Computer Communications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

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

ARTICLE INFO

Article history:
Received 24 April 2014
Received in revised form 13 November 2014
Accepted 26 January 2015
Available online xxxx

Keywords:
DDoS
DRDoS
DNS
Darknet
Cyber threats

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.

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.

* Corresponding author at: Computer Security Laboratory, Concordia University, Canada. Tel.: +1 514 848 2424x3166.
E-mail address: c_fachkh@encs.concordia.ca (C. Fachkha).

http://dx.doi.org/10.1016/j.comcom.2015.01.016
0140-3664/

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 DRDoS 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 returns 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 botnets (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 DRDoS 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 DRDoS 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 DRDoS 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

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

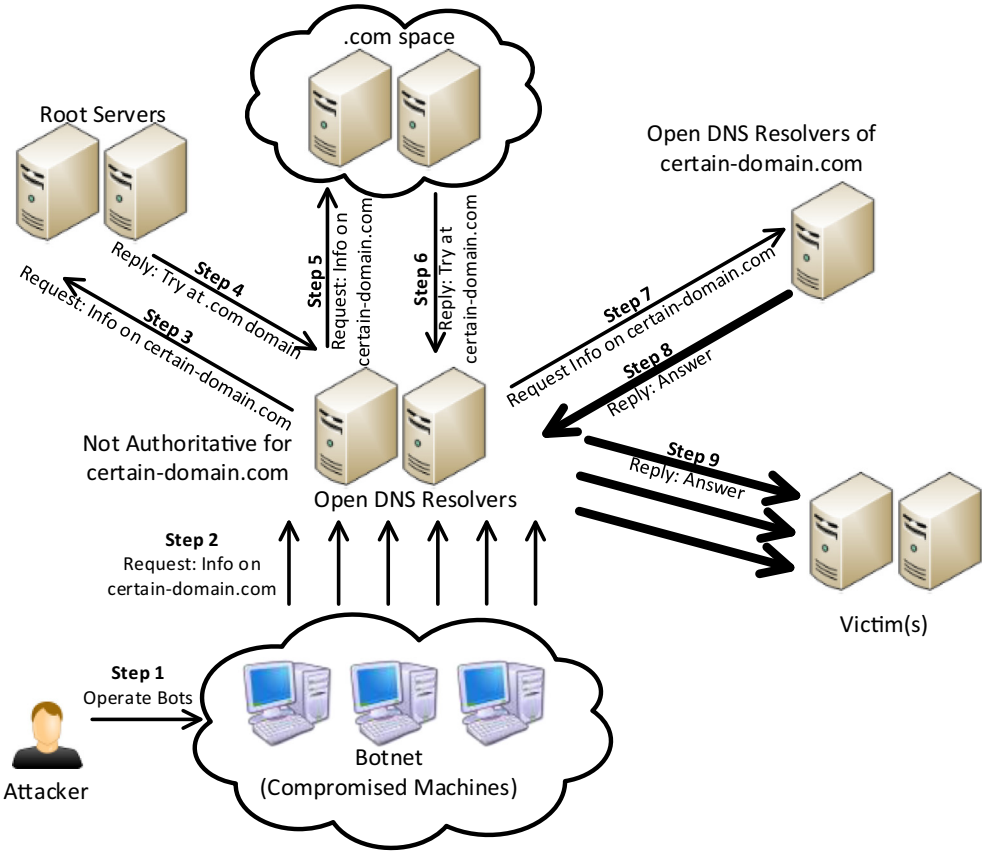
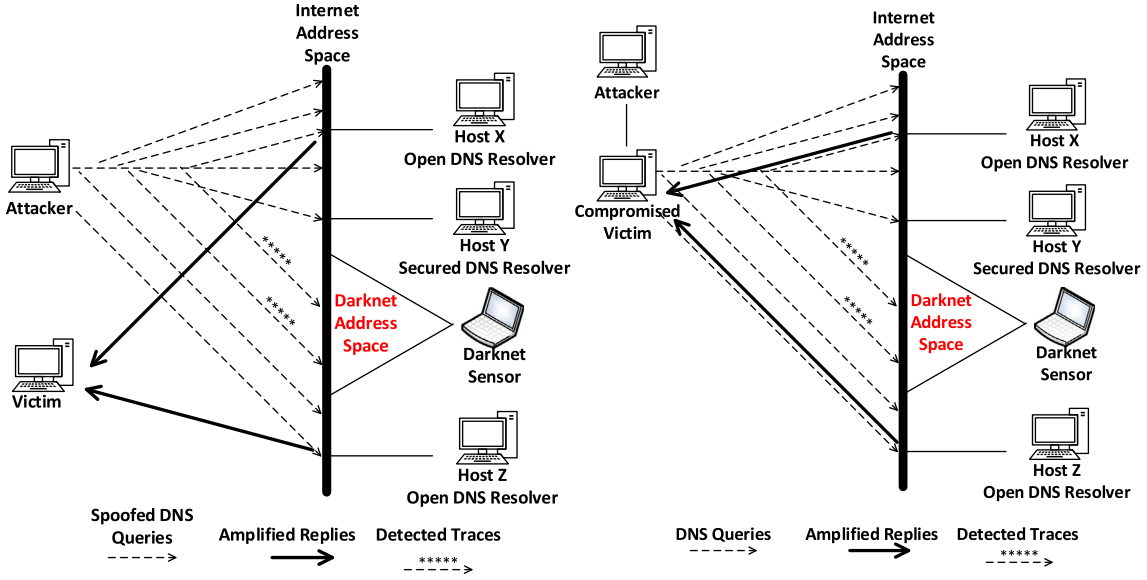


Fig. 1. DNS-based DRDoS scenario.



