As the Net Churns: Fast-Flux Botnet Observations

Jose Nazario
Arbor Networks
jose@arbor.net

Thorsten Holz
University of Mannheim
holz@uni-mannheim.de

September 5, 2008

Abstract

While botnets themselves provide a rich platform for financial gain for the botnet master, the use of the infected hosts as webservers can provide an additional botnet use. Botnet herders often use *fast-flux DNS* techniques to host unwanted or illegal content within a botnet. These techniques change the mapping of the domain name to different bots within the botnet with constant shifting, while the bots simply relay content back to a central server. This can give the attackers additional stepping stones to thwart takedown and can obscure their true origins.

Evidence suggests that more attackers are adopting fast-flux techniques, but very little data has been gathered to discover what these botnets are being used for. To address this gap in understanding, we have been mining live traffic to discover new fast-flux domains and then tracking those botnets with active measurements for several months. We have identified over 900 fast-flux domain names from early to mid 2008 and monitored their use across the Internet to discern fast-flux botnet behaviors. We found that the active lifetimes of fast-flux botnets vary from less than one day to months, domains that are used in fast-flux operations are often registered but dormant for months prior to activation, that these botnets are associated with a broad range of online fraud and crime including pharmacy sites, phishing and malware distribution, and that we can identify distinct botnets across multiple domain names. We support our findings through an in-depth examination of an Internet-scale data continuously collected for hundreds of domain names over several months.

1 Introduction

Botnet-based hosting has become one of the many ways that botnet operators can make money from their botnet by turning it into a high availability, global content distribution network. Because the content is unwelcome or sometimes illegal, it must be defended from disruption. *Fast-flux* is a technique to cycle the mappings of domain names to IP addresses of hosts participating in a botnet, often with short lifetime mappings [16]. These bots, in turn, simply relay the content from the botnet endpoint to a central location (often termed *mothership*). This makes disruption of the content hosting significantly harder than blocking IP addresses, as a fast-flux botnet may contain hundreds of thousands of endpoints. To globally disrupt a fast-flux hosting operation, one must shut down the domain name at the registrar level. Unfortunately, this is often a tedious and time-consuming job, especially given the fact that not all registrars respond to abuse complaints. Alternatively, a fast flux botnet can be shut down if the *mothership* systems are taken offline. This has proven difficult to achieve in practice because these systems are often located in complicit or hostile networks which do not respond to abuse complaints.

Hosts become part of a fast-flux botnet through malware infections and then participate in the botnet. This botnet may be a traditional IRC-based botnet, or based on other protocols. The botnet then manages itself to screen for candidate hosts that should be listed in the fast-flux DNS results. Criteria for using the infected host include a globally unique IP address with global accessibility, which means that hosts behind a NAT device can not be used in a fast-flux network. The botnet may also use a node's bandwidth and uptime as criteria for including a bot in the pool of fast-flux servers. Because only a subset of the bots match these criteria, this means that the botnet is only partially visible through fast-flux DNS mappings. How a botnet manages its member nodes in the fast-flux network has only been studied from the outside, as no researchers, to our knowledge, have been able to capture and analyze the fast-flux management toolkits used by a live botnet.

In practice we have seen fast-flux DNS techniques used in phishing and malware hosting operations. Two of the most well known fast-flux service networks are the Storm Worm network [13], which sometimes uses DNS-based websites to distribute the malware via spam lures, and the Rock Phishing network, which has a long history of using fast-flux hosting for its phishing sites [11].

Fast-flux techniques and their use in malicious operations has been described elsewhere [4, 6, 16]. Experience suggests that fast-flux botnet techniques have become more popular in the past year as it is adopted by additional groups. However, a number of key questions about the actual operations and measurements of fast-flux botnets remain unanswered. In this paper, we address these open questions and make the following three contributions: first, we examine the efficacy of a real-time fast-flux qualification process, the use of domain names in fast-flux operations in the wild, and demonstrates that it is possible to distinguish individual botnets using fast-flux hosting techniques for specific purposes based solely on DNS data mining. Second, we present measurement results for tracking more than 900 fast-flux domains for a period of several months. We study for example how popular fast-flux botnet operations are among the malcode and botnet underground community, and what activities these botnets support. These findings highlight changes in the past year in fast-flux botnet operations. Third, we introduce a method to identify distinct fast-flux botnets. Our results indicate that only a small number of distinct botnets that use fast-flux techniques exist in the wild.

