.*, aggressive change of connectivity, short BGP announcements) with the goal to detect malicious ASes. Prakash *et al.* developed BGPlens, which monitors anomalies by observing statistical anomalies in BGP updates based on analysis of several features, including self-similarity, power-law, and lognormal marginals [78]. Similarly, Mai [66], Zhang [112] and Al-Rousan [3] have examined BGP update messages using tools based on self-similarity and wavelets analysis hidden Markov models to design anomaly detection mechanisms.

2.4 Designing disincentives for cyber-criminal activities

Leontiadis *et al.* [57] studied empirically different types of online criminal networks with the goal to: a) identify common business characteristics across cyber-criminal networks, and b) identify common economic pressure points that may help to make cyber-crime less profitable. More specifically, Leontiadis *et al.* studied the following types of criminal activities: 1) online prescription drug trade [59–61], 2) exploitation of trending news topics and of prescription drugs [69], and 3) WHOIS misuse [58].

The authors empirically identified the following common components across these three types of criminal businesses: 1) search engines; they play an important role providing the means for cyber-criminals to engage victims, 2) payment networks; they provide means of monetization of the activities, 3) registrars and internet service providers; they provide the necessary resources, 4) law enforcement: the authors suggest that international cooperation along with targeted efforts may have the best results for taking down cyber-criminal business that span across countries and hosting providers.

CHAPTER III

FAST FLUX SERVICE NETWORKS: DYNAMICS AND ROLES IN HOSTING ONLINE SCAMS

3.1 Introduction

Online scams require victims to contact a point-of-sale Web site, which must be both highly available and dynamic enough to evade detection and blocking. Until recently, many sites for a scam were hosted by a single IP address for a considerable amount of time (i.e., up to a week) [6]; unfortunately, the relatively static nature of these sites made it possible to block scams with simple countermeasures, such as blocking the IP address. To maintain sites that are both dynamic and highly available, cybercriminals are increasingly using *fast flux*—a DNS-based technique used by botnets to rapidly change these delivery sites.

In this chapter, we find that the scam infrastructure has become considerably more sophisticated and dynamic. Indeed, in this chapter we show that attackers have developed a sophisticated infrastructure for directing victims to scam sites that move around frequently to evade detection and blocking. Attackers that mount scam campaigns appear to be making extensive use of fast-flux service networks [40], which can dynamically (and quickly) redirect clients to different sites for hosting scams. The machines that host content are typically ephemeral (i.e., they may simply be compromised machines) and distinct from the controllers that provide content and control redirections.

This chapter studies the dynamics of fast-flux service networks as they are used to host point-of-sale sites for email scam campaigns. We study the scam sites that were hosted by more than 350 domains as part of 21 scam campaigns in over 115,000 emails collected over the course of a month at a large spam simkhole. We study characteristics of *dynamics* of the infrastructure hosting fast-flux service networks, the *roles* that various machines play in hosting online scams, and the effectiveness of various blacklists at identifying IP addresses and URLs of scam sites.

Previous work has studied the rates at which fast-flux networks change DNS A-record mappings (i.e., name to IP address mappings) and the rate at which new IP addresses are accumulated [33]; this chapter presents many new classes of findings. First, we study fast-flux networks *by campaign* to determine whether dynamics differ across campaigns, and whether distinct spam campaigns share fast-flux service infrastructure. Second, we perform

Table 1: Summary of results.

Finding	Table/Figure	Implications
Dynamics		
Rates of change. DNS records change more quickly than TTL values. NS records are more stable than A records or IPs of NS records. DNS records for fast flux domains change more quickly than those from "legitimate" popular domains.	Fig. 5, Fig. 6	Blacklisting authoritative name server names may help with fast-flux mitigation.
Rates of accumulation. Different scam campaigns (and URLs for those campaigns) recruit new IP addresses at different rates	Fig. 8	The rate at which a URL "accumulates" new IP addresses may help detect fast flux networks and also identify scam campaigns.
Location of change in DNS hierarchy. Different fast-flux domains change at different locations in the DNS hierarchy (i.e., A records, IP of NS record, NS record).	Tab. 3.4.2	
Roles		
Sharing. Different scam campaigns share fast-flux infrastructure	Tab. 3, Tab. 4, Tab. 9	Identifying fast-flux *infrastructure* may help with early detection of scam campaigns.
Distribution across /24s. Fast flux domains return A records that are distributed over a far larger set of /24s than legitimate popular Web sites (as seen when queried from a single DNS location).	Fig. 11	The distribution of query results across IP address space may be useful for detecting fast-flux activity.
Distribution in IP address space. A and NS records are distributed across IP address space, but some regions have a high density of both fast-flux agents and spammers.	Fig. 9	Detection of spammers might also help detect fast-flux networks, and vice versa.
Blacklists. Some IP addresses that appear as flux agents appear in spam and exploit blacklists weeks later.	Fig. 12, Tab. 10, Tab. 11	Identification of FF infrastructures can help towards earlier blacklisting of spam/exploit IPs and vice versa.

continual and iterative DNS monitoring to discover the locations in the DNS hierarchy where fast-flux networks dynamically redirect clients. Finally, we study the roles of fast-flux nodes in hosting different parts of the infrastructure (e.g., authoritative name server, Web server, or spammer) and how these roles evolve over time.

Table 3.1 summarizes the findings of our study and possible implications for these findings. We present findings regarding the following aspects of fast-flux networks:

- *Rate of change.* We examine the rates at which fast-flux networks redirect clients to different authoritative name servers (either by changing the authoritative name-server's name or IP address), or to different Web sites entirely. We find that, while the DNS TTL values do not differ fundamentally from other sites that do DNS-based load balancing, the rates of change (1) differ fundamentally from legitimate load balancing activities; (2) differ across individual scam campaigns.
- *Rate of accumulation ("recruit").* We study the extent to which individual fast-flux networks "recruit" new IP addresses and how the rate of growth varies across different scam campaigns. We find that, while there is a considerable amount of sharing of IP addresses across different scam campaigns, different campaigns accumulate new IP addresses at different rates.
- *Location of change.* We study the extent to which fast-flux networks change the Web servers to which clients are redirected. We infer the location of change by monitoring

any changes of (1) the authoritative nameservers for the domains that clients resolve (the NS record, or the IP address associated with an NS record) or of (2) the mapping of the domain name to the IP address itself (the A record for the name). We find that behavior differs by campaign, but that many scam campaigns redirect clients by changing *all* three types of mappings, whereas most legitimate load-balancing activities only involve changes to A records.

- *Use and sharing of infrastructure.* We study the geographical and topological locations of fast-flux hosts (both authoritative nameservers and Web servers), as well as how fast-flux infrastructure is shared over time, across scam campaigns, and between spamming and hosting infrastructure. We find that different scam campaigns share fast-flux infrastructure; we also find overlap between spamming infrastructure and online scam hosting infrastructure.

Our findings lend insights into the operation of fast-flux networks that may ultimately lead to more effective mitigation techniques: Although scam campaigns are short-lived, the infrastructure that hosts these scams (i.e., the fast-flux network or networks) appears to have relatively invariant features that may prove useful for identifying scams and the spam messages that advertise them.

The rest of this chapter is organized as follows. Section 3.2 describes background on fast-flux networks, current understanding about their roles in hosting online scams, and related work in studying fast-flux networks. Section 3.3 describes our data collection methods, as well as various limitations of our dataset. Section 3.4 describes the dynamics of fast-flux service networks that we observed hosting 21 different spam campaigns over the course of a month. Section 3.5 describes the roles that we observed each IP address playing in the fast flux networks, the locations of spammers and fast flux infrastructure in the IP address space, and the sharing of infrastructure across different roles. Section 3.6 describes the relationships between when various blacklists listed IP addresses and when these IP addresses were seen in the fast-flux hosting infrastructure that we observed in our data. Section 3.7 concludes with a summary and discussion of future work.

3.2 Background

We describe main redirection techniques commonly employed by fast flux service networks and show an example illustrating how this technique can be observed from DNS responses. We then discuss related work.

3.2.1 Fast-Flux Mechanics

Fast flux is a DNS-based method that cybercriminals use in order to organise, sustain and protect their service infrastructures such as illegal web hosting and spamming [102]. Multiple cybercriminal families have been observed to use the fast flux techniques for illegal or fake online businesses, phishing sites, adult content sites [91]. Somewhat similar to a technique used by content distribution networks (CDNs) such as Akamai, a fast-flux domain is served by many distributed machines and short time-to-live (TTL) values are used to quickly change a mapping between a domain and an IP address. However, the hosts involved for serving a fast-flux domain are botnet zombie drones and instead of hosting actual content, these zombies often act as front-end proxies that relay messages between a client and a “mothership” node [102]. Consequently, using this fast flux technique, cybercriminals can easily throw in and out a large number of compromised hosts as needed while effectively hiding their mothership node.

Variations of the technique also exist [102]. In addition to fluctuating address records (A records), a fast-flux domain can have changing name servers records (NS records or IP addresses of NS records). In practice, any combinations of DNS record fluctuations can be used for flexible and resilient operations. Moreover, as we will show in Section 3.4.3, many hosts exploited by fast-flux service networks are found to play the role of both a hosting server (or a front end proxy of it) and an authoritative name server (or a front end proxy of it).

The dynamics of fast-flux service networks make ineffective the existing mitigation scheme that relies on blacklisting offending hosts. Operators of such networks can simply swap out blacklisted hosts. Moreover, by constantly monitoring the “health” of individual hosts, the operators can increase service availability from likely unstable compromised machines. To demonstrate how quickly fast-flux service networks change, we show an example of a fast-flux domain that we monitored on January 20, 2008 at 20:51:52 GMT. The fast-flux domain is called `pathsouth.com` and at that time it was pointing to one of illegal pharmaceutical companies called Canadian Pharmacy [95]. Table 3.2.1 shows the DNS records resulted from two lookups with seven minutes apart. The first column shows the IP addresses of spam sources, from which our spamtrap received copies of spam containing the fast-flux domain. The ten records in bold show that the domain swapped in nine new hosts for serving content and one new name server.

Table 2: DNS lookup results of the `pathsouth.com` fluxing domain: The IP addresses in bold highlight changes between the two lookups six minutes apart.

<table>
<tr><th colspan="7">Domain name: pathsouth.com & responding authoritative nameserver: 218.236.53.11</th></tr>
<tr><td></td><td colspan="6">Time: 20:51:52 (GMT)</td></tr>
<tr><td>Spam sent by IPs</td><td>A records</td><td>TTL</td><td>NS records</td><td>TTL</td><td>IP addresses of NS records</td><td>TTL</td></tr>
<tr><td rowspan="4">88.234.185.68,
88.229.212.225,
212.156.205.188,
189.4.136.197,
89.132.220.6,
125.27.211.201,
85.109.43.187,
83.27.32.75,
80.94.175.76,
84.115.175.16,
88.65.174.73,
195.8.27.96,
124.28.82.112</td><td>77.178.224.156, 79.120.37.38, 79.120.63.225, 79.120.72.0, 79.120.101.244, 79.120.107.25, 85.216.198.225, 87.228.106.92, 89.20.146.249, 89.20.159.226, 89.176.63.78, 89.208.2.199, 89.208.5.106, 213.141.146.83, 220.208.7.115</td><td>300</td><td>ns0.nameedns.com,
ns0.nameedns1.com,
ns0.renewwdns.com,
ns0.renewwdns1.com</td><td>172800</td><td>218.236.53.11,
89.29.35.218,
78.107.123.140,
79.120.86.168</td><td>172800</td></tr>
<tr><td colspan="6">Time: 20:57:49 (GMT)</td></tr>
<tr><td>A records</td><td>TTL</td><td>NS records</td><td>TTL</td><td>IP addresses of NS records</td><td>TTL</td></tr>
<tr><td><b>61.105.185.90, 69.228.33.128,</b> 79.120.37.38, 79.120.108.136, 85.216.198.225, 87.228.106.92, 89.20.146.249, <b>89.20.159.178</b>, <b>89.29.35.218</b>, <b>91.122.121.88</b>, <b>213.220.251.97</b>, <b>218.254.157.62, 218.255.10.103,</b> 220.208.7.115, <b>222.5.114.183</b></td><td>300</td><td>ns0.nameedns.com,
ns0.nameedns1.com,
ns0.renewwdns.com,
ns0.renewwdns1.com</td><td>172800</td><td>218.236.53.11,
89.29.35.218,
78.107.123.140,
<b>213.248.28.235</b></td><td>172800</td></tr>
</table>

3.3 Data

Accordingly, our data collection and processing involves three steps: (1) passive collection of spam data; (2) active DNS monitoring of domains for scams contained in those spam messages; (3) clustering of spam and DNS data by campaign. This section describes our method for collecting spam data and monitoring DNS record changes associated with the associated scam campaigns. We also describe how we monitor the DNS dynamics of popular Web sites to use as a baseline for comparison. We then explain how we postprocess the data to group spam email messages (and the associated DNS data) according to common campaigns. Finally, we discuss potential limitations of our data collection and analysis methods.

3.3.1 Data Collection

To amass a collection of domains to monitor for fast-flux behavior, we collected 3,360 distinct domain names that appeared at spam email messages which were collected at a

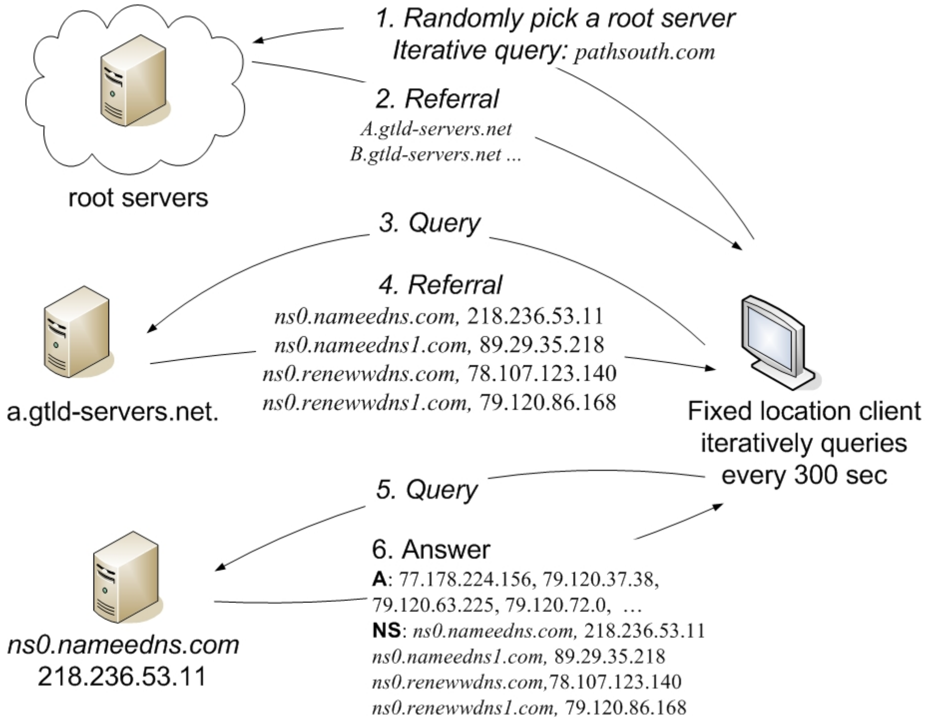

1. Randomly pick a root server
Iterative query: pathsouth.com
2. Referral
A.gtld-servers.net
B.gtld-servers.net ...
root servers
3. Query
4. Referral
ns0.nameedns.com, 218.236.53.11
ns0.nameedns1.com, 89.29.35.218
ns0.renewwdns.com, 78.107.123.140
ns0.renewwdns1.com, 79.120.86.168
a.gtld-servers.net.
Fixed location client
iteratively queries
every 300 sec
5. Query
6. Answer
A: 77.178.224.156, 79.120.37.38,
79.120.63.225, 79.120.72.0, ...
NS: ns0.nameedns.com, 218.236.53.11
ns0.nameedns1.com, 89.29.35.218
ns0.renewwdns.com,78.107.123.140
ns0.renewwdns1.com, 79.120.86.168
ns0.nameedns.com
218.236.53.11

Figure 3: Diagram of the data collection; a fixed location iterative resolver is set up. The resolver starts the queries from a randomly selected root server, every 300 sec for every domain. Here we feature the same fluxing domain `pathsouth.com` as in Table 3.2.1. Our iterative resolver logs all referrals and the DNS records that are returned for every query, at each level of the DNS hierarchy.

large spam sink hole. To obtain this list of URLs, we used a simple URL pattern matcher to extract URLs from the bodies of the messages received at the spam trap. We collected these domains over a period of three months, from October 1, 2007 to December 31, 2007.

Next, we implemented an iterative resolver (at a fixed location) to resolve every domain name from this set once every five minutes. Figure 3 shows the method by which our resolver recorded DNS mappings at each level of DNS resolution, which allows us to monitor fast-flux networks for DNS changes at *three distinct locations in the hierarchy*: (1) the A record; (2) the NS record; and (3) the IP addresses corresponding to the names in the NS record. To avoid possible caching effects, the resolver randomly selected a DNS root server at which to start the DNS query. The iterative resolver recorded the answers received at every level of the DNS hierarchy; we recorded all the referrals and the answers by the queried DNS servers for every domain.

Due to the sheer number of DNS lookups required to monitor the domains arriving at the spam trap, the resolver proceeded through the list of domains sequentially: We began by resolving the first 120 domains received at the spam trap each day. Every day the resolver began resolving the next 120 domains on the list. After each domain had been resolved continously for three weeks, we removed the domain from the list. The resolver operated from January 14, 2008 to February 9, 2008.

To compare the dynamics of the domains received at the spam trap as part of online scam campaigns to the DNS dynamics of "legitimate" domains, we used the same iterative resolution process to study the dynamics of the 500 most popular domains, according to Alexa [4].

3.3.2 Postprocessing: Scam Campaigns

After collecting spam and DNS data, we restricted our analysis to the domains that had reachable Web sites and for which we had observed at least one change in any DNS record. We then clustered the spam messages into *scam campaigns*. To perform this clustering, we retrieve content from the URLs in the email messages and cluster emails whose URLs retrieve common content:

- *Snapshots and Web page sources.* We used AutoIT [11] to sequentially open each URL on a browser, wait until the page is loaded, and take a snapshot of the current page[1]. While doing so, HTTP Analyzer [41] captures all the HTTP requests and responses for further analysis.

- *Clustering by snapshots.* We manually went through snapshot images and cluster URLs if the site is selling the same products under the same brand name using a similar page layout. The clustering is manual and subjective but fairly straightforward.

- *Clustering supplemented with file comparison.* The image comparison fails in the case of partially download pages. For example, `pathsouth.com`, one of the Canadian Pharmacy [95] sites downloads 88 files, of which 85 are jpeg and gif image files. Slow response, which is often observed from fast flux service networks, allows only a few small none image files to be received until the somewhat generous 30 second timeout expires, generating an empty looking page when a snapshot is taken. To make up this shortcoming, for the URLs that are not classified, we check to see whether the downloaded file names of each URL is a subset of those of already identified campaign. We find that most of partially downloaded pages

[1] We used a 30 second timeout value to move on to a next URL if the current site is not reachable. The AutoIT script was executed on a virtual machine running Windows XP to avoid possible drive-by infection. We also disabled most security features that display warnings or prompt for approvals as these interfere with automation.

are caused by the Canadian Pharmacy campaign sites and that all of them request `canadian_pharmacy_2_style.css` in common.

Table 3 shows the summary data of 21 campaigns. We denote each campaign with a *category-ID*.[2] The second column is the number of domain names that we found changing during our one month measurement period (fluxing domain). The following two columns show the number of total spam emails containing the fluxing domains that we received at our spam trap and the total number of sender IPs of those spam emails. The last three columns summarize our measurement data: the first two numbers are the total distinct number of IPs returned as A records of domains ($\text{IP}_{\text{domains}}$) and that of IPs returned as A records of name servers ($\text{IP}_{\text{nameservers}}$). the last number is the total distinct number of IPs from the combined sets ($\text{IP}_{\text{domains}} \cup \text{IP}_{\text{nameservers}}$) . For comparison, Table 4 summarizes the Alexa dataset.

Top campaigns. The top campaign, by the number of hosting servers, is Pharmacy-A. We believe that it is one of Canadian Pharmacy scam campaigns [95]. The campaign swapped in at least of 9,448 distinct IP addresses as hosting servers (or front end proxies of them) for 149 domains over one month. The next two followers are Watch-A [94] and Watch-B [93], both of which offer replica watches. We note that for these top three campaigns, the average ratio of A records associated with a domain name is over 50, allowing the scammers to freely move around among available hosting servers. However, the remaining campaigns are rather modest and we even see the sharing of a few hosting servers by multiple domains (e.g., Pharmacy-D and Links-B appear to have only 5 hosting servers for 50 and 35 domains respectively). Nonetheless, all 21 campaigns exhibited fluxing behavior in their DNS records to some extent during the measurement period.

Registrars. The fast-flux domains in our dataset are mostly .com (348, or 90.6%). The rest 8.4% are .net (32), .ph (3) and .su (1). However, over 99% of these domains are unique (e.g., a.com, b.com)[3], requiring separate registrations with the corresponding top-level domains. Table 5 shows registrar information of the 384 fast-flux domains that we found on May 7, 2008 via `jwhois` queries. 70% of the domains are still marked as active and registered with eight registrars in China, India, and US. Among these, the three registrars in China are responsible for 257 fast-flux domains (66% of total or 95% of the active ones). Surprisingly, all but `paycenter.com.cn` are ICANN-accredited registrars [38].

[2] We looked at each Web page snapshot and assigned a category based on offered products.

[3] Only 4 out of 384 fast-flux domains have the same second-level domain name.

Table 3: Statistics for fast-flux networks hosting scam campaigns. Campaigns are sorted by the total number of IP addresses returned from A records as the number indicates the size of the underlying infrastructure.

