Dynamics of Online Scam Hosting Infrastructure

Maria Konte[1], Nick Feamster[1], and Jaeyeon Jung[2]

[1] Georgia Institute of Technology
[2] Intel Research
{mkonte,feamster}@cc.gatech.edu, jaeyeon.jung@intel.com

Abstract. This paper studies the dynamics of scam hosting infrastructure, with an emphasis on the role of fast-flux service networks. By monitoring changes in DNS records of over 350 distinct spam-advertised domains collected from URLs in 115,000 spam emails received at a large spam sinkhole, we measure the rates and locations of remapping DNS records, and the rates at which "fresh" IP addresses are used. We find that, unlike the short-lived nature of the scams themselves, the infrastructure that hosts these scams has relatively persistent features that may ultimately assist detection.

1 Introduction

Online scam hosting infrastructure is critical to spam's profit cycle; victims must contact point-of-sale Web sites, 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) [2]. However, simple countermeasures, such as blocking the IP address called for more sophisticated techniques. For example, the past year has seen the rise of "fast-flux service networks" [5], which allow the sites that host online scams to change rapidly.

This paper studies the dynamics of the Internet infrastructure that hosts point-of-sale sites for email scam campaigns. We focus on how fast-flux service networks are used to host these online scams. Beyond offering a better understanding of the characteristics of the infrastructure, our study discovers invariant features of the infrastructure that may ultimately help identify scams and the spam messages that advertise them faster than existing methods.

We study the scam sites that were hosted by 384 domains as part of 21 scam campaigns in over 115,000 emails collected over the course of a month at a large spam sinkhole. This paper studies two aspects of the dynamics:

- *What are the rates and extent of change?* We examine the rates at which scam infrastructures, via the use of fast-flux service networks, redirect clients to different authoritative name servers (either by changing the authoritative nameserver's name or IP address), or to different Web sites entirely. We find that, while the scam sites' DNS TTL values do not differ significantly from other sites that perform DNS-based load balancing, the rates of change (1) differ from legitimate load balancing activities; and (2) differ across individual scam campaigns.
- *How are dynamics implemented?* We study the mechanics by which scam hosting infrastructures change the Web servers to which clients are redirected. We determine the location of change by monitoring any changes of (1) the authoritative

Time: 20:51:52 (GMT)					
A records	TTL	NS records	TTL	IPs of NS records	TTL
77.178.224.156, 79.120.37.38, 79.120.72.0, 85.216.198.225, 87.228.106.92, 89.20.146.249, 213.141.146.83, 220.208.7.115	300	ns0.nameedns.com, ns0.nameedns1.com, ns0.renewwdns.com, ns0.renewwdns1.com	172800	218.236.53.11, 89.29.35.218, 78.107.123.140, 79.120.86.168	172800
Time: 20:57:49 (GMT)					
A records	TTL	NS records	TTL	IPs of NS records	TTL
61.105.185.90, **69.228.33.128,** 79.120.37.38, 87.228.106.92, 89.20.146.249, **89.20.159.178**, **89.29.35.218**, **91.122.121.88**,	300	ns0.nameedns.com, ns0.nameedns1.com, ns0.renewwdns.com, ns0.renewwdns1.com	172800	218.236.53.11, 89.29.35.218, 78.107.123.140, **213.248.28.235**	172800

Table 1. DNS lookup results for the domain `pathsouth.com` (responding authoritative nameserver was 218.236.53.11): The IP addresses in bold highlight changes between the two lookups taken six minutes apart. For the full list of IPs for this domain, see our technical report [7].

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 analyze both on the basis of individual spam-advertised domains and campaigns that are formed after domain clustering. 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. We also study the infrastructures in terms of the geographical and topological locations of scam hosts and the country in which the domains were registered.

Background. Fast-flux is a DNS-based method that cybercriminals use to organize, sustain, and protect their service infrastructures such as illegal Web hosting and spamming. 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 allow a controller to quickly change the mapping between a domain name and its A records, its NS records, or the IP addresses of its NS records) [13]. Cybercriminals can rotate through compromised hosts, which renders traditional blacklisting largely ineffective. We show an example of a fast-flux domain, called `pathsouth.com`, that we monitored on January 20, 2008 (Table 1).

