Exploring Botnet-based DDoS Strategies Behind the Scenes: Analysis, Modeling, and Tool Development

Although enormous efforts have been made continuously from both academia and industry to defend against DDoS attacks, the fact that DDoS attacks are still very prevalent on the Internet today suggests that there are still a lot of challenges yet to overcome in order to build better defenses. DDoS attacks, by nature, are difficult to defend against since it is hard to know when the attack is launched, where the attacking machines are from, how many attacking machines are involved, how long the attack will last, etc.

Today most Internet DDoS attacks are attributed to various botnets, which enable a thriving ecosystem guided by economical profit in what has been coined as "Botnet-as-a-Service" (BaaS). Many of today's botnets are designed and developed to be loaned easily to third parties. Reportedly botnet controllers can make a large amount of money by loaning the service in the mature underground market. For such botnet-based commercialized DDoS attacks, are these attacks totally random? how do the attackers, e.g., bostmasters, manage the attack army? If there are some patterns in these attacks, can we learn them so that we can utilize to improve the existing defenses?

To answer these questions, a fine view of today's botnet behaviors and their contribution to various DDoS attacks is necessity. Such information is often difficult to obtain. In this project, based on a real-world DDoS attack workload, we propose to explore botnet-based DDoS attacks collectively in order to investigate the underlying strategies of attackers, in the hope to help DDoS mitigation and defenses. The workload that we will use in this project contains a total of *confirmed* 50,704 different Internet DDoS attacks directly observed in a 7-month period (from 2012 to 2013) from over 370 major ISPs all over the world. These attacks were launched by 674 botnets from 23 different botnet families with a total of 9026 victim IPs belonging to 1074 organizations in 186 countries. For this purpose, we propose three research tasks in this project. The ***intellectual merit*** of the project includes three major research thrusts, summarized as follows:

- **Measurement and Analysis.** In this research thrust, we first plan to conduct an in-depth study of the botnet-based DDoS characteristics, including attack protocols, attack frequency, attack duration, attack magnitude, attack source, targeted organizational and geographical distributions, intra- and inter-family collaborative attacks, etc. We will focus on new findings with the hope that such findings and insights can be directly or indirectly leveraged for DDoS defenses.
- **Modeling of DDoS Patterns.** In this research thrust, we will explore models for characterizing patterns of DDoS attacks. Previous studies often make outdated or simplified assumptions about DDoS attacks (e.g., IP addresses are random) . As we find from the preliminary workload analysis, today botmasters use very sophisticated techniques to control their bot army for attacks. For example, bots are rotated periodically in one attack and different attacks. We aim to study the underlying strategies of botmasters and construct models to characterize them whenever possible. We expect such models can allow us to accurately predict some properties of DDoS attacks and aid mitigation.
- **Tool Development**: We utilize the models for DDoS patterns and their prediction to develop a DDoS simulation tool. There are a number of DDoS simulators and traffic generators available. However, most of them are based on some high-level or simplified models for the attacker IP origins, attack duration, and attack magnitude, etc. Thus, the traffic generated loses many properties exhibited by real-world attack traffic. We propose to build our findings into a simulator that hopefully can provide more fine-grained details for the community to research on DDoS attacks.

The potential ***broader impacts*** of the proposed projects are as follows. (i) The proposed research will provide a better understanding of critical operational issues for one of the most prevalent security threats. (ii) The fundamental mathematical models and associated tool out of this research will help improve Internet

DDoS attack defenses. (iii) The developed techniques and tools will be available online for free and wide usage. (iv) Anonymized data used in this project, and all data artifacts from models and simulations in their entirety, will be shared with the research community in ways that will facilitate reproducibility of results. (v) the PIs will also enhance network and system security course teaching at SUNY Buffalo and George Mason University. The PIs will integrate the research results into these courses. The PIs will encourage qualified undergraduate, underrepresented and female students to participate in this project.

Keywords: DDoS, Botnet, Modeling, Simulation.

Contents

1 Introduction

Today, persistent Internet Distributed Denial of Services (DDoS) attacks are prevalent largely due to the ease of access to large numbers of compromised machines, collectively known as botnets. According to a report released in Oct. 2013 by Neustar, a DDoS mitigation provider [20], on average, "a DDoS attack is not detected until 4.5 hours after its commencement; and a further 4.9 hours passes before mitigation can commence". Furthermore, "it can cost an Internet reliant company $1 million before the company starts to mitigate the attack". More recently, and according to a report released in Feb. 2015 by Verisign on DDoS trends [22], the duration, intensity, and diversity of attacks are on the rise: a year-over-year analysis shows that the average DDoS attack size has increased by 245% in the fourth quarter of 2014, compared to the same quarter of 2013, with an average attack of 7.39 Gbps. Furthermore, the same report shows that all industry verticals are targeted by DDoS attacks.

Widely reported incidents of DDoS attacks further highlight their actual size and the devastating operational impact they can have on large online services on the Internet. For example, it has been recently reported that 3,000 open domain name service (DNS) resolvers were capable of generating 300 Gbps DDoS traffic [37], and taking down Spamhaus, a popular spam tracking service. Also, it has been recently reported that an amplification attack utilizing 4,529 network time protocol (NTP) servers was capable of generating a 325-400 Gbps of persistent attack traffic [70].

The rise of such attacks has enabled a thriving ecosystem guided by economical profit in what has been coined as "botnet-as-a-service" [48]. Botnets utilized for DDoS attacks have become a mainstream commodity in the cybercrime ecosystem. Reportedly, botnet controllers can make a large amount of money in the underground markets [71, 78]. Furthermore, today's botnets are not limited to sophisticated machines, like servers and personal computers: with the Internet of Things (IoT) promising to revolutionize our use of technology, recent DDoS attacks were reportedly observed from fridges [65], and other scanning activities used embedded devices, including IP monitoring cameras and security doors [77]. This evolution, while highlights new trends in technology, also calls for detailed investigations.

To this end, enormous efforts have been made continuously from both academia and industry to understand, model, and defend against DDoS attacks (c.f. related work). With ever-improving defenses, the attacking strategies are constantly changing as well, making DDoS attacks one of the most severe threats on the Internet. DDoS attacks, by nature, are challenging to defend against since it is hard to predict when an attack is launched, where the attacking machines are from, how many attacking machines are involved, and how long the attack will last. This is particularly true for the DDoS attacks used for "demonstration of skills" or when driven by non-commercial reasons (e.g., hacktivism).

1.1 Research Need

Most Internet DDoS attacks today are attributed to various botnets, and understanding their behavioral traits is key to defending against them. For such botnet-based DDoS attacks, various research questions are particularly intriguing. (i) Are those attacks totally random? (ii) How do attackers, e.g., bostmasters, manage their attack armies? (iii) Can we *predict* the attack origins, sizes, duration, start time, and magnitude? (iv) If there are some patterns in these attacks, *can we learn them and utilize that knowledge to improve existing defenses*? Measuring and understanding the latest attacking *strategies* is certainly key to effective defenses. However, most of the existing measurement and analysis on DDoS attacks are based on indirect traffic measures, such as backscatters, or traffic collected locally, such as in an ISP or by infiltration into a botnet. Both approaches, while valuable, result in inferences, rather than direct measurements. A large-scale and timely view of today's Internet DDoS attacks is missing due to the unavailability of data. This makes it harder, if not impossible, to research better and more effective DDoS defense schemes.

1.2 Research Approach

In this project, the main goal is to enhance our understanding of botnet-based DDoS attacks and seek to move the DDoS defense posture several steps forward using *data-driven analyses* of real-world DDoS attack workload. The workload is used in our preliminary study in [75, 26, 27], and was obtained from operational security efforts led by Team Cymru. With this data, we are able to (i) conduct this research, (ii) publicize the findings, and (iii) share derivatives of the dataset with interested parties (more details in the data management plan). The dataset is highly annotated, and contains a total of *confirmed* 50,704 different Internet DDoS attacks directly observed in a 7-month period (from 2012 to 2013) from over 370 major ISPs all over the world. These attacks were launched by 674 botnets from 23 different botnet families with a total of 9,026 victim IPs belonging to 1,074 organizations in 186 countries. To achieve the main goal of this project, we propose to conduct research activities in three thrusts, namely measurements and analysis, modeling of DDoS patterns, and tools building. We summarize those activities as follows.

- **Measurement and Analysis.** In this research thrust, we first plan to conduct an in-depth study of the botnet-based DDoS characteristics, including attack protocols, attack frequency, attack duration, attack magnitude, attack source, targeted organizational and geographical distributions, intra- and inter-family collaborative attacks, etc. We will focus on new findings with the hope that such findings and insights can be directly or indirectly leveraged for DDoS defenses.
- **Modeling of DDoS Patterns.** In this research thrust, we will explore models for characterizing patterns of DDoS attacks. Previous studies often make outdated or simplified assumptions about DDoS attacks (e.g., IP addresses are random) . As we find from the preliminary workload analysis, today botmasters use very sophisticated techniques to control their bot army for attacks. For example, bots are rotated periodically in one attack and different attacks. We aim to study the underlying strategies of botmasters and construct models to characterize them whenever possible. We expect such models can allow us to accurately predict some properties of DDoS attacks and aid mitigation.
- **Tool Development**: In this research thrust, we plan to utilize the models for DDoS patterns and their prediction to develop a DDoS simulation tool. There are a number of DDoS simulators and traffic generators available. However, most of them are based on some high-level or simplified models for the attacker IP *origins*, attack duration, magnitude, and *strategy*. Thus, the traffic generated loses many properties exhibited by real-world attack traffic. We propose to build our findings into a simulator that can provide more fine-grained details for the community to research on DDoS attacks.

We note that while some features obtained from our analysis to build our simulator are arguably easy to manipulate by adversaries, some others, like core strategies, rotation of resources to evade detection, attack duration (for successful attacks), affinity, attacker's network location, etc., are more difficult to manipulate. If successful, those findings would improve common practices used for simulating adversaries in the research community (e.g., compared to simplistic and theoretical adversary assumptions, as in [68, 41, 61, 86]. For further details, see §3 and associated remedies.

