Chapter 26
On Inferring and Characterizing Large-Scale Probing and DDoS Campaigns

Elias Bou-Harb and Claude Fachkha

26.1 Introduction

Cyberspace is the electronic world created by interconnected networks of Information Technology (IT) and the information on those networks. It can be defined as the interdependent network of IT infrastructure, including the Internet, telecommunication networks, computer systems, and embedded industrial processors and controllers. Cyberspace is a global commons where more than 1.7 billion people are linked together to exchange ideas and services [1]. Moreover, it underpins almost every facet of a modern society and provides critical support for the economy, civil infrastructure, public safety, and national security.

However, recent events have indeed demonstrated that cyberspace could be subjected, at the speed of light and in full anonymity, to severe attacks with drastic consequences. One particular research revealed that 90% of corporations have been the target of a cyber attack, with 80% suffering a significant financial loss [2]. In addition, the cyber security report [1] elaborated that in a recent 1 year period, 86% of large North American organizations had suffered a cyber attack where the loss of intellectual property as a result of these attacks doubled between 2011 and 2015. Moreover, the report alarmed that more than 60% of all the malicious code ever detected, originating from more than 190 countries, was introduced into cyberspace solely in 2016.

E. Bou-Harb (✉)
Cyber Threat Intelligence Lab, Florida Atlantic University, Boca Raton, FL, USA
e-mail: ebouharb@fau.edu

C. Fachkha
University of Dubai, Dubai, United Arab Emirates
e-mail: cfachkha@ud.ac.ae

K. Daimi (ed.), *Computer and Network Security Essentials*,
DOI 10.1007/978-3-319-58424-9_26

To this end, generating effective cyber threat intelligence is indeed an effective approach that would aid in preventing, inferring, characterizing, analyzing, and mitigating various Internet-scale malicious activities. Thus, in this chapter, we aim to generate such cyber threat intelligence related to two specific types of malicious actives, namely probing and DDoS events, and their corresponding orchestrated campaigns, by analyzing the darknet IP space.

26.1.1 Background

In this section, we provide brief yet relevant background information related to the concerned topics.

Probing Activities Probing or scanning events [3] could be defined by the task of executing reconnaissance activities towards enterprise networks or Internet-wide services, searching for vulnerabilities or ways to infiltrate IT assets. Such events are commonly the primary stage of an intrusion attempt that enables an attacker to remotely locate, target, and subsequently exploit vulnerable systems [4]. They are basically a core technique and a facilitating factor of various subsequent cyber attacks. Readers that are further interested in inner details related to probing activities are kindly referred to the following surveys [5, 6].

DDoS Activities Denial of Service (DoS) attacks are characterized by an explicit attempt to prevent the legitimate use of a service. Distributed DoS (DDoS) attacks employ multiple attacking entities (i.e., compromised machines/bots) to achieve their intended aim. DDoS attacks could be related to flooding attempts, in which the bots directly attack the victim, or they could be rendered by amplification attempts, where the attacker employs third party servers known as open amplifiers to indirectly launch the attack towards the victim. Readers that are interested in more details related to DDoS activities are kindly referred to [7].

Darknets A network telescope, also commonly referred to as a darknet or an Internet sink [8], is a set of routable and allocated yet unused IP addresses [9]. It represents a partial view of the entire Internet address space. From a design perceptive, network telescopes are transparent and indistinguishable compared with the rest of the Internet space. From a deployment perspective, it is rendered by network sensors that are implemented and dispersed on numerous strategic points throughout the Internet. Such sensors are often distributed and are typically hosted by various global entities, including Internet Service Providers (ISPs), academic and research facilities, and backbone networks. The aim of a darknet is to provide a lens on Internet-wide malicious traffic; since darknet IP addresses are unused, any traffic targeting them represents a continuous view of anomalous unsolicited traffic.

Orchestrated Campaigns A number of malicious activities could operate within the context of large-scale campaigns. These render a new era of such malicious

events, since they are distinguished from previous independent incidents as (1) the population of the participating bots is several orders of magnitude larger, (2) the target scope is generally the entire IP address space, and (3) the bots adopt well-orchestrated, often botmaster coordinated, stealth scan strategies that maximize targets' coverage while minimizing redundancy and overlap. Readers that are further interested in inner details related to large-scale orchestrated malicious campaigns are kindly referred to [10].

