Darknet as a Source of Cyber Intelligence: Survey, Taxonomy and Characterization

Article *in* IEEE Communications Surveys & Tutorials · January 2015

DOI: 10.1109/COMST.2015.2497690

2 authors:

Claude Fachkha
New York University
13 PUBLICATIONS **125** CITATIONS
SEE PROFILE

Mourad Debbabi
Concordia University Montreal
323 PUBLICATIONS **2,734** CITATIONS
SEE PROFILE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Darknet as a Source of Cyber Intelligence: Survey, Taxonomy and Characterization

Claude Fachkha and Mourad Debbabi
Concordia University & NCFTA Canada
Montreal, Quebec, Canada
c_fachkh@encs.concordia.ca, debbabi@encs.concordia.ca

***Abstract*—Today, the Internet security community is largely emphasizing on cyberspace monitoring for the purpose of generating cyber intelligence. In this paper, we present a survey on darknet. The latter is an effective approach to observe Internet activities and cyber attacks via passive monitoring. We primarily define and characterize darknet and indicate its alternative names. We further list other trap-based monitoring systems and compare them to darknet. Moreover, in order to provide realistic measures and analysis of darknet information, we report case studies, namely, Conficker worm in 2008 and 2009, Sality SIP scan botnet in 2011 and the largest amplification attack in 2014. Finally, we provide a taxonomy in relation to darknet technologies and identify research gaps that are related to three main darknet categories: deployment, traffic analysis, and visualization. Darknet projects are found to monitor various cyber threat activities and are distributed in one third of the global Internet. We further identify that Honeyd is probably the most practical tool to implement darknet sensors, and future deployment of darknet will include mobile-based VOIP technology. In addition, as far as darknet analysis is considered, computer worms and scanning activities are found to be the most common threats that can be investigated throughout darknet; Code Red and Slammer/Sapphire are the most analyzed worms. Furthermore, our study uncovers various lacks in darknet research. For instance, less than 1% of the contributions tackled Distributed Reflection Denial of Service (DRDoS) amplification investigations and at most 2% of research works pinpointed spoofing activities. Last but not least, our survey identifies specific darknet areas, such as IPv6 darknet, event monitoring and game engine visualization methods, that require a significantly greater amount of attention from the research community.**

***Index Terms*—Cyber, Darknet, Threats, Security, Intelligence, Cyber Attacks, Distributed Denial of Service (DDoS), Distributed Reflection Denial of Service (DRDoS), Botnet, Worms, Probing.**

I. Introduction

The Internet has become highly integrated into our current everyday lives. Society has gone digital through communication, socializing, learning and doing business online. Furthermore, in order to improve the quality and efficiency of work, governments are leveraging the Internet to operate their critical infrastructure. While cyberspace provides major benefits, our increasing reliance on it is producing new, significant vulnerabilities. As such, any security breach has the potential to result in debilitating effects on security, economy, public health, or safety. Furthermore, cyber attacks are dramatically increasing in size and number. Recent online threats demonstrated that organizations and governmental agencies could be subjected, nearly instantaneously and in full anonymity, to large-scale disrupting and orchestrated attacks with the potential to lead to severe security, privacy, and economic consequences (e.g., cyber-terrorism, DDoS, DRDoS, information theft, spam, fraud, child exploitation, etc.) [1]. For instance, a nuclear power plant in Iran was targeted for the first time by Stuxnet [2], a complex malware discovered in 2010. In 2012, a sophisticated malware known as Flame [3] was discovered with massive espionage capabilities. In addition, in recent years, there has been an increasing trend of DDoS attacks, specifically DRDoS, which have been used to exhaust and deny services of large organizations through the flood of amplified network traffic to the targeted victim. For example, in 2014, the Internet experienced the largest DRDoS attack in history which peaked at an uncontrollable rate of 400 Gigabits per second [4]. In addition, orchestrated cyber campaigns, which occurs when a given cyber force conducts a series of planned and coordinated cyber attack(s), leverage botnet (networks of orchestrated and infected computers) to communicate and execute attacks. Such threats have caused over $110 billion in losses worldwide [5]. These events constitute a serious threat with the potential to endanger human lives. Moreover, the existence of widely available encryption and anonymity techniques and tools greatly increases the difficulty of surveillance, investigation and attribution of cyber attacks. In such a context, the availability of relevant cyber intelligence is of paramount importance.

One of the effective ways to watch Internet activity is to employ passive monitoring using sensors or traps such as darknet [6, 7]. Darknet data is defined as traffic targeting advertised, but unused, IP addresses. Since these network addresses are unused, they represent new hosts that have never been communicating with other devices, neither for benign or legitimate communication. As such, any observed traffic destined to these non-interactive hosts raises suspicion and hence necessitates investigation. These darknet-based monitoring systems are designed through these unused IP address to attract or trap attackers for intelligence gathering. For instance, darknet has been used in the past to extract insights on: 1) probes or scanning activities [8] due to worms, bots and other automated exploit tools; 2) DDoS attacks due to victims' reply (backscatter) packets to spoofed IP addresses [9]; and 3) other activities, such as misconfiguration [10], and political events [11]. Darknet is an asset for network security. Several deployment techniques [12] were invented, various projects (e.g., CAIDA [13]) were built, and numerous visualization

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

techniques were used in order to observe the data. As a result, this topic has grown to include various terminologies, concepts, tools and techniques.

