Understanding Adversarial Strategies from Bot Recruitment to Scheduling

Wentao Chang[1], Aziz Mohaisen[2], An Wang[1], Songqing Chen[1]

[1] George Mason University
[2] The University of Central Florida

Abstract. Today botnets are still one of the most prevalent and devastating attacking platforms that cyber criminals rely on to launch large scale Internet attacks. Botmasters behind the scenes are becoming more agile and discreet, and some new and sophisticated strategies are adopted to recruit bots and schedule their activities to evade detection more effectively. In this paper, we conduct a measurement study of 23 active botnet families to uncover some new botmaster strategies based on an operational dataset collected over a period of seven months. Our analysis shows that different from the common perception that bots are randomly recruited in a best-effort manner, bots recruitment has strong geographical and organizational locality, offering defenses a direction and priority when attempting to shut down these botnets. Furthermore, our study to measure dynamics of botnet activity reveals that botmasters start to deliberately schedule their bots to hibernate and alternate in attacks so that the detection window becomes smaller and smaller.

1 Introduction

Botnets are collections of networks of infected machines (aka bots) that are widely used to carry out a variety of malicious activities as instructed by a botmaster. As a result, botnets are notoriously known as one of the primary attack and threat vectors utilized against critical infrastructures and services in activities that include distributed denial of service (DDoS), spam distribution, phishing, scanning and network exploration, among others. Such malicious activities utilize vulnerabilities in existing protocols, and capitalize on their power to disturb large services.

The advent of botnets is often associated with vandalism. However, recent years have witnessed the rise of other uses of botnets, including “hacktivism” [1] and “botnet-as-a-service” [5]. Botnets have been used as a mean of promoting political ends, such as targeting political and ideological opponents, stealing precious data from their networks, or for bringing their networks down. OpIsrael [2], DarkSeoul [3], and OpUSA [4] are recent prominent examples of hacktivism, where ideas and political beliefs influenced botnet-based cybersecurity events and driven them. The rise of such a direction has facilitated a thriving ecosystem guided by economical profit in what has been coined as botnet-as-a-service [5]. In such a model, botnets are designed to be “rented” easily to underground users, where botmasters are reportedly making large sums of money in underground marketplaces [6].

Understanding botnets through analyses and measurements has been a goal in the research community since their arrival. Such analyses is geared towards understanding attacks, guiding defenses, and helping with bots containment and disinfection by chronologizing their lifecycle. The first and foremost step in the lifecycle of botnets is to recruit and manage a dedicated pool of bots. Such step is done by either recruiting a new group of bots via infection or by renting a network of already infected machines in the botnet-as-a-service marketplace. Once recruited, botmasters utilize their bots (in a given botnet) to launch attacks. Considering them as valuable resources, botmasters want to maximize the return on investment by launching as many attacks as possible without being detected by a defender. To this end, bot scheduling is a critical aspect of botnet management, and further insights into how botmasters schedule their bots could potentially unveil patterns in this ecosystem that could lead to 1) better understanding of botnets, and 2) guide defenses.

In this paper, we advance the state-of-the-art by analyzing the botmasters strategies in recruiting and managing bots based on a large workload collected from more than 300 Internet vantage points across the globe covering 23 most active botnets for a continuous 7 months. Our study reveals several interesting and previously unreported recruitment strategies by botnets in the wild. A highlight of the new sophisticated techniques adopted by modern botmasters includes (c.f. §5 for implications):

- Our geographical analysis shows that most dedicated bots reside in a small number of countries and organizations. This provides some helpful insights for defenses. For example, pushback models [8–10] of defenses can be guided by this insight in deploying routing-based monitoring closer to the sources of the attack.
- Bots recruitments are not purely random but rather targeted with per-family unique characteristics. Further analysis shows that different botnet families have their unique per-family characteristics (i.e., affinity). This insight can be utilized in postmortem host cleaning. For example, upon taking down a botmaster, cleaning disconnected bots becomes a challenge, and knowing the affinity would guide efforts of disinfection and cleaning.
- Bots are not always active. Instead, they are recruited and used with a clear alternation pattern, and longer periods of hibernation in between. This can effectively minimize the detection window of detection tools and thwart them. This pattern and trend can be utilized to guide defenses: a defense that utilizes the distribution of activity window of bots is more likely to detect an attack earlier than one that uses a fixed (and potentially large) time window.

To the best of our knowledge, *many of the recruitment strategies uncovered in this study are novel and not reported before*, making them interesting in their own right. As we are still unfolding the use of the recruitment strategies, we suggest to leverage such insights to devise new defense and mitigation schemes. While there has been a large body of literature on the problem (e.g., [11–20]; c.f. §6), all of the prior work draws conclusion on behavior of botnets by analyzing a single botnet (or a limited number of them). To our knowledge, this is the first study that tries to understand recruitment and scheduling patterns by performing a meta-analysis over a large number of botnets and associated behavior.

2 Data and Collection Methodology

Prior work on botnet measurements have mainly focused on their taxonomy and classification by analyzing botnet behavior and common characteristics, including architecture, command and control (C2), communication protocols, and evasion techniques [21, 22]. Such efforts have mainly been done via passive measurement or infiltration, and are usually focused on specific botnets. Different from these approaches, our work relies on data provided by the monitoring and attribution unit a DDoS mitigation company, with partnerships of traffic sharing with a large number of major Internet service providers across the globe. The dataset is previously utilized by Wang *et al.* [7] for analyzing trends in DDoS attacks.

2.1 Collection Methodology

The unit constantly monitors Internet attacking traffic to aid the mitigation efforts of its clients, using both active and passive measurement techniques. For active measurements and attribution, malware families used in launching various attacks are reverse engineered, and labeled to a known malware family using best practices. A honeypot is then created to emulate the operation of the reverse-engineered malware sample that belongs to a given botnet and to enumerate all bots across the globe participating in that particular botnet.

As each botnet evolves over time, new generations are marked by their unique (MD5 and SHA-1) hashes. Traces of traffic associated with various botnets are then collected at various anchor points on the Internet, via the cooperation of many ISPs all over the world, and analyzed to attribute and characterize attacks. The collection of traffic is guided by two general principles: 1) that the source of the traffic is an infected host participating in a botnet attack, and 2) the destination of the traffic is a targeted client, as concluded from eavesdropping on C2 of the campaign using a live sample.