The remainder of this paper is organized as follows. In Section 2, we discuss how we identify fast-flux domain names and store the data for the botnets. Then, in Section 3, we analyze the data collected by ATLAS, an automated system to track fast-flux botnets, to determine the behaviors of these networks. We explore the size, lifetimes, and interactions of fast-flux domains. In Section 4, we provide some discussion and further analysis of the data before concluding the paper and discussing future directions in Section 5.

2 Tracking Fast-Flux Domains

We used the ATLAS system from Arbor Networks to automatically identify and track new fast-flux networks [12]. ATLAS is a data repository and globally deployed network of honeypots, to which fast-flux tracking was added in early 2008. The goal of tracking fast-flux networks is to gain insights into botnets and infected clients around the world.

ATLAS builds a list of fast-flux domain names and then loops over them, performing DNS queries to identify participating bots until the domain name is shut down by the registrar. Our approach assumes a few features of a typical fast-flux botnet. First, we assume that the DNS names will be advertised in one form or another, typically through spam email. We do make accommodations for some botnets lacking visibility by using multiple sources to discover fast-flux domains, as described below. Secondly, we assume that the botnet will self-manage the IP addresses it should be providing in the DNS queries, both for accuracy and for completeness.

2.1 Identifying and Qualifying Domains

Candidate fast-flux domains are identified through at least three different means. The first, and where most of the domains we have identified were found, is via spam trap analysis. All of the URLs mentioned in spam emails are analyzed to discover the domains and then passed to the qualifier, described below. The second mechanism is through different domain blacklists such as [2, 10], or via internal malware analysis. The third mechanism is manual analysis to identify candidate domains. We can act on anti-spam and anti-malware community reports to discover new domains and pass them to the screening process for possible tracking.

All candidate domains then go through a set of heuristics to qualify them as fast-flux domains. These heuristics are similar to those used by Holz *et al.* [4]. We have developed our own set of heuristics based on real-world experience when building the system. The qualification process works in different phases. First, the fully qualified domain name candidate is stripped down to the domain name only to ensure consistency. Then several DNS queries are made, including A-record queries, NS-record queries- and SOA-record queries, along with BGP queries to discover the origin autonomous system number (ASN). The results of these queries are scored as such, with a "point" being added for every criteria being met:

- TTL measurements – If the time to live of the A records is under 900 seconds (15 minutes). If the TTL is under two seconds (as is common with Storm Worm fast-flux domain names), it is repeatedly queried at this stage to discover more IP address mappings and to look for distinct replies.
- If we discover more than five unique IP addresses in the A record queries.
- If the average "distance" between each of the IP addresses in the result set is more than 65,535 addresses (equivalent to distinct /16 netblocks).
- If we find more than eight IPs in the A record result set and the average distance between the addresses is more than 65,535 IP addresses apart.
- If the A record results highlight more than two distinct ASNs.
- If the nameserver records are more than 65,355 IP addresses apart.
- If there are more than three nameserver entries.
- If the nameservers are located in more than two distinct ASNs.
- If the minimum "retry" listed is under 900 seconds, as determined via the SOA query.

If more than four of the above behaviors are observed the domain is marked "fluxy". Similar to the metric by Holz *et al.* [4], we weight different characteristics of the DNS responses to decide whether or not a given domain belongs to a fast-flux botnet. The intuition behind our metric is that we assume that any successful fast-flux network will set the A record DNS mappings for the content servers to be short lived and widely spread across the Internet to provide maximum availability for the botnet in response to endpoint disruptions. Nameserver records, in contrast, may have longer TTL values, but we will see fewer NS records per domain and they will tend to migrate less.

The above heuristics are designed to balance accuracy of a true fast-flux domain name together with the sensitivity needed to prevent legitimate, benign domain names used by content distribution networks which use DNS load balancing techniques that may appear fast-flux-like [8, 9].

A domain *whitelist* for exclusions is then checked to ensure that legitimate operations using fast-flux-style techniques for load balancing are not falsely tracked. Once all of these criteria are met, the domain is added to the list of domains for ATLAS to check for new IP address records via DNS A record queries. Fast-flux domains that appear to be novel are shared with the anti-spam community via the SURBL project [15].

We find that most, if not all, fast-flux domain names use wildcard DNS techniques to match any hostname under the domain name. Because of this, we can maintain a list of domain names and ignore the hostname portions, which greatly simplifies analysis and data storage.

