Inferring distributed reflection denial of service attacks from darknet

Claude Fachkha *, Elias Bou-Harb, Mourad Debbabi

Computer Security Laboratory, Concordia University, Canada
National Cyber-Forensics and Training Alliance, Canada

ABSTRACT

This work proposes a novel approach to infer and characterize Internet-scale DNS Distributed Reflection Denial of Service (DRDoS) attacks by leveraging the darknet space. Complementary to the pioneer work on inferring Distributed Denial of Service (DDoS) activities using darknet, this work shows that we can extract DDoS activities without relying on backscattered analysis. The aim of this work is to extract cyber security intelligence related to DRDoS activities such as intensity, rate and geo-location in addition to various network-layer and flow-based insights. To achieve this task, the proposed approach exploits certain DDoS parameters to detect the attacks and the expectation maximization and k-means clustering techniques in an attempt to identify campaigns of DRDoS Attacks. We empirically evaluate the proposed approach using 1.44 TB of real darknet data collected from a/13 address space during a recent several months period. Our analysis reveals that the approach was successful in inferring significant DNS amplification DRDoS activities including the recent prominent attack that targeted one of the largest anti-spam organizations. Moreover, the analysis disclosed the mechanism of such DNS amplification attacks. Further, the results uncover high-speed and stealthy attempts that were never previously documented. The extracted insights from various validated DNS DRDoS case studies lead to a better understanding of the nature and scale of this threat and can generate inferences that could contribute in detecting, preventing, assessing, mitigating and even attributing of DRDoS activities.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Cyber attacks continue to threaten today's information technology. These threats are growing dramatically in terms of size and impact targeting large organizations, Internet service providers and governments. A DDoS attack is one of the major cyber attacks that attempts to make a computer or network resources unavailable. DDoS activities, indeed, dominate today's attack landscape. In a recent report by Arbor Networks [1], it was concluded that 48% of all cyber threats are DDoS. Further, it was stated that the top 4 perceived threats for the next 12 months will be DDoS related, targeting customers, network and service infrastructure. Governmental organizations, corporations as well as critical infrastructure were also recently deemed as DDoS victims [2–4].

A DNS-based DRDoS attack is a form of DDoS that relies on the use of publicly accessible open recursive DNS servers to overwhelm a victim system with DNS response traffic [5]. A recent event demonstrated that even a cyber security organization became a victim of the largest (i.e., 300 Gbps) DNS amplification DDoS attack in history [6]. The above facts concur that DDoS attacks in general, and DRDoS in particular, are and will continue to be a significant cyber security issue, causing momentous damage to a targeted victim as well as negatively affecting, by means of collateral damage, the network infrastructure (i.e., routers, links, etc.), the finance, the trust in, and the reputation of the organization under attack.

In this work, we tackle the following questions:

1. How to infer large-scale DNS-based DRDoS activities?
2. What are the characteristics of DNS amplification DRDoS attacks?
3. What inferences can we extract from analyzing DNS DRDoS traces?

Answering those questions would aid computer security response teams, law enforcement agencies and governments to build a darknet-based central infrastructure to scrutinize DNS-based amplification traffic in order to contribute in understanding, detecting, preventing, assessing, mitigating and even attributing of DRDoS attacks.
In this context, we frame this paper's contributions as follows:

- Proposing a systematic flow-based approach for inferring DNS amplification DDoS activities by leveraging DNS queries to darknets.
- Characterizing the inferred DDoS threats during several months period.
- Applying clustering and similarity algorithms in an attempt to identify campaigns of DNS amplified DDoS attacks.

The remainder of this paper is organized as follows: In Section 2, we provide an overview and background information on DNS amplification attacks and the darknet space. In Section 3, we present the proposed approach and elaborate on various aspects of its components. In Section 4, we empirically evaluate the approach and disclose several DNS amplified DDoS case studies. In Section 5, we survey the related work. Finally, Section 6 summarizes the paper, pinpoints some lessons learned and discusses the future work.

2. Background

In this section, we provide some background information related to the mechanism of DNS amplified DDoS attacks, the darknet space and DNS queries targeting the darknet.

2.1. DNS-based DDoS attacks

A DNS amplification attack is a well known practice of a DDoS, in which malicious users abuse open DNS servers to bombard a victim with DNS reply traffic [5]. The technique consists of an invader directing a DNS name lookup query to an open DNS server having the source IP spoofed to be the victim’s address. Subsequently, all DNS server responses will be sent to the targeted victim. In general, malicious users will request domains that cover a large zone to increase the amplification factor. In order to have a high impact on the victim, the attackers use DNS requests of type ANY to return all possible known information to the victim, and hence increase the amplification of the attack. Moreover, in order to increase the size of the attack with little effort, attackers use bots (i.e., campaigns) [7] to synchronize an army of bots and order them to send the DNS requests. Based on such concepts, Fig. 1 depicts a basic DNS amplification attack with recursive DNS. In the first two steps, the attacker uses a botnet to generate spoofed DNS lookup requests to the Internet. In step 3–7, the internal and external DNS servers collaborate in order to provide an answer to the requester. Finally, in step 8 and 9, the amplified replies congest the victim’s computer and network resources with a large flood of traffic.

2.2. Darknet space

In a nutshell, darknet traffic is Internet traffic destined to unused Internet addresses (i.e., dark sensors). Since these addresses are unallocated, any traffic targeting such space is suspicious. Darknet analysis has shown to be an effective method to generate cyber threat intelligence [8, 9]. Darknet traffic is typically composed of three types of traffic, namely, scanning, backscattered and misconfiguration [10]. Scanning arises from bots and worms while backscattered traffic commonly refers to unsolicited traffic that is the result of responses to DDoS attacks with spoofed source IP addresses. On the other hand, misconfiguration traffic is due to network/routing or hardware/software faults causing such traffic to be sent to the darknet sensors.

2.3. DNS queries on darknet

On the darknet space, one can also observe a significant number of DNS queries that could be sent by the following sources:

- Attacker spoofing the victim’s IP: This scenario is depicted in Fig. 2a. In this case, the attacker sends spoofed DNS queries on the Internet address space using the victim’s IP address. All replies from the open DNS resolvers (i.e., hosts X and Z) will bounce back towards the victim.
- Compromised victim: This scenario is depicted in Fig. 2b. In this case, the attacker uses the victim’s machine to send DNS queries. The attacker might use several techniques to control the victim’s machine, including malware infection and/or vulnerability exploitation. This scenario does not involve spoofed DNS queries.
- Scanner: In this scenario, the attacker scans the Internet to infer the locations of open DNS resolvers. This task requires collecting information from the reply packets and hence, a non-spoofed address is used by the scanners.
- Others: Other hosts may include firewalls to reduce the impact of the attack or misconfigured devices, etc.