2.2 Botnet Families

There are 23 known botnet families in the wild captured in our dataset. Those botnet families are (using their publicly known names assigned by antivirus vendors [7,23]) `Aldibot`, `Armageddon`, `Asprox` , `Blackenergy`, `Colddeath`, `Conficker`, `Darkcomet`, `Darkshell`, `Ddoser`, `Dirtjumper`, `Gumblar`, `Illusion`, `Myloader`, `Nitol`, `Optima`, `Pandora`, `Redgirl`, `Storm`, `Tdss`, `Torpig`, `Waledac`, `Yzf` and `Zeus`. From the dataset multiple botnets are identified for each family, and each botnet is potentially owned by different botmasters. By tracking bots' temporal activities,the monitors of the company generate a log dump every hour. There are 24 hourly reports per day for each botnet family. The set of bots or controllers listed in each report are cumulative over past 24 hours. The 24-hour time span is counted from time stamp of last known bot activity and time of log dump. The log covers the period from 08/29/2012 to 03/24/2013, a total of 207 days.

2.3 Caveats and Comparisons

While the dataset we use in this paper is comparable in size to other dataset previously used in the literature, it provides a timely insight into the recent state of botnet operations, as opposed to the state of botnets many years ago. Furthermore, the efforts of identifying malware that is used for operating a botnet family provide high fidelity: the techniques involve a combination of dynamic and static analysis utilizing deep understanding and reverse-engineering of the studied families.

We note that some of the hosts infected by the studied malware families may not be included in our data for a few reasons. For example, they may not be included if they do not participate in an attack against a monitored resource, or if they do not contact the C2 server of the studied family. However, we believe that those hosts are of less interest, since they are isolated and do not contribute to the potential attack activity of the botnet. They do not contribute to the recruitment and scheduling aspects studied in this paper, and their disinfection and cleaning is a secondary issue to this study.

3 Bots Recruitment

During about 7 months of our data collection and analysis, over 2 million unique bots across 23 malware families are identified in our dataset. The purpose of our botnet study is to gain insights into active botnets' nature so that security analysts and experts can effectively take down existing botnets by disinfection, or prevent benign hosts from infections for suspicious sources (as previously done in other work; e.g., Stone-Gross *et al.* [24] and Gu *et al.* [25]).

One of the primary properties of bots that interests us most is their physical location and how the location shifts in regions across different phases of the botnet's life cycles. We are also interested in whether a certain subset of bots play a more critical role than others in the bots recruitment. Our conjecture is that bots recruitment as a process may not be purely random but rather targeted with per-family unique characteristics, and the geographical distribution analysis confirms our conjecture.

In this section we examine all known bots in our dataset by mapping their IP addresses to a list of countries where they reside, and identifying organizations that such IP addresses belong to. We perform the mapping of the IP addresses using a highly-accurate commercial grade geo-mapping dataset by Digital Envoy (Digital Element services [26]), which provides—besides the country—the individual *city*, and *organization* of each queried IP address.

Addressing NAT effect. Dynamic IP addresses and NAT constitute a significant portion of the Internet [24, 27], preventing a one-to-one mapping between bot and IPs. Addressing NAT is a challenging problem, which falls out of the scope of this work. However, we follow a similar approach to [28] to minimize its impact on our findings. While the NAT effect leads to undercounting bots, such undercounting is corrected by churns, resulting in overcounting, due to DHCP. Thus, in preprocessing, and for each botnet, we aggregate the different bots with unique IP addresses that have distinct patterns into unique bots. For passive IP churn using DHCP at the ISP level, we aggregate the unique IP addresses over shorter hourly time periods to minimize the potential of DHCP churn [28, 29]. A recent study [30] shows that the distribution of dynamic IP addresses is not uniform but rather biased towards regions or ISPs. By analyzing IP addresses at the country and organization level, we conclude that the estimated number of bots should be considered as a lower bound, thus minimizing the impact on our recruitment findings.

3.1 Bots Country Preference

Figure 1 shows the heat map of bots' geographical distribution. The darker the color of the country is, the more bots are found to be located in that country. We can see that those two million bots are widely spread all over the world with several harder-hit areas in the darker green regions.

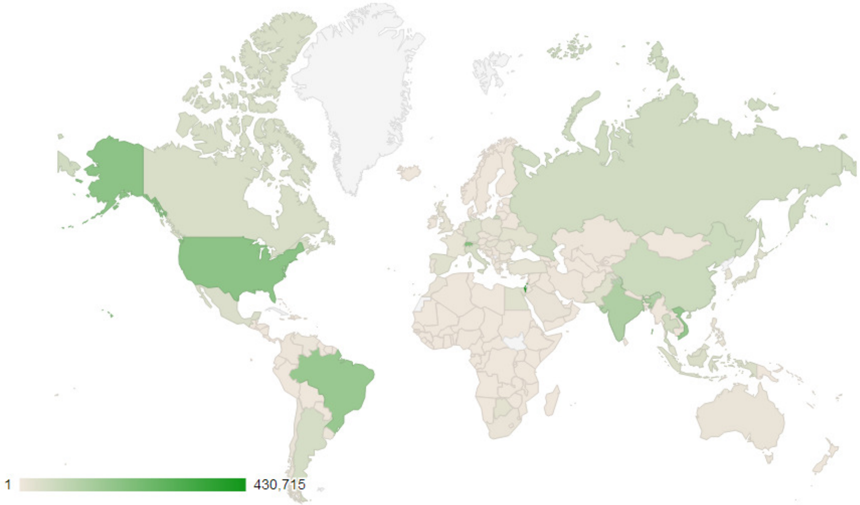

Fig. 1. Geographical distribution of bots

Table 1. Top 10 countries with most bots.

Country Name	Number of bots
Israel	430,715
Switzerland	207,386
U.S.A	187,483
Vietnam	172,066
Brazil	163,983
India	121,949
China	66,742
Russia	59,158
Thailand	54,135
Argentina	48,760

In Table 1, we list the top 10 countries with most bots. These ten countries together host 66.3% of all bots in our dataset (i.e., 1,512,377 out of 2,280,389). One of the surprising findings in the table is that Israel and Switzerland lead the rank of all countries, with a combined share of 28.0% of bots, despite that they are neither far-flung countries in area nor large countries in population. A reasonable explanation is that our data provider might parterner with major ISPs that have dominant existence in Europe, or during 7-month collection period botnets from aforementioned countries are involved in active campaigns. Drilling down to the per-family bots distribution, some per-family unique characteristics of the country preference are revealed:

Bots preferential attachment. We notice that for most families their bots are concentrated in a few preferred regions, and those preferred regions tend to vary significantly across different families. In general, the majority of botnet families have tangible bots existence in the top countries we listed in Table 1. Interestingly every family also has their own set of preferred countries. Take the family `Optima` as an example, we find that the top 5 countries for `Optima`'s 343,524 bots are Israel (21.0%), India (14.5%), United States (12.6%), Switzerland (9.4%), and Brazil (6.1%). The first three countries in the list contain more than 48% of the total bots for Optima, and all these 5 countries can also be found on the top 10 overall list—although in different ordering.