(a) IP Spoofing Scenario (b) Compromised Victim Scenario

Fig. 2. Two possible scenarios.

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

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

Table 1
DNS amplified DDoS identification parameters.

Parameter	Value
Packet count	>21 (experimental) >29 (practical)
Targeted IPs	>29
DNS query type	ANY
Requested domain	Found in Root_DNS_DB

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.

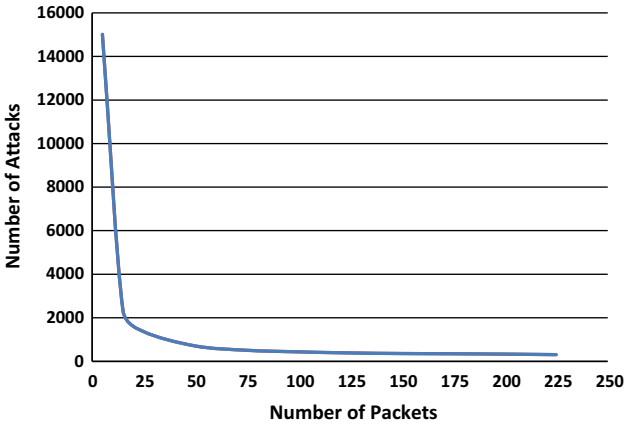


Fig. 3. Packet count parameter estimation.

- *DNS query type*: One of the major strengths of DNS DDoS attacks is rendered by their amplification factor. In the majority of DNS amplification DDoS attacks, DNS query type **ANY** is used [5]. This type of DNS query returns all known information about a DNS zone in a single request to the victim. This technique is an attempt to amplify the attack. In this work, we impose that all DNS amplification DDoS traces have **ANY** as the DNS query type.
- *Requested domain*: DNS amplification attempts are known to request root and Top Level Domain (TLD) name server operators [13]. We built a database containing a list of all known root and TLD domains. In general, these domains contain several DNS records. Therefore, DNS **ANY** queries targeting these servers trigger a large (amplified) reply. In this work, we corroborate that all DNS amplification DDoS activities request domains from the assembled database.

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 DNS 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 2
Classification per attack rate.

Attack rate category	Value (pps)
Low	Rate $\leqslant$ 0.5
Medium	0.5 < rate < 4700
High	Rate $\geqslant$ 4700

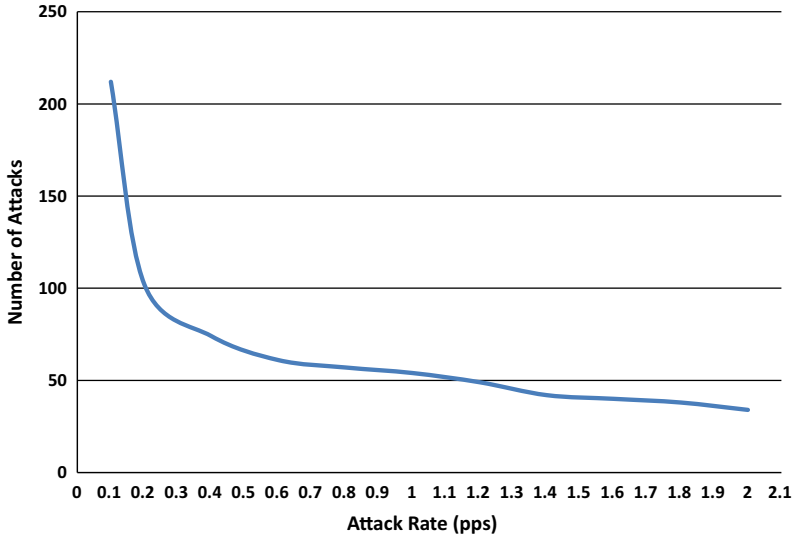


Fig. 4. Rate threshold.

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 and described in Table 3.

Indeed, we have initially analyzed hundreds of attributes from different network layers (i.e., IP/UDP/DNS) in addition to numerous flow-based features (i.e., attack duration, average packet size, etc.). However, we have leveraged a ranker [16] to evaluate the information gain of all the attributes and have chosen the top 10 as shown in Table 3. This allowed us to filter out those attributes that were not applicable or has no or low information gain.

Table 3
Chosen clustering attributes.

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

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.[1] 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

[1] https://www.farsightsecurity.com/.