In our work, we assert that high speed ANY DNS queries [5] will be sent from an attacker spoofing the victim’s IP and/or compromised victim but not from a scanner. In other words, scanners might send ANY DNS queries to the Internet but with low-speed rate to avoid receiving the amplified flood of replies.

3. Proposed approach

This section presents and elaborates on our proposed approach that aims at generating darknet flows and inferring DNS-based DDoS activities by leveraging darknet data. The approach exploits the idea of analyzing DNS queries that target the darknet space that were originally intended by the attacker to reach Internet open DNS resolvers [11]. Please note that our work leverages the dark space to infer and characterize amplification attacks. Intuitively, such an approach will not be able to pinpoint attacks that do not target such space; this limitation, however, is a generic drawback with any work that employs darknet to infer malicious activity [12]. In this case, our approach could be used in conjunction with other approaches that infer amplification attacks using operational non-dark spaces to provide a more comprehensive view of such attacks. Indeed, the approach takes as input darknet traffic and outputs inferred DNS amplification DDoS insights. It is based on several components, namely, the flows generation, the detection, the rate classification and the clustering components. We discuss these components in what follows.

3.1. Flow generation

The flow generation component takes an input darknet traffic to produce flows of traffic on a daily basis. A flow is defined as a series of consecutive packets sharing the same source IP address targeting darknet addresses. In order to generate such flow, (1) we collect network traces that consist of a unique source and destination IP pair, and (2) merge all flows that belong to the same source IP.

3.2. Detection component

The detection component takes as input darknet traffic and outputs DNS-based DDoS flows. To achieve the detection task, we base our detection component on analyzing DNS queries targeting darknet addresses. These DNS queries are attempts towards port 53. In order to detect DNS amplification DDoS, we built our
approach in accordance with the parameters of Table 1. We describe below each of those parameters next.

- **Packet count**: The packet count parameter defines the minimum number of packets sent per one source to our/13 darknet space. This parameter is useful to extract DDoS attacks with high impact in addition to providing an estimate of its scale. For instance, a flow that possesses thousands of packets sent to the darknet space is larger and more effective than a flow with 50 packets. In order to estimate a suitable packet count parameter for the attack flows, we execute an experiment as shown in Fig. 3. The experiment is based on
inferred darknet DDoS attacks and the investigation of their corresponding number of packets. For such attack flows, we fix the number of packets as perceived by the telescope and compute the number of attack flows that have at least such a number of packets. It is evident that below 21 packets, the attack flows will dramatically increase, while above that number, such flows will not decrease sharply. Thus, in this work, we decided to chose 21 packets as the packet count parameter for a DDoS attack flow. We assert that this threshold is a conservative number between false positives and false negatives. It is very significant to note that in [12], the authors also perform such experiment to extract DDoS attack flows; they found that 25 packets is suitable in their case which was in 2006. We postulate that the slight decrease in packet threshold that we found is due the recent rise of stealthy attacks that employ lower number of packets per unit of time to achieve their attack while attempting to avoid detection.

- **Targeted IPs**: Inspecting the number of targeted IPs verifies that the packets sent are not targeting only one IP address but distinct ones. Moreover, this permits the filtering of misconfiguration traffic (i.e., a host sending packets to only 1 unused IP address) and identifies the scanning mechanism for open DNS resolvers. To approximate a threshold for the number of targeted IPs, we semi-automatically (i.e., using a script and manual analysis and observation) investigated 1000 random DDoS attacks that were inferred by analyzing the darknet space using the open source network intrusion detection system Snort. The average of all those attacks were shown to target at least 29 different IPs. Thus in this work, we assert that the inferred DDoS attempts involve at least 29 distinct open DNS resolvers; this is based on the realistic assumption that an attempt of contacting at least 29 unused IP address out of half a million darknet IP addresses in order to amplify an attack has a similar intention to contacting at least 29 distinct open resolvers on the Internet space. Please note that imposed by the latter, and in practice, one should adopt the minimum packet count to be at least 29 packets. Note that, we could have also added other parameters such as attack-duration and packet-rate to our detection component. However, we avoid using time-based constraints; we have detected some flash attempts [14] that targeted thousands of distinct unused IPs within seconds and other stealthy scanning activities [15] that persisted for several weeks.

In a nutshell, our detection component labels a flow of traffic as a DDoS amplification DDoS attack if it has sent at least 21 DNS query of type ANY to at least 29 distinct unused dark IP addresses. Further, the flow must have requested domains that exist in root and TLD database.

### 3.3. Rate classification component

The rate of the attack is one of the major characteristics of DDoS activities [12]. After inferring DNS amplification flows, we noticed the existence of a large deviation among DNS amplification DDoS attack rates. For example, some flow rates reached more than 50 thousand packets per second (pps) whereas others were below 1 pps. Therefore, in order to understand more this large deviation and to group attacks per attack rates, we executed a rate classification exercise based on the values found in Table 2. Please note that in order to compute the rate as well as the other parameters of Table 1, we employ a time-out metric, which is the case when a source in a particular flow ceases to send packets towards the network telescope.

Going back to the rate classification procedure, the three attack rate categories are explained as follows:

- **Low attack rate**: To differentiate between low and medium attacks, we have executed an experiment with a number of confirmed attack flows as depicted in Fig. 4. We also follow a conservative approach by choosing 0.5 pps as the threshold. Please note that the latter is only used to cluster the attacks per rate and thus is not employed in the detection component that was discussed in the previous section.

- **High attack rate**: This category contains high rate attempts that are commonly referred to as flash attacks [14]. We have chosen a threshold of 4700 pps, which is the average rate of the Slammer worm propagation [14], to differentiate between medium and high rate attacks. In this exercise, we

### Table 1

**DNS amplified DDoS identification parameters.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet count</td>
<td>&gt;21 (experimental)</td>
</tr>
<tr>
<td></td>
<td>&gt;29 (practical)</td>
</tr>
<tr>
<td>Targeted IPs</td>
<td>&gt;29</td>
</tr>
<tr>
<td>DNS query type</td>
<td>ANY</td>
</tr>
<tr>
<td>Requested domain</td>
<td>Found in Root_DNS_DB</td>
</tr>
</tbody>
</table>

### Table 2

**Classification per attack rate.**

<table>
<thead>
<tr>
<th>Attack rate category</th>
<th>Value (pps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Rate ≤ 0.5</td>
</tr>
<tr>
<td>Medium</td>
<td>0.5 &lt; rate ≤ 4700</td>
</tr>
<tr>
<td>High</td>
<td>Rate &gt; 4700</td>
</tr>
</tbody>
</table>
assume that the average rate of the fastest worm propagation in 2003 will have, at least, similar rates as flash attacks in 2014. Please note that in general, on one hand, worm propagation performs scans for vulnerabilities on hosts in an attempt to exploit or infect the victims. On the other hand, in relation to DNS amplification DDoS attempts, the attackers generate, in only one step, similar attempts to infer open DNS resolvers and execute the amplification attack. Recall, that the latter technique does not aim at searching for a vulnerability to exploit, but instead sends benign DNS ANY queries to abuse open DNS resolvers in order to amplify the replies on the victims.

