Delving into Internet DDoS Attacks by Botnets: Characterization and Analysis

An Wang, *Student Member, IEEE,* Wentao Chang, *Member, IEEE,*
Songqing Chen, *Senior Member, IEEE* Aziz Mohaisen, *Senior Member, IEEE*

***Abstract*—Internet Distributed Denial of Service (DDoS) attacks are prevalent but hard to defend against, partially due to the volatility of the attacking methods and patterns used by attackers. Understanding the latest DDoS attacks can provide new insights for effective defense. But most of existing understandings are based on indirect traffic measures (e.g., backscatters) or traffic seen locally. In this study, we present an in-depth analysis based on 50,704 different Internet DDoS attacks directly observed in a seven-month period. 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. Our analysis reveals several interesting findings about today's Internet DDoS attacks. Some highlights include: (1) geolocation analysis shows that the geospatial distribution of the attacking sources follows certain patterns, which enables very accurate source prediction of future attacks for most active botnet families; (2) from the target perspective, multiple attacks to the same target also exhibit strong patterns of inter-attack time interval, allowing accurate start time prediction of the next anticipated attacks from certain botnet families; (3) there is a trend for different botnets to launch DDoS attacks targeting the same victim, simultaneously or in turn. These findings add to the existing literature on the understanding of today's Internet DDoS attacks, and offer new insights for designing new defense schemes at different levels.**

I. Introduction

TODAY, Internet Distributed Denial of Services (DDoS) attacks are prevalent with the ease of access to large numbers of infected machines, collectively called botnets [2], [3]. According to a recent report [4], the duration, intensity, and diversity of attacks are on the rise: an annual 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, and by 14% from the previous quarter of the same year, with an average attack of 7.39 Gbps. Furthermore, the same report shows that all industry verticals are targeted by attacks. Another report reveals a clear increase in the average duration of DDoS attacks from 60 minutes in the first quarter of 2014 to 72 minutes in second quarter of the same year, which translates to 20% increase [5]. Additionally, recent DDoS attacks have witnessed an uptrend in operational impact, size, and consequences [6], [7], with the largest reported attacks exceeding 500 Gbps [8]. Today's malicious actors are not limited to sophisticated machines, like servers and personal computers; recent DDoS attacks were reportedly utilizing fridges [9], and other massive scanning activities were done using embedded devices, including monitoring cameras and security doors [10]. Recently, a large body of research work [11], [12], [13] also highlight the trend of mobile devices involved in botnet activities.

Security researchers in academia and industry devoted enormous efforts to understanding DDoS attacks and defending against them. The arms race between the attackers and the guardians keeps evolving driven by demands. Understanding the current trends in today's DDoS attacks and their attack vectors is an important phase in devising effective defenses. Existing studies in this regard are based on indirect traffic analyses and artifacts, such as backscatters, or traffic collected locally, or by infiltrating into a botnet. A large scale view of today's Internet DDoS attacks is missing in the literature and calls for further investigation.

In this paper, we present our study of DDoS attacks analysis. As most of the DDoS attacks nowadays are launched by botnets, the dataset utilized in this study focuses on DDoS attacks launched by various botnet families across the Internet. A comprehensive analysis of the botnet families in the dataset could be found in our previous work in [14]. In a seven-month period captured in our dataset, a total of 50,704 different DDoS attacks were observed, which were launched by 674 different botnets coming from 23 different botnet families. These attacks targeted 9,026 different IPs that belong to 1,074 organizations in 186 countries.

Our analyses revealed several interesting observations about today's Internet botnet DDoS attacks. 1) Geolocation analysis shows that the geospatial distribution of the attacking sources follows certain patterns, which enables very accurate source prediction of future attacks for most active botnet families. 2) From the target perspective, multiple attacks to the same target also exhibit strong patterns of inter-attack time interval, allowing accurate start time prediction of the next anticipated attacks from certain botnet families. 4) There is a trend for different botnets to launch DDoS attacks targeting the same victim, simultaneously or in turn.

These findings offer new insights on trends for different malactors, which align well: affinities, collaborative behavior, etc. are all indicators that can shed light on cross-family behaviors: once learned in one family they can be used

A preliminary version of this work modeling the shift patterns has appeared in DSN 2015 [1]. This work is partially supported by National Science Foundation (NSF) under grants CNS-1524462 and CNS-1809000, as well as the Global Research Lab. (GRL) Program of the National Research Foundation (NRF) under grant NRF-2016K1A1A2912757.

A. Wang is with the Computer Science Department at Case Western Reserve University, Cleveland, OH, USA. This work was done while she was at GMU.

S. Chen is with the Department of Computer Science at George Mason University, Fairfax, VA, USA.

W. Chang is with Google. The work was done while he was with GMU.

A. Mohaisen is with the Department of Computer Science at the University of Central Florida, FL, USA.

to understand behavior in other families. Establishing this behavior through the observations in a systematic study will be of significance. Furthermore, even if the observations provided in our study on the state of botnet-driven DDoS attacks do not hold five or ten years from the time of the attack/study, the work at hand still provides a great intellectual contribution and service to the community: it provides an overview of the state of DDoS attacks as of the time of executing the research. For future studies to understand the change in behavior of botnets, they would benefit greatly from this study as a baseline. Aside from the observation, the methods used in the study also can be reused for further analyses (by us and others). Finally, some of the findings can provide insights for designing effective and/or customized defense schemes at different levels.

Organization. In Section II, we describe our dataset including the overall data statistics and the data fields we utilized to do our analysis. In Section III, we present an overview of these DDoS attacks. In Section IV, we analyze the geolocation affinity of attacking sources and their targets. In Section V, we present in depth collaboration analyses between different botnets in a family or across families. We discuss related work in Section VI and conclude with a concise summary of our analyses and their implications in Section VII.

II. Dataset Collection and Methodology

A. Dataset

Our dataset is provided by a monitoring service, using both active and passive measurement techniques. This monitoring service helps the enterprises gain better understanding of the trends in the evolvements of the botnet families. For this purpose, they have deployed infrastructures to provide automated tracking and reporting of known botnets. Also, they have analysts focusing on investigating new malware families and variants of those families.

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. For example, their unique behavioral patterns could be employed for labelling, including custom protocols and custom encryption schemes, as well as threat indicators of attribution (e.g., infrastructure utilized by various malware families). Hosts participating in the given botnet, by communicating with pieces of infrastructure infected by that malware family (e.g. the command and control) are then enumerated and monitored over time, and their activities are logged and analyzed.

As for the attributes of the data we utilized in this study, there are three separate schemas: a Botlist schema, a Botnetlist schema and a DDoSattack schema; all of them collectively are used to capture to profile the malicious activities of botnet families. For the Botlist schema, it contains information related to Bots including the IP, BGP and GeoIP information related to each bot. The Botnetlist schema contains information related to botnets, including the type of the botnet, the infected hosts that belong to that botnet and the details about the host being used to control the botnet. The DDoSattack list contains information related to the DDoS attacks. Each DDoS record represents a separate attack recorded by the monitoring systems. For our analyses, we associate three schemas to create a comprehensive dataset with a focus on the DDoS attacks launched by these bots. An overview of this dataset could be found in Table I and we will discuss the details in the following sections.

B. Collection methodology

As each botnet evolves over time, new generations are marked by their unique (MD5 and SHA-1) hashes. The hash values are assigned by the vendor providing the data. The hash value is computed over the binary (of the malware), captured and analyzed, used for launching the attack at that point in time. Traces of traffic associated with various botnets are collected at various points on the Internet in cooperation with various ISPs. Traffic logs are then analyzed to attribute and characterize attacks. The collection and analysis are guided by two general principles: 1) the source of the traffic is an infected host participating in a botnet attack, and 2) that the destination of the traffic is a targeted client, as concluded from eavesdropping on command and control of the campaign using a live malware samples.

By tracking temporal activities of 23 different known botnet families, the dataset captures a snapshot of each family every hour from 08/29/2012 to 03/24/2013, a total of 207 days, or about seven months. There are 24 hourly reports per day for each botnet family. The set of bots or controllers listed in each report are cumulative over the past 24 hours. The 24-hour time span is measured using the timestamp of the last known bot activity and the time of logged snapshot.

The analysis is high level in nature to cope with the high volume of ingest traffic at peak attack times. As shown later, on average, there was 243 simultaneous verified DDoS attacks launched by the different botnets studied in this work. High level statistics associated with the various botnets and DDoS attacks are recorded every one hour. The workload we obtained ranges from August 29, 2012 to March 24, 2013, a total of 207 days (about seven months of valid and marked attack logs). In the log, each 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 are shown in Table I. We note that botnet family identification and DDoS attacks labeling falls out of the scope of this paper, as it has been addressed in a large body of the literature [15]. In short, labeling is performed using state-of-the-art techniques by professional companies offering the DDoS shielding business combining dynamic analysis, static analysis, and threat sharing. The likelihood of false labelling is very small with the support of techniques that help identify C&C communication channel of botnets.