2.2 Measuring Fast-Flux Botnets

Queries are made against a small number of local, caching DNS servers using standard UDP-based DNS query libraries. We found that we had to stagger the query order to prevent throttling by the servers in some cases we hit attack thresholds in some of the servers we used.

The back-end query architecture uses a timer to schedule domain queries, storing the results of A-record queries in one hour working sets before logging them. Duplicate IP addresses are pruned in the hour's set to ensure minimal data explosion. All results are logged with the first timestamp of the observation for the hour, in addition to the ASN where the IP address is located and the country code via an updated geo-location mapping service.

ATLAS is able to expire domains that become inactive by looking for a failure to update new IP addresses for a period of 24 hours. This keeps the working set pruned to a minimum and can identify, with reasonable confidence, when a domain became inactive or simply "parked" on a common website. A typical active working set for ATLAS at present (summer, 2008) is approximately 400 domain names. This number is similar to the number of active fast-flux domain names being tracked by the FluXOR project, another community fast-flux activity tracking service which appeared concurrent to our work [3].

3 Fast-Flux Botnet Behaviors

The system introduced in the previous section is in operational use since several months and used to observe a wide range of fast-flux botnets. We now look at the data collected by ATLAS in the timeframe of late January, 2008, until the end of May, 2008. In this time period ATLAS identified 928 distinct fast-flux domain names and collected 15,080,044 IP addresses to domain name mappings, which also includes repeated mappings for an IP address to a domain name at different times. Part of our in-depth analysis was performed on one day's data (May 30, 2008).

In the following we provide an overview of fast-flux botnet activity on the Internet as measured by our DNS mining tools in ATLAS. We aim to provide insight into the lifetimes of common fast-flux domain names, botnet sizes and activities. We break the data down into five main categories: *discovery*, *lifetimes*, *membership*, *visibility*, and *distinct botnets*. All five of these categories highlight how fast-flux botnets operate and what kinds of activities they host.

3.1 Fast-Flux Domain Features

During our measurement period, ATLAS was able to discover 928 qualified fast-flux domain names. Of these, the majority (76%) used `.com` as the top level domain (TLD), with `.cn` (14.3%) and `.net` (5.6%) the next most popular TLDs seen. In total, we saw eleven TLDs used in fast-flux operations, as shown in Table 1. This is a wider TLD distribution that was reported earlier by Holz *et al.* [4], which covered measurements from July and August, 2007. Infected hosts appearing in our measurements were found in broadly distributed countries and autonomous systems around the world, similar to the findings of Holz *et al.* [4].

These results, when compared to earlier results [4], suggest that the awareness campaigns about fast-flux to DNS registrars is having an effect [6]. This broader distribution of TLDs used in fast-flux reveals that attackers have to go to new locations to find friendly registrars. While this makes remediating the problem through registrars more challenging, it is a key vulnerability for fast-flux operators.

To assess the efficacy of our methods to discover and track new fast-flux domains, we looked at the time between a domain name registration or creation by a registrar and the first appearance of it in ATLAS' fast-flux data logs. We performed this analysis by collecting *whois* data on all of the domains, which shows the domain's creation date, and our full data set from January, 2008, until the present. We

Number of Observations	Observed TLD
705	com
133	cn
52	net
7	uk
3	info
3	in
2	us
2	org
2	kg
2	biz
1	ph

Table 1: Cumulative number of top level domain (TLD) names for distinct fast-flux domain names tracked by ATLAS during this data collection period.

then measured the interval between the stated domain name inception date and the time it took for ATLAS to begin measuring the botnet associated with it.

We find that the average interval between the domain's creation and the first appearance of it in ATLAS is just over four weeks at 28 days. The longest appearance is 236 days, or over 33 weeks, while the shortest interval is under 1 day. Based on this comparison, we find that 90 (9.7%) of the domains were created and seen in ATLAS less than 7 days later, and 154 (16.6%) were seen activated as fast-flux domains within one month of their registration.

The two main reasons that may cause this wide range in the intervals between the domain's creation and its first appearance in ATLAS are due to possible poor visibility by ATLAS and the use of "sleeper" domains, which have a significant delay between their inception and use in a fast-flux operation. The ATLAS email spam feed, used for malicious code detection and spam URL analysis, receives hundreds of messages an hour from a variety of sources, but clearly lacks a global scope that you would find in an anti-spam company with globally deployed sensors. Based on long term malcode and phishing campaign analysis, we usually find that the ATLAS email collection system receives representative spam samples within a day of their appearance on the Internet. Thus, a more likely scenario is the use of domains that have been registered, perhaps in bulk, and are activated later by the botnet operators. This assumption is also supported by a deeper manual analysis of a specific fast-flux botnet named *Asprox*/*Damnec* which activates new domains on demand.