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

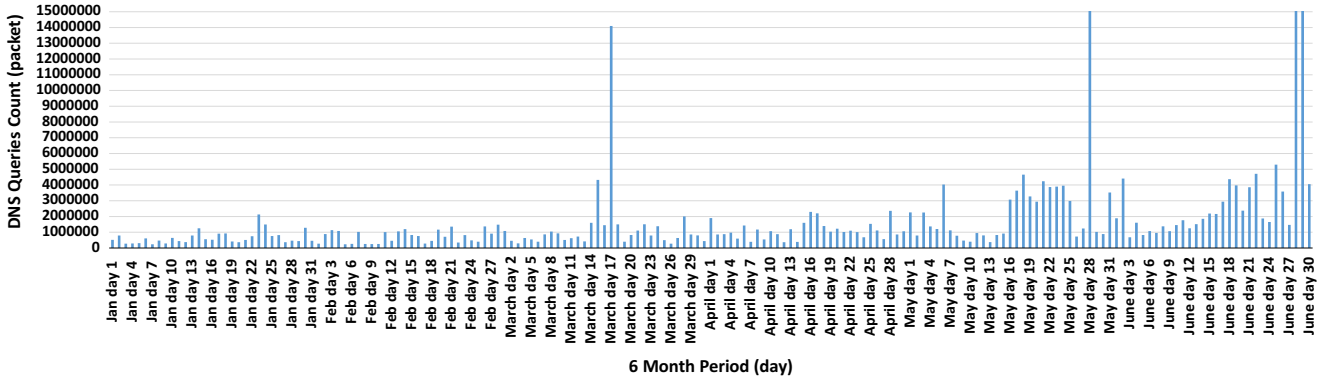


Fig. 5. DNS queries distribution – Semiannual 2013 data.

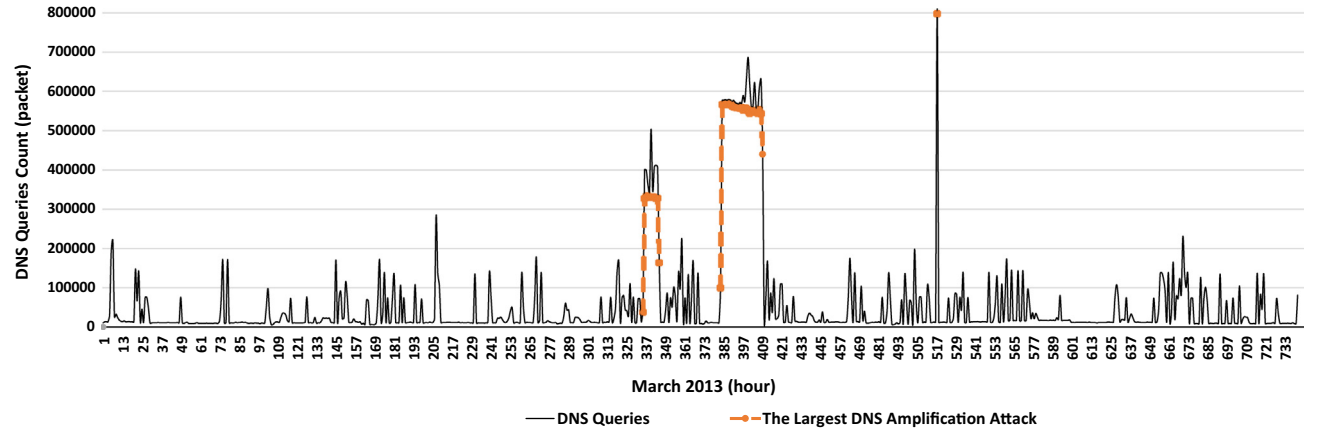


Fig. 6. DNS queries distribution – March 2013 data.

Table 4
Top 5 DNS query type – 2013 semiannual darknet data.

January Packet_Count (%)	February Packet_Count (%)	March Packet_Count (%)	April Packet_Count (%)	May Packet_Count (%)	June Packet_Count (%)
9,717,559 A (48.91%)	10,047,038 A (49.02%)	27,649,274 ANY (64.23%)	18,378,685 ANY (54.60%)	71,798,518 ANY (86.14%)	7,174,182 ANY (81.08%)
6,738,709 ANY (33.91%)	7,763,817 ANY (37.88%)	11,310,058 A (26.28%)	11,595,908 A (34.45%)	10,966,132 A (13.15%)	19,876,332 A (18.48%)
3,323,599 TXT (16.72%)	2,479,572 TXT (12.10%)	2,459,257 TXT (5.71%)	3,402,073 TXT (10.11%)	473,973 TXT (0.56%)	410,547 TXT (0.38%)
50,473 MX (0.25%)	100,463 MX (0.49%)	500,143 MX (1.16%)	180,779 MX (0.54%)	69,117 MX (0.08%)	30,130 AAAA (0.02%)
36,438 PTR (0.18%)	29,232 PTR (0.14%)	63,340 RRSIG (0.15%)	28,716 AAAA (0.09%)	37,052 AAAA (0.04%)	15,441 MX (0.01%)

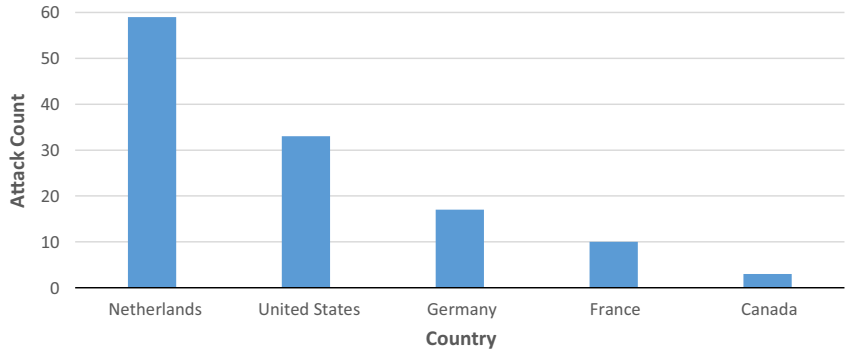


Fig. 7. Top 5 source countries (attacks).