C. Discussions

One may argue that the dataset used in this study is not up-to-date and may not reflect the latest behaviors of DDoS attacks, such as Mirai [16] and the Dyn attacks [17]. However, we argue that our dataset covers DDoS attacks launched by some of the most active botnet families as 2013, which have been still active on the Internet as of 2016. For example,

TABLE I
INFORMATION OF WORKLOAD ENTRIES

Field	Description
ddos_id	a global unique identifier for the specified DDoS attack
botnet_id	unique identification of each botnet
category	description of the nature of the attack
target_ip	IP address of the victim host
timestamp	the time when the attack started
end_time	the time when the attack ended
botnet_ip	the IP address of botnets involved in the attacks
asn	autonomous system number
cc	country in which the target resides (ISO3166-1 alpha-2)
city	city and/or state in which the target resides
latitude	latitude of target
longitude	longitude of target

the most recent attacks launched by botnet *Blackenergy* date to Jan 2016 [18]. Thus, studying their attacking strategies and behaviors is still important, particularly to shed light on the landscape of traditional network attacks. Furthermore, the economics of the botnets may result in similar behaviors of different botnet families, especially since those botnets actually utilize similar connection-oriented transport as Mirai. To this end, the collaborations and the geolocation affinity could be general to all botnet families including the most recent botnet such as Mirai. This work aims to learn from the history to understand the reality. For the geolocation mapping, we used a commercial-grade mapping service provided by Digital Envoy (https://www.digitalenvoy.com/).

D. Features and statistics

In the following we introduces features and general statistics of our dataset. One interesting feature, as shown in Table I, is the attack category, which refers to the nature of the DDoS attacks by classifying them into various types based on the protocol utilized for launching them; HTTP, TCP, UDP, Undetermined, ICMP, Unknown, and SYN. Different from *Unknown*, *Undetermined* means that the attack type could not be determined based on the available information.

1) Attack mechanisms: Based on the traffic type information, Figure 1 shows the statistic of different protocols. Clearly, the dominant protocol used in these attacks is HTTP, followed by UDP and TCP. Based on the latest reports conducted by the Arbor Networks [19] and Kaspersky Lab [20], TCP-based attacks are still very active and prevalent in today's Internet, though the UDP based reflection/amplification attacks have predominant share of the attack traffic volume. Our work provides complementary analyses and explorations for the community to understand the behaviors of such attack activities. Table II shows the breakdown of transport types of different botnet families. The last column in the table shows the number of attacks belonging to each type. Note that a botnet could utilize multiple attack types. For example, *Blackenergy* supports different transport mechanisms of attack traffic, including HTTP, TCP, UDP, ICMP and SYN. The variety of transport mechanisms explains the family's popularity. Furthermore, the dominance of HTTP as the attacking mechanism in this family highlights the preferred target of attacks, namely application deficiencies instead of infrastructure

TABLE II
PROTOCOL PREFERENCES OF EACH BOTNET FAMILY

Protocol	botnet family	# of attacks
HTTP	colddeath	826
	darkshell	999
	dirtjumper	34620
	blackenergy	3048
	nitol	591
	optima	567
	pandora	6906
	yzf	177
TCP	blackenergy	199
	nitol	345
	yzf	182
	aldibot	26
UDP	blackenergy	71
	ddoser	126
	yzf	187
UNDETERMINED	darkshell	1530
ICMP	blackenergy	147
UNKNOWN	optima	126
SYN	blackenergy	31

vulnerabilities. The dominance also implies that there are no reflection or amplification attacks in our dataset. Most of the reflection attacks utilize the UDP protocol, such as DNS and NTP, because the TCP protocol is connection oriented.

2) Geolocation information: The *longitude* and *latitude* of each IP address in Table I are obtained using a highly-accurate geo-mapping service during the trace collection. The mapping of the IP addresses is a real-time process, making it resistive to IP dynamics. Beside the longitude and latitude, we also generate the individual *city* and *organization* of each IP address involved in an attack using a highly-accurate commercial grade geo-mapping dataset by Digital Envoy (Digital Element services [21]). We use such information for geographical analysis as presented later.

Table III sums up some statistics of our dataset, including information from both the attacker and the target sides. Target statistics are illuminating. Over a period of 28 weeks, 50,704 different DDoS attacks were observed. Each DDoS attack record is differentiated by a unique attack ID and each is associated with a start timestamp and end timestamp. The target IP could also be utilized as an indicator of different attacks. In our analysis, we discovered some periodic pattern of the DDoS attacks as shown in Section V-B. However, for attacks whose interval exceeds 60 seconds, we consider them as different attacks. Note that we defined this attack interval for an in-depth study of the periodic patterns of the DDoS attacks. This does not mean that DDoS attacks could not last longer than 60 seconds. We choose 60 seconds based on two considerations: (1) From the results shown in Fig 7 in Section III, less than 10% of the attacks last less than 60 seconds, meaning that we include majority of the attacks with this interval value; (2) Using smaller interval value help reduce the false positives when identifying collaboration activities, which we will discuss in Section V, since the purpose of collaborations is often maximizing the attacking force by launching attacks almost simultaneously. The actual duration of the DDoS attacks could be calculated with 'timestamp' and 'end_time' from Table I. These attacks were launched by 674

TABLE III
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
# of ddos_id	50704		
# of botnet_id	674		
# of traffic types	7		

different botnets. These attacks targeted victims located in 84 different countries, 616 cities, involving 1,074 organizations, and residing in 1,260 different autonomous systems (ASes).

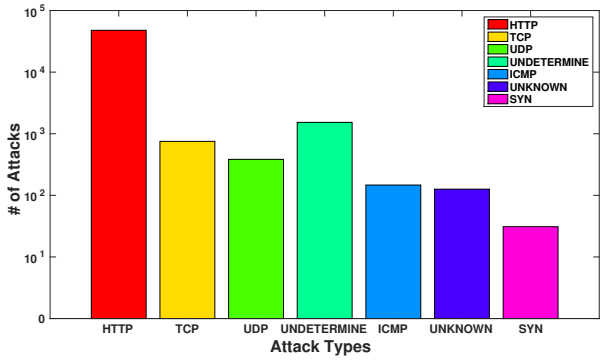


Fig. 1. Popularity of attack types, with most of attacks carried over HTTP, followed by TCP. Undermined implies an attack using multiple protocols, whereas unknown indicate traffic of unknown type. Notice that the majority of attacks are carried over a connection-oriented transport.

E. Comparison and limitations

Several works [22], [23], [24] are on radiation and port scanning measurements. However, most of them are concerned with a single network (Tier-1 ISP [24], sinkhole traffic [23]). Our work is on DDoS attack characterization at a larger scale, making it difficult to directly compare it with the prior literature. Towards the limitations of our data collection, one may argue that not covering all ISPs on the Internet for data collection may bias our data, and thus our findings. 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 [24], our study characterizes more than $50,000$ verified attacks over seven months observation period (compared to 31,612 *alarms* over a period of four weeks in the prior work). Note, the fundamental difference between attacks and alarms is that a large number of triggered alarms in anomaly detection systems could be false alarms, while attacks are verified alarms

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 at the time of the data collection and reversed engineered, and traffic sources utilized for launching the attacks are enumerated by active measurement. To that end, we believe that our data collection 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.

III. OVERVIEW OF DDOS ATTACKS

In this section, we present an overview of DDoS attacks logged in our dataset. We recognize that not all of the 23 botnets logged in our dataset are active all the time. Among them, 10 families are more active than others – a complete analysis of all 23 botnet families can be found in [14]. To this end, in this section we focus on analyzing and characterizing attacks launched by those 10 active families. Namely, we study the DDoS attacks launched by *Aldibot*, *Blackenergy*, *Colddeath*, *Darkshell*, *Ddoser*, *Dirtjumper*, *Nitol*, *Optima*, *Pandora*, and *YZF*.