Motivated by the above, we first provide background information on darknet research, including definitions, alternative names, sensor types and extracted threats. We further characterize significant volumes of real-life darknet traffic for profiling purposes in terms of protocols, applications, and threats. In addition, in order to provide realistic measures and analysis of darknet traffic, we report case studies based on real darknet data analysis: 1) Conficker worm in 2008 and 2009; 2) Sality SIP scan botnet in 2011; and 3) the largest DRDoS attack in 2014. Second, we produce a taxonomy based on three major categories: deployment, traffic analysis and visualization of data. Furthermore, we explore the evolution of this network traffic monitoring technique, which began in 1992 and demonstrate the trends in this area over the past thirteen years. Our goal is to summarize, categorize, compare, and highlight overlaps in darknet research. Moreover, the aim is to identify research gaps. We can thus provide with recommendations and mitigating techniques to cope with cyber events. Finally, we discuss our findings, identify research gaps and consider future work in darknet research.

The main contributions of this paper are to:

- Provide a survey on passive monitoring systems by studying their roots and principal components, deployments, projects, tools and techniques since the year 2000.
- Interpret large volumes of real darknet data to provide insights about the nature of its traffic in terms of protocols, applications, threats, etc.
- Provide case studies on large-scale cyber events based on the analysis of real darknet traffic from numerous countries/sources.
- Propose a taxonomy that classifies darknet research proposals and highlights the overlap among them.
- Identify research gaps and provide discussions and directions for future work.

The remainder of this paper is organized as follows: Section II provides an overview of darknet and highlights the differences between darknet and other trap-based monitoring systems. Furthermore, it shows real darknet operation scenarios and characterization of its data. Section III provides darknet measurements and real case studies on worms, scanning botnet, DDoS and DRDoS cyber threats through the analysis of large volumes of real darknet traffic emanating from numerous organizations and sources. Sections IV, V and VI present our taxonomy, which is distributed over darknet deployment, traffic analysis and visualization techniques. Section VII yields a discussion on several topics. Section VIII presents the related work in this area. Finally, Section IX concludes our work with a discussion on trends and future direction in darknet research.

II. Background

This section provides an overview of darknet and highlights the focus of our survey by: (1) providing definitions that list the alternative names; (2) discussing the differences between darknet and other trap-based monitoring systems; (3) providing some examples of darknet and its operation on the Internet; and finally (4) describing the darknet taxonomy and its organization.

A. Darknet Definitions

The term darknet can refer to the following:

- Any communication system that operates by stealth and conceals its users' identity. Freenet [14] and BitTorrent [15] software are two examples that fit in this category.
- Servers and programs used to illegally distribute copyrighted material [16]. Such systems can include peer-to-peer file sharing technologies such as Napster and Gnutella [17].
- Servers configured to trap adversaries and collect suspicious data. This type of darknet runs in passive mode without interacting with attackers. This is similar to the darknet project of Team Cymru [18].

In this work, we refer to darknet as per the last definition. Since these servers run in passive mode and correspond to unused hosts or devices, any observed traffic destined to them raises suspicion and hence necessitates investigation.

It is noteworthy to mention that the word darknet has been known under various alternative terms, including darkspace, blackhole monitors, unused IP addresses, network telescopes, unsolicited network traffic, unwanted traffic, non-productive or non-responsive traffic, spurious traffic, Internet background radiation (IBR), unallocated but reachable IP addresses and unassigned IP addresses. To harmonize the terminology, we use the word darknet throughout this paper.

B. Trap-Based Monitoring Systems

Trap-based monitoring systems aim to deploy online sensors to trick and trap adversaries to collect malicious activities. Several systems leverage this approach such as darknet [19] and greynet [20]. A thin line separates various forms of trap-based network monitoring systems. In this subsection, several monitoring systems are contrasted and classified based on their types, interactivity levels, complexity, data collection and security aspects.

- *Darknet*: An IP address block configured in passive mode. Most of the darknet sensors return "unreachable" errors when a request is sent to listening hosts. This error explains that a certain host or port is not reachable. Darknet implementation is considered simple since these sensors do not communicate with the initiator of the communication. The captured traffic therefore consists mostly of the first request in communication.
- *IP Gray Space*: These addresses refer to devices that are not assigned to any host throughout a given time period (e.g., 1 hour, 1 day). Conceptually, IP gray space is similar to darknet; the only difference is that IP gray space addresses are unused only for a limited time, whereas darknet addresses are permanently unused. Unlike darknet, IP gray space might prove more difficult

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Monitoring System	Type	Interactivity	Complexity	Data Collection	Security
Darknet	passive IP	null	low	low	secure
IP Gray Space	temporarily passive IP	null	low	low	secure
Low-Interactive Honeypot	active IP	low	low	low	vulnerable
Medium-Interactive Honeypot	active IP	medium	medium	medium	vulnerable
High-Interactive Honeypot	active IP	high	high	high	vulnerable
Greynet	active/passive	low/medium/high	low/medium/high	low/medium/high	secure-vulnerable

TABLE I: Trap-based Monitoring Systems - Comparison

to be detected by an attacker since the underlying hosts might be active and operating as a regular machine during various periods of time. The aim is to imitate regular hosts.

- *Honeypot*: This is an interactive computer system, mostly connected to the Internet, that is configured to trap attackers. Honeypots are similar in nature to darknet but with more specific goals. Honeypots require more resources than darknet, since they interact during communication. As far as interaction is concerned, there are three major types of honeypots, namely, low, medium and highly interactive. A low-interactive honeypot is configured to interact with the initiator of the communication by emulating basic services such as replying to Internet Control Message Protocol (ICMP) ECHO request (e.g., Ping). A medium-interactive honeypot is similar to a low-interactive one but with further interactions and a greater number of emulated services for more data capturing and analysis. A highly interactive honeypot is a computer system that does not emulate services; it instead runs a fully-fledged, potentially vulnerable, operating systems, services and applications.
- *Honeynet*: This network is simply a group of honeypots used to deploy distributed trap-based network monitoring systems for large-scale data collection and analysis.
- *Greynet*: This network is populated with active IP addresses interspersed with darknet addresses. In other words, greynet uses both darknet (passive) and honeypots (active) in the same monitoring IP address space. The purpose is to make the monitored IP space appear as a more attractive trap for adversaries. Take for example, a range of IPs that have both darknet and other active sensors running fake services. This scenario imitates a typical organization network that hosts both running and unused IP addresses, which may trick the attacker into thinking that the whole range of IPs in the monitored block is an appropriate target.

