This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2808344, IEEE Transactions on Dependable and Secure Computing

A Data-Driven Study of DDoS Attacks and Their Dynamics

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

Abstract—Despite continuous defense efforts, DDoS attacks are still very prevalent on the Internet. In such arms races, attackers are becoming more agile and their strategies are more sophisticated to escape from detection. Effective defenses demand in-depth understanding of such strategies. In this paper, we set to investigate the DDoS landscape from the perspective of the attackers. We focus on the dynamics of the attacking force, aiming to explore the strategies behind the scenes, if any. Our study is based on 50,704 different Internet DDoS attacks across the globe in a seven-month period. Our results indicate that attackers deliberately schedule their controlled bots in a dynamic fashion, and such dynamics can be well captured by statistical distributions. Furthermore, different botnet families exhibit similar scheduling patterns, strongly suggesting their close relationship and potential collaborations. Such collaborations are further confirmed by bots rotating in multiple families, and such rotation patterns are examined and confirmed at various levels. These findings lay a promising foundation for predicting DDoS attacks in the future and aid mitigation efforts.

Index Terms—DDoS, Measurements, Attack Scheduling.

1 Introduction

Internet Distributed Denial of Service (DDoS) attacks have been a challenge for many years. Today, many DDoS attacks are launched via different botnets, set of hosts connected to the Internet and infected by a malicious software, i.e., malware. Recent years have witnessed the rapid increase of such DDoS attacks in terms of both their numbers and the volumes, and various studies and reports highlighted their devastating operational impact [2], [3], [4], [5], [6]. For example, according to a recent report [7], 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, and by 14% from the previous quarter of the same year, with an average attack of 7.39 Gbps.

While it is very difficult to estimate the actual monetary loss due to DDoS attacks, since targeted victims are often very secretive about their losses, one can view that in the grand scheme of losses due to malware and associated cyberspace activities: according to a recent study [8], direct and indirect costs due to breaches of malware are estimated at $491 billion in 2014 alone. Given that DDoS attacks are one major security threat, we anticipate that they contribute greatly to those figures of losses. As a matter of fact, today botnet-based DDoS attacks have become a mainstream commodity in the cybercrime ecosystem, where they could be rented or loaned to launch malicious activities, and botmasters can make sizable income by utilizing those botnets.

Simple and conventional DDoS attacks are easy to mitigate, thus DDoS attackers evolved over time to make it harder for a defender to address their threat. Driven by profits generated to the cybercriminals launching those attacks and the lifted bar by the ever-improving defense mechanisms, DDoS attacks and attacking strategies are becoming increasingly sophisticated [9], [10]. Therefore, a timely and an in-depth understanding of the latest trends of DDoS attack and strategies utilized for launching them is a key to a deeper insight in this essential phenomenon, to improve existing defenses, and to realize new ones.

To understand the fundamentals of DDoS attacks, their operation, and potentially to defend against them, enormous efforts are continuously made in both academia and industry, which resulted in many published results and findings [11], [12], [13], [14] (more on that is in section 6). However, most of our understanding of DDoS attacks driven from the state-of-the-art is based on indirect traffic measurements and static characterization of DDoS attacks [12], [14], [15], [16], [17], [18]. On the contrary, our previous study shows that most DDoS attacks today are not widely distributed, but are rather highly regionalized [14]. Furthermore, most of such characterizations only touch the surface of attackers' strategies, making them far from sufficient for us to design effective defenses against many attacks, or even understand DDoS attacks sufficiently. This paper builds on such prior work to better understand and model DDoS attacks and their dynamics.

Contributions. Motivated by the current state of the related art, and to help guide (and perhaps win) the arms race of the DDoS attack and defense, we set out to investigate the attacking strategies of typical and recent DDoS attacks behind the scenes. For this purpose, we explore the attackers' strategies in deploying the attack force, focusing on the dynamics and control of the attack forces in different DDoS attacks. Towards that, we contribute a model for characterizing geographical dynamics of DDoS attackers (§3) and use it to characterize both country-level bot rotation (§4) and family-level bot scheduling (§5). Our study is based on a DDoS dataset collected for a period of 7 continuous months. Our dataset is provided by the monitoring and attribution unit in a DDoS mitigation company in the United States, with partnerships

- *A preliminary version of this work on modeling the shift patterns has appeared in DIMVA 2015 [1].*
- *A. Wang, W. Chang, and S. Chen are with the Department of Computer Science at George Mason University, Fairfax, VA, USA.*
- *A. Mohaisen (corresponding author; e-mail: mohaisen@buffalo.edu) is with the Department of Computer Science and Engineering at the University at Buffalo, State University of New York, Buffalo, NY, USA.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2808344, IEEE Transactions on Dependable and Secure Computing

with various major and global ISPs. The data was collected from August 28, 2012 to March 24, 2013, a total of 209 days, over a period of about seven months of valid and marked attack logs. In this seven-month period, a total of 50,704 different DDoS attacks were observed. Through our analysis, we find several interesting results. Some highlights of those findings include new characterizations of bot scheduling patterns, botnet shift patterns, and attack patterns, summarized as follows:

- **Bot scheduling patterns.** We found that a botnet family often uses a limited number of sophisticated patterns in dynamically scheduling bots to participate in various DDoS attacks. This dynamic scheduling is featured by the shifting patterns of participating bots.
- **Botnet shift patterns.** We observe that bot shifting patterns in different botnet families can be well captured by certain statistical distributions, with parameters of such a distribution dependent on the corresponding family. We highlight such distributions, and discuss their potential in further modeling DDoS attacks and botnet behavior.
- **Attack patterns.** Among all bots participating in DDoS attacks, some bots have a periodic attacking pattern. In particular, they switch between multiple botnet families to evade defense, and such patterns can also be mathematically characterized and modeled.

Implications and Broader Impact. The behavioral patterns that we unveil in this study are direct results of the management strategies of attackers in using their bots to launch the various attacks and attack campaigns. These findings not only refresh and refine our understanding of today's Internet DDoS attacks—a finding that is important in and of itself, particularly for attributing DDoS attacks based on their behavioral characteristics—but also offer new insights for security analysts to identify botnet families and help predict how the attacking forces evolve over time during attacks. This insight can help further enhance the existing defense mechanisms, and perhaps highlight new defense avenues. To the best of our knowledge, this study is first of its kind in unveiling the attacking strategies and dynamic patterns in DDoS attacks, especially patterns that are observed from real-world attack data of various botnet families at the same time.

Organization. The rest of the paper is organized as follows. In Section 2, we describe our dataset including the overall data statistics and the data fields we utilized for our analysis. In Section 3, we define our dynamic model at the country-level, including the notions of shift pattern. In Section 4, we study the country level bots rotation behavior, including intra-family analysis, bot shift modeling, and inter-family analysis. In section 5, we study the family-level bots rotation, including intra- and cross-family rotation and characteristics. We discuss related work in Section 6 and conclude with a concise summary of our analyses and their implications in Section 7.

2 Dataset and Collection Methodology

The dataset we use in this study is based on a constant monitoring of Internet critical infrastructure to aid intelligence gathering concerning the state of the DDoS attack posture, using both active and passive measurement techniques. In the following we review the data collection, criteria, and high-level characteristics.

2.1 Data Collection

The unit responsible for collecting the data for intelligence purposes constantly monitors Internet attacks and associated attacking traffic to aid the mitigation efforts of its customers, 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 [19], [20]. An enumeration technique, similar to the ones proposed by Kang et al. [21], and combining other command and control signals, as in [22], are used to enumerate bots participating in each botnet studied in this paper. As each botnet evolves over time, new generations are identified and marked by their unique hash values, and used for grouping bots and attributing their activities to the given generation.

While we believe that no dataset collection method can eliminate all possible sources of bias, and even though there might be some potential skewness in our dataset, our preliminary studies using the dataset indicate that our dataset still preserves various invariant features that facilitate accurate and insightful analysis. Such invariants particularly include geographical features of botnet and associated families [2], [14], [15]. In mapping the geographical location of the bots and augmenting the dataset to include network features, such as operator and Autonomous System (AS) number, we use a commercial grade and timely mapping service that is updated daily [23] to address potential IP dynamics on the accuracy of location analysis.

Collection Criteria. Traces of traffic associated with various DDoS campaigns are collected at various anchor points across the globe in cooperation with various ISPs. The traces are then analyzed to attribute and characterize attacks on various targets. The collection criteria followed in obtaining the dataset utilized in our study aim to reduce the amount of unnecessary traffic analysis. To this end, the collection of traffic is guided by two general principles that are consistent during the entire period of data collection. First, the source of the traffic logged in our data should be an infected host participating in a DDoS campaign and belongs to a family among the families we are interested in analyzing and understanding. Second, the destination (target) of the traffic is a targeted client, as concluded from eavesdropping on C&C of the campaign using a live sample [24], or where the end-host is a customer of the said DDoS mitigation company.

2.2 High-level Characteristics

The analysis of the collected traces is high level in nature to cope with the high volume of ingested traffic at peak attack times; as shown later, on average there were 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 hour. The workload we obtained covers the period from August 28, 2012 to March 24, 2013, providing a total of 209 days of activities (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 a given DDoS malware family on a given target.

Table 1 sums up some of the essential and high level statistics of our dataset, including information at both the attacker and the target sides. Over a period of 28 weeks, 50,704 different DDoS attacks were observed. These attacks were launched by 674 different botnets (where each generation as defined previously with a unique hash value is counted as a unique botnet). These

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

attacks targeted victims located in 84 different countries, in over 600 cities, involving over 1000 organizations, residing in 1260 different autonomous systems. In our analysis, we focus on the botnets involved in DDoS attacks. However, our prior work in [2], which is orthogonal to this work, contains more details on each botnet family, their activities, and associated patterns.