A. Attack Distribution

More than $50,000$ DDoS attacks launched by 10 active botnet families were observed during the period of 28 weeks' collection. The attack density distribution is an important feature to measure the activity levels of a botnet family. For that, we extract the beginning time of each attack and plotted the aggregate number of attacks over the period in Figure 2. In this figure, the y-axis represents the number of aggregated DDoS attacks for multiple botnet families, and the x-axis represents the time (date). We find that on average there are 243 DDoS attacks launched by the 10 botnet families every day. The maximum number of simultaneous DDoS attacks per day was 983 attacks, which happened on August 30, 2012. All of these attacks were launched by *Dirtjumper* and the targets were located in the same subnet in Russia, suggesting a strong relationship between the different attacks. From comparisons of different families, we can observe that botnet activity patterns are defined by both active time and the attack volumes. For example, *Dirtjumper* presents most aggresiveness due to its constant activities and major contributions to the DDoS attacks. *Blackenergy*, on the other hand, only stays active for about 1/3 of the period. Behind the scenes, the activity level could suggest the proliferation capability and the viability of the botnet malwares.

Although we observe fluctuations in the number of attacks over time, we did not find any obvious daily, weekly, or monthly patterns in Figure 2 that are common in other Internet activities (e.g., diurnal patterns in web access). This is, however, anticipated since DDoS attacks typically are not user-driven, thus lack periodic patterns.

B. Attack Intervals

We further extract the intervals between DDoS attacks. We define the intervals between two DDoS attacks similar to that of the inter-arrival time: the time interval between any two consecutive attacks launched by the same botnet family (or on the same target; across multiple families). Figure 3 shows

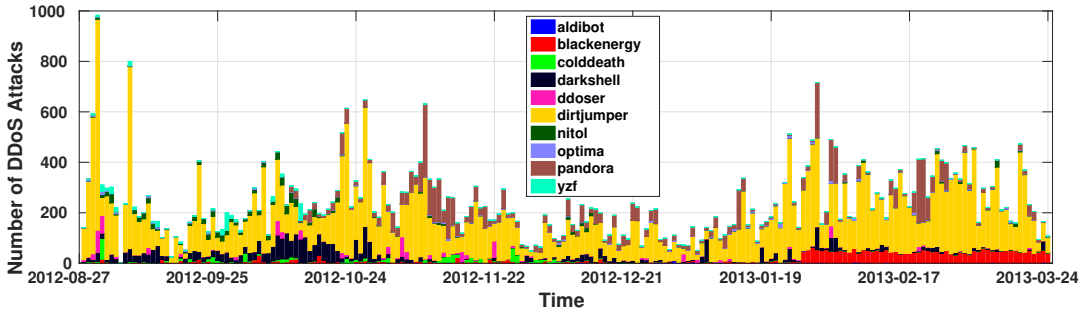


Fig. 2. The daily attack distribution. The number of attacks fluctuates over time, although the numbers do not exhibit any obvious pattern as seen in other online services. On the other hand, while their source varied, many of the attacks happening in the same day were launched against networks in the same network; e.g., in the same country or residing in the same autonomous system.

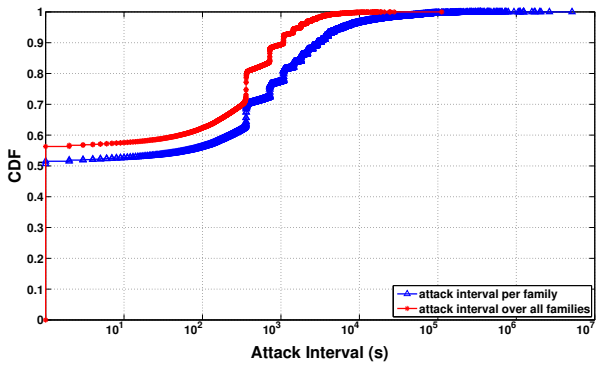


Fig. 3. Attack interval, comparing all attacks and family-based interval distribution (CDF). Notice that more than 50% of the attacks are concurrent when the characterization is confined to the same family, and more than 55% of the attacks are concurrent when characterization is done across all families. The interval on the x-axis is in seconds and in log scale.

the CDF of the attack intervals across all attacks and attacks launched by each family. Note that x-axis is in log scale.

Attack intervals observed from all attacks and family-based attacks show consistent patterns. More than half of the attacks are launched simultaneously. For family based attacks, we found that the longest attack interval was 59 days, almost two months. Also, 80% of the attack intervals lasted less than 1081 seconds, which is roughly 18 minutes. The average DDoS attack interval was 3060 seconds and the standard deviation was 39140 seconds. Those numbers, and by observing the CDF in Figure 3, tell that the attack intervals follow two extremes: except for 15% of the attack falling in the $[1,000, 10,000]$ seconds interval, the majority of the attacks (about 50%) are concurrent, with less than 1% of the attacks at least one order of magnitude larger than the rest of attack intervals.

In this paper, we assume bots do not spoof their IP addresses. This assumption is supported by the following arguments. First, it has been shown that IP spoofing for botnet-launched DDoS attacks is not common [24]. Second, with our traces of attack data, the majority of attacks were using connection-oriented protocols (HTTP), as shown in Fig. 1, making spoofing almost impossible. Thus, we use the number of IPs involved in an attack to estimate one aspect of the corresponding attack magnitude. Thirdly, by aggregating the source IP address and the destination port number 53, we further verify that there are no reflection or amplifications attacks in our dataset, where source IP addresses could belong to the targets. In addition, it is also very unlikely to have anonymized IPs in this dataset since the dataset is provided as part of the service agreements with the enterprises. With the deployment of monitoring systems within ISPs, it is plausible that our dataset captures the malicious behaviors on the attackers' side before they use any proxy mechanisms. Based on the above discussions, we eliminate the possibilities of IP spoofing in our dataset.

As these concurrent attacks are very interesting, we take a closer look at them. We find that they can be classified into two categories: attacks launched by a single botnet family and attacks launched by multiple families. Attacks in the first category happened 3692 times and attacks in the second category happened 956 times.

For the first category, we found that seven out of the 10 botnet families exhibit such behavior. Among all families, *Dirtjumper* is the most active in launching simultaneous attacks; 10% of the attacks launched by *Dirtjumper* are simultaneous. For the second category, we found that most common combinations were *Dirtjumper* with *Blackenergy* and *Dirtjumper* with *Pandora*, which happened 391 and 338 times respectively. This finding is very interesting, and further investigation is dedicated to understand it in §V.

From families' perspective, Figure 4 further shows the interval distributions of all DDoS attacks launched by each botnet family. DDoS attacks are arranged in chronological order for the calculation of attack intervals and simultaneous attacks are eliminated for this analysis. Futher the calculated intervals are grouped into four clusters with different time units, i.e. minutes, hours, days, weeks and months, based on their lengths. From this figure, we observe that the attack intervals present random distributions for all botnet families. However, intervals of 6-7 min, 20-40 min and 2-3 hrs are most commonly shared by all botnet families than others, which suggests predictive attacking strategies utilized by the botmasters. These observations also highlight the possible open time slots for effective mitigations of DDoS attacks.

Figure 5 further shows the attack interval CDF for each family, where the x-axis represents the attack intervals in seconds and each color represents a single family. Note that the x-axis is in log scale (base 2) to highlight the trend and

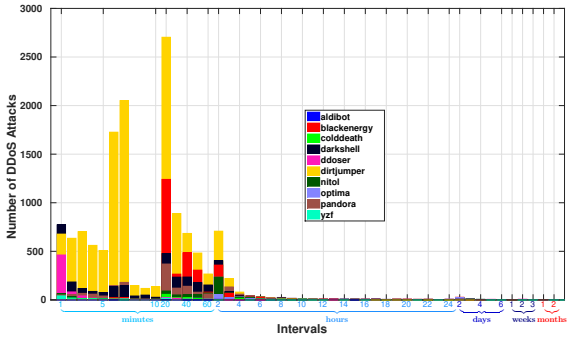


Fig. 4. Attack interval distributions, where the characterization is carried out in chronological order per each family and the simultaneous attacks are eliminated. Notice that while the attack interval is generally random, 6-7 minutes, 20-40 minutes, and 2-3 hours are most common intervals among all characterized families.

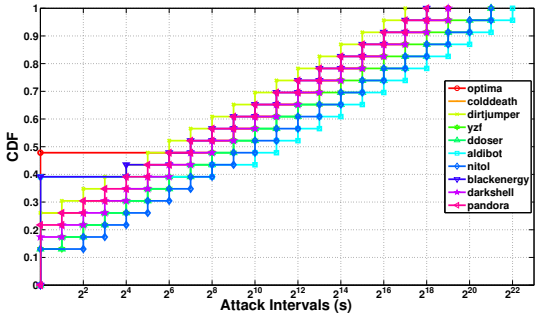


Fig. 5. Per-family CDF characterization of the DDoS attack intervals. Notice that this characterization reveals some of the unique features of various families: while some families have a a large number of attacks that are simultaneous, other families do not have any attacks happening within less than 60 seconds in between of them (e.g., Aldibot and Optima).

pattern in the intervals for the various families. Different from Figure 4, simultaneous attacks are included in this profiling. From this figure we observe that *Blackenergy*, *Aldibot* and *Optima* launch 40%-50% of attacks simultaneously or within a short time frame. We also observe that both *Aldibot* and *Optima* have no attacks with intervals that are less than 60 seconds. This could be a strategy utilized to evade detections. Finally, from the same figure, we observe that the activeness of botnets differ by an order of magnitude, with *Nitol* and *Aldibot* being the least active ones.