Table I provides a comparison of trap-based network monitoring systems based on several features: type of sensor, interactivity with the initiator, deployment complexity, data collection, and security of the monitoring IP address space. First, as mentioned earlier, darknet and IP gray space share similar features. These two trap-based monitoring systems are considered secure since they do not interact with the adversary. Furthermore, since they run in passive mode (null interactivity), their deployment difficulty and data gathering features are considered low compared to other monitoring systems. Second, regarding honeypots, the interactivity, the complexity and the data gathering features are mostly proportional to each other. For instance, the more interaction there is with the adversary, the more complex the implementation to setup and the greater is the amount of data that needs to be collected. However, all honeypots with an interactive feature are potentially vulnerable in terms of security. Finally, since greynet consists of darknet and honeypots, it is considered a more comprehensive monitoring system and could therefore have more possibilities in terms of interactivity, complexity, data collection and security.

Figure 1 provides a graphical comparison between these major trap-based monitoring systems.

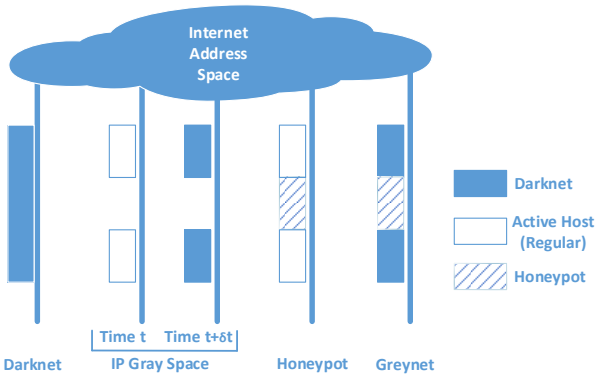


Fig. 1: Trap-based Monitoring Systems - Address Space Distribution

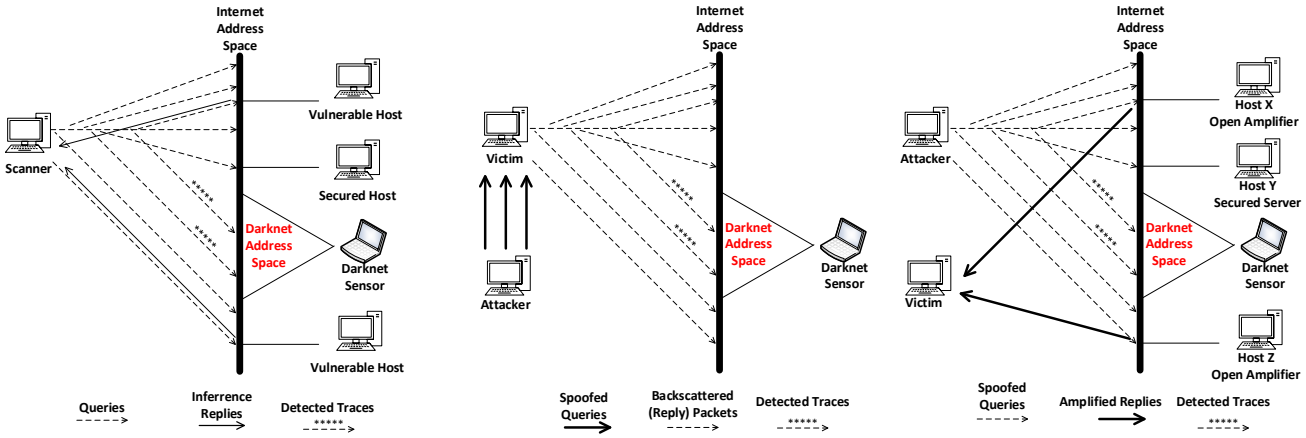
First, the darknet IP address space contains only unused addresses running in passive (inactive) mode. Second, the IP gray space (at time $t + \delta t$) is similar to darknet. However, the same address space was already active in a previous period of time (time t). Third, honeypots can run in various modes, either solely on a network or with other active/passive hosts. The latter case represents the greynet address distribution.

It is worthy to mention that some monitoring systems have the capability to run in both darknet and honeypot mode, a feature, which allows honeypots to capture more data. Given that our aim is to investigate passive monitoring of unused IP addresses, the scope of our work covers the study of mainly darknet, greynet, IP gray space and few honeypots that solely target unused address space, such as Honeyd [21] and LaBrea Tarpit [22].

C. Darknet Operation

In this section, we provide some brief background information related to some darknet scenarios. In particular, we show how darknet can be exploited on the Internet to generate various elements of cyber threat intelligence, including probing, DDoS and DRDoS activities.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials



(a) Probing Activities (b) DDoS Activities (c) DRDoS Activities

Fig. 2: Cyber threat intelligence by leveraging darknet

A darknet is indeed an effective approach to infer various Internet-scale probing activities [23]. Figure 2a presents an illustrative example in which a probing machine is scanning the Internet. Such machine could have been previously infected by a worm that is trying to propagate, or perhaps is participating in automated Internet-scale scanning [24]. Some of these network probing packets can hit the network telescope and thus are subsequently captured. Recall, that the probing machine, while spraying its probes across the Internet, cannot probably avoid the network telescope as it does have any knowledge about its existence. Further, it has been shown in [25] that it is extremely rare if not impossible for a probing source to have any capability dedicated to such avoidance.