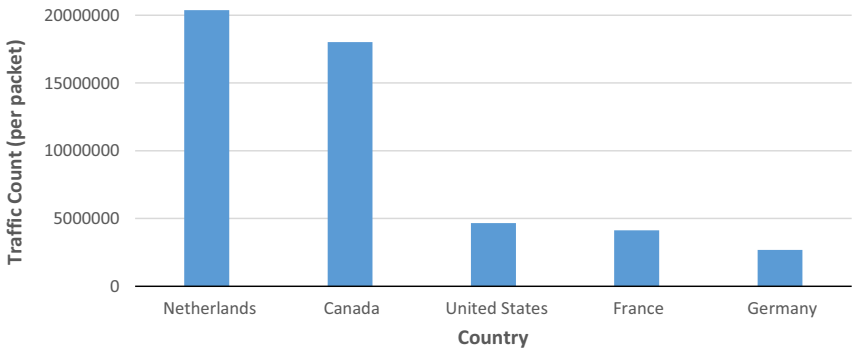


Fig. 8. Top 5 source countries (generated traffic).

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.

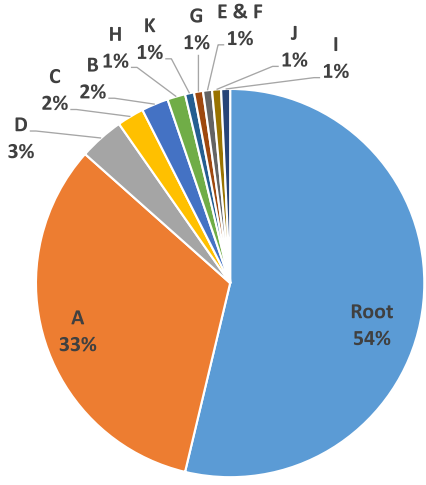


Fig. 9. Top requested domains.

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

Table 5
k-means clustered instances.

Cluster	*k*-means instances
0	31 (57%)
1	4 (7%)
2	12 (22%)
3	5 (9%)
4	2 (4%)

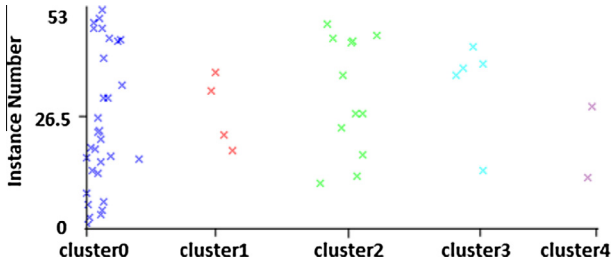


Fig. 10. *k*-means clustering of DNS amplified DDoS attacks.

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 [30]. 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[3] 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 resorted to publicly accessible Dshield [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 Dshield 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/ccsu_courses/DataMining-Ex3.html.

[3] The map was generated using Gephi [31], an open source visualization tool.

Table 6
k-means training cluster centroids.

Attribute	Cluster 0 (49)	Cluster 1 (8)	Cluster 2 (14)	Cluster 3 (5)	Cluster 4 (4)
high.asn.numb	ASN-V	ASN-W	ASN-X	ASN-Y	ASN-Y
ip.flag	0x02	0x00	0x02	0x00	0x02
ip.flags.df	0	1	0	1	0
ip.len	56	45	64	64	64
ip.ttl	<60	<60	<60	>100	<60
udp.length	36	34	44	44	44
dns.qry.name	Root	B	A	A	A
flow.avg.pkt.size	70	68	78	78	78
flow.attack.duration	<1 day	<1 day	<1 day	<1 day	btw-day-1 week

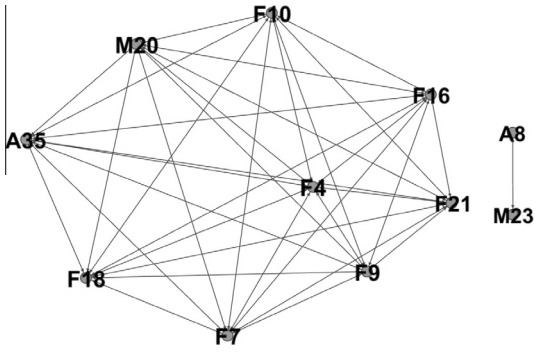


Fig. 11. IPs sharing at least 90% darknet space.

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 darknet 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 360000 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,[4] 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 300 Gb 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

[4] As of November 2013, this is very probable as there is around 32 million open DNS servers on the entire Internet [36].

Table 7
Summary of the Analyzed DNS Amplification DDoS Traces (February 2013).