Related Work. The operation of fast-flux service networks and the use of these platforms to send spam was first described in detail by the Honeynet Project [13]. Compared to other studies of fast-flux networks [4, 8, 15], we focus on fast-flux networks as they relate to hosting online scams. This paper is the largest such study (it is an order of magnitude larger than the previous study [4]), and it is the first to (1) study the location (within the DNS hierarchy) of dynamics, (2) the behavior of hosting infrastructure across campaigns. We examine scam hosting infrastructure using both spam trap data and content-based scam campaign clustering; we draw on previous studies that analyzed spam trap data [6, 11, 14] or performed content-based analysis [2, 4, 9], albeit for different purposes. Previous work has used passive DNS monitoring to study the dynamics of botnets [3, 10], some of which are now believed to used to host fast-flux networks.

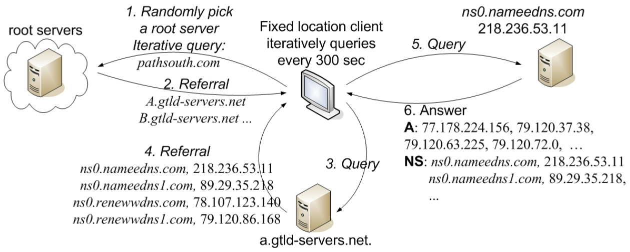


Fig. 1. Data collection diagram. The resolver records all DNS mappings at each level of DNS resolution. Here, we feature the same domain as in Table 1.

The rest of this paper is organized as follows. Section 2 describes our data and its limitations. Section 3 describes the dynamics of scam infrastructure and fast-flux service networks that we observed hosting 21 different spam campaigns over the course of a month. Section 4 refers to topological and geographic location of the observed infrastructures. Section 5 concludes with a summary and discussion of future work.

2 Data Collection

Our data collection and processing involves three steps: (1) passive collection of spam data; (2) active DNS monitoring of domains for scam sites contained in those spam messages; (3) clustering of spam and DNS data by scam campaign. This section describes these methods in detail, our use of popular Web sites as a baseline for comparison, and the limitations of our dataset.

We collected 3,360 distinct domain names that appeared at spam email messages from a large spam sink hole from October 1, 2007 to December 31, 2007. We used a simple URL pattern matcher to extract URLs from the message body. Next, we implemented an iterative resolver (at a fixed location) to resolve every domain name from this set once every five minutes. Figure 1 illustrates the process 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 each query. The iterative resolver recorded responses received at every level of the DNS hierarchy, including all referrals and answers.

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 added 120 new domains to the list. After each domain had been resolved continuously for three weeks, we removed the domain from the list. The resolver operated from January 14, 2008 to February 9, 2008. We picked the domains and the campaigns they mapped to, by restricting our analysis to the domains that had reachable Web sites and for which we had observed at least one change in any DNS record. To compare

Campaign	Spam emails	Spam advertising IPs	Campaign domains	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
All campaigns	115198	77030	465	384	9521	2421	9821
Alexa data set				500	1048	852	1877

Table 2. Statistics for the top three scam campaigns compared to Alexa domains. Campaigns are sorted by the total number of IP addresses returned from A records.

the dynamics those of "legitimate" domains, we used the same iterative resolution process to study the dynamics of the 500 most popular Web server domains, according to Alexa [1].

Clustering spam by scam campaign. To cluster the spam messages into *scam campaigns*, we retrieved content from the URLs in the email messages and cluster emails whose URLs retrieve common content. 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. In the case of slow response or when only a few small non-image files are received, we checked whether the downloaded file names of each URL is a subset of those of already identified campaign.

All 21 campaigns exhibited fluxing behavior in their DNS records to some extent during the measurement period. Table 2 shows the summary data for the three campaigns that used the most hosting servers. We denote each campaign with a *category-ID*, which we assigned based on the products offered on the Web site. The first 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 third column is the total number of domains for that campaign, and the fourth column is the number of domain names that we found changing ("fluxing domains"). The last three columns show (1) the distinct number of IPs returned as A records of domains ($\text{IP}_{\text{domains}}$); (2) the number of IPs returned as A records of name servers ($\text{IP}_{\text{nameservers}}$); and (3) the total distinct number of IPs from the combined sets ($\text{IP}_{\text{domains}} \cup \text{IP}_{\text{nameservers}}$).

The top campaign is Pharmacy-A, one of the Canadian Pharmacy scam campaigns [12]. The campaign used 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 (Exquisite Replica) [12] and Watch-B (Diamond Replicas) [12], both of which offer replica watches. For these campaigns, the average number of A records associated with a single domain name is over 50, demonstrating a lot of activity in moving scam sites. We also witnessed multiple domains that shared a few hosting servers.