- **Medium** attack rate: Intuitively, this class captures those attacks that are in between the low and high rate categories.

### 3.4. Clustering component

In an attempt to uncover and cluster similar DNS amplification DDoS traces that might be executed by similar authors/code/botnet/campaign, we resort to data mining clustering approaches. This exercise can aid in detecting patterns, trends and links among attack traces. To achieve this task, we have selected and extracted a number of attributes as shown in Table 3. This allowed us to filter out those attributes that were not applicable or has no or low information gain.

In order to perform the clustering, we have leveraged two algorithms, namely, the Expectation Maximization (EM) [17] and the **k**-means [18].

**The EM algorithm:** This popular iterative refinement algorithm is a standard procedure for maximum likelihood estimation. This procedure has two stages; the first, which is the expectation step, is used to mine the association between current estimates of the parameters and the latent variables by calculating subsequent probabilities. The second step, which is the maximization step, is employed to update the parameters based on an expected complete data log-likelihood [19].

**The k-means algorithm:** One of the most well-known and commonly used clustering technique is the k-means. First, the algorithm randomly selects k of the objects (i.e., values of extracted attributes), each of which initially represents a cluster mean or center. As for the remaining objects, based on the cluster mean, they are allocated to the closest cluster. Consequently, the algorithm calculates the new mean for every cluster. This process continues through other iterations until the criterion function converges.

We have chosen the above mentioned algorithms for several reasons. In addition to being well-known in tackling the data clustering problem, the k-means algorithm has been successfully used to detect anomalies [20] and DDoS [21]. On the other side, the expectation maximization, which extends the k-means paradigm using a probabilistic approach, has also been leveraged in clustering attacks [22,23] and has been shown to yield promising results. For more information regarding the inner workings of the aforementioned clustering algorithms, we kindly refer the reader to [24].

### 4. Empirical evaluation

The evaluation is based on a real darknet dataset during a 6 months period between January and June, 2013. In general, we possess real darknet data that we receive from a trusted party.

The darknet traps monitor/13 address blocks (i.e., ≈ half a million dark IPs). The analyzed data consists of an average of 1.44 TB of

#### Table 3

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ip.flag</td>
<td>IP Flags</td>
</tr>
<tr>
<td>ip.flag.df</td>
<td>Do not fragment</td>
</tr>
<tr>
<td>ip.len</td>
<td>Total IP length</td>
</tr>
<tr>
<td>ip.ttl</td>
<td>Time to live</td>
</tr>
<tr>
<td>udp.len</td>
<td>UDP length</td>
</tr>
<tr>
<td>dns.count.add.rr</td>
<td>DNS additional RRs</td>
</tr>
<tr>
<td>dns.qry.name</td>
<td>DNS query name</td>
</tr>
<tr>
<td>flow.avg.pkt.size</td>
<td>Average packet size</td>
</tr>
<tr>
<td>flow.attack.duration</td>
<td>Attack duration</td>
</tr>
<tr>
<td>high.asn.numb</td>
<td>Autonomous system #</td>
</tr>
</tbody>
</table>

![Fig. 4. Rate threshold.](image-url)
one-way communications to unused IPs. Note that the proposed DNS amplification inference approach is capable of processing and inferring attacks in around 90 s per 20 GB of darknet traffic. The latter advocates that the proposed approach is practically viable in operational environments. In regards to our data mining exercises, our analysis is based on Weka [25], which is a data mining tool implemented in Java. We abide and closely follow the steps of our proposed approach that was discussed in Section 3 to elaborate on our analysis, which is based on three main elements, namely, the characterization, the insights generation and a case study. In total, our approach identified a total of 134 DNS amplification DDoS attacks including high-speed, medium and stealthy attacks (please refer to the Appendix A).

4.1. DNS amplification DDoS characterization

In this section, we present the overall DNS amplification DDoS statistics related to our analyzed dataset. The semiannual DNS queries distribution is shown in Fig. 5. The outcome clearly demonstrates the effectiveness of the proposed detection approach by fingerprinting large-scale amplified DDoS attacks including the famous reported event, which occurred in March 2013 [26]. On the other hand, in order to have a closer look at the latter attack, we depict Fig. 6 that illustrates the distribution of the queries for the month of March. Please note that the other peaks which resemble various unreported amplified attacks as shown in Fig. 5, will be analyzed and elaborated in future work. The average DNS queries arrival time per hour is approximately 58,050 packets. Obviously, several large-scale DNS amplified DDoS attacks caused some peaks at some periods such as at hours 340, 400 and 517 in which the distribution of packets was raised to 503,995, 686,774 and 798,192 packets, respectively. More explanation on these peaks are discussed in Section 4.3.

4.1.1. Query type distribution

In order to understand the types of DNS queries received on the monitored dark space, we list in Table 4 the DNS query type distribution of the analyzed dataset. As expected, the vast majority of these are ANY queries. Note that the top 4 records are the same for the entire 6 months period. Further, in contrast with the results in 2007 by [27] that found that ANY records scored only 0.0199% of the entire perceived records, we record 59.64% as observed on the darknet space. As a result, we can arguably assume that the recent trend of DNS amplification attacks are behind the increase of ANY records found on the darknet in the current year [26].

4.1.2. Top countries

Figs. 7 and 8 respectively show the top 5 source countries of DNS amplification DDoS attacks and their corresponding generated traffic. Note that in what follows, we focus our analysis during the three months of February, March and April, 2013.

Netherlands was ranked first in terms of both traffic sent and attack counts. Our results cross validate with the investigation in [28] and the news in [29]. Since Netherlands was mainly involved in the attack, it is normal to see victims and even scanners located in Netherlands. The United States was also found in our result as one of the top most involved countries. For Canada, notice the low number of attacks but the large amount of generated traffic. The reason behind this difference is that, although few of the Canadian IPs were found involved, yet they generated

![Fig. 5. DNS queries distribution – Semiannual 2013 data.](image)

![Fig. 6. DNS queries distribution – March 2013 data.](image)
huge amount of traffic. This corroborates the fact that DNS amplified attacks are very powerful since they allow attackers to create an immense amount of traffic (i.e., the amplification factor) with very little effort (i.e., very few number of leveraged bots). After manual inspection, some of these Canadian IPs were found involved in the largest DDoS attack [6]. More on this is discussed in Section 4.3.