Victim/ scanner ID	Requested domain name	Detection period	Analyzed attack duration (s)	Intensity (packet)	Contacted unique dark IPs	Avg. packet size (bytes)	Avg. rate (pps)	Rate category
F1	A	February 19	0	34,410	34,410	78	79565.67	High
F2	G	February 14	4477	129,206	129,206	85	28.86	Medium
F3	A	February 21	29,174	690,219	305,544	78	23.66	Medium
F4	Root	February 26	17,084	351,617	351,617	70	20.58	Medium
F5	Root	February 19	16,245	290,590	290,590	70	17.89	Medium
F6	Root	February 26	9389	162,513	162,513	70	17.31	Medium
F7	Root	February 11–12	25,052	349,692	349,692	70	13.96	Medium
F8	Root	February 20	15,215	187,886	187,886	70	12.35	Medium
F9	Root	February 13	61,591	660,473	356,162	70	10.72	Medium
F10	Root	February 16–17	33,602	355,188	355,188	70	10.57	Medium
F11	Root	February 3	6625	64,726	64,726	70	9.8	Medium
F12	Root	February 23	11,412	96,216	96,216	70	8.4	Medium
F13	Root	February 2–3	93,268	633,886	357,497	70	6.8	Medium
F14	A	February 3	19,872	128,297	128,297	78	6.46	Medium
F15	Root	February 7	2107	12,965	12,965	70	6.15	Medium
F16	Root	February 23–27	401,266	804,348	359,868	70	2	Medium
F17	Root	February 11–15	311,301	316,425	316,425	70	1.02	Medium
F18	Root	February 4–19	1,322,119	869,395	360,666	70	0.66	Medium
F19	Root	February 4–14	853,983	540,412	356,117	70	0.63	Medium
F20	A	February 3	10,634	6632	6632	78	0.62	Medium
F21	A	February 3–16	1,138,804	683,321	359,470	78	0.6	Medium
F22	Root	February 20–28	766,810	378,289	319,668	70	0.49	Low
F23	Root	February 5	27,832	9645	8123	70	0.35	Low
F24	A	February 19	50,374	16,393	16,393	78	0.33	Low
F25	A	February 4	16,353	5306	5306	78	0.32	Low
F26	Root	February 6–26	1,706,728	191,562	191,329	70	0.11	Low
F27	Root	February 15–26	970,150	19,636	19,636	70	0.02	Low
F28	A	February 9–28	1,691,139	16,845	16,845	78	0.01	Low
F29	A	February 15–22	640,165	966	966	78	0	Low

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 portsweep 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 NIDS. Further, our approach, leveraging the darknet space, can infer DNS amplification DDoS activities while an NIDS is limited to pinpointing scanning activities.

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.

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 factor 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, all 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.

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

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

Victim/ scanner ID	Requested domain name	Detection period	Analyzed attack duration (s)	Intensity (packet)	Contacted unique dark IPs	Avg. packet size (bytes)	Avg. rate (pps)	Rate Category
M1	A	March 18	1	50,257	50,257	78.00	46677.36	High
M2	A	March 31	26	63,543	63,543	78.00	2419.83	Medium
M3	E & F	March 22	620	798,192	65,025	73.00	1287.41	Medium
M4	A	March 20	402	91,042	91,042	67.00	226.21	Medium
M5	B	March 17–18	93,508	14,464,427	360,705	68.00	154.69	Medium
M6	Root	March 3	572	64,956	64,956	70.00	113.53	Medium
M7	Root	March 23	662	64,230	64,230	70.00	97.00	Medium
M8	Root	March 30	610	58,104	58,104	70.00	95.19	Medium
M9	Root	March 24	665	63,139	63,139	70.00	94.99	Medium
M10	B	March 15	34,605	3,176,785	360,683	68.00	91.80	Medium
M11	Root	March 1	769	63,342	63,342	70.00	82.33	Medium
M12	A	March 25	985	79,333	54,632	78.00	80.52	Medium
M13	Root	March 12	581	40,364	37,160	70.00	69.46	Medium
M14	Root	March 1–2	2685	161,847	154,905	70.00	60.28	Medium
M15	C	March 25	1	60	60	77.00	58.69	Medium
M16	A	March 9	8884	504,794	270,352	78.00	56.82	Medium
M17	A	March 30	1963	63,623	63,623	78.00	32.41	Medium
M18	Root	March 21	10,255	254,285	254,285	70.00	24.80	Medium
M19	Root	March 7	13,572	247,483	247,483	70.00	18.23	Medium
M20	Root	March 2	25,314	355,675	355,675	70.00	14.05	Medium
M21	Root	March 13	9796	128,147	128,147	70.00	13.08	Medium
M22	Root	March 27	24,391	286,664	286,664	70.00	11.75	Medium
M23	Root	March 8	33,354	346,244	346,244	70.00	10.38	Medium
M24	Root	March 28–29	33,280	342,941	342,941	70.00	10.30	Medium
M25	A	March 17–18	71,943	358,931	267,826	78.00	4.99	Medium
M26	A	March 30	13,667	61,269	51,999	78.00	4.48	Medium
M27	Root	March 14–17	342,024	1,396,535	360,701	70.00	4.08	Medium
M28	Root	March 28–29	56,305	224,327	224,327	70.00	3.98	Medium
M29	Root	March 11	73,864	248,582	129,708	70.00	3.37	Medium
M30	A	March 24	213	663	663	78.00	3.12	Medium
M31	Root	March 28–29	85,385	221,213	221,213	70.00	2.59	Medium
M32	A	March 30	163	397	396	78.00	2.43	Medium
M33	A	March 29–30	82,278	159,295	159,295	78.00	1.94	Medium
M34	A	March 30	330	640	639	78.00	1.94	Medium
M35	Root	March 24–25	69,590	127,214	127,214	70.00	1.83	Medium
M36	A	March 31	38,596	63,553	63,311	78.00	1.65	Medium
M37	Root	March 21–24	182,116	254,529	130,964	60.00	1.40	Medium
M38	Root	March 4–5	140,455	184,555	159,959	70.00	1.31	Medium
M39	Root	March 22–25	276,510	352,012	352,011	70.00	1.27	Medium
M40	Root	March 22–23	116,870	118,871	65,213	70.00	1.02	Medium
M41	Root	March 15–29	1,207,792	1,171,393	360,697	70.00	0.97	Medium
M42	Root	March 22–29	563,031	404,882	351,862	70.00	0.72	Medium
M43	A	March 1	21,616	7107	7107	78.00	0.33	Low
M44	A	March 15	52,584	17,013	17,013	78.00	0.32	Low
M45	A	March 1–7	466,136	92,176	89,073	78.00	0.20	Low
M46	A	March 15–31	1,393,227	152,254	134,270	78.00	0.11	Low
M47	A	March 6–30	2,119,713	194,209	65,792	78.00	0.09	Low
M48	A	March 13	24,521	2297	2117	78.00	0.09	Low
M49	Root	March 6–24	1,570,323	64,062	63,698	70.00	0.04	Low
M50	A	March 18–28	642,350	278	236	78.00	0.00	Low
M51	D	March 27–28	41,548	44	44	70.00	0.00	Low
M52	D	March 27–28	75,803	42	42	70.00	0.00	Low
M53	D	March 27–28	90,128	39	39	70.00	0.00	Low
M54	D	March 27–28	56,874	37	37	70.00	0.00	Low