Campaign	Spam emails	Spamvertising IPs	Domains in campaigns	Fluxing domains	IPs of A rec	IPs of NS rec	IPs of both A+NS rec
Pharmacy-A	18459	11670	149	149	9448	2340	9705
Watch-A	40681	30411	34	30	1516	225	1572
Watch-B	454	427	43	19	1204	219	1267
Pharmacy-B	371	223	86	52	15	13	22
Casino-A	317	226	6	6	12	12	16
Pharmacy-C	30	4	6	6	12	11	12
Casino-B	15	8	2	1	11	10	17
Links-A	15	8	2	2	10	14	22
Casino-C	4652	4150	9	5	10	14	18
E-Marketing-A	32	4	6	4	8	2	10
Pharmacy-D	37472	28340	52	50	5	5	6
Pharmacy-E	32	25	4	4	5	7	12
Links-B	5663	4573	38	35	5	5	6
Pharmacy-F	2	1	2	2	4	6	10
Pharmacy-G	208	205	2	2	3	8	8
Links-B	4	2	4	2	3	8	11
Service-A	9	1	3	1	2	4	4
Software-A	950	463	5	5	2	4	5
Watch-C	6226	4154	7	5	2	2	2
DomainNames-A	3	3	3	3	2	4	6
Service-B	26	2	2	1	1	3	4
All campaigns	115198	77030	465	384	9521	2421	9821

Table 4: Alexa dataset.

	Domains	IPs of A rec	IPs of NS rec	IPs of A+NS rec
Total	500	1048	852	1877

Figure 4 shows the month when these domains were registered. Because our data collection was done before February 2008, all the domains were registered before then. Interestingly, however, over 40% were registered in January 2008 and immediately put in use for serving scams. Further, these domains are still active even after four months and the 2% of the domains had been active for over 7 months at the time of our measurement. Unfortunately, our WHOIS lookup is after the fact and thus we are unable to tell whether the 30% inactive domains as of May 2008 are due to registration expiration or some other reasons.

3.3.3 Limitations

Our data is derived from spam collected at a single spam trap; different spam traps might receive different distributions of spam emails from different locations. The relatively high volume of emails received at our spam trap (6,247,937 messages from the period of October 2007 through February 2008) suggests that the data we have collected may be representative of spam and scam campaigns seen at other networks, although our trap may certainly induce some geogprahic bias (for example, spam traps located in other countries

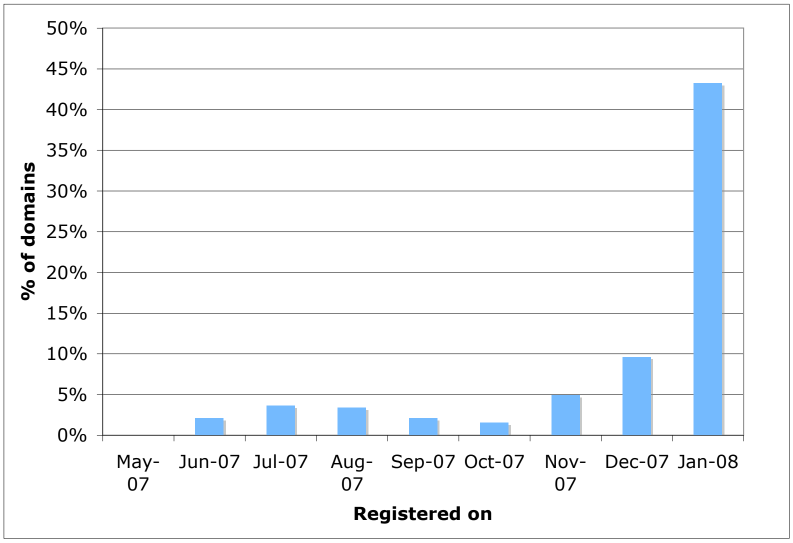


Figure 4: Record creation date for the 384 fast-flux domains: Y-axis is the percentage of the fast-flux domains that were registered on the particular month.

Table 5: Registrars of the 384 fast-flux domains as of May 7, 2008.

Registrar	Country	Domains
dns.com.cn	China	180 (46.9%)
paycenter.com.cn	China	65 (16.9%)
todaynic.com	China	12 (3.1%)
signdomains.com	India	7 (1.8%)
leadnetworks.com	India	3 (0.8%)
coolhandle.com	US	2 (0.5%)
webair.com	US	1 (0.3%)
stargateinc.com	US	1 (0.3%)
total active domains		271 (70.6%)

may receive different scams). We sampled our dataset further by only actively monitoring a subset of the domains contained in URLs received in spam messages at the spam trap; in particular, we did not analyze domain names that we could not explicitly group into a scam campaign. Many of the domains that we could not group into a campaign also exhibited highly dynamic behavior, but we omitted these domains from the analysis because we could not confirm their participation in an online scam.

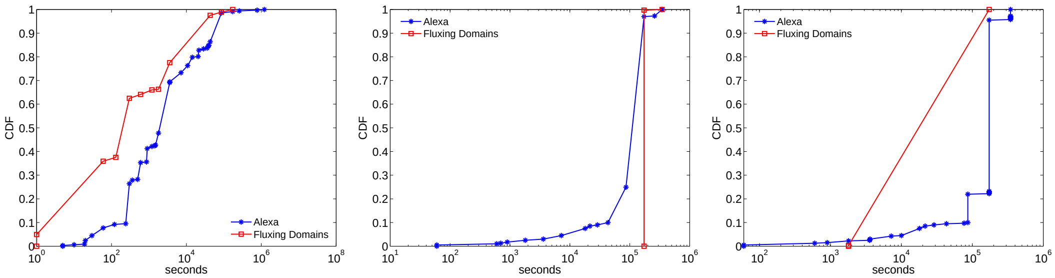
Unlike a legitimate site, a scam campaign operates under many different domain names (e.g., pathsouth.com, yesself.com), possibly to avoid URL blacklisting and perhaps to spread the risk of being detected and terminated by vigilant registrars. Our dataset may cover only a subset of these names for each campaign and we are unaware of any clear way to find out the true number of registered domain names for each scam campaign. However, in most cases, each domain within a campaign appears to show similar behavior.

In some cases, our DNS resolution process occurred months after the spam message corresponding to the scam was received at the spam trap. It is possible that the dynamics of fast-flux networks may differ close to the time of the receipt of the actual spam; however, our measurements suggest that the dynamics of fast-flux networks for each scam campaign (i.e., the rate of change of DNS records, the rate of accumulation of new IP addresses) remain stable over the course of a month, so it may be reasonable to expect similar dynamic behavior closer to the original receipt of the mail. We were surprised that most domains remained not only resolvable, but also reachable, even months after receipt of the original spam email associated with the campaign. This behavior differs from the dynamics of scam hosting sites observed in previous studies [6], which observed that many scam sites remain active for only a week; the difference may be due to the rise of fast-flux networks.

Our clustering technique assigns a URL to a *single* campaign based on snapshots of the Web site's content for a single snapshot. It is possible that, over time, a single domain could be used to host multiple campaigns; in these cases, our analysis would attribute behavior to a single campaign (i.e., the one corresponding to our snapshot) when, in fact, the domain was hosting multiple campaigns over the course of our analysis. We did not collect frequent enough snapshots to detect such behavior.

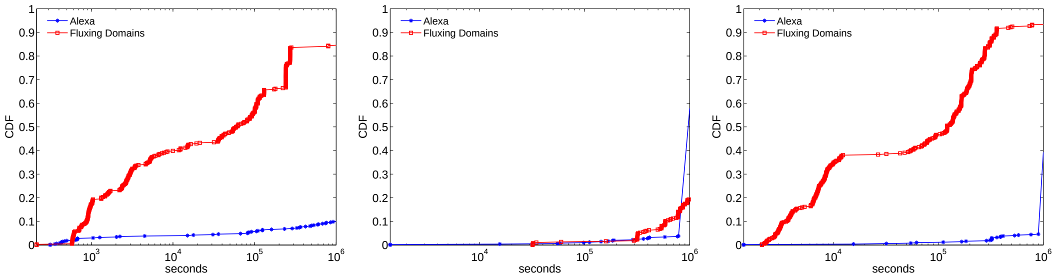
3.4 Dynamics

This section presents our findings concerning the dynamics of fast-flux service networks. We study three aspects of dynamics: (1) the rate at which DNS records change at each level of the hierarchy; (2) the rate at which fast flux networks accumulate new IP addresses (both overall, and by campaign); and (3) the location in the DNS hierarchy where dynamics are taking place. To understand the nature of these features with respect to "legitimate" load balancing behavior, we also analyze the same set of features for 500 popular sites listed by Alexa [4] as a baseline. We find many aspects of dynamics that are distinct to fast flux service networks.



(a) A records (b) NS records (c) IP of NS records

Figure 5: Cumulative distribution of TTLs values of A, NS, and IP of NS records.



(a) A records (b) NS records (c) IP of NS records

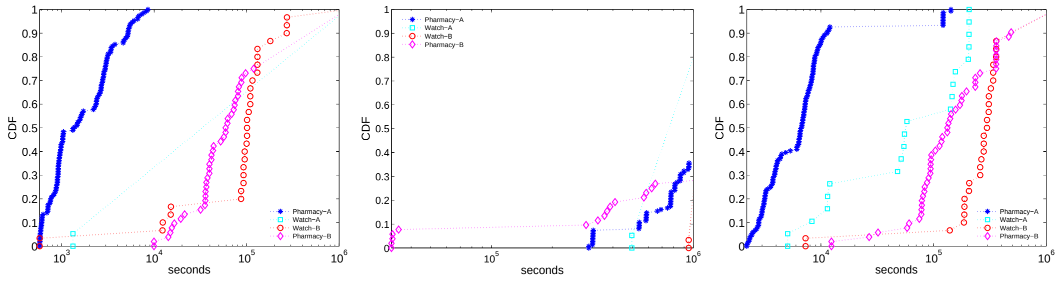
Figure 6: Cumulative distribution of the average time between changes of A, NS, and IP of NS records.

3.4.1 Rate of Change

We studied the rates at which domains for online scams changed DNS record mappings and the corresponding TTL values for these records. We expected that fast-flux domains would both have short TTL values and exhibit frequent changes in name-to-IP address mappings.

Figure 5 compares the distributions of TTLs between the fluxing domains and the domains in the Alexa data set. The disritubiotn of A record TTLs shows that scam sites have slightly shorter TTL values than popular Web sites; however, both classes of Web sites have A records with a wide range TTL values. Even more surprisingly, about 30% of popular Web sites maintain NS records with TTL values of less than a day, but almost all fast-flux domains we analyzed had TTL values for NS records of longer than a day. In hindsight, these results do make sense: many clients visiting scam sites will visit a particular domain infrequently, and only a small number of times, so the TTL value is less important than the rate at which the mapping itself is changing (i.e., for *new* clients that attempt to resolve the domain).

To detect changes that may be related to fast-flux behavior, we record both the A records



(a) A records

(b) NS records

(c) IP of NS records

Figure 7: Cumulative distributions of the average time between changes of A, NS, and IP of NS records for Pharmacy-A, Watch-A, Watch-B, and Pharmacy-B.

that are returned at Step 6 in Figure 3, and NS names and IP addresses of NS names that are returned at Step 4. The reason why we record these two separate pieces of information is because NS names and IP addresses of NS names are not always returned with the A records of the answer at Step 6; the lack of complete information about the sequence of lookups in the DNS hierarchy will make it difficult to observe all aspects of the dynamics.

To account for possible load-balancing mechanisms at a higher level of the DNS hierarhcy, we group the responses according to the authoritative server that provided them. We then perform pairwise comparisons across each group of records. In the case of A records responses and NS record responses, we consider response as a change if at least one new record appears since last answer, or if the number of records returned has otherwise changed since the last response. (We do not consider reordering the records as a change.) In the case ofIP addresses of NS records, we consider the response to be a change if either NS names appear with different IPs or a new NS name shows up since last reply.

Fast-flux domains change on shorter time intervals than their TTL values. Figure 6 shows a distribution of the average time between changes for each domain across all 21 scam campaigns; each point on the line represents the average time between changes for a particular domain that we monitored. The distribution shows that fast-flux domains change hosting servers (A records) and name servers (IP addresses of NS records) more frequently than popular Web servers do. In particular, the rate of change of IP of NS records is much more frequent than TTL values of these records, causing possible service disruption for returning clients. In some cases, for example the IP addresses of NS records, the changes are significantly more frequent than the TTL values would suggest. The incongruence of DNS TTL values with the rates at which these records are actually changing could also prove to be a useful feature for detecting fast-flux behavior.

Domains in the same campaign tend to show similar rates of change. We also analyzed the rate of change of DNS records after clustering the fast-flux domains according to campaign. Figure 7 shows these results for the top 4 campaigns (ranked by the number of distinct IPs returned in A records for domains hosting the campaigns). The results are striking: different scam campaigns rotate DNS record mappings at distinct rates, and the rates at which DNS records for a particular campaign are remapped are similar across all domains for a particular scam. This finding suggests that it may be possible to use rates of flux at different levels in the DNS hierarchy as a type of signature for a scam campaign.

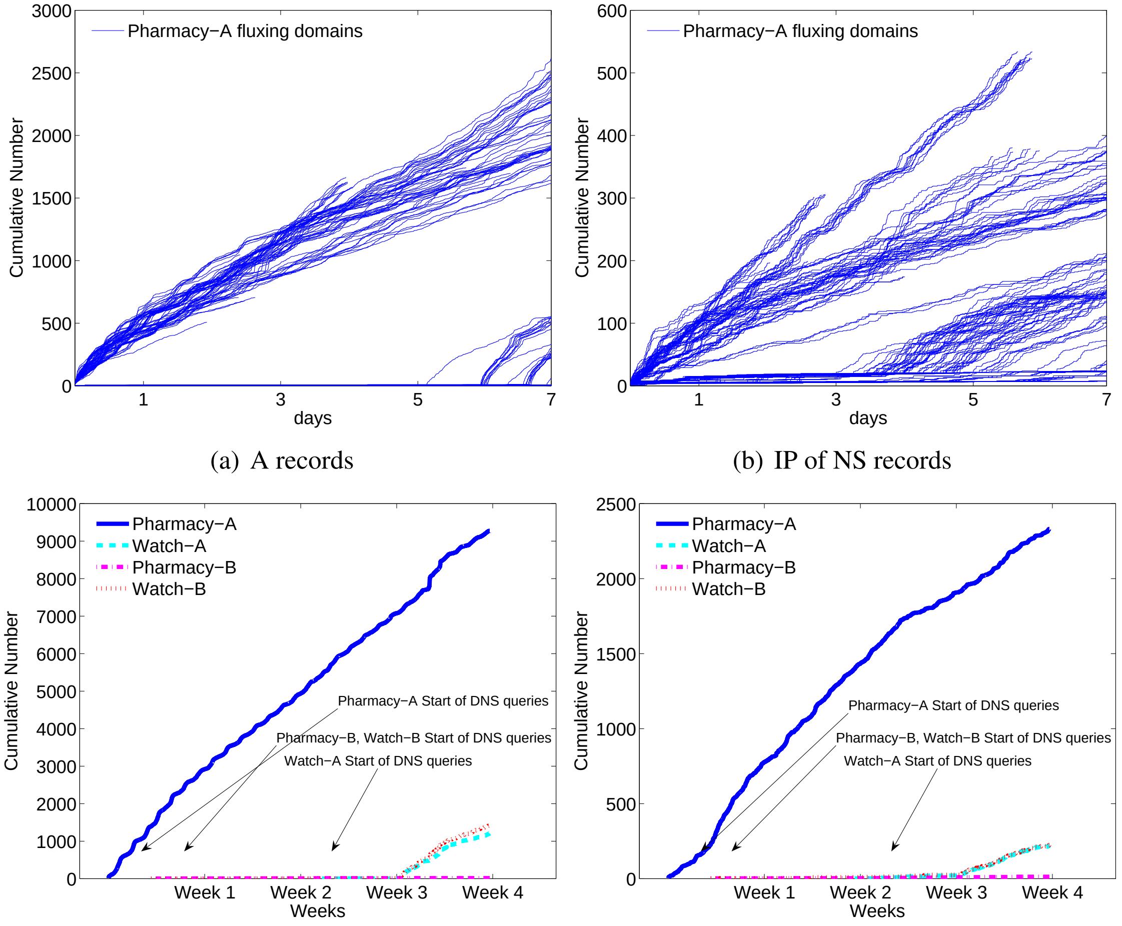
3.4.2 Rate of Accumulation

Ideally, we wish to know the size of a fast flux service network at a given moment and to measure the rate at which the network grows over time. However, in practice, our measurement is limited by the rate at which a flux domain updates its DNS records and what we present in this section is the rate at which a previously unseen host becomes an active hosting server (A records of a domain) or a name server (IP addresses of names returned by NS records).

Using a method similar to the one used by Holz *et al.* [33], we determine the rate of “flux” by repeatedly resolving each domain and assigning an increasing sequential ID to each previously unseen IP address. Figures 8(a) and 8(b) show the total number of distinct IPs for each fast-flux domain (the y-value of the end of each line) over the first week of our data collection period (first 2,000 iterations, 300 seconds apart from each other) and how fast each domain accumulated new hosts (slope). A steeper slope incidates more rapid accumulation of new IP addresses for that domain. Figure 8(a) shows this statistic for A records and Figure 8(b) shows the same statistic for IP addresses of NS records of the domains that belong to campaign Pharmacy-A (top campaign). Interestingly, the rate of accumulation is much slower (almost an order of magnitude) for hosts used as name servers, as shown in Figure 8(b).

Many domains in the same campaign have similar accumulation rates. We see many domains with similar slopes throughout the month. These domains tend to belong to a same campaign. However, not all the fluxing domains belonging to the same campaign have similar slops (See Figure 8(c)). One reasonable explanation is that a scam campaign runs on multiple fast flux networks, each of which has a different rate of recruiting and swapping in a new host. In any case, it is alarming to see that many fluxing domains can easily throw in thousands of hosting servers and hundreds of name servers over a month.

Some domains only begin accumulating IP addresses after some period of dormancy. Some domains appear to exhaust available hosts for a while (days to weeks) before



(a) A records

(b) IP of NS records

(c) A records - Top 4

(d) IP of NS records - Top 4

Figure 8: Cumulative number of distinct IPs for the A records and IP addresses of NS records, for the first week (first 2,000 iterations of data collection) for the Pharmacy-A domains, and for the top 4 campaigns across the four weeks of data collection.

accumulating new IP addresses. We examined two campaigns that exhibited rapid accumulation of IP addresses after some dormancy. In both cases, only one domain per campaign begins accumulating IP addresses. These two domains shared exactly the same set of NS names. These 8 NS names are doing all the work for those two campaigns. We observed other scams (e.g., the canadian pharmacy as well) where a few domains accumulate IP addresses faster than others. In addition to accumulation, we also saw attrition: 10% of fluxing domains became unreachable in the while we were monitoring them. These domains may have been blacklisted and so removed by registrars or by scammers themselves.

Rates of accumulation differ across campaigns. Figures 8(c) and 8(d) show the rate of accumulation of IP addresses for the top four campaigns for the IP addresses of A records

Table 6: Location of change for all campaigns, sorted by the total number of distinct IPs of A records.

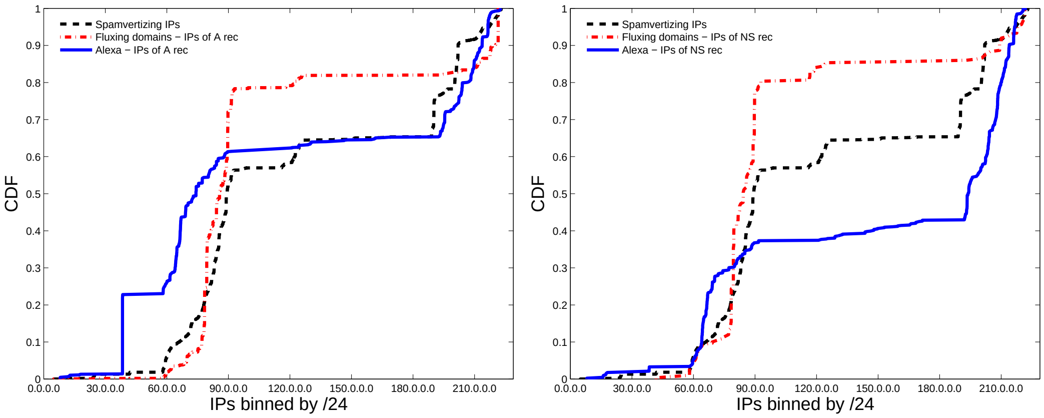
Campaign	Fluxing domains	Location of change						
		A	[IP of NS]	NS	A+ [IP of NS]	A+ NS	NS+ [IP of NS]	A+NS +[IP of NS]
Pharmacy-A	149				77			72
Watch-A	30	4	1		24			1
Watch-B	19		18					1
Pharmacy-B	52	5	13		19			15
Casino-A	6		1		5			
Pharmacy-C	6		6					
Casino-B	1				1			
Links-A	2					1		1
Casino-C	5				5			
E-Marketing-A	4	4						
Pharmacy-D	50	2	3		45			
Pharmacy-E	4				4			
Links-B	35		1		34			
Pharmacy-F	2				2			
Pharmacy-G	2				1			1
Links-B	2			2				
Service-A	1				1			
Software-A	5		5					
Watch-C	5		4		1			
DomainNames-A	3	3						
Service-B	1			1				
Total	384	18	52	3	219	1		91
Alexa	500 (domains)	37	5	15	4	1	1	

and NS records, respectively. The rate of accumulation for each campaign is higher than that of each fluxing domain. Because of resource sharing across the domains in a campaign, the total number of distinct IP addresses for a campaign is fewer than the sum of that of an individual domain. Section 3.5 will discuss how infrastructure is shared across domains and across campaigns in more detail.

3.4.3 Location of Change in DNS hierarchy

While it is possible to change any record of your own domain in DNS hierarchy (A, NS, and IP addresses of NS records), it is substantially more difficult to change NS records or A records of name server names as this requires updating a parent domain's (often a top level domain such as .com) zone file. However, we see many fast-flux domains that freely change NS records or IP of NS records separately or in combination of other records.

Campaigns change DNS record mappings at different levels of the DNS hierarchy. Table 3.4.2 shows the type of change for each campaign. In contrast to previous studies [40, 102], we observe many different types of changes in addition to single flux (A records) and double flux (A + IP of NS). Another notable point is that each campaign tends to use mixes of techniques (e.g., For Pharmacy-A, 52% of domains are double flux and 48% change all three types of records). We believe that this is another indication that a campaign operates



(a) A records

(b) IP of NS records

Figure 9: Distribution of the IPs of A rec, of autoritative servers and the IPs that sent the originating spam.

on multiple fast flux service networks.