4.1.3. Requested domains

Last but not least, we illustrate the top requested DNS domains as shown in Fig. 9. We anonymize TLDs for sensitivity issues. Fig. 9 shows that Root is the most requested domain name as perceived by the monitored darknet. Recall that attackers will typically submit a request for as much zone information as possible to maximize the amplification effect. Hence, the use of Root as the requested domain name. Note that, from our data, the second top requested domain (labeled as A) is a TLD that belongs to one of the largest Internet-scale DNS operators.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>January Packet_Count (%)</th>
<th>February Packet_Count (%)</th>
<th>March Packet_Count (%)</th>
<th>April Packet_Count (%)</th>
<th>May Packet_Count (%)</th>
<th>June Packet_Count (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9,717,559 A</td>
<td>(48.91%)</td>
<td>10,047,038 A</td>
<td>(49.02%)</td>
<td>27,649,274 ANY</td>
<td>(64.23%)</td>
<td>18,378,685 ANY</td>
</tr>
<tr>
<td>6,738,709 ANY</td>
<td>(33.91%)</td>
<td>7,763,817 ANY</td>
<td>(37.88%)</td>
<td>11,310,058 A</td>
<td>(26.28%)</td>
<td>11,595,908 A</td>
</tr>
<tr>
<td>3,323,598 TXT</td>
<td>(16.22%)</td>
<td>2,479,577 TXT</td>
<td>(12.10%)</td>
<td>2,459,257 TXT</td>
<td>(5.71%)</td>
<td>3,402,073 TXT</td>
</tr>
<tr>
<td>50,473 MX</td>
<td>(0.25%)</td>
<td>100,463 MX</td>
<td>(0.49%)</td>
<td>500,143 MX</td>
<td>(1.16%)</td>
<td>180,779 MX</td>
</tr>
<tr>
<td>36,438 PTR</td>
<td>(0.18%)</td>
<td>29,232 PTR</td>
<td>(0.14%)</td>
<td>63,340 RRSIG</td>
<td>(0.15%)</td>
<td>28,716 AAAA</td>
</tr>
</tbody>
</table>

Fig. 7. Top 5 source countries (attacks).

Fig. 8. Top 5 source countries (generated traffic).

Fig. 9. Top requested domains.
The map was generated using Gephi\(^\text{\textregistered}\), an open source visualization tool.

Different locations could be the result of one campaign using different ASNs from different other ASNs with some specific attributes. For instance, in regards to the clusters during the testing mode, based on the majority value of the class attribute and generates the clustering. Then it assigns classes based on the training set of the data. Cross validation is a technique used to evaluate machine learning models. It involves partitioning the dataset into training and testing sets. In this case, the dataset is split into 60% training data and 40% for testing as it achieved the minimum classification error. Based on this technique, we have achieved a 82% accuracy. In other words, our model incorrectly classified 18% of the traces to their corresponding clusters. We aim, in our future work, to analyze more data and run more complex algorithm to improve our clustering result.

Please note, that although we do not have a decisive proof of whether each cluster represent a campaign or a botnet of DNS amplification DDoS, we relatively succeeded in this task by pinpointing similarities of features among the DNS amplification DDoS traces.

4.2. Clustering Insights

This section highlights our clustering results. Recall that the aim is to cluster similar DNS amplification DDoS traces that might be executed by similar authors/code/botnet/campaign.

Since we had no prior knowledge on the number of clusters, we first run the EM algorithm to only infer the number of clusters by cross validation. We executed the algorithm in several cluster modes, using a training set and several percentage split tasks. We compared all the results and chose the model with the highest log likelihood for the best fit. After retrieving the number of clusters, we run the k-means with that number of clusters for further analysis. Again, we run several experiments (40%, 50%, 60%, 70% and 80% split) using the k-means algorithms and chose the model with 60% training data and 40% for testing as it achieved the minimum cluster sum of squared errors. Based on our testing data, Table 5 lists our summarized instances per clusters while Fig. 10 visualizes the final k-means output.

Next, we disclose the attributes that formed the clusters. Table 6 shows the cluster centroids of the k-means algorithm. This table is based on the training set of the data.

It is shown that our model clustered the traces based on 4 different ASNs with some specific attributes. For instance, in regards to cluster 0, all the DDoS attacks have source IPs within ASN-V and have the DF flag not set in the IP header. Moreover, the same flow must have an IP length of 56 bytes and a TTL value less than 60. In addition, the UDP length must be 36 bytes while the requested domain is root. Additionally, all the attacks that belong to cluster 0 should be launched within a 1 day period and possess an entire encapsulated DNS packets of an average size of 70 bytes. Through manual inspection, we found that the majority of IPs that fall within cluster 0 are originating from Netherlands which is coherent with the investigation in [28]. Similar concept applies for other clusters. Note the similarities between cluster 2, 3 and 4 which could be the result of one campaign using different ASNs from different locations.

After the clustering exercise, in order to evaluate our model, we run the cluster evaluation algorithm in Weka.\(^2\) First it ignores the class attribute and generates the clustering. Then it assigns classes to the clusters during the testing mode, based on the majority value of the class attribute within each cluster. Then it calculates the classification error. Based on this technique, we have achieved a 82% accuracy. In other words, our model incorrectly classified 18% of the traces to their corresponding clusters. We aim, in our future work, to analyze more data and run more complex algorithm to improve our clustering result.

Please note, that although we do not have a decisive proof of whether each cluster represent a campaign or a botnet of DNS amplification DDoS, we relatively succeeded in this task by pinpointing similarities of features among the DNS amplification DDoS traces.

4.2.1. Similarity insights

This exercise aims at inferring insights related to the used darknet address space. The aim is to provide a more core element to our clustering approach. The rationale behind this task states that since bots in the same campaign typically utilize the same list of IPs when launching their attacks, it would be interesting to capture the similarity of use related to these IP lists. By accomplishing this, we can possibly infer campaigns or at least detect similarities in attack mechanisms. To achieve the intended goal, we executed an experiment to represent attacks that exchange at least 90% of dark IPs. Fig. 11 depicts an IP map\(^1\) that satisfies the latter condition.

It is disclosed that two groups of IPs share at least 90% of dark IPs. Please refer to the tables in the appendix for attack references. The smaller group consists of 2 IPs from different months (March and April). Our analysis identified that these two sources share not just dark IP usage, but also country, ASN number, speed range, requested domain, and many other attributes as previously identified in Section 4.2 in cluster 0. As for the second group, 7 out of 8 originate from the same ASN number. All of the attacks in this group are initiated from Europe, specifically from Netherlands; this finding is corroborated in [28]. Similar to the first group, these attacks share similarities in clustering attributes and 55.56% of these traces are found also in cluster 0. One of the interesting point uncovered by analyzing this group is that all its members are sharing a specific address space range, possibly highlighting a DDoS campaign.

4.3. Case studies

We discuss below some major case studies that belong to three different attack rates.