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, the discussed case studies elaborated on three attack categories and provided significant cyber security intelligence related to them. As for future work, we aim to execute our model on a larger data set and experiment with more complex data mining exercises to improve our clustering model. 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.

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

Table 9
Summary of the Analyzed DNS Amplification DDoS Traces (April 2013).

Victim/ scanner ID	Requested domain name	Detection period	Analyzed attack duration (s)	Intensity (packet)	Contacted unique dark IPs	Avg. packet size (bytes)	Avg. rate (pps)	Rate category
A1	A	April 15	3	61,859	61,859	78	21672.18	High
A2	H	April 13	136	64,485	64,485	70	472.64	Medium
A3	Root	April 10	70	18,718	18,718	70	266.8	Medium
A4	A	April 21	4463	479,863	264,283	78	107.51	Medium
A5	Root	April 25	4023	151,894	151,894	70	37.76	Medium
A6	Root	April 20	325	11,068	11,068	70	34.05	Medium
A7	C	April 28	1274	40,903	40,903	77	32.11	Medium
A8	Root	April 4	6927	218,917	218,917	70	31.6	Medium
A9	Root	April 25	3171	57,837	42,578	70	18.24	Medium
A10	A	April 4	3791	68,039	56,211	78	17.95	Medium
A11	Root	April 16	8723	154,154	154,154	70	17.67	Medium
A12	Root	April 11	24,015	350,275	350,275	70	14.59	Medium
A13	I	April 1	23,608	340,905	340,905	92	14.44	Medium
A14	Root	April 25	39,305	408,596	408,596	70	10.4	Medium
A15	Root	April 16–17	27,760	284,387	284,386	70	10.24	Medium
A16	Root	April 12	6821	64,299	64,299	70	9.43	Medium
A17	Root	April 16–17	65,224	610,166	355,290	70	9.35	Medium
A18	Root	April 13–14	11,834	95,117	95,117	70	8.04	Medium
A19	B	April 5–6	73,456	345,133	343,652	79	4.7	Medium
A20	Root	April 14–15	42,560	182,836	182,834	60	4.3	Medium
A21	A	April 20–21	55,680	237,640	190,915	67	4.27	Medium
A22	Root	April 6–8	179,271	695,695	360,267	60	3.88	Medium
A23	A	April 15–16	89,471	346,554	346,554	78	3.87	Medium
A24	Root	April 1–2	135,389	507,427	291,844	70	3.75	Medium
A25	A	April 18	23	85	85	78	3.75	Medium
A26	A	April 24–30	568,658	1,601,134	357,930	78	2.82	Medium
A27	Root	April 1–2	120,727	316,718	224,789	70	2.62	Medium
A28	A	April 21	46,328	116,129	65,563	78	2.51	Medium
A29	Root	April 2–3	90,532	222,416	222,416	70	2.46	Medium
A30	Root	April 13–15	184,882	408,581	228,422	70	2.21	Medium
A31	Root	April 22–23	145,929	321,446	257,906	70	2.2	Medium
A32	A	April 3–4	56,113	120,662	120,662	78	2.15	Medium
A33	Root	April 1–29	2,463,203	3,495,104	360,705	70	1.42	Medium
A34	Root	April 13–22	777,630	1,049,946	360,690	70	1.35	Medium
A35	Root	April 3–8	463,324	593,142	357,414	70	1.28	Medium
A36	Root	April 7–11	295,595	316,685	225,376	70	1.07	Medium
A37	A	April 10–20	839,737	746,958	297,831	78	0.89	Medium
A38	Root	April 27–28	91,306	64,338	64,338	70	0.7	Medium
A39	A	April 12	18,587	6049	6049	78	0.33	Low
A40	A	April 5–20	1,312,707	385,495	65,792	78	0.29	Low
A41	A	April 25–30	431,330	119,938	65,642	78	0.28	Low
A42	C	April 17–19	158,580	40,362	40,362	77	0.25	Low
A43	Root	April 13–20	543,326	129,962	95,477	70	0.24	Low
A44	A	April 1–4	288,469	60,878	60,878	78	0.21	Low
A45	A	April 17–26	831,476	131,106	109,673	78	0.16	Low
A46	Root	April 14–20	496,168	63,559	40,901	70	0.13	Low
A47	Root	April 5–10	426,625	35,125	35,125	70	0.08	Low
A48	J	April 2–23	1,828,890	81,868	3744	75.49	0.04	Low
A49	H	April 9–10	96,970	1077	1074	70	0.01	Low
A50	K	April 23–30	640,451	8964	7871	68	0.01	Low
A51	D	April 15–17	156,226	63	47	71.02	0	Low

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 **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

[1] Arbor Networks, 2012 Infrastructure Security Report. <http://tinyurl.com/ag6tht4> (accessed 15.11.13).
[2] Forbes, Testing The Limits, LulzSec Takes Down CIA's Website. <http://tinyurl.com/bfhzbta> (accessed 15.11.13).