The attackers' IP information enables us to study the geolocation distribution and associated features of each botnet family. Contrary to the traditional understanding of DDoS attacks featured by a group of vastly distributed malicious actors, the attacks, characterized by the participating bots and affected targets, are not very distributed but are rather highly regionalized, as shown in [14]. Furthermore, each family has its own geolocation preferences. Among all the families, we notice that *Dirtjumper* covers the largest number of countries; a total of 164 countries, followed by Optima's spread over 153 countries. Even though these families have very broad country coverages, the average number of bots participating in each attack pertaining to those botnets is small.

3 Attack Dynamics: Model

To seek an in-depth understanding of attackers' strategies and dynamics, we set out to explore attacks from the adversary's perspective. By doing that, we are motivated to find out how their controlled bots are scheduled and then used to participate in various DDoS attacks. To this end, we use the IP information of the bots captured in our dataset to perform such analysis. Our analysis starts off with two different perspectives, namely the bot shift pattern dynamics and the multi-owned bot attacking interval, both of which are related to DDoS attack strategies. We do this analysis by measuring the country-level bots rotation (§4) and family-level bots rotation (§5).

DDoS attacks evolve over time in terms of their attacking power (a.k.a. force, which is measured by the number of attacking bots that indicate the attack magnitude [13]). In our dataset, each entry represents a snapshot of the DDoS attacks captured at that point in time. We thus can represent the dynamic of each DDoS attack by analyzing each data record, and draw conclusions on the dynamics of the attacking force at various levels.

3.1 Dynamics Characterization

For each entry in our dataset corresponding to an attack, we have the IP address information of all the bots participating in that DDoS attack at that given time, of which the country code (cc) could also be obtained from such information (keep in mind that the snapshots are updated hourly as discussed in §2). After further organizing the bots based on their country code, each entry in the dataset can be denoted by

$$vec_j = \langle cc_1^j : n_1^j, cc_2^j : n_2^j, \ldots, cc_m^j : n_m^j \rangle, \quad (1)$$

where each $cc_i^j, i \in [1 \ldots m]$ represents the country code where the bots are located at time j. We use n_i^j, where $i \in [1 \ldots m]$ to denote the number of bots located in cc_i^j, for $i \in [1 \ldots m]$, at time j. Since each of such vectors represents a time series of snapshot of the activities of a botnet with a certain interval, we can use them to understand the dynamics of the botnet activities as follows

First, we align all vectors belonging to the same DDoS attack together based on the country code, and observe the dynamics (deployment difference of bots) by comparing the number of bots in each country and the number of countries involved in this attack. For example, let vec_j and vec_k be two arbitrary vectors defined as in Eq. (1). We calculate the change in the numbers of bots participating in the attack and indexed by the country as:

$$vec_j - vec_k = \langle cc_1^{j-k} : \Delta_1, \ldots, cc_r^{j-k} : \Delta_r \rangle = vec_{\Delta_v}. \quad (2)$$

Notice that the lengths of vec_j and vec_k (e.g., m_j and m_k) in Eq. (2) may not be the same and the length of vec_{Δ_v}, namely $r = |vec_{\Delta_v}|$, will be the size of the union of vec_j and vec_k (i.e., the length will be the total number of unique countries in both vectors that make the result up). Furthermore, we define the length of vec_{Δ_v} as $r = |vec_{\Delta_v}|$ to quantify the changes in the number of bots involved in an attack. In Eq. (2), we define Δ_z for an arbitrary index z where $1 \leq z \leq r$ as:

$$\Delta_z = \begin{cases} n_j & \text{if } cc_z^j \in vec_j \text{ and } cc_z^k \notin vec_k \\ n_k & \text{if } cc_z^k \in vec_k \text{ and } cc_z^j \notin vec_j \\ |n_k - n_j| & \text{if } cc_z^k \in vec_k \text{ and } cc_z^j \in vec_j \end{cases} \quad (3)$$

An alternative approach to the representation of those vectors is to normalize the vectors by expressing each of them with all existing countries so that they have the same length. However, due to the locality feature presented by the attacking source [25], most elements of the vector would be zero in such normalized representation, which results in a poor utilization with a sparse representation of the location information—although calculations would be made simpler. Since the set of the involved countries are usually unknown a priori, it is more reasonable to use variable length vectors. Otherwise, we would have to use the entire 195 countries to represent each vector for our case. In addition, fixed length vectors are aligned country by country. Thus Euclidean distance could only capture the universal similarity of the shift vectors. Instead, variable length vectors generate optimal pattern matching via dynamic warping. As a result, individual similarity of the shift vectors, which presents a stronger indication of the attackers' strategies and dynamics, could be characterized as well. To this end, we use the variable length vectors and their difference to reflect the changes of the bots numbers at the country level for the given attacks.

3.2 Shift Expectation

To further quantify such changes and dynamics more abstractly and in a comparable way, we use the notion of *shift expectation* to represent each attacking force shift. As the name of the notion indicates, each vector described above as multiple values is represented by a single value called the *shift expectation*. In this way, each DDoS attack can be denoted by a vector whose elements are shift expectation, denoted as

$$\langle E_{shift_1}, E_{shift_2}, \ldots, E_{shift_m} \rangle.$$

Justification. This characterization is particularly reasonable since each DDoS attack can be denoted by a time series of snapshots

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2808344, IEEE Transactions on Dependable and Secure Computing

capturing the attack forces at various points in time. The value of the above vector is determined by both the number of attacking bots (magnitude changes, i.e. vec_{Δ_v}) that have happened in each attack, as well as the number of snapshots of each attack (i.e. length of the vector, which indicates the length of the attack in hours, as highlighted in §2).

Calculation. The *shift expectation* is calculated as follows:

$$E_{shift} = \sum_{i=1}^{m} p_i \times \Delta_i, \tag{4}$$

where Δ_i is obtained from vec_{Δ_v} in Eq. (2) and p_i denotes the probability estimator of the shift. The values of the probability estimator p_i in Eq. (4) are computed as follows:

1) From our dataset, we obtain the geolocation information of all bots involved in the DDoS attacks.
2) For each family, we generate a table that has two columns, where the first column contains all country codes where bots participating in this family reside while the second column has the corresponding number of bots that are located in that country. Each entry in this table is denoted by (cc_i, n_i), for $i \in [1 \dots l]$.
3) p_i, where $i \in [1 \dots l]$, is calculated as

$$p_i = \frac{n_i}{\sum_{j=1}^{l} n_j}. \tag{5}$$

With both p_i in Eq. (5) and Δ_i in Eq. (2), the expectation of each shift E_{shift} can be calculated according to Eq. (4).

After converting each DDoS attack into a time series vector, we have all the vectors with various lengths for all the DDoS attacks in our dataset. Our following analyses is built on top of these vectors and using the notion of shift expectation in Eq. (4).

4 Country-level Bots Rotation

With the previously outlined model of shift and dynamics, we perform intra-family analysis (in §4.1, where we show that participating bots in attacks have the same shifting pattern), bot shift patterns characterization (§4.2, where we show that bot shift patterns can be well captured by certain probability distributions) and inter-family analysis (§4.3, where we highlight indicators of collaborations across multiple families).

4.1 Intra-family Analysis

To use the shift expectation for understanding the intra-family patterns, we need a normalization step to reduce the number of patterns into a manageable size. One possible approach for normalization and to reduce the dimensionality of the data is via clustering. While there are multiple approaches to perform clustering, including DBSCAN and hierarchical clustering, we chose to use the K-means clustering algorithm. The reason we choose K-means over other possible clustering algorithms is that both hierarchical clustering and DBSCAN require certain knowledge about the data we want to cluster beforehand. For example, the data density for DBSCAN or a measure of dissimilarity between sets for hierarchical clustering are required for utilizing the aforementioned algorithms. As a result, we find K-means clustering a more intuitive and simpler way to perform clustering based on the *number of botnet families* we have.

Since the lengths of vectors may vary, we cannot calculate the Euclidean distance between vectors directly. On the other hand, the Dynamic Time Warping (DTW) has been widely used for shape matching and time series classification, where compared vectors in that domain are not necessarily with the same length, similar to the problem settings at hand. Accordingly, we use the DTW to calculate the distance and similarity between the various attack vectors as represented earlier. To reduce the distortion under the influence of attack magnitude, we normalize the vectors using the Euclidean norm before we calculate the DTW distance. In the following we present the results of analysis and their implications.

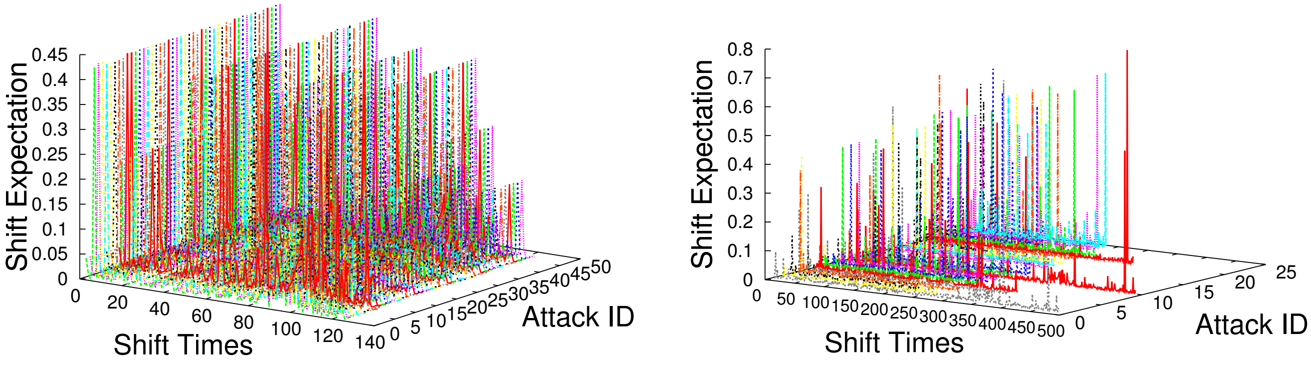
4.1.1 Results

The K-mean clustering algorithm aims to group an arbitrary number of observations into a fixed number of clusters, K, and requires such a parameter to be fixed in advance for its operation. We cluster the various vectors into 5, 10, and 20 clusters. We however observe that clustering them into 10 clusters yields better results, which we select as our parameter for the number of clusters, and present the results for the two largest clusters as a demonstration of the findings, and for brevity. Notice that such criterion of selecting K is widely accepted and known in the literature, which bases the final number of the quality of the obtained clusters.