3.2 Active Lifetimes

We define the *active lifetime* of a fast-flux domain as the time interval between the first data captured by ATLAS and the time of its termination as an inactive domain. We assume that the reason for inactivity is domain name suspension by a registrar. In our experience, we find that registrars are made aware of a fast-flux domain name and, following a successful investigation, they may suspend its activities. The time to which the domain is suspended depends on how aggressive the anti-malware alerting service is, as well as the responsiveness of the domain name registrar. Domain activity may be suspended through techniques such as *parking*, where it is mapped to a small number of IP addresses serving up registrar controlled content, or the domain may be locked by the registrar and the nameserver mappings deleted, ceasing any further resolution to the botnet.

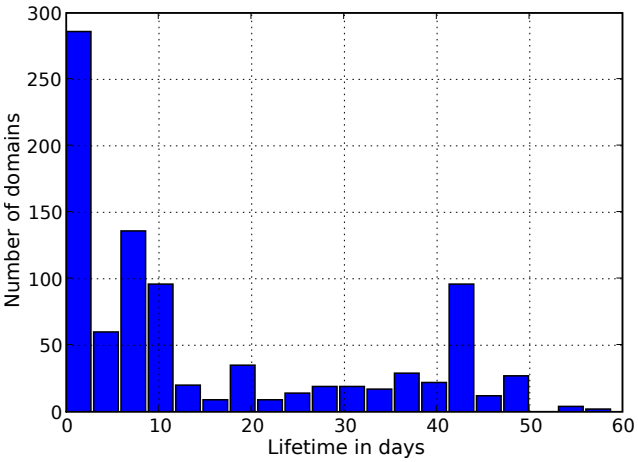

Figure 1: Distribution of lifetime in days for all monitored fast-flux domains used in our study.

Figure 1 depicts the distribution of lifetime in days for all the fast-flux domains we monitored. The average active lifetime of a fast-flux domain seen in ATLAS is 18.5 days. The longest three domains that we tracked in this timeframe were the Storm Worm hosted phishing site `ibank-halifax.com` [7], active for 60 days, `armsummer.com`, active for 59 days, and `croptriangle.com`, active for 57 days. Over one third of all of the domain names we monitored in this timeframe – nearly 400 – were active for less than a week.

3.3 Fast-Flux Botnet Membership

Fast-flux botnet DNS mining may be an effective way to estimate how large a given botnet is from the outside, based on measurements from a completely different angle compared to previous approaches. The drawback of DNS min-

Domain	Hosts	Lifetime (days)
ibank-halifax.com	100,379	59.95
armsummer.com	14,233	58.80
boardhour.com	11,900	54.92
swimhad.com	11,719	56.85
thickour.com	11,711	56.85
croptriangle.com	11,648	56.88
systemsuggest.com	11,136	50.96
minuteabove.com	11,134	50.96
momentten.com	11,123	50.96
spokewatch.com	11,110	50.96

Table 2: Cumulative botnet sizes in unique IP addresses for the largest fast-flux botnet domain names tracked by ATLAS during the data collection period.

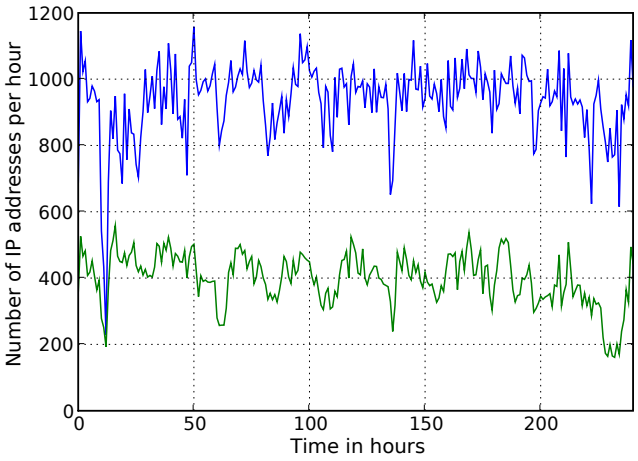

Figure 2: Number of IP addresses per hour (upper) and number of unique IP addresses per hour (lower) for one of the domains used by Storm Worm (ibank-halifax.com).

ing is the fact that hosts on RFC 1918 private address space will not be advertised and we can thus not include them in our size estimations. Based on over four months of cumulative data, from late January until the end of May, 2008, we found that the average fast-flux domain had a cumulative total of 2,683 distinct IP addresses associated with it. The largest networks were `ibank-halifax.com` with 100,379 unique IP addresses seen in this time period, `armsummer.com` with 14,233 IP addresses associated with it, and `boardhour.com` with 11,900 separate IP addresses associated with it in this time period. An overview of the ten largest fast-flux botnets by domain name is provided in Table 2, together with the lifetime in days.