Darknet traffic analysis is as well effective technique in pinpointing victims of DDoS attacks [26]. Figure 2b illustrates such scenario. The attacker is directed to launch a DDoS attack towards the victim. To hide its identity, the attacker spoofs its address and replaces it with a random IP address. Such random address could happen to be that of the darknet. When the attack is launched, the reply packets from the victim will be directed towards some dark IP address. Traces that hit the darknet are often dubbed as backscattered packets [26] and could be effectively employed to infer that the victim has been the target of a DDoS attack.

In the last scenario, a darknet is leveraged to infer DRDoS attacks [27]. Indeed, as previously mentioned, such attacks are an emerging form of DDoS attacks that rely on the use of publicly accessible UDP servers [28][1], which act as "open amplifiers" of the attack. The bandwidth amplification factors are function of the instrumented protocol. The idea is to send small queries to such amplifiers in which the replies, that aim at flooding the victim, are orders of magnitude larger. Recall that such approaches are behind the notorious 300 and 400 Gbps attacks that hit the Internet in the last couple years [28]. Figure 2c depicts this scenario. Commonly, the attacker will spray the Internet with such spoofed queries in a hope to reach as many open amplifiers as possible in order to achieve a large amplification factor. This case will occur in the scenario where attackers do not know in advance the IP addresses of Internet open amplifiers. We argue that such generated requests are not probes intended to gather information, since the attackers in this case do not aim to build/manage a list of open amplifiers nor does they want to jeopardize being detected (by using their real IP addresses, which is typical in probing activities). Intuitively, some of those requests will hit the darknet and hence will be captured. Requests that actually reach those servers will be amplified and directed towards the victim.

[1]Although TCP amplifiers could be vulnerable to such abuse [29]

D. Darknet Taxonomy

Despite the fact that the idea of monitoring unused IP addresses started in the early 90's [8, 30], this survey mainly focuses on the study of darknet research during the past thirteen years. The reason behind choosing this period is that the major contributions started after 2001.

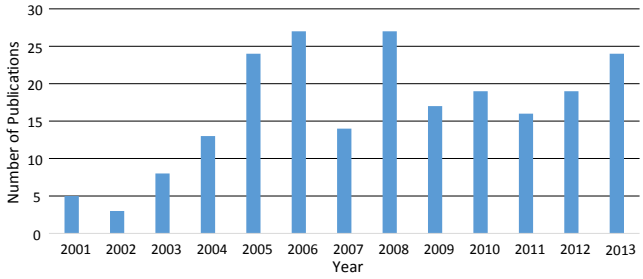


Fig. 3: Trend of Publications Per Year

Figure 3 represents the trend of the darknet research from 2001 to 2013 in terms of research publications. Some of the important contributions include the discovery of the relationship between backscatter traffic and DDoS attacks, which emerged in 2001 [31], the trend of worms propagation and analysis between 2003 and 2005 [32–34], the use of time series and data mining techniques on darknet traffic raised in 2008 [35], and finally the monitoring of large-scale cyber events [11], which began in the past few years.

Our taxonomy classifies current darknet research into three major areas, namely, darknet deployment and setup, analysis

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

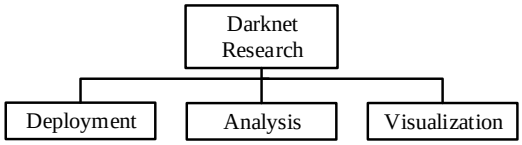


Fig. 4: Darknet Research Taxonomy

and measurement of darknet data through deployed sensors, and tools and techniques for the visualization and representation of its traffic. A high level overview of the proposed taxonomy is shown in Figure 4.

E. Related Work

As previously discussed, a thin line distinguishes darknet from other trap-based monitoring systems such as IP gray space [36, 37], greynet [20], honeytokens [38] and darkports [39]. However, two main groups of surveys can be related to our work. The first group focused solely on specific technology or threat whereas the second elaborated on trap-based monitoring systems.

First, various surveys tackled the detection techniques in network traffic such as NIDS [40], threats such as DDoS [41], botnet [42, 43] and worms [44], and malicious activities such as scanning [45]. Compared to our work, this group of research focused on a specific technology or a threat only whereas ours was more comprehensive. For instance, our survey included not only a study on DDoS threats, but also provided an overview on several darknet topics that can be leveraged to infer various insights from the Internet, including threats, events, techniques and tools.

Second, in regard to surveys that tackled trap-based monitoring systems, Zhang et al. [46] were among the first to classify honeypots in 2003. They highlighted data capture and data control in honeypots. Furthermore, they provided a classification of these traps based on security and application purposes. Furthermore, Seifert et al. [47] presented a taxonomy of honeypots. The authors described a classification of honeypots based on several schemes and were able to distinguish between seven types of honeypots (e.g., low and high interactive). In 2012, Bringer et al. [48] divided honeypot research into 5 major areas: types of honeypots, analysis of data, configuration, detection of sensors, and legal and ethical issues. The main difference between the works in [46–48] and ours is the scope of the survey. This group of works focused on honeypots, including active monitoring. Complementary, our work focused solely on passive monitoring of unused IP addresses. The only work that touched darknet research is [47] by discussing darknet and comparing it to other monitoring systems (low and high interactive honeypots). Our work is more comprehensive in regard to darknet study as it covers development, data analysis, and visualization.

Therefore, our survey is more close to the second group of contributions which tackled trap-based monitoring systems. Our survey complements the aforementioned related research works. Furthermore, the realistic analysis and investigation of real data provides more understanding and hands-on investigation experience on darknet data and threat analysis. We provided a guideline to develop, analyze and visualize real cyber insights by leveraging darknet data. The extracted darknet knowledge in our work can help in building a cyber intelligence platform for Internet monitoring. We are not aware of any similar contribution.