Figure 1a and Figure 1b illustrate two of the four largest clusters discovered by the K-means algorithm of the *Dirtjumper* family, where $K = 10$. The two clusters contain 54 and 24 attacks, respectively. In each figure, the x-axis represents the length of the attack vector, i.e., the shifts happened in a single attack; the y-axis represents the unique DDoS ID; and the z-axis represents the shift expectation of each shift. Note that since *Dirjumper* has too many DDoS attacks with different lengths of shifts, we first group the attacks by size. In this study, we focus on the analysis of attack vectors that have more than 100 shifts, which include 242 attacks launched by *Dirtjumper*.

In these figures, the expectations should be discrete values. To more clearly show the changes, we use lines to connect these dots. Figure 1a shows that in these attacks, bots are being scheduled with the exact same pattern in different attacks, while Figure 1b indicates a similar pattern—although not identical—in different attacks. With further inspection, we find that in Figure 1a there are 46 simultaneous DDoS attacks ongoing towards the same target located in Finland, which is a company providing communication services from basic broadband to high-speed fiber connections.

Notice that the type of attacks analyzed and modeled in this paper are still of paramount importance today, despite the existence, equal importance, and prevalence of other significant types of attacks that are also worth studying. In particular, we limit our attention to those attacks for several reasons. First, we do so because they are more likely to cause significant damages in real world applications instantaneously, as opposed to low-rate attacks intended for interrupting services by the mere persistence. Second, modeling low-rate and lightweight DDoS attacks, while important, might not generate meaningful nor interesting results on the various aspects highlighted earlier in this paper, and demonstrated in the majority of the attacks observed in this study. We note that while we use a single dataset to drive the main modeling analysis, the number of verified attacks is large enough to reveal generalized results and benefit the community in various ways through characterization and guided defenses.



(a) *Dirtjumper*: attacks with same bot shift pattern

(b) *Dirtjumper*: attacks with similar bot shift pattern

Fig. 1: Shift pattern of the *Dirtjumper* DDoS botnet family.

4.1.2 Implications

These results suggest that the attacking forces are not randomly scheduled by the attackers in *Dirtjumper*. Also, simultaneous attacks cannot be arranged by a completely random deployment strategy. There has to be certain strategies behind DDoS attacks launched by each family. To see if such a pattern is specific to *Dirtjumper* or generalizable to others, we examine other families. Figure 2a and Figure 2b illustrate two clusters of another active botnet family, namely *Pandora*. We use the same K-means clustering technique with 10 clusters for attacks and more than 100 shifts as before. We have similar observations on *Pandora* as on *Dirtjumper*. Other families also exhibit the same pattern (figures are omitted). These findings of consistent pattern across multiple families we studied also suggest that such characteristics can be perhaps leveraged to detect DDoS attacks based on these shift behaviors. But this only will be possible if we can precisely model these pattern, which is the aim of our next contribution.

4.2 Bot Shift Patterns Modeling

Findings in the previous section on understanding the characteristics of shift patterns of bots are intriguing, and may potentially be meaningful in devising techniques to thwart attacks. However, such findings would be more meaningful only if one can predict those patterns. One way to pursue such direction is to understand how various mathematical distributions can capture the shift patterns. 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. We use *Dirtjumper* as an example to highlight this process, since it is the most active family.

4.2.1 Results

The cumulative distribution function (CDF) of the distance distribution for *Dirtjumper* is shown in Figure 3. In this figure, each curve represents a cluster. By carefully observing these curves, we find that the distances seem to follow the normal distribution very well except for cluster-1. To verify the distribution, we further fit the data into multiple distributions, including *tlocationscale distribution*, *normal distribution*, *logistic distribution* and *extreme value distribution*. The fitting results are shown in Figure 4.

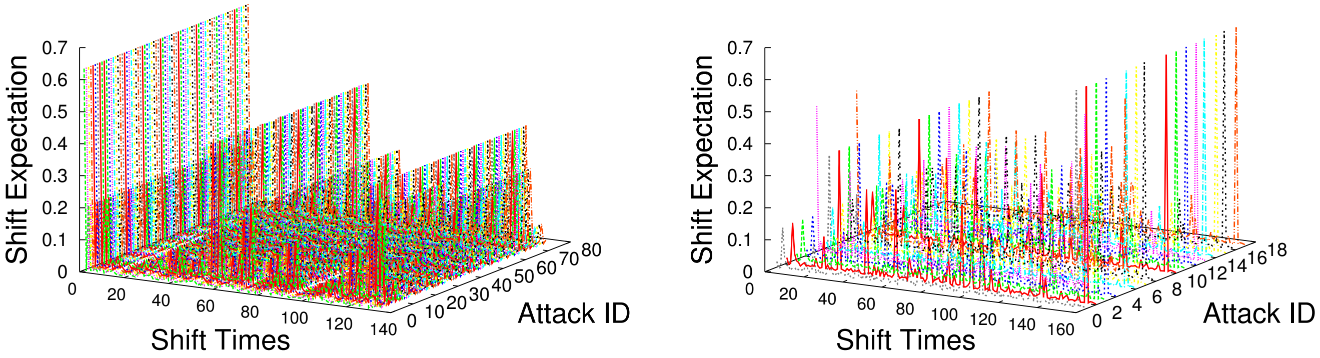
Except for the *extreme value distribution*, all other distribution functions are symmetric distributions. Figure 4 shows that the data fit the *tlocationscale distribution* best. *tlocationscale distribution* is the generalized *Student's t-distribution* into location-scale family. Location-scale family is a family of univariate probability distributions parameterized by a location parameter and a non-negative scale parameter. The *tlocationscale distribution* is useful for modeling data distribution with heavier tails than the *normal distribution*, meaning that it is more prone to producing values that fall far from its mean. This makes it useful for understanding the statistical behavior of certain types of ratios of random quantities. Nonetheless, this still indicates that the shift behaviors are predictable with the help of advanced time series modeling tools, such as Autoregressive Integrated Moving Average (ARIMA) algorithm, among others. In this case, the distribution describes the distances between multiple shift patterns of botnets. It also means that if we use the centroids of different clusters as baseline, we can learn and predict how bots will shift based on this distribution.

4.2.2 Potential Explanation

While there is potentially many explanations for this trend, none of those explanations alludes to an arbitrary behavior, but rather a controlled and systematic behavior. One explanation is that it is likely that attackers are utilizing this feature to arrange and control bots during attacks, especially with a large number of bots. Intuitively, such control is determined by the number of active bots during the attack, and has perhaps little to do with the time of the day in isolation. A further analysis that includes the time of the day in which the attacks happen is out of the scope of this paper, although perhaps worth investigating in the future.

From the defense perspective, such shift information can be very useful. On one hand, with this information—even though there might be more than one shift pattern per family—we can predict how attacks shift based on the distribution. On the other hand, we can simulate DDoS attacks behaviors, not only based on traffic volume but also by incorporating dynamics behind them.

Generalization to Other Families. We further explore by measurements whether such trend is applicable to other families. First, we apply the same analysis to the *Pandora* family. Similar to Figure 3, we also plotted a CDF for *Pandora*'s clusters, which



(a) *Pandora*: attacks with same bot shift pattern

(b) *Pandora*: attacks with similar bot shift pattern

Fig. 2: Shift pattern of the *Pandora* DDoS botnet family.

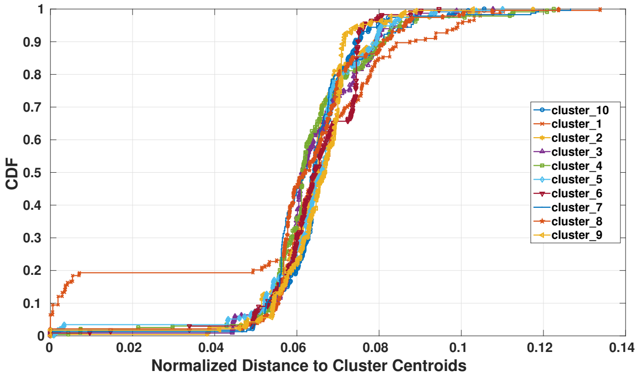


Fig. 3: Vector Distances CDF

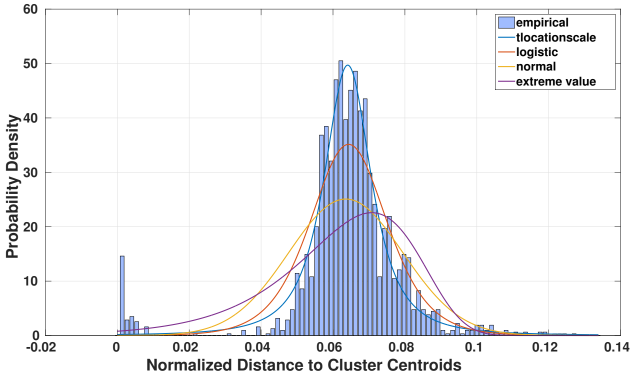


Fig. 4: Distribution Fit

confirmed similar behavioral patterns. However, compared to *Dirtjumper*, *Pandora* exhibits a slight deviation in the distribution, perhaps due to the smaller number of attacks launched by the *Pandora* botnet family compared to those launched by *Dirtjumper*.