Note that the sizes of these botnets may be grossly inflated due to IP address churn on broadband and consumer connections. This is especially true for some of the more long-lived fast-flux domains such as `ibank-halifax.com` and `armsummer.com`, where the same infected customer may be assigned a new IP address multiple times a day under some conditions.

We also looked at the *promiscuity* of an individual IP address, or how many fast-flux domains has it been associated with. All of the data collected by our system was aggregated by IP address and analyzed to determine which domain names were ever associated with an IP address. We found that the average IP address had 14 fast-flux domain names associated with it, while the maximum seen was 392 domains, which is probably due to a *domain parking* IP address for frozen domain names.

3.4 Estimating Fast-Flux Botnet Sizes

A unique characteristic of the Storm Worm botnet is that the time to live of answers to DNS queries is very low (typically 0) and thus repeated DNS queries lead to the discovery of new IP addresses that are part of the botnet. Therefore we obtain an estimate of the current size of the fast-flux part of this botnet by continuously monitoring a domain which belongs to this botnet.

Figure 2 depicts the number of (unique) IP addresses observed per hour over time for the period between January 18, 2008 at 2:00 pm, and January 28, 2008 at 1:00 pm for the domain `ibank-halifax.com` which belongs to the Storm Worm botnet. During this ten day period, we observed typically between 800 and 1,100 IP addresses per hour as results of the DNS queries. From the set of all IP addresses, between 300 and 500 were unique, i.e., this number of IP addresses is *active* for the Storm Worm botnet during the given hour and used for hosting content with the help of fast-flux techniques. The figure also shows dynamic, recurring patterns within the number of observed IP addresses. At the beginning of the measurement period we observe a short interval were almost all returned IP addresses are unique. This could be due to internal restructuring within the botnet since we observed a similar behavior later on a couple of times.

Using a crawler-based method [5], we observed between 25,000 and 35,000 IP addresses per hour active in the Storm botnet. This indicates that only a fraction of the total number of bots - approximately 1% - is used for the fast-flux part of the network. There are mainly two reasons for this behavior: First, hosts with an address from the RFC 1918 private address space can not be used for fast-flux hosting and can thus not be observed with the help of DNS mining. Second, not all bots with a globally unique IP address meet the criteria of fast-flux hosting, e.g., only a short uptime, small bandwidth, or other criteria. Nevertheless, DNS min-

ing provides us with a technique to estimate the size of the fast-flux network from a completely different viewpoint.

If a fast-flux botnet uses higher time to live values for answers to DNS queries, e.g., if the answer is valid for 600 seconds, then our technique may lead to imprecise size estimations: since we can then only observe a small number of IP addresses, we would underestimate the total number of bots within the fast-flux botnet.

3.5 Identifying Distinct Fast-Flux Botnets

Our observations indicate that fast-flux botnets usually have multiple domain names associated with their activities. This provides the botnet with even more resilience to takedown or disruption. However, it may not always be apparent which domain names are used for the same botnet. Based strictly on our data set of IP address and domain name mappings, we were able to identify how many distinct botnets are using fast-flux techniques in practice.

Using a simple set-based approach as described in Equation 1, we can do all-pairs analysis to identify the same hosts across different domain names, indicating the same botnet in use. This analysis assumes that the same IP addresses would be used from the botnet across multiple domain names and that no partitioning is occurring. We performed this analysis on 428 active fast-flux domain names from ATLAS using a 24 hour data snapshot from May 30, 2008.

$$\frac{net1 \cap net2}{net1 \cup net2} \quad (1)$$

Net_1 and Net_2 are the set of unique IP addresses for different domain names within the 24 hour time period under study. As the equation above reaches 0, the two domains are from entirely different botnets. When the equation reaches 1, the two domains have identified the same botnet. We focused strictly on the complete intersection of the two networks, where the above equation was 1.0.