The first case study represents high-speed (i.e., flash) DNS amplification DDoS detected attacks. In our dataset, we have found 3 attacks that fall within this category; ID F1, M1 and A1. These are shown in the first rows of Tables 7 and 8, respectively. These attacks are found to be focused; intensity is equal to the contacted unique dark IPs or, in other words, the host/attacker sends only 1 packet per open DNS resolver. First, attack F1 is the fastest detected attack. It was launched from the United States, California on February 19th. The detected attack has a rate of 79565.67 pps. This propagation speed is 17 times faster than the Slammer worm [14]. This attack targeted 6.5% of our darknet space in less than 1 s. Assuming the intent of the attacker is to send one packet for each IP, a malware with this speed can target the whole IPv4 Internet address space in less than a week (6 days and few hours). In order to validate the occurrence of this flash DNS amplified DDoS attack, we recourse to publicly accessible Dshiel[32] data and inspected port 53 for the 3 days before and after the 19th of February. We have noticed a significant increase at this specific date. According to Dshielf data, the average incident reports measured on port 53 was 14.28% for the surrounded 7 days of this attack. However, on February 19th, the average reached 38.19% with a 10,347,879

---

\(^2\) http://www.cs.ccsu.edu/markov/ccc_courses/DataMining-Ex3.html

\(^1\) The map was generated using Gephi\(^\text{\textregistered}\), an open source visualization tool.

---
increase in reports from the previous day. Second, attack M1 was launched from Taiwan on March 18th. This date is the same date of the largest DDoS attack as declared in [26]. This flash attack sent probes to 50,257 unique dark IP (9.5% of the our/13 darkspace) within 1 s with an average rate of 46677.36 pps. This speed is almost 10 times faster than the Slammer worm. With this speed, this DDoS can target 16 millions IPv4 hosts (/8) on the Internet in less than 6 min. Third, attack A1 was also launched from the United States, California on April 15th. The attack possesses a rate of 21672.18 pps. This attack targeted 11.7% of our darken address space.

The second case study, which involves medium speed attacks, is one of the major inferred DNS amplification DDoS in terms of size and impact. Compared to the previous case study, this attack is not focused (intensity is not equal to the contacted unique dark IP or sending at least 1 packet per open DNS resolver). This attack targeted one victim using 2 hosts (ID M5 and M10 of Table 8). This attack targeted around 360,000 unique dark IPs (68% of the monitored/13 darknet), and hence could be considered the most comprehensive compared to all other threats. Our analysis linked these traces to the largest DNS amplification DDoS [6] for the following reasons: (1) in addition to the use of the ANY DNS query, the traces of this attack targeted the “ripe.net” domain name; this domain was used in the largest DDoS as declared in a blog posted by the victim [26]; (2) the timing of the traces from the host with ID M10 started on March 15th, whereas those of the host with ID M5 started on March 17th. The two mentioned dates could be found in the media [33, 34] and were posted on Twitter on March 17th by a company support personnel [35]. In order to depict this distributed attack, in Fig. 6, we highlighted the threat using a colored dashed-line. The first and/or second peaks are likely performed as testing before actually executing the largest DDoS as demonstrated by the third peak. Our result matches the ascending order of peaks as discussed by the victims [26]. In order to predict or provide an approximation of the number of machines that were involved in the aforementioned largest DNS amplification attack, we assume the following: Consider M5 as a sample of victim (spoofed IP or compromised machine). The average attempts sent on the darknet is 14,464,427 packets over 360,705 open DNS resolver which is around 40 requests for each dark IP. Recall that each dark IP might be considered as an open DNS resolver. Also, assume that the amplification factor is 75 [26] and each request has a size of 68 byte. Moreover, assuming only 1% (3607) of the 360,705 reached successfully open DNS resolvers, then using a regular machine with a dedicated Internet service, only 1 host can generate amplified reply of 5.482 gigabits (Gb) through 3607 open DNS resolvers within 1 s. Therefore, to generate a 75 or 300Gb DNS amplified DDoS attack, only 14 or 55 synchronized machines (bots) are needed, respectively.

The above two mentioned case studies are probably executed by an attacker using spoofed IP address of the victims or using compromised machines (recall Fig. 2a and b); we unlikely consider these activities as scanning event that are using legitimate addresses (i.e., the intention is not to DDoS themselves but other targeted victims).

The third case study involves slow rate attacks such as hosts with ID M51 to M54 in Table 8. This analysis targets stealthy focused attempts; these attacks have low sending rate and are typically hard to detect using a firewall and/or a typical intrusion detection system [15]. From Table 8, all information regarding these 4 hosts appears very similar or the same. Therefore, they are mostly generated by the same author/code/campaign. Although we cannot claim the orchestration among these hosts, our data highlights some shared characteristics among such stealthy threats. Note that the requested domain names within these attacks is a top-notch organization that deals with securing online transactions. Another group of stealthy attempts that are of interest are IDs A48 and A51 that are shown in Table 9. The hosts behind these activities scan slowly with an unprecedented average packet rate. For instance, ID A48 remains online for almost 3 weeks. Future analysis on this group of stealthy attempts might pinpoint to certain suspicious unknown activities. Unfortunately, it is very hard to validate our stealthy scanning activities with other security repositories or media as their impact is in the information gain rather than the maliciousness of their acts. In contrast to the previous two case studies, the attackers in such stealthy scenarios can use their legitimate addresses. The reason behind this assumption is that it is almost impossible to execute a powerful DNS amplified DDoS attack through a low-speed propagation. However, in these attacks, we reason that attackers will attempt to locate open DNS resolvers and/or build a DNS hierarchy table retrieved from the ANY replies before executing their attacks.

In addition to performing several validation of our results through DShield and the media, we execute a renowned Network

---

**Table 6**

$k$-means training cluster centroids.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Cluster 0 (49)</th>
<th>Cluster 1 (8)</th>
<th>Cluster 2 (14)</th>
<th>Cluster 3 (5)</th>
<th>Cluster 4 (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>high.asn numb</td>
<td>ASN-V</td>
<td>ASN-W</td>
<td>ASN-X</td>
<td>ASN-Y</td>
<td>ASN-Y</td>
</tr>
<tr>
<td>ip.flags</td>
<td>0x02</td>
<td>0x00</td>
<td>0x02</td>
<td>0x00</td>
<td>0x02</td>
</tr>
<tr>
<td>ip.flags.df</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ip.len</td>
<td>56</td>
<td>45</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>ip.ttl</td>
<td>&lt;60</td>
<td>&lt;60</td>
<td>&lt;60</td>
<td>&gt;100</td>
<td>&lt;60</td>
</tr>
<tr>
<td>udp.length</td>
<td>36</td>
<td>34</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>dns.qry.name</td>
<td>Root</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>flow.avg.pkt.size</td>
<td>&lt;1 day</td>
<td>&lt;1 day</td>
<td>&lt;1 day</td>
<td>&lt;1 day</td>
<td>btw-day-1 week</td>
</tr>
</tbody>
</table>

---

**Fig. 11.** IPs sharing at least 90% darken space.
Intrusion and Detection System (NIDS) (i.e., Snort [37]) on the whole traces to see if we can detect such malicious activities. The NIDS labeled 129 out of the inferred 134 (96%) threats as executing filtered portswipe probes. We have found that the 5 undetected attacks refer to the third case study (i.e., slow rate attacks, namely, IDs M51 to M54 and A51) that was previously discussed. After manual inspection, the M51 and A51 attacks turned out to be originating from the same source who is executing stealthy scans but in different time periods. Moreover, all these attacks are requesting one organization’s domain. In summary, we can claim that our approach that aims at inferring DNS amplified attacks yielded zero false negative in comparison with a leading IDS. Snort [37].