In this chapter, we aim to infer and characterize probing and DDoS orchestrated campaigns by uniquely analyzing darknet traffic.

26.1.2 Organization

The remaining of this chapter is organized as follows. In the next section, we address the problem of inferring independent and orchestrated probing events while in Sect. 26.3, we focus on inferring and characterizing DDoS events and large-scale campaigns. In Sect. 26.4, we review some literature work to demonstrate the uniqueness of the presented work. We conclude this chapter in Sect. 26.5 by summarizing the offered contributions and pinpointing several topics that are worthy of being investigated in the future.

26.2 Probing Campaigns

In this section, we present methods to infer independent and orchestrated probing events by scrutinizing darknet data. Further, we present some results characterizing such events.

26.2.1 Inferring Probing Events

Motivated by recent cyber attacks that were facilitated through probing [11], limited cyber security intelligence, and the lack of accuracy that is provided by scanning detection systems, this section presents a new approach to fingerprint Internet-scale probing activities. The rationale of the proposed method states that regardless of the source, strategy, and aim of the probing, the reconnaissance activity should have been generated using a certain literature-known scanning technique (i.e., TCP SYN, UDP, ACK, etc. [5]). We observe that a number of those probing techniques demonstrate a similar temporal correlation and similarity when generating their corresponding probing traffic. In other words, the observation states that we can cluster the scanning techniques based on their traffic correlation

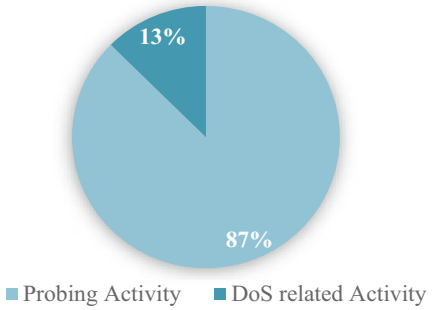

Fig. 26.1 Sessions distribution

statuses. Subsequently, we can differentiate between probing and other darknet malicious traffic based on the possessed traffic correlation status. We can as well attribute the probing traffic to a certain cluster of scanning techniques (i.e., the probing activity, after confirmed as probing, can be identified as being generated by a certain cluster of techniques that possess similar traffic correlation status). To identify exactly which scanning technique has been employed in the probing, we statistically estimate the relative closeness of the probing traffic in comparison with the techniques found in that cluster. To enable the capturing of traffic signals correlation statuses, the proposed method employs the Detrended Fluctuation Analysis (DFA) technique [12]. Elaborative details about the modus operandi of the proposed inference method could be found in [13].

Empirical Results We employ around 10 GB of real darknet data to evaluate the inference approach. We first applied the approach to attempt to differentiate between scanning and darknet backscattered traffic (i.e., DoS related activity). Figure 26.1 represents how the 700 sessions were distributed and fingerprinted. It is shown that probing activity corresponds to 87% (612) of all the sessions. This scanning to backscattered traffic ratio is somehow coherent with other darknet studies [14]. To evaluate the scanning fingerprinting capabilities of our approach, we experimented with Snort's sfPortscan preprocessor using the same 612 sessions that were fingerprinted as probing. Snort's sfPortscan detected 590 scans. After a semi-automated analysis and comparison that was based on the logged scanning traffic flows, we identified that all the 612 scans that our approach fingerprinted as probing activity include sfPortscan's 590 scans. Therefore, relative to this technique and experimenting with this specific data set, we confirm that our approach yielded no false negative, with only 2% as false positives.

26.2.2 Inferring and Characterizing Probing Campaigns

To infer orchestrated probing campaigns, for each of the previously inferred probing event, we generate their feature vectors as summarized in Table 26.1. The machinery that would generate such vectors is summarized in [15].