C. Attack Duration

The duration of an attack is one aspect that measures its strength and longevity. In our dataset, the measurement of duration is in a way aggregate and does not differentiate between providers and their capability. Figure 6 depicts the durations of all DDoS attacks, where the x-axis represents the attacks along time on daily basis shown in different colors while the y-axis represents the attack duration in seconds. Simultaneous attacks are ordered based on IP addresses. As shown, from the density of the duration distributions, most attacks last between 100 seconds to 10000 seconds. Nonetheless, the attack duration varies significantly: while the average duration is $10,308$ seconds, the median is only $1,766$ seconds, with a standard deviation of $18,475$ seconds (which indicates wide-spread).

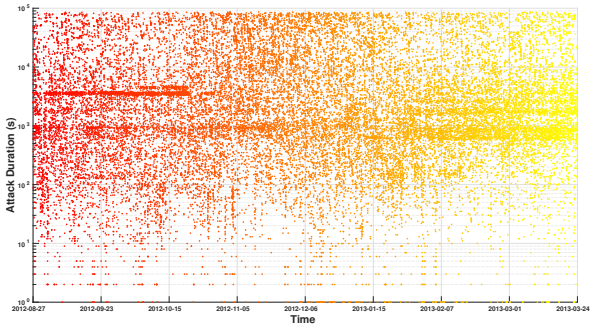


Fig. 6. Attack duration, defined as the duration between the start and the end of the observed attacks in second (log scale) over time. Notice that the majority of attacks' intervals lay between 100 and 10000 seconds.

Figure 7 further shows the corresponding CDF of the attack duration. As shown, 80% of the attacks last for less than $13,882$ seconds (about four hours). Choosing four hours as the cut-off for the majority of attacks duration is perhaps not arbitrary. This value suggests that four hours might be a reasonable time window for DDoS attacks detections and mitigations. An adaptive attacker using such a strategy would evade detection for the longest possible time for most attacks. That is, the longer the attack lasts, the higher its chances are of being detected. By limiting attack to four hours, the attacker can successfully reduce the detection rate, and thus can repetitively launch more attacks later without risking being blacklisted. Compared with the literature [24], where it was shown that 80% of attacks in a comparable study last for less than 1.25 hours, this finding is interesting in itself: DDoS attacks are becoming more persistent by lasting longer; however, their duration is still smaller than the required time frame for detections.

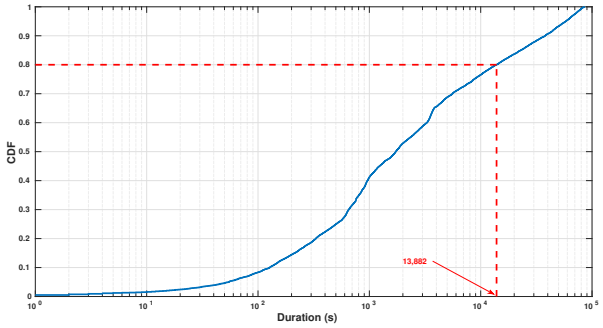


Fig. 7. Duration distribution as CDF across all families, where 80% of the attacks last for less than about 4 hours.

D. Summary

DDoS attacks nowadays are most likely to be event- or profit-driven, demonstrating by the sporadic distribution patterns. Further, multiple rounds of attacks could be launched against the same target within a short interval of up to several

hours. Different attacking intervals suggest various strategies utilized by the botmasters. For the attacking durations, 80% of the attacks have a duration less than four hours, where targets are constantly attacked. This is more likely to be a strategy, rather than the effectiveness of defenses. Above discussions further motivate *automatic* detection and defense instead of any *semi-automatic* or *manual* approaches. Only the former can effectively respond in such a short time frame. Without such an automatic system in place, the detection is not possible for one-time attacking targets. For targets that are repetitively attacked, investigation of the attack intervals may be helpful.

IV. Analysis and Prediction: Target and Source

A. Source Analysis

Geolocation affinity is a direct indicator of how an attacker is geo-spatially distributed. To further quantify the geolocation affinity, we extract all the bots involved in DDoS attacks for each family and aggregate the number of these bots per week. Thus, we are able to observe the attack source and their migrations over weeks. We define such changes as a shift pattern. Figure 8 shows the dynamic shift patterns per week for each botnet family. Shifts are categorized into two clusters based on their destination locations, existing countries or new countries. Two clusters are represented by bars with different patterns on the left and right, respectively and the stacked bars aggregate the total shifts introduced by all botnet families. From this figure we can see clearly that most of the attack sources will be limited to the same group of countries (notice that there are different count units for two clusters, 10^4 and 10^3), confirming that most of these attacks are highly regionalized. Also, this observation applies across multiple botnet families. Next, we explore how the geolocations of different bots participating in attacks change over time.

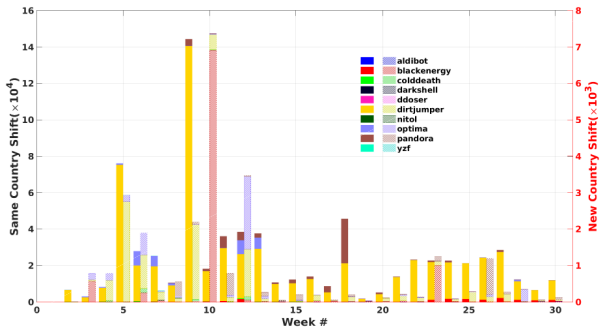


Fig. 8. Source analysis by tracking the botnet shift patterns over time (weekly). The shift pattern capture the number of bots used for attacks and their mobility over countries (origin of attacks). We notice that the botnet shift patterns have a strong affinity within a fixed set of countries, and very few bots are recruited from out of those countries (the right-side y-axis).

In our dataset, each DDoS attack could be illustrated by a series of snapshots along time. In each snapshot, as discussed in §II, IP addresses of all bots evolved at the given time were recorded. Since every IP address corresponds to a single location (longitude and latitude pair), we are able to pin down the locations of all the bots involved on a map. We use such information to characterize source location distributions. First, we find the geological center point of the various locations of IP addresses at any time. Then, we calculate the distance between each bot and this center point (using Haversine formula), and add the distances together. In our analysis, the distance has a sign to indicate direction: positive indicates east or north, and negative indicates west and south. For simplicity, we consider the absolute value of the sum of all distances; a sum of zero means that participating bots are geographically symmetric. We use these distances to represent the geolocation distribution of the bots. These values help profile the dispersions of the attacking sources. However, the actual distance contributes very little to the modeling accuracy of the attackers' locations. Thus, the preference captured by the distance could be applied for predictions as well. We calculate this value across all the families and plot the CDF of geolocation distributions in Figure 9.

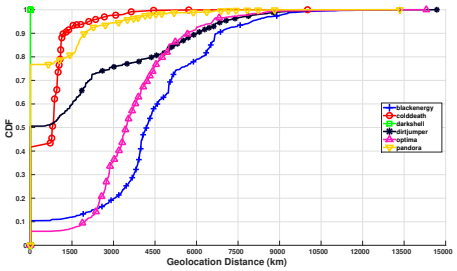


Fig. 9. The geolocation distribution as a CDF of various active botnet families, characterized by the geolocation distance (in km).

In this figure, six families with at least 10 snapshots (with active attacks for more than 10 days) are reported. From Figure 9, we observe that the average distance between the attackers and the targets is about 3,500 kilometers. The number 3,500 itself is incidental. However, after comparing among different botnet families, we find that all the botnet families present predictable patterns in terms of the distance between the involved bots and the target host. We could leverage this information together with the geolocation affinity characteristics of each botnet family to narrow down the candidate pool during the detections, therefore improving mitigation accuracy. Since the attack data was collected globally, it is not surprising that the attackers are located far away from the targets of the DDoS attacks. . We also observe that not all the families follow the same distribution of location proximity. For the families *Optima* and *Blackenergy*, the distances exhibit a normal distribution, whereas other families have a skewed distribution. The families *Dirtjumper* and *Pandora* both have more than 40% distribution distances of zero, indicating complete geographical symmetry. Later, we will show that *Dirtjumper* and *Pandora* collaborate with each other closely, which may explain the similar distribution of their geolocation distances. Furthermore, the different distribution patterns suggest that geolocation distribution is less likely to be random, but rather part of the attack and infection strategy, which could be further confirmed later.