Future work could consider the latter task. For example, the approximate number of infections. In other work, Paxson, Dagon and Anagnostopoulos [7] introduced a new technique to execute DNS amplification attacks through DNSSEC-powered servers. The attack can reach up to 44 amplification factors in an undetectable manner. In comparison to our work, Oberheide et al. [27] have not linked or investigated any DNS DDoS activities such as the approximate number of infections. However, their work is different from this category as their methodology is only based on reply packets and do not include request packets such as DNS traces. Hence, DNS amplified activities cannot be inferred using their approach.

Second, in the area of DNS traffic analysis, the most related work is rendered by Oberheide et al. [27] who analyze DNS queries that target darknet sensors. The authors characterize these traces and propose a mechanism to implement a secure DNS service on darknet sensors. Moreover, Paxson [56] was among the first to pinpoint the threats of DNS reflectors on making DDoS attacks harder to defend. In another work, Dagon et al. [13] analyze corrupted DNS resolution paths and pinpoint an increase in malware that modified these paths and threatened DNS authorities. Further, Anagnostopoulos et al. [7] introduced a new technique to execute DNS amplification attacks through DNSSEC-powered servers. The attack can reach up to 44 amplification factors in an undetectable manner. In comparison to our work, Oberheide et al. have not linked or investigated any DNS DDoS traces through their analysis but solely focused on analyzing DNS traffic. On the other hand, Paxson, Dagon and Anagnostopoulos did not investigate darknet data in the context of DNS amplification attack inference and characterization. Therefore, DNS amplification traces destined to unused IP addresses (darknet) cannot be detected through their analysis. However, darknet and other sources of data (i.e., passive DNS) could be correlated to extract further intelligence on DNS amplification DDoS activities such as the approximate number of infections. Future work could consider the latter task.

5. Related work

Cyber security experts and researchers employ darknet analysis for several purposes, namely, monitoring and inferring of large-scale Internet events, including, DDoS [38], probing activities [39,40], worm propagation [41], analyzing events [42], measuring misconfiguration [43] and implementing monitoring sensors [44]. Since our work deals with cyber threats in general and DNS amplification DDoS in particular, we subsequently pinpoint the major related work in the areas of backscattered traffic analysis and DNS traffic investigation.

First, the use of darknet to infer DDoS activities owes much to the pioneer work carried out by Moore et al. in [38] that was revisited in [12]. The key observation behind the authors’ technique is that attackers, before executing a DDoS attack, spoof their addresses using random IPs. Hence, once the attack is executed, all the victims’ replies (i.e., backscattered packets) are bounced back to the fake IP addresses, which could be in the monitored darknet space. Their work is operated by CAIDA [45], which provide backscattered data for researchers. Numerous research work has been performed on such data to analyze DDoS activities. The majority focus on implementing new detection techniques to infer DDoS attacks [46–49], tracing-back the sources of attacks [50,51], investigating spoofed attacks [52] and visualizing attacks [53–55]. Our work is different from this category as their methodology is only based on reply packets and do not include request packets such as DNS queries. Hence, DNS amplified activities cannot be inferred using their approach.