III. Darknet Measurements

Before we start our darknet taxonomy, and in order to better understand the nature of darknet data, in this section, we primarily provide an overview of darknet traffic and insights on large volumes of darknet traffic emanating from numerous organization. As such, we describe darknet traffic and characterize its traffic. Second, we discuss three case studies related to separate events, namely, probing, botnet and DRDoS activities. Our dataset is collected from several sources such as CAIDA[2] and DShield[3].

A. Inside Darknet

Generally, darknet data is composed of scanning, backscatter, and misconfiguration traffic. Network scanning forms the majority of darknet data. Scanning activities are the result of reconnaissance and DRDoS activities. Generally, attackers scan the Internet to identify vulnerabilities and running services with an intent to compromise them [8, 30]. Furthermore, researchers are recently leveraging scanning activities to darknet to infer DRDoS activities [27]. In a typical DRDoS scenario, adversaries attempt to abuse Internet services to generate a flood of amplified traffic to the victim [49, 50]. Furthermore, darknet contains backscatter traffic, which commonly refers to reply packets in a network protocol (e.g., SYN-ACK, ACK). Backscatter packets are the result of replies to an attack with spoofed source of IP address. Backscatter data is used to infer activities of DDoS victims. In a typical DDoS scenario, attackers spoof source IP addresses to execute cyber attacks. Subsequently, victims reply to the flood by sending backscatter packets to the spoofed addresses. Darknet consists of unused IP addresses that can be randomly used by attackers for spoofing purposes. Such backscatter traffic can be leveraged to infer the IPs of DDoS victims as reported in [9]. The rest of traffic packets is labeled as misconfiguration.

To understand the nature of darknet data, we provide an overview of its traffic. The dataset is pure darknet data captured during a five-year period from a single unused /8 network address block [51].

Count	TCP	UDP	ICMP
Packet	76.6%	19.9%	2.8%
Bytes	55.82%	40.82%	2.66%

TABLE II: Protocols Distribution - Inspired by [51]

Table II lists the distribution of darknet transport and network layer protocols. It is shown that the majority of darknet

[2]CAIDA Dataset: http://www.caida.org/data/
[3]DShield: https://www.dshield.org/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

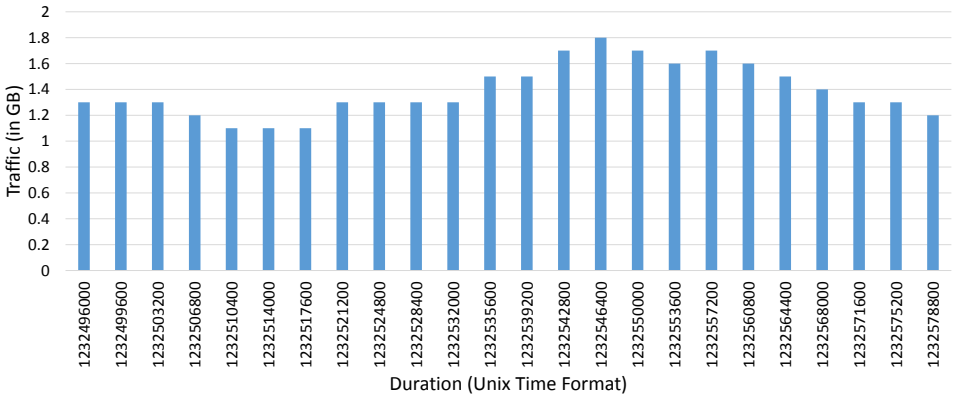


Fig. 5: Conficker Worm in 2009 - Traffic Distribution (1 hour interval)

traffic consists of TCP packets. Several facts can explain the TCP dominance. First, TCP provides various scanning techniques (e.g., SYN, Fragmentation, SYN-ACK) [52]. Second, generating TCP scanning is generally more feasible than UDP [53]. Finally, as noted in [54], well-known cyber attacks are specifically targeting TCP services.

We further list top application protocols found on darknet.

Port	Service
445	microsoft-ds
139	NetBIOS
4662	eDonkey
80	HTTP
135	Endpoint Mapper

TABLE III: Top TCP-based Services

Table III depicts the top 5 TCP-based services that have been observed based on [51]. The results demonstrate that the Microsoft Directory Service (microsoft-ds) is leading while the NetBIOS is ranked second. The former service is known to be abused by malware such as Conficker worm [55]. More information on the Conficker worm can be found next.

B. Case Studies

Base on real darknet data analysis, we provide three case studies on separate events, namely, Conficker worm in 2008 and 2009, Sality SIP scan botnet in 2011, and the largest DRDoS attack in 2014.

1) Case Study 1 - Conficker Worm in 2008 and 2009: In 2008, a new exploit targeted Windows services. As such, Microsoft announced a security update (MS08-067) to resolve the issue. The threat originated from a malicious TCP scanning behavior by a worm named Conficker. The latter is a malware designed to exploit victim machines by exploiting TCP port 445 (Microsoft Directory Services). Conficker infected millions of computers in over 200 countries, which render it one of the largest known computer worms. In this case study, we show the outcome of the darknet analysis that inferred random scans generated by this worm. The dataset of the attack in January 2009 is shown in Figure 5.

Several versions of Conficker (A and B) were involved in the attack. It is noteworthy to mention that the figures depicts the peak at 2 pm in the analyzed 2009 dataset. However, based on the analysis done by other researchers, the attack also peaked at 2 pm during the 2008 dataset [56]. This confirms the orchestration and automation of the machinery behind the attack, which is shown as a diurnal pattern in [56, 57].