Table 7: Top 10 ASes by number of IPs.

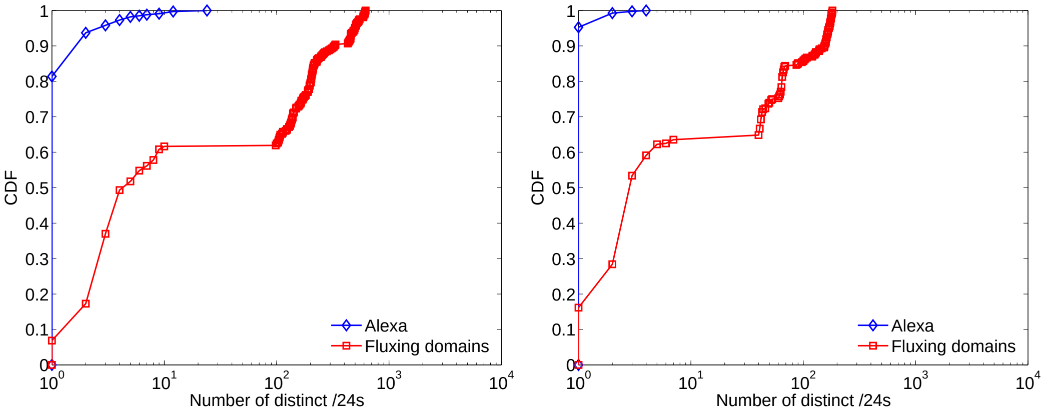
Top ASes by A Rec	Top ASes by IP of NS Rec	Top ASes by Spamming IPs
8402 - CORBINA-AS (1232)	12714 TI-AS NetByNet Holding (365)	9121 TTNET (6566)
12714 - TI-AS NetByNet Holding (1127)	3904 - HUTCHISON-AS-AP (260)	6147 Telefonica del Peru (3173)
9304 - HUTCHISON-AS-AP (951)	8402 - CORBINA-AS (224)	22927 Telefonica de Argentina (2726)
7132 - AT& T (511)	7132 - AT& T (121)	5617 TPNET Polish Telecom (2356)
6855 - SK SLOVAK TELECOM (345)	9908 - HKCABLE2-HK-AP (91)	19262 VZGNI-TRANSIT Verizon (2107)
13184 - HANSENET (332)	12695 - DINET-AS (81)	4837 CHINA169-BACKBONE (1697)
12695 - DINET-AS (307)	20597 - ELTEL-AS (72)	7738 Telecomunicacoes da Bahia (1524)
3209 - Arcor IP-Network (270)	13184 - HANSENET(66)	8359 COMSTAR (1436)
8615 - CNT-AS (252)	4766 - KIXS-AS-KR Korea Tel. (60)	4134 CHINANET-BACKBONE (1344)
3320 - DTAG (203)	30784 - ISKRATELECOM-AS(59)	9829 BSNL-NIB (1340)

3.5 Roles

This section describes the roles (e.g., content hosting, name service, spamming) played by hosts in fast flux service networks and how these roles evolve over time. We first examine the geographic and topological locations of fast-flux nodes; we also compare these locations to the spamming hosts that mount the messages in the scam campagins. We then explore how the roles of fast-flux nodes evolve over time, and how the fast-flux infrastructure is shared across different scam campaigns.

3.5.1 Location

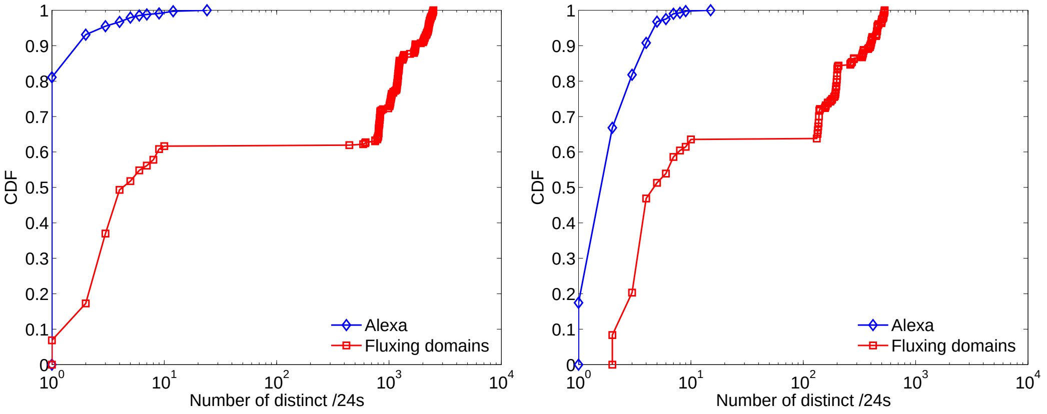
In this section, we examine the network and geographic location of fast flux hosts and compare them to both legitimate Web sites and the spammers who advertise the scams.



(a) A records

(b) IP of NS records

Figure 10: Distribution of unique /24s that appeared as the *first* record in a reply.



(a) A records

(b) IP of NS records

Figure 11: Distribution of unique /24s that appeared for *all* records in a reply.

3.5.1.1 Network Location

This section describes how fast-flux IP addressess are spread across IP address space. To examime whether fast-flux service networks use different portions of the IP space than the top 500 domains, we plotted the distribution of the IPs across the whole IP range. Figure 9 shows that fast-flux networks use a different portion of the IP space than sites that host popular legitimate content: The IPs that host legitimate sites are considerably more distributed: and more than 30% of these sites are hosted in the 30/8-60/8 IP address range, which hosted almost none of the scam sites observed in our study.

Fast-flux hosts are concentrated in small regions of IP address space; some spammers are concentrated in slightly different regions. Interestingly, the IP address space that hosts fast-flux domains is even more concentrated than that which sends spam advertising the scam campaigns. Although the distributions are ismilarly concentrated in the 80/8 - 90/8 range, there is a much higher concentration of spammers in the 200/8 - 210/8 range (ranges allocated to Latin America and Asia, respectively). These differing distributions suggest that hosts in different regions of the IP address space do in fact play different "roles" in spam campaigns.

DNS lookups for fast-flux domains often return much more widely distributed IP addresses than lookups for legitimate Web sites. Our intuition was that fast-flux networks that hosted scame sites would be more distributed across the network than legitimate Web hosting sites, particularly from the perspective of DNS queries from a single client (even in the case of a distributed content distribution network, DNS queries would tend to map a single client to nearby Web cache). Figures 10 and 11 show the distribution of distinct /24s that appear at the answer section of the DNS replies) for the first record in the reply and for all records in the reply, respectively. It turns out that a few legitimate domains that are hosted by content distribution networks appear to have the largest number of distinct /24s contained in a single DNS reply. In particular, `www.runescape.com`, `www.statcounter.com`, `www.yahoo.co.jp`, `www.monografias.com` returns IP addresses in 12, 8, 7, and 6 distinct /24s respectively. We also observed several legitimate domains which showed a large number of distinct /24s for their IPs of NS records (which actually reflects good network design because it introduces redundancy). Examples include `www.altavista.com`, `www.geocities.com`, `www.runescape.com`, `www.php.net` which had 11, 9, 8, and 7 distinct /24s in IPs of NS records.

Fast flux domains tend to return IP addresses that are distributed across a larger number of distinct /24s than legitimate domains. Indeed, roughly 40% of all A records returned for fast-flux domains were distributed across at least 300 distinct /24s, and many were distributed across thousands of /24s. In contrast, domains for popular Web sites were never distributed across more than 12 distinct /24s (when queried from a single location). Thus, overly widespread distribution of query replies may serve as a strong indicator that a domain is indeed hosted by a fast-flux network.

The predominant networks that host fast-flux infrastructure differ from those that host spammers for the corresponding scam campaigns. Table 7 shows the top ten ASes by the number of IP addresses for A records (i.e., hosting sites), NS records (i.e., nameservers), and spammers (as observed in the spam trap). Interestingly, although there is some overlap between the ASes that host the scam sites and those that host authoritative nameservers, there is almost no overlap between the ASes that host the sites and nameservers for the scams do not overlap much with the ASes hosting the spamming IP addresses. Indeed, Figure 9 also shows that spammers for the campaigns we observed are more heavily concentrated in Latin America, Turkey, and the United States, whereas fast-flux hosts are more concentrated in Asia. The fact that significant differences exist between networks that host fast-flux infrastructure and those that host spammers suggest that scammers may have divided the infrastructure into different roles (in Section 3.6, we see that many fast-flux hosts are not listed on spam blacklists, which is consistent with this observation).

3.5.1.2 Geographic location

Hosting servers and name servers are widely distributed. Table 8 lists country names in which fast flux nodes are hosted, according to the country of the AS in which they are hosted. In total, we observed IP addresses for A records in 283 ASes across 50 countries, IP addresses for NS records in 191 ASes across 40 countries, and IP addresses for spammers for the corresponding scam campaigns across 2,976 IP addresses across 157 countries. Although many fast flux nodes appear to be in Russia, Germany, and the US, the long list of ASes and countries shows that fast flux service networks are truly distributed; this kind of geographical distribution may be necessary to accommodate the diurnal pattern of compromised hosts' uptime [23]. Interestingly, the countries that are referred to by the most A records are not the same set of countries that host authoritative nameservers for those domains (as indicated by IP addresses of NS records). In particular, Slovakia, Israel, and Romania appear to host relatively more nameservers than sites, and China appears to host relatively more nameservers. This difference in distribution deserves further study; one possible explanation is that nameserver infrastructure for fast-flux networks must be more robust than the sites that host scams (which might be relatively transient). countries

3.5.2 Sharing Across Campaigns

In this section, we describe our findings regarding the sharing of the same fast-flux infrastructure across multiple scam campaigns.

Many fast-flux machines have dual roles, and different campaigns share hosting infrastructure. Referring back to Table 3, the last three columns indicate that many hosting servers double as name servers (and vice versa). 16 out of 21 campaigns (76%) show such sharing. On the contrary, we see a clear role separation of the hosts associated with the domains of the popular Web sites listed by Alexa. We also find significant overlaps among the hosts involved for the top four campaigns. Table 9 shows that Watch-A and Watch-B are likely to share the underlying infrastructure—99% of hosting servers, 80% of

Table 8: Top 10 countries by number of IPs.

Top Countries by A Rec	Top Countries by IP of NS Rec	Top Countries by Spamvertising IPs
Russia (4025)	Russia (982)	US (6972)
Germany (1207)	Hong Kong (425)	Turkey (6580)
Hong Kong (1207)	Germany (216)	Russia (5914)
US (606)	US (168)	Brazil (4606)
Slovakia (391)	Korea (154)	Argentina (4268)
Korea (350)	China (77)	China (4041)
Israel (337)	Japan (64)	Poland (3424)
Japan (248)	Taiwan (48)	India (3302)
Ukraine (247)	Ukraine (40)	Peru (3214)
Romania (131)	Slovakia (39)	Germany (3122)

NS records, and 98% of name servers of Watch-B are common with those used for Watch-A. Moreover, both campaigns share many of the servers and NS records with Pharmacy-A. This overlap strongly suggests that the all three campaigns involve same fast-flux service networks. Interestingly, our observation is consistent with Spam Trackers [91], which attributes all the three scam campaigns to Yambo Financials [92].

3.6 Relationship to Blacklists

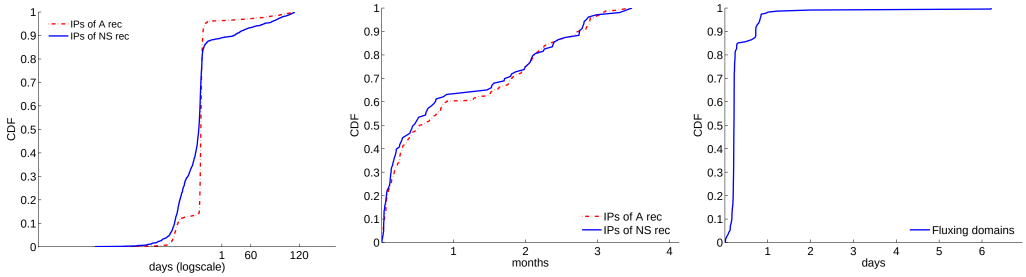
In this section, we evaluate whether the IPs that show up as part of the fast-flux network hosting infrastructure appear on various blacklists: (1) the Spamhaus spam blacklist (SBL/PBL) [98]; (2) the Spamhaus exploit blacklist (XBL) [99]; and (3) the URI blacklist (URIBL) [105]. We find, generally, that the time to blacklisting varies significantly by blacklist, and that many fast-flux IP addresses are not listed in the SBL; those that are tend to be listed both before and after we observed fast-flux activity.

Method. To determine whether the IP addresses in our dataset are blacklisted at the time that we witness them as part of fast-flux infrastructure, or whether they become blacklisted at some later point, we query each blacklists database for historical information about listing. Georgia Tech actively runs mirrors for SpamHaus SBL/PBL/XBL and for URIBL, which gives us access to precise information about when each IP address or domain is listed in the database. We query the following databases:

- XBL, a real-time database of IP addresses of infected machines including open proxies worms/viruses with built-in spam engines, and other types exploits.
- SBL, a realtime database of IP addresses of verified spam sources and spam operations.

Table 9: Sharing among the top 4 campaigns.

Sharing of A records				
	Pharmacy-A	Watch-A	Watch-B	Pharmacy-B
Total per campaign	9448	1516	1204	15
Pharmacy-A	-	1510	1203	1
Watch-A	1510	-	1203	1
Watch-B	1203	1203	-	1
Pharmacy-B	1	1	1	-
Sharing of NS records				
	Pharmacy-A	Watch-A	Watch-B	Pharmacy-B
Total per campaign	52	14	10	10
Pharmacy-A	-	8	8	0
Watch-A	8	-	8	0
Watch-B	8	8	-	0
Pharmacy-B	0	0	0	-
Sharing of IPs of NS records				
	Pharmacy-A	Watch-A	Watch-B	Pharmacy-B
Total per campaign	2340	225	219	13
Pharmacy-A	-	220	215	9
Watch-A	220	-	215	9
Watch-B	215	215	-	6
Pharmacy-B	9	9	6	-



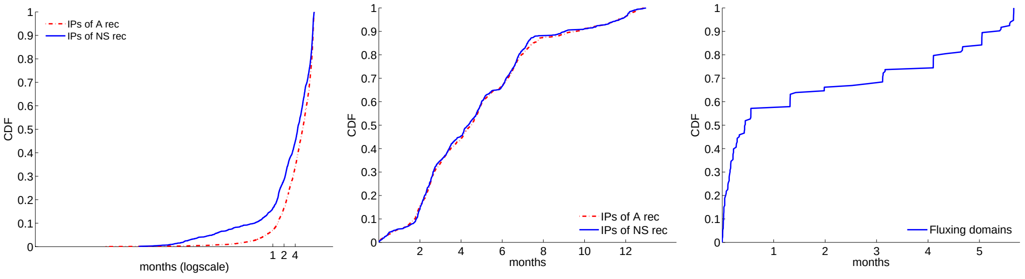
(a) XBL

(b) SBL/PBL

(c) URIBL

Figure 12: CDF of time elapsed between the appearance of an IP address in our dataset, either as IP of A record or IP of NS record of a fluxing domain and the timestamp of appearance at Spamhaus BL. Also the same for the fluxing domains and the elapsed time before they were blacklisted at URIBL.

- PBL, a database of end-user IP address ranges that should not be delivering unauthenticated SMTP email to any Internet mail server except those provided for specifically by an ISP for that customer's use.
- URIBL, a realtime blacklist that lists domains that appear in spam and are likely phishing or scam sites.



(a) XBL (b) SBL/PBL (c) URIBL

Figure 13: CDF of time elapsed between the appearance of an IP address, either as IP of A record or IP of NS record of a fluxing domain at a blacklist, and the timestamp of appearance in our dataset (The “opposite” from Figure 12).)

Table 10: IPs of A records, IPs of NS records and domains which were blacklisted before (B), after (A) or before and after (B+A) when they appeared at our collection of DNS records.

	SBL/PBL				XBL				
	Never	B	A	B+A	Never	B	A	B+A	Total
A	1692	29	283	7517	265	244	2648	6364	9521
NS	547	7	80	1787	183	98	481	1659	2421

URIBL

	Never	B	A	B+A	Total
Domains	113	0	138	133	384

Many fast-flux IP addresses and domains do not appear in blacklists at the time when their activity is first observed. We queried the blacklist data at the end of April 2008 for historical information (back to February 2007) for each IP address and domain from our dataset. Table 10 shows the number of IPs that were already blacklisted before we observed them at our dataset, IPs that were blacklisted after we observed them in our dataset, IPs that were blacklisted as active before and after we observed, and finally IPs that were never blacklisted (by the time we querried the BLs database). Table 10 shows that a significant fraction of IP addresses hosting scam infrastructure (more than 17%) were never listed in the SBL; considerably higher fractions were listed in the XBL and URIBL, although many of the IP addresses and domains listed in the XBL and URIBL respectively were only listed *after* we observed activity from those IP addresses and domains. The lack of these IP addresses in the SBL could suggest one of two things: (1) the SpamHaus SBL is incomplete; or (2) the SBL may simply not list this fraction of IP addresses because it was never used to spam (i.e., it only hosted scam infrastructure).

Table 11: IPs of A records, IPs of NS records and domains which appeared at our spamtrap as spam relays before (B), after (A) or before and after (B+A) their time of appearance at our collection of DNS records.

	Not appeared	Before	After	Bef.+After	Total
IPs of A rec	9417	11	92	1	9521
IPs of NS rec	2420	5	16	0	2421

Time to listing after activity is observed can vary from hours to weeks, depending on the blacklist. IP addresses tend to take longer to show up in the Spamhaus SBL. To determine how long it takes for IP addresses to appear in various blacklists after we observed their activity, we measured the time between when we observed the IP addresses participating in fast-flux activity and the time when they were first blacklisted. Figure 12 shows the distribution of these delays. We plot the CDFs of the elapsed times between appearance and listing for XBL, SBL/PBL, and URIBL.

Most IP addresses are listed relatively quickly (if they are not already listed when we observe either activity), but for some IP addresses and domains, the time that elapses between the time we first observe activity and the time an IP address or domain is listed is on the order of weeks. These delays in listing IP addresses in the SBL suggests that there are parts of fast-flux networks that are used first as flux agents and later as spam relays. In these cases, monitoring hosts for fast-flux activity may be useful for predicting future spamming activity. Figure 13 shows the same distribution, but for IP addresses that were listed before we observed activity from them in our dataset. Interestingly, most IP addresses that were listed before we observed their activity were listed in the XBL weeks to months before we observed them (IP addresses for A records were listed sooner).

We observe a small amount of overlap between IP address that host fast-flux infrastructure and those that send spam to our spam trap. To further understand the relationship between spamming infrastructure and the scam hosting infrastructure, we examined the overlap between IP addresses that spam and those that host infrastructure: For each IP address that we observed (IPs of A records and IPs of NS records), we checked to see whether the IP had sent any spam emails to the same spam trap from which we extracted the fluxing domains over the period of October 2007 through February 2008 (i.e., from nearly three months before the start of our collection of fast flux data until 1 month after our data collection). Table 11 shows that a very small fraction of IP addresses (about 1%) sent spam to our spam trap either before or after the time when we observed them as part of the scam hosting infrastructure. The small overlap may simply reflect the fact that our trap only sees a fraction of all spam (and spammers). These spamming IPs advertise the same fast-flux domains that they are hosting, which suggests that the spamming infrastructure and the hosting infrastructure may be shared.

3.7 Conclusion

This chapter has presented an empirical study of the dynamics and roles of fast-flux networks in mounting scam campaigns. We actively monitored the DNS records for URLs for scam campaigns received at a large spam sinkhole over a one-month period to study the rates of change in fast-flux networks, the locations in the DNS hierarchy that change, and the extent to which the fast-flux network infrastructure is shared across different campaigns. We also contrast the dynamics observed in these networks to that used for load balancing for popular Web sites. Our findings suggest that monitoring the infrastructure for unusual changes in DNS mappings may be helpful for detecting scams hosted on fast-flux networks. In future work, we plan to use these features design a detection scheme that can automatically identify scam campaigns based on invariant properties of the infrastructure. We expect that doing so may allow us to detect online scams automatically, and considerably faster than today's manual blacklisting mechanisms.

CHAPTER IV

RE-WIRING ACTIVITY OF MALICIOUS NETWORKS

4.1 Introduction

Securing the Internet's routing system has been a concern for both network operators and protocol designers for nearly fifteen years. One of the frequently stated reasons for securing the Internet's interdomain routing protocol, the Border Gateway Protocol (BGP), is that attackers may use BGP to launch their attacks and hide their traces. Previous work has exposed some specific techniques that attackers use; for example, Ramachandran *et al.* observed that some attackers send spam from short-lived prefixes [80]. Nevertheless, little is known about how ASes that host these attackers exploit BGP to provide them with further protection. There have been some publicized cases of malicious ASes—major hubs of illegal activity—namely Atrivo/Intercage and VolgaHost, that were observed to get frequently de-peered. Eventually they were officially reported and cut off by all providers in September 2008 and January 2011, respectively. This practice of frequent change in upstream connectivity and eventual cut-off may constitute a feature that potentially characterizes malicious ASes, regardless of the type of attacks they launch which varies across time. We refer to the activity that is related with change of connectivity of an AS as *re-wiring activity* of this AS.

This chapter presents the first systematic study of the re-wiring activity of malicious ASes, with the goal of improving our understanding of how malicious networks exploit interconnection through different upstream ASes to cover their tracks. Rather than attempting to detect any individual type of attack (e.g., spam, denial of service), we characterize the re-wiring activity of malicious networks that are primarily responsible for attacker activities. We identify features of the re-wiring behavior that may be more stable across time than the characteristics of any single attack. We believe that ultimately certain aspects of routing behavior may serve as invariants for detecting malicious infrastructure, even as the attacks themselves evolve.

We draw the following conclusions from our study:

- *Malicious Enterprise Customers (ECs) on average change their upstream connectivity more aggressively than non reported ECs.* We offer a new class of observations on the AS-level re-wiring activity of malicious ECs. ECs are typically stub networks

(Section 4.2). We find that, on average, malicious ECs change their upstream connectivity more frequently and link with a larger set of providers throughout their lifetime than non-reported ECs. We observe that malicious ECs link on average with a total of twelve providers from 1998–2010, whereas the rest of ECs link with only about four providers on average. Also, malicious ECs' peering sessions last for less time; the top 5% of CP links formed by malicious ECs are observed for 14–31 consecutive snapshots of the AS graph, whereas the top 5% of the CP links formed by the rest of ECs are observed for 27–51 consecutive snapshots.