Country preferential attachment unveils activity correlations. We observe that for some families bot's geographical preference is somehow preserved at different stages of botnet activities. For example, the activity curve of `Optima` highlights three sudden spikes dated at 10/18/2012, 10/29/2012 and 11/10/2013, respectively, which could be attributed to 3 active campaigns that were launched around that time frame with a large number of bots participation. Thus, we examine all bots involved in those spike events to expose their preferential attachment. It is evident that 1st and 3rd spike events are correlated, because the overwhelmingly majority of bots in these two spike events originate from Israel, Switzerland, United States, Botswana, and Canada. The high resemblance of bots distribution between the 1st and 3rd spikes implies that they are very likely to be two stages of the same attack. On the contrary, by examining the bots from

2nd spike we find that its bot distribution is significantly different: 37.8% bots (23,499) originate from India and 16.1%(10,027) originate from United States. This can be explained that the second spike is potentially another independent campaign launched by a different botmaster.

Another family shows strong activity correlation is Dirtjumper. The top 5 countries of Dirtjumper's 818,452 bots are Israel (37.5%), Switzerland (15.2%), United States (10.5%), Brazil (9.9%), India (4.4%). We find two correlated spike events for this family because they both exhibit the same geographical distribution patterns. The majority of bots in these two spikes originate from Israel, Switzerland and the United States. The other two countries (Brazil and India) in the top 5 list did not contribute much to the spike events.

Local botnets. The geographical distribution of Illusion does not conform to the all bots distribution chart. Pakistan, which has an unnoticeable presence in the overall country ranking of most bots, contributes a dominantly large number of bots to Illusion. This finding strongly suggests that Illusion either prefers to or gain privileges to recruit most of its bots from Pakistan. Similarly, bots that belong to Pandora show a significantly biased existence in Mexico and Thailand.

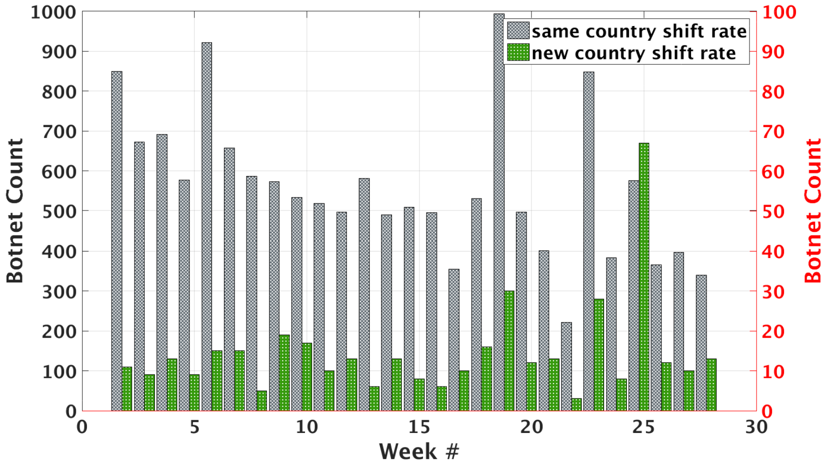

Fig. 2. Bots shift patterns for Conficker

Mobility within preferred regions. We explore how bots of each family shift over time. In this analysis we aim to identify whether the newly arrived bots originate from the same country or from different countries. The results show that the majority of bots only shift within their preferred regions. Left y-axis in the Figure 2 represents the shift rate of Conficker within the same country, while right y-axis represents the shift rate across countries. For Conficker the arrival rate of bots from the same country is 20 to 40 times higher than that of bots from a different country. This localized shift pattern can be further validated by our bots alternation analysis in Section 4.2. The set of active bots controlled by the same family has a strong location affinity.

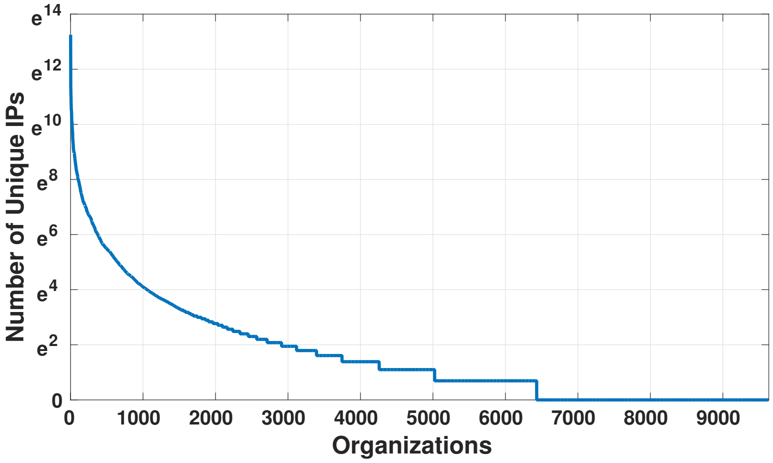

Fig. 3. Number of unique IPs in each organization

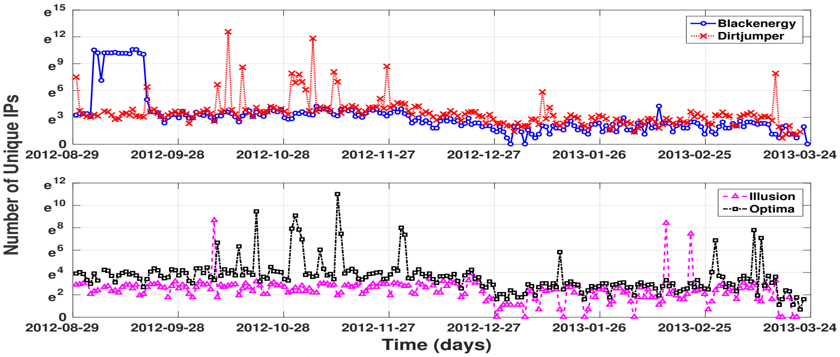

Fig. 4. Activities from top 2 organizations

3.2 Bots Organization Preference

Using the IP mapping dataset, we are also able to identify the organizations that own the IP addresses hosting the infections, which represent the bots we studied in our dataset. Across all botnet families, with 2,280,389 unique IPs in our dataset, we enumerated 9,633 different organizations. We sort the organizations in descending order of the number of associated bots and display them in Figure 3. The x-axis represents 9,633 organizations, and the y-axis denotes the logarithm number of unique IPs. We learn from the figure that the distribution is heavy-tailed, and that the top 20% of organizations in our study contain more than 90% of all bots. With the impact factor of different organization size considered, we can still draw the conclusion that bots organizational distribution is not purely random, certain organizations become easy targets for bots recruitment or at least are comparably tardy for the mitigation of botnet infection. We list the top 10 organizations with most bots in Table 2. While some of the organizations are service providers, we notice that some of the organizations are businesses that do not provide Internet services to customers–thus highlighting an interesting dimension in such a distribution.