1.3 Broader Impacts of the Proposed Research

If the project is successful, we expect to have the following technical and societal impacts. (i) The research output through this project will enable a better understanding of the Internet DDoS attacks and potentially better mitigation and defense schemes. This can lead to much less economic loss and a more reliable and safer Internet environment. (ii) The tool we will develop through this project can be directly used by other researcher in the community for their botnet and DDoS research. For example, the simulator can be used to test the effectiveness of the new defense schemes. (iii) The research results and the developed tool will be

freely available online for wide usage. Along with the project's progress, we will make our results and the corresponding tool available online so that they can be widely commented and adopted. (iv) The research in this project will contribute to the curriculum development and student training in two institutions. The PIs will jointly encourage qualified undergraduate, underrepresented, and female students to participate in this research project at their respective institutes.

1.4 Comparison with previous measurement-based projects

A large body of the lietrature is on radiation and port scanning measurements [57, 80, 47, 25, 24] concerned with a single network (Tier-1 ISP [47], sinkhole traffic [80, 24]), rather than DDoS attacks characterizations and modeling or tool development. Towards the limitations of our data collection, one may argue its bias for not covering all ISPs on the Internet in our data collection. We note that, however, our data collection also incorporates at-destination data collection, thus all statistics of interest are gathered in the process. For the data size, and in comparison to [47], our study is based on more than 50,000 attacks over 7 months observation period (compared to 31,612 *alarms* over a 4 weeks period in [47]). Note the fundamental difference between attacks and alarms, since a large number of triggered alarms in anomaly detection could be false alarms, while attacks are verified.

Note that our data collection method is not subject to the shortcoming of locality bias highlighted in [25]: all malware families used for launching attacks that we study are well-understood and reversed engineered [52], and traffic sources utilized for launching the attacks are enumerated by active participation. To this end, we believe that our data is representative to the characterized events, and that the length of the observation period is sufficient to draw some conclusions on DDoS attacks on the Internet today.

1.5 Related Work

Research on prevention and mitigation of DDoS attacks remains one of the hottest topics in the security research community [68, 41, 61, 86, 40, 76, 81], despite being one of the oldest, too [56]. DDoS attacks have been intensively investigated and numerous countermeasures have been proposed to defend against them. The related work addresses various issues, including incentives, filtering, takedown, detection, defenses, and measurements. The following are *only samples* of the related work.

Huang et al. [35] addressed the lack of motivations for organizations to adopt the existent cooperative solutions to defeat DDoS attacks by fixing the *incentive* chain. To look closer at the botnet take-down problem, Nadji et al. [55] proposed a *takedown* analysis and recommendation system called rza, which not only allows a postmortem analysis of past takedown but also provides recommendations for future takedown actions. As a proactive solution to DDoS attacks, several *filtering* schemes [58, 45, 38, 83, 84], which must execute on IP routers, have been proposed to prevent flooding packets from reaching target victims. Chen et al. [29] proposed a new defense system that can detect DDoS attacks *over multiple network domains*. *Overlay-based protection* systems such as Secure Overlay Services [43] offer another attractive alternative, as it requires no changes to existing network routing infrastructure and minimal collaboration from Internet Service Provider. In their follow-up work [66] Stavrou and Keromytis proposed a novel, *multiple-path overlay network* that adopts a spread-spectrum-like communication paradigm to address the limitations in existing overlay-based approaches. *Statistical approaches* [33, 39, 46, 44] are applied for DDoS detection.

Detection techniques vary. Feinstein et al. [33] based detection of DDoS attacks on the *entropy* and *frequency-sorted distributions* of selected packet attributes in live traffic. Lee et al. [44] based detection on *cluster analysis*; cubic clustering criterion (CCC) is used on selected traffic variables. Walfish et al. [73] advocated DDoS *defense by offense*; they implemented an application-level defense named speak-up, in which victimized server encourages all clients to automatically send higher volumes of traffic to attackers.

Broadly, defense and detection methods are classified into *at-destination* and *at-source* mechanisms. Historically, most defense systems such as Cisco IDSM-2 [30] and LADS [62] are deployed at destination since it suffers most of the impact. Mirkovic et al. [50] proposed D-WARD, a DDoS defense system deployed at source-end networks that autonomously detects and stops attacks originating from these networks. Akella et al. [23] proposed at-source mechanism to help ISP networks to detect attacks on itself.

Although a lot of work is focused on defense development, the prerequisite to good defense is *understanding characteristics* of DDoS attacks; their types, durations, and patterns. Wood et al. extracted distinct features of DDoS attacks in wireless sensor networks [79], while Geng et al. [34] focused on unique DDoS features in ad hoc networks. Mirkovic et al. [51] and Specht et al. [64] proposed taxonomies of DDoS attacks and defenses. Douligeris et al. [31] highlighted features of various attack and defense systems and outlined their advantages and disadvantages. Peng et al. [59] presented a comprehensive survey of the causes of DoS attacks and the state-of-art detection methods.

To understand the nature of DDoS attacks towards effective defenses, Mao et al. [47] presented findings from a measurement study of DDoS attacks using both *flow-level direct* and *backscatter-based indirect* measurements. Moore et al. [54] presented a backscatter analysis for quantitatively estimating attack activity on the Internet. They applied their approach to three-week long dataset to study attacks, including their size, length and other characteristics. They also studied victim classification including their geographical distribution. Their study is outdated, since the security landscape has changed significantly since 2006.

Source IP spoofing and reflection are two common techniques used in launching DDoS attacks. Rossow [60] identified protocols that are susceptible to amplification attacks. 14 protocols of various services including network services such as NTP, SNMP, legacy services, P2P file sharing protocols, among others, were shown to be vulnerable and can be abused by distributed reflective denial-of-service (DRDoS) attacks.

Because of network address translation and firewalls, much of the Internet was unseen in earlier measurements. Casado et al. [25] proposed an opportunistic measurement approach that leverages sources of spurious traffic and backscatter to unveil unseen portion of the Internet. Pang et al. [57] conducted a study of broad characteristics of Internet background radiation in four large unused subnets. Both filtering techniques and active transponders are used for analysis of Internet traffic. Wustrow et al. [80] revisited the topic and characterized the current state of background radiation with significant differences; rapid growth outpacing the growth in productive network traffic, trends toward increasing SYN and decreasing SYN-ACK traffic, etc. Bailey et al. [24] introduced the Internet Motion Sensor (IMS), a globally scoped Internet monitoring system to detect Internet threats. IMS includes a distributed blackhole network with a lightweight responder and a novel payload signature. Zou et al. [87] introduced another monitoring system for early detection of worms using non-threshold based "trend detection" method.

Some prior work focused *traffic characterization* of IP backbones to build general behavior profiles. Xu et al. [82] built behavior profiles of Internet backbone traffic capturing communication patterns of end-hosts and services. Their work used data mining techniques to automatically discover patterns from link-level traffic data with plausible interpretations. By far the most commonly used tools to collect intelligence on new malware families and data are honeypots. In [72], a honey farm system was proposed to improve honeypot scalability by up to six orders of magnitude while still offering quantitatively similar fidelity.

Finally, measurements of botnets are seen in various works, including *exemplary* work in [63]. [32], [69], and [67]. The measurements focus on a single family each and do not identifying botnets from malicious activities, like DDoS, but rather through enumeration efforts (honeypots, takeover, etc.)

2 Proposed Research

In this section, we outline in details our proposed research after a brief introduction of the DDoS workload.

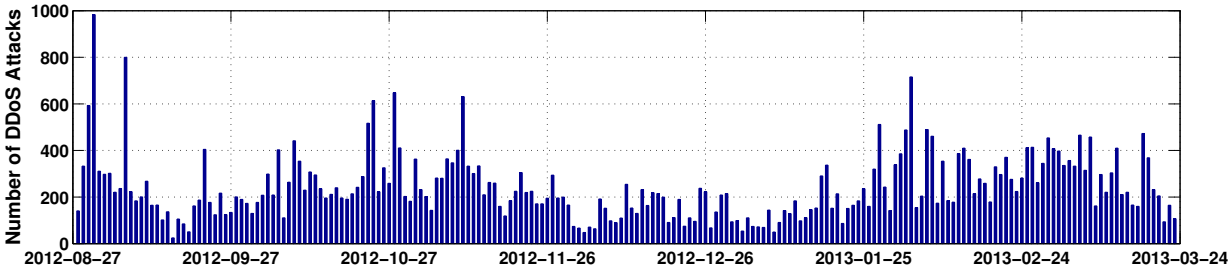

Figure 1: Daily Attack Distribution.

2.1 Dataset and Collection Method

Our dataset is provided by the monitoring and attribution unit in Team Cymru, which operates with partnerships of traffic sharing with more than 300 major ISPs across the globe. 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 the 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 and to enumerate all bots across the globe participating in the particular botnet. As each botnet evolves over time, new generations are marked by their unique hashes.

Traces of traffic associated with various DDoS campaigns are then collected at various anchor points located at more than 300 major ISPs across the globe: North and South America, Asia, Europe, and Africa. The traces are then analyzed to attribute and characterize attacks on various targets. The collection of traffic is guided by two general principles: (i) that the source of the traffic is an infected host participating in a DDoS campaign, and (ii) the destination of the traffic is a targeted client, as concluded from eavesdropping on the command and control (C&C) channel of the campaign using a live sample, or where the end-host is a customer of the said DDoS mitigation company.

2.2 General characteristics

On average there were 243 simultaneous verified DDoS attacks launched by the different botnets everyday. The workload we obtained ranges from August 28, 2012 to March 24, 2013, a total of 209 days (about seven months of valid and marked attack logs). In the log, a DDoS attack is labeled with a unique DDoS identifier, corresponding to an attack by given DDoS malware family on a given target. Other attributes and statistics of the dataset include the DDoS˙ID, botnet˙ID, traffic category (i.e., TCP, UDP, HTTP, SYN, ICMP, etc.), Target_IP, timestamp, end_time, bot_IP, AS Number, organization, country, city, latitude and longitude of target, etc. The longitude and latitude of each IP address are obtained using a highly-accurate industrial geo-mapping service during trace collection. The mapping of the IP addresses happens in real time, making it resistive to IP dynamics. city and organization of each IP address involved in an attack are based on a highly-accurate commercial grade geo-mapping dataset by Digital Envoy [21].