Registrars. To determine the registrar responsible for each of the 384 scam hosting domains, we performed `jwhois` queries on May 7, 2008 for each domain. Table 3 shows that about 70% of these domains are still marked as active and registered with just eight registrars in China, India, and US. Among these, the three registrars in China are

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

Table 3. Registrars of the 384 scam domains as of May 7, 2008.

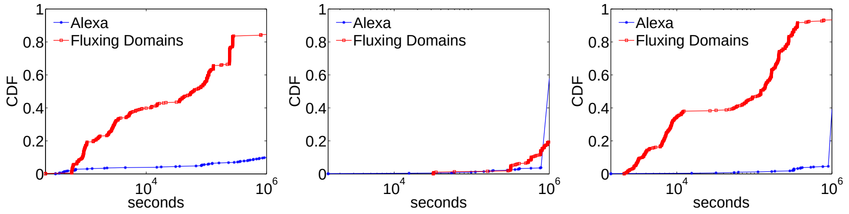
responsible for 257 domains (66% of the total or 95% of the active ones). Our data collection was done before February 2008, so all domains were registered before that time. All 384 domains were all active after four months, and 2% of the domains had been active for over 7 months. Interestingly, over 40% of these domains were registered in January 2008, just before the scams themselves were hosted; thus, a newly registered domain might also ultimately serve as a useful indicator for detecting scam hosting.

Limitations. Our data is derived from spam collected at a single spam trap, which receives a relatively high number of spam messages (6,247,937 messages from October 2007 through February 2008) but may still reflect some bias in the spam it receives. Because we are primarily looking to analyze the dynamics of widespread campaigns (i.e., domains that are likely visible at many traps), this limitation should not greatly affect our results. The main limitation is that our data may not contain all domains for a particular scam. Some of our measurements occurred months after the spam was received, but our results suggest that the dynamics of these domains remain relatively consistent over the month that we monitored them.

3 Dynamics

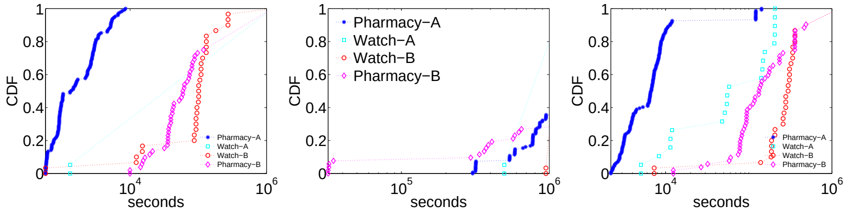
We studied three aspects of dynamics: (1) the rate at which DNS records change at each level of the hierarchy; (2) the rate at which scam hosting infrastructure accumulates new IP addresses (both overall and by campaign); and (3) the location in the DNS hierarchy where changes take pace. To understand the nature of these features with respect to "legitimate" load balancing behavior, we also analyzed the same set of features for 500 popular sites listed by Alexa [1] as a baseline.

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 compared with the TTLs for domains listed by Alexa. (Our technical report includes the TTL distribution graphs [7].) The distribution 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, almost all scam domains we analyzed had TTL values for NS records of longer than a day. These results 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).



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

Fig. 2. Distribution of the average time between changes of A, NS, and IP of NS records.

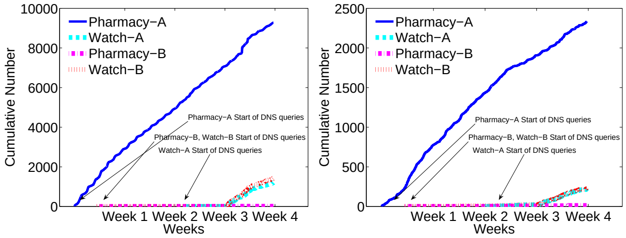


(a) A records (b) NS records (c) IP of NS records

Fig. 3. 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.

We grouped the responses according to the authoritative server that provided them to account for possible load balancing. We then performed pairwise comparisons across each group of records. In the case of A and NS-record responses, we considered a response to be a change if at least one new record appears or if the number of records returned has otherwise changed since the last response; we did not consider reordering the records as a change. In the case of IP addresses of NS records, we considered the response to be a change if either NS names appear with different IPs or a new NS name appears. We discovered the following two characteristics, both of which might ultimately help automatically detect scams:

– *Scam domains change on shorter time intervals than their TTL values.* Figure 2 shows the cumulative distribution of 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 scam domains change hosting servers (A records) and name servers (IP addresses of NS records) more frequently than popular Web servers do, and also much more frequent than TTL values of the records.
– *Domains in the same campaign exhibit similar rates of change.* We also analyzed the rate of change of DNS records after clustering the scam domains according to campaign. Figure 3 shows these results for the top 4 campaigns (ranked by the



(a) A records - Top 4 (b) IP of NS records - Top 4

Fig. 4. Cumulative number of distinct IPs for the A records and IP addresses of NS records for the top 4 campaigns across the four weeks of data collection.