<table>
<thead>
<tr>
<th>Victim/scanner ID</th>
<th>Requested domain name</th>
<th>Detection period</th>
<th>Analyzed attack duration (s)</th>
<th>Intensity (packet)</th>
<th>Contacted unique dark IPs</th>
<th>Avg. packet size (bytes)</th>
<th>Avg. rate (pps)</th>
<th>Rate category</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>A</td>
<td>February 19</td>
<td>0</td>
<td>34,410</td>
<td>34,410</td>
<td>78</td>
<td>79565.67</td>
<td>High</td>
</tr>
<tr>
<td>F2</td>
<td>G</td>
<td>February 14</td>
<td>4477</td>
<td>129,206</td>
<td>129,206</td>
<td>85</td>
<td>28.86</td>
<td>Medium</td>
</tr>
<tr>
<td>F3</td>
<td>A</td>
<td>February 21</td>
<td>29,174</td>
<td>690,219</td>
<td>305,344</td>
<td>78</td>
<td>23.66</td>
<td>Medium</td>
</tr>
<tr>
<td>F4</td>
<td>Root</td>
<td>February 26</td>
<td>17,084</td>
<td>351,617</td>
<td>351,617</td>
<td>70</td>
<td>20.58</td>
<td>Medium</td>
</tr>
<tr>
<td>F5</td>
<td>Root</td>
<td>February 19</td>
<td>16,245</td>
<td>259,590</td>
<td>259,590</td>
<td>70</td>
<td>17.89</td>
<td>Medium</td>
</tr>
<tr>
<td>F6</td>
<td>Root</td>
<td>February 26</td>
<td>9389</td>
<td>162,513</td>
<td>162,513</td>
<td>70</td>
<td>17.31</td>
<td>Medium</td>
</tr>
<tr>
<td>F7</td>
<td>Root</td>
<td>February 11–12</td>
<td>25,052</td>
<td>349,692</td>
<td>349,692</td>
<td>70</td>
<td>13.96</td>
<td>Medium</td>
</tr>
<tr>
<td>F8</td>
<td>Root</td>
<td>February 20</td>
<td>15,215</td>
<td>187,886</td>
<td>187,886</td>
<td>70</td>
<td>12.35</td>
<td>Medium</td>
</tr>
<tr>
<td>F9</td>
<td>Root</td>
<td>February 13</td>
<td>61,591</td>
<td>660,473</td>
<td>356,162</td>
<td>70</td>
<td>10.72</td>
<td>Medium</td>
</tr>
<tr>
<td>F10</td>
<td>Root</td>
<td>February 16–17</td>
<td>33,602</td>
<td>355,188</td>
<td>355,188</td>
<td>70</td>
<td>10.57</td>
<td>Medium</td>
</tr>
<tr>
<td>F11</td>
<td>Root</td>
<td>February 3</td>
<td>6625</td>
<td>64,726</td>
<td>64,726</td>
<td>70</td>
<td>9.8</td>
<td>Medium</td>
</tr>
<tr>
<td>F12</td>
<td>Root</td>
<td>February 23</td>
<td>11,412</td>
<td>96,216</td>
<td>96,216</td>
<td>70</td>
<td>8.4</td>
<td>Medium</td>
</tr>
<tr>
<td>F13</td>
<td>Root</td>
<td>February 2–3</td>
<td>93,268</td>
<td>633,886</td>
<td>357,497</td>
<td>70</td>
<td>6.8</td>
<td>Medium</td>
</tr>
<tr>
<td>F14</td>
<td>A</td>
<td>February 3</td>
<td>19,872</td>
<td>128,297</td>
<td>128,297</td>
<td>78</td>
<td>6.46</td>
<td>Medium</td>
</tr>
<tr>
<td>F15</td>
<td>Root</td>
<td>February 7</td>
<td>2107</td>
<td>12,965</td>
<td>12,965</td>
<td>70</td>
<td>6.15</td>
<td>Medium</td>
</tr>
<tr>
<td>F16</td>
<td>Root</td>
<td>February 23–27</td>
<td>401,266</td>
<td>804,348</td>
<td>359,686</td>
<td>70</td>
<td>2</td>
<td>Medium</td>
</tr>
<tr>
<td>F17</td>
<td>Root</td>
<td>February 11–15</td>
<td>311,301</td>
<td>316,425</td>
<td>316,425</td>
<td>70</td>
<td>1.02</td>
<td>Medium</td>
</tr>
<tr>
<td>F18</td>
<td>Root</td>
<td>February 4–19</td>
<td>1,322,119</td>
<td>869,395</td>
<td>360,666</td>
<td>70</td>
<td>0.66</td>
<td>Medium</td>
</tr>
<tr>
<td>F19</td>
<td>Root</td>
<td>February 4–14</td>
<td>853,983</td>
<td>540,412</td>
<td>356,117</td>
<td>70</td>
<td>0.63</td>
<td>Medium</td>
</tr>
<tr>
<td>F20</td>
<td>A</td>
<td>February 3</td>
<td>10,634</td>
<td>6632</td>
<td>6632</td>
<td>78</td>
<td>0.62</td>
<td>Medium</td>
</tr>
<tr>
<td>F21</td>
<td>A</td>
<td>February 3–16</td>
<td>1,138,804</td>
<td>683,321</td>
<td>359,470</td>
<td>78</td>
<td>0.6</td>
<td>Medium</td>
</tr>
<tr>
<td>F22</td>
<td>Root</td>
<td>February 20–28</td>
<td>766,810</td>
<td>378,289</td>
<td>319,668</td>
<td>70</td>
<td>0.49</td>
<td>Low</td>
</tr>
<tr>
<td>F23</td>
<td>Root</td>
<td>February 5</td>
<td>27,832</td>
<td>9645</td>
<td>8123</td>
<td>70</td>
<td>0.35</td>
<td>Low</td>
</tr>
<tr>
<td>F24</td>
<td>A</td>
<td>February 19</td>
<td>50,374</td>
<td>16,393</td>
<td>16,393</td>
<td>78</td>
<td>0.33</td>
<td>Low</td>
</tr>
<tr>
<td>F25</td>
<td>A</td>
<td>February 4</td>
<td>16,353</td>
<td>5306</td>
<td>5306</td>
<td>78</td>
<td>0.32</td>
<td>Low</td>
</tr>
<tr>
<td>F26</td>
<td>Root</td>
<td>February 6–26</td>
<td>1,706,728</td>
<td>191,562</td>
<td>191,329</td>
<td>70</td>
<td>0.11</td>
<td>Low</td>
</tr>
<tr>
<td>F27</td>
<td>Root</td>
<td>February 15–26</td>
<td>970,150</td>
<td>19,636</td>
<td>19,636</td>
<td>70</td>
<td>0.02</td>
<td>Low</td>
</tr>
<tr>
<td>F28</td>
<td>A</td>
<td>February 9–28</td>
<td>1,691,139</td>
<td>16,845</td>
<td>16,845</td>
<td>78</td>
<td>0.01</td>
<td>Low</td>
</tr>
<tr>
<td>F29</td>
<td>A</td>
<td>February 15–22</td>
<td>640,165</td>
<td>966</td>
<td>966</td>
<td>78</td>
<td>0.0</td>
<td>Low</td>
</tr>
</tbody>
</table>
6. Conclusion

This work presented a new approach to infer Internet DNS Amplification Denial of Service activities by leveraging the darknet space. The approach corroborated the fact that one can infer DDoS attacks without relying on backscattered analysis. The detection module is based on certain parameters to fingerprint network flows as DNS amplification DDoS related. The classification module amalgamates the attacks based on their possessed rate while the clustering component attempts to identify flows that share similarity features in an attempt to disclose campaigns of DNS Amplification DDoS. The analysis was based on 1.44 TB of real darknet traffic collected during several month period. The results disclose 134 DNS amplified DDoS activities, including flash and stealthy attacks. The clustering and similarity exercises provided insights and inferences that permit the detection of DNS amplification DDoS campaign activities. Moreover, we would like to investigate other protocols than DNS that could also be vulnerable to amplification attacks such as NTP, SSDP, SNMP, NTP [57] and implement our proposed approach in a near real-time fashion.

Table 8
Summary of the Analyzed DNS Amplification DDoS Traces (March 2013).