Figure 1 plots the aggregate number of attacks over the period of 28 weeks. In this figure, the y-axis represents the number of different DDoS attacks, and the x-axis represents the time in days. These attacks are launched by 674 different botnets. These attacks targeted victims located in 84 different countries, 616 cities, involving 1074 organizations, residing in 1260 different ASes. Table 2.2 provides a brief summary of the workload and associated features, while the following subsections further outline the proposed work.

To this end, in the following we highlight the three main thrusts of this research projects: measurements and analysis (§2.3), modeling (§2.4) and simulation (§2.5).

Table 1: Summary of the workload information

Summary of Attackers		Summary of Victims	
description	count	description	count
# of bot_ips	310950	# of target_ip	9026
# of cities	2897	# of cities	616
# of countries	186	# of countries	84
# of organizations	3498	# of organizations	1074
# of asn	3973	# of asn	1260

2.3 Measurement and Analysis of DDoS Patterns

In the first research thrust, we aim to conduct measurements and deep analysis to understand unique patterns of real-world DDoS attacks. In this activity, we hope to find some meaningful and new insights that will be leveraged later on in building reliable models for characterizing DDoS attacks in the next activity. To conduct this activity, we aim to quantify findings on the characteristics of DDoS attacks using the proper metrics, wherever possible. After an initial processing, we have already started our analysis work in various preliminary studies, which reveal many quite interesting findings towards this activity. Some of these findings are discussed in the following.

2.3.1 Regionalized Attacks (source analysis)

Traditionally, DDoS attacks are believed to be widely distributed in terms of their attack sources. Such belief is highly employed in defenses that, for example, employ cross-network mechanisms. To examine whether this is the case in today's botnet based DDoS attacks, we first look into the geographical distributions of these DDoS attacks from attacker's perspective. For this purpose, we obtained the mapping database of IP addresses to organizations and countries. Then for these attacks, we extracted the IP addresses and mapped them into the corresponding organization.

Figure 2 shows the CDF of country coverage of each single DDoS attack launched by each of the top-10 botnet family (for clarity of the figure). From this figure, we can see that for more than 80% of the attacks, attackers are from less than seven countries, which indicates that *most DDoS attacks are not widely distributed*. And we can also tell that Optima has the broadest coverage among all these families. We plot the similar figure for organizations, shown in Figure 3. These two figures show very similar trends. This indicates that for a single country, there are usually *only a few organizations involved in launching attacks*.

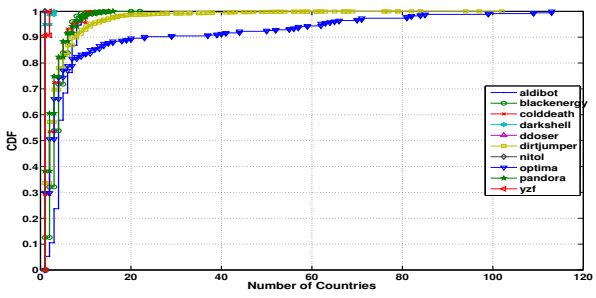

Figure 2: Country-level CDF of botnet families.

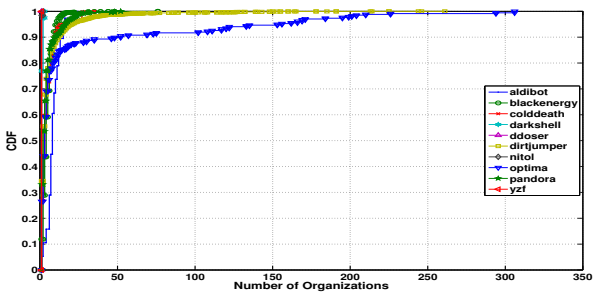

Figure 3: Organization-level CDF of botnet families.

We use Dirtjumper as an example to further highlight such trend in more details. Figure 4 shows one of the DDoS attacks launched by Dirtjumper towards a Russian target starting at 17:12:03 on November 9th,

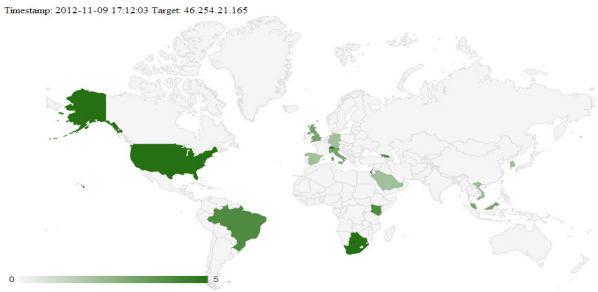

Figure 4: Country-level attacker preference.

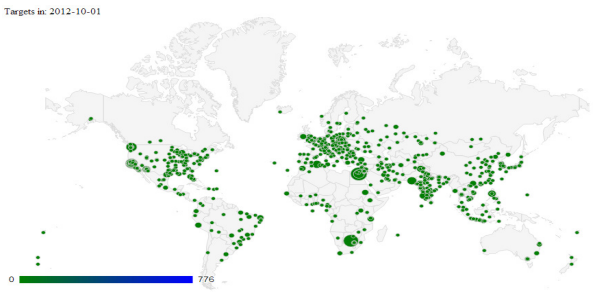

Figure 5: Organization-level attacker preference.

2012. To make it clearer, we use log scale on the numbers of bots involved. For this attack, we find that most of the bots used in this attack were located in US and Israel.

Figure 5 shows the affinity at the organization level. This figure aggregates all the attacks launched in October 2012 by Dirtjumper. In Figure 5, the size of the circles represents the number of attacks launched from that specific organization. Even though there are some small dots scattered randomly, generally most attackers were located in the US, South Africa, Europe and Southeast Asia, highlighting such affinity.

Target Preference (target analysis). (Different botnet families have strong target preferences in the same area). After finding that these attacks are originated from attackers in the same or close by regions, we conduct a similar analysis on the targets of these DDoS attacks. Interestingly, we found a similar affinity pattern of these targets. In our dataset, there are 9,026 unique targets in total. Our study shows that while 40% of the targets were attacked only once, the most popular target was attacked 940 times over the same period. After looking up the most popular target IP address, we found that this IP address belongs to the domain of HostGator, a Houston-based web hosting service, indicating that the real target could be an organization hosted by this service. These results indicate that most botnet families have some target preference, perhaps dictated by its business vertical.

We consider one of the most active families, Dirtjumper, as an example. Figure 6 shows the target country distribution for activities by Dirtjumper in October 2012, where grade of color represents the popularity of different target areas. The deeper the color, the more attacks happened in that area. From this figure, we can see clearly that most attacks targeted the US and Russian, supporting our initial characterization.

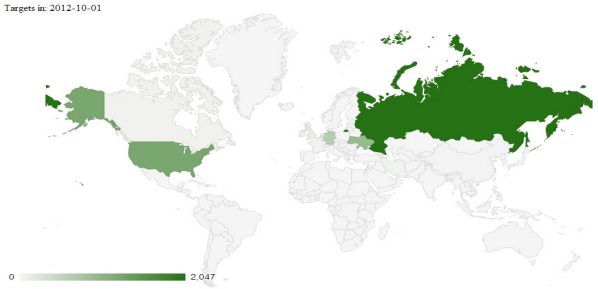

Figure 6: Country-level target preference.

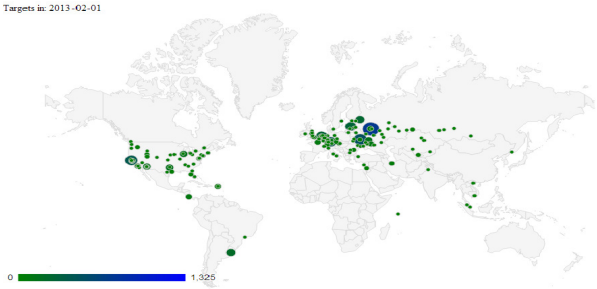

Figure 7: Organization-level target preference.

Similarly, Figure 7 shows the organization-level analysis result for Feburary, 2013, where the size of the circles on the map represents the number of attacks that targeted a specific target. From this figure we can see that among all the targets there are some hotspots. Our further analysis indicates that most attacks were aimed towards web hosting services, large-scale cloud providers and data centers, Internet domain registrars, and backbone autonomous system. They normally possess massive network resources and play a critical function in the operation of other Internet services, making them a desirable target.

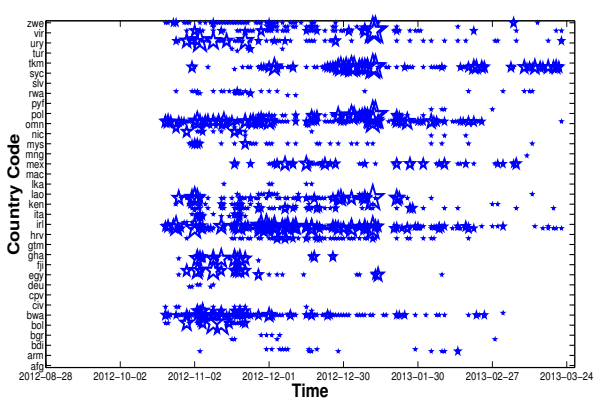

Figure 8: Pandora bots shift (country-level).