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.

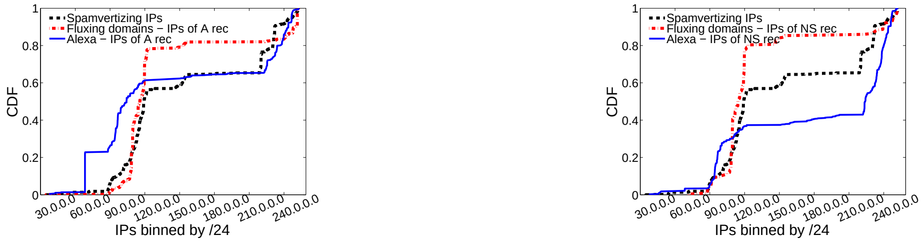
Rate of Accumulation. We measure the rate at which the network grows over time. In practice, our measurement is limited by the rate at which a domain updates its DNS records and what we present in this section is the rate at which a previously unseen host becomes an active hosting server (A records of 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.* [4], we determined the rate of growth by repeatedly resolving each domain and assigning an increasing sequential ID to each previously unseen IP address. Holz *et al.* performed this analysis for A records of domains without regard to campaign; we performed this analysis for A records and IP addresses of NS records, both at the level of individual domains and at the level of campaigns:

- *Rates of accumulation differ across campaigns.* Figures 4(a) and 4(b) show the total number of distinct IPs for each scam domain (the *y*-value of the end of each line) over the four weeks of data collection (all iterations, 300 seconds apart from each other) and how fast each campaign accumulated new hosts (slope), for the IP addresses of A records and NS records, respectively. A steeper slope indicates more rapid accumulation of new IP addresses for that campaign.
- *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 accumulating new IP addresses. We examined two campaigns that exhibited rapid accumulation of IP addresses after some dormancy. Interestingly, only one domain from each campaign begins accumulating IP addresses. These two campaigns shared exactly the same set of NS names. In addition to accumulation, we also saw attrition: 10% of scam domains became unreachable in the while we were monitoring them. These domains may have been blacklisted and removed by registrars or the scammers.

Campaign	Domains	Location of change						
		A	[NS IP]	NS	A+[NS IP]	A+NS	NS+[NS IP]	A+NS+[NS IP]
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	-	-	-
Total	384	18	52	3	219	1	-	91
Alexa	500	37	5	15	4	1	1	-

Table 4. Location of change for the top five campaigns, sorted by the total number of distinct IPs of A records.



(a) A records

(b) IP of NS records

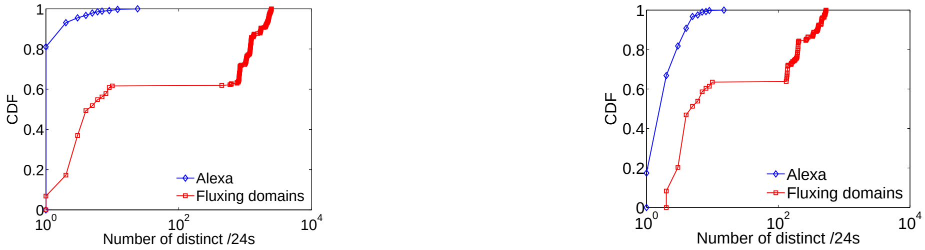
Fig. 5. Distribution of the IPs of A rec of authoritative servers spam senders.

Location of Change in DNS Hierarchy. We observed many scam domains with NS records or IP addresses of NS records that change rapidly, often in conjunction with other records: Campaigns change DNS record mappings at different levels of the DNS hierarchy. Table 4 shows the type of change for the top five campaign. In contrast to previous studies [5, 13], we observed many different types of changes in addition to single flux (A records) and double flux (A, and IP address of NS). Another notable point is that each campaign tends to combine techniques: For Pharmacy-A, 52% of domains are double flux and 48% change all three types of records. This result indicates that a single campaign can operate using multiple infrastructures.

4 Location

In this section, we examine the network and geographic locations of hosts that are hosting scam Web sites or serving as name servers; we also compare these locations to those of both spamming hosts and legitimate Web sites hosts.