Finally, results that were obtained by analyzing other families reveal similar findings, and highlight the power of mathematical distributions in fairly easily characterizing the shift patterns and revealing scheduling strategies.

TABLE 2: Statistic Information of *Pandora* Cluster

Size	Max_Diss	Avg_Diss	Diameter	Separation	Avg_Exp	Max_Exp	Std
97	3.69	0.18	3.73	3.46	0.03	0.74	0.076
74	0.06	0.02	0.09	3.62	0.03	0.63	0.08
20	4.52	1.46	4.99	2.39	0.025	0.88	0.073
17	3.48	0.41	3.48	2.39	0.028	0.68	0.075
7	4.18	1.06	4.30	3.57	0.03	0.88	0.08

4.2.3 Other Statistical Characteristics

Besides the pattern clustering graphs, we statistically analyze the different observed clusters. For that, Table 2 summarizes some statistical information about *Pandora*'s clusters. In this table the following are defined. 1) *Size* shows the size of each cluster. 2) *Max_Diss* represents the maximum distance between any two vectors in the same cluster. 3) *Diameter* shows how large are the different clusters. 4) *Separation* represents the minimal dissimilarity between an observation of the cluster and an observation of another cluster. 5) *Avg_Exp* shows the average shift expectation of each cluster. 6) *Std* is the standard deviation of expectations of each cluster. Statistically, the smaller the *Diameter*, the better the cluster. From this table, we can easily see that the second cluster is the best, which also conforms with Figure 2a. Though *Max_Diss* is larger than *Separation* in some clusters, both values are dictated by extreme values for each cluster. Thus, *Avg_Diss* and *Diameter* provide better reference for measurement of goodness of clusters. Another observation from this table is that for most clusters the *Diameter* is larger than the *Separation*, meaning that these clusters are not totally isolated. A total isolation means that the patterns might be attack-specific. However, results show the opposite: each cluster still shares some similarities with other clusters. This further indicates that there might be certain dynamic mechanisms behind each family call for further investigation.

4.2.4 Shift versus Attack Duration

One of the most important indicators of the persistence of an attack is the attack duration. For that, we attempt to understand how the attack duration is related to the shift pattern. Interestingly, we observe that the total shift expectation of a DDoS attack is inversely proportional to the length of the attacks, as shown in Figure 5. To make it clear, we truncate the figure by eliminating

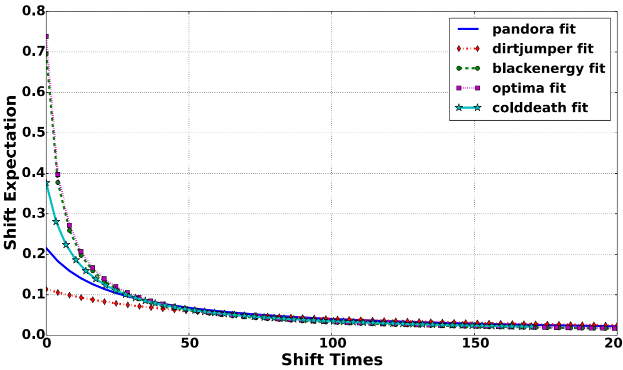


Fig. 5: Total shift expectation curves

attacks longer than 200. We confirm that this finding is consistent across multiple families, as highlighted in the following.

First, and to highlight the finding across multiple families, we include five active families in this measurement: *Blackenergy*, *Colddeath*, *Dirtjumper*, *Optima* and *Pandora*. As before, each DDoS attack is represented as a vector of shift expectations. The total expectation of each DDoS attack can be calculated as the sum of the whole vector. We then classify these vectors into different groups based on their lengths and calculate the average shift expectation within the same group. Thus, these attacks will be represented by groups of data pairs, denoted by the length of the attack and the averaged shift expectation. Finally, we use the non-linear least squares [26] to fit a function to these data points as shown in Figure 5.

Observations. First we calculate the standard error of the estimate for each family to measure the accuracy of the fitting. For all five families, the standard distances of the data points from the fitted lines are about 0.033, which are less than the average values and indicate that the fitted results are representative. Accordingly, the result clearly shows that the total shift expectation of DDoS attacks is inversely proportional to the length of the attacks. This means that the total shift expectation of attacks are basically the same no matter how long the attacks may last. This again can help us predict how the attacks are going to evolve by using the non-linear function. While the curves of *Dirtjumper* and *Pandora* are almost identical, the same pattern is seen between *Optima* and *Blackenergy*. These findings highlight two pairs of families with very high similarity, possibly because of related malware generation and collaborations. Thus, if we already know the relationship between multiple families, we can predict how similar the shift patterns will be for these families. Conversely, as more and more botnets start to collaborate with each other, this will help reduce the complexity of analyzing and detecting them.

Implications. This finding is particularly useful to targets that suffer persistent and long DDoS attacks because as the attack goes on, the attacking forces tend to stay stable, making it easier to identify the attackers. On the other hand, it also suggests that the early stage of DDoS attacks is very crucial for defense since the probability of success decreases dramatically afterwards.

4.3 Inter-family Analysis

In all results thus far, attack vectors have been shown to be composed of periodic-like spikes, which makes us wonder whether these vectors could be shared across multiple families based on their resemblance. Also, from our previous static analysis of various features of attacks, we have discovered collaborating botnet families [14]. To verify that they are actually collaborating, we follow a similar approach, and highlight further supporting evidence to such findings.

Notice that the purpose of the analysis in this section (as well as the analysis in Section 5) is to show the phenomenon of collaboration; inter and intra-family, as well as resources rotation (at the family level). Such phenomenon is not applicable to all attacks, and only finding attacks that exhibit the phenomenon using the tools proposed in this work is sufficient for our analysis, justifying the restriction of our analysis to those attacks that have those characteristics. We do not claim that the phenomenon is universal.

4.3.1 Results

For this analysis, we first conduct the clustering on the two most active families, *Dirtjumper* and *Pandora*. Figure 6a and Figure 6b show two of these clusters. This analysis of collaboration shows 449 attacks launched by *Dirtjumper* and *Pandora* in total, of which 234 were launched by *Dirtjumper* and the rest were launched by *Pandora*. We cluster these attacks into 20 clusters and then check which clusters involve collaborations. For the first collaboration, there were seven attacks from *Dirtjumper* and 74 from *Pandora*. For the second, there were 17 attacks from *Dirtjumper* and one attack from *Pandora*. Similarly, Figure 7a and Figure 7b show the results of clustering of *Dirtjumper* and *Blackenergy*. They are differentiated by different line styles with different colors; *Dirtjumper* is represented by a solid green line, whereas *Pandora* is represented by a dotted blue line and *Blackenergy* is represented by a dotted red line.

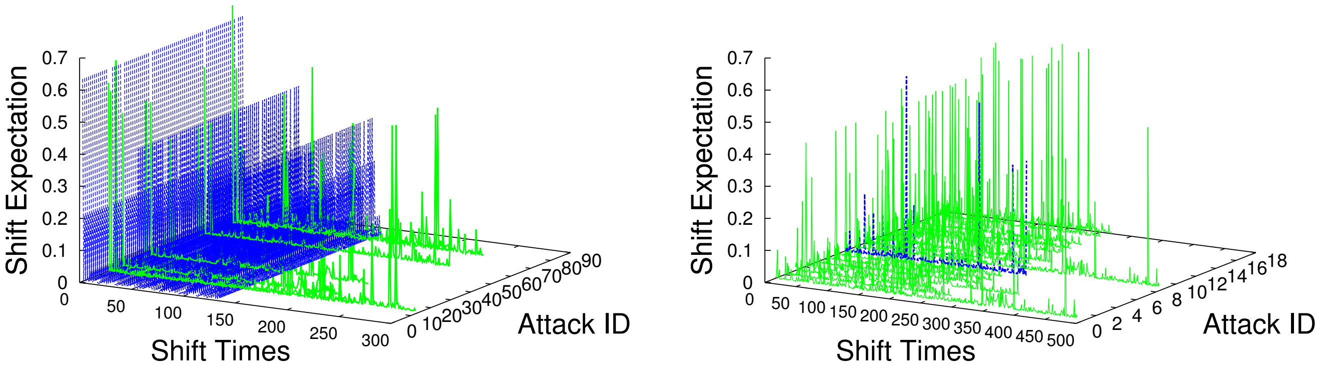
Collaborations involving more than two families can be discovered by classifying attacks into different groups based on the bot shift pattern. Figure 8a and Figure 8b show two examples. Figure 8a shows the clustering results of attacks whose length is smaller than 50 shifts: this collaboration involves seven *Optima* attacks, one *Colddeath* attack and one *Blackenergy* attack. Same as in the previous figures, different families have been denoted by different colors. Except for *Dirtjumper*, *Pandora* and *Blackenergy*, *Optima* is denoted by dotted cyan line and *Colddeath* by dotted black line.

4.3.2 Results Interpretation

The results confirm that shared shift patterns exist in different families. Looking into these attacks, we also find that these attacks were targeting different targets and are launched at different times. Traditionally, we could not detect collaborations between botnet families unless they launch attacks towards the same target or at the same time. However, such sophisticated collaborations can be revealed by the shared shift pattern among families. Furthermore, for the first collaboration, these exact 15 *Dirtjumper* attacks are also launched in a collaboration with *Pandora* (not shown in Figure 6a nor Figure 6b). This means that *Pandora* and *Blackenergy* have potential collaborations. More importantly, the same shift pattern applied to multiple attacks is more likely to come from the same group of bots, which may be very active in launching DDoS attacks. We will further examine these bots in Section 5.