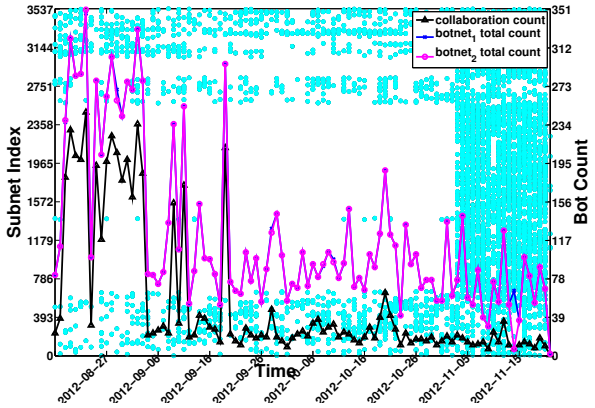

Figure 9: Intra-family collaboration for Blackenergy.

2.3.2 Shift Patterns

(Bots also shift mostly in the same country). Our analysis shows that botmaster strategically rotates the usage of the bots in the attacks launched by the same malware family. For example, a bot is activated for four hours, and then deactivated (hibernated) for one week, and then reactivated for a similar period, and so on. In the meantime, other bots are used to replace the hibernated ones. We refer to this trend as the shift pattern of bots, and highlight it by an example through measurements. For this example, we use Pandora, another active botnet. Figure 8 shows the result of our analysis, where the x-axis represents the whole 28 weeks timespan and the y-axis represents all the countries involved in the attacks. The size of the markers in the figure represents the number of bots involved and located in the corresponding countries. This figure further confirms the geolocation affinity since most of the shifts happened within the same countries. Note here we aggregate the shifts of all the attacks that happened at the same time.

2.3.3 Collaborative Attacks

(Collaborative attacks are observed and conducted by different botnets in a concurrent fashion or serialized; different botnets take turn). Another interesting trend we found through preliminary analyses is the collaboration of different botnets, inside of the same family or across different families. Some different botnet identifiers within the same family have extremely high concurrent usage over the same set of bots. We speculate that attackers employ multiple botnets to launch the same attack. Figure 9 shows an example, demonstrating the collaborations of two botnets within the Blackenergy family. In this figure, the x-axis represents daily timestamps when collaboration happened, the y-axis on the left represents the index of subnet involved in the collaborations, and the y-axis on the right represents the count of bots. For the scatter plot, each dot with different color shows which botnet the subnet belongs to, and the cyan color indicates that they are employed concurrently by both botnets. Three curves for the total number of bots from each botnet and the number of collaborating bots are also plotted. From the collaboration curve and the total count curve, we can clearly see that activities of both botnets are well synchronized. A dedicated group of bots that belong to both botnets are responsible for the surge events.

In other attacks, we notice different botnet groups share the same bots, but those bots are rarely used concurrently. Instead, they were solely used by one botnet at a time, and later majority of those bots are transferred to another botnet. We suspect the same set of bots are leveraged to participate in different campaigns. This temporal pattern suggests that different botnets could be controlled by the same botmaster.

2.3.4 Further Research Activities

These are some exemplar preliminary findings we have found through preliminary analysis. Some are published in SIGCOMM'14 poster [28], CCS'14 poster [74], and ASIA CCS'15 [26]. So far, these exemplar findings are based on DDoS attacks by a specific botnet or a small number of botnets. Some of these findings can be directly applied to practical defense designs. (i) Since botnets are not widely distributed and their targets have strong geographical preferences, different defenders from different regions can prioritize their efforts in dealing with attacks from some different botnet families (such resource allocation philosophy for defense is getting some tractions [42, 49]). (ii) Organizations can also better provision their limited defense resources to maximize the protection surface and capabilities.

To extend beyond this activity, we will address the following directions. (i) Do other botnets behave similarly? (ii) Are there other properties that are worth exploring? (iii) What are the metrics to use to compare various DDoS activities, other than visuals? (iii) What are the methods to execute on such insight of attacks, and metrics to use for evaluating such localized defenses and resource provisioning and their affectiveness? (iv) Along this direction, we will further investigate other characteristics of DDoS attacks; the attack duration, attack size, attack magnitude, attack frequency, etc. Furthermore, these findings indicate some patterns that we plan to further explore. Ultimately, while the insight presented here is preliminary, we aim to have a fine understanding of more complex characteristics that would result in faithful adversary models (c.f. §2.4). If successful, this would be very helpful to the DDoS mitigation in the long run.

2.4 DDoS Pattern Modeling

In this thrust, we will explore models that can capture behavioral traits of bots and botmasters to properly characterize DDoS attacks they launch. Accordingly, more effective mitigation and defense schemes could be conceived. Along this direction, we next discuss some preliminary success we have achieved so far.

2.4.1 Modeling Bot Geographical Distribution

The preliminary findings in §2.3 motivated us to conduct an analysis on the attack sources to explore the geolocation affinity of DDoS attacks. If attackers are not widely distributed, then (i) how do their geolocations change over time? and (ii) can we model such changes?

In our dataset, each DDoS attack could be illustrated by a series of snapshots along time, which correspond to a series of database records. Since every IP address corresponds to one single location, which could be indicated by the longitude and latitude pair, we are able to identify the locations of all the bots involved on a map. Therefore, we can first find the geological center point of these locations. Then we calculate the distance between each bot and this center point and sum the distances together. We use this value to represent the geolocation distribution of the bots. We calculate this value across all the families and plot the CDF of geolocation distributions. We further arrange all the geolocation distribution values of all the DDoS attacks launched by each family in time order. Then we plot the geolocation distances along time. Figure 10 and Figure 11 show the result for *Pandora* and *Blackenergy*, respectively.

In these two figures, the x-axis represents the time when each snapshot was taken, and the y-axis represents the geolocation distance value. From these figures, we can find that there exist periodic patterns in both cases and the distance values present in stationary states, meaning that the values vary around a certain mean value. This indicates that these values are predictable or even stable.

To verify our conjecture, we start with a prediction model over this data. To build the model, we use the Autoregressive Integrated Moving Average (ARIMA) model, a popular linear models in time series forecasting. The ARIMA models are easy to use, and flexible for various types of time series [85].

Methodology. To evaluate the results of our prediction model, we split our data into two parts, the first half is for training and the remaining half is used for prediction and evaluation. The prediction results for

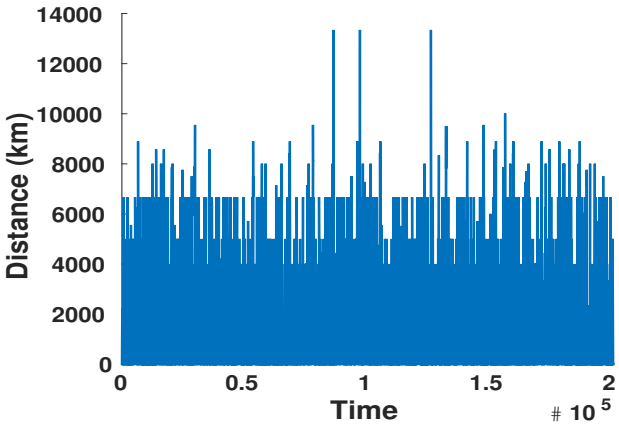

Figure 10: *Pandora* geolocation distances

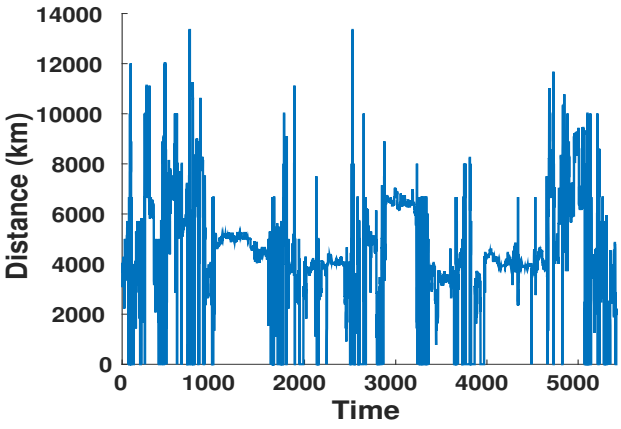

Figure 11: *Blackenergy* geolocation distances

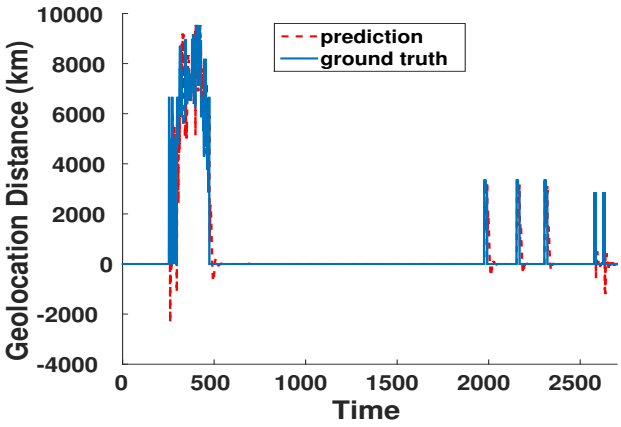

Figure 12: Pandora's geolocation prediction.

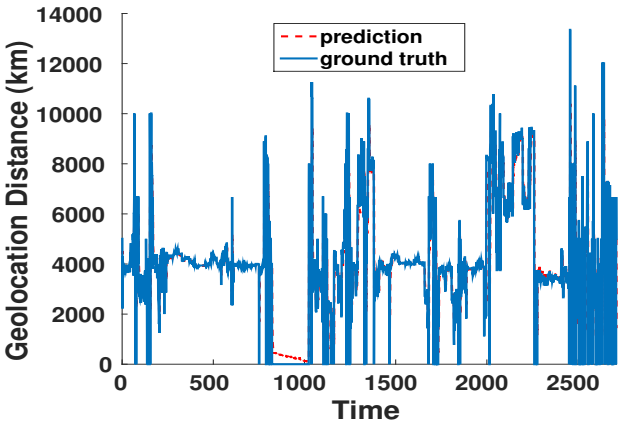

Figure 13: Blackenergy's geolocation prediction.

Pandora and *Blackenergy* are shown in Figure 12 and Figure 13. In these figures, the x-axis represents the predicted points while the y-axis represents the geolocation distance value. The predicted results are shown in dotted red curve and the ground truth values are marked in blue lines. From these figures, we can clearly observe that the predicted results are almost identical with the ground truth value.