- *Malicious ECs prefer to attach to popular providers as non reported ECs do when they first appear, but they attach to less popular providers when they re-wire later in their lifetime.* In 2009–2010, all malicious ECs that were observed for the first time used the *most-preferred providers* (the providers that are responsible for the most CP links generated by non-reported ECs at that snapshot) for transit, as the rest of the ECs did (Section 4.3). During the same time period, the malicious ECs that were observed to re-wire, attached to providers from which only a fraction of 30–50% were among the top providers preferred by non-reported ECs.

Our results highlight some findings that have implications for the design of more effective defense mechanisms against Internet attacks, which we discuss in more detail in Section 4.4. Even though an AS may be acting maliciously, network administrators in other networks may not be aware of it. Instead, if AS rewiring activity is monitored and potential malfeasance is reported, then this information may be useful to network administrators. For example, in the case of AS-Atrivo, even if not all network administrators knew that AS-Atrivo was acting maliciously or the details of the attacks, it may have been helpful to know that its re-wiring activity was suspicious, in case they receive suspicious traffic originating from that AS or if AS-Atrivo tries to connect directly with them. The re-wiring features are independent of the evolving nature of the attacks and also more difficult for the attackers to tamper, incorporating them may provide significant gains over existing defense mechanisms or complement them.

The rest of this chapter is organized as follows. In Section 4.2, we describe the datasets we used in our analysis. Section 4.3 presents our findings and Section 4.4 offers some observations about the implications of our findings with regard to existing defense mechanisms. We conclude in Section 4.5.

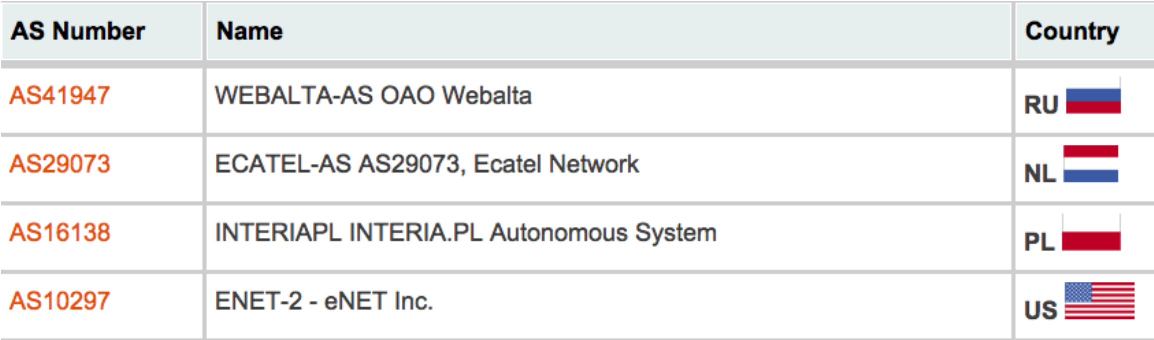
4.2 Data

Our primary dataset consists of a list of networks that are reported as top in Internet criminal activities, in a period of three years (2009–2011) according to Hostexploit. There are a total of 129 distinct ASes. We hypothesize that malicious ASes have different business functions and incentives than most ASes, which may be reflected on the links they form with other ASes in the Internet. To understand the re-wiring behavior of these networks we obtain a publicly available dataset of customer-provider links formed among ASes over ten years (available here [25]), over which we track them. We considerASes that have not ever been reported by Hostexploit to be legitimate.

Hostexploit reports. To study the behavior of malicious networks that are mostly engaged in Internet nefarious activities, we collected the list of ASes reported as most malicious networks by Hostexploit in global Internet activity reports issued in quarters of 2009–2011. Hostexploit is an open organization of international volunteers who are Internet professionals within the areas of web hosting, server management, DNS, Internet security, and intrusion detection systems, with a focus on creating awareness of cybercrime activity. Hostexploit integrates and correlates data from multiple sources such as spam, malware, malicious URLs, spam bots, botnet command and control servers, phishing servers, exploit servers, and cyber-warfare intelligence provided by industry partners. To provide a list of the worst networks, they rate each AS with a Hostexploit Index (HE), based on the activity of the AS weighted by the size of its allocated address space. HE is a proportional rather than an absolute index of AS maliciousness. Hostexploit has been issuing reports on top networks by their HE index periodically since 2009.

Figure 14 shows an example Hostexploit report for the first quarter of 2011. The report is accompanied by an analysis of the quarterly results. For example, they compare the new ranking results with the previous quarterly ranking, show the top worst networks by sector (spam, botnet hosting, etc.), show an analysis by country, etc.

Historic AS relationships. To understand the re-wiring behavior of malicious networks, we tracked these ASes' rewiring activity over a ten years. Dhamdhere *et al.* [25] collected BGP AS paths from BGP table dumps obtained from repositories at RouteViews and RIPE. They collected snapshots of AS paths; a snapshot refers to a period of 21 days (not an instant). During each snapshot, they collected at five different times the unique AS paths from all active monitors. Then, to obtain the primary Internet links (used most of the time) and to filter out the backup links (used during failures or overload conditions), they kept



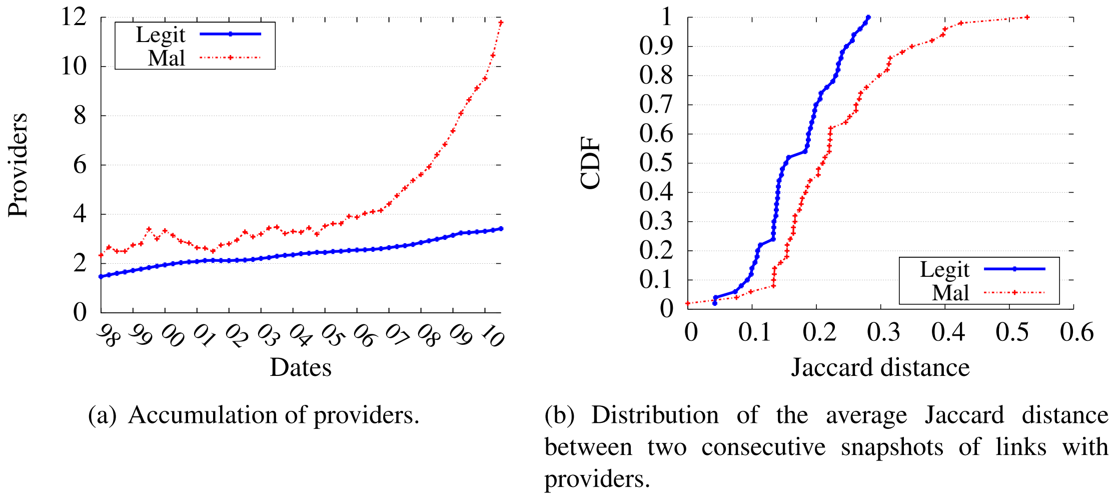
AS Number	Name	Country
AS41947	WEBALTA-AS OAO Webalta	RU
AS29073	ECATEL-AS AS29073, Ecatel Network	NL
AS16138	INTERIAPL INTERIA.PL Autonomous System	PL
AS10297	ENET-2 - eNET Inc.	US

Figure 14: A quarterly report from Hostexploit from the 1st quarter of 2011. The report ranks ASes by an index (HE) based on the activity of the AS, weighted by the size of its allocated address space.

only the AS paths that appeared in the majority of the samples and ignored the rest (majority filtering algorithm). Then, they used the AS paths in each snapshot (those that had passed the majority filtering process) to infer the underlying AS topology and the relationships between adjacent ASes. Finally, they used the well-known algorithm proposed by Gao [32]. Gao's algorithm resulted in four types of AS relationships: Customer-Provider (CP), peering, sibling, and unknown. To understand the evolution of the Internet ecosystem they classified ASes into types depending on their function and business type, based on their observable topological properties. They classified ASes into Enterprise Customers (ECs), small transit provider, large transit provider, content access and hosting provider.

In this study, we focus only on the CP links of ECs (62 in total). We note that based on the method that the classes were derived, ASes of the same class are close in terms of the number of customers and peers they link with at each snapshot. As far as the number of providers they link with at each snapshot there is larger variability among the ASes, but our analysis is not relying on this metric. Instead, we are looking at the accumulation of distinct providers across time, the CP link duration and other features (see Section 4.3).

Dataset limitations. Our data has the following limitations: (1) The Hostexploit data may be inaccurate. (2) The methods that were used to infer the AS graph [25] have been shown to to be inadequate. Most ASes are detected but a significant fraction of peering and backup links at the edges of the Internet are missed. (3) Dhamdhere *et al.*'s AS classification algorithm may introduce errors.



(a) Accumulation of providers.

(b) Distribution of the average Jaccard distance between two consecutive snapshots of links with providers.

Figure 15: Malicious ECs on average attach to more providers and change their providers more frequently than legitimate ECs.

4.3 Results

We hypothesize that malicious ASes have different business functions and incentives than the rest of the ASes, and that these differences may be reflected on their re-wiring activity. For example, we suspect that malicious ASes may get frequently de-peered from their providers and that they need to re-wire more frequently than legitimate ASes. To test our hypothesis, we survey the wiring trends of malicious networks; (1) re-wiring frequency and (2) attachment preference. To determine whether our findings may be suitable for characterizing network re-wiring activity, we compare the behavior of malicious ASes to that of non-reported ASes over a ten-year period. We present the results of our analysis for Enterprise Customers (ECs) only.

4.3.1 Re-Wiring Frequency

To understand the re-wiring dynamics of malicious networks and, more specifically, how these networks change their upstream connectivity, we examine three features; (1) the number of providers they link with throughout their lifetime, (2) how frequently they change providers and (3) how long their CP links last. To determine whether these features may be suitable for characterizing the re-wiring activity of malicious ECs, we compare the behavior of malicious ECs to that of the rest of the ECs over ten years.

Malicious ECs on average connect to more providers than legitimate ECs do. To investigate how malicious ECs link with their providers we first look at the total number of providers they link with throughout their lifetime. To compute the cumulative number of providers that an EC has, on average, first we track each EC for the duration of its lifetime; we add the total number of providers it links with to the total number of providers for its group (malicious or rest of EC). Then, for every snapshot, we compute the average by dividing the cumulative total number of providers of the EC group with the number of ASes in the EC class. We note that across time the size of the group increases as new ECs appear, and as does the cumulative number of providers for the group of EC. The rate with which the two numbers increase does not stay the same across time, so some snapshots exhibit a small decrease in the average cumulative number of providers. Figure 15(a) shows the cumulative number of providers that ECs attach to, on average, for two cases: malicious ECs and the rest of ECs. On average, malicious ECs link with more providers throughout their lifetime than the rest of the ECs do.

Malicious ECs link on average with a total of twelve providers during the period 1998–2010, whereas the rest of ECs link on average with only about four providers. The top ECs by the total number of providers are AS 3, which connected to 83 distinct providers; AS 26415, which linked with a total of 68 providers; and the private AS 65000, linked to 20 different providers. On the other hand, the malicious EC that we observed to link with the most providers (a total of 83) through its lifetime was AS 23456. We note that this AS is reserved by IANA and is used for backward compatibility between old (2-byte AS support) and new (4-byte AS support) BGP speakers. Because it is reserved for backwards compatibility, this AS is likely not a single AS, but rather a group of ASes; we must investigate this further to better understand the individual ASes that are perpetrating suspicious activity.

Malicious ECs on average change upstream providers more frequently than legitimate ECs. To quantify the aggressiveness of an EC in changing its upstream connectivity, we consider the distance between the set of the AS's providers for two consecutive snapshots. We use the Jaccard distance as a metric of distance between the set of providers of two consecutive snapshots. For example, a Jaccard distance of 0.8 indicates that 80% of the links seen in the two snapshots are observed in one of the two snapshots but not in both. We calculated the Jaccard distance between two consecutive snapshots for each AS throughout its lifetime. Figure 15(b) shows the distribution of the Jaccard distance values for malicious ECs and the rest of the ECs throughout their lifetime. We observe that malicious ECs are more aggressive on average in changing their upstream connectivity than the rest of ECs.

The fact that malicious ECs change their upstream connectivity more frequently than legitimate ECs may be a reflection of the following: (1) malicious ECs are noticed by their providers regarding their attacker activities, they get de-peered and they are forced to change upstream connectivity more frequently than legitimate ECs, (2) malicious ECs

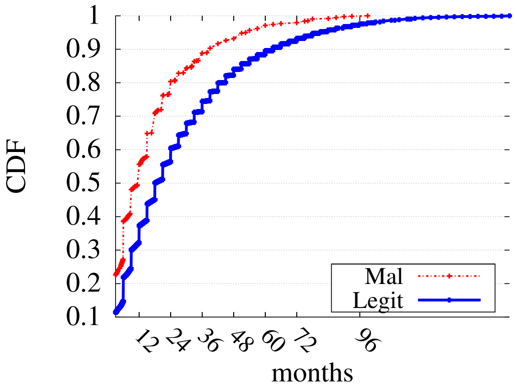
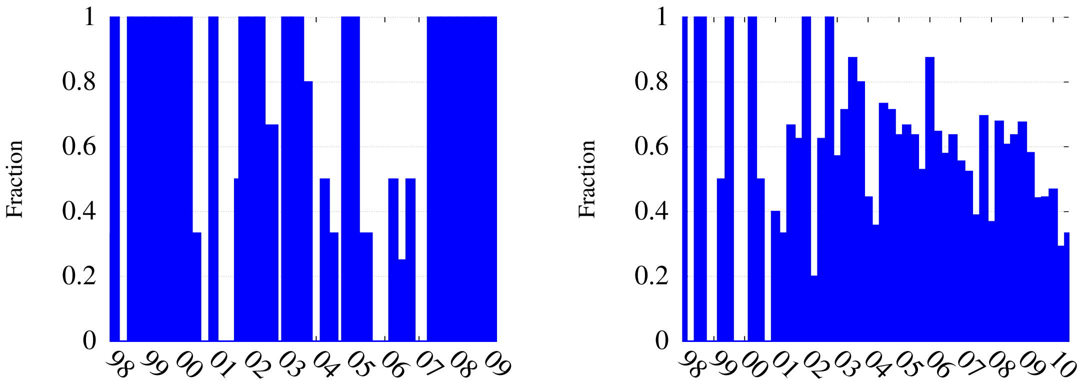


Figure 16: Distribution of the average duration of CP links formed by ECs with their providers. Malicious ECs on average form shorter duration links than the rest of the ECs.

attempt to find providers with less strict restrictions, which may make it easier for them to launch their attacks, (3) malicious ECs attempt to avoid accountability or legal consequences of their activities.

Malicious ECs on average form shorter duration CP links than legitimate ECs. To better understand the behavior of malicious ECs, we consider the duration of the links they form with their providers. To compute the CP link duration, we proceed as follows: First, for every snapshot, we determine the links that are present. Second, for each link, we determine whether it is present in the previous snapshot. Finally, for each link, we measure the total number of consecutive snapshots it was present. In cases where a link appears multiple times throughout 1998–2010 we consider the average duration of that link. Figure 16 shows the distribution of average CP link duration of malicious ECs and the rest of ECs. We observe that malicious ECs on average form CP links that are shorter in duration than the CP links that legitimate ECs form.

Figure 16 shows that approximately 22% of the CP links formed by malicious EC and about 10% of CP links formed by the rest of ECs were observed for only one snapshot. For example ASes in AS 23456, which linked with the largest number of providers, formed CP links with ASes 30083, 5617, and 7643, which appeared for only one snapshot. Examples of short-lived CP links formed by non-malicious ECs include AS 20195–AS 6395, AS 39709–AS 28870, and AS 3748–AS 4554. These links were observed for only one snapshot. On the other hand, as far as long-lived CP links, the top 5% of CP links formed by malicious ECs are observed for 14–31 consecutive snapshots, whereas the top 5% of the CP links formed by non-malicious ECs are observed for 27–51 consecutive snapshots. The most long-lived CP links by malicious ECs are links AS 14280–AS 6327 which was



(a) First attachment.

(b) Re-wiring.

Figure 17: Fraction of malicious ASes providers' that belong to the most preferred providers of all ECs for each snapshot. Malicious ECs in some cases attach to the most popular providers when they first appear, but not when they rewire. This is especially true for the more recent years.

observed for 33 consecutive snapshots, AS 17974–AS 7713 for 31 consecutive snapshots and AS 13174-AS 1299 for 30 consecutive snapshots. The most long-lived CP links by legitimate ECs were observed for 51 consecutive snapshots (i.e., the entire period from 1998 to 2010); some examples are: AS 8581–AS 5408, AS 2508–AS 2907, AS 7065–AS 701, AS 10357–AS 7066, and AS 1104–AS 1103.

4.3.2 Attachment Preference

In this section, we study the providers that malicious ECs connect to. Our goal is to determine whether malicious ECs have different attachment preferences than the rest of the ECs. We define the *most preferred providers* as the providers that are responsible for the most CP links generated by non-reported ECs at that snapshot. We find that malicious ECs do not connect to the most preferred providers as non-reported ECs do.

First, at each snapshot, we determine the most preferred providers as follows: we calculate the total number of CP links generated by non-reported ECs at each snapshot and we extract the providers that are responsible for at least 60% of the total at that snapshot. This set comprises the most preferred providers of non-reported ECs at each snapshot. Second, for each snapshot, we determine the providers that malicious ECs link with and whether those links are first-time attachments or re-wirings. Finally, at each snapshot, we calculate the fraction of the providers that malicious ECs link with that also belong to the most preferred providers of non-reported ECs. We calculate this fraction separately for the cases of first-attachment (when an EC is observed for the first time) and re-wiring (when an EC re-wires).

Figure 17 shows the fractions of the providers of malicious ECs that belong to the most preferred providers of non-reported ECs, for the cases of first attachment and re-wiring across time. We observe that for the case of first attachment, all or almost all of the providers they link with are also among the most preferred providers of non-reported ECs. In contrast, for the case of re-wiring, only a percentage of the providers that malicious ECs link with are also among the most preferred providers of non-reported ECs. We observe that for the years 2009–2010, approximately 30–50% of the providers that malicious ECs link with are also among the most preferred providers of non-reported ECs.

4.4 Recommendations and Future Work

Although we have focused on *characterization* of re-wiring activity, we believe that routing behavior associated with malicious networks has properties that can lead to both better protection against Internet attacks and more efficient detection of the networks that host the attackers. Specifically, because it is independent of any particular attack, routing behavior may serve as invariant behavior that can help identify networks that perpetrate a wider variety of attacks. Routing data is also publicly available and can complement other types of detection mechanisms. In the remainder of this section, first we discuss various lessons from this characterization study that may ultimately inform detection methods and second we discuss future work.

We first note that networks that consistently host malicious activity have re-wiring behaviors that are distinct from other networks. This may point to possibilities for stemming the tide of attack traffic in the future. In the past, we have seen dramatic de-peering events of a single AS (e.g., Atrivo/Intercage), which have resulted in a precipitous drop in spam traffic, which returned at a later date, likely as the spammers re-established upstream connectivity with other upstream ASes. Our analysis shows that the stub ASes that tend to originate a significant amount of attack traffic tend to re-wire with upstream ASes that are not "preferred" upstream ASes. In the future, more effective AS reputation systems might incorporate information not only about the AS itself that originates the traffic, but also the *upstream ASes to which the network connects*. Given the aggressive re-wiring in which malicious ASes participate, another possible way forward might be to encourage some type of registration or verification process by which an AS (or its upstream) is vetted as a reputable provider.

In future work, we plan to evaluate (1) the possibility to classify malicious ASes using re-wiring activity features we observed in this study and (2) the efficacy of using routing information as input to an AS reputation system. We also plan to investigate the re-wiring activity of malicious ASes by class (e.g., small transit providers, large transit providers and content providers) and also by the type of attack they appear to be engaged with (e.g., some malicious ASes may appear mostly to send spam during a specific period of time whereas other ASes may appear to be hosting command-and-control infrastructure). It remains to be seen whether a blacklisting system that is based on re-wiring behavior would be sufficiently different than one that uses observations of the attacks themselves as input.

4.5 Conclusion

Although limited empirical studies have suggested that attackers exploit BGP routing to help cloak their attacks, there has been no detailed longitudinal study of how malicious networks may interconnect differently from other ASes. In this chapter, we have analyzed more than ten years of BGP data in conjunction with reports of malicious ASes from Hostexploit and found that ASes that are known to host malicious traffic consistently exhibit different re-wiring behavior than other ASes. We believe that our findings may ultimately serve as useful features for other reputation or attack detection mechanisms. The fact that routing dynamics are a property of the network hosting the attack, rather than of any specific attack, may ultimately prove advantageous in this regard. In particular, using BGP routing data as an input to such an AS-based reputation system is a promising area for future work.

CHAPTER V

ASWATCH: AN AS REPUTATION SYSTEM TO EXPOSE BULLETPROOF HOSTING ASES

5.1 Introduction

Today's cyber-criminals must carefully manage their network resources to evade detection and maintain profitable illicit businesses. For example, botmasters need to protect their botnet command-and-control (C&C) servers from takedowns, spammers need to rotate IP addresses to evade trivial blacklisting, and rogue online businesses need to set up proxies to mask scam hosting servers. Often, cyber-criminals accomplish these goals by hosting their services within a malicious autonomous system (AS) owned by an Internet service provider that willingly hosts and protects illicit activities. Such service providers are usually referred to as *bulletproof hosting* [16], due to their reluctance to address repeated abuse complaints regarding their customers and the illegal services they run. Notorious cases of malicious ASes include McColo [54], Intercage [47], Troyak [68], and Vline [2] (these ASes were taken down by law enforcement between 2008 and 2011). According to Host-exploit's reports [35], these types of ASes continue to appear in many regions around the world—mostly in smaller countries with lower levels of regulation, but also in the United States—to support activities ranging from hosting botnet command-and-control to phishing attacks [36]. For example, the Russian Business Network [83], one of the most notorious and still active cybercrime organizations, have decentralized their operations across multiple ASes. In most cases, nobody notices bulletproof hosting ASes until they have become hubs of illegal activities, at which point they are de-peered from their upstream providers. For example, Intercage [47] was de-peered more than ten times before it reached notoriety and was cut off from all upstream providers.