Table 2. Top 10 organizations with most infected hosts

Organization Name	Bots
ilan ISP (Israel)	568,200
Switch Swiss Education and Research Network	224,810
Viettel Corporation (Vietnam)	80,872
The Corporation for Financing Promoting Tech. (Vietnam)	79,341
Telebahia (Brazil)	66,154
African Network Information Center (Uganda)	60,403
National Internet Backbone (India)	58,283
Cogent Communications	45,264
Independent Electricity System Operator (Canada)	44,599
China Telecom	37,119

Positive organization preference. Similar to the country-level analysis, we explore organizations-preference. The result is shown in Table 3. We find that 14.3% of the bots of `Aldibot` are located in Canada-based organizations, for example, Rogers Cable Communications inc. and Bell Canada. India's "National internet backbone (NIB)" is another favorite organizations for 4 botnet families: `Colddeath`, `Darkcomet`, `Darkshell` and `Yzf`. For these 4 families, the number of bots from NIB is in absolute dominance compared to other organizations. We also find that the organization with most bots for `Ddoser` is telecom argentina stet-france telecom s.a., and `Nitol` is found to have a tendency to recruit bots from organization te-as. Similarly the Australia-based organization, Telstra Pty Ltd, owns a large majority of bots for `Torpig`. Organizations with substantial bots existence for `Zeus` are quite a few, and `Zeus` is the botnet family that is discovered from most organizations in our study. The total number of organizations accounted for `Zeus` botnet activity is as high as 5,541—possibly because `Zeus` is a mass-market credential stealing botnet.

Organization preferential attachment unveils activity correlations. We choose 5 botnet families that were very active during the 7-month data collection period as our analysis candidates. They are `Blackenergy` , `Conficker`, `Dirtjumper`, `Illusion` and `Optima`. After a closer look, we realize that the majority of bots that contribute to 4 of those 5 families come from the top 2 organizations we listed in Table 2, `Conficker` is the only exception due to the multi-variants nature of the family. Bots owned by `Conficker` are widely distributed over 3,522 different organizations and thus it has no clear culprit organizations. To this end, we plot the activity curve of those 4 families considering only bots coming from these 2 organizations. As Figure 4 shows, the bots from those 2 organizations stay hibernated most of the time during our observation period. The timing of their sudden wake-up coincides with the peak events of the botnet families. This evident behavior strongly suggests the bots

Table 3. Organization-level bots preference

Aldibot	Rogers Cable Communications Inc.	1,056
	Bell Canada	675
	Google Inc.	521
	Uninet, S.a. de C.v. (Mexico)	513
	Cox Communications Inc.	429
Blackenergy	ilan ISP (Israel)	179,619
	Viettel Corporation (Vietnam)	62,250
	Switch Swiss Education and Research Network	61,739
	The Corporation for Financing Promoting Technology (Vietnam)	55,337
	Independent Electricity System Operator (Canada)	13,687
Conficker	China Telecom	32,477
	Data Communication Business Group (Taiwan)	27,036
	Telefonica de Argentina	23,013
	TM Net (Malaysia)	21,659
	Telecom Italia	18,975
Dirtjumper	ilan ISP (Israel)	306,144
	Switch Swiss Education and Research Network	124,158
	Telebahia (Brazil)	50,310
	African Network Information Center (Uganda)	27,522
	independent electricity system operator	23,943
Illusion	ilan ISP (Israel)	7,731
	Cyber Internet Services Ltd. (Pakistan)	5,778
	Switch Swiss Education and Research Network	5,564
	African Network Information Center (Uganda)	2,531
	National Internet Backbone (India)	2,076
Nitol	Telecom Egypt	6,539
	National Internet Backbone (India)	2,417
	Tata Teleservices Ltd	533
	China Telecom	319
	China Networks Inter-exchange	304
Optima	ilan ISP (Israel)	71,177
	Switch Swiss Education and Research Network	31,750
	National Internet Backbone (India)	22,003
	African Network Information Center (Uganda)	16,582
	VNPT corp (Vietnam)	7,476
Pandora	Uninet, S.a. de C.v. (Mexico)	1,284
	Cyber Internet Services Ltd. (Pakistan)	1,241
	African Network Information Center (Uganda)	1,041
	National Internet Backbone (India)	991
	San Paulo Research Foundation (Brazil)	522
YZF	National Internet Backbone (India)	2,291
	Nepal Telecommunications Corporation	462
	African Network Information Center (Uganda)	415
	Tata Teleservices Ltd	295
Zeus	Turk Telecommunications	8,833
	Viettel Corporation (Vietnam)	8,663
	Maroc Telecom (Morocco)	7,841
	Cox Communications Inc.	7,142
	The Corporation for Financing Promoting Technology (Vietnam)	6,903

from these two organizations are coordinated to perform attacks on purpose. It is very likely bots in these two organizations are zombies, dedicated machines controlled by remote attackers to conduct cyber attacks, which explains the fact that they are infected by multiple instances of botnet families.

4 Bots Scheduling Strategies

In this section, we perform an in-depth study of botnets' dynamics to expose the latest bots scheduling strategies with three aims in mind. First, we closely monitor the bots dynamics in 7-month observation window and conduct a lifespan analysis of all bots in our dataset to understand their involvement in botnet activity. Second, we dive deep into exposing unique activity patterns of short-lived bots, which strongly implies a deliberate action of bots alternation and re-occurrence when scheduling bots. Last but not the least, we find that a substantial number of bots are recruited and reused by more than one botnet families, and we are interested in behind-the-scenes reasons why those bots are favored by botmasters in recruitment and scheduling.