Similarly, and for collaborations that include more than two families, similar interpretations are produced on Figure 8a and Figure 8b. All of these figures show the same patterns as in the previous figures, although applied to new families, and indicate

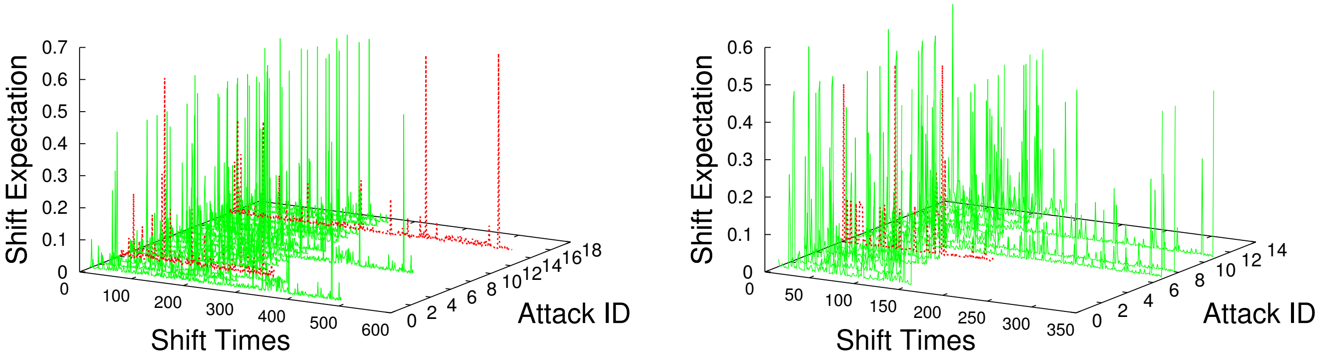
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2808344, IEEE Transactions on Dependable and Secure Computing



(a) Clustering finds out *Dirtjumper-Pandora* collaborations

(b) Clustering finds out *Dirtjumper-Pandora* collaboration

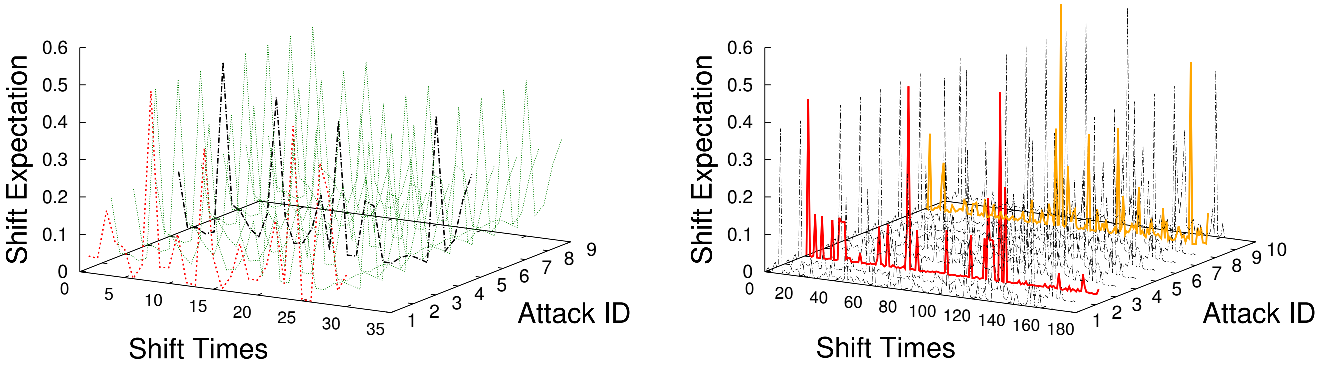
Fig. 6: The *Dirtjumper-Pandora* collaboration.



(a) Clustering finds out *Dirtjumper-Blackenergy* collaboration

(b) Clustering finds out *Dirtjumper-Blackenergy* collaboration

Fig. 7: The *Dirtjumper-Blackenergy* collaboration.



(a) Multiple Botnet family collaboration

(b) Multiple Botnet family collaboration

Fig. 8: Collaboration of more than two families.

that the potential collaboration is a universal feature among botnet families in launching DDoS attacks. The bots involved in collaborations might be the main attacking forces of this botnet family, they may live longer than other bots, and they may also cause more damage. We will discuss that more in Section 5.

In summary, this analysis shows that collaborations for launching attacks are prevalent, and are done between two or more families. As a result, these collaborations will inevitably make the defense more difficult than when done against a single family, especially if the defense included attribution in the wild. Finally, as a recommendation, we notice from our analysis that the targets that are often attacked by a certain botnet family should stay alert of DDoS attacks from other collaborating families as well.

5 Family-level bots rotation

In section 4, we discussed the country-level bots rotation, where the rotating bots might be distributed across several countries. Rotation is also an aspect that can be associated with a single family, and studying individual family's rotation patterns might highlight the various (and different) behavioral traits of different families. Accordingly, in this section we focus on measuring and discussing the implications of family-level rotation of multi-owned bots. Multi-owned bots here refer to the bots involved in multiple DDoS attacks at different times. Such DDoS attacks could be launched by a single botnet family (c.f. §5.1) or by multiple families (c.f. §5.2).

Feature Extraction and Families Selection. We extract information associated with all bots belonging to each family and sort them based on the timestamps when the DDoS attacks happened while they are involved in such attack. In this way, each bot in the dataset has a string of timestamps indicating its attack participation history. Our analysis starts from intra-family rotation. For highlighting the findings, we focus on *Dirtjumper*, *Pandora*, *Blackenergy* and *Optima* because they have the largest number of multi-owned bots.

5.1 Intra-family Analysis

In the following, we review findings and implications of multi-owned bots, their activity level, and rotation patterns.

5.1.1 Multi-owned Bots

We first study basic statistics of multi-owned bots for these four families. The results are shown in Table 3. In Table 3, the second column represents the percentage of multi-owned bots among all bots belonging to that family; and the third column represents the percentage of DDoS attacks which involved multi-owned bots among all the DDoS attacks launched by that family. From this table, we first observe that for *Pandora* more than 60 percent of the bots reappeared, which may indicate that the defense mechanisms targeting *Pandora* are not very effective or that *Pandora* botnet is adaptive, and is successful in covering its behavioral and detectable features and trails. While for the other three families, and even though the percentages of multi-owned bots are not very high, they were involved in over one-third of all attacks. This means that studying the behavior of multi-owned bots is necessary and may help mitigate DDoS attacks.

TABLE 3: Statistic Information of Multi-owned Bots

Family	Multi-owned bots (%)	Attacks by multi-owned bots
Pandora	62.66%	92.16%
Optima	18.48%	43.16%
Dirtjumper	12.67%	38.75%
Blackenergy	14.08%	32.79%

5.1.2 Activity Level

We now characterize these bots based on the length of their "history", which is the number of DDoS attacks they are involved in. We choose this approach to analysis mainly because it is easier to perform. Accordingly, we classify these bots into five different categories based on the length: $[2, 10)$, $[10, 20)$, $[20, 40)$, $[40, 80)$ and $[80, \infty)$. The statistical result is shown in Figure 9 and explained in the following.