To defend against these crime-friendly ASes, the community has developed several AS reputation systems that monitor *data-plane* traffic for illicit activities. Existing AS reputation systems typically monitor network traffic from different vantage points to detect the presence of either malware-infected machines that contact their C&C servers, send spam, host phishing or scam websites, or perform other illicit activities. These systems establish AS reputation by measuring the "density" of malicious network activities hosted within an AS. For instance, FIRE [101] tracks the number of botnet C&C and drive-by malware download servers within an AS. ASes that host a large concentration of malware-related servers are then assigned a low reputation. Similarly, Hostexploit [35] and BGP Ranking [12] compute the reputation of an AS based on data collected from sources such as DShield [26] and a variety of IP and domain name blacklists.

Unfortunately, these existing AS reputation systems have a number of limitations: (1) They cannot distinguish between *malicious* and *legitimate but abused* ASes. Legitimate ASes often unwillingly host malicious network activities (*e.g.*, C&C servers, phishing sites) simply because the machines that they host are abused. For example, AS 26496 (GoDaddy) and AS 15169 (Google) repeatedly appeared for years among the ASes with lowest reputation, as reported by Hostexploit. Although these ASes are legitimate and typically respond to abuse complaints with corrective actions, they may simply be unable to keep pace with the level of abuse within their network. On the other hand, *malicious* ASes are typically unresponsive to security complaints and subject to law-enforcement takedown. (2) Because of the inability to distinguish between *malicious* and *legitimate but abused* ASes, it is not clear how to use the existing AS rankings to defend against *malicious* ASes. (3) Existing AS reputation systems require direct observation of malicious activity from many different vantage points and for an extended period of time, thus delaying detection.

We present a fundamentally different approach to establishing AS reputation. We design a system, *ASwatch*, that aims to identify malicious ASes using exclusively *control-plane* data (*i.e.*, the BGP routing control messages exchanged between ASes using BGP). Unlike existing *data-plane* based reputation systems, *ASwatch* explicitly aims to identify *malicious* ASes, rather than assigning low reputation to legitimate ASes that have unfortunately been abused.

Our work is motivated by the practical help that an AS reputation system, which accurately identifies malicious ASes, may offer: (1) Network administrators may handle traffic appropriately from ASes that are likely operated by cyber criminals. (2) Upstream providers may use reliable AS reputation in the peering decision process (e.g. charge higher a low reputation customer, or even de-peer early). (3) Law enforcement practitioners may prioritize their investigations and start early monitoring on ASes, which will likely need remediation steps.

The main intuition behind *ASwatch* is that malicious ASes may manipulate the Internet routing system, in ways that legitimate ASes do not, in an attempt to evade current detection and remediation efforts. For example, malicious ASes "rewire" with one another, forming groups of ASes, often for a relatively short period of time [49]. Only one AS from the group connects to a legitimate upstream provider, to ensure connectivity and protection for the group. Alternatively, they may connect directly to a legitimate upstream provider, in which case they may need to change upstream providers frequently, to avoid being de-peered and isolated from the rest of the internet. Changing providers is necessary because a legitimate upstream provider typically responds (albeit often slowly) to repeated abuse complaints concerning its customer ASes. Another example is that a malicious AS may advertise and use small blocks of its IP address space, so that as soon as one small block of IP addresses is blocked or blacklisted, a new block can be advertised and used to support malicious activities. To capture this intuition, we derive a collection of *control-plane features* that is evident solely from BGP traffic observed via Routeviews [86]. We then incorporate these features into a supervised learning algorithm, that automatically distinguishes malicious ASes from legitimate ones.

We offer the following contributions:

- We present *ASwatch*, an AS reputation system that aims to identify malicious ASes by monitoring their *control plane behavior*.
- We identify three families of features that aim to capture different aspects of the "agile" control plane behavior typical of malicious ASes. (1) *AS rewiring* captures aggressive changes in AS connectivity; (2) *BGP routing dynamics* capture routing behavior that may reflect criminal illicit operations; and (3) *Fragmentation and churn of the advertised IP address space* capture the partition and rotation of the advertised IP address space.
- We evaluate *ASwatch* on real cases of malicious ASes. We collect *ground truth* information about numerous malicious and legitimate ASes, and we show that *ASwatch* can achieve high true positive rates with reasonably low false positives. We evaluate our statistical features and find that the rewiring features are the most important.
- We compare the performance of *ASwatch* with BGP Ranking, a state-of-the-art AS reputation system that relies on data-plane information. Our analysis over nearly three years shows that *ASwatch* detects about 72% of the malicious ASes that were observable over this time period, whereas BGP Ranking detects only about 34%.

The rest of the chapter is organized as follows. Section 5.2 offers background information about bulletproof hosting ASes. Section 5.3 describes the features we devised and an overview of our system. Section 5.4 discusses the evaluation of the system. Section 5.5 discusses various limitations of our work and Section 5.6 concludes.

5.2 Background

In this section, we describe more precisely the differences between malicious (bulletproof hosting) and legitimate ASes. We provide background information, with an emphasis on characteristics that are common across most confirmed cases of malicious ASes. We also discuss how malicious ASes tend to connect with one another, and how some ISPs (some of which are themselves malicious) provide these ASes with upstream connectivity and protection. To illustrate this behavior, we explore a case study that shows how malicious ASes may be established and "rewired" in an attempt to evade current detection and takedown efforts.

Malicious vs. Legitimate ASes: We call an AS *malicious*, if it is managed and operated by cyber-criminals, and if its main purpose is to support illicit network activities (*e.g.*, phishing, malware distribution, botnets). In contrast, we refer to an AS as *legitimate*, if its main purpose is to provide legitimate Internet services. In some cases, a legitimate AS's IP address space may be abused by cyber-criminals to host malicious activities (*e.g.*, sending spam, hosting a botnet command-and-control server). Such abuse is distinct from those cases where cyber-criminals operate and manage the AS. *ASwatch* focuses on distinguishing between malicious and legitimate ASes; we aim to label *legitimate but abused* ASes as legitimate. Our approach is thus a significant departure from existing data-plane based AS reputation systems, which are limited to computing reputation by primarily focusing on data-plane abuse, rather than establishing if an AS is actually malicious.

Malicious AS Relationships: Bulletproof hosting ASes provide cyber-criminals with a safe environment to operate. Sometimes, malicious ASes form business relationships with one another to ensure upstream connectivity and protection. For example, they may connect to upstream providers that are themselves operated in part with criminal intent. In turn, these upstream ASes connect to legitimate ISPs, effectively providing cover for the bulletproof hosting ASes [2]. These "masking" upstream providers may not be actively engaged in cyber-criminal activity themselves (as observed from the data-plane). Consequently, network operators at legitimate ISPs may be unaware of the partnership among these "shady" upstream providers and bulletproof hosting ASes, making detection and remediation efforts more difficult.

Efforts to take down bulletproof hosting ASes have been ongoing since at least 2007, when upstream ISPs of the Russian Business Network (RBN) refused to route its traffic [53]. Many organizations track rogue ASes and report tens to hundreds of new rogue

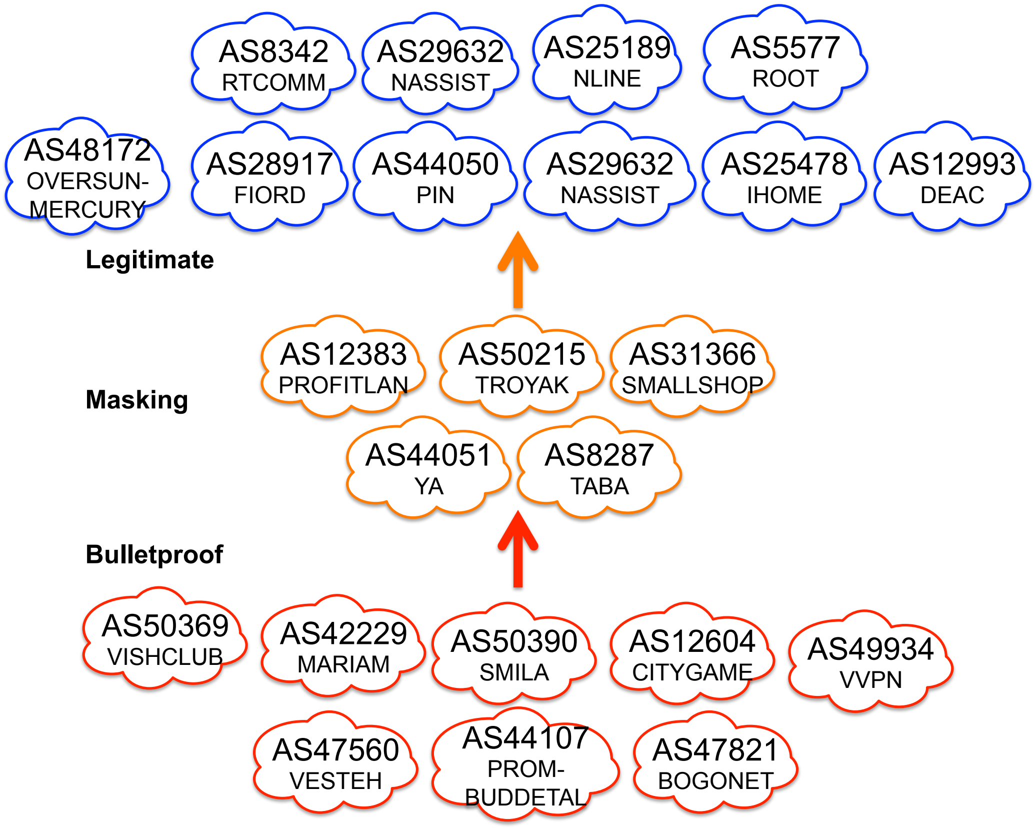
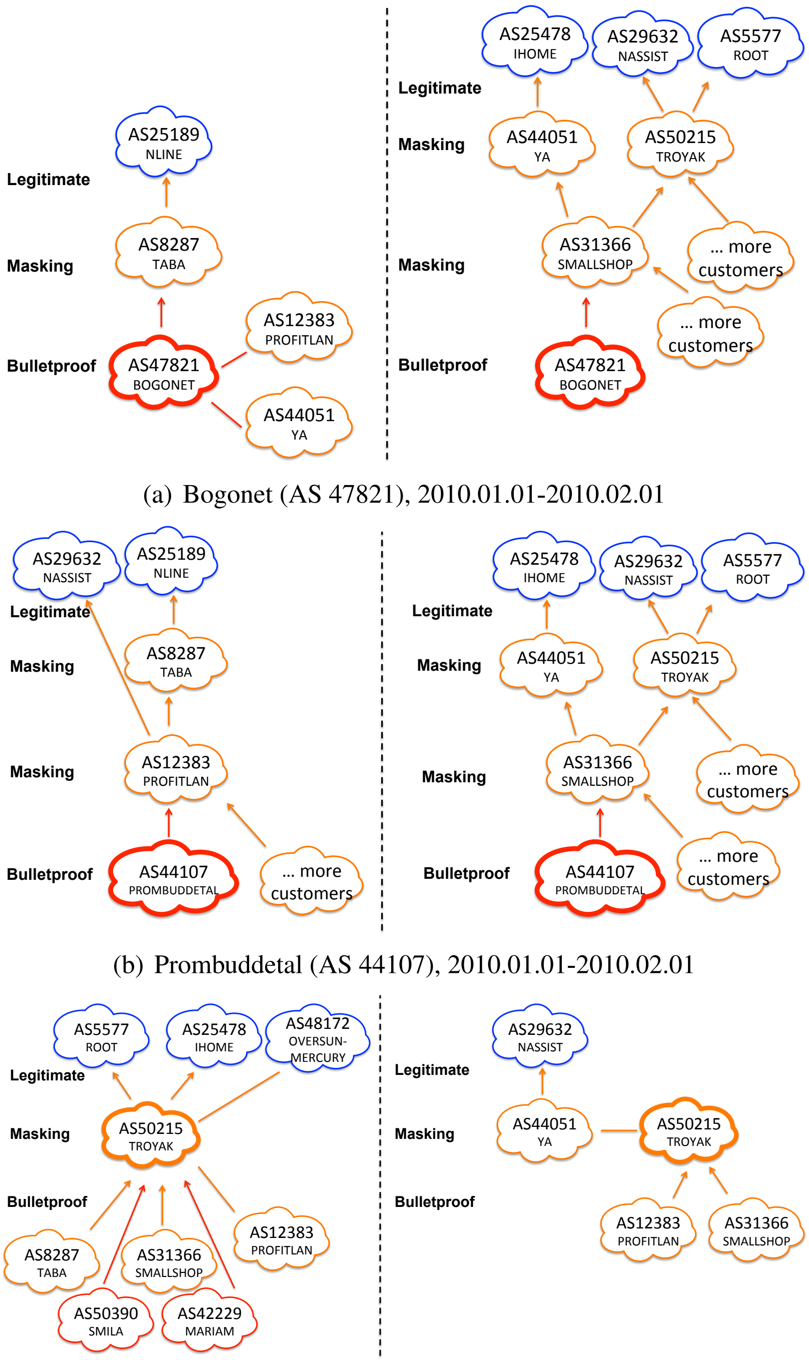


Figure 18: The AS-TROYAK infrastructure (malicious ASes identified by `blogs.rsa.com`). The core of the infrastructure comprises eight bulletproof networks, which connect to legitimate ASes via a set of intermediate ”masking” providers.

ASes every year [36]. Takedown efforts often result in a malicious AS moving to new upstream ISPs; for example, RBN now operates on many different ISP networks.

Case Study - Behavior of Malicious ASes: Figure 18 shows an example of a real network of eight bulletproof hosting ASes that connect to legitimate ASes via a set of intermediate “masking” providers. Notice that while we label the malicious ASes in this case study, based on ground truth provided by `blogs.rsa.com`, we independently derive and analyze the relationships between the ASes from routing information. At the time they were reported by `blogs.rsa.com` (March 2010), the eight bulletproof ASes hosted a range of malware, including Zeus Trojans, RockPhish JabberZeus servers, and Gozi Trojan servers. We chose this as a case study because it represents one of the most well documented cases of known bulletproof hosting ASes, and is representative of other less well known incidents.

The bulletproof hosting ASes switched between five upstream providers, which served



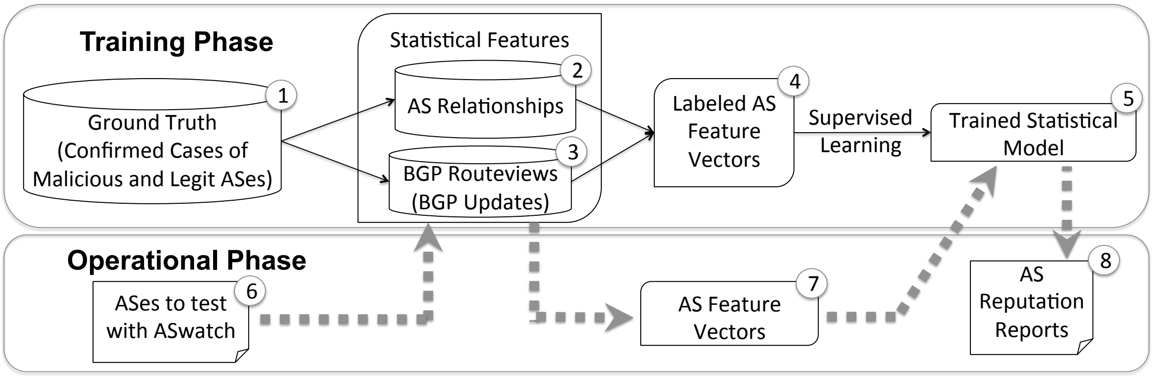
(a) Bogonet (AS 47821), 2010.01.01-2010.02.01

(b) Prombuddetal (AS 44107), 2010.01.01-2010.02.01

(c) Troyak (AS 50215), 2010.02.01-2010.04.01

Figure 19: Connectivity snapshots of three cases of ASes which are operated by cyber-criminals. All connected to a "masking" upstream provider. Directed edges represent customer-provider relationships; undirected edges represent peering relationships.

as intermediaries to connect to the legitimate ASes. In turn, the upstream "masking" providers were customers of nine different legitimate ISPs.



Training Phase
Statistical Features
1
Ground Truth (Confirmed Cases of Malicious and Legit ASes)
2
AS Relationships
3
BGP Routeviews (BGP Updates)
4
Labeled AS Feature Vectors
Supervised Learning
5
Trained Statistical Model
Operational Phase
6
ASes to test with ASwatch
7
AS Feature Vectors
8
AS Reputation Reports

Figure 20: *ASwatch system architecture.*

To understand how malicious ASes form business relationships and how these relationships evolve over time, we tracked the upstream and downstream connectivity of the malicious ASes, as shown in Figures 18 and 19 (the figures show an activity period from January to April 2010; the malicious ASes went offline in March 2010).

We tracked the connectivity of one "masking" AS, Troyak (AS 50215), and two bulletproof hosting ASes, Bogonet (AS 47821) and Prombuddetal (AS 44107), that belong to the Troyak infrastructure. To track their upstream and downstream connectivity, we used a publicly available dataset from CAIDA, which provides snapshots of the AS graph, annotated with business relationships [65]. Figure 19 shows snapshots of the connectivity for the reported ASes.

All of these malicious ASes connected to a "masking" upstream provider, thus avoiding direct connectivity with legitimate ISPs, and also they change their connectivity between one another. For example, before takedown, Troyak had three upstream providers: Root, Ihome, and Oversun-Mercury. After the blog report on March 2010, Troyak lost *all* of its upstream providers and relied on a peering relationship with Ya for connectivity. After April 2010, Troyak and its customers went offline. Bogonet switched from Taba to Smallshop, and Prombuddetal switched from Profitlan to Smallshop, before going offline.

5.3 ASwatch

ASwatch monitors globally visible BGP routing activity and AS relationships, to determine which ASes exhibit *control plane behavior* typical of malicious ASes. Because of the nature of their operations (criminal activity) and their need to fend off detection and possible take-down efforts, malicious ASes tend to exhibit control-plane behavior that is different from that of legitimate ASes. We now discuss how *ASwatch* works, including a detailed description of the features we used to differentiate between malicious and legitimate ASes, and our intuition for choosing each feature.

5.3.1 System Overview

Figure 20 presents an overview of *ASwatch*. The system has a training phase (Section 5.3.3.1) and an operational phase (Section 5.3.3.2). During the training phase, *ASwatch* learns the control-plane behavior of malicious and legitimate ASes. We provide the system with ① a list of known malicious and legitimate ASes (Section 5.4.1 describes this dataset). *ASwatch* tracks the control-plane behavior of the legitimate and malicious ASes over time using two sources of information: ② business relationships between ASes, and ③ BGP updates (from RouteViews). *ASwatch* then computes statistical features (Section 5.3.2 describes this process) from the previous inputs. Each AS is represented by a feature vector based on these statistical features ④. *ASwatch* uses these labeled feature vectors and a supervised learning algorithm to ⑤ train a statistical model. During the operational phase, we provide *ASwatch* with a list of new (not yet labeled) ASes ⑥ to be classified as legitimate or malicious using the same statistical features over the given time period. Then, *ASwatch* ⑦ computes the new AS feature vectors and ⑤ tests them against the previously trained statistical model. Finally, ⑧ the system assigns a reputation score to each AS.

5.3.2 Statistical Features

In this section, we describe the features we compute and the intuition for choosing them. Table 12 gives an overview of our feature families, and the most important group of features for each family. Given an AS, A, and time window, T, *ASwatch* monitors A's control-plane behavior and translates it into a feature vector consisting of three groups of features: rewiring activity, IP fragmentation and churn, and BGP routing dynamics.

Some of the behavioral characteristics we measure can be naturally described by a probability distribution, rather than a single numerical feature. In these cases, to capture the behavioral characteristics in a way that is more suitable for input to a statistical classifier, we translate each probability distribution into three numerical features that approximately describe the *shape* of the distribution. Specifically, we compute its 5th percentile, 95th percentile, and median. In the following, we refer to such features as *distribution characteristics*. We include these three values as features in the overall feature vector, and repeat this process for all behavioral characteristics that can be described as a probability distribution.

Notice that even though more values may more accurately summarize a distribution's shape, such a representation would significantly increase the overall size of the feature vector used to describe an AS. For this reason, we chose to only use three representative values, which we found to work well in practice.

We now explain in detail the features that *ASwatch* uses to establish AS reputation and motivate how we selected them.

5.3.2.1 Rewiring Activity

This group of features aims to capture the changes in A's connectivity. Our intuition is that malicious ASes have different connectivity behavior than legitimate ASes, because they tend to: (1) change providers more frequently to make detection and remediation more difficult; (2) connect with less popular providers, which may have less strict security procedures and may respond less promptly to abuse complaints, (3) have longer periods of downtime, possibly due to short-duration contracts or even de-peering from a legitimate upstream provider. In contrast, legitimate ASes tend to change their connectivity less frequently, typically due to business considerations (*e.g.*, a less expensive contract with a new provider).

To capture rewiring activity, *ASwatch* tracks changes to AS relationships (Step 2 in Figure 20). We use periodic snapshots of historic AS relationships, with one snapshot per month (Section 5.4.1 describes the data sets in more detail). A snapshot S_i contains the AS links annotated with the type of relationships, as observed at a given time t_i (*e.g.*, one snapshot is produced on the first day of each month).

AS presence and overall activity. Let A be the AS for which we want to compute our features. Given a sequence of N consecutive snapshots $\{S_i\}_{i=1}^{N}$, we capture the *presence* of an AS by measuring the total number of snapshots, C, and the maximum number of contiguous snapshots, M, in which A was present, the fraction C/N, and M/N (four features in total). To capture the overall activity of A, we measure the distribution (over time) of the number of customers, providers, and peers A links with for each snapshot. To summarize each of these distributions, we extract the distribution characteristics (5th percentile, 95th percentile, and median), as described earlier. This yields a total of nine features (three for each of the three types of AS relationships). We also count the total number and fraction (*i.e.*, normalized by C) of distinct customers, providers, and peers that A has linked with across all C snapshots when it was present, yielding another six features.

Link stability. We capture the *stability* of different types of relationships that an AS forms over time. For each of the C snapshots where A was present, we track all relationships

Table 12: Overview of *ASwatch* feature families and the most important feature for each family.