4.1 Bot Lifespan Analysis

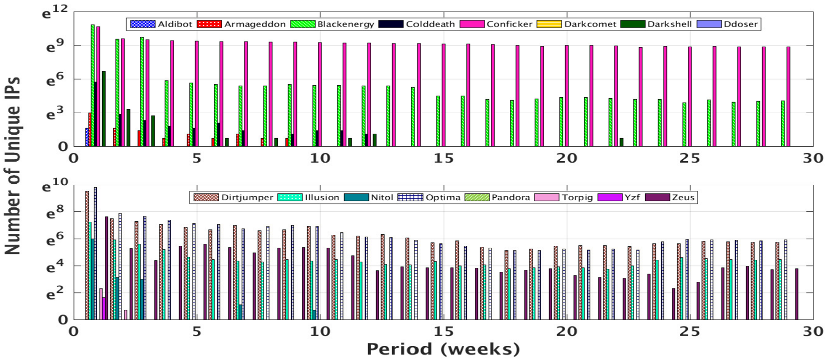

Fig. 5. Bots lifespan analysis

The lifespan of a bot is an important indicator of bots' involvement in the botnet activity. To this end, we conduct a weekly pattern analysis for every bot in our dataset to understand their presence and evolvement. Given the confined context of our dataset and the fact that mitigation techniques might have already been in place to take down active bots, it is not far-fetched to speculate that many bots in the dataset are short-lived, and this will be presented as those bots active in week i become dormant in week $i + 1$. Our speculation is confirmed in the analysis result that in general less bots are found with longer lifespan. The number of bots with various lifespan is depicted in Figure 5 and the short-lived statement is held true for all botnet families in our dataset.

Besides short-lived bots, our analysis reveals a small but steady group of bots that stay active for an extended period of time, in many cases several weeks. We call them "always-on" bots. For a subset of botnet families we observe, there exist a very small set of "always-on" bots (i.e., single-digit) whose lifespan cover the entire data collection period. One exception is `Conficker`, one of the largest known computer worm infection [13]. The number of "always-on" bots for `Conficker` is approximately 4,400 unique IP addresses, by far the largest among all families we studied. The fact that `Conficker` is a very well-represented botnet with multiple variants in existence, making it difficult to remove from end users' computers as effectively as done for other families.

The role played by bots with a longer lifespan is possibly different from that played by bots with a shorter lifespan. A potential role such bots play include a shadow botmaster (to mitigate failure) and to serve as a dedicated bot. From a defense perspective, it takes precedence to shut down those long-lived bots than others when mitigating large-scale botnet activity. Our bot's lifespan analysis over the large-scale dataset provides a firsthand information of what bots defenders should target to remove in priority. However, the assumption that the consequence of mitigation is the sole reason for many bots' short lifespan is doubtful, because this does not explain the existence of in-negligible amount of long-lived bots, and the number of bots for each family does not always decline linearly as the lifespan increases in Figure 5. This raises suspicion that those bots might be deliberately hibernated by the botmasters as a countermeasure to thwart detection efforts, thus we will further investigate whether bots hibernation are scheduled purposely and how in following subsections.

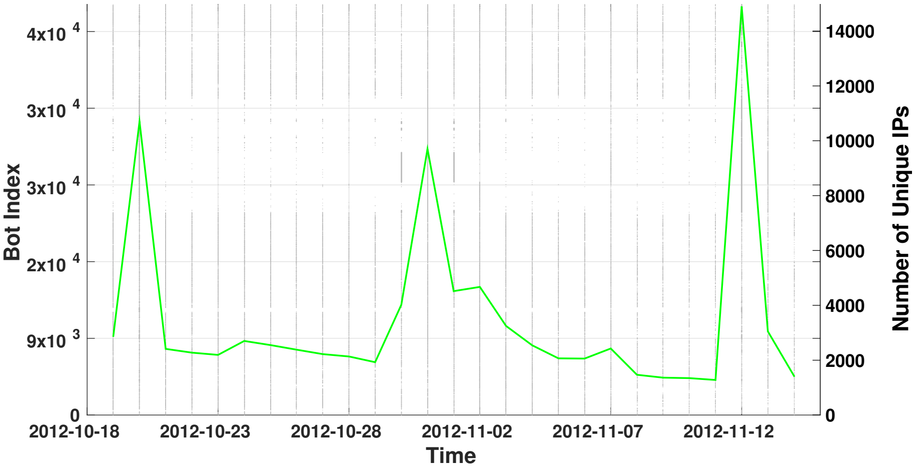

Fig. 6. Fast bot recruitment for Optima

4.2 Bots Fast Recruitment and Active Bots Alternation

As shown earlier, the lifespan of most bots is usually short. Thus, to perform large-scale attacks, botmasters need to recruit a large number of bots. In our study, we observe that the majority of bots only appear once in our traces. The fast bots recruitment for `Optima` is depicted in Figure 6, the dotted vertical line in the figure marks whether corresponding bots from the left y-axis are active or not at the given day, and the curve in the right y-axis represents the total number of active bots at a given day. Note that the bots from the left y-axis are sorted in the ascending order of their IP addresses. During those 3 peak events tens of thousands of bots suddenly become active and disappear after the completion of a major campaign. Our bots organizational preference analysis in Section 3.2 confirms that the abrupt surge of bots is primarily due to a temporary recruitment of bots from other botnets to launch a highly intensive attack—including the borrowing of dedicated bots found at some organizations with vulnerable defenses.

Diving deep into the composition of active bots we find a strong level of bots alternation. We learn from our lifespan analysis that, for all families even in their seemingly stable periods, in which the total number of simultaneously live bots does not change much over one or several weeks, the majority of bots still remain active for less than one week. Analysis results show that the new bots activated by botmasters compensate the loss of old bots. It is unlikely this unnatural harmony of bots alternation is merely due to the effectiveness of bots mitigation, and we believe it is a side effect of countermeasures to defeat defense that botmasters voluntarily utilize, which is to iterate bots from their pool of slave bots to complicate the process of take-down mitigations.

4.3 Bots Re-occurrence Patterns

Excluding a small number of "always-on" bots, majority of bots are only observed once except for a number of bots that consistently reoccur in the 29-week observation window. To understand the root cause of those bots' uncommon behavior, we divide bots into groups by their occurrence count. Note that we count the weekly occurrence for all bots, and if the same bot occurs in two or more consecutive weeks, we only count it as one occurrence. As Figure 7 shows, a common patten across all families is that the larger the occurrence count is, the less number of bots there are. Another pattern is that a proportional relationship exists across families between bot count per occurrence group and total bot count in their respective family. Given the current data in hand we're still investigating what are the criteria used by botmasters to select bots to occur more than once in their lifecycle.