<table>
<thead>
<tr>
<th>Victim/scanner ID</th>
<th>Requested domain name</th>
<th>Detection period</th>
<th>Analyzed attack duration (s)</th>
<th>Intensity (packet)</th>
<th>Contacted unique dark IPs</th>
<th>Avg. packet size (bytes)</th>
<th>Avg. rate (pps)</th>
<th>Rate Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>A</td>
<td>March 18</td>
<td>1</td>
<td>50,257</td>
<td>50,257</td>
<td>78.00</td>
<td>46677.36</td>
<td>High</td>
</tr>
<tr>
<td>M2</td>
<td>A</td>
<td>March 31</td>
<td>26</td>
<td>63,543</td>
<td>63,543</td>
<td>78.00</td>
<td>24198.83</td>
<td>Medium</td>
</tr>
<tr>
<td>M3</td>
<td>E &amp; F</td>
<td>March 22</td>
<td>620</td>
<td>798,192</td>
<td>65,025</td>
<td>73.00</td>
<td>1287.41</td>
<td>Medium</td>
</tr>
<tr>
<td>M4</td>
<td>A</td>
<td>March 10</td>
<td>39</td>
<td>91,042</td>
<td>91,042</td>
<td>67.00</td>
<td>226.21</td>
<td>Medium</td>
</tr>
<tr>
<td>M5</td>
<td>B</td>
<td>March 17–18</td>
<td>93,508</td>
<td>14,464,427</td>
<td>360,705</td>
<td>68.00</td>
<td>154.69</td>
<td>Medium</td>
</tr>
<tr>
<td>M6</td>
<td>Root</td>
<td>March 3</td>
<td>572</td>
<td>64,956</td>
<td>64,956</td>
<td>70.00</td>
<td>113.53</td>
<td>Medium</td>
</tr>
<tr>
<td>M7</td>
<td>Root</td>
<td>March 23</td>
<td>662</td>
<td>64,230</td>
<td>64,230</td>
<td>70.00</td>
<td>97.00</td>
<td>Medium</td>
</tr>
<tr>
<td>M8</td>
<td>Root</td>
<td>March 30</td>
<td>610</td>
<td>58,104</td>
<td>58,104</td>
<td>70.00</td>
<td>95.19</td>
<td>Medium</td>
</tr>
<tr>
<td>M9</td>
<td>Root</td>
<td>March 24</td>
<td>665</td>
<td>63,139</td>
<td>63,139</td>
<td>70.00</td>
<td>94.99</td>
<td>Medium</td>
</tr>
<tr>
<td>M10</td>
<td>B</td>
<td>March 15</td>
<td>34,605</td>
<td>3,176,785</td>
<td>360,683</td>
<td>68.00</td>
<td>91.80</td>
<td>Medium</td>
</tr>
<tr>
<td>M11</td>
<td>Root</td>
<td>March 1</td>
<td>769</td>
<td>63,342</td>
<td>63,342</td>
<td>70.00</td>
<td>82.33</td>
<td>Medium</td>
</tr>
<tr>
<td>M12</td>
<td>A</td>
<td>March 25</td>
<td>985</td>
<td>79,333</td>
<td>54,632</td>
<td>70.00</td>
<td>80.52</td>
<td>Medium</td>
</tr>
<tr>
<td>M13</td>
<td>Root</td>
<td>March 12</td>
<td>581</td>
<td>40,364</td>
<td>37,160</td>
<td>70.00</td>
<td>69.46</td>
<td>Medium</td>
</tr>
<tr>
<td>M14</td>
<td>Root</td>
<td>March 1–2</td>
<td>2685</td>
<td>161,847</td>
<td>154,905</td>
<td>70.00</td>
<td>60.28</td>
<td>Medium</td>
</tr>
<tr>
<td>M15</td>
<td>C</td>
<td>March 25</td>
<td>1</td>
<td>60</td>
<td>60</td>
<td>77.00</td>
<td>58.69</td>
<td>Medium</td>
</tr>
<tr>
<td>M16</td>
<td>A</td>
<td>March 9</td>
<td>8884</td>
<td>504,794</td>
<td>270,352</td>
<td>78.00</td>
<td>56.82</td>
<td>Medium</td>
</tr>
<tr>
<td>M17</td>
<td>A</td>
<td>March 30</td>
<td>1963</td>
<td>63,623</td>
<td>63,623</td>
<td>78.00</td>
<td>32.41</td>
<td>Medium</td>
</tr>
<tr>
<td>M18</td>
<td>Root</td>
<td>March 21</td>
<td>10,255</td>
<td>254,285</td>
<td>254,285</td>
<td>78.00</td>
<td>24.80</td>
<td>Medium</td>
</tr>
<tr>
<td>M19</td>
<td>Root</td>
<td>March 7</td>
<td>13,572</td>
<td>247,483</td>
<td>247,483</td>
<td>78.00</td>
<td>18.23</td>
<td>Medium</td>
</tr>
<tr>
<td>M20</td>
<td>Root</td>
<td>March 2</td>
<td>25,314</td>
<td>355,675</td>
<td>355,675</td>
<td>78.00</td>
<td>14.05</td>
<td>Medium</td>
</tr>
<tr>
<td>M21</td>
<td>Root</td>
<td>March 13</td>
<td>9796</td>
<td>63,342</td>
<td>63,342</td>
<td>78.00</td>
<td>8.23</td>
<td>Medium</td>
</tr>
<tr>
<td>M22</td>
<td>Root</td>
<td>March 30</td>
<td>581</td>
<td>39,644</td>
<td>37,160</td>
<td>70.00</td>
<td>69.46</td>
<td>Medium</td>
</tr>
<tr>
<td>M23</td>
<td>Root</td>
<td>March 24</td>
<td>665</td>
<td>63,139</td>
<td>63,139</td>
<td>70.00</td>
<td>94.99</td>
<td>Medium</td>
</tr>
<tr>
<td>M24</td>
<td>Root</td>
<td>March 25</td>
<td>1</td>
<td>60</td>
<td>60</td>
<td>77.00</td>
<td>58.69</td>
<td>Medium</td>
</tr>
<tr>
<td>M25</td>
<td>A</td>
<td>March 9</td>
<td>8884</td>
<td>504,794</td>
<td>270,352</td>
<td>78.00</td>
<td>56.82</td>
<td>Medium</td>
</tr>
<tr>
<td>M26</td>
<td>A</td>
<td>March 30</td>
<td>1963</td>
<td>63,623</td>
<td>63,623</td>
<td>78.00</td>
<td>32.41</td>
<td>Medium</td>
</tr>
<tr>
<td>M27</td>
<td>Root</td>
<td>March 21</td>
<td>10,255</td>
<td>254,285</td>
<td>254,285</td>
<td>78.00</td>
<td>24.80</td>
<td>Medium</td>
</tr>
<tr>
<td>M28</td>
<td>Root</td>
<td>March 7</td>
<td>13,572</td>
<td>247,483</td>
<td>247,483</td>
<td>78.00</td>
<td>18.23</td>
<td>Medium</td>
</tr>
<tr>
<td>M29</td>
<td>Root</td>
<td>March 24</td>
<td>25,314</td>
<td>355,675</td>
<td>355,675</td>
<td>78.00</td>
<td>14.05</td>
<td>Medium</td>
</tr>
<tr>
<td>M30</td>
<td>Root</td>
<td>March 25</td>
<td>1</td>
<td>60</td>
<td>60</td>
<td>77.00</td>
<td>58.69</td>
<td>Medium</td>
</tr>
</tbody>
</table>
6.1. Lessons learned

From this work, we can extract the following insights related to DNS amplification attacks: First, when compared to previous years, we have found that the DNS amplification attacks are behind the increase of DNS queries of type \textit{ANY} on the Internet. Second, we have pinpointed that the majority of the attacks target the root domain. Third, we have inferred that DNS amplified attack rates can range from very low to high speeds. High speeds attacks pinpoint victims of spoofed attacks and compromised machines whereas the very slow attacks reflects stealthy scans. Last but not least, we have unexpectedly uncover a UDP-based mechanism used by DNS amplification attackers to execute DNS amplification attacks in a highly rapid manner without collecting information about open DNS resolvers. In other words, we have inferred that unlike typical DDoS attempts that scan for vulnerable machines and then execute the attack, the largest DNS amplification analyzed was executed in only one step through a small number of machines; benign DNS queries are sent to the Internet with the intention to reach open DNS resolvers, which subsequently trigger an amplified reply to the victim.

Appendix A

The summary of the Analyzed DNS Amplification DDoS Traces of February, March and April 2013 is shown in Tables 7–9 respectively.

References