Table 26.1 Probing feature vectors

Employed probing technique
Probing traffic (random vs patterns)
Employed pattern
Adopted probing strategy
Nature of probing source
Type of probing (targeted vs dispersed)
Signs of malware infection
Exact malware type/variant
Probing rate
Ratio of destination overlaps
Target port

To automatically infer orchestrated probing events, the approach leverages all the previously extracted inferences and insights related to the probing sessions/sources to build and parse a Frequent Pattern (FP) tree. In such a tree, each node after the root represents a feature extracted from the probing sessions, which is shared by the sub-trees beneath. Each path in the tree represents sets of features that co-occur in the sessions, in non-increasing order of frequency of occurrences. Thus, two sessions that have several frequent features in common and are different just on infrequent features will share a common path in the tree. The proposed approach also employs the FP tree-based mining method, FP-growth, for mining the complete set of generated frequent patterns. As an outcome, the generated patterns represent frequent and similar probing behavioral characteristics that correlate the probing sources into orchestrated probing events.

Empirical Results We evaluate the proposed approach using 330 GB of darknet data. We visualize the outcome of the feature vectors as depicted in Fig. 26.2. Such "flower-based" result intuitively and creatively illustrates how the FP-tree is constructed. Recall that the tree depicts frequent probing features that co-occur in the probing sessions, which are generated by the probing behavioral analytics. One can notice several groupings or clusters that depict probing events sharing various common machinery. For the sake of this work, we have devised a parsing algorithm that automatically build patterns from the FP-tree that aim at capturing orchestrated probing events that probe horizontally; probe all IPs by focusing on specific ports.

The proposed approach automatically inferred the pattern that is summarized in Table 26.2. The pattern permitted the detection, identification, and correlation of 846 unique probing bots into a well-defined orchestrated probing event that targeted the VoIP (SIP) service. It is shown that this event adopted UDP scanning, probed around 65% of the monitored dark space (i.e., 300,000 dark IPs) where all its bots did not follow a certain pattern when generating their probing traffic. Further, the results demonstrate that the bots employed a reverse IP-sequential probing strategy when probing their targets. Moreover, the malware responsible for this event was shown to be attributed to the Sality malware.

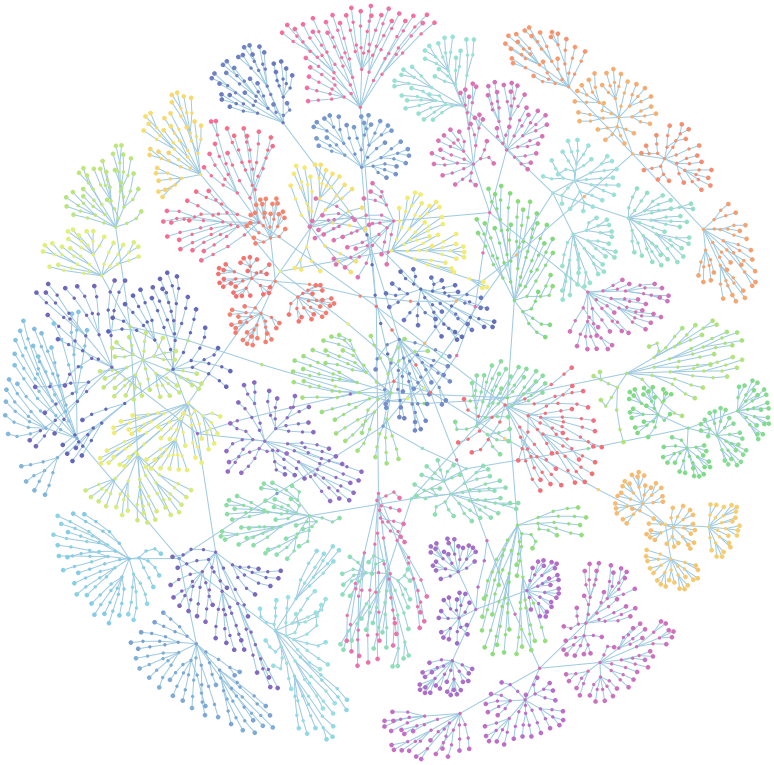

Fig. 26.2 Visualization of the outcome of the probing behavioral analytics in the FP-tree

Table 26.2 The inferred pattern capturing a large-scale orchestrated probing event

Employed probing technique: UDP
Probing traffic (random vs patterns): Random
Employed pattern: Null
Adopted probing strategy: Reverse IP-sequential
Nature of probing source: Bot
Type of probing (targeted vs dispersed): Dispersed
Signs of malware infection: Yes
Exact malware type/variant: Virus.Win32.Sality.bh
Probing rate: 12 pps
Target port: 5060

26.3 DDoS Campaigns

In this section, we present techniques to infer distributed and orchestrated DDoS events by analyzing real darknet data. In addition, we present some results characterizing these large-scale activities.