We also conduct a per-bot re-occurrence analysis to measure their re-occurrence distance. The term "re-occurrence" in this context is used to describe bots that are active in week i, become dormant in week $i+1$, but are brought back to life in week $i+j$ (where $j > 1$). In this sense, we define j as the re-occurrence distance. Because our observation window is only 7 months rather than years, the chance that the same bot is taken down through disinfection, but becomes re-infected by the same malware, is low. Therefore the impact of false positives is negligible for this analysis. Bots with longer re-occurrence distance could be attributed to either dedicated bots or zombie machines that existing mitigation efforts fail to completely disinfect. As Figure 8 shows, as the re-occurrence distance increases the number of bots declines near linearly except for 2 families, `Aldibot` and `Dirtjumper`. A closer look at these 2 families in the figure reveals that they both own a relatively large number of bots with long re-occurrence distances compared to others. `Dirtjumper` has over 1000 bots with the re-occurrence distance as high as 18 and 21, while `Aldibot` has a sudden surge at the re-occurrence distance of 16. The re-occurrence pattern associated with both families highlights the lack of response to active hosts in malicious activities, where resources utilized in botnets stay for long time infected—thus reused after a relatively long time.

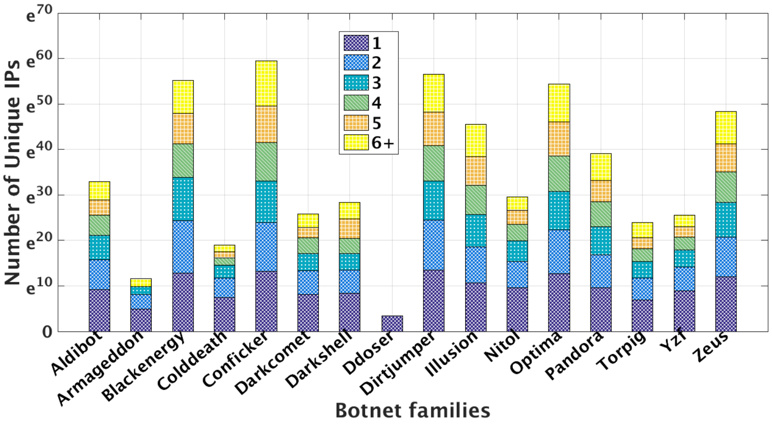

Fig. 7. Number of occurrence for bots

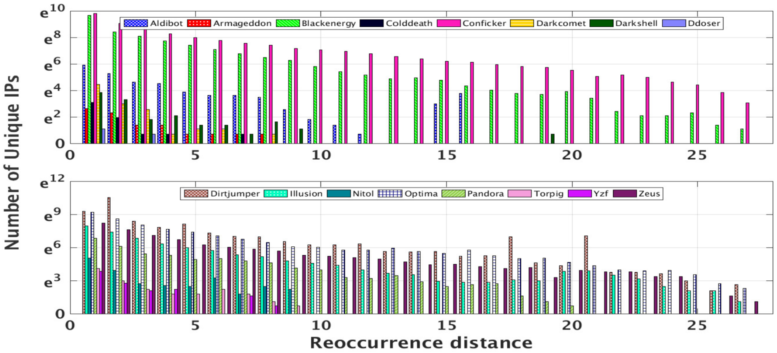

Fig. 8. Bots re-occurrence distance

4.4 Reused Bots Analysis

Another interesting observation we revealed when examining the bots in our dataset is the presence of a large amount of reused bots – the bots belonging to multiple botnet families, which could be due to multiple infections or due to using paid infrastructure (i.e., pay-per-install) [31]. In reality, while one would expect normally a single host to be utilized for a single malicious activity of a certain type, we hypothesize that hosts with multiple infections are often utilized by multiple botnets to perform various types of malicious activities and to participate in many campaigns. This hypothesis, validated through our analysis results below, is of particular interest to the security community for multiple reasons. While the understanding of reuse may shed light on the genealogy of malware and their associations, it most importantly highlights the differential roles that various types of bots play in the cyber underground world, where reusable bots may play much bigger roles in cyber attacks launched via botnets. Host machines that serve as reused bots are more threatening, which means they could be further leveraged to participate in other campaigns. Also, having various infections may highlight those hosts tendency not to disinfect from a compromise over a long period of time (honeypot is an exception). Such nature of reused bots indicates that they are long-living and possibly a good candidate to serve as nodes for botnet C&C channels, or "always-on" piece of the botnet infrastructure. Thus, by correctly identifying those bots, cyber defenders may leverage such information to effectively defend against cyber attacks by guiding efforts of disinfection in a feasible way.

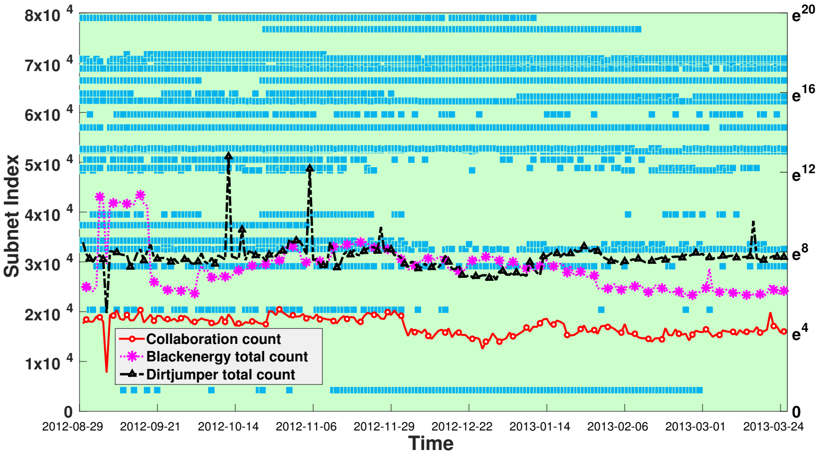

Fig. 9. Reused Bots. (An illustration using the case of Blackenergy and Dirtjumper botnet families.)

4.4.1 Reused bots scheduling strategies Among the 2,280,389 unique IPs identified as infected hosts, 320,340 IPs, accounting for roughly 14.0% of total bots, are confirmed to be reused by at least two families during the 7 months. The average reuse ratio across different families varies significantly, where statistics of reuse unveil that some families tend to have a higher reuse ratio of its bots than other families. The number of reused bots and their shift pattern are an important metric to measure the collaboration efforts among different botnet families. It would be crucial to recognize the correlation between the magnitude of reused bots and the overall activity level of individual families. Thus, in this subsection we inspect the interactions and collaborations via reused bots between two specific families over time.