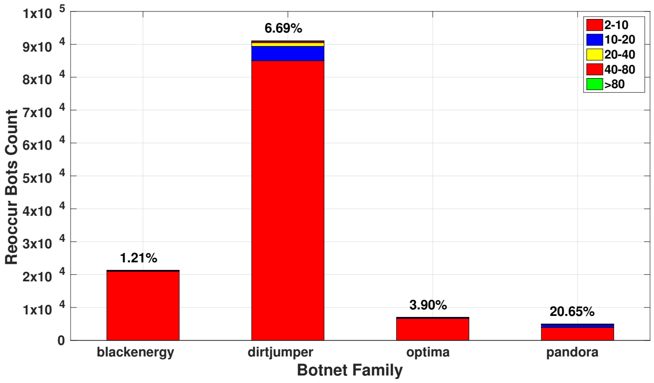


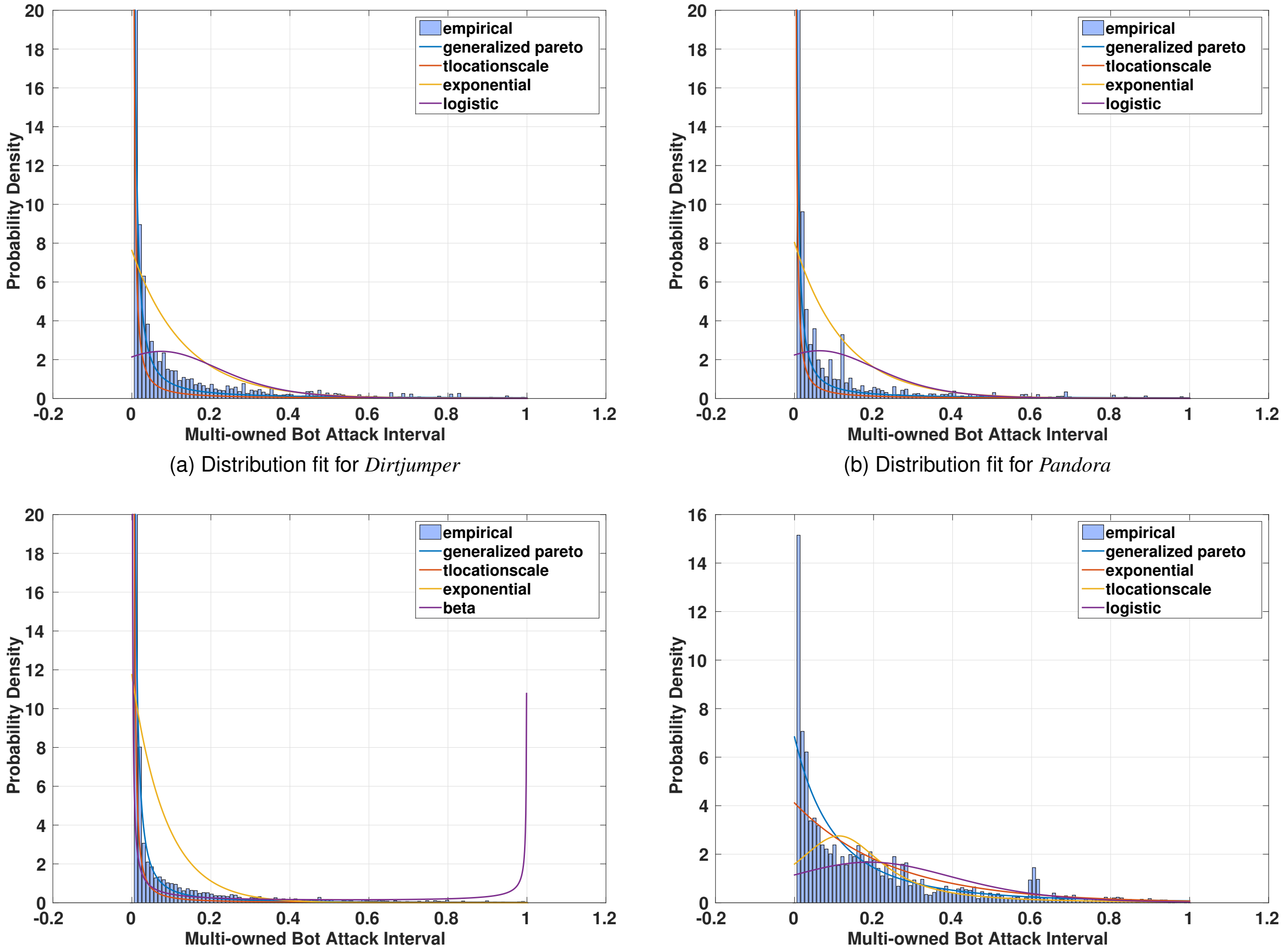
Fig. 9: Statistics of Multi-owned Bots

In Figure 9, each bar on the x-axis represents a botnet family while the y-axis represents the number of multi-owned bots and the number on top of each bar represents the percentage of bots that have involved in more than 10 DDoS attacks. It is evident that the bots in category $[2, 10)$ are much more than in any other category. After further looking into the data, we found that in category $[2, 10)$, most bots recorded got involved in only two DDoS attacks. Not only that it is not very helpful for the analysis with two attack intervals, but also bots involved in more DDoS attacks are much more critical for defense since they will cause more damage. As such, all of the following analyses will focus on bots that participated in more than 10 DDoS attacks.

5.1.3 Rotation Patterns

To better understand and characterize multi-owned bots behavioral rotation pattern, we will use attack intervals to illustrate rotation activities. Since each DDoS attack comes with a timestamp, it is convenient to calculate the time interval between two successive DDoS attacks. The attack interval combined with the total lengths of attacks is a metric we use to measure the rotation activities of multi-owned bots. Based on that, we first extract all bots belonging to the category described above. Then we can obtain a sequence of time intervals for each multi-owned bot.

Next, similar to what we did for the shift pattern analysis, we first try to fit these attack intervals with different distributions, including the *generalized pareto distribution*, the *exponential distribution*, the *tlocationscale distribution* and the *logistic distribution*. The results of all four families are shown in Figure 10a through Figure 10d.



(a) Distribution fit for *Dirtjumper*

(b) Distribution fit for *Pandora*

(c) Distribution fit for *Blackenergy*

(d) Distribution fit for *Optima*

Fig. 10: Multi-owned bots and their attack intervals (used for indicating the activity level).

In Figure 10a through Figure 10d, the bars represent the empirical probability density while the curves represent the fitted distribution based on the empirical results. From these four figures, we can observe that all of the intervals follow the *generalized pareto distribution* best. Similar to the *exponential distribution*, the *generalized pareto distribution* is often used to model the tail of another distribution. Usually, the model built on top of the *generalized pareto distribution* is used to fit extremes of complex data. The *generalized pareto distribution* also plays a vital role in network modeling. Most QoS research assumes exponential arrivals for ease of modeling. Some previous work, *e.g.*, Feldmann *et al.*'s [27] and Harchol-Balter *et al.*'s [28], show that most of the Internet traffic is better modeled with a heavy-tailed distribution such as the *generalized pareto distribution*.

5.1.4 Interpretation and Implications

Such trend in distribution indicates that the botmaster maintains a certain attacking strategy to manage his own bots because of the diverse patterns of attack intervals of multi-owned bots to avoid detection. On the positive side, this means that the defense mechanisms can utilize such patterns of multi-owned bots to make fine-grained botnet attack simulations. Combined with previous shift patterns, DDoS attacks could be reconstructed more accurately. Besides the intra-family multi-owned bots, there are also a large portion of bots rotating across-multiple families. These bots make the detection more difficult. Thus, it is important that we study their behaviors as well.

5.2 Cross-family Analysis

To analyze the behavior of cross-family bots, we first need to extract this kind of bots from our dataset and the basic statistical results are shown in Table 4. In this table, we only list the four families we studied in last subsection. However, bots sharing is common among all botnet families. From this table, we can observe that family *Dirtjumper*/*Pandora*, *Dirtjumper*/*Optima* and *Dirtjumper*/*Blackenergy* share the most bots, indicating that they may have more collaborations with each other than any other family pairs. Some collaborations have been confirmed in 4.3.

5.2.1 Attack Interval and Activity Level

The statistical result of cross-family multi-owned bots is shown in Figure 11. Based on the same reasoning, we focus on the analyses of bots that are involved in more than 10 DDoS attacks for cross-family multi-owned bots. Note that among these bots, there are 909 bots active in more than two families. 669 of them have been active in three families and the rest have been in all four families. Our following analyses target the bots that only rotate between two different families.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2808344, IEEE Transactions on Dependable and Secure Computing

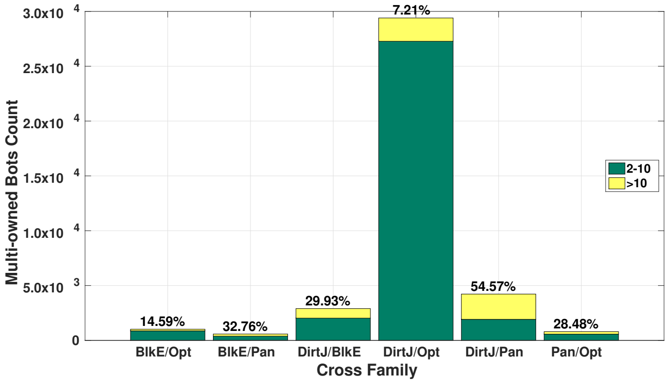


Fig. 11: Statistics of Cross-family Multi-owned Bots

TABLE 4: Statistic Information of Cross-family Multi-owned Bots

-	*Optima*	*Dirtjumper*	*Blackenergy*
Pandora	797	4224	580
Optima	-	29400	1021
Dirtjumper	-	-	2900

Similar to the inter-family analysis, we use the attack interval to represent rotation activities. We again fit the attack intervals of cross-family bots to multiple distributions. As a result, we obtained almost identical results to the intra-family analysis. We omit the figures for brevity. As before, the intervals also follow the *generalized pareto distribution*, but with different parameters. This further confirms that some attacking strategies are also applied to multi-owned bots across families since it is less likely that they switch to another family on their own. The rotations between different families make the attacking behavior more complex and thus more difficult to defend against. Knowing the rotation interval pattern, however, sheds some light on predicting the rotation behavior.

In addition to the attack intervals, another very important factor for the cross-family multi-owned bots is their activity levels, denoted by how often they switch to another family. To quantify their activity levels, we sort the attack activities from both families of multi-owned bots we got previously in time order. Then we label each activity in this sequence with either 0 or 1 depending on which family they belong to when the attack happened. In this way, the activities of these bots are simplified into a binary sequence. Next, we quantify the activity level through the changes in the sequences, either from 0 to 1 or the other way around. Thus, the activity level is calculated as the ratio of the number of family switches over the total activity counts. Accordingly, the activity level should be a number between 0 and 1, and is interpreted as the closer it gets to 1, the more active the bot is. The CDF of the activity levels is shown in Figure 12

In Figure 12, the x-axis represents the activity levels, while the y-axis represents the CDF. From this figure, we observe that about 80% of the bots have an activity level less than 0.45, meaning that bots are not very active in terms of family switches. For defenses, this is a good indicator, since if bots are more likely to stay in one family, then their behaviors do not change rapidly either. Of all the bot activities, the most active bot switches between family *Dirtjumper* and *Optima* alternatively. It even stays active in both families during the same time interval involving two different DDoS attacks launched by these two families, respectively. Such behaviors help bots evade detection easily.