To further explore the dynamics behind the geolocation changes of each DDoS attack, we 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 distributions for *Pandora* and *Blackenergy*, respectively.

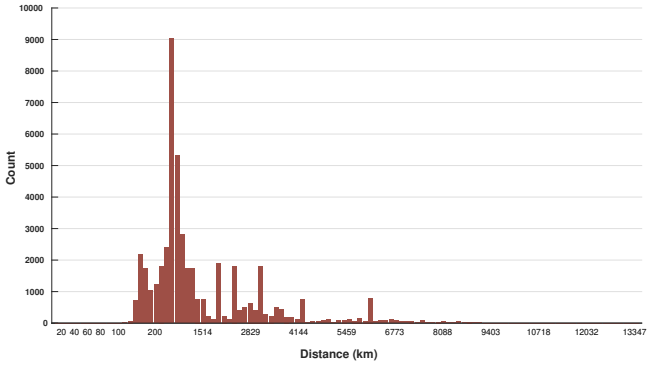


Fig. 10. A histogram of the geolocation of the source of attack, capturing the geolocation distribution of source of the *Pandora* family.

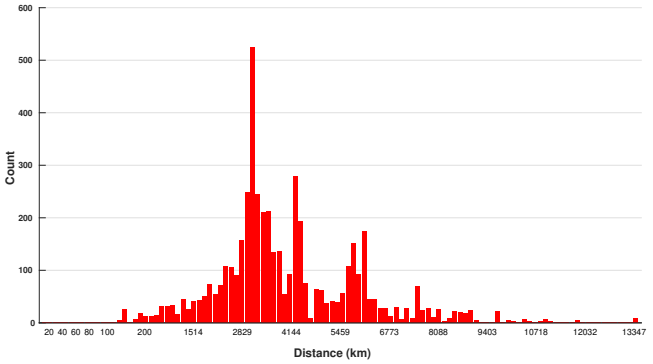


Fig. 11. A histogram of the geolocation of the source of attack, capturing the geolocation distribution of source of the *Blackenergy* family.

In both figures, we remove the symmetric distributions to make clear demonstrations of the asymmetric ones since symmetric distributions dominate the overall distributions, with 76.7% for *Pandora* and 89.5% for *Blackenergy*. After that, the x-axis represents the bins of distances in kilometers, and the y-axis represents the according counts of the specific values. From the above figure, we observe distribution patterns in both cases. The distances appear in stationary states by varying around certain mean values, 566 for *Pandora* and 4304 for *Blackenergy*. This indicates that these values are stable even predictable.

To verify our conjecture, we next build a prediction model over this data. To build the model, we use the Autoregressive Integrated Moving Average (ARIMA) model, which is one of the popular linear models in time series forecasting. The popularity of the ARIMA model is because of its statistical properties in the model building process. In addition, ARIMA models are quite flexible in that they can present several different types of time series [26].

To evaluate the results of our prediction model, we split our data into two parts, the first half is for training and the other half is used for prediction and evaluation. For the prediction part, we use the last 2,700 values (2,700 is a randomly picked number. This value shouldn't affect our prediction results). Again, due to the space limit, we only present the results for the same two families *Pandora* and *Blackenergy*. The prediction results are shown in Figure 12 and Figure 13.

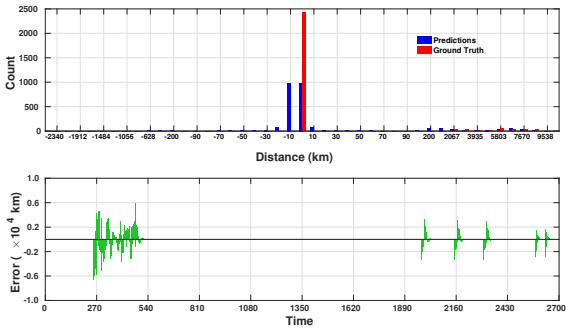


Fig. 12. *Pandora* geolocation distance prediction, with the upper figure showing the actual versus predicted distance as a histogram, and the lower figure showing the error rate over time.

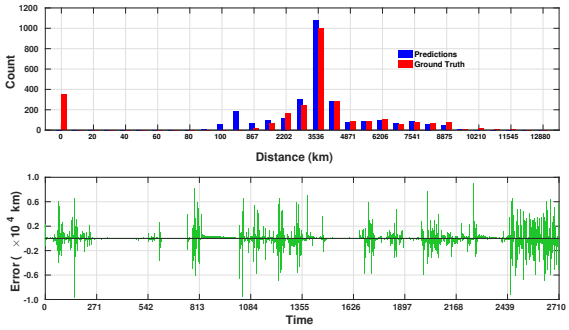


Fig. 13. *Blackenergy* geolocation distance prediction, with the upper figure showing the actual versus predicted distance as a histogram, and the lower figure showing the error rate over time.

In these figures, both the comparisons of geolocation distributions for predictions and ground truth and the corresponding deviations are represented. For the errors, they are shown in chronological order for each predicted data point in the bottom figure. From these figures, we can clearly observe that the predicted results are almost identical with the ground truth value from their distributions. Most errors are caused by the extreme values occured during the attacks. We further calculate the numerical statistics for all the families except for *Darkshell* since there are not enough data points for training the model. The results are listed in Table IV.

We compare two groups of artifacts in this table: the prediction and the ground truth values. We calculated the mean value and the standard deviation value of both groups. Further, we compared these two groups by calculating their cosine similarity with each other. From this table, we can see that for all the families, both the mean value and the standard deviation are close to those of the ground truth, except for family *Dirtjumper* and *Colddeath*; the predicted results represent more than 90% similarity to the ground truth.

TABLE IV
STATISTICS FOR GEOLOCATION DISTANCE PREDICTION. SIMILARITY STANDS FOR THE COSINE SIMILARITY, AND STD STANDS FOR THE STANDARD DEVIATION.

Family	Group	Mean	std	Similarity
Blackenergy	prediction	3968.4	1955.5	0.960
	ground truth	3970.6	2294.4	
Pandora	prediction	562.6	1809.2	0.946
	ground truth	569.2	1842.5	
Dirtjumper	prediction	1203.9	925.8	0.848
	ground truth	1229.1	1033.7	
Optima	prediction	3526.6	1150.1	0.941
	ground truth	3545.8	1717.8	
Colddeath	prediction	356.5	753.2	0.809
	ground truth	341.6	933.8	

Insight into defenses: These results reveal several insights including: (1) The geolocation dynamics of bots involved in DDoS attacks exhibit certain patterns for different botnet families. (2) Attack source geolocation changes can be accurately predicted by using a proper model. accurately predicted by using a proper model. (3) Such information combined with changes of the attack volumes can be used for forecasting how DDoS attacks evolve over time, thus allowing one to deploy or adjust defenses accordingly.

B. Target Analysis

1) Country-level analysis: Now, we turn our attention to the country-level preference of families and their victims. The third column in Table V shows the top five popular targeted countries of each active family. Most families have a specific preference over specific areas or organizations. The top five most popular target countries are the United States of America (USA), targeted by $13,738$ attacks, Russia, targeted by $11,451$ attacks, Germany, targeted by $5,048$ attacks, Ukraine, targeted by $4,078$ attacks, and the Netherlands, targeted by $2,816$ attacks. The *Aldibot* and *Dirtjumper* families' preferred target country is the USA; *Colddeath*'s is India; the *Optima*, *Pandor* and *YZF* families' is Russia; the *Darkshell* and *Nitol* families' is China and *Ddoser*'s is Mexico.

2) Organization-level analysis: Similar to country-level analysis, we have also conducted organization-level analysis. Our results show that the targets were narrowly distributed within several organizations. Figure 14 shows the organization-level analysis in February 2013 for *Pandora*. In this figure, the size of the markers on the map represents the number of attacks toward a specific target. From this figure, we can easily identify some hotspots in Russia and the USA. Among all the families, Dirtjumper has a wider presence by attacking more organizations than any other family. Also, we found that most attacks were aimed towards web hosting services, large-scale cloud providers and data centers, Internet domain registers and backbone autonomous systems, where massive network resources are possessed and play a critical function in the operations of other Internet services.