Selecting only these pairs of domains that have a full intersection, the data yielded 26 distinct clusters, indicating that at most 26 distinct fast-flux botnets were being tracked by ATLAS. This represents the upper bound of the active fast-flux botnets we are tracking in ATLAS. There could be partial overlap in our data set where the botnet is using the different domains found in different clusters if the botnet is advertising overlapping but different subgroups of the hosts.

Using post-facto analysis (the domains had almost all been deactivated by the time the analysis was done) we were able to infer the purposes of the clusters. This analysis was facilitated by online investigation reports from teams such as CastleCops [1] and PhishTank [14]. The clusters grouped together into the following broad categories:

- 1 cluster was used to advertise online casino sites
- 1 cluster was used as a website for an "enlargement" product
- 4 clusters were used for malware distribution
- 10 clusters were used for "pharmacy" products
- 13 clusters were using in phishing attacks

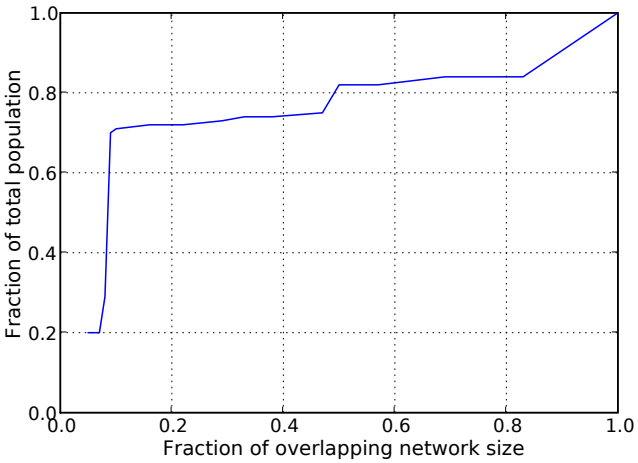

Figure 3: Fast-flux network overlap in fraction of total size vs. the fraction of all domain pairs analyzed in this study. This data indicates that a significant number of domain name pairs have some membership overlaps, most likely due to our partial visibility into a single botnet using multiple domain names. This data was from a 24 hour time period on May 30 2008 to minimize IP address churn for the same infected machine.

Note that some of the clusters had more than one purpose, including a combined phishing and malware distribution network. The total number of clusters is therefore larger as the number of distinct botnets we have identified.

We found a handful of domains had between 10% and 83% overlap when we did this set-based analysis. This result could be due to multiple infections by different botnet tools on the same hosts or by partial advertisements by the botnet. We discount this possibility, however, because we find that the proxy software used by fast-flux operations interferes with other, similar malcode from different authors. Therefore, a more likely explanation is that the botnet is partitioned for multiple uses, meaning that the fast-flux domains do not map to the entire botnet but only a subset of the bots. The partial overlap result would appear if some individual IP addresses are in more than one partition, but not all are in the same partitions.

4 Discussion

The data presented in this paper provides us with a rich picture of fast-flux botnets and their use in today's Internet. We showed that some fast-flux domains can remain registered and active for weeks or longer, and that some may be associated with tens of thousands of IP addresses over time. We also showed that a large pool of domain names that are associated with fast-flux have been registered but remain dormant for more than one moth before their use. Also, a variety of web-based attacks can be committed using fast-flux techniques, including fraud and impersonation as well as malcode infections. To the best of our knowledge, this is the first such study of real world fast-flux botnet practices.

As expected, a number of botnet clusters were used for phishing attacks. Phishing teams such as the Rock Phish group have been using fast-flux techniques for over two years with great success [11], but other groups may be mimicking their techniques. We also anticipated that a number of botnets were distributing malware using fast-flux hosting techniques, including the Storm Worm botnet [5]. We did not expect to see so many fraudulent pharmacy sites hosted in fast-flux botnets, however.

As registrars become aware of fast-flux operations, they are increasingly taking measures to combat the technique. As such, we are seeing more global top level domains (TLDs) used than ever. As we noted above, we see more TLDs in use than were previously seen by Holz *et. al* [5]. We expect this trend to continue as fast-flux users must find more registrars who are unable to thwart their registration and use, and they will gravitate to TLD operators who lag behind best security practices. We have been working with the ICANN SSAC efforts on combatting fast-flux operations at the registrar level [6].