Feature Family	Description	Most Important Feature
Rewiring Activity	Changes in AS's connectivity (*e.g.*, frequent change of providers, customers or peers)	Link stability
IP Space Fragmentation & Churn	IP space partitioning in small prefixes & rotation of advertised prefixes	IP space fragmentation
BGP Routing Dynamics	BGP announcements patters (*e.g.*, short prefix announcements)	Prefix reachability

between A and any other AS. Assuming A appeared as an upstream provider for another AS, say A^k, in v out of C snapshots, we compute the fraction $F^k = v/C$. We repeat this for all ASes where A appears as a provider at least once within C snapshots, thus obtaining a distribution of the F^k values. Finally, we summarize this distribution of the F^k values, computing the distribution characteristics as described above. We repeat this process, considering all ASes that appear as the upstream provider for A (*i.e.*, A is their customer), and for all ASes that have peering relationships with A. Overall, we compute nine features that summarize three different distributions (three features for each type of relationship).

Upstream connectivity. We attempt to capture *change* in the set of providers. Assume that from the i-th snapshot S_i we observed a total of M_i upstream providers for A, and call $\{A_i^k\}_{k=1}^{M_i}$ the set of upstream provider ASes. Then, for each pair of contiguous snapshots, S_i and S_{i+1}, we measure the Jaccard similarity coefficient $J_{i,i+1}$ between the sets $\{A_i^k\}$ and $\{A_{i+1}^k\}$. We repeat for all available $(N-1)$ pairs of consecutive snapshots, thus obtaining a distribution of Jaccard similarity coefficients. To summarize this distribution, we compute the distribution characteristics as described above, yielding three features. Figure 21 shows the CDF of the minimum Jaccard similarity, for the malicious and the legitimate ASes. Overall, the legitimate ASes tend to have higher values of the Jaccard similarity metric, which indicates fewer changes in their upstream providers.

Attachment to popular providers. We aim to capture an AS's *preference* for "popular" providers. As previous work has shown [49], malicious ASes tend to connect more often with less prominent providers, which may have less strict security procedures and may respond less promptly to abuse complaints.

We compute the popularity of each provider per snapshot and across all snapshots. To this end, we first empirically derive the distribution of the number of customers per provider. We then consider a provider to be (a) *very popular*, if it belongs to the top 1% of all providers overall; (b) *popular*, if it belongs to the top 5%; (c) *very popular with respect to a snapshot* S_i, if it belongs to the top 1% in S_i, and (d) *popular with respect to a snapshot* S_i, if it belongs to the top 5% in S_i.

We then gather all upstream providers that A has used and compute the fraction of these providers that fall into each of the four categories described above (thus yielding four features). Finally, we compute the fraction of snapshots in which A has linked to at least one provider falling into one of the above categories; we do this for each category, thus obtaining four more features.

We capture the overall rewiring behavior of an AS with a total number of thirty five features.

5.3.2.2 IP Space Fragmentation and Churn

Malicious ASes tend to partition their IP address space into small BGP prefixes and to advertise only some of these prefixes at any given time. One possible explanation for this behavior may be that they attempt to avoid having their entire IP address space blacklisted at once. For example, if a number of IP addresses within a given BGP prefix are detected as hosting malicious activities, a blacklist operator (*e.g.*, Spamhaus [96]) may decide to blacklist the entire prefix where the IP addresses reside. By fragmenting the IP address space and advertising only a subset of their BGP prefixes, the operators of a malicious AS may be able to quickly move malicious activities to a "fresh" space. They perform this maneuver by leveraging not-yet-blacklisted IP addresses within newly advertised prefixes. On the other hand, legitimate ASes tend to consistently advertise their available IP address space in less fragmented prefixes, as they do not need to attempt to evade blacklisting.

IP Space Fragmentation and Churn Features. We attempt to capture IP address *fragmentation* with the following features. Given a snapshot, we group the advertised BGP prefixes into contiguous IP *blocks*. For each, AS we count the number of BGP prefixes and the number of distinct `/8`, `/16`, and `/24` prefixes within each IP block. To capture the *churn* in the advertisement of the IP address space, we proceed as follows. Given a pair of adjacent snapshots for an AS, we measure the Jaccard similarity among the sets of

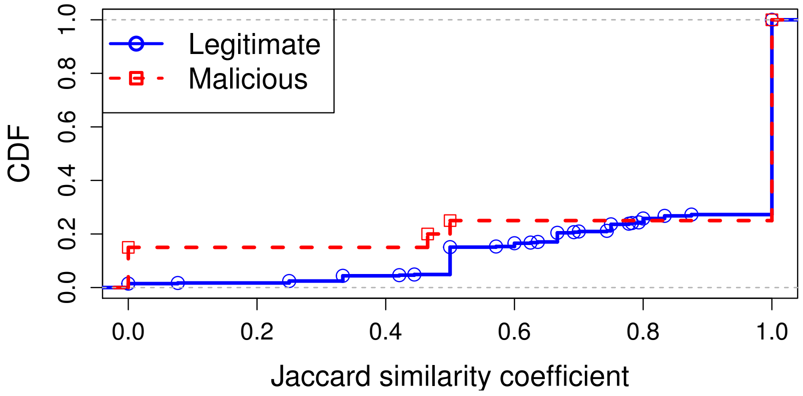


Figure 21: Malicious ASes rewire more frequently. Distribution of the 5th percentile of the Jaccard similarity coefficient between consecutive snapshots of an AS's upstream providers. Higher values indicate fewer changes in upstream connectivity.

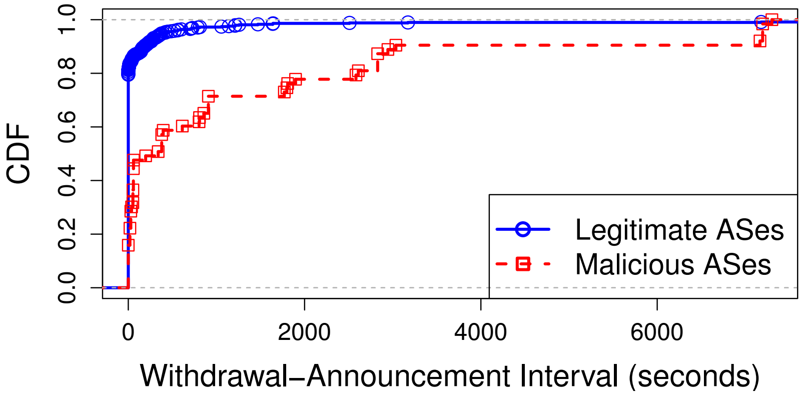


Figure 22: Malicious ASes withdraw prefixes for longer periods. The distribution of the median interval between a prefix withdrawal and re-announcement across 15 contiguous epochs.

BGP prefixes advertised by the AS in the two snapshots. Similarly, we compute the Jaccard index among the sets of /8, /16, and /24 prefixes. We summarize the above four distributions using the *distribution characteristics* that we described earlier, thus obtaining a total of twelve features.

5.3.2.3 BGP Routing Dynamics

These features attempt to capture abnormal BGP announcement and withdrawal patterns. For example, to support aggressive IP address space fragmentation and churn and avoid easy blacklisting, malicious ASes may periodically announce certain prefixes for short periods of time. On the contrary, the pattern of BGP announcements and withdrawals for legitimate ASes is mainly driven by normal network operations (*e.g.*, traffic load balancing, local policy changes), and should thus exhibit BGP routing dynamics that are different to those of malicious ASes.

Prefix reachability. We aim to capture the fraction of time that prefixes advertised by A remain reachable, which we define as *reachability*. First, we measure the time that elapses between an announcement and a withdrawal for every advertised prefix. Given the distribution of these time intervals, we extract the distribution characteristics as described above. Second, we track the time for a prefix to become reachable again after a withdrawal. Third, we measure the inter-arrival time (IAT) between withdrawals, for each of the prefixes that A announces, and compute the IAT distribution. As before, we extract the distribution characteristics for each of the three distributions, yielding a total of nine features. Figure 22 shows the CDF of the median reachability value for the malicious and the legitimate ASes over the course of one day, and over 15 days. Higher values of this feature suggest that malicious ASes tend to re-advertise their prefixes after longer delays.

Topology and policy changes. We track the *topology and policy changes*, defined as in Li *et al.* [63], that are associated with each prefix. We define a *policy change* as follows: after a path to a destination is announced, a second BGP announcement is observed with the same AS path and next-hop, yet one or more of the other attributes (such as MED or community) is different. Similarly, we define a *topology change* event as follows: after a path to a destination is announced, a second announcement follows with an alternate route (implicit withdrawal) or after a route to a destination is explicitly withdrawn, a different route (with different AS path or next-hop attributes) to the same destination is announced (explicit withdrawal).

To capture and summarize the topology and policy changes per AS, we group the prefixes per origin AS (the origin AS appears as the last AS in the AS path). We track the policy change events for each prefix, and we measure the inter-arrival time between the events per prefix. Then, we analyze the collection of inter-arrival times of the policy events for all prefixes advertised by the same AS. For each AS, we form the distribution of such intervals, and we extract the distribution characteristics as described above. We also compute the total number of events and the total number of events divided by the total prefixes advertised by the AS. We repeat this process for the topology change events. We compute a total of ten features.

5.3.3 System Operation

We now describe *ASwatch*'s training and operation.

5.3.3.1 Training Phase

To train the classifier (Steps 6 and 7 in Figure 20), we first prepare a training dataset with labeled feature vectors related to known malicious and legitimate ASes. We start with a ground truth dataset that includes confirmed cases of *malicious* ASes, and legitimate ASes (described in more details is Section 5.4.1).

We compute the statistical features for each labeled AS using two sources of data: BGP announcements and withdrawals from Routeviews [86], and information from a publicly available dataset [65] about the relationships between ASes. We compute the feature vectors over m contiguous epochs (in our experiments, each epoch is one day). More specifically, we maintain a sliding window of size m epochs, which advances one epoch at a time. Using this sliding window, we can compute multiple feature vectors for each AS (one per window). Then, we associate a label to each feature vector, according to the ground truth related to the AS from which a vector was computed.

Finally, to build the statistical classifier, we use the Random Forest (RF) algorithm. We experimented with different algorithms, but we chose RF because it can be trained efficiently and has been shown to perform competitively with respect to other algorithms for a variety of problems [15].

5.3.3.2 Operational Phase

Once the statistical classifier has been trained, *ASwatch* can assign a reputation score to new ASes (*i.e.*, ASes for which no ground truth is yet available). *ASwatch* computes a reputation score for each new AS observed in the BGP messages from Routeviews. Suppose that we want to compute the reputation of an AS, A, over some time period, T. First, we compute A's features (as explained in Section 5.3.2) over period T, using a sliding window procedure as in the training phase. Namely, a feature vector is computed for each window within T. Second, we classify an AS as malicious, if *ASwatch* consistently assigns it a low reputation score for several days in a row.

More specifically, let T_i be the current day of observations, f_{A,T_i} be the corresponding feature vector for A, and $s(f_{A,T_i})$ be the *bad reputation* score output by the classifier at the end of T_i. Also let $W_i = (T_i, T_{i+1}, \ldots, T_{(i+m-1)})$ be a period of m consecutive days. We report A as malicious if: (a) score $s(f_{A,T_i}) > \theta$ for 90% of the days in period W_i, where θ is a predefined threshold that can be learned during the training period; and (b) condition (a) holds for at least l consecutive periods $W_i, W_{i+1}, \ldots, W_{i+l}$.

We note that we have experimented with multiple values for m and l (see Section 5.4.3 for detailed discussion on parameter selection).

5.4 Evaluation

We now describe the data we collected and the setup for our evaluation of *ASwatch*, where we evaluate the system's accuracy. Our results show that *ASwatch* achieves a high detection rate for a reasonably low false positive rate, can detect malicious ASes before they are publicly reported by others, and can complement existing AS reputation systems that rely solely on data-plane observations. Furthermore, we find that *ASwatch* detects nearly double the fraction of confirmed cases of malicious ASes compared to BGP Ranking, a data-plane based AS reputation system.

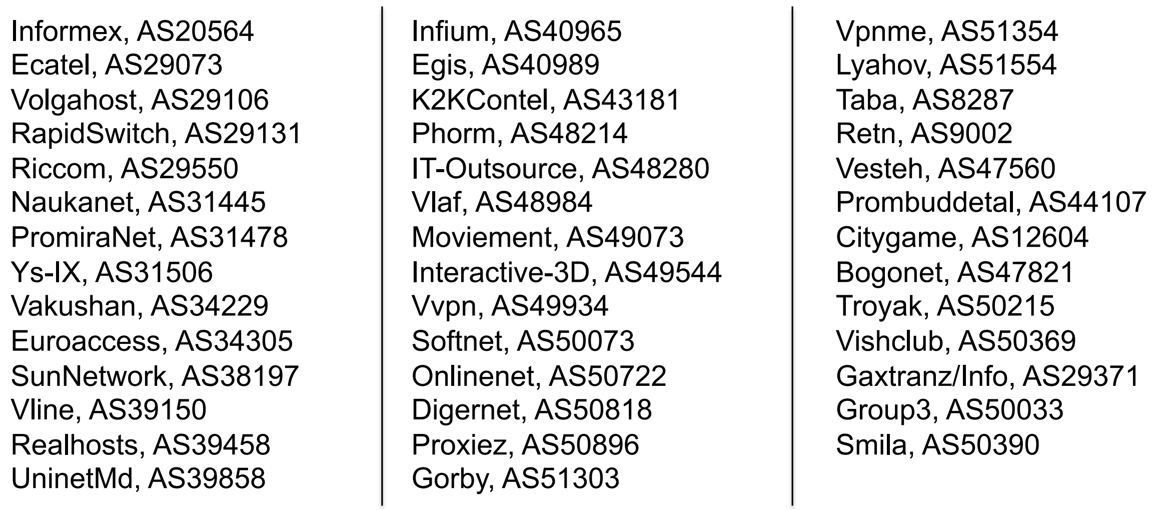
5.4.1 Data

Labeling malicious ASes. Collecting reliable ground truth about malicious ASes is extremely challenging, due to the utter lack of public information available about such cases. Nonetheless, through extensive manual search and review efforts, we managed to collect a set of ASes for which there exists publicly available evidence of malicious behavior. For example, we identified a reasonable set of malicious ASes that were at some point seized by law enforcement or disconnected by other network operators.

To obtain our dataset of malicious ASes, we searched through websites that are operated by cyber-security professionals (*e.g.*, `www.abuse.ch`, `blogs.rsa.com` [1,2,21,34,37, 55]) and carefully reviewed articles about ASes known to be operated by cyber-criminals.

We observed the following common characteristics across all articles and blog reports we considered: (1) the reported ASes hosted a variety of cyber-criminal activities (*e.g.*, botnet C&C hosting, malware domains, phishing), (2) several ASes were associated with each other, either directly (*e.g.*, customer-provider relationship) or indirectly (*e.g.*, they shared the same upstream provider), (3) the operators of these ASes were uncooperative and unresponsive (*e.g.*, would not respond to abuse complaints or attempts by other AS operators to communicate with them), (4) some ASes were prosecuted by law enforcement and taken down, (5) many of these disappeared only for a relatively short time before resurfacing. From each blog report, we extracted the ASes involved and the dates when they were active. Overall, we collected forty one known malicious ASes. We provide our list of ASes in Figure 23.

Labeling legitimate ASes. To collect a set of legitimate ASes, we proceeded as follows. Every day for one year, we collected the list of top one million domain names from `alexa.com`. For each of these domains, we calculated the average daily ranking; we selected the domain names that had an average ranking above 10,000. In other words, we



Informex, AS20564
Ecatel, AS29073
Volgahost, AS29106
RapidSwitch, AS29131
Riccom, AS29550
Naukanet, AS31445
PromiraNet, AS31478
Ys-IX, AS31506
Vakushan, AS34229
Euroaccess, AS34305
SunNetwork, AS38197
Vline, AS39150
Realhosts, AS39458
UninetMd, AS39858

Infium, AS40965
Egis, AS40989
K2KContel, AS43181
Phorm, AS48214
IT-Outsource, AS48280
Vlaf, AS48984
Moviement, AS49073
Interactive-3D, AS49544
Vvpn, AS49934
Softnet, AS50073
Onlinenet, AS50722
Digernet, AS50818
Proxiez, AS50896
Gorby, AS51303

Vpnme, AS51354
Lyahov, AS51554
Taba, AS8287
Retn, AS9002
Vesteh, AS47560
Prombuddetal, AS44107
Citygame, AS12604
Bogonet, AS47821
Troyak, AS50215
Vishclub, AS50369
Gaxtranz/Info, AS29371
Group3, AS50033
Smila, AS50390

Figure 23: Malicious ASes we collected from blogs.

selected only those domains that were consistently very popular. Finally, we mapped each domain name to its resolved IP addresses and mapped those IP addresses to the AS that hosted them. Overall, we collected a total of 389 ASes, which we label as *legitimate*.

Although we cannot be absolutely certain that our labeling of legitimate ASes contains no noise, we rely on two reasonable assumptions. First, we assume that websites that are consistently popular are unlikely to be offering malicious services. Intuitively, a malicious site that becomes highly popular would also have a high number of victims, and would rapidly attract attention for take-down. As a result, the site would be quickly blocked or taken down and would thus not remain consistently popular. Second, we assume that the administrators of the most popular websites are unlikely to host their services within malicious ASes. Intuitively, if they relied on malicious ASes, they would risk damaging their own reputation, not to mention extended downtimes if the hosting ASes were taken down due to abuse complaints.

Finally, to ensure that our set of legitimate ASes consists of ASes that are similar in size to the malicious ASes, we keep only those legitimate ASes that have no customers, or whose customers are all stub ASes.

AS rewiring and relationships data (CAIDA). To track how malicious ASes change their connectivity, we use a publicly available dataset that reports AS business relationships. The dataset reports one snapshot of the AS graph per month, from 1998 to 2013.

Luckie *et al.* [65] provide an AS graph built by inferring business relationships among ASes, based on AS customer cones. Although this dataset has its own limitations (see

Section 5.5), it provides a reasonably accurate view of AS relationships, allowing us to estimate our rewiring features that we presented in Section 5.3.2.

BGP routing dynamics (Routeviews). To further capture the control-plane behavior of malicious and legitimate ASes, we monitored the BGP messages that originate from these ASes using the Routeviews dataset. We use this dataset to measure both the dynamics of BGP updates and the IP fragmentation and churn features.

5.4.2 Experiment Setup

In the following section, we describe the training and the evaluation of our system. The training period extends from January 2010 to March 2010, while the evaluation experiments extend from January 2011 to December 2013.

Computing AS feature vectors. Given a period of time (*i.e.*, m contiguous epochs) over which we want to capture the behavior of an AS, we construct the AS feature vector as follows: (1) *Rewiring activity*: We compute the rewiring features over the most recent k snapshots of the AS relationships dataset, prior to the period of interest. Our source of AS relationships provides only one snapshot per month. Given this limitation, we select a reasonable number of snapshots to capture the most recent rewiring activity of an AS. For our experiments we set $k = 4$ (see Section 5.4.3 on parameter selection); (2) *BGP routing activity*: To compute BGP routing dynamics features, IP address space fragmentation and churn, we collect the BGP announcements and withdrawals originating from the AS during the period of interest. We note that BGP Routeviews offers a large number of monitors. Our pilot experiments over a number of different monitors indicated that changing the monitor selection did not significantly affect the overall performance of our classifier. Therefore, to compute our routing activity features, we select one monitor and consistently use it, throughout all the experiments.

Training the AS reputation model. Because our data is derived from cases of malicious ASes publicly reported by others, we rely on the report dates for an approximate period of time when the ASes were likely to be actively used by the attackers. For example, if an AS was reported as malicious on a given day d, we assume the AS was operated by criminals for at least a few months before d (in fact, it typically takes time for security operators to detect, track, confirm, and take down a malicious AS). For the purpose of computing our labeled feature vectors and training our system, we selected a period of time with the highest concentration of active malicious ASes. This period extends from January–March

2010, during which we identified a total of 15 active malicious ASes. Even though this period may appear somewhat dated, it allows us to capture the agile behavior of several known malicious ASes within one consistent time frame, enabling a "clean" evaluation setup. Our evaluation detects a large fraction of malicious ASes that we have observed over a longer, more recent time period (2011–2013). In the future, we plan to investigate more sources of ground truth and identify additional periods of time that can be used to train our model (see Section 5.5 for further discussion).

Performing cross-validation tests. During the three-month training period mentioned above, we maintain a sliding window of fifteen contiguous days (epochs), sliding the window one day at a time (*i.e.*, two consecutive windows overlap by 14 days). For each sliding window, we compute the feature vector for each AS and we perform three-fold cross-validation as follows: (1) We separate the ASes into three subsets, using two subsets to train our reputation model, and one for testing. (2) For each training subset, we balance the two classes by oversampling from the underrepresented class. After balancing, the number of feature vectors of the two classes are equal. (3) We train the model using a Random Forest classifier [15]. (4) Finally, we test all feature vectors that belong to the third fold against the model, as we described in Section 5.3.3. Cross-validation yields the scores from the testing phase and the true label for each AS feature vector. We plot the receiver operating characteristic (ROC), which illustrates the performance of the classifier for different values of the detection threshold. Because we perform our testing once for each sliding window, we plot a similar ROC for each sliding window. The results are reported in Section 5.4.3.

Evaluating ASwatch across a nearly three-year period. After the cross-validation experiments, we use our model to test new ASes whose BGP behavior was observed outside the training period over nearly three years, from 2011 to 2013. We perform this evaluation for two reasons: a) to test how well *ASwatch* performs to detect new malicious ASes (outside of the training period), and b) to compare the performance of *ASwatch* with other AS reputation systems (*e.g.*, BGP Ranking) over an extended period of time. For each (previously unseen) AS we want to test against *ASwatch*, we classify it as malicious if it has *multiple* feature vectors that are *consistently* assigned a "bad reputation" score (see Section 5.3.3). The results are reported in Section 5.4.3.