Insight into defenses: The country and organization level target analyses provide insights for defenses. For example, findings concerning the country-level characterization can set some guidelines on country-level prioritization of disinfection and botnet takedowns. Organization-level characterization and findings associated with that can hint on the possible role provisioning can play in maximizing protection capabilities.

TABLE V
COUNTRY-LEVEL DDOS *target* STATISTICS

Family	Countries	Top 5	Count
Aldibot	14	USA	32
		France	11
		Spain	8
		Venezuela	8
		Germany	4
Blackenergy	20	Netherlands	949
		USA	820
		Singapore	729
		Russian	262
		Germany	219
Colddeath	16	India	801
		Pakistan	345
		Botswana	125
		Thailand	117
		Indonesia	112
Darkshell	13	China	1880
		South Korea	1004
		USA	694
		Hong Kong	385
		Japan	86
Ddoser	19	Mexico	452
		Venezuela	191
		Uruguay	83
		Chile	66
		USA	48
Dirtjumper	71	USA	9674
		Russian	8391
		Germany	3750
		Ukraine	3412
		Netherlands	1626
Nitol	12	China	778
		USA	176
		Canada	15
		United Kingdom	10
		Netherlands	6
Optima	12	Russian	171
		Germany	155
		USA	123
		Ukraine	9
		Kyrgyzstan	7
Pandora	43	Russian	2115
		Germany	155
		USA	123
		Ukraine	9
		Kyrgyzstan	7
YZF	11	Russian	120
		Ukraine	105
		USA	65
		Germany	39
		Netherlands	19

V. ANALYSIS OF COLLABORATIVE ATTACKS

So far, DDoS attacks were analyzed individually. Based on the target analysis discussed earlier, we found that different botnets (in the same family corresponding to different generations, or from different families) may collaborate to attack the same target. They may launch attacks at the same time or alternate their attacks in a way that indicates collaboration. In the following, we elaborate on this collaboration.

Table VI shows the collaboration results using both intra-family and cross-family collaborations. Basically, if different botnets are targeting the same target, and their starting time is simultaneous (or within a 60 second timeframe from each other), and their duration difference is within half an hour, then they are regarded as collaborations. As shown in this table, 121 of the detected collaborations are between different

TABLE VI
BOTNETS COLLABORATION STATISTICS

Collaboration Type	Blackenergy	Colddeath	Darkshell	Ddoser	Dirtjumper	Nitol	Optima	Pandora	YZF
Intra-Family	0	0	253	134	756	17	1	10	66
Inter-Family	1	1	0	0	121	0	1	118	0

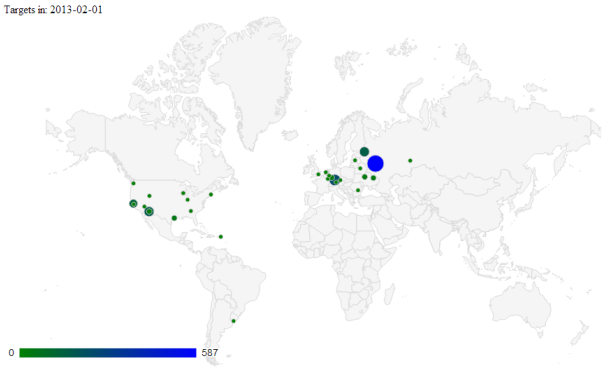


Fig. 14. Target affinity characterization of the *Pandora* malware family over the globe, with larger blue-ish points corresponding to the targets hit the most by bots of the studied family, and smaller green small points are marginal targets. The characterization is done at the organization-level, and the organizations are mapped geographically to their home coordinates (city).

families. Among these collaborations, we observe that two families, namely *Dirtjumper* and *Darkshell*, have the most intra-family collaborations. Next, we look into these intra-family collaborations (between different botnet IDs of the same family) and inter-family collaborations in details.

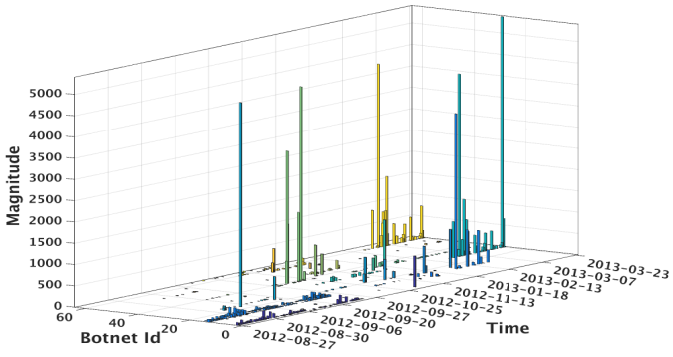


Fig. 15. Intra-family collaborations of *Dirtjumper*, where various generations of the same family (identified by a different botnet identifiers) collaborate to launch an attack against the same target within a confined timeframe.

A. Concurrent Attacks

Figure 15 shows the collaboration attack magnitude by the family *Dirtjumper*. For clarity with respect to the multiple variables, we plot a three dimensional (3D) figure characterizing *Dirtjumper*: the x-axis represents each unique botnet ID, the y-axis represents the date of collaboration, and the z-axis represents the attack volume. From this figure, we can see that for most collaborations, there are two botnets involved, where the average number of botnets involved in the collaboration is 2.19. Such collaborations may be due to a guided action by botmasters, or as instrumented by bots themselves (e.g., multiple entities behind various attacks coincided to utilize the same resources to attack the same target at random). Looking into Figure 15 , we also find that for most bars along the same timestamp, they have the same height. Such an observation reduces the likelihood of involvement of the previously mentioned entities in these collaborations. That is, for all the botnets involved in the collaboration, detailed instructions were perhaps given for the attack magnitude. While that being a random coincidence is possible, it is not plausible, and that further highlights the potential of close collaborations between different botnets.

In addition to the collaborative attacks launched by botnets from the same family, we found that there are attacks launched by botnets from different botnet families. From Table VI, we can see that all families involved in inter-family collaborations had collaborated with *Dirtjumper*. Among these collaborations, *Dirtjumper* and *Pandora* collaborated with each other the most. Our next analysis will focus on those two families.

The collaborations between *Dirtjumper* and *Pandora* involved 96 unique targets, which were located in 16 countries, 58 organizations and 61 ASes. Among the 16 countries, the most popular three countries were Russia, the USA and Germany; with 31, 26 and 14 attacks per country, respectively. On the other hand, for *Pandora*, the average duration of an attack was $6,420$ seconds (107 minutes), while the duration was $5,083$ seconds (87.7 minutes) per attack for *Dirtjumper*.

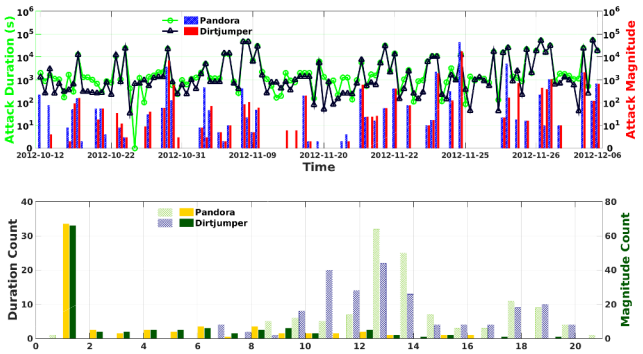


Fig. 16. Inter-family collaborations between *Dirtjumper* and *Pandora*, where bots of different botnet families coordinate their attack against the same target within a confined timeframe.

Figure 16 shows the duration and attack magnitude comparisons of collaborations between *Dirtjumper* and *Pandora*. The top figure shows the value change for both families over time, with left y-axis representing the attack duration while the right y-axis representing the attack magnitude. Both of the y-axes are in log scale. The histogram shows the attack magnitude and the curve shows the attack durations. The bottom figure shows the value distributions for both

duration and magnitude presenting in different patterns. From this figure, we observe that the attack magnitude for these two families are almost equal for most of the attacks, and the duration of these two families are almost identical. This could be further verified by the value distributions. Another observation we make is that the attack magnitudes are not very high for both families except for an outlier. Finally, we observe that the time span of collaboration lasted from October 2012 until December 2012, covering nearly 16 weeks. This long-term collaboration between *Dirtjumper* and *Pandora* highlights a close tie between the two families.

B. Multistage Attacks

Thus far, we consider the collaboration as multiple individual DDoS attacks are launched at the same time. Besides this kind of collaboration, another form of collaboration could be multiple DDoS attacks happening continuously one after another. Next, we investigated this type of collaboration among botnets. For this purpose, we extract the DDoS attacks on a given target that happen consecutively (i.e., the second attack happens at the end of the first attack, or within 60 second margin over overlap). For this type of attack, the results show that only intra-family collaborations were involved. Furthermore, we found that four families had this type of collaboration; *Darkshell*, *Ddoser*, *Dirtjumper* and *Nitol*.