Findings. The results reveals several insights including: (i) The geolocation dynamics of bots involved in DDoS attacks exhibit certain patterns for different families. (ii) The changes of attack source geolocation can be accurately predicted using a proper model. (iii) Such information combined with changes of attack volumes can be used for forecasting how DDoS attacks evolve over time, thus deploy or adjust defenses.

2.4.2 Modeling Attack Interval

We model the attack interval in order to predict the start time of next attack. Besides the geolocation, we also conduct preliminary analyses of attack intervals of each target of each family. Similar to the analysis of the target geolocation change, we sort the attacks in time order to calculate the attack intervals between consecutive attacks towards the same target. In this way, we obtain a series of attack intervals for each target. This information is also time-related, suggesting that it might be explored and utilized to characterize and model attacking behaviors. Figure 14 displays two examples of targets by family *Blackenergy*.

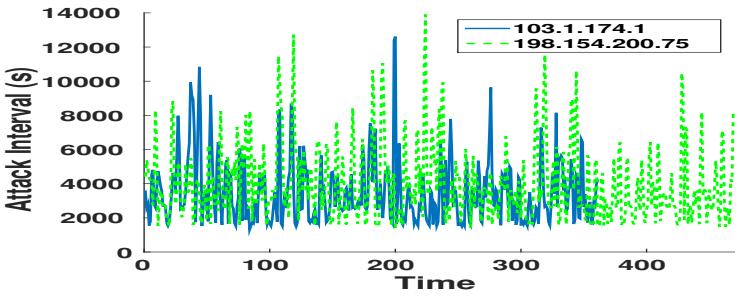

Figure 14: Blackenergy attack intervals.

In Figure 14, the x-axis represents the attack interval along time and the y-axis represents the interval value in seconds. The figure shows some repeated patterns in the peaks and valleys of the series. Besides the periodic pattern, they also present stationary state concerning the mean value of the attack interval values. This means that we may be able to characterize these series by modeling with an ARIMA model and predict the next attack interval value, thus the start time of the next attack.

Methodology and Results. To verify our conjecture, we construct the model as before and Figure 15 and Figure 16 show the prediction results (for a given target). In both cases, we split the data into two equal half, one for the training pool and the other for prediction and evaluation. In both figures, the x-axis represents the time and the y-axis represents predicted values or ground truth values. The ground truth values are marked by dotted curves while the predicted values are marked by solid blue lines. At the first glance of these figures, it is clear that the predicted values match the ground truth consistently.

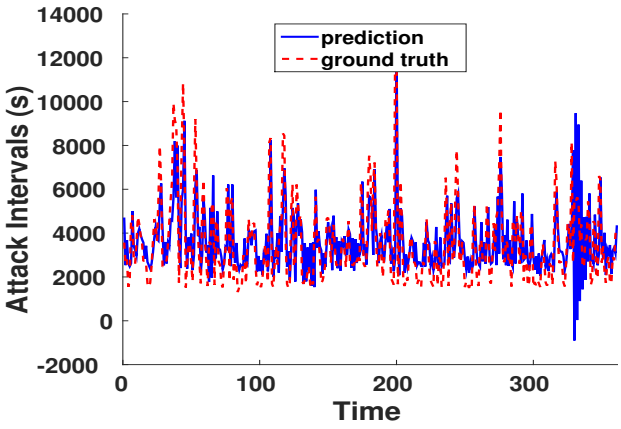

Figure 15: *Blackenergy* attacking interval prediction.

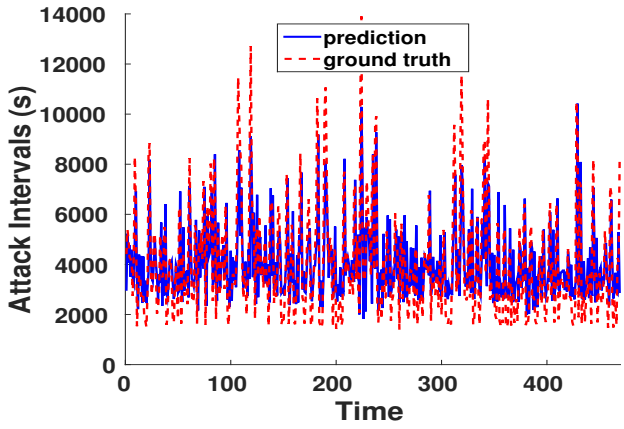

Figure 16: *Blackenergy* attacking interval prediction.

Table 2: Statistics for attacking interval prediction

Target	Group	Mean	Standard Deviation	Cosine Similarity
103.1.174.1	prediction	3579.07596949	1435.59818583	0.92620508
	ground truth	3534.82825485	1901.6606734	
198.154.200.75	prediction	4019.27512744	1473.30582867	0.93394388
	ground truth	4040.81449893	2187.42513432	

We also calculate the statistical results for both cases listed in Table 2. For both groups, the prediction results present more than 90% similarities to the ground truth. These statistical results confirm our conjecture that the attack intervals can be predicted accurately in some cases. After further investigating these

two targets, we found that the studied target in both figures is common among both of Blackenergy and Dirtjumper for multiple times, suggesting that this might be a more popular target.

In conclusion, the attack interval pattern enables the defenders to provision for attacks beforehand. This will be even more helpful with the knowledge of DDoS attack evolution along time.

2.4.3 Modeling Bot Rotation

In our preliminary results, we find that bot rotation management is driven by normal distributions. Previously we have discussed that botmasters rotate their bots in attacks. A potential reason is to minimize the detection window. To seek an in-depth understanding of attackers' strategies, we are motivated to find out how their controlled bots are scheduled to participate in attacks. For this purpose, we use the IP address information of the bots captured in our dataset. For each entry in our dataset, we have the IP address information of all the bots participating in that DDoS attack at that moment, of which the country code (cc) could also be obtained from such information. The information is updated hourly. Thus, the dataset contains all the IP information of bots involved. To this end, in the following, we describe tools for modeling bot rotation.

Method. For each record, such information could be denoted by $\langle cc_1 : n_1, cc_2 : n_2, \ldots, cc_m : n_m \rangle$, where $n_i, i \in [1 \ldots m]$ denotes the number of bots located in $cc_i, i \in [1 \ldots m]$. If we have two such records, denoted by vec_1 and vec_2, the change could be denoted by $vec_2 - vec_1 = \langle cc_1 : \Delta_1, cc_2 : \Delta_2, \ldots, cc_j : \Delta_j \rangle = vec_{\Delta_v}$. Notice that the lengths of vec_1 and vec_2 may not be equal and the length of vec_{Δ_v} will be the same as the longer one. Thus, the difference vector reflects the changes of the bots numbers at the country level for the the given attack. To quantify such changes, we use the notion of *shift expectation* to represent each attack magnitude change. Each DDoS attack is denoted by a vector whose elements are shift expectation,
i.e. $\langle E_{shift_1}, E_{shift_2}, \ldots, E_{shift_m} \rangle$. And the length of this vector is determined by the number of magnitude changes happened in each attack.

Metric. The *shift expectation* is calculated as $\sum_{i=1}^{m} p_i \times \Delta_i$, where Δ_i is obtained from vec_{Δ_v} and p_i denotes the probability of the shift. From our dataset, we obtain the geolocation information of all bots involved in the DDoS attacks. For each family, we generate a table that has two columns; the first column contains all the country codes that are covered by this family while the second one has the corresponding number of bots located in that country. Each entry in this table is denoted by (cc_i, n_i), for $i \in [1 \ldots l]$. On the other hand, $p_i, i \in [1 \ldots l]$, is approximated as $\frac{n_i}{\sum_{j=1}^{l} n_j}$. Thus, the expectation of each shift E_{shift} could be calculated.

Grouping and Analysis. First, we use K-means clustering on all the attack vectors of each family. Since the lengths of vectors may vary, we cannot calculate Euclidean distance between vectors directly. The Dynamic Time Warping (DTW) has been widely used for shape matching and time series classification. Accordingly, we use DTW to calculate the distance and similarity between attack vectors. To reduce the distortion under the influence of attack magnitude, we normalize the vector before we calculate the DTW distance on them. To further explore the pattern behind these vectors, we first find the centroid vector of each cluster and then calculate the distance between each attack vector in that cluster and the centroid.

Preliminary Results. The CDF of the distance distribution for *Dirtjumper* is shown in Figure 17. In this figure, each curve represents a cluster. Besides the curves for clusters, there is also a thick solid CDF curve without markers for a generated collection of numbers that conforms the Normal distribution as a reference. Obviously, the distances follow the normal distribution very well except for cluster-2.

Model Advantages. The operational advantage of normal distribution over other alternatives is that it is easy to manipulate algebraically. It can be used easily to derive formulae of utility. This means that it is possible to derive results that can be easily and automatically applied to bots operation if a normal distribution is utilized. To this end, it is likely that botmasters also utilize this feature to manage bots, especially with such large number of bots. From a defense perspective, such information could be very useful. On one hand, with this information—even though there might be more than one shift pattern for each family—we can

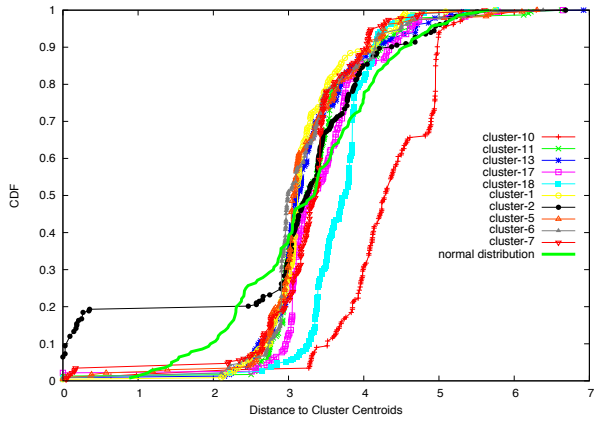

Figure 17: Vector Distances CDF

still predict how attacks shift based on the distribution. On the other hand, we could simulate DDoS attacks behaviors, not only based on traffic volume, but also by incorporating dynamics behind them. This is the result for *Dirtjumper*. We plan to investigate other families as well.