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016

[3] PcWorld, Hacker Arrested for DDoS Attacks on Amazon.com. <http://tinyurl.com/d22myng> (accessed 15.11.13).
[4] ITPRO, InfoSection 2011: Energy Firms Pummelled by DDoS Attacks. <http://tinyurl.com/cpqodbx> (accessed 15.11.13).
[5] US-CERT, DNS Amplification Attacks – Alert (TA13-088A). <http://tinyurl.com/bp4xud4> (accessed 15.11.13).
[6] Ars Technica, When Spammers go to War: Behind the Spamhaus DDoS. <http://tinyurl.com/d9vkegg> (accessed 15.11.13).
[7] M. Anagnostopoulos, G. Kambourakis, P. Kopanos, G. Louloudakis, S. Gritzalis, Dns amplification attack revisited, Comput. Secur. Part B (39) (2013) 475–485. <http://www.sciencedirect.com/science/article/pii/S0167404813001405>.
[8] M. Bailey, E. Cooke, F. Jahanian, J. Nazario, D. Watson, et al., The internet motion sensor: a distributed blackhole monitoring system, in: Proceedings of the 12th ISOC Symposium on Network and Distributed Systems Security (SNDSS), 2005, pp. 167–179.
[9] Yegneswaran, Vinod et al., On the design and use of internet sinks for network abuse monitoring, in: Recent Advances in Intrusion Detection, 2004.
[10] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, G. Huston, Internet background radiation revisited, in: Proceedings of the 10th Annual Conference on Internet Measurement, ACM, 2010, pp. 62–74.
[11] C. Rossow, Amplification Hell: Revisiting Network Protocols for DDoS Abuse.
[12] D. Moore, C. Shannon, D.J. Brown, G.M. Voelker, S. Savage, Inferring internet denial-of-service activity, ACM Trans. Comput. Syst. (TOCS) 24 (2) (2006) 115–139.
[13] D. Dagon, N. Provos, C.P. Lee, W. Lee, Corrupted dns resolution paths: the rise of a malicious resolution authority, in: NDSS, 2008.
[14] S. Staniford, D. Moore, V. Paxson, N. Weaver, The top speed of flash worms, in: Proceedings of the 2004 ACM Workshop on Rapid Malcode, WORM '04, ACM, New York, NY, USA, 2004, pp. 33–42, http://dx.doi.org/10.1145/1029618.1029624.
[15] S. Staniford, J.A. Hoagland, J.M. McAlerney, Practical automated detection of stealthy portscans, J. Comput. Secur. 10 (1) (2002) 105–136.
[16] The University of Waikato, WEKA Select Attributes: Ranker. <http://tinyurl.com/kt6rccu> (accessed 15.11.13).
[17] T.K. Moon, The expectation-maximization algorithm, Signal Process. Mag. IEEE 13 (6) (1996) 47–60.
[18] J.A. Hartigan, M.A. Wong, Algorithm as 136: a *k*-means clustering algorithm, J. Roy. Stat. Soc. Ser. C (Appl. Stat.) 28 (1) (1979) 100–108.
[19] D. Hudson, Interval estimation from the likelihood function, J. Roy. Stat. Soc. Ser. B (Methodol.) (1971) 256–262.
[20] G. Münz, S. Li, G. Carle, Traffic anomaly detection using k-means clustering, in: GI/ITG Workshop MMBnet, 2007.
[21] J. Yu, Z. Li, H. Chen, X. Chen, A detection and offense mechanism to defend against application layer ddos attacks, in: Third International Conference on Networking and Services, 2007, ICNS, IEEE, 2007, p. 54.
[22] S. Zhong, T.M. Khoshgoftaar, N. Seliya, Clustering-based network intrusion detection, Int. J. Reliab. Qual. Safety Eng. 14 (02) (2007) 169–187.
[23] A. Nucci, Architecture, Systems and Methods to Detect Efficiently DoS and DDoS Attacks for Large Scale Internet, US Patent 7,584,507, September 1 2009.
[24] Han, Jiawei, Kamber, Micheline, Pei, Jian, Data Mining: Concepts and Techniques, 2006.
[25] The University of Waikato, Weka: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/> (accessed 15.11.13).
[26] CloudFlare blog, The DDoS That Knocked Spamhaus Offline (And How We Mitigated It). <http://tinyurl.com/d46gpkj> (accessed 14.04.14).
[27] J. Oberheide, M. Karir, Z.M. Mao, Characterizing dark dns behavior, in: Fourth GI International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA), 2007.
[28] Spamhaus, An Arrest in Response to March DDoS Attacks on Spamhaus. <http://www.spamhaus.org/news/article/698/> (accessed 15.11.13).
[29] The New York Times, Firm Is Accused of Sending Spam, and Fight Jams Internet. <http://tinyurl.com/bnkmr4c> (accessed 15.11.13).
[30] P. Smyth, Model selection for probabilistic clustering using cross-validated likelihood, Stat. Comput. 10 (1) (2000) 63–72.
[31] Gephi, Gephi: An Open Source Graph Visualization and Manipulation Software. <https://gephi.org/> (accessed 14.04.14).
[32] DShield, Dshield Port Report. <http://www1.dshield.org> (accessed 15.11.13).
[33] Bloomberg News, Dutch Man Arrested in Spain in Probe of Spamhaus Attack. <http://tinyurl.com/puru9l2> (accessed 14.04.14).