5.4.3 Results

How accurate is ASwatch? Evaluation with cross-validation: Figure 24 shows the detection and false positive rates for one cross-validation run. The detection rate and false

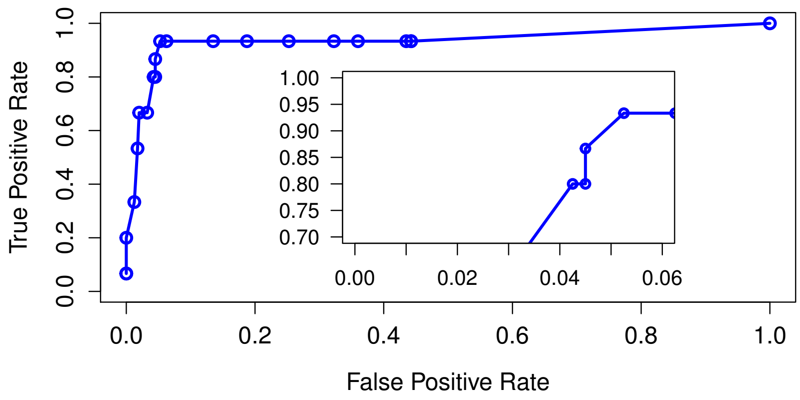


Figure 24: The cross-validation detection and false positive rates of *ASwatch*.

positives reported on the ROC correspond to the fraction of *malicious* feature vectors that are correctly classified and *legitimate* feature vectors that are incorrectly classified, respectively. As shown by the ROC curve, *ASwatch* can achieve a detection rate of 93.33% (correctly classifying 14 out of 15 ASes as malicious), with a reasonably low false positive rate of 5.25% (20 falsely detected ASes). In practice, we believe this false positive rate is manageable, as it represents 20 falsely detected ASes over a three-month period, or one every few days. Although this false positive rate is clearly too high to automate critical decisions such as take-down efforts, *ASwatch* can still be used to significantly narrow down the set of ASes for further investigation considerably, and can thus help both law enforcement focus their investigation efforts, and network administrators make decisions on who to peer with or which abuse complaints to prioritize.

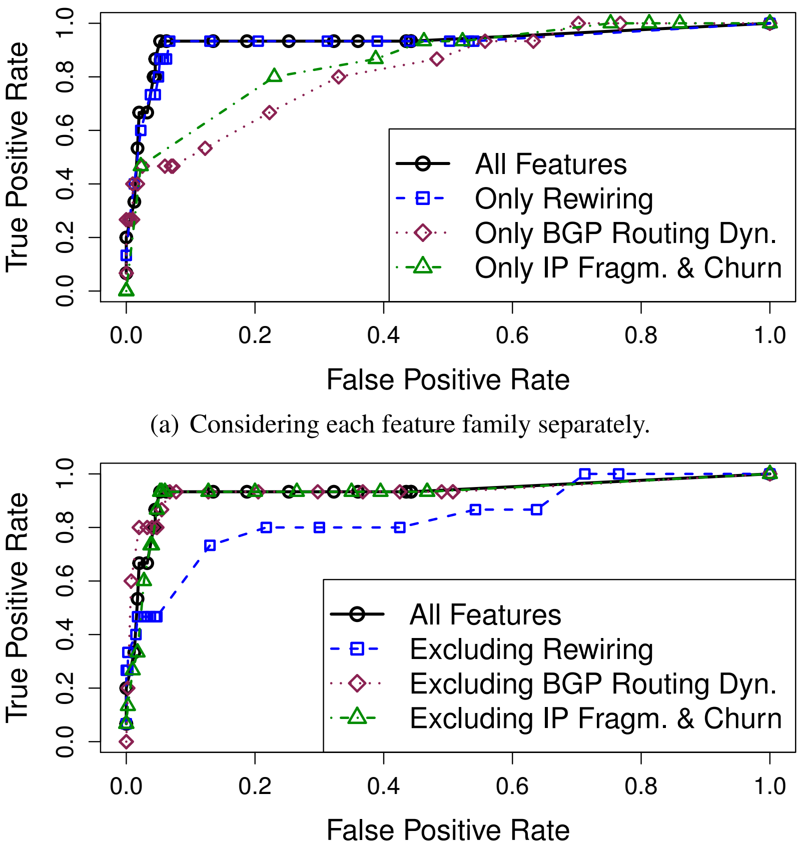
Evaluation outside the training period, over nearly three years: As described in Section 5.4.1, we use our model to test new ASes observed after the training period, over nearly three years, from 2011 to 2013. It is important to notice that, from a control-plane point of view, malicious ASes may not always be behaving maliciously across a three year period of time. Our ground truth information does not allow us to distinguish between the periods of activity and periods of “dormancy”. Nonetheless, over time an AS operated by cybercriminals will likely behave in a noticeably different way, compared to legitimate ASes, allowing us to detect it. Figure 27 shows the *cumulative* true positive rate of detected ASes over the testing period. At the end of this nearly three years period, *ASwatch* reached a true positive rate of 72% (21 out of 29 ASes correctly flagged as malicious).

To compute the false positives, for each month we count the number of distinct ASes that were detected as malicious. The false positives reach at most ten to fifteen ASes per month, which we believe is a manageable number, because these cases can be further reviewed by network operators and law enforcement. For instance, the upstream providers of an AS that is flagged as malicious by *ASwatch* may take a closer look at its customer's activities and time-to-response for abuse complaints. Furthermore, the output of *ASwatch* could be combined with the reputation score assigned by existing data-plane based AS reputation systems. The intuition is that if an AS behaves maliciously both at the control plane (as detected by *ASwatch*) and at the data plane (as detected by existing reputation systems), it is more likely that the AS is in fact operated by cyber-criminals.

How early can ASwatch detect malicious ASes before they are widely noticed? We want to evaluate if *ASwatch* can detect malicious ASes *before* they were reported by blog articles. For each of the 14 malicious ASes that *ASwatch* detected during the cross-validation experiments discussed earlier, we took note of the day that *ASwatch* first detected the malicious AS, and we measured the number of days between the time *ASwatch* detected the AS and the day the blog story was published. About 85% of the detected malicious ASes were detected by *ASwatch* 50 to 60 days before their story became public.

Which features are the most important? We evaluate the strength of each family of features that *ASwatch* uses. To understand which features are most important for *ASwatch*, we evaluate each family's contribution to the overall true and false positive rates. In particular, we want to study the effect of each family of features on the detection of malicious ASes, independently from the other families, and the effect of each family on the false positives when those features are excluded. To this end, we repeated the experiment described previously by excluding one family of features at a time. We repeated the experiment four times, once for each family of features, and we calculated the overall detection and false positive rates. Figure 25 shows the results of our experiments, which suggest that the rewiring features are very important, because excluding them significantly lowers the detection rate. The BGP dynamics and IP address space churn and fragmentation features help reduce the false positives slightly (the "Only Rewiring" ROC in Figure 25(a) is slightly shifted to the right). We followed a similar procedure to identify which features are most important for each family of features. Table 12 shows the most important features for each family.

Is ASwatch sensitive to parameter tuning? As explained in Sections 5.3.3.2, 5.4.2 we use the following parameters to classify an AS as malicious: (1) *feature vectors window size*: we compute feature vectors for an AS for a window of m consecutive days (one feature vector per day), and we repeat the feature computation over l consecutive sliding



(a) Considering each feature family separately.

(b) Excluding one feature family at a time.

Figure 25: Relative importance of different types of features. The rewiring features contribute the most to the overall detection rate; other features contribute to lower false positive rates.

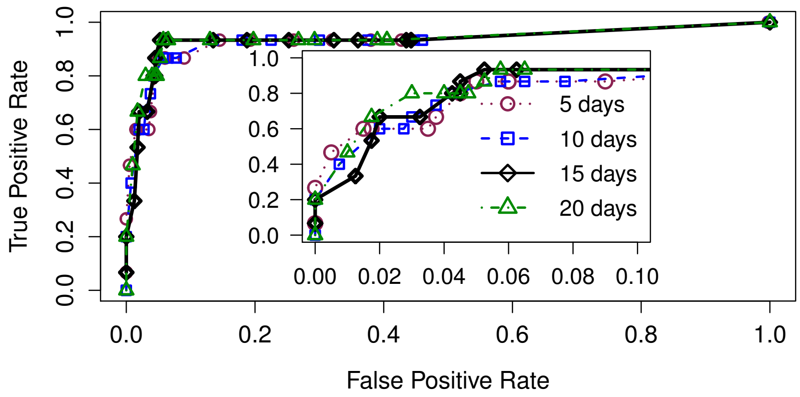


Figure 26: The detection and false positive rates for *ASwatch*, if we vary the size of the sliding window. Our experiments show that the performance is not greatly affected.

windows of size m. (2) *recent snapshots*: we compute the rewiring features for an AS over the k most recent snapshots of AS relationships.

To tune our parameters, we performed several pilot experiments, rather than an exhaustive search over the entire parameter space. Our pilot experiments showed that *ASwatch*'s performance is robust to both parameters m and l. Due to space limitations, we only show our experiments for the parameter m. Figure 26 shows the performance for window sizes of 5, 10, 15, and 20 days. Our results show that the accuracy of *ASwatch* is not overly sensitive to the choice of window size m. The ROC plots in Figure 26 show that $m = 15$ gives a higher true positive rate with a reasonable false positive rate. We therefore set $m = 15$. Using a similar approach, we set $l = 5$. We classify an AS as malicious, if it scores lower than the detection threshold over five consecutive periods of 15 days.

After we have selected parameters m and l, we proceed to set parameter k. Suppose that we want to compute the reputation of an AS A, over period T. Then, parameter k is the number of most recent AS relationship snapshots, prior to T, over which we compute the rewiring features for A (notice that our AS relationships dataset consists of one snapshot per month, as mentioned in Section 5.4.1). In other words, k denotes "how much" history we consider, to capture the rewiring behavior for A. Ideally, we want to accurately capture A's rewiring behavior while using a small number of snapshots. We performed experiments using different values of k (*i.e.*, 1, 2, 3, 4). We then selected $k = 4$, because further increasing its value did not produce a significant increase in classification accuracy.

5.4.4 Comparison to BGP Ranking

We now compare *ASwatch* with BGP Ranking. In contrast to *ASwatch*, BGP Ranking is an AS reputation system based on *data-plane* features (*e.g.*, observations of attack traffic enabled by machines hosted within an AS). Clearly, BGP Ranking is an AS reputation system that is designed differently from *ASwatch*, because it aims to report ASes that are most heavily abused by cyber-criminals, but not necessarily operated by cyber-criminals. We compare the two systems for two reasons: (1) to test how many of the malicious ASes that are operated by cyber-criminals show enough data-plane evidence of maliciousness and get detected by existing data-plane based AS reputation systems; and (2) to evaluate whether the control-plane based approach can effectively complement data-plane based AS reputation systems.

Results summary. We found that *ASwatch* detected 72% of our set of malicious ASes over a three year period, and BGP Ranking detected about 34%. Both systems reported the same rate of false positives (on average 2.5% per month, which is ten to fifteen ASes per month). Combining the two systems we were able to detect only 14% of the malicious ASes, but we were able to reduce the false positives to 0.08% per month (12 ASes in total across the three year period).

BGP Ranking reports. BGP Ranking [13] has been making its AS reputation scores publicly available since 2011, along with a description of the approach used to compute the scores. BGP Ranking currently has information for a total of 14k ASes, and they announce a daily list of the worst 100 ASes by reputation score. The BGP Ranking score has a minimum value of 1 (which indicates that the AS hosts benign activity) but no maximum value (the more malicious traffic hosted by the AS, the higher the score).

Using our list of confirmed cases of malicious ASes (Section 5.4.1), we checked which ASes are visible from BGP Routeviews starting from 2011. We found a total of 29 ASes. We chose to check which ASes are active since January 2011, because this is the oldest date for which BGP Ranking has data available. Then, we tracked these ASes until November 2013, because the historic AS relationships dataset from CAIDA has a gap from November 2013 to August 2014. Therefore, we collected the historical scores for each active known malicious AS from BGP Ranking, from January 2011 until the end of 2013.

ASwatch setup. Using *ASwatch*, we generate the feature vectors for each AS in our list, starting from January 2011 until November 2013. To generate the feature vectors, we follow the same procedure as described in Section 5.3.3.2. We train *ASwatch* as previously described (on training data collected in 2010) and test the ASes observed from 2011 to 2013 against the model.

Comparing BGP Ranking with ASwatch. As mentioned earlier, BGP Ranking is not a detection system *per se*, in that it aims to report ASes that host a high concentration of malicious activities, and does not focus on distinguishing between abused ASes and ASes that are instead owned and operated by cyber-criminals. Nonetheless, for the sake of comparison it is possible to obtain a detection system by setting a threshold on the score output by BGP Ranking. BGP Ranking publishes the set of "worst" 100 ASes and their scores, which are updated daily (to obtain the historic scores for any other non-top-100 AS, one has to make explicit queries through the web portal). It also reports the average AS score per country or region, and ranks the countries that host the ASes with the lowest reputation. The four top ("worst") countries are Russia, Ukraine, Hong Kong, and the US. Using the above information we consider five distinct detection thresholds as follows: (1) average score for ASes in Russia (BGP Ranking Russia cut-off), (2) average score for

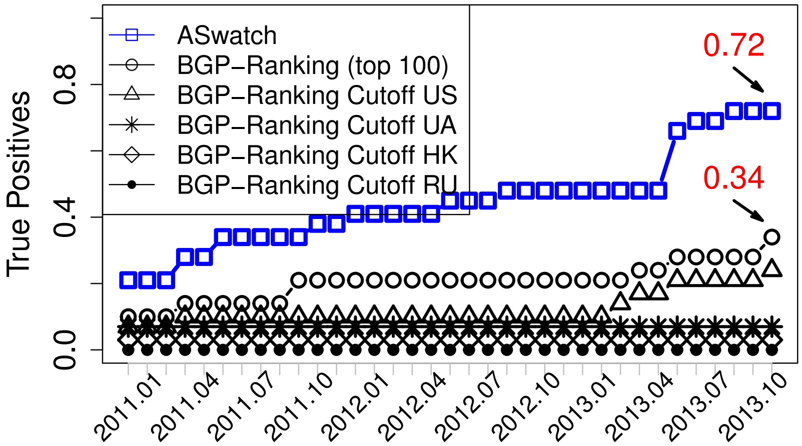


Figure 27: True positive rates for *ASwatch* and BGP Ranking. Accumulation of detected ASes over nearly three years.

ASes in Ukraine (BGP Ranking Ukraine cut-off), (3) average score for Hong Kong (BGP Ranking Hong Kong cut-off), and (4) average score for ASes in the US (BGP Ranking US cut-off). We also set a threshold based on the average score of the 100th worst AS (BGP Ranking top 100) collected from the daily reports. Figure 27 shows the detection results using these thresholds.

We then compared BGP Ranking's detection with that of *ASwatch*. Figure 27 shows the fraction of ASes that *ASwatch* and BGP Ranking detected. We show the cumulative fraction of detected ASes, from January 2011 to November 2013. At the end of the 35-month period, *ASwatch* detected about 72% of the set of ASes we tracked, while BGP Ranking detected about 34%. *We found that 72% of the malicious ASes were detected by monitoring their control-plane behavior, but only 34% of the malicious ASes showed enough data-plane activity to be detected by BGP Ranking.* BGP Ranking may have only limited visibility of malicious activities in the data plane across the entire Internet, and thus may completely miss the malicious activities of certain ASes. Naturally, it is challenging to deploy a large number of sensors dedicated to detecting malicious network communications over the entire Internet. On the other hand, *ASwatch* monitors BGP behavior, and may therefore compensate the limited visibility of data-plane based approaches.

We also compared the false positive rates of BGP Ranking and *ASwatch*. Our motivation is to see if the false positives are manageable within a reasonable period of time (*e.g.*one month). We collected the *ASwatch* scores and the BGP Ranking scores for our set of legitimate ASes (see Section 5.4.1). For each system, we counted the number of legitimate ASes that *ASwatch* detected per month. We found that both systems produce only ten to fifteen false positives per month on average over the total of 389 known legitimate ASes in our dataset. As we have mentioned earlier, BGP Ranking is designed differently from *ASwatch*. Although the rate we calculated does not represent the actual false positive rate for BGP ranking, it does provide an estimate of the false positive that an operator would need to deal with, if BGP Ranking were used to detect malicious ASes.

Combining control-plane with data-plane. Finally, we evaluated how the two systems would perform if we used them together. To this end, we label an AS as malicious if it was reported by both systems, with each two report dates to be at most six months apart from each other. For BGP Ranking we used the BGP Ranking top 100 threshold. We found that combining the two systems, we were able to detect 14% of our malicious ASes. This means that of 14% of the known malicious ASes exhibited both control plane and data plane malicious behavior within six months. The fraction of legitimate ASes that both systems detected as malicious is only 3% (*i.e.*, 12 ASes out of 389) for the whole three year period (which is on average 0.08% per month). Finally, five out of the 29 known malicious ASes that were active in the three year observation period were missed by both systems. For example, AS 49544 (Interactive 3D) and AS 39858 (UninetMd, now Comstar Volga Arzamas) are among the top worst ASes that both systems detected.

5.5 Discussion

ASwatch reputation scores in practice . *ASwatch* may help the work of network operators and security practitioners as follows: (1) *Prioritize traffic*: knowing what ASes have suspicious (low reputation) control-plane behavior may help administrators to appropriately handle traffic originating from such ASes; (2) *Peering decisions*: Upstream providers could use AS reputation scores as an additional source of information to make peering decisions, for example by charging higher costs to compensate for the risk of having a low reputation customer or even de-peer early if reputation scores drop significantly; (3) *Prioritize investigations*: law enforcement and security practitioners may prioritize their investigations and start early monitoring on low reputation ASes; (4) *Complement data-plane based systems*: *ASwatch* could be used in combination with data-plane based reputation systems, so that ASes that exhibit malicious behavior both from the control-and data-plane points of view can be prioritized first; (5) *Strengthen existing defenses*: furthermore, reputation could be used as input to other network defenses (*e.g.*, spam filters, botnet detection systems) to improve their detection accuracy.

Working with limited ground truth. We briefly summarize the challenges that we faced due to limited ground truth, and how we mitigated them. (1) *Highly unbalanced dataset*: The ratio of malicious ASes to legitimate ASes produced a highly unbalanced dataset. Before training we used well-known data mining approaches to balance the dataset, by oversampling the underrepresented class of malicious ASes (Section 5.4.1). (2) *Limited time period for training*: We relied on the date of the ground truth reports to estimate the period of time in which the ASes were likely to be actively used by the attackers. We were not able to obtain additional information about the activity periods (or dormancy periods) outside the report dates. Therefore, we designed AS *ASwatch* so that it does not make a final decision for an AS based on a single observation (*i.e.*, a single feature vector). Instead, we introduced parameters to ensure that we label an AS as malicious only if it is assigned consistently low scores for an extended period of time. (3) *Model update with adaptive training*: Because of the lack of information on the activity periods (or dormancy periods) outside the report dates, we were not able to periodically update our model. Therefore, we performed a one-time training on our model using a period of time (January–March 2010) for which we had "clean" data. Even though *ASwatch* uses observations of cases of malicious ASes in 2010, we believe that it effectively models fundamental characteristics of malicious ASes that are still reflected on today's cases. This belief is supported in part by the results of correlating *ASwatch*'s output with recent BGP Ranking reports (see Section 5.4). In our future work, we plan to investigate more sources of ground truth and identify other periods of time that could be included in our training.

Limitations of the AS relationships dataset. To measure our rewiring features, we relied on a dataset that provides snapshots of AS relationships over years (see Section 5.4.1). The relationship inference algorithm is based on the idea of customer cones—the set of ASes an AS can reach through its customer links. This dataset has its own set of limitations. For example, each pair of ASes is assigned only a single relationship, and visibility is limited to the monitoring points publicly available via Routeviews. It is possible that some business relationships may be missing, or that some false relationships are reported. Moreover, since the dataset is provided in snapshots (one snapshot per month), we are not able to observe rewiring activity that may be happening at a finer time scales. Nevertheless, this AS relationships dataset has the largest validated collection of AS relationships gathered to date, with about 44,000 (34.6%) of the inferences validated, and it reports the AS relationships over years, which allowed us to track our ground truth ASes over an extended period of time.

Evasion. Naturally, as for any other detection system, *ASwatch* may face the challenge of sophisticated attackers who attempt to evade it. For example, an attacker may attempt to manage her AS to mimic the BGP behavior of legitimate ASes. However, we should notice that *ASwatch* relies heavily on rewiring features, which capture how an AS connects with other ASes, including upstream providers. Mimicking legitimate behavior to evade *ASwatch* would mean that the malicious AS has to become "less agile". In turn, being less agile may expose the AS to de-peering by its upstream providers as a consequence of accumulating abuse complaints. For example, if McColo (which was taken down in 2008) had not changed ten upstream providers before it was taken down, it might have been taken down much sooner.

Future work. We plan to expand our set of features to capture other types of behavior, such as making peering arrangements for specific prefixes. We intend to expand our sources of bullet-proof hosting ASes, so that we test *ASwatch* over larger datasets and longer periods of time. We also plan to explore how we may combine our set of control plane features with data plane features.

5.6 Conclusion

This chapter presented *ASwatch*, the first system to derive AS reputation based on control-plane behavior. *ASwatch* is based on the intuition that malicious ASes exhibit "agile" control-plane behavior (*e.g.*, short-lived routes, aggressive rewiring). We evaluated *ASwatch* on known malicious ASes and found that it detected 93% of malicious ASes with a 5% false positive rate. When comparing to BGP Ranking, the current state-of-the-art AS reputation system, we found that *ASwatch* detected 72% of reported malicious ASes, whereas BGP ranking detected only 34%. These results suggest that *ASwatch* can better help network operators and law enforcement take swifter action against these ASes that continue to remain sources of malicious activities. Possible remediations could be assessing the risk of peering with a particular AS, prioritizing investigations, and complementing existing defenses that incorporate other datasets.