26.3.1 Inferring DDoS Events

This section leverages darknets to identify independent DDoS attacks. To achieve its aim, our approach adopts three steps: (1) selecting backscattered packets from victims' replies; (2) extracting session flows corresponding to malicious activities; and (3) inferring DDoS attacks by employing a detection algorithm. First, in order to select backscattered packets, we adopt the technique from [16] that relies on flags in packet headers, such as TCP SYN+ACK, RST, RST+ACK, and ACK. However, this technique might cause misconfiguration as well as scanning probes (i.e., SYN/ACK Scan) to co-occur within the backscattered packets. In order to filter out the misconfiguration, we use a simple metric that records the average number of sources per destination darknet address. This metric should be significantly larger for misconfiguration than scanning traffic [17]. Second, in order to filter out scanning activities, we split the connections into separate session flows, each of which consists of a unique source and destination IP/port pair. The rationale for this is that DDoS attempts possess a much greater number of packets sent to one destination (i.e., flood) whereas portsweep scanners have one or few attempts towards one destination (i.e., probe). Third, we aim to confirm that all the extracted sessions in fact reflect real DDoS attempts. To accomplish this, we employ a modified version of the DDoS detection parameters from [18] to label a session as a single DoS attack. Algorithm 1 displays our detection mechanism. We proceed by merging all the previously extracted sessions that have the same source IP (i.e., victim) to extract DDoS attacks.

Empirical Results Similar to the probing analysis, the data is based on the previously darknet data set collected during the same period. We inferred thousands of DDoS attacks, as per Fig. 26.3a, where the majority were shown to abuse TCP services (62%), ICMP (21%), and UDP (17%). Furthermore, as shown in Fig. 26.3b, these attack types are distributed as follows: 82% for TCP flooding, 14% for DNS flooding, and the remaining are ICMP flooding events.

26.3.2 Inferring and Characterizing DDoS Campaigns

In the previous sections, we elaborated on the components of the systematic approach for inferring DDoS activities targeting a unique organization. In this section, we extend the approach by proposing a clustering approach to infer orchestrated DDoS campaigns that target multiple victims. This permits the fingerprinting of the nature of such campaigns. For example, it could be identified that a specific DDoS campaign is specialized in targeting financial institutions while another campaign is focused on targeting critical infrastructure. Further, such clustering approach allows the elaboration on the actual scope of the DDoS campaign to provide cyber security situational awareness; how large is the campaign and what is

Algorithm 1 DDoS detection engine

In the algorithm:
Each flow f contains packet count (*pkt_cnt*) and rate (*rate*)
Tw: Time Window
p_th: Packet Threshold
r_th: Rate Threshold
Tn: Time of packet number *n* in a flow
pkt: Packet

Input: A set of darknet flows F where each f in F is composed of a pair of <source IP, destination IP> leveraging a series of consecutive packets that share the same source IP address.
Output: DDoS attack flows

for each f in F **do**
  $attack\_flag = 0$
  $pkt\_cnt = 0$
  $T1 =$ pkt_gettime(1)
  $Tf = T1 + Tw$
  while *pkt* in f **do**
    $Tn =$ pkt_gettime()
    if $Tn < Tf$ **then**
      pkt_cnt++
    end if
  end while
  $rate = \frac{pkt\_cnt}{Tw}$
  if $pkt\_cnt >$ p_th & $rate > r\_th$ **then**
    attack_flag = 1
  end if
end for

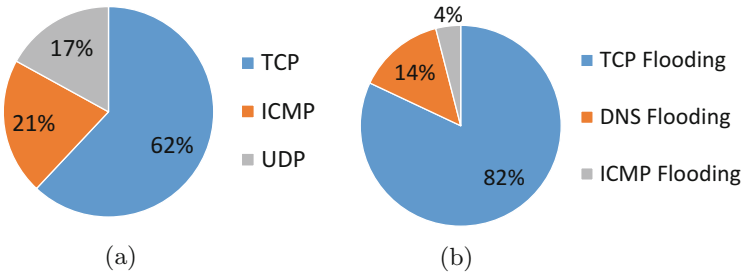

Fig. 26.3 DDoS: major protocols and distribution. (**a**) Abused protocols, (**b**) Attack distribution

its employed rates, when attacking the various victims. Additionally, the proposed approach could be leveraged to predict the campaign's features in terms of rate and number of involved machines.