Another interesting finding in our data is that the number of "sleeper" fast-flux domains is so high. Only 16.5% of all fast-flux domains we tracked during this time period were activated as fast-flux domains within one month of their creation, meaning that a large number – over 80% – lie dormant for longer. If we can identify what sleeper domains are present and likely to be used in fast-flux operations in the future, operators can proactively block them using DNS blacklisting techniques or monitor them for signs of active fast-flux use. This may be accomplished using nameserver mining, `whois` data analysis, or registration tracking to determine related domains. Our data shows that most fast-flux domains are dormant for more than one month before their use, meaning that *proactively* identifying future fast-flux domains has the potential to disrupt the online fraud and crime that is hosted in these networks.

At present, we do not know of any authoritative, unequivocal, and well-maintained list of fast-flux domain names and networks, making a comparison between ATLAS and other systems impossible. We suspect that ATLAS has not been able to identify *all* of the fast-flux domains acting on the Internet, but we are unable to determine how many it is missing. As such, we cannot measure the completeness of coverage afforded by our approach. However, we have not seen any challenges to our base assumptions that lead us to discover new fast-flux domain names advertised in spam or revealed in malcode analysis. Therefore, the techniques described in this paper can – if necessary – be scaled to observe more fast-flux domain names.

Our technique for identifying active botnets is based solely on IP address collections. An underlying assumption is that no partitioning takes place within the botnets, and that all IP addresses are advertised equally. Based on discussions with other researchers, this may not be the case. They have been able to look at the connections from the infected client host to the central content server and have discovered about 10 active fast-flux botnets working at any one time. Both numbers, our original estimate of 26 fast-flux botnets and the more conservative estimate of 10 active fast-flux botnets, are much lower than we would have expected if fast-flux hosting were growing in popularity. This result suggests that while it has proven itself effective, fast-flux techniques are still not a significant enough gain for attackers to set up and maintain. Alternatively, less sophisticated botnet herders lack access to fast-flux DNS management tools and are unwilling to build their own.

Finally, we have no visibility into the actual size of some of the botnets being studied here, most notably the Rock Phish botnet. To our knowledge, no one has been able to successfully determine how many hosts are in the Rock Phish botnet, but several research groups have estimated the size of another well known botnet using fast-flux, the Storm Worm botnet, at approximately 30,000 hosts active per hour [5]. Our measurement results in Section 3.4 reveal approximately 1% of the botnet in any given hour, indicating that DNS mining can only provide a limited overview of the total number of bots in a given botnet. This is presumably due to the fact that the botnet screens for candidate hosts that should be included in the DNS results and not all bots match these criteria, thus preventing us from seeing these bots with our technique. It is unclear to us how consistent this botnet visibility will be across different botnets. Fast-flux botnet size estimates based on DNS data mining does not appear to be a consistent method to measure the size of a botnet.

5 Conclusions and Future Work

Fast-flux hosting techniques have become increasingly popular on the Internet as a way to monetize a botnet, providing "bullet-proof" hosting that is resistant to endpoint takedown. Investigations into the true content hosting system

can be slowed down, with fast-flux techniques giving the attackers an additional buffer of anonymity. Furthermore, the resilience can be increased since taking down fast-flux botnets is hard.

We have found that continuous data mining of fast-flux DNS records can yield insights into the size and scope of a fast-flux operation. We have also found that we can begin to disambiguate the true number of active fast-flux botnets by collecting a significant amount of IP addresses associated with a set of domain names and, through direct, all-pairs data analysis, identify which ones are linked and part of the same fast-flux botnet.

Our analysis highlights two methods by which we can combat fast-flux botnet hosting using DNS techniques. This is the key weakness for fast-flux hosting and the best avenue for addressing the problem. The first technique requires the involvement of a local registrar or global top level domain registrar to de-activate the domain. The second technique can be applied locally using local authoritative DNS servers.

Most importantly, we have shown that a vast majority of all domain names that display fast-flux behaviors are dormant for a significant period of time. This gives the DNS registrar community tremendous opportunities to proactively monitor or disrupt fast-flux operations when related domains can be identified. At present the DNS operator community is beginning to examine how they can best combat fast-flux botnet activities [6], and we hope that this finding from our data can yield methods to disrupt future fast-flux operations.

An additional technique that local network security administrators can apply as a counter-measure to the fast-flux problem is through the use of a local DNS server. DNS blacklisting services such as SURBL, the Malware Domain List or the Malware Domain Blocklist [2,10,15] can be used to block all access to malicious domains including fast flux domains. We are presently exploring how to share our data with such services for such purposes.

Our system presented in this paper, based on the ATLAS honeypot data storage system, provides us with a flexible data collection and reporting mechanism. However, the fast-flux data collection system is in its early stages and a number of limitations and omissions are apparent.