CHAPTER VI

CONCLUSION

6.1 Summary of contributions

This thesis has demonstrated that it is possible to counteract internet infrastructures that support cybercrime, by designing an AS reputation system that, unlike existing approaches, monitors the control-plane behavior of ASes. Below we summarize our contributions:

- Empirical study of the dynamics of fast-flux service networks. We studied a representative DNS based infrastructure, as it was used to host point-of-sale sites for email scam campaigns. We actively monitored the DNS records for URLs for spam advertised scam campaigns. We studied the rates of change in fast-flux networks, the locations in the DNS hierarchy that change, and the extent to which the fast-flux network infrastructure is shared across different campaigns.
- We presented the first systematic study of the re-wiring dynamics of malicious ASes. We tracked Hostexploit-listed ASes, and compared their AS-level re-wiring dynamics with non-listed ASes. We used a publicly available dataset of Customer-Provider (CP) relations in the Internets AS graph, we studied how interconnection between autonomous systems evolves, both for reported and non-reported ASes.
- We presented a fundamentally different approach to establishing AS reputation. We designed and implemented a system, *ASwatch*, that aims to identify malicious ASes using exclusively *control-plane* data (*i.e.*, the BGP routing control messages exchanged between ASes using BGP). Unlike existing *data-plane* based reputation systems, *ASwatch* explicitly aims to identify *malicious* ASes, rather than assigning low reputation to legitimate ASes that have unfortunately been abused.

6.2 Lessons learned and future work

Below we list some lessons we have learned, that may be useful for designing systems to defend against cybercriminal infrastructures:

Defenses against victim engagement. This thesis focused on counteracting cybercrime infrastructures. Even though it is important to design systems that help towards this goal, it is also important to design defenses to help against victim engagement. An important part of cybercriminals' efforts is to engage end users into various activities, with the ultimate goal to make profit: (1) either directly, for example engage users to buy products from illicit businesses, or (2) indirectly, for example engage a user to install malware that compromises his machine and operate as a spam. It would help to design early warning systems, to warn the end users in a timely manner when such operations are under the way.

Need for systems that target cybercrime monetization operations and infrastructure. Leontiadis *et al.* [57–61] empirically identified the following common components across these different types of criminal businesses. To defend against cybercriminal infrastructures, it would help to design defense systems that target to identify components of cybercriminal infrastructures that play an important role for monetization purposes.

Bringing together complementary defense systems. The research community has developed multiple defense approaches that operate on different, often non-overlapping, areas. For example, significant research effort has focused on critical parts of DNS-based infrastructures [8, 71, 75, 77]. Other efforts have focused on identifying evidence of maliciousness [12, 90, 101], or mismanaged networks [113]. Even though each approach is very effective in targeting specific aspects of cybercrime operations, it may be useful, if we bring those systems together. For example, it may be useful to design an approach that attempts to counteract cybercriminal infrastructures on multiple levels by: (1) monitoring evidence of maliciousness on the routing level, (2) collecting evidence of maliciousness on the data-plane level (traffic monitoring evidence), (3) collecting indications of mismanagement, and (4) identifying parts of cybercrime infrastructures that are critical for monetization purposes.

REFERENCES

[1] ABUSE.CH, "And Another Bulletproof Hosting AS Goes Offline," Mar. 2010. http://www.abuse.ch/?p=2496.

[2] ABUSE.CH, "2011: A Bad Start For Cybercriminals: 14 Rogue ISPs Disconnected." http://www.abuse.ch/?tag=vline-telecom, Jan. 2011.

[3] AL-ROUSAN, N. M. and TRAJKOVIC, L., "Machine learning models for classification of BGP anomalies," in *High Performance Switching and Routing (HPSR)*, pp. 103–108, 2012.

[4] "Alexa top web sites," June 2011. http://www.alexa.com/topsites.

[5] ANDERSON, D. S., FLEIZACH, C., SAVAGE, S., and VOELKER, G. M., "Spamscatter: Characterizing Internet scam hosting infrastructure.," in *14th conference on USENIX Security Symposium*, 2007.

[6] ANDERSON, D. S., FLEIZACH, C., SAVAGE, S., and VOELKER, G. M., "Spamscatter: Characterizing Internet Scam Hosting Infrastructure," in *USENIX Security Symposium*, Aug. 2007.

[7] ANDERSON, D. S., FLEIZACH, C., SAVAGE, S., and VOELKER, G. M., *Spamscatter: Characterizing internet scam hosting infrastructure.* PhD thesis, University of California, San Diego, 2007.

[8] ANTONAKAKIS, M., PERDISCI, R., DAGON, D., LEE, W., and FEAMSTER, N., "Building a dynamic reputation system for dns.," in *USENIX security symposium*, pp. 273–290, 2010.

[9] ANTONAKAKIS, M., PERDISCI, R., LEE, W., VASILOGLOU II, N., and DAGON, D., "Detecting malware domains at the upper dns hierarchy.," in *USENIX Security Symposium*, p. 16, 2011.

[10] ANTONAKAKIS, M., PERDISCI, R., NADJI, Y., VASILOGLOU II, N., ABU-NIMEH, S., LEE, W., and DAGON, D., "From throw-away traffic to bots: Detecting the rise of dga-based malware.," in *USENIX Security Symposium*, pp. 491–506, 2012.

[11] "Autoit." http://www.autoitscript.com/autoit3/.

[12] "BGP Ranking." http://bgpranking.circl.lu/.

[13] "Bgp ranking reports.." http://bgpranking.circl.lu/.

[14] BILGE, L., KIRDA, E., KRUEGEL, C., and BALDUZZI, M., “Exposure: Finding malicious domains using passive dns analysis.,” in *NDSS*, 2011.

[15] BREIMAN, L., “Random Forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[16] “Bulletproof Hosting.” `http://en.wikipedia.org/wiki/Bulletproof_hosting`. Wikipedia.

[17] CHIANG, K. and LLOYD, L., “A case study of the Rustock rootkit and spam bot.,” in *The First Workshop in Understanding Botnets*, 2007.

[18] CISCO SECURITY BLOG, “Threat Spotlight: PoSeidon, A Deep Dive Into Point of Sale Malware.” `http://blogs.cisco.com/security/talos/poseidon`, Mar. 2015.

[19] CLAUSET, A., NEWMAN, M. E. J., and MOORE, C., “Finding community structure in very large networks,” in *Physical Review E 70, 066111. (arxiv:cond-mat/0408187)*, 2004.

[20] COLLINS, M. P., SHIMEALL, T. J., FABER, S., JANIES, J., WEAVER, R., DE SHON, M., and KADANE, J., “Using uncleanliness to predict future botnet addresses,” in *ACM SIGCOMM Internet Measurement Conference*, pp. 93–104, 2007.

[21] “Criminal service providers..” `http://cyberthreat.wordpress.com/category/criminal-service-providers/`.

[22] DAGON, D., ZOU, C., and LEE, W., “Modeling Botnet Propagation Using Time Zones,” in *The 13th Annual Network and Distributed System Security Symposium (NDSS 2006)*, (San Diego, CA), Feb. 2006.

[23] DAGON, D., ZOU, C., and LEE., W., “Modeling Botnet Propagation Using Time Zones,” in *NDSS*, Feb. 2006.

[24] DEREK, G., DONAL, D., and PADRAIG, C., “Tracking the evolution of communities in dynamic social networks,” in *International Conference on Advances in Social Network Analysis and Mining*, 2010.

[25] DHAMDHERE, A. and DOVROLIS., C., “Ten Years in the Evolution of the Internet Ecosystem.,” in *Proceedings of ACM SIGCOMM/USENIX Internet Measurement Conference (IMC).*, 2008.

[26] “DShield: Internet Storm Center - Internet Security.” `www.dshield.org/?`

[27] FEAMSTER, N., “Open problems in BGP anomaly detection.,” in *CAIDA Workshop on Internet Signal Processing.*, 2004.

[28] FEAMSTER, N., JUNG, J., and BALAKRISHNAN, H., “An Empirical Study of Bogon Route Advertisements.,” in *ACM Computer Communications Review.*, 2004.

[29] FETTERLY, D., MANASSE, M., NAJORK, M., and WIENER, J. L., "A large-scale study of the evolution of web pages.," in *Softw. Pract. Exper.*, 2004.

[30] F.LI and HSIEH, M., "An empirical study of clustering behavior of spammers and group-based anti-spam strategies.," in *CEAS 2006:Proceedings of the 3rd conference on email and anti-spam*, 2006.

[31] FRANKLIN, J., PAXSON, V., PERRIG, A., and SAVAGE, S., "An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants," in *ACM CCS*, Nov. 2008.

[32] GAO., L., "On Inferring Autonomous System Relationships in the Internet.," in *IEEE/ACM Transactions on Networking.*, 2001.

[33] HOLZ, T., CORECKI, C., RIECK, K., and FREILING, F. C., "Measuring and Detecting Fast-Flux Service Networks," in *NDSS*, Feb. 2008.

[34] HOSTEXPLOIT, "AS50896-PROXIEZ Overview of a crime server," May 2010. `http://goo.gl/AYGKAQ`.

[35] "Hostexploit," June 2011. `http://hostexploit.com/`.

[36] HOSTEXPLOIT, "World Hosts Report," tech. rep., Mar. 2014. `http://hostexploit.com/downloads/summary/7-public-reports/52-world-hosts-report-march-2014.html`.

[37] "Crimeware-firendly ISPs." `http://hphosts.blogspot.com/2010/02/crimeware-friendly-isps-cogent-psi.html`.

[38] ICANN, "ICANN-Accredited Registrars." `http://www.icann.org/registrars/accredited-list.html`, 2008.

[39] "Icann-accredited registrars." $http://www.icann.org/registrars/accredited-list.html$, 2008.

[40] ICANN SECURITY AND STABILITY ADVISORY COMMITTEE, "SSAC Advisory on Fast Flux Hosting and DNS." `http://www.icann.org/committees/security/sac025.pdf`, Mar. 2008. `http://www.spamhaus.org/sbl`.

[41] IEINSPECTOR SOFTWARE LLC., "IEInspector HTTP Analyzer — HTTP Sniffer, HTTP Monitor, HTTP Trace, HTTP Debug." `http://www.ieinspector.com/httpanalyzer/`, 2007.

[42] JOHNSON, B., CHUANG, J., GROSSKLAGS, J., and CHRISTIN, N., "Metrics for Measuring ISP Badness: The Case of Spam," in *Financial Cryptography and Data Security*, pp. 89–97, Springer, 2012.

[43] JUNG, J. and SIT., E., " An Empirical Study of Spam Traffic and the Use of DNS Black Lists.," in *Proc. ACM SIGCOMM Internet Measurement Conference.*, 2004.

[44] JUNG, J. and SIT, E., "An Empirical Study of Spam Traffic and the Use of DNS Black Lists," in *Internet Measurement Conference*, (Taormina, Italy), October 2004.

[45] KALAFUT, A. J., SHUE, C. A., and GUPTA, M., "Malicious Hubs: Detecting Abnormally Malicious Autonomous Systems," in *IEEE INFOCOM*, pp. 1–5, IEEE, 2010.

[46] KIM, H. A. and KARP., B., "Autograph: Toward automated, distributed worm signature detection.," in *the 13th conference on USENIX Security Symposium*, 2004.

[47] KIRK, J., "ISP Cut Off from Internet After Security Concerns." http://www.pcworld.com/article/153734/mccolo_isp_security.html, Nov. 2008. PC World.

[48] KONTE, M. and FEAMSTER, N., "Wide-area routing dynamics of malicious networks," vol. 41, pp. 432–433, ACM, 2011.

[49] KONTE, M. and FEAMSTER, N., "Re-wiring Activity of Malicious Networks," in *Passive and Active Measurement*, pp. 116–125, Springer, 2012.

[50] KONTE, M., FEAMSTER, N., and JUNG, J., "Fast Flux Service Networks: Dynamics and Roles in Online Scam Hosting Infrastructure," Tech. Rep. GT-CS-08-07, Sept. 2008. http://www.cc.gatech.edu/~feamster/papers/fastflux-tr08.pdf.

[51] KONTE, M., FEAMSTER, N., and JUNG, J., "Dynamics of online scam hosting infrastructure," in *Passive and active network measurement*, pp. 219–228, 2009.

[52] KONTE, M., PERDISCI, R., and FEAMSTER, N., "ASwatch: An AS Reputation System to Expose Bulletproof Hosting ASes," in *Proceedings of the 2015 ACM SIGCOMM*, pp. 625–638, ACM, 2015.

[53] KREBS, B., "Russian Business Network: Down, But Not Out." http://goo.gl/6ITJwP, Nov. 2007. Washington Post.

[54] KREBS, B., "Host of Internet Spam Groups Is Cut Off." http://goo.gl/8J5P89, Nov. 2008. Washington Post.

[55] KREBS, B., "Dozens of ZeuS Botnets Knocked Offline," Mar. 2010. http://krebsonsecurity.com/2010/03/dozens-of-zeus-botnets-knocked-offline/.

[56] KREIBICH, C. and CROWCROFT., J., "Honeycomb: Creating intrusion detection signatures using honeypots.," in *2nd Workshop on Hot Topics in Networks (HotNets-II)*, 2003.

[57] LEONTIADIS, N., *Structuring disincentives for online criminals.* PhD thesis, Carnegie Mellon University Pittsburgh, PA, 2014.

[58] Leontiadis, N. and Christin, N., “Empirically measuring whois misuse,” in *Computer Security-ESORICS 2014*, pp. 19–36, Springer, 2014.

[59] Leontiadis, N., Moore, T., and Christin, N., “Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade.,” in *USENIX Security Symposium*, 2011.

[60] Leontiadis, N., Moore, T., and Christin, N., “Pick your poison: pricing and inventories at unlicensed online pharmacies,” in *fdksfndklfdsProceedings of the fourteenth ACM conference on Electronic commerce*, pp. 621–638, ACM, 2013.

[61] Leontiadis, N., Moore, T., and Christin, N., “A nearly four-year longitudinal study of search-engine poisoning,” in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 930–941, ACM, 2014.

[62] Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Félegyházi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., Liu, H., and others, “Click trajectories: End-to-end analysis of the spam value chain,” in *Security and Privacy (SP), 2011 IEEE Symposium on*, pp. 431–446, IEEE, 2011.

[63] Li, J., Guidero, M., Wu, Z., Purpus, E., and Ehrenkranz, T., “BGP Routing Dynamics Revisited,” *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 2, pp. 5–16, 2007.

[64] Li, Z., Sanghi, M., Chen, Y., Kao, M., and Chavez., B., “Hamsa: Fast signature generation for zero-day polymorphic worm with provable attack resilience.,” in *In IEEE Symposium on Security and Privacy*, 2006.

[65] Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., and others, “AS Relationships, Customer Cones, and Validation,” in *Proceedings of ACM SIGCOMM Internet Measurement Conference*, pp. 243–256, ACM, 2013.

[66] Mai, J., Yuan, L., and Chuah, C.-N., “Detecting BGP anomalies with wavelet,” in *IEEE Network Operations and Management Symposium*, pp. 465–472, IEEE, 2008.

[67] Maria , Roberto Perdisci, Nick Feamster, “ASwatch an AS reputation system to expose bulletproof hosting ASes.” `https://www.nanog.org/meetings/nanog62/agenda`, Oct. 2014.

[68] McMillan, R., “After Takedown, Botnet-linked ISP Troyak Resurfaces.” `http://goo.gl/5k0OV1`, Mar. 2010. Computer World.

[69] Moore, T., Leontiadis, N., and Christin, N., “Fashion crimes: trending-term exploitation on the web,” in *Proceedings of the 18th ACM conference on Computer and communications security*, pp. 455–466, ACM, 2011.

[70] MOURA, G. C. M., SADRE, R., SPEROTTO, A., and PRAS, A., "Internet bad neighborhoods aggregation," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, pp. 343–350, IEEE, 2012.

[71] NADJI, Y., ANTONAKAKIS, M., PERDISCI, R., DAGON, D., and LEE, W., "Beheading hydras: performing effective botnet takedowns," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 121–132, ACM, 2013.

[72] NADJI, Y., ANTONAKAKIS, M., PERDISCI, R., and LEE, W., "Connected colors: Unveiling the structure of criminal networks," in *Research in Attacks, Intrusions, and Defenses*, pp. 390–410, Springer, 2013.

[73] NAZARIO, J. and HOLZ, T., "As the net churns: Fast-flux botnet observations," in *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*, pp. 24–31, IEEE, 2008.

[74] NEWSOME, J., KARP, B., and SONG, D., "Polygraph: Automatically generating signatures for polymorphic worms.," in *Proceedings of the 2005 IEEE Symposium on Security and Privacy*, 2005.

[75] PASSERINI, E., PALEARI, R., MARTIGNONI, L., and BRUSCHI, D., "FluXOR: detecting and monitoring fast-flux service networks," in *DIMVA*, July 2008.

[76] PATHAK, A., HU, Y. C., and MAO, Z. M., "Peeking into Spammer Behavior from a Unique Vantage Point," in *First USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, (San Francisco, CA), Apr. 2008.

[77] PERDISCI, R., CORONA, I., DAGON, D., and LEE, W., "Detecting malicious flux service networks through passive analysis of recursive dns traces," in *Computer Security Applications Conference, 2009. ACSAC'09. Annual*, pp. 311–320, IEEE, 2009.

[78] PRAKASH, B. A., VALLER, N., ANDERSEN, D., FALOUTSOS, M., and FALOUTSOS, C., "BGP-lens: Patterns and Anomalies in Internet Routing Updates," in *ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pp. 1315–1324, 2009.

[79] RAJAB, M., ZARFOSS, J., MONROSE, F., and TERZIS, A., "A Multifaceted Approach to Understanding the Botnet Phenomenon," in *ACM SIGCOMM/USENIX Internet Measurement Conference*, (Brazil), Oct. 2006.

[80] RAMACHANDRAN, A. and FEAMSTER, N., "Understanding the network-level behavior of spammers.," in *Proceedings of Sigcomm*, 2006.

[81] RAMACHANDRAN, A., FEAMSTER, N., and VEMPALA, S., "Filtering spam with behavioral blacklisting.," in *Proceedings of the 14th ACM conference on computer and communications security*, 2007.

[82] RAMACHANDRAN, A. and FEAMSTER, N., "Understanding the Network-Level Behavior of Spammers," in *SIGCOMM*, Sept. 2006.

[83] "The Russian Business Network." http://en.wikipedia.org/wiki/Russian_Business_Network.

[84] REKHTER, Y., LI, T., and HARES, S., *A Border Gateway Protocol 4 (BGP-4)*, Jan. 2006. RFC 4271.

[85] REXFORD, J., WANG, J., XIAO, Z., and ZHANG, Y., "BGP routing stability of popular destinations," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pp. 197–202, ACM, 2002.

[86] "The RouteViews Project." www.routeviews.org/.

[87] ROVETA, F., CAVIGLIA, G., DI MARIO, L., ZANERO, S., MAGGI, F., and CIUCCARELLI, P., "Burn: Baring unknown rogue networks," in *International Symposium on Visualization for Cyber Security (VizSec)*, 2011.

[88] SHADOWSERVER, "Botnets: Command and Control Mechanisms." https://www.shadowserver.org/wiki/pmwiki.php/Information/Botnets.

[89] SINGH, S., ESTAN, C., VARGHESE, G., and SAVAGE, S., "Automated worm fingerprinting.," in *OSDI*, 2004.

[90] "Sitevet." http://sitevet.com/.

[91] SPAM TRACKERS, "Fast-flux." http://spamtrackers.eu/wiki/index.php?title=Fast-flux, Oct. 2007.

[92] SPAM TRACKERS, "Category:Yambo family." http://spamtrackers.eu/wiki/index.php?title=Category:Yambo_family, Mar. 2008.

[93] SPAM TRACKERS, "Diamond Replicas." http://spamtrackers.eu/wiki/index.php?title=Diamond_Replicas, Apr. 2008.

[94] SPAM TRACKERS, "Exquisite Replica." http://spamtrackers.eu/wiki/index.php?title=ExquisiteReplica, Mar. 2008.

[95] SPAM TRACKERS, "Canadian Pharmacy." http://spamtrackers.eu/wiki/index.php?title=Canadian_Pharmacy, Apr. 2009.

[96] "Spamhaus.." www.spamhaus.org.

[97] "Spamhaus PBL." http://www.spamhaus.org/pbl.

[98] "Spamhaus SBL." http://www.spamhaus.org/sbl.

[99] "Spamhaus XBL." http://www.spamhaus.org/xbl.

[100] SPAMMER-X, "Inside the Spam Cartel.." Syngress., 2004.

[101] Stone-Gross, B., Kruegel, C., Almeroth, K., Moser, A., and Kirda, E., "FIRE: Finding rogue networks," in *IEEE Computer Security Applications Conference (ACSAC)*, pp. 231–240, IEEE, 2009.

[102] The Honeynet Project, "Know Your Enemy: Fast-Flux Service Networks." `http://www.honeynet.org/papers/ff/`, July 2007.

[103] Todd, J., "AS number inconsistencies.." $http : //www.merit.edu/mail.archives/nanog/2002 - 07/msg00259.html.$, 2002.

[104] "AS-Troyak Exposes a Large Cybercrime Infrastructure." `https://blogs.rsa.com/as-troyak-exposes-a-large-cybercrime-infrastructure/`.

[105] "URIBL." `http://www.uribl.org/`.

[106] "Vmware." `http://www.vmware.com/`.

[107] Wagner, C., François, J., State, R., Dulaunoy, A., Engel, T., and Massen, G., "ASMATRA: Ranking ASes providing transit service to malware hosters," in *IFIP/IEEE International Symposium on Integrated Network Management*, pp. 260–268, 2013.

[108] Xie, Y., Yu, F., Achan, K., Gillum, E., Goldszmidt, M., and Wobber, T., "How dynamic are IP addresses?," in *ACM Sigcomm*, 2007.

[109] Xie, Y., Yu, F., Achan, K., Gillum, E., Goldszmidt, M., and Wobber, T., "A Multifaceted Approach to Understanding the Botnet Phenomenon," in *ACM SIGCOMM*, (Kyoto, Japan), Aug. 2007.

[110] Xie, Y., Yu, F., Achan, K., Panigrahy, R., and Hulten, G., "Spamming botnets: signatures and characteristics," in *In SIGCOMM*, 2008.

[111] Zdrnja, B., Brownlee, N., and Wessels, D., "Passive Monitoring of DNS Anomalies," in *DIMVA*, July 2007.

[112] Zhang, J., Rexford, J., and Feigenbaum, J., "Learning-based anomaly detection in BGP updates," in *ACM SIGCOMM Workshop on Mining Network Data (MineNet)*, pp. 219–220, ACM, 2005.

[113] Zhang, J., Durumeric, Z., Bailey, M., Karir, M., and Liu, M., "On the Mismanagement and Maliciousness of Networks," in *Proceedings of the 21st Annual Network & Distributed System Security Symposium (NDSS '14)*, (San Diego, California, USA), February 2013.