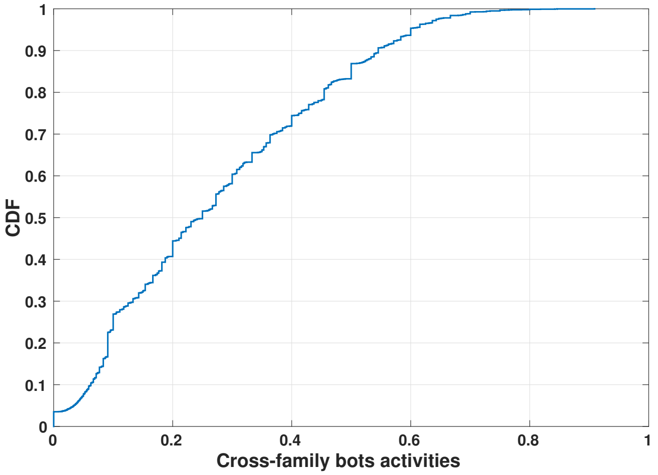


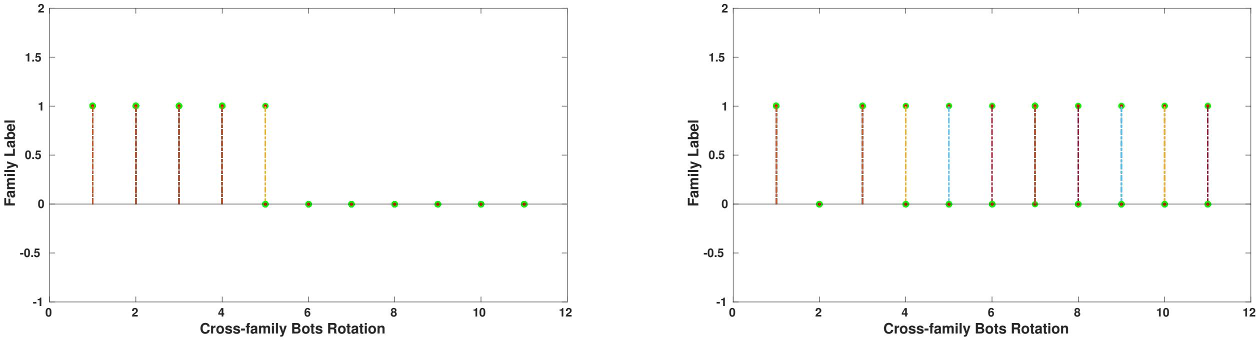
Fig. 12: CDF of cross-family bots activities

5.2.2 Comparing Activity Levels

Now that all the activities of cross-family multi-owned bots are simplified as binary sequences, we are able to compare these sequences numerically. Because there are too many sequences involved, we first classify them into different categories based on their length. From the basic statistical results we obtain for these different categories, we find that the bots that have between 10 and 20 activities are the majority. Also, bots activities are much more intensive in category $[10, 20)$ than any other intervals. After further investigation, we found that there are 515 bots that have 11 activities in group $[10, 20)$. For demonstration of results, our next analysis focuses on this particular group.

Since all bots in our studied groups have the same length, we are able to compare them by calculating the distances between them via the Hamming distance. In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. In other words, the Hamming distance measures the minimum number of substitutions required to change one string into another, or the minimum number of errors that could have transformed one string into another. In our case, we use the Hamming distance to measure the similarity of two binary sequences, where a higher similarity indicates higher rotation consistency. Next, we use K-means to cluster this group of bots into multiple clusters, where the distance matrix is obtained by calculating the Hamming distance between each pair of bots. The bots belonging to this group can be clustered into 10 isolated clusters. The clustering results of the two largest clusters are shown in Figure 13a and Figure 13b.

In Figure 13a and Figure 13b, the x-axis represents the length of bots activities, while the y-axis represents the label for different families, which is either 0 or 1. For a clear visualization of the results, the y-axis has been adjusted to $[-1, 2]$ in the figure. Each marker in the figure represents one activity and each bar connecting the markers indicates family label switching. The different colors of the bars denote different rotation patterns. For example, in Figure 13a, there are two colors of the bars, red and yellow, which means there are two rotation patterns in this cluster. However, only one of the bars has markers on both ends, meaning that these two rotation patterns are almost identical except one activity. Based on this observation, we can infer that the bots in



(a) Cluster with identical rotations

(b) Cluster with very similar rotations

Fig. 13: Rotation patterns for various clusters.

this cluster rotate almost identically. We further looked into the dataset and found that there are 114 bots in this cluster, and they belong to the botnet families *Pandora* or *Dirtjumper*. Figure 13b is similar to Figure 13a, although with more rotation patterns. There are four different rotation patterns and they belong to all four botnet families.

Note here that we have only analyzed bots of the same activity length since the Hamming distance can only be calculated between two sequences of the same length. We also tried the Dynamic Time Warping as we did with the bot shift pattern, but initial findings indicate that it is not efficient and effective when dealing with a large amount of binary sequences. To this end, we leave addressing this issue as a potential future work, where we will try other similarity metrics to further explore the rotation behavior of cross-family multi-owned bots.

5.2.3 Further Interpretation and Implications

The above analysis shows how important it is to understand the behaviors of multi-owned bots, both at the intra-family and inter-family levels. Not only because that the multi-owned bots can cause more damage, but also because their repeated appearances prove the difficulty to take them down. They also provide a new angle to look into the attacking strategy utilized by the attacker, which will benefit the security community by refreshing our knowledge of botnet DDoS attack behaviors. From the defense perspective, the behavior of single bot might be more of interest since there is no effective method to isolate malicious bots from legitimate requests. The existence of such behavior pattern provides a possibility to achieve this goal.

6 Related Work

DDoS attacks have been intensively investigated and numerous measurement works have been done to help achieve better understanding of them. In 2006, Mao *et al.* [13] presented their measurement work of DDoS attacks relying on both direct measurement of flow-level information and more traditional indirect measurements using backscatter analysis. Findings in this work are a decade old, and our findings in this paper update such results in various ways. Moore *et al.* [29] conducted a backscatter analysis for quantitatively estimating DoS activity in the Internet based on a three-week dataset. Due to the growth of network address translation and firewall techniques, much of the Internet was precluded from the study by the traditional network measurement techniques. Our study relies on 7 months observation of large number of botnets.

In 2005, Casado *et al.* [11] proposed an opportunistic measurement approach that leverages sources of spurious traffic, such as worms and DDoS backscatter, to unveil unseen portion of Internet. The monitoring of packets destined for unused Internet addresses, termed as “background radiation”, proved to be another useful technique to measure Internet phenomenon. In 2004, Pang *et al.* [30] conducted an initial study of broad characteristics of Internet background radiation by measuring traffic from four large unused subnets. A recent study [31] revisited the same topic and characterized the current state of background radiation specifically highlighting those which exhibit significant differences. Our work serves as a revisit to those studies with new insights, and utilizes direct measurements of DDoS attacks, as indirect measurements through inferences using backscatter.

Similar in purpose to the tool utilized in obtaining our data, Bailey *et al.* [32] designed and implemented the Internet Motion Sensors (IMS), a globally scoped Internet monitoring system to detect Internet threats, which includes a distributed blackhole network with a lightweight responder and a novel payload signature and caching mechanism. Xu *et al.* [33] presented a general methodology to build behavior profiles of Internet backbone traffic in terms of communication patterns of end-hosts and services.

All of the aforementioned studies rely on static analysis of DDoS attacks. So far little has been done to analyze the dynamics of DDoS attacks. Gu *et al.* [6] designed and developed a new detection framework by utilizing clustering analysis of botnet communication patterns in 2008. Based on this work, in 2010, Perdisci *et al.* [34] presented a novel network-level behavioral malware clustering system. Lu *et al.* [5] proposed a new approach for detecting and clustering botnet traffic on large-scale network traffic payload signatures. The emergence of these malware behavior analyses indicates that researchers start to expand the traditional static measurements and explore the dynamics. Our work explores yet unrevealed aspect of today’s DDoS attacks through their dynamics.

In our work, we focus on DDoS dynamics. We use several techniques including K-means clustering and Dynamic Time Warping (DTW). DTW was first introduced in the data mining community in the context of mining time series proposed by Berndt *et al.* [35]. Several techniques have been introduced to speed up DTW and to reduce the space overhead [36], [37]. The

K-means clustering methods we use were first proposed by Lloyd *et al.* [38], and remains a very popular method for clustering after many years perhaps due to its simplicity and effectiveness in practice. These techniques successfully helped us discover the principles of the dynamics behind the scenes.

While remotely related, there has been a lot of work in the literature on various aspects of the design of (DDoS) attacks and defenses. This line of work includes the kinds of Kang *et al.*'s [39], [40], [41], Schuchard *et al.*'s [42], [43], Li *et al.*'s [44], Walfish *et al.*'s [45], among others [46], [47], [48], [49], [50]. While broadly related, they mainly treat protocol-level characteristics for defenses, and do not handle or take into account strategies of the attacker. We believe that findings in this study can guide defenses proposed in those works and validate (or invalidate) their attacks and defenses. Pursuing such research direction of a data-driven approach to defenses is an open direction that we would like to pursue in the future.

7 Conclusion

DDoS attacks remain one of the most challenging threats on the Internet, despite numerous efforts to characterize, model, and defend against them. This indicates that increasingly sophisticated strategies are being employed by the DDoS attackers. Successful defenses demand in-depth understanding of their strategies. In this work, we have conducted an analysis on a large scale DDoS dataset, aiming to uncover the dynamics of the DDoS attack strategies behind the scenes. With the help of Dynamic Time Warping and clustering, we have found that attackers are deliberately and dynamically deploying their attack forces in individual or collaborative attacks, indicating the strong bond and organization of different botnet families in various attacks. Furthermore, such dynamics can be well captured by statistical distributions. Contrary to conventional understanding of botnet, we discovered that some bots rotate in different families, and such rotating patterns can also be mathematically characterized. These results add to the existing literature of DDoS characterization and understanding. More importantly, they lay a promising foundation for us to predict the dynamics during a DDoS attack in the future, which could be utilized to enhance existing defenses.

Acknowledgments

The authors would like to thank Gianluca Stringhini for his valuable feedback on the preliminary results in [1]. This work is partially supported by National Science Foundation (NSF) under grants CNS-1117300 and CNS-1643207, and the Global Research Lab. (GRL) Program of the National Research Foundation (NRF) funded by Ministry of Science, ICT (Information and Communication Technologies) and Future Planning (NRF-2016K1A1A2912757). The views and opinions expressed in this paper are the views of the authors, and do not necessarily represent the policy or position of the funding agencies.