We choose the `Blackenergy` and `Dirjumper` since they are two of the most active families in our dataset. In our data pre-processing step we aggregate the /24 subnets to group various IP addresses, thus reducing the total number of addresses in our analysis. Figure 9 shows how the activity level of those two families correlates with their collaboration. The x-axis and the left y-axis mark the date and the subnet index, respectively. The right y-axis represents the total number of active subnet in log scale. The dotted horizontal line represents whether the corresponding subnet from left y-axis is reused by both families across time. From the figure it is evident that tens of different subnets are reused by these two families. The three curves represent the number of active bots for `Blackenergy`, `Dirtjumper`, and the reused subnets, respectively. We observe that when there're no spike events, the number of active bots for `Blackenergy` and `Dirtjumper` is in the same order of magnitude, and the number of reused subnets fluctuates roughly in the same pace as that of those two families. Even when the spike events of either families occur, there are no noticeable surges of reused subnet count. These similar behaviors between these two families are less likely to be only coincidental, therefore we infer that the huge number of new bots in spike events is recruited for one-time use of specific campaigns, while reused subnets are treated as backbone of botnet activities. This subtle relationship implies that a master-slave relationship might exist between them.

5 Insights to botnet detection and defense

In previous sections our in-depth analysis of a large botnet dataset expose some new bot recruitment patterns and various sophisticated botnet scheduling strategies. By understanding the trending techniques adopted by bot-masters, security researchers could devise more effective defense mechanisms to detect and mitigate botnet attacks.

Prediction of bots origins based on their family. In Section 3 county preference study, we learn that many different bot families have their own per-family unique characteristics regarding recruitment preference. For example, some bot families tend to concentrate in a few preferred countries, and some bot families only exist in one or two dedicated countries. These recruitment preference patterns persist during our 7-month data collection period, thus with confidence we could predict participating bots' origins for a campaign launched via certain bot family. This new capability will definitely boost defense to effectively identify attacking traffics from normal ones. For example, when the host machine detects itself under DDoS attacks from bots in `Illusion`, one defense mechanism is to activate a specific firewall rule to block all connections from Pakistan to alleviate the system burden, since we know `Illusion` family exhibits local botnet characteristic. Also, as noted earlier, pushback models, such as the work of Ioannidis *et al.* [8], Chen *et al.* [9] and Kang *et al.* [10], can benefit from this insight in determining where pushback and filtering (at the Internet-level) are done.

Vulnerable organizations. Our organization preference analysis shows that bots distribution is not purely random, but rather targeted. Some organizations contain significantly more bots than others, for perhaps having more vulnerable machines, which makes them easy targets for bot recruitments, or for being backbone network service providers. With the list of vulnerable organizations, security researchers can perform a more thorough security audit of host network environment, and even urge third party organizations to improve their security guarantees. This insight can be useful in postmortem host cleaning: upon the take-down of C&C channels, it would be useful to clean hosts. Knowing the affinity of botnets and certain malware families to certain organization would guide such cleaning efforts.

Bots ordering of severity. In Section 4 we discussed the latest bots scheduling strategies we identified from the dataset. In addition to static attributes such as bots' geographic information, these dynamic attributes of bots increased the complexity of dis-infection efforts. Due to the limited resources and time-sensitive defense requirements to recover from attacks, it calls for a meaningful ordering of bots per their severity. We believe long-lived bots could potentially serve as shadow botmasters (to mitigate failure) or command and control channels for botnets. Thus when mitigating a large-scale botnet activity, it takes precedence to shut down long-lived bots. We also learnt that some short-lived bots are bots shared in a dedicated pool and coordinated by bot-masters to participate in botnet activities.

It is essential to enumerate all short-lived bots in the pool and shut down them all in once, if possible. Those short-lived bots with apparent alternation or re-concurrence patterns will need to be assigned higher level of severity. Reused bots are another important and interesting finding in our study. Different botnets collaborate to some extent to perform malicious actions via reused bots. Taking down one reused bot would mitigate threats from multiple botnets, thus reused bots should be given higher level of severity as well when cyber defenders plan their dis-infection efforts.

6 Related Work

Previous research efforts on botnet measurements have mainly focused on the taxonomy and classification of botnets by analyzing botnet behavior and common characteristics, such as architecture, command and control channels, communication protocols, and evasion techniques. These efforts have mainly been done via infiltration [18], as done by Bacher *et al.* [32, 33] or passive measurement, as done by Rajab *et al.* [34]. Many early studies looked at the most common IRC-based bots relying on a centralized control, as shown by karasaridis *et al.* [35] and Barford *et al.* [36]. Later on numerous new botnets began to use http-based C&C channels and leverage the more stable P2P based communication architecture, per Wang *et al.* [17] and Holz *et al.* [37], to mitigate failure due to centralization. Other work focused on in-depth case study of individual botnet families, as done by Binsalleeh *et al.* [19], Andrade *et al.* [14] and Shin *et al.* [38].

Recent work focused on Internet or large scale measurement study of network traffic to develop methods for revealing more properties of botnets, such as their size [29] and activeness [24, 38]. In our work, we analyze the bots' static properties and dynamic behaviors from a different angle, focusing on botmaster strategies behind the scenes. By conducting a large-scale measurement study of bots activity from multiple well-known botnet families, we uncover several new bot recruitment and scheduling strategies. To the best of our knowledge, some of our findings, such as bot recruitment preferences and bot resources scheduling, are not reported before. Similar to our study, Chang *et al.* [39] conducted a measurement study of a commercial dataset to reveal the latest botmasters' strategies. While both works share a common theme of understanding botnets utilized by DDoS attacks, the focus of each work is different. In particular, Chang *et al.* outlined measurement highlights collectively, focusing on botnet collaboration observations. On the other hand, we study the bot recruitment at both the country and organization levels, and bot scheduling – fast recruitment, bot alternation, bot recurrence – from where we can infer the strategies of botmasters. Furthermore, our work focuses on bot dynamics, which is a topic of independent interest.

Although we utilize the same dataset used by Wang *et al.* [7], our approach of data analysis is completely different from theirs: the findings and contributions of our work do not share any common ground with theirs. Wang *et al.* utilized the dataset to understand state-of-the-art of DDoS attacks, while we utilize it to understand the source and tool used for the attacks; botnets. For this purpose they analyzed the geo-distribution of attack sources for many DDoS attacks, the temporal patterns and collaboration trends between botnet families to launch attacks, etc. On the other hand, whereas we focus on a meta-study of various botnets, our goal is to reveal both static attributes and dynamic patterns of all botnets from 23 known families to understand the constantly advancing strategies adopted by botmasters.