To achieve this task, our approach employs the following statistical-based mechanism. First, backscattered sessions are extracted as previously discussed. Second,

the notion of fuzzy hashing [19] between the different sessions is applied. Fuzzy hashing is advantageous in comparison with typical hashing as it can provide a percentage of similarity between two traffic samples rather than producing a null value if the samples are different. This popular technique is derived from the digital forensics research field and is typically applied on files or images [19]. Our approach explores the capabilities of this technique on backscattered DDoS traffic. We select the sessions that demonstrate at least 20% similarity. We concur that this threshold is a reasonable starting point and aids in reducing false negatives. Third, from those similar sessions, we employ two statistical tests, namely the Euclidean and the Kolmogorov–Smirnov tests [20] to measure the distance between the selected sessions. We extract those sessions that minimize the statistical distance after executing both tests. The rationale of the latter approach stems from the need to cluster the sessions belonging to multiple victims that share a maximized similar traffic behavior while minimizing the false positives by confirming such similarity using various tests. Note that, we hereafter refer to the use of the previous two techniques as the fusion technique. The outcome of the proposed approach are clustered diverse victims that are inferred to be the target of the same orchestrated DDoS campaign.

Empirical Results In this section, we present the empirical evaluation results. We employ the DDoS campaign clustering model as discussed in the previous section to demonstrate how multiple victims could be modeled as being the target of the same campaign.

26.3.2.1 TCP SYN Flooding on Multiple HTTP Servers

To demonstrate the effectiveness of the approach, we experiment with a 1 day sample retrieved from our darknet data set. We extract more than 600 backscattered DDoS sessions and apply fuzzy hashing between the sessions, by leveraging deep-toad, a fuzzy hashing implementation. The outcome of this operation is depicted in Fig. 26.4a, where the victims are represented by round circles while directed arrows illustrate how the various victims were shown to be statistically close to other targeted victims. It is important to note that we anonymize the real identity of the victims due to sensitivity and legal reasons. Subsequently, the Euclidean and the Kolmogorov–Smirnov tests are executed to exactly pinpoint and cluster the victims that demonstrate significant traffic similarity. Figure 26.4b shows such result while Table 26.3 summarizes the outcome of the proposed DDoS campaign clustering approach. From Fig. 26.4b, one can notice the formation of root nodes, advocating that the approach is successful in clustering various victims that are the target of the same DDoS campaign.

In general, the approach yielded, for 1 day data set, 13 unique campaigns where each campaign clusters a number of victims ranging from 2 to 125 targets.

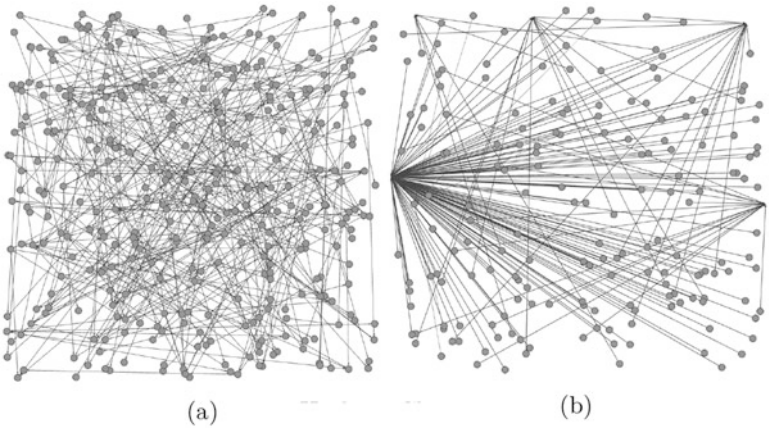

Fig. 26.4 DDoS clustering. (**a**) Fuzzy hashing clustering. (**b**) Fusion technique clustering

Table 26.3 Summary of the DDoS campaign clustering approach

Technique	Unique campaign count	Campaign of 2 victim machines	Campaign of 3 victim machines	Campaign of 4 victim machines	Campaign of 5 victim machines	Campaign of 6 victim machines	Campaign of 125 victim machines
Euclidean	16	6	2	3	3	1	1
KS	16	6	2	3	2	2	1
Fusion	13	6	1	2	2	1	1

26.4 Related Work

In this section, we review the related work on various concerned topics.