2) Case Study 2 - Sality Botnet SIP Scan in 2011: In February 2011, the Sality botnet executed a $/0$ SIP scan through the whole IPv4 address space. This 12-day event involved 3 million unique IP addresses in one of the most coordinated cyber scanning campaigns ever. The botnet generated 20 million scans to 14.5 million addresses, which is almost 86.6% of the whole $/8$ monitors. This campaign targeted SIP services, which run on port 5060 and threatened the voice communications infrastructure. The darknet observation of this event is depicted in Figure 6. The campaign initiated in January and ended in February, 2011. The attack peaked at 21,000 hosts within a 5 minute interval. More on this attack can be found in [58].

3) Case Study 3 - The Largest DRDoS Attacks in 2014: In 2013, a 300 Gbps DRDoS attack targeted Spamhaus [59]. In February 2014, the largest DRDoS attack in history, which peaked at 400 Gbps of bandwidth, hit the Internet infrastructure. We have depicted the latter attack through the DShield data in Figure 7. This image shows the source distribution of UDP-based packets on port 123-NTP (in yellow) with the corresponding generated reports (in blue). The graph depicts the increase in NTP packets and reports during the attack.

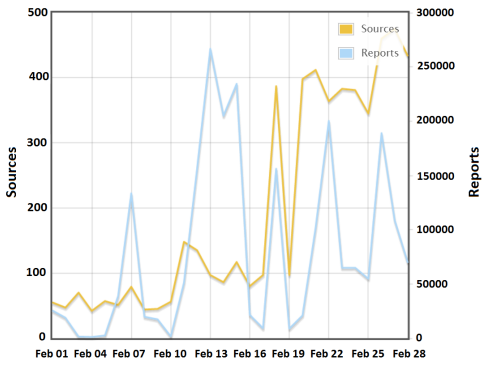


Fig. 7: The Largest NTP-based DRDoS Attack in History

Typically, in an NTP amplification attack, the adversary generates a flood of spoofed UDP network packets. This large

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

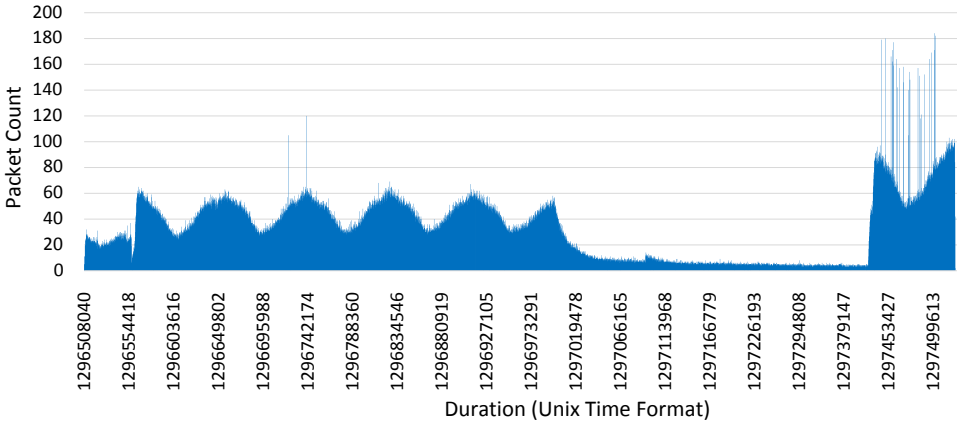


Fig. 6: Sality Botnet SIP Scan in 2011 - Traffic Distribution (12 days)