7 Conclusion

Botnets today are responsible for most large-scale attacks on the Internet. Thus, it is essential to understand their latest behavioral traits for insight into defenses. In this paper, we have performed a measurement study of bots activity from 23 known botnet families for about 7 months. By conducting a series of in-depth analysis of bots' static properties and dynamic behaviors, we have uncovered that today botmasters have adopted several new strategies to recruit and schedule bots. As we still investigate the potential consequence of those strategies in bots recruitment, we suggest to leverage such insights to devise new defense and mitigation schemes.

8 Acknowledgment

We appreciate constructive comments from anonymous referees. This work is partially supported by an ARO grant W911NF-15-1-0262, a NIST grant 70NANB16H166, NSF grants CNS-1524462 and CNS-1643207, and NRF grant 2016K1A1A2912757 (Global Research Laboratory).

References

1. Wikipedia. Hacktivism. `http://bit.ly/1kM2Vos`, 2014.
2. Wikipedia. Operation Israel. `http://bit.ly/1noDUlI`, 2014.
3. Symantec Security Response. Four years of darkseoul cyberattacks against south korea continue on anniversary of korean war. `http://bit.ly/1fbGlFm`, June 2013.
4. Bank Info Security. Opusa threatens banks, government. `http://bit.ly/1kP3Urt`, May 2013.
5. Paul McDougall. `http://ubm.io/MtCSr7`Microsoft: Kelihos ring sold 'botnet-as-a-service'. September 2011.
6. Marissa Vicario. Four ways cybercriminals profit from botnets. Nov 2010.
7. A. Wang, A. Mohaisen, W. Chang, and S. Chen. Delving into internet DDoS attacks by botnets. In *IEEE DSN* 2015.
8. J. Ioannidis and S. Bellovin. Implementing pushback: Router-based defense against ddos attacks. 2002.
9. Y. Chen, Y. Kwok, and K. Hwang. MAFIC: adaptive packet dropping for cutting malicious flows to push back ddos attacks. In *ICDCS*, 2005.
10. M. Kang and V. D Gligor. Routing bottlenecks in the internet: Causes, exploits, and countermeasures. In *Proc. of ACM SIGSAC*, 2014.
11. Wikipedia. Carna botnet 2014.
12. M. Starr. Fridge caught sending spam emails in botnet attack 2014.
13. M. Thomas and A. Mohaisen. Kindred domains: detecting and clustering botnet domains using dns traffic. In *Proc. of WWW*, 2014.
14. M Andrade and N Vlajic. Dirt jumper: A key player in today's botnet-for-ddos market. In *WorldCIS* 2012.
15. L. Song, Z. Jin, and G. Sun. Modeling and analyzing of botnet interactions. *Proc. of Physica A*, 390(2):347–358, 2011.
16. Z. Li, A. Goyal, Y. Chen, and V. Paxson. Towards situational awareness of large-scale botnet probing events. *IEEE TIFS*, 6(1):175–188, 2011.
17. P. Wang, S. Sparks, and C. Zou. An advanced hybrid peer-to-peer botnet. *TDSC*, 2010.
18. C. Cho, J. Caballero, C. Grier, V. Paxson, and D. Song. Insights from the inside: A view of botnet management from infiltration. *LEET* 2010.
19. H. Binsalleeh, T. Ormerod, A. Boukhtouta, P. Sinha, A. Youssef, M. Debbabi, and L. Wang. On the analysis of the zeus botnet crimeware toolkit. In *IEEE PST* 2010.
20. J. Caballero, P. Poosankam, C. Kreibich, and D. Song. Dispatcher: Enabling active botnet infiltration using automatic protocol reverse-engineering. In *Proc. of ACM CCS*, 2009.
21. C. P. Lee D. Dagon, G. Gu and W. Lee. A taxonomy of botnet structures. In *Proc. of ACM ACSCA*, 2007.

22. L. Jing, X. Yang, G. Kaveh, D. Hongmei. Botnet: classification, attacks, detection, tracing, and preventive measures. *JWCN*, 2009.
23. A. Mohaisen and O. Alrawi. AV-meter: An evaluation of antivirus scans and labels. In *Proc. of DIMVA* 2014.
24. B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna. Your botnet is my botnet: analysis of a botnet takeover. In *Prof. of ACM CCS* 2009.
25. G. Gu, R. Perdisci, J. Zhang, W. Lee, et al. Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. In *Proc. of USENIX Security* 2008.
26. Digital Envoy. Digital element services.
27. Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov. Spamming botnets: signatures and characteristics. In *SIGCOMM* 2008.
28. M. Maertens, H. Asghari, M. van Eeten, and P. Van Mieghem. A time-dependent sis-model for long-term computer worm evolution. In *Proc. of IEEE CNS* 2016.
29. M. Rajab, J. Zarfoss, F. Monrose, and A. Terzis. My botnet is bigger than yours (maybe, better than yours): Why size estimates remain challenging. In *Proc. of USENIX HotBots*, 2007.
30. Y. Xie, Fang Yu, K. Achan, E. Gillum, M. Goldszmidt, and T. Wobber. How dynamic are ip addresses? In *ACM SIGCOMM CCR*, 2007.
31. J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring pay-per-install: The commoditization of malware distribution. In *Proc. of USENIX Security*, 2011.
32. P. Bacher, T. Holz, M. Kotter, and G. Wicherski. Know your enemy: Tracking botnets. 2005.
33. P. Baecher, M. Koetter, T. Holz, M. Dornseif, and F. Freiling. The nepenthes platform: An efficient approach to collect malware. In *Proc. of RAID* 2006.
34. M. Abu Rajab, J. Zarfoss, F. Monrose, and A. Terzis. A multifaceted approach to understanding the botnet phenomenon. In *IMC*, 2006.
35. A. Karasaridis, B. Rexroad, and D. Hoeflin. Wide-scale botnet detection and characterization. In *Proc. of USENIX HotBots* 2007.
36. P. Barford and V. Yegneswaran. An inside look at botnets. In *Proc. of Malware Detection* 2007.
37. T. Holz, M. Steiner, F. Dahl, E. Biersack, and F. C Freiling. Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm. In *USENIX LEET*, 2008.
38. S. Shin and G. Gu. Conficker and beyond: a large-scale empirical study. In *Proc. of ACM ACSAC* 2010.
39. W. Chang, A. Mohaisen, A. Wang, and S. Chen. Measuring botnets in the wild: Some new trends. In *ACM ASIACCS*, 2015.