[34] The Irish Times, Spam Dispute Results in Biggest Ever Cyber Attack. <http://tinyurl.com/d2xwv9w> (accessed 14.04.14).
[35] Luc Rossini, Twitter Tweet: Spamhaus is Currently Under a DDoS Attack Against Our Website which We are Working on Mitigating. Our DNSBLs are Not Affected. <http://tinyurl.com/m7dayxr> (accessed 14.04.14).
[36] Open Resolver Project. <http://openresolverproject.org/> (accessed 15.11.13).
[37] Sourcefire, Snort NIDS. <http://www.snort.org/> (accessed 15.11.13).
[38] D. Moore, G. Voelker, S. Savage, Inferring internet denial-of-service activity, in: Proceedings of the 10th Usenix Security Symposium, 2001.
[39] S.M. Bellovin, There be dragons, in: Proceedings of the Third Usenix Unix Security Symposium, 1992, pp. 1–16.
[40] A. Dainotti, A. King, F. Papale, A. Pescapè, et al., Analysis of a/0 stealth scan from a botnet, in: Proceedings of the 2012 ACM Conference on Internet Measurement Conference, ACM, 2012, pp. 1–14.
[41] M. Bailey, E. Cooke, F. Jahanian, D. Watson, J. Nazario, The blaster worm: then and now, Secur. Privacy IEEE 3 (4) (2005) 26–31.
[42] A. Dainotti, C. Squarcella, E. Aben, K.C. Claffy, M. Chiesa, M. Russo, A. Pescape, Analysis of country-wide internet outages caused by censorship, in: Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC, 2011, pp. 1–17.
[43] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, S. Sinha, Practical darknet measurement, in: 2006 40th Annual Conference on Information Sciences and Systems, 2006, p. 14961501.
[44] M. Bailey, E. Cooke, F. Jahanian, N. Provos, K. Rosaen, D. Watson, Data reduction for the scalable automated analysis of distributed darknet traffic, in: Proceedings of the USENIX/ACM Internet Measurement Conference, 2005.
[45] CAIDA, Cooperative Association for Internet Data Analysis. <http://www.caida.org> (accessed 15.11.13).
[46] P.A.R. Kumar, S. Selvakumar, Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems, Comput. Commun. 36 (3) (2013) 303–319.
[47] S. Bhatia, D. Schmidt, G. Mohay, Ensemble-based ddos detection and mitigation model, in: Proceedings of the Fifth International Conference on Security of Information and Networks, SIN '12, ACM, New York, NY, USA, 2012, pp. 79–86, http://dx.doi.org/10.1145/2388576.2388587.
[48] D.Q. Le, T. Jeong, H.E. Roman, J.W.-K. Hong, in: Proceedings of the Second Symposium on Information and Communication Technology, SoICT '11, ACM, New York, NY, USA, 2011, pp. 36–41, http://dx.doi.org/10.1145/2069216.2069227.
[49] Z.M. Fadlullah, T. Taleb, A.V. Vasilakos, M. Guizani, N. Kato, Dtrab: combating against attacks on encrypted protocols through traffic-feature analysis, IEEE/ACM Trans. Network. (TON) 18 (4) (2010) 1234–1247.
[50] Z.H. Aghaei Foroushani, TDFA: traceback-based defense against DDoS flooding attacks, in: 28th International Conference on Advanced Information Networking and Applications, AINA, IEEE, 2014, pp. 710–715.
[51] G. Yao, J. Bi, Z. Zhou, Passive IP traceback: capturing the origin of anonymous traffic through network telescopes, ACM SIGCOMM Computer Communication Review, vol. 40, ACM, 2010, pp. 413–414.
[52] J. Bi, P. Hu, P. Li, Study on classification and characteristics of source address spoofing attacks in the internet, in: Ninth International Conference on Networks (ICN), 2010, IEEE, 2010, pp. 226–230.
[53] E. Gansner, B. Krishnamurthy, W. Willinger, F. Bustamante, M. Snchez, Demo abstract: towards extracting semantics by visualizing large traceroute datasets, Computing 96 (1) (2014) 81–83, http://dx.doi.org/10.1007/s00607-013-0290-8.
[54] S. Papadopoulos, G. Theodoridis, D. Tzovaras, Bgpfuse: Using visual feature fusion for the detection and attribution of bgp anomalies, in: Proceedings of the Tenth Workshop on Visualization for Cyber Security, VizSec '13, ACM, New York, NY, USA, 2013, pp. 57–64, http://dx.doi.org/10.1145/2517957.2517965.
[55] B. Irwin, N. Pilkington, High Level Internet Scale Traffic Visualization using Hilbert Curve Mapping, Springer, 2008.
[56] V. Paxson, An analysis of using reflectors for distributed denial-of-service attacks, SIGCOMM Comput. Commun. Rev. 31 (3) (2001) 38–47, http://dx.doi.org/10.1145/505659.505664.
[57] CloudFlare blog, Technical Details Behind a 400 Gbps NTP Amplification DDoS Attack. <http://tinyurl.com/p3exvnc> (accessed 14.04.14).

Please cite this article in press as: C. Fachkha et al., Inferring distributed reflection denial of service attacks from darknet, Comput. Commun. (2015), http://dx.doi.org/10.1016/j.comcom.2015.01.016