2.4.4 Further Research Activities

So far we discussed some of our preliminary modeling results, based on the DDoS attacks by some specific botnet families. While these findings are promising, they motivate for further activities in this research thrust (i) do other botnets follow the similar strategies in launching attacks? (ii) how collective analysis of attacks (from various botnets or at different targets) affect the robust and quality of the models for various modeled criteria? (iii) do these models fit and standard models? (iv) are there other properties that can be characterized by modeling?

In this project, we will seek answers to these specific questions. We also plan to progress the modeling the movement of the target and sources. We will conduct such modeling on both the target and the source at both organizational and geographical levels. In short, our plan is as follows.

- Spatial Analysis on the Attack Target and Sources: The preliminary result shows some location dependency between the DDoS targets and the attacking sources based on IP addresses. Such dependency exists at both the organizational and the country level. We plan to use advanced spatial analysis to investigate such dependency for both the attack target and the attack sources.

- Time Series Analysis on the Attack Properties: Simply predicting the occurrence of the next DDoS attacks is good, but it is more desirable to also predict the attack duration, attack magnitude, etc. Such information can prepare the service providers with sufficient defense resources. In the project, we plan to apply time series analysis techniques for this purpose.

With the models for different aspects of DDoS attacks, we aim to integrate into a complete model that can (at least partially) characterize the main features of DDoS attacks by botnets.

2.5 Simulation Tool Development

While the output of both direct measurement and analysis of the workload, and the modeling of the DDoS attacks can enhance our understanding and help improve the DDoS mitigation and defenses (e.g., to predict the start time of next attack in a continuous attack sequence as we discussed in the last subsection), in this project, we also propose to utilize such models to build a botnet based DDoS attack simulator.

2.5.1 Why Another Simulator?

As one of the most prevalent attacks on the Internet, profit-driven DDoS attacks are expected to persist, at least in the near-term. As such, the security community, from both academia and industry, has been making continuous efforts to detect and defend against them, as we discussed in the related work. However, a constraint we have is that once some innovative defense schemes are designed, it is hard to evaluate their effectiveness. A reason is that we are still lacking an effective DDoS simulator or DDoS traffic generator that we can use to generate realistic DDoS traffic considering the topology, size, duration, magnitude, preferences, bot rotation, etc. Often, we have to rely on some simple traffic generators or DDoS simulators. For example, one of the most popular simulators is ddosim [1]. It can be used in a laboratory environment to simulate a DDoS attack against a target server by simulating several zombie hosts (having random IP addresses) that create full TCP connections to the target server. After completing the connection, ddosim starts the conversation with the listening application (e.g. HTTP server). It allows users to specify the number of connections to establish and the delay (in milliseconds) between SYN packets. However, it does not support new attacking strategies as we discussed before. Other traffic generators, such as Trinoo [2], TFN [3], Stacheldraht [4], can generate DoS attacks instead of DDoS attacks. Traffic parameters are often pre-defined, and users can only specify the target and the duration. Trinoo generates UDP flood to deplete bandwidth via a buffer overflow exploit. TFN launches ICMP flood, SYN flood, UDP flood, and Smurf style attacks. Stacheldraht is a combinition of Trinoo and TFN with extra encrypted communication between the attacker and the Stacheldraht masters. These tools have great limitations because (1) they usually can only generate DoS attacks, and (2) they only have limited or specific types of attacks. Thus they only exploit the functional vulnerabilities of the targets.

2.5.2 Commercial Tools

In addition to open source tools, there are also many commercial tools. For example, hping [5] (a newer version is hping3) is one of the de facto tools for security auditing and testing of firewalls and networks. Currently, it supports IP/TCP/UDP/ICMP protocols. Some fields of the protocols can be specified through command line, e.g., spoofing IP addresses. Users can specify the sending speed, the intervals between two consecutive packets, and after how many received responses it will stop. Again, such parameters can only emulate simplified DDoS attacks. Ixia BreakingPoint [6] offers an all-in-one security and performance testing platform for massive-scale, stateful layer 4-7 application and security testing. It provides real-world application traffic and real attacks and DDoS and botnet simulation. In order to simulate a DDoS of an arbitrary number of sources directing traffic at a single or small group of targets, both the source and target domains will be specified. To simulate a botnet DDoS attack, the available address range is customized and this range should reflect the size of the DDoS attackers botnet that will be simulated. It includes pre-built botnet simulations such as Cutwail, Zeus, SpyEye, ZeroAccess, Duqu, and BlackEnergy. These simulations are accomplished through utilizing real malware to generate attack traffic. This is the closest testing tool available on the market, although it is not clear if it can simulate the details at the level we have discussed, such as bot rotations and botnet collaborations. In addition, it is a commercial tool. Spirent Test Center [7] provides purpose-built, end-to-end security testing ranging from Terabit-scale, at line rate speeds to emulate daily business traffic, to validate security capabilities via fuzzing testing, DDoS replication, etc. However, Spirent Test Center does not specialize in generating attacking traffic.

2.5.3 General Purpose Tools

Some general purpose traffic generators have also been used in DDoS research. But they often have more constraints. For example, AnetTest [8] is an integrated packet generator and sniffer for Ethernet. Some of its main features include generating any packets at channel level for Ethernet networks, working with

TCP sessions and using (creating) own headers for any network protocol or packet's template. Like most packet generators, it can specify the physical interface through which the packets are sent and assign values to packet fields of various network protocols. However, no other parameters like duration, interval or sending speed could be set. Similarly, other tools, such as Bit-Twist [9], Cat Karat packet builder [10], CommView Packet Generator [11], Nemesis [12], Ostinato [13], Pktgen [14], packETH [15], pierf [16], Scapy [17], Mausezahn [18], and Nping [19], often focus on single packet or single flow features while ignoring other critical features during DDoS attacks like network topology etc. So they may work well for simulating/testing a single network device, but not a network.

2.5.4 The New Simulator

Overall, we find that some of these tools can generate specific attacks such as TCP SYN flood attack, UDP flooding attack etc. Even though some tools try to simulate multiple attackers by using multithreads, they fail to simulate the attacking traffic patterns, which makes them less effective for simulation or penetration tests. Even for the professional testing devices, the flexibilities are very limited as well. Most of these devices rely on users' specification of the traffic patterns generated. As a result, users without domain knowledges cannot effectively simulate real attacks.

To this end, we propose to build an open source DDoS simulator based on our analysis and models developed in the previous stages. We plan to start from scratch considering the limitations of existing tools. We will integrate the models we develop through our research into the simulator. We hope this tool can advance the DDoS research a step forward.

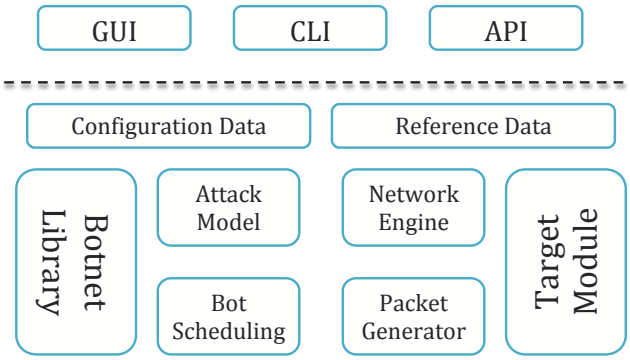

Figure 18: Architecture of the DDoS simulator: α version

High-level design. Figure 18 sketches the high-level design of the simulator. For this simulator, a GUI (Graphical User Interface) will be provided for users' configuration input to specify some parameters of the DDoS attacks, which is saved into the Configuration Data module. Users will also be provided a Command Line Interface (CLI), which is commonly available in the above mentioned tools. In addition to that, we plan to build an API, so that the simulator can directly interact with other programs.

Features and Feeds. Besides the users input, the simulator provides reference data based on the existing over 50 thousands attacks in our workload. The Botnet Library provides the implementation of the 23 active different botnet families. After the selection of botnets, the users can also decide the proper Attack Model, which refers to continuous attack, intermittent attack, long-term attack, short-term attack, collaborative attack, or stand-alone attack, etc. Bot Scheduling provides a scheduling and management option of using the bots. The Network Engine determines the network protocol used for transporting the attack traffic, the network topology, the geographical distribution of bots, the attack rate, duration, size, etc. Once all these parameters are determined, the Packet Generator generates packets and send to the target. Similar to the bot side, the Target Module determines the characteristics of the target.

2.5.5 Research Tasks

To the end, the research activities we envision around this thrust are: (i) system (simulator) requirements, (ii) geographical distribution models (source and target) development, (iii) time modules development (iv) volumetric features module and development, (v) glueing code modules development, and (vi) evaluation and experiments. Those tasks will be aligned with outcomes of the research tasks in the first and second thrusts in this proposal (analysis and modeling). Both (ii) and (iii) will include strategies.

3 Project Risk and Remedies

A natural concern on the proposed research is about the timeliness of the results from the project. For example, one may argue that once the results are known to the community, botmasters can use other strategies to control and manage their bots, and they may change the models that we derive from the real workload, thus making our findings, models, and tools outdated. While we total agree with this projection and expect that to happen (sooner or later), we strongly believe that our project can still help DDoS research and defenses in the systematic approaches we will pursue. Furthermore: (i) on one hand, without our models and the simulator that we will develop, DDoS research may still have to rely on even more outdated tools, if one is needed, based on over-simplified assumptions that do not hold for the modern DDoS attacks; and (ii) on the other hand, DDoS attacks and defenses, much alike other research on security threats and their evolution, are arms' race between the defenders and the attackers. If our research can be used in defenses or to guide an immediate goal, it will lift the bar for attackers, and be considered a measurable success.