Topological Location. To examine whether scam sites use different portions of the IP space than the top 500 domains, we studied the distribution of the IPs across the whole IP range. Figure 5 shows that scam 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. 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:



(a) A records

(b) IP of NS records

Fig. 6. Distribution of unique /24s that appeared for *all* records in a reply.

- *The predominant networks that host scam sites differ from those that host spammers for the corresponding scam campaigns (Figure 5).* Our technical report lists for 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) [7]. Interestingly, there is almost no overlap between the ASes hosting the scam sites and the nameservers (mostly Asia) and the ASes hosting the spamming IP addresses (mostly Latin America, Turkey, and US). The fact that significant differences exist between networks of scam infrastructure and those of spammers suggest that hosts in different regions of the IP address space do in fact play different "roles" in spam campaigns.
- *DNS lookups for scam domains often return much more widely distributed IP addresses than lookups for legitimate Web sites.* Our intuition was that fast-flux networks that hosted scam 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 typically map a single client to a nearby Web cache). Figure 6 shows the distribution of distinct /24s that appear at the answer section of the DNS replies) for all records in the reply. Roughly 40% of all A records returned for scam domains were distributed across at least 300 distinct /24s, and many were distributed across thousands of /24s. An overly widespread distribution of query replies may indicate that a domain is indeed suspicious (e.g., a fast-flux network).

Geographic Location. Hosting servers and name servers are widely distributed. 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 across 2,976 IP addresses across 157 countries. Although many scam nodes appear to be in Russia, Germany, and the US, the long list of ASes and countries shows that scam networks are truly distributed; this geographical distribution may be necessary to accommodate the diurnal pattern of compromised hosts' uptime [3]. 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 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).

5 Summary

This paper studied 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 dynamics features of fast-flux service networks as they are used to host online scam and contrast to the dynamics used for load balancing for popular Web sites. Our findings suggest that monitoring the infrastructure for unusual, invariant changes in DNS mappings may be helpful for automating detection. We plan to explore this possibility in future work.

References

1. Alexa. Alexa the Web Information Company. `http://www.alexa.com/`, 2008.
2. D. S. Anderson, C. Fleizach, S. Savage, and G. M. Voelker. Spamscatter: Characterizing Internet Scam Hosting Infrastructure. In *USENIX Security Symposium*, Aug. 2007.
3. D. Dagon, C. Zou, and W. Lee. Modeling Botnet Propagation Using Time Zones. In *The 13th Annual Network and Distributed System Security Symposium (NDSS 2006)*, San Diego, CA, Feb. 2006.
4. T. Holz, C. Corecki, K. Rieck, and F. C. Freiling. Measuring and Detecting Fast-Flux Service Networks. In *NDSS*, Feb. 2008.
5. ICANN Security and Stability Advisory Committee. SSAC Advisory on Fast Flux Hosting and DNS. `http://www.icann.org/committees/security/sac025.pdf`, Mar. 2008.
6. J. Jung and E. Sit. An Empirical Study of Spam Traffic and the Use of DNS Black Lists. In *Internet Measurement Conference*, Taormina, Italy, October 2004.
7. M. Konte, N. Feamster, and J. Jung. Fast Flux Service Networks: Dynamics and Roles in Online Scam Hosting Infrastructure. Technical Report GT-CS-08-07, Sept. 2008. `http://www.cc.gatech.edu/~feamster/papers/fastflux-tr08.pdf`.
8. E. Passerini, R. Paleari, L. Martignoni, and D. Bruschi. FluXOR: detecting and monitoring fast-flux service networks. In *DIMVA*, July 2008.
9. A. Pathak, Y. C. Hu, and Z. M. Mao. 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.
10. M. Rajab, J. Zarfoss, F. Monrose, and A. Terzis. A Multifaceted Approach to Understanding the Botnet Phenomenon. In *ACM SIGCOMM/USENIX Internet Measurement Conference*, Brazil, Oct. 2006.
11. A. Ramachandran and N. Feamster. Understanding the Network-Level Behavior of Spammers. In *SIGCOMM*, Sept. 2006.
12. Spam Trackers. `http://spamtrackers.eu/wiki/index.php?title=Main_Page`.
13. The Honeynet Project. Know Your Enemy: Fast-Flux Service Networks. `http://www.honeynet.org/papers/ff/`, July 2007.
14. Y. Xie, F. Yu, K. Achan, E. Gillum, M. Goldszmidt, and T. Wobber. How dynamic are IP addresses? In *ACM SIGCOMM*, Kyoto, Japan, Aug. 2007.
15. B. Zdrnja, N. Brownlee, and D. Wessels. Passive Monitoring of DNS Anomalies. In *DIMVA*, July 2007.