One minor optimization that we could make in the short term would be to query additional, geographically dispersed DNS servers to obtain more results. We repeatedly hit the query cache of the limited number of DNS servers we are using, so this change may yield additional results. Alternatively, ATLAS could incorporate passive DNS replication data feeds to gather a diverse set of observations [17, 18].

The current ATLAS fast-flux measurement system is missing several pieces of information. Some of these were design choices, and some of these were oversights in the original designs. Many of them would be valuable future directions to pursue within ATLAS or in another fast-flux measurement system.

The first piece of missing data in our fast-flux data archive is the name of the registrar used by the domain. This would be useful in helping to identify problem registrars and assisting in remediating via domain disabling. Long term, it may help identify effective fast-flux domain disruption techniques used by various registrars, and any "hot spots" that have emerged.

We are also missing information about the malcode family behind the botnet. It would be useful if this information was included so that we could assess the size of the botnets over time and track them by malcode family, not just by domain name. This additional info may be useful to disambiguate fast-flux botnets and to look at their use and partitioning.

As discussed in Section 3.5, we did post-facto analysis on the botnets to discover what kind of content they were hosting. This was sometimes difficult as the domains had been disabled and we had to rely on third party reports for this information. Again, this would enable tracking of fast-flux botnets by purpose, not just by domain name.

Finally, we do not keep track of the NS records for the domains. If we did so, we could identify *double flux* scenarios, where the nameservers are also hosted in a botnet [16]. Also, we could identify common nameservers and possibly mine them for additional domain names which may be fast-flux related. At present, we are not aware of any measurements of how popular double-flux methodologies are.

Acknowledgments

The authors would like to thank Robert Danford for his discussion with the authors while developing the screening heuristics, and the ATLAS engineering team for improvements in the fast-flux monitoring prototype developed by one of the authors. Danny McPherson and John Kristoff provided helpful feedback during the preparation of this manuscript.

References

[1] CastleCops. `http://www.castlecops.com/`, 2008.

[2] DNS-BH – Malware Domain Blocklist. `http://malwaredomains.com/`, 2008.

[3] FluXOR. `http://fluxor.laser.dico.unimi.it/~fluxor/`, 2008.

[4] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling. Measuring and Detecting Fast-Flux Service Networks.

In *Proceedings of the 15th Annual Network & Distributed System Security Symposium (NDSS)*, 2008.

[5] T. Holz, M. Steiner, F. Dahl, E. Biersack, and F. Freiling. Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm . In *Proceedings of the First USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '08)*, 2008.

[6] ICANN Security and Stability Advisory Committee (SSAC). SAC 025: SSAC Advisory on Fast Flux Hosting and DNS, 2008.

[7] Jeremy. Storm Worm Fast Fast Flux domain "superdrugtesting.com", 2008. `http://www.sudosecure.net/archives/36`.

[8] K. L. Johnson, J. F. Carr, M. S. Day, and M. F. Kaashoek. The measured performance of content distribution networks. *Computer Communications*, 24(2):202–206, 2001.

[9] B. Krishnamurthy, C. E. Wills, and Y. Zhang. On the use and performance of content distribution networks. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pages 169–182, 2001.

[10] Malware Domain List. `http://www.malwaredomainlist.com/`, 2008.

[11] T. Moore and R. Clayton. An Empirical Analysis of the Current State of Phishing Attack and Defence. In *Proceedings of the 2007 Workshop on the Economics of Information Security (WEIS'07)*, 2007.

[12] A. Networks. `http://atlas.arbor.net/`, 2008.

[13] Phillip Porras, Hassen Saidi and Vinod Yegneswaran. A Multi-perspective Analysis of the Storm (Peacomm) Worm, 2007. `http://www.cyber-ta.org/pubs/StormWorm/`.

[14] PhishTank. `http://www.phishtank.com/`, 2008.

[15] SURBL. `http://www.surbl.org/`, 2008.

[16] The Honeynet Project. Know Your Enemy: Fast-Flux Service Networks, 2007. `http://www.honeynet.org/papers/ff/fast-flux.pdf`.

[17] F. Weimer. Passive DNS Replication. In *Proceedings of 17th Annual FIRST Conference on Computer Security Incident Handling*, 2005.

[18] B. Zdrnja, N. Brownlee, and D. Wessels. Passive Monitoring of DNS Anomalies. In *Proceedings of the 4th Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, pages 129–39. Springer Berlin, 2007.