Among all the families and collaborations, *Ddoser* has the longest consecutive DDoS attack involving 22 continuous attacks that lasted for more than 18 minutes on August 30, 2012. On average, the mean interval between two consecutive attacks was 0.11 seconds (a median of three seconds) with a standard deviation of 23 seconds (bursty period)

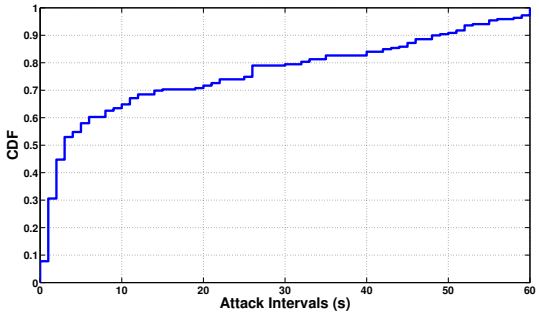


Fig. 17. The distribution of time of consecutive attacks, captured by the CDF. Notice that more than 65% of the consecutive attacks happened within only 10 seconds.

Figure 17 displays the CDF of the intervals between two consecutive attacks. By our definition, we observe that nearly 80% of the consecutive attacks happened within 30 seconds. In practice, this anticipated, and highlights the potential intelligence behind those coordinated attacks: a longer interval would potentially allow targets to deploy various defense mechanisms, and is not likely to be logged in our dataset.

Figure 18 shows the attack magnitude of *all* consecutive DDoS attacks. In this figure, the x-axis represents the 28 week timespan of our dataset, and the y-axis represents all the targets attacked by these consecutive DDoS attacks. Each

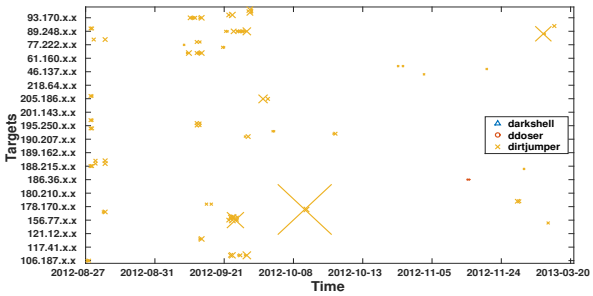


Fig. 18. In-depth analysis of consecutive attacks over time for three botnets: larger keys correspond to larger attacks in size.

dot represents a single DDoS attack. In this figure, the dots displayed consecutively in a row indicate that the attacks happened consecutively. Finally, the size of each marker represents the attack magnitude of each DDoS attack and the different colors represent different families. We observe that the attack magnitudes of different collaborating families are relatively stable during the consecutive attacks, except for *Dirtjumper* that has several attacks of a very large magnitude.

Summary. Intra- and inter-family collaborations could be due to an underlying ecosystem, the evolution of a botnet family, or the evolution of defense mechanisms, which all make defending against them daunting tasks. Devising defenses that employ this insight for attack attribution with an in-depth understanding of the participating hosts in each family is imperative. For example, if we could model the consecutive patterns of DDoS attacks, then the defender could leverage this information to prepare for the next rounds of attacks, e.g., by utilizing a blacklist. For the active defense mechanisms, Zhou *et al.* [27] proposed a solution to detect collaborative attacks, including DDoS attacks.

VI. Related Work

DDoS attacks have been intensively investigated in the literature. Jérôme et al. [28] designed and implemented a collaborative system to detect flooding DDoS attacks as far as possible from the victim host and as close as possible to the attack source(s) at the Internet service provider (ISP) level. It relies on a distributed architecture that is composed of multiple IPSs forming overlay networks to protect subscribed customers. Bilge et al. [29] introduced EXPOSURE, a system that employs passive DNS analysis techniques to detect malicious domain names. Similarly, Sharifnya et al. [30] proposed a negative reputation system that considers the history of both suspicious group activities and suspicious failures in DNS traffic to detect domain-flux botnets. Plohmann et al. [31] recently reverse-engineered 43 malware families and variants that use Domain Generation Algorithms (DGAs). A comprehensive measurement and analysis of behaviors of different botnet families are provided in [14]. Welzel et al. [32] also measures the impact of attacks by DDoS botnets to the victims by analyzing C&C servers of 14 DirtJumper and Yoddos botnets. To look closer to the botnet take-down problem, Nadji et al. [33] proposed a take-down analysis and

recommendation system called rza, which not only allows a postmortem analysis of past take-downs but also provides recommendations for future take-down actions.

There have been several works on understanding unique characteristic of DDoS attacks. Czyz et al. [34] characterize the advent and evolution of DDoS attacks based on Network Time Protocol (NTP) via 5 distinct datasets. They discovered that a large fraction of NTP DDoS attacks are perpetrated against gamers by analyzing the attacked port numbers. Jonker et al. [35] introduced a framework for macroscopic characterization of attacks, attack targets, and DDoS Protection Services (DPSs). They also discovered that the targets are often simultaneously hit by different types of attacks.

Giotis et al. [36] proposed to leverage the OpenFlow protocol as a means to enhance the legacy Remote Triggered Black-Hole (RTBH) routing approach, towards DDoS attack mitigation. Their scheme preserves normal operation of the victim while pushing the mitigation process upstream towards the edge of the network. Lee et al. [37] also proposed to integrate an anomaly detection development framework into SDN to support sophisticated anomaly detection services. Kang et al. [38] designed and implemented a SDN based system to mitigate link flooding attacks with traffic engineering algorithms. A similar framework is built by Liaskos et al. [39] to continuously re-route traffic in a manner that makes persistent participation to link-flooding events highly improbable.

Benson et al. [40] explored the utility of Internet Background Radiation (IBR) as a data source of Internet-wide measurements. They showed that IBR can supplement existing techniques by improving coverage and/or diversity of analyzable networks while reducing measurement overhead. Durumeric et al. [41] analyzed the scanning behavior triggered by vulnerabilities in Linksys routers, OpenSSL, and NTP. They found that large horizontal scanning is common and is responsible for almost 80% of nonConficker scan traffic. In another similar work, Rossow [42] revisited other UDP-based network protocols and identified protocols that are susceptible to amplification attacks. 14 protocols of various services including network services such as Network Time Protocol, Simple Network Management Protocol, legacy services, p2p file sharing network and so on were shown to be vulnerable and can be abused by distributed reflective denial-of-service(DRDos) attacks. A more recent study [43] identifies DNS backscatter as a new source of information about networkwide activity. They used information about the queriers to classify originator activity using machine-learning. Pan et al. [44] proposed a software-defined infrastructure that simplifies and incentivizes collaborative measurement and monitoring of cyber-threat activity.

VII. Conclusion

DDoS attacks are frequently launched on the Internet. While most of the existing studies have mainly focused on designing various defense schemes, the measurement and analysis of large scale Internet DDoS attacks are not very common, although understanding DDoS attacks patterns is the key to defending against them. In this study, with the access to a large scale dataset, we were able to collectively characterize today's Internet DDoS attacks from different perspectives. Our in-depth investigation of these DDoS attacks reveals several interesting findings about today's botnet based DDoS attacks. These results provide new insights for understanding and defending against modern DDoS attacks at different levels (e.g., organization and country). While this study focuses on DDoS characterization, in the future, we plan to leverage these findings to design more effective defense schemes.

References

[1] A. Wang, A. Mohaisen, W. Chang, and S. Chen, "Delving into internet ddos attacks by botnets: characterization and analysis," in *Dependable Systems and Networks (DSN), 2015 45th Annual IEEE/IFIP International Conference on*. IEEE, 2015, pp. 379–390.

[2] A. Wang, A. Mohaisen, and S. Chen, "An adversary-centric behavior modeling of ddos attacks," in *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, GA, USA, June 5-8, 2017*, 2017, pp. 1126–1136.

[3] A. Wang, W. Chang, S. Chen, and A. Mohaisen, "A data-driven study of ddos attacks and their dynamics," *IEEE Transactions on Dependable and Secure Computing*, 2018.

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

[5] ——, http://news.softpedia.com/news/Volumetric-DDoS-Attacks-Decrease-in-Q2-2014-Compared-to-Q1-451160.shtml, July 2014.