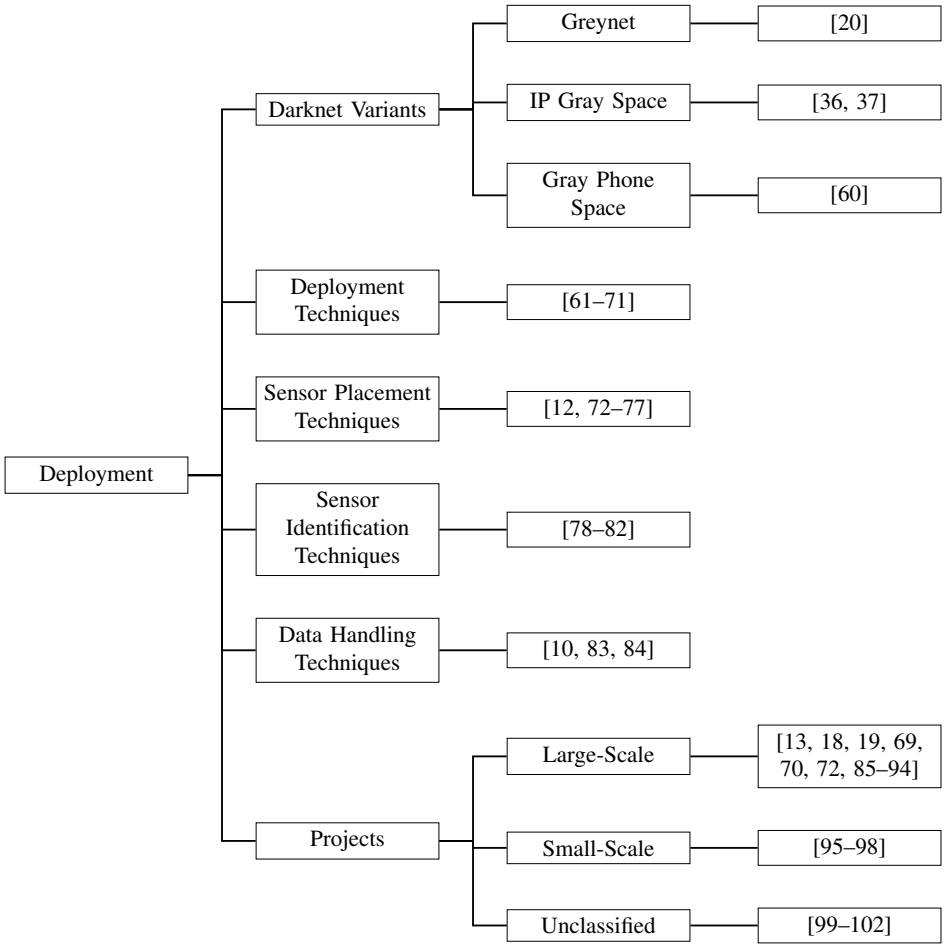


Fig. 8: Deployment Research Taxonomy - Overview

amount of traffic is sent to open Network Time Protocol servers, which operates at port number 123. This attack abuse the MONLIST service in NTP with an aim to send amplified traffic to the victim. More on NTP amplification DRDoS attacks in the context of darknet can be found in [103]. Similar to NTP amplification DRDoS attack, DNS service can also be abused to generate amplification/DRDoS attack. More on DNS amplification and DRDoS activities through darknet analysis can be found in [27, 104].

The aforementioned darknet measurements and case studies provide a basic understanding on what is darknet and how it is leveraged to generate cyber intelligence. Next, as depicted in Figure 4, we start a taxonomy of darknet research based on three categories, namely, development, analysis and visualization.

IV. Darknet Deployment

The first step in darknet monitoring is the deployment of sensors, which aims to capture network traffic. This exercise

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

requires an understanding of the network architecture and a careful configuration of the dynamic host server or the upstream router to forward unreachable packets to darknet sensors. A basic darknet deployment architecture is shown in Figure 9.

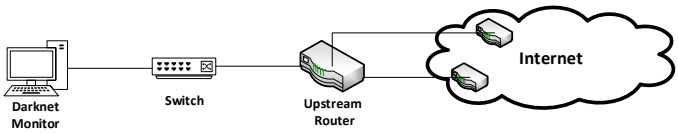


Fig. 9: Basic Darknet Deployment - Inspired by [10]

This section provides insights on the elements of darknet deployment, namely, darknet variants, as well as techniques such sensor placement/identification and data handling, and projects. Figure 8 provides a taxonomy of deployment research works based on the aforementioned elements.

A. Darknet Variants

Recalling Section II-B, darknet variants are the deployment mechanisms of trap-based monitoring systems using techniques similar to those of a darknet. This part thus includes deployment of IP gray address space and greynet monitors. Table IV summarizes the papers on darknet variants. For instance, Harrop et al. [20] define and assess the concept of a greynet, a network address space that is populated with darknet addresses mixed with active IP addresses. Using data collected from a university network, the authors evaluate their concept and show how small number of dark IP addresses can increase the efficiency of network scanning detection. Furthermore, Jin et al. [36, 37] are among the pioneers to use IP gray space in passive monitoring. This work applies a heuristic algorithm to identify IP gray space addresses. The authors investigate the behavior of such traffic. This study tackles patterns such as dominant and random behaviors. The approach identifies the usefulness of IP gray space to uncover insights on the behavior of malicious activities as well as their intentions. The result identified several malicious activities such as scanning, worm propagation as well as spam. Finally, in a unique work, Jiang et al. [60] investigate passive monitoring in mobile communication. This work presents a novel approach to detect SMS spammers on a cellular network. The approach is based on greystar technology and employs a statistical model to infer spam size through fingerprinting. The proposed approach also has the capability to reduce the spam traffic by 75% during peak periods. The authors analyzed five months of SMS data from large cellular networks and inferred thousands of unreported spam activities.

B. Deployment Techniques

In this section, we discuss the research works that mainly target the techniques of deploying passive monitoring systems. Table V summarizes these contributions. In this category, research works are mainly leveraging Intrusion Detection Systems (IDS) and hybrid techniques.

Publications	Approach/Technique	Tool/Project
[61]	IDS	honeyd/Dshield/DOMINO
[62]	Sink - IDS	iSink
[63]	Sink - IDS	honeyd/iSink
[64]	Statistics - IDS	Custom
[65]	Mobile-Based AS Data	honeyd/Mohonk
[66]	Hybrid	honeyd
[10]	Hybrid	IMS
[71]	Hybrid	Custom
[68]	Comparative Study	honeyd
[67]	Hybrid	IMS