References

[1] A. Wang, A. Mohaisen, W. Chang, and S. Chen, "Revealing ddos attack dynamics behind the scenes," in *Proc. of DIMVA*, 2015.

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

[3] A. Welzel, C. Rossow, and H. Bos, "On measuring the impact of ddos botnets," in *Proc. of EuroSec*. ACM, 2014, p. 3.

[4] Y. Nadji, M. Antonakakis, R. Perdisci, D. Dagon, and W. Lee, "Beheading hydras: performing effective botnet takedowns," *Proc. of ACM SIGSAC*, pp. 121–132, Nov. 2013.

[5] W. Lu, G. Rammidi, and A. A.Ghorbani, "Clustering botnet communication traffic based on n-gram feature selection," in *Computer Communications*, 2011.

[6] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol- and structure-independent botnet detection," in *Proc. of USENIX Security*, 2008.

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

[8] T. Robinson, "Breaches, malware to cost $491 billion in 2014, study says," http://bit.ly/1gNXu90, 2014.

[9] M. Karami and D. McCoy, "Understanding the emerging threat of ddos-as-a-service." in *Proc. of USENIX LEET*, 2013.

[10] A. Büscher and T. Holz, "Tracking ddos attacks: Insights into the business of disrupting the web," in *Proc. of USENIX LEET*, 2012.

[11] M. Casado, T. Garfinkel, W. Cui, V. Paxson, and S. Savage, "Opportunistic measurement: Extracting insight from spurious traffic," in *Proc. of ACM Hotnets*, 2005.

[12] S. Jin and D. Yeung, "A covariance analysis model for ddos attack detection," *IEEE ICC*, 2004.

[13] Z. M. Mao, V. Sekar, O. Spatscheck, J. van der Merwe, and R. Vasudevan, "Analyzing Large DDoS Attacks using Multiple Data Sources," *Proc. of ACM SIGCOMM LSAD*, 2006.

[14] A. Wang, W. Chang, A. Mohaisen, and S. Chen, "How distributed are today's ddos attacks?" in *Proc. of ACM CCS*, 2014.

[15] W. Chang, A. Wang, A. Mohaisen, and S. Chen, "Characterizing botnets-as-a-service," in *Proc. of ACM SIGCOMM*, 2014.

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

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

[18] K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim, "DDoS attack detection method using cluster analysis," *Expert Systems with Applications*, vol. 34, pp. 1659–1665, 2008.

[19] A. Mohaisen and O. Alrawi, "Av-meter: An evaluation of antivirus scans and labels," in *Proc. of DIMVA*, ser. LNCS, vol. 8550. Springer, 2014, pp. 112–131.

[20] ——, "Amal: High-fidelity, behavior-based automated malware analysis and classification," *Computers & Security*, 2015.

[21] B. B. Kang, E. Chan-Tin, C. P. Lee, J. Tyra, H. J. Kang, C. Nunnery, Z. Wadler, G. Sinclair, N. Hopper, D. Dagon *et al.*, "Towards complete node enumeration in a peer-to-peer botnet," in *Proc. of ACM ASIA CCS*. ACM, 2009, pp. 23–34.

[22] M. Thomas and A. Mohaisen, "Kindred domains: detecting and clustering botnet domains using DNS traffic," in *Proc. of WWW*, 2014, pp. 707–712.

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

[24] A. G. West and A. Mohaisen, "Metadata-driven threat classification of network endpoints appearing in malware," in *Proc. of DIMVA*, vol. 8550, 2014, pp. 152–171.

[25] A. Wang, A. Mohaisen, W. Chang, and S. Chen, "Delving into Internet DDoS Attacks by Botnets: Characterization and Analysis," *Proc. of IEEE/IFIP DSN*, 2015.

[26] K. Levenberg, "A method for the solution of certain non–linear problems in least squares," 1944.

[27] A. Feldmann, "Characteristics of tcp connection arrivals," *Self-Similar Network Traffic and Performance Evaluation*, pp. 367–397, 2000.

[28] M. Harchol-Balter, "Network analysis without exponentiality assumptions," Ph.D. dissertation, UNIVERSITY of CALIFORNIA at BERKELEY, 1996.

[29] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, "Inferring internet denial-of-service activity," *ACM TOCS*, vol. 24, no. 2, pp. 115–139, 2006.

[30] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of internet background radiation," in *Proc. of ACM IMC*. ACM, 2004, pp. 27–40.

[31] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, "Internet background radiation revisited," in *Proc. of ACM IMC*. ACM, 2010, pp. 62–74.

[32] M. Bailey, E. Cooke, F. Jahanian, J. Nazario, D. Watson *et al.*, "The internet motion sensor-a distributed blackhole monitoring system." in *Proc. of NDSS*, 2005.

[33] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, “Profiling internet backbone traffic: behavior models and applications,” in *Proc. of ACM SIGCOMM CCR*, vol. 35, no. 4. ACM, 2005, pp. 169–180.

[34] R. Perdisci, W. Lee, and N. Feamster, “Behavioral clustering of http-based malware and signature generation using malicious network traces,” in *Proc. of NSDI*, 2010.

[35] D. J.Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” in *Proc. of KDD Workshops*, 1994.

[36] S.-W. Kim, S. Park, and W. W.Chu, “An index-based approach for similarity search supporting time warping in large sequence databases,” in *Proc. of IEEE ICDE*, 2001.

[37] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” in *Knowledge and Information Systems*, 2005.

[38] S. P.Lloyd, “Least squares quantization in pcm,” in *IEEE Transactions on Information Theory*, 1982.

[39] M. S. Kang, S. B. Lee, and V. D. Gligor, “The crossfire attack,” in *Proc. of IEEE S&P*. IEEE, 2013, pp. 127–141.

[40] M. Kang, V. D. Gligor, and V. Sekar, “Spiffy: Inducing cost-detectability tradeoffs for persistent link-flooding attacks.” in *Proc. of NDSS*, 2016.

[41] M. S. Kang, V. D. Gligor, and V. Sekar, “Spiffy: Inducing cost-detectability tradeoffs for persistent link-flooding attacks,” in *Proc. of NDSS*, 2016.

[42] M. Schuchard, A. Mohaisen, D. F. Kune, N. Hopper, Y. Kim, and E. Y. Vasserman, “Losing control of the internet: Using the data plane to attack the control plane,” in *Proc. of NDSS*, 2011.

[43] M. Schuchard, C. Thompson, N. Hopper, and Y. Kim, “Peer pressure: exerting malicious influence on routers at a distance,” in *Proc. of ICDCS*. IEEE, 2013, pp. 571–580.

[44] Q. Li, X. Zhang, X. Zhang, and P. Su, “Invalidating idealized bgp security proposals and countermeasures,” *IEEE TDSC*, vol. 12, no. 3, pp. 298–311, 2015.

[45] M. Walfish, M. Vutukuru, H. Balakrishnan, D. Karger, and S. Shenke, “DDoS defense by offense,” *Proc. of ACM SIGCOMM*, pp. 303–314, 2006.

[46] J. Ioannidis and S. M. Bellovin, “Implementing pushback: Router-based defense against DDoS attacks,” in *Proc. of NDSS*, 2002. [Online]. Available: https://www.cs.columbia.edu/∼smb/papers/pushback-impl.pdf

[47] 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.

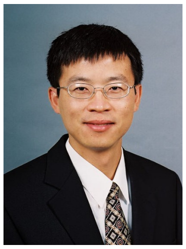
[48] J. Mirkovic, M. Robinson, and P. Reiher, “Alliance formation for ddos defense,” in *Proc. of NSPW*. ACM, 2003, pp. 11–18.

[49] G. Oikonomou, J. Mirkovic, P. Reiher, and M. Robinson, “A framework for a collaborative ddos defense,” in *Proc. of ACM ACSAC*. IEEE, 2006, pp. 33–42.

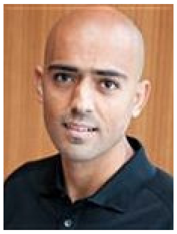
[50] A. Merlo, M. Migliardi, N. Gobbo, F. Palmieri, and A. Castiglione, “A denial of service attack to umts networks using sim-less devices,” *IEEE TDSC*, vol. 11, no. 3, pp. 280–291, 2014.

An Wang is a PhD student in Computer Science at George Mason University. She received the B.S. degree in Computer Science from Jilin University, China in June 2012. Her research interests include system and network security and software-defined networks.

Wentao Chang 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 towards the Ph.D. degree in Computer Security at George Mason University. 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 an associate professor of computer science at George Mason University. His research interests include the Internet content delivery systems, Internet measurement and modeling, operating systems and system security, and distributed systems and high performance computing. 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 Assistant Professor in the Computer Science Department at the University at Buffalo, SUNY. Previously, he was a Senior Research Scientist at Verisign Labs (2012 to 2015) and a Member of Engineering Staff at ETRI in Daejeon, South Korea (2007 to 2009). Starting in the Fall of 2017, he will be an Associate Professor in the Department of Computer Science, with a joint appointment in the Department of Electrical and Computer Engineering, at the University of Central Florida. His research interests are in the areas of systems, security, privacy, and measurements. He won the Summer Faculty Fellowship from the United States Air Force Office of Scientific Research (2016), the Best Paper Nominee at ICDCS (2017), the Best Paper Award at WISA (2014), the Best Poster Award at IEEE CNS (2014), and a Doctoral Dissertation Fellowship from the University of Minnesota (2011-2012), among other honors. He was recognized for his service to IEEE INFOCOM (2017) and IEEE CNS (2016). His research work has been supported by various awards from NSF, AFOSR, and AFRL, and his research results have been featured in popular media, including MIT Technology Review, the New Scientist, Minnesota Daily, Slashdot, The Verge, Deep Dot Web, and Slate, among others. He is a member of ACM and a senior member of IEEE.