[6] Info Security Magazine, "Spamhaus suffers largest DDoS attack in history – entire internet affected," March 2013. [Online]. Available: http://www.infosecurity-magazine.com/news/spamhaus-suffers-largest-ddos-attack-in-history/

[7] S. J. Vaughan-Nichols, "Worst DDoS attack of all time hits french site," February 2014. [Online]. Available: http://www.zdnet.com/article/worst-ddos-attack-of-all-time-hits-french-site/

[8] P. Olson, "The largest cyber attack in history has been hitting hong kong sites," Forbes, November 2014.

[9] M. Starr, "Fridge caught sending spam emails in botnet attack," http://bit.ly/1j5Jac1, Jan 2014.

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

[11] S. Zhao, P. P. Lee, J. Lui, X. Guan, X. Ma, and J. Tao, "Cloud-based push-styled mobile botnets: a case study of exploiting the cloud to device messaging service," in *Proc of ACM ACSAC*. ACM, 2012, pp. 119–128.

[12] R. Hasan, N. Saxena, T. Haleviz, S. Zawoad, and D. Rinehart, "Sensing-enabled channels for hard-to-detect command and control of mobile devices," in *Proc of ACM ASIA CCS*. ACM, 2013, pp. 469–480.

[13] T. Wang, Y. Jang, Y. Chen, S. Chung, B. Lau, and W. Lee, "On the feasibility of large-scale infections of ios devices," in *Proc of USENIX Security*, 2014, pp. 79–93.

[14] W. Chang, A. Mohaisen, A. Wang, and S. Chen, "Measuring botnets in the wild: Some new trends," in *Proc of ACM ASIA CCS*, 2015.

[15] A. Mohaisen and O. 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*, ser. LNCS, vol. 8550. Springer, 2014, pp. 112–131.

[16] Wikipedia, "Mirai (malware)," https://goo.gl/su1a4N, 2017.

[17] ——, "2016 dyn cyberattacks," https://goo.gl/RkkDux, 2016.

[18] "Blackenergy trojan strikes again: Attacks ukrainian electric power industry," https://goo.gl/zjpwDK.

[19] A. Networks, "13th_worldwide_infrastructure_security_report.pdf," https://pages.arbornetworks.com/rs/082-KNA-087/images/13th_Worldwide_Infrastructure_Security_Report.pdf.

[20] K. Lab, "DDoS attacks in Q1 2018 - Securelist," https://securelist.com/ddos-report-in-q1-2018/85373/.

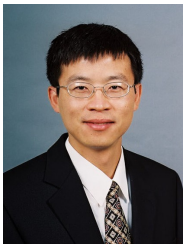
[21] —, "NetAcuity and NetAcuity Edge IP Location Technology," http://www.digitalelement.com/, Feb 2014.

[22] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of internet background radiation," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. ACM, 2004, pp. 27–40.

[23] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, "Internet background radiation revisited," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 2010, pp. 62–74.
[24] Z. M. Mao, V. Sekar, O. Spatscheck, J. van der Merwe, and R. Vasudevan, "Analyzing Large DDoS Attacks using Multiple Data Sources," *In Proceedings of ACM SIGCOMM Workshop on Large-Scale Attack Defense*, 2006.
[25] M. Casado, T. Garfinkel, W. Cui, V. Paxson, and S. Savage, "Opportunistic measurement: Extracting insight from spurious traffic," in *Proc. 4th ACM Workshop on Hot Topics in Networks (Hotnets-IV)*, 2005.
[26] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," in *Neurocomputing*, 2003, pp. 159–175.
[27] C. V. Zhou, S. Karunasekera, and C. Leckie, "Evaluation of a decentralized architecture for large scale collaborative intrusion detection," in *Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on*, 2007.
[28] J. François, I. Aib, and R. Boutaba, "Firecol: a collaborative protection network for the detection of flooding ddos attacks," *IEEE-ACM Transactions on Networking*, 2012.
[29] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "Exposure: A passive dns analysis service to detect and report malicious domains," *ACM Transactions on Information and System Security (TISSEC)*, 2014.
[30] R. Sharifnya and M. Abadi, "Dfbotkiller: domain-flux botnet detection based on the history of group activities and failures in dns traffic," *Elsevier Digital Investigation*, 2015.
[31] D. Plohmann, K. Yakdan, M. Klatt, J. Bader, and E. Gerhards-Padilla, "A comprehensive measurement study of domain generating malware." in *Proc of USENIX Security*, 2016.
[32] A. Welzel, C. Rossow, and H. Bos, "On measuring the impact of ddos botnets," in *Proc of ACM European Workshop on System Security*, 2014.
[33] Y. Nadji, M. Antonakakis, R. Perdisci, D. Dagon, and W. Lee, "Beheading hydras: performing effective botnet takedowns," *In Proceedings of the 2013 ACM SIGSAC conference on Computer and Communications Security*, pp. 121–132, Nov. 2013.
[34] J. Czyz, M. Kallitsis, M. Gharaibeh, C. Papadopoulos, M. Bailey, and M. Karir, "Taming the 800 pound gorilla: The rise and decline of ntp ddos attacks," in *Proc of the ACM Internet Measurement Conference (IMC)*, 2014.
[35] M. Jonker, A. King, J. Krupp, C. Rossow, A. Sperotto, and A. Dainotti, "Millions of targets under attack: a macroscopic characterization of the dos ecosystem," in *Proc of ACM Internet Measurement Conference (IMC)*, 2017.
[36] K. Giotis, G. Androulidakis, and V. Maglaris, "Leveraging sdn for efficient anomaly detection and mitigation on legacy networks," in *Proc of IEEE European Workshop on Software Defined Networks (EWSDN)*, 2014.
[37] S. Lee, J. Kim, S. Shin, P. Porras, and V. Yegneswaran, "Athena: A framework for scalable anomaly detection in software-defined networks," in *Proc of IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2017.
[38] M. S. Kang, V. D. Gligor, V. Sekar *et al.*, "Spiffy: Inducing cost-detectability tradeoffs for persistent link-flooding attacks." in *NDSS*, 2016.
[39] C. Liaskos, V. Kotronis, and X. Dimitropoulos, "A novel framework for modeling and mitigating distributed link flooding attacks," in *Proc of IEEE International Conference on Computer Communications (INFOCOM)*, 2016.
[40] K. Benson, A. Dainotti, A. C. Snoeren, M. Kallitsis *et al.*, "Leveraging internet background radiation for opportunistic network analysis," in *Proc of ACM Internet Measurement Conference (IMC)*, 2015.
[41] Z. Durumeric, M. Bailey, and J. A. Halderman, "An internet-wide view of internet-wide scanning." in *Proc of USENIX Security*, 2014.
[42] C. Rossow, "Amplification hell: Revisiting network protocols for DDoS abuse," in *NDSS Symposium 2014*, 2014.
[43] K. Fukuda, J. Heidemann, and A. Qadeer, "Detecting malicious activity with dns backscatter over time," *IEEE-ACM Transactions on Networking*, 2017.
[44] X. Pan, V. Yegneswaran, Y. Chen, P. Porras, and S. Shin, "Hogmap: using sdns to incentivize collaborative security monitoring," in *Proc of ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*, 2016.

An Wang received her Ph.D. in Computer Science from George Mason University in 2018. Before that, she received her BS in Computer Science from Jilin University in China in 2012. She is currently an assistant professor in the Department of Electrical Engineering and Computer Science at Case Western Reserve University. Her research interests lie in the areas of security for networked systems and network virtualization, focusing on Software-Defined Networking (SDN) and cloud systems, and large-scale network attacks.

Wentao Chang received his Ph.D. in Computer Science from George Mason University in 2018. Before that, he received the B.S. degree (Hons.) in Computer Science and Technology from Nanjing University, Nanjing, China in 2006, M.S. degree in Computer Science from George Mason University in 2010. He is currently working as a Software Engineer at Google. His research interests include botnet analysis, browser and web security.



Songqing Chen received BS and MS degrees in computer science from Huazhong University of Science and Technology in 1997 and 1999, respectively, and the Ph.D. degree in computer science from the College of William and Mary in 2004. He is currently a professor of computer science at George Mason University. His research interests include the Internet content delivery, Internet measurement and modeling, system security, Edge Computing, and distributed systems. He is a recipient of the US NSF CAREER Award and the AFOSR YIP Award.

Aziz Mohaisen earned his M.Sc. and Ph.D. degrees from the University of Minnesota in 2012. He is currently an Associate Professor of Computer Science at the University of Central Florida. Before joining UCF, he was on the faculty of SUNY Buffalo, a senior research scientist at Verisign Labs, and a researcher at ETRI (a large research institute in South Korea). His research interests are in the areas of systems, security, privacy, and measurements. He is on the editorial board of IEEE Transactions on Mobile Computing. He is a member of ACM and a senior member of IEEE.