Extracting Probing Events Li et al. [21] considered large spikes of unique source counts as probing events. The authors extracted those events from darknet traffic using time-series analysis; they first automatically identified and extracted the rough boundaries of events and then manually refined the event starting and ending times. At this point, they used manual analysis and visualization techniques to extract the event. In an alternate work, Jin et al. [22] considered any incoming flow that touches any temporary dark (grey) IP address as potentially suspicious. The authors narrowed down the flows with sustained suspicious activities and investigated whether certain source or destination ports are repeatedly used in those activities. Using these ports, the authors separated the probing activities of an external host from other traffic that is generated from the same host. In contrast, in this work, we propose a method that exploits a unique observation related to the signal correlation status of probing events. By leveraging this, we are able to differentiate between probing and other events and subsequently extract the former from incoming darknet traffic.

Analyzing Probing Events The authors of [22, 23] studied probing activities towards a large campus network using netflow data. Their goal was to infer the probing strategies of scanners and thereby assess the harmfulness of their actions. They introduced the notion of gray IP space, developed techniques to identify potential scanners, and subsequently studied their scanning behaviors. In another work, the authors of [21, 24, 25] presented an analysis that drew upon extensive honeynet data to explore the prevalence of different types of scanning. Additionally, they designed mathematical and observational schemes to extrapolate the global properties of scanning events including total population and target scope. In contrary, we aim at inferring large-scale probing and DDoS campaigns rather than focusing on analyzing probing events.

Probing Measurement Studies In addition to [26, 27], Benoit et al. [28] presented the world's first Web census while Heidemann et al. [29] were among the first to survey edge hosts in the visible Internet. Further, Pryadkin et al. [30] offered an empirical evaluation of IP address space occupancy whereas Cui and Stolfo [31] presented a quantitative analysis of the insecurity of embedded network devices obtained from a wide-area scan. In a slightly different work, Leonary and Loguinov [32] demonstrated IRLscanner, a tool which aimed at maximizing politeness yet provided scanning rates that achieved coverage of the Internet in minutes. In this work, as previously mentioned, we strive to infer large-scale campaigns rather than solely providing measurements of particular events.

Botnet Detection Frameworks A number of botnet detection systems have been proposed in the literature [33–36]. Some investigates specific channels, others might require deep packet inspection or training periods, while the majority depends on malware infections and/or attack life-cycles. To the best of our knowledge, none of the proposals is dedicated to tackle the problem of inferring large-scale probing and DDoS campaigns. Further, in this work, we aim to achieve that task by analyzing the dark IP space and by focusing on the machinery and netflow characteristics of the received darknet traffic, without requiring content analysis or training periods.

26.5 Concluding Remarks

This chapter aims at generating effective cyber threat intelligence to aid in proactive and defensive protection of cyberspace. To this end, several techniques to detect and identify large-scale orchestrated probing and DDoS campaigns by leveraging real darknet data were elaborated. On one hand, we presented approaches that addressed the problem of inferring probing activities, which are typically the precursors of future cyber attacks. In particularity, we discussed an approach rooted in time-series fluctuation analysis to identity probing activities as well as attribute such events to a certain technique. Further, we leveraged this inference approach to detect orchestrating probing events, by proposing a feature generation and clustering approach. The latter is based on a set of behavioral data analytics

and the employment of a data mining method. On the other hand, we designed and developed darknet-based techniques to infer and characterize independent DDoS attacks. Additionally, we addressed the problem of DDoS campaigns by exploiting fuzzy and statistical methods. Empirical evaluations based on real darknet data demonstrated that the devised techniques, methods, and approaches are effective in inferring and characterizing such stealthy and devastating events.