Furthermore, we also plan to consider such risks in our project and remedy the risk at various stages using various methods. (i) The PIs will seek to acquire additional datasets that capture behavioral traits of attackers—wherever possible. The dataset used in this preliminary work along with other datasets highlighted in the data management plan and acquired by the PIs leveraging their industrial credentials are all strong indicators of success that support this direction. (ii) We will consider such evolutions and updates in the attacker behavior in the planned simulator design and development. For example, for bot rotation, besides normal distributions, we will also build other potential or plausible models of distribution that botmasters could leverage. (iii) In a similar manner, while building the simulator based on our findings, we will leave the modules as open as possible so that they could be easily expanded later. (iv) The outcomes of the research, including the code of the simulator, will be published as an open source to enable other researchers in the community to build on top of it in a timely fashion.

4 Broader Impact of the Proposed Work

4.1 Broader Technical and Societal Impact

The broader impact of the work proposed in this project is multifaceted. Measuring by the significance and timeliness of the topic treated in this project, we anticipate an equally significant broader potential impact. Such impact is anticipated at multiple fronts, including technical and societal

4.1.1 Technical broader impact

The technical broader impact of this project includes great potential in both the academic and industrial arenas. In the *academic community*, the expected potential impact includes (i) a timely and better understanding of critical operational issues related to one of the most significant threats on the Internet today, DDoS attacks, (ii) fundamental mathematical models and associated tool out of this research that will help improve Internet DDoS attack defenses, (iii) developed techniques and tools that will be available online for free shared with the community through various means in order to allow their wide usage, (iv) anonymized

data used in this project, and all data artifacts from models and simulations in their entirety, will be shared with the community in ways that will facilitate reproducibility of results.

Both (i) and (ii) will be shared in traditional publication venues (peer-reviewed journal and conference publications), whereas (iii) and (iv) will be shared with interested researchers through a dedicated project homepage. List of other teams that will use the research artifacts and their corresponding work will be maintained, as an outcome of this project.

In the *industrial community*, and building on the prior experience of the PIs in communicating their research with that community [53], we anticipate two activities that will result in measurable impact: (i) influencing the design and development of ongoing standard initiatives related to DDoS defense and mitigation, such as the DDoS Open Threat Signaling (DOTS) [36], and increasing awareness of such trends in the proper operators community, such as North American Network Operators Group (NANOG) and DNS Operations, Analysis, and Research Center (DNS-OARC) in their workshops and semi-annual meetings.

4.1.2 Societal impact

The direct *societal* impact of this project is as follows. DDoS attacks, among other malicious activities launched on ordinary Internet, negatively affect Internet users' experience. The development of an arsenal of models to predict and tame such attacks, when deployed, will positively improve such experience, and make the Internet a more pleasant space. With the potential technical impacts outlined above in mind, we aim to pursue such societal impact with confidence.

4.2 Educational Impact

The proposed research and education activities will be conducted at SUNY Buffalo and the George Mason University, which are both rapidly growing universities in both research and education, serving large and diverse student populations in West New York and Northern Virginia, respectively.

With close to 30,000 students, SUNY Buffalo is a member of the AAU, and is a Center of Excellence in Information Systems Assurance Research and Education designated by NSA. At SUNY Buffalo, significant investments have been made in security education as a core area of specialization, with emphasis on teaching practical and timely security curriculum to prepare the next generation of cybersecurity leaders.

George Mason University is also one of the earliest National Centers of Academic Excellence in Information Assurance Education designated by NSA. At GMU, long term efforts have been made to promote an active environment to encourage the development of expertise in both theoretical and applied aspects of systems security. To this end, the educational impact of this project will be significant from several aspects:

- **New curriculum development**. The proposed research will contribute to the PIs current development of new graduate classes entitled *Advanced Topics in Computer Security* and *Internet Computing and Security*. PI Mohaisen is a major security instructor, teaching two *newly designed* courses of computer security at the graduate and undergraduate levels (Fall 2015; `http://cse709.net/`, and Spring 2016). At GMU, PI Chen is a major instructor of networking and systems classes at both undergraduate and graduate levels. For both PIs, the new findings, models constructed, and the tool developed from this research will be introduced in those classes in a timely manner. This will help students better understand various security issues. PI Mohaisen is experimenting with such newly developed course in Fall 2015, and is planning to teach it every Fall thereafter with additional results from this project. PI Chen plans to offer this course in Spring 2016 and expects to teach it once every year. In this course, the PIs will cover system and security related issues for Internet computing, where Botnets and DDoS attacks are included topics. Semester-long course projects will be designed to facilitate learning.

- **Undergraduate participation**. The PIs will encourage highly motivated students to conduct their undergraduate honor projects related to the proposed research. The students will be selected from PIs undergraduate system and security classes, where the students learn the concepts of distributed system, network security, and Internet insightfully, and will have good hands-on experience of computer systems, security, and performance evaluation.
- **Female and underrepresented student participation**. According to the most recent CRA Taulbee report, less than 18% of Computer Science Ph.D. graduates are women. Furthermore, only a small percentage of female Ph.D. students pursue research careers in computer security. *PI Chen has graduated one female Ph.D. student who has started as an assistant professor at Department of Computer Science at SUNY Binghamton*. He supervised one female Ph.D student in industrial internship (Maliheh Shirvanian) and is on the dissertation committee of other one female Ph.D. students. PI Mohaisen has advised several female students, and is on the dissertation committee of a female Ph.D. student. Both PIs will continue to recruit more female and underrepresented students into this research project.

References

[1] https://stormsecurity.wordpress.com/2009/03/03/application-layer-ddos-simulator/.

[2] http://packetstormsecurity.com/distributed/trinoo.analysis.txt.

[3] http://staff.washington.edu/dittrich/misc/tfn.analysis.

[4] http://staff.washington.edu/dittrich/misc/stacheldraht.analysis.

[5] http://linux.die.net/man/8/hping3.

[6] http://www.ixiacom.com/products/ixia-breakingpoint.

[7] http://www.spirent.com/Ethernet_Testing/Software/TestCenter.

[8] http://anettest.sourceforge.net/.

[9] http://bittwist.sourceforge.net/.

[10] http://packetbuilder.net/Home/Intro.

[11] http://www.tamos.com/htmlhelp/commview/pgen.htm.

[12] http://nemesis.sourceforge.net/.

[13] https://code.google.com/p/ostinato/.

[14] http://www.linuxfoundation.org/collaborate/workgroups/networking/pktgen.

[15] http://packeth.sourceforge.net/packeth/Home.html.

[16] http://pierf.sourceforge.net/.

[17] http://www.secdev.org/projects/scapy/.

[18] http://www.perihel.at/sec/mz/.

[19] http://nmap.org/nping/.

[20] —. A ddos attack could cost $1 million before mitigation even starts. http://bit.ly/MUXadv, October 2013.

[21] —. NetAcuity and NetAcuity Edge IP Location Technology. http://www.digitalelement.com/, Feb 2014.

[22] —. Verisign distributed denial of service trends report. http://www.verisigninc.com/en_US/cyber-security/ddos-protection/ddos-report/index.xhtml, February 2015.

[23] Aditya Akella, Ashwin Bharambe, Mike Reiter, and Srinivasan Seshan. Detecting DDoS Attacks on ISP Networks. *ACM SIGMOD/PODS Workshop on Management and Processing of Data Streams (MPDS)*, 2003.

[24] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, David Watson, et al. The internet motion sensor-a distributed blackhole monitoring system. In *NDSS*, 2005.

[25] Martin Casado, Tal Garfinkel, Weidong Cui, Vern Paxson, and Stefan Savage. Opportunistic measurement: Extracting insight from spurious traffic. In *Proc. 4th ACM Workshop on Hot Topics in Networks (Hotnets-IV)*, 2005.

[26] Wentao Chang, Aziz Mohaisen, An Wang, and Songqing Chen. Measuring botnets in the wild: Some new trends. In *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, ASIA CCS '15, pages 645–650, New York, NY, USA, 2015. ACM.

[27] Wentao Chang, An Wang, Aziz Mohaisen, and Songqing Chen. Characterizing botnets-as-a-service. *SIGCOMM Comput. Commun. Rev.*, 44(4):585–586, August 2014.

[28] Wentao Chang, An Wang, Aziz Mohaisen, and Songqing Chen. Characterizing botnets-as-a-service. In *Proceedings of the ACM SIGCOMM (poster)*, Chicago, IL, Aug. 17-22 2014.

[29] Yu Chen, Kai Hwang, and Wei-Shinn Ku. Collaborative Detection of DDoS Attacks over Multiple Network Domains. *IEEE Transactions ons Parallel and Distributed Systems*, 18:1649–1662, 2007.

[30] Cisco. Cisco Catalyst 6500 Series Intrusion Detection System. `http://bit.ly/1hspyy9`, Feb 2014.

[31] Christos Douligeris and Aikaterini Mitrokotsa. DoS attacks and defense mechanisms: classification and state-of-the-art. *Computer Networks*, 2004.

[32] MARJZ Fabian and Monrose Andreas Terzis. My botnet is bigger than yours (maybe, better than yours): why size estimates remain challenging. In *Proceedings of the 1st USENIX Workshop on Hot Topics in Understanding Botnets, Cambridge, USA*, 2007.

[33] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred. Statistical approaches to DDoS attack detection and response. In *DARPA Information Survivability Conference and Exposition*, 2003.

[34] Xianjun Geng, Yun Huang, and Andrew B. Whinston. Defending wireless infrastructure against the challenge of DDoS attacks. *Mobile Networks and Applications*, 7(3):213 – 223, 2002.

[35] Yun Huang, Xianjun Geng, and Andrew B. Whinston. Defeating DDoS attacks by fixing the incentive chain. *ACM Transactions on Internet Technology*, 7(1), 2007.

[36] IETF. DDoS Open Threat Signaling (dots). `https://datatracker.ietf.org/wg/dots/charter/`, 9 2015.

[37] Info Security Magazine. Spamhaus suffers largest ddos attack in history – entire internet affected. `http://bit.ly/1bfx3ZH`, March 2013.

[38] John Ioannidis and Steven M. Bellovin. Implementing pushback: Router-based defense against DDoS attacks. In *Proc. of Internet Society Symposium on Network and Distributed System Security*, 2002.

[39] S. Jin and D.S. Yeung. A covariance analysis model for ddos attack detection. *IEEE International Conference on Communications*, 2004.

[40] Mattijs Jonker and Anna Sperotto. Mitigating ddos attacks using openflow-based software defined networking. In *Intelligent Mechanisms for Network Configuration and Security*, pages 129–133. Springer, 2015.

[41] Min Suk Kang, Soo Bum Lee, and Virgil D Gligor. The crossfire attack. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 127–141. IEEE, 2013.

[42] Alex Kantchelian, Michael Carl Tschantz, Sadia Afroz, Brad Miller, Vaishaal Shankar, Rekha Bachwani, Anthony D Joseph, and JD Tygar. Better malware ground truth: Techniques for weighting anti-virus vendor labels. In *Proceedings of the 2015 ACM Workshop on Artificial Intelligence and Security. AISec*, volume 15.

[43] A. D. Keromytis, A. D. Misra, and D. Rubenstein. SOS: An Architecture For Mitigating DDoS Attacks. *IEEE Journal on Selected Areas of Communications*, 2004.

[44] Keunsoo Lee, Juhyun Kim, Ki Hoon Kwon, Younggoo Han, and Sehun Kim. DDoS attack detection method using cluster analysis. *Expert Systems with Applications*, 34:1659–1665, 2008.

[45] J. Li, J. Mirkovic, M. Wang, P. Reiher, and L. Zhang. Save: Source address validity enforcement protocol. In *Proc. of IEEE International Conference on Computer Communications*, 2002.

[46] M. Li. Change trend of averaged hurst parameter of traffic under ddos flood attacks. *Computers and Security*, 2006.

[47] Z. Morley Mao, Vyas Sekar, Oliver Spatscheck, Jacobus van der Merwe, and Rangarajan Vasudevan. Analyzing Large DDoS Attacks using Multiple Data Sources. *In Proceedings of ACM SIGCOMM Workshop on Large-Scale Attack Defense*, 2006.

[48] Paul McDougall. Microsoft: Kelihos ring sold 'botnet-as-a-service'. `http://ubm.io/MtCSr7`, September 2011.

[49] Bradley Miller. Scalable platform for malicious content detection integrating machine learning and manual review. Technical report, UC Berkeley, 2015.

[50] Jelena Mirkovic, Gregory Prier, and Peter Reiher. Attacking DDoS at the Source. *In Proceedings of 10th IEEE International Conference on Network Protocols*, pages 312–321, November 2002.

[51] Jelena Mirkovic and Peter Reiher. A Taxonomy of DDoS Attack and DDoS Defense Mechanisms. *ACM SIGCOMM Computer Communications Review*, 34:39–54, April 2004.

[52] Aziz Mohaisen and Omar Alrawi. Av-meter: An evaluation of antivirus scans and labels. In *Detection of Intrusions and Malware, and Vulnerability Assessment - 11th International Conference, DIMVA 2014, Egham, UK, July 10-11, 2014. Proceedings*, pages 112–131, 2014.

[53] Aziz Mohaisen and Allison Mankin. Evaluation of privacy for dns private exchange. IETF DPRIVE WG, 7 2015.

[54] David Moore, Colleen Shannon, Douglas J Brown, Geoffrey M Voelker, and Stefan Savage. Inferring internet denial-of-service activity. *ACM Transactions on Computer Systems (TOCS)*, 24(2):115–139, 2006.

[55] Yacin Nadji, Manos Antonakakis, Roberto Perdisci, David Dagon, and Wenke Lee. Beheading hydras: performing effective botnet takedowns. *In Proceedings of the 2013 ACM SIGSAC conference on Computer and Communications Security*, pages 121–132, November 2013.

[56] Roger M Needham. Denial of service. In *Proceedings of the 1st ACM Conference on Computer and Communications Security*, pages 151–153. ACM, 1993.

[57] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 27–40. ACM, 2004.

[58] K. Park and H. Lee. On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law Internets. *In Proceedings of ACM SIGCOMM*, 2001.

[59] Tao Peng, Christopher Leckie, and Kotagiri Ramamohanarao. Survey of network-based defense mechanisms countering the dos and ddos problems. *ACM Comput. Surv.*, 39, 2007.

[60] Christian Rossow. Amplification hell: Revisiting network protocols for ddos abuse. In *NDSS Symposium 2014*, 2014.

[61] Max Schuchard, Abedelaziz Mohaisen, Denis Foo Kune, Nicholas Hopper, Yongdae Kim, and Eugene Y. Vasserman. Losing control of the internet: Using the data plane to attack the control plane. In *NDSS*, 2011.

[62] Vyas Sekar, Nick Duffield, Oliver Spatscheck, Jacobus van der Merwe, and Hui Zhang. Lads: Large-scale automated ddos detection system. In *Proc. of USENIX Annual Technical Conference*, pages 171–184, 2006.

[63] Seungwon Shin and Guofei Gu. Conficker and beyond: a large-scale empirical study. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 151–160. ACM, 2010.

[64] Stephen M. Specht and Ruby B. Lee. Distributed Denial of Service:Taxonomies of Attacks, Tools and Countermeasures. *International Workshop on Security in Parallel and Distributed Systems*, pages 543–550, September 2004.

[65] Michelle Starr. Fridge caught sending spam emails in botnet attack. `http://bit.ly/1j5Jac1`, Jan 2014.

[66] Angelos Stavrou and Angelos D. Keromytis. Countering DoS Attacks With Stateless Multipath Overlays. *In Proceedings of the 12th ACM Conference on Computer and Communications Security*, pages 249–259, 2005.

[67] Brett Stone-Gross, Marco Cova, Lorenzo Cavallaro, Bob Gilbert, Martin Szydlowski, Richard Kemmerer, Christopher Kruegel, and Giovanni Vigna. Your botnet is my botnet: analysis of a botnet takeover. In *Proceedings of the 16th ACM conference on Computer and communications security*, pages 635–647. ACM, 2009.

[68] Ahren Studer and Adrian Perrig. The coremelt attack. In *Proceedings of the 14th European Conference on Research in Computer Security*, ESORICS'09, pages 37–52, Berlin, Heidelberg, 2009. Springer-Verlag.

[69] Matthew Thomas and Aziz Mohaisen. Kindred domains: detecting and clustering botnet domains using dns traffic. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 707–712. International World Wide Web Conferences Steering Committee, 2014.

[70] Steven J. Vaughan-Nichols. Worst ddos attack of all time hits french site. `http://zd.net/1kFDurZ`, February 2014.

[71] Marissa Vicario. Four ways cybercriminals profit from botnets. http://bit.ly/1e1SIiP, Nov 2010.

[72] Michael Vrable, Justin Ma, Jay Chen, David Moore, Erik Vandekieft, Alex C Snoeren, Geoffrey M Voelker, and Stefan Savage. Scalability, fidelity, and containment in the potemkin virtual honeyfarm. *ACM SIGOPS Operating Systems Review*, 39(5):148–162, 2005.

[73] Michael Walfish, Mythili Vutukuru, Hari Balakrishnan, David Karger, and Scott Shenke. DDoS defense by offense. *In Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 303–314, 2006.

[74] An Wang, Wentao Chang, Aziz Mohaisen, and Songqing Chen. How distributed are today's ddos attacks? In *Proceedings of the ACM CCS (poster)*, Scottsdale, AZ, Nov. 3 - 7 2014.

[75] An Wang, Aziz Mohaisen, Wentao Chang, and Songqing Chen. Capturing ddos attack dynamics behind the scenes. In *Detection of Intrusions and Malware, and Vulnerability Assessment - 12th International Conference, DIMVA 2015, Milan, Italy, July 9-10, 2015, Proceedings*, pages 205–215, 2015.

[76] Bing Wang, Yao Zheng, Wenjing Lou, and Y Thomas Hou. Ddos attack protection in the era of cloud computing and software-defined networking. *Computer Networks*, 81:308–319, 2015.

[77] Wikipedia. Carna botnet. http://bit.ly/1slx1E6, 2014.

[78] Wayne Williams. Want to launch your own ddos attacks on a website? $200 will get you everything you need. http://bit.ly/1E6Ie5t, October 2014.

[79] A. D. Wood and J. A. Stankovic. A Taxonomy for Denial-of-Service Attacks in Wireless Sensor Networks. *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, 2004.

[80] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. Internet background radiation revisited. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 62–74. ACM, 2010.

[81] Peng Xiao, Wenyu Qu, Heng Qi, and Zhiyang Li. Detecting ddos attacks against data center with correlation analysis. *Computer Communications*, 67:66–74, 2015.

[82] Kuai Xu, Zhi-Li Zhang, and Supratik Bhattacharyya. Profiling internet backbone traffic: behavior models and applications. In *ACM SIGCOMM Computer Communication Review*, volume 35, pages 169–180. ACM, 2005.

[83] A. Yaar, A. Perrig, and D. Song. Siff: a stateless internet flow filter to mitigate ddos flooding attacks. *IEEE Symposium on Security and Privacy*, 2004.

[84] A. Yaar, A. Perrig, and D. Song. Stackpi: New packet marking and filtering mechanisms for ddos and ip spoofing defense. *IEEE Journal on Selected Areas in Communications*, 2006.

[85] G. Peter Zhang. Time series forecasting using a hybrid arima and neural network model. In *Neurocomputing*, pages 159–175, 2003.

[86] Ying Zhang, Zhuoqing Morley Mao, and Jia Wang. Low-rate tcp-targeted dos attack disrupts internet routing. In *NDSS*, 2007.

[87] Cliff Changchun Zou, Weibo Gong, Don Towsley, and LX Gao. The monitoring and early detection of internet worms. *IEEE-ACM Transactions on Networking*, 13(5):961–974, 2005.