26.5.1 Considerations and Research Gaps

Developing and deploying cyber security capabilities to combat contemporary threats in general, and cyber campaigns in particular, require several considerations. First, characterizing security information requires access to real attack data sets, which is relatively difficult to access or obtain. Second, developing techniques might not be as simple as deploying them. For instance, deploying darknet-based models require access to real hardware devices. Furthermore, deploying such techniques must be approved by authorities (network administrators, Internet Service Providers, etc.) and therefore necessitate significant collaborative effort and coordination. Our future plan is to deploy our models in real-time and leverage such capabilities to develop an Internet-scale situation awareness system, working closely with our partners and affiliations.

From the conducted research, we can extract the following points/research gaps:

- Inferring and attributing botnets or malicious campaigns by solely monitoring the dark IP space is very challenging due to the passive nature of such IP space. Therefore, other interactive techniques such as honeypots could be used in conjunction with darknet analysis to enhance botnet investigation.
- Packet analysis is the only technique employed on darknet data to investigate spoofing activities. This method is rendered by inspecting ICMP packets and TTL values. Minimal research has been executed to study spoofing events through darknet analysis. Therefore, spoofing is still a noteworthy malicious activity that needs more attention from the security research community.
- Despite the existence of few collaborative darknet projects, more darknet resources and information sharing efforts should emerge to infer and attribute large-scale cyber activities. Indeed, establishing a worldwide darknet information exchange is a capability that requires collaboration and trust; however, this collaboration necessitates the implementation of numerous global policies and undoubtedly would raise serious privacy concerns.
- There exists a need to explore darknet data to generate cyber threat intelligence for other evolving paradigms, include the Internet-of-Things (IoT) and Cyber-Physical Systems (CPS).

Acknowledgements The authors would like to acknowledge the computer security lab at Concordia University, Canada where most of the presented work was conducted. The authors are also grateful to the anonymous reviewers for their insightful comments and suggestions.

References

1. Government of Canada. (2010). Canada's cyber security strategy report, http://www.capb.ca/uploads/files/documents/Cyber_Security_Strategy.pdf.
2. Hinde, S. (2003). The law, cybercrime, risk assessment and cyber protection. *Computers & Security, 22*, 90–95.
3. Bou-Harb, E., Debbabi, M., & Assi, C. (2013). A statistical approach for fingerprinting probing activities. In *2013 Eighth International Conference on Availability, Reliability and Security (ARES)* (pp. 21–30), Sept 2013.
4. Bou-Harb, E., Lakhdari, N. -E., Binsalleeh, H., & Debbabi, M. (2014). Multidimensional investigation of source port 0 probing. *Digital Investigation, 11*(Supplement 2), S114–S123; Fourteenth Annual {DFRWS} Conference.
5. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2010). Surveying port scans and their detection methodologies. *The Computer Journal, 54*(10), 1565–1581.
6. Bou-Harb, E., Debbabi, M., & Assi, C. (2014). Cyber scanning: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 16*(3), 1496–1519.
7. Rossow, C. (2014). Amplification hell: Revisiting network protocols for DDoS abuse. In *NDSS*.
8. Fachkha, C., & Debbabi, M. (2016). Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization. *IEEE Communications Surveys & Tutorials, 18*(2), 1197–1227.
9. Moore, D., Shannon, C., Voelker, G. M., & Savage, S. (2004). Network Telescopes: Technical Report. Department of Computer Science and Engineering, University of California, San Diego.
10. Bou-Harb, E., Assi, C., & Debbabi, M. (2016). Csc-detector: A system to infer large-scale probing campaigns. *IEEE Transactions on Dependable and Secure Computing, PP*(99), 1
11. Bou-Harb, E., Debbabi, M., & Assi, C. (2013). A systematic approach for detecting and clustering distributed cyber scanning. *Computer Networks, 57*(18), 3826–3839
12. Peng, C. -K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Phys. Rev. E, 49*, 1685–1689.
13. Bou-Harb, E., Debbabi, M., & Assi, C. (2014). On fingerprinting probing activities. *Computers & Security, 43*, 35–48.
14. Wustrow, E., Karir, M., Bailey, M., Jahanian, F., Huston, G. (2010). Internet background radiation revisited. In *Proceedings of the 10th Annual Conference on Internet Measurement* (pp 62–74). New York, NY: ACM.
15. Bou-Harb, E., Debbabi, M., & Assi, C. (2014) Behavioral analytics for inferring large-scale orchestrated probing events. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 506–511). New York, NY: IEEE.
16. Moore, D., Voelker, G. M., & Savage, S. (2001). Inferring internet denial-of-service activity. Technical Report, DTIC Document.
17. Li, Z., Goyal, A., Chen, Y., & Paxson, V. (2011). Towards situational awareness of large-scale botnet probing events. *IEEE Transactions on Information Forensics and Security, 6*(1), 175–188.
18. Moore, D., Shannon, C., Brown, D.J., Voelker, G.M., & Savage, S. (2006). Inferring internet denial-of-service activity. *ACM Transactions on Computer Systems (TOCS), 24*(2), 115–139
19. Kornblum, J. (2006). Identifying almost identical files using context triggered piecewise hashing. *Digital Investigation, 3*(Supplement), 91–97; The Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS'06).

20. Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*(318), 399–402.
21. Li, Z., Goyal, A., Chen, Y., & Paxson, V. (2011). Towards situational awareness of large-scale botnet probing events. *IEEE Transactions on Information Forensics and Security, 6*(1), 175–188
22. Jin, Y., Simon, G., Xu, K., Zhang, Z.-L., & Kumar, V. (2007). Gray's anatomy: Dissecting scanning activities using IP gray space analysis. In *Usenix SysML07*.
23. Jin, Y., Zhang, Z. -L., Xu, K., Cao, F., & Sahu, S. (2007). Identifying and tracking suspicious activities through IP gray space analysis. In *Proceedings of the 3rd Annual ACM Workshop on Mining Network Data, MineNet'07* (pp. 7–12). New York, NY: ACM.
24. Li, Z., Goyal, A., Chen, Y., & Paxson, V. (2009). Automating analysis of large-scale botnet probing events. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security, ASIACCS'09* (pp. 11–22). New York, NY: ACM.
25. Yegneswaran, V., Barford, P., & Paxson, V. (2005). Using honeynets for internet situational awareness. In *Proceedings of ACM Hotnets IV*.
26. Dainotti, A., King, A., Claffy, K., Papale, F., & Pescapé, A. (2014). Analysis of a "/0" Stealth Scan from a Botnet. *IEEE/ACM Transactions on Networking, 23*, 341–354.
27. Internet Census 2012-Port scanning /0 using insecure embedded devices, http://tinyurl.com/c8af8lt.
28. Benoit, D., Trudel, A. (2007). World's first web census. *International Journal of Web Information Systems, 3*(4), 378.
29. Heidemann, J., Pradkin, Y., Govindan, R., Papadopoulos, C., Bartlett, G., & Bannister, J. (2008). Census and survey of the visible internet. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement, IMC'08* (pp. 169–182). New York, NY: ACM.
30. Pryadkin, Y., Lindell, R., Bannister, J., & Govindan, R. (2004). *An empirical evaluation of ip address space occupancy*. USC/ISI Technical Report ISI-TR, 598.
31. Cui, A., & Stolfo, S. J. (2010). A quantitative analysis of the insecurity of embedded network devices: Results of a wide-area scan. In *Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC'10* (pp. 97–106). New York, NY: ACM.
32. Leonard, D., & Loguinov, D. (2010). Demystifying service discovery: Implementing an internet-wide scanner. In *The 10th ACM SIGCOMM Conference on Internet Measurement*. New York, NY: ACM.
33. Gu, G., Porras, P., Yegneswaran, V., Fong, M., & Lee, W. (2007). Bothunter: Detecting malware infection through ids-driven dialog correlation. In *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium, SS'07* (pp. 12:1–12:16). Berkeley, CA: USENIX Association.
34. Goebel, J., & Holz, T. (2007). Rishi: Identify bot contaminated hosts by irc nickname evaluation. In *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets (USENIX HotBots)*, Cambridge, MA (pp. 8–8).
35. Wurzinger, P., Bilge, L., Holz, T., Goebel, J., Kruegel, C., & Kirda, E. (2009). Automatically generating models for botnet detection. In M. Backes, & P. Ning, (Eds.), *Computer security – ESORICS 2009*. *Lecture notes in computer science* (Vol. 5789, pp. 232–249). Berlin: Springer.
36. Tegeler, F., Fu, X., Vigna, G., & Kruegel, C. (2012). Botfinder: Finding bots in network traffic without deep packet inspection. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, CoNEXT'12* (pp. 349–360). New York, NY: ACM.