TABLE V: Deployment Techniques Research Papers - Summary

For instance, Yegneswaran et al. [61] introduce a scalable, heterogeneous, and robust Distributed Overlay for Monitoring InterNet Outbreaks (DOMINO). The proposed approach provides an architecture for collaboration of distributed IDS data on different nodes on an overlay network. In an overlay design, a network is built on the top of another one. One of DOMINO's components is the use of active nodes, which measure connections targeting unused IP addresses. The authors emphasize the importance of the approach in detecting sources of IP spoofing, classifying cyber attacks, generating updated blacklists and reducing false positives. Moreover, Yegneswaran et al. [62] introduce iSink and elaborate on a darknet case study to analyze attack traces. The study is composed of various components such as the analysis of backscatter packets and the investigation of unique periodic probes. iSink deployment proved its relevancy in detecting worms such as Sasser. Through iSink, the authors managed to observe different worm variants and malware. Furthermore, Yegneswaran et al. [63] explore ways to integrate trap-based monitoring information, including darknet data, into daily network security monitoring with the goal of sufficiently classifying and summarizing the data to provide ongoing situational awareness. To this end, the authors develop a system based on honeynets, analyzers that leverage Network-based Intrusion Detection System (NIDS), and a back-end database to facilitate the analysis of honeynet data. The system is able to capture and identify numerous malicious activities including botnet and worms. Furthermore, Choi et al. [64] propose a framework to monitor and respond to security events. The approach aims to trace potential attackers using darknet. The approach was evaluated using a $/24$ darknet IP address block and other alert logs. Several attack trends and patterns were identified. In addition, the approach showed capabilities to detect zero-day attacks. Furthermore, Krishnamurthy et al. [65] propose a mobile darknet-based mechanism that allows unwanted traffic to be detected significantly closer to the origin source of attack. The scheme is based on two pieces of information: the additional data that is made accessible to the upstream autonomous systems (AS) and the changes in the advertised dark address space set. Such shared data among ASes can identify and minimize unwanted traffic between these entities. Furthermore, Bailey et al. [66] propose a hybrid monitoring architecture that uses low-interaction honeypots (honeyd) as front-end filters and high-interaction honeypots as a back-end for further investigation. In order to reduce loads on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Publications	Approach/Technique	Contribution	Tool/Project
[20]	Defining and Characterizing	Greynet Development	Custom
[36]	Heuristic Algorithm	IP Gray Space Development	Custom
[37]	Heuristic Algorithm	IP Gray Space Development	Custom
[60]	Statistics	Gray Phone Space Development	Greystar

TABLE IV: Darknet Variants Research Papers - Summary

back-ends, a filtering mechanism is used coupled with a novel hand-off mechanism. The authors use five months of data to demonstrate the efficiency, scalability and robustness of their work. Pouget et al. [68] provide a thorough comparison between honeypots based on their level of interaction. Using honeyd as a low-interactive honeypot, this qualitative and quantitative comparison uncover interesting classification and correlation among detected threats. Bailey et al. [67] examine the singular and distributed passive monitoring sensors to effectively build a scalable hybrid monitoring system. The authors demonstrated that the majority of the threats coming to darknet were based on a limited number of source hosts, and proposed a new source-distribution approach to reduce the number of events found on darknet. The analysis listed several threats including worms and scanning activities. In addition, Bailey et al. [10] discuss the major elements of darknet deployment setup, namely, the storage and network requirements and the deployment techniques. They further review the methods to collect darknet data and list the most suitable formats. They propose three major darknet deployment approaches to build darknet sensors. Finally, Komisarczuk et al. [71] discuss the opportunities and research direction in the Internet sensor grid for detecting and analyzing malicious behaviors online. The authors review the developments of monitoring sensors in active and passive modes. They further share their experiences in sensor deployments.

C. Sensor Placement Techniques

This category includes techniques that are used to improve sensor placement and setup. Table VI lists the relevant research works.

Publications	Approach/Technique	Tool/Project
[72]	Hybrid	IMS
[73]	Comparative Study	honeyd/Leurre.com
[76]	Multiscale Density Estimation	iSink/Dshield
[75]	Comparative Study	Dshield
[74]	Empirical Analysis	Netflow
[77]	Sampling	Custom
[12]	Comparative Study	IMS

TABLE VI: Sensor Placement Research Papers - Summary

Several techniques have been used to improve darknet monitors placements. For example, Cooke et al. [12] examine variations observed on different network blocks. The authors showed evidence that distributed address blocks exhibit significant changes in traffic patterns. They further demonstrated changes over protocols (services) and specific worm signatures. Moreover, Bailey et al. [72] examine the properties of individual and distributed darknet sensors to test the effectiveness of deploying hybrid systems (darknet with honeypots). The authors used source-based techniques to reduce redundant actions generated by individual darknet and hence lowered the evaluated connections by over 90%. They also expanded source-distribution based techniques to detect a variety of global attacks. Furthermore, Chen et al. [73] demonstrate the importance of deploying multiple sensors in different locations. The authors deployed two identical sensors, having the same configurations, in two different locations and compared various parameters. While analyzing data from a six-month period, the analysis revealed different anomalies. Likewise, Berthier et al. [74] focus on the size and the location of darknet sensors to perform an empirical analysis and to increase the efficiency of darknet monitors. In addition, Abu Rajab et al. [75] quantify the importance behind the design of a distributed monitoring system and evaluate the applicability of this approach. In order to achieve their goals, the authors propose a worm propagation model to evaluate the locations of monitors, the size of the monitored IP addresses, and the impact of worm detection time. Over 1.5 billion suspicious connection attempts were observed through many detection systems across the Internet. The results showed that distributed monitoring systems were better than centralized ones. In terms of speed, a distributed monitor system was found to act four times faster than a centralized one. Furthermore, the authors mentioned that monitor placement can be improved by having partial knowledge of the vulnerable population density. In some cases, exploiting information related to vulnerable host locations can help decrease the detection time by seven times compared to random monitoring deployment. Furthermore, Barford et al. [76] present a study of source locations of hosts that send unwanted traffic through dark addresses. The researchers use a multi-scale density estimation method which allowed them to see a small number of tight clusters that are formed by darknet source addresses. The authors propose a multiplicative model for darknet host locations that can be used to generate data with the same distributed property as empirical data. Their model can be used for testing, evaluating, measuring, simulating and analyzing traffic for the purpose of reducing darknet pollution. Finally, Pemberton et al. [77] outline results from a $/16$ darknet network by experimenting with various sampling techniques and applying them to arrival density measures. The authors found that current darknet deployments using continuous lists of IP addresses were inefficient in predicting threats. They further studied three other address space allocation techniques and discovered better accuracy. The researchers claim that business users as well as Internet Service Providers (ISPs)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

can use these techniques to enhance darknet deployment in the future.

D. Sensor Identification Techniques

This part includes research targeting the identification of darknet sensors. From the adversary's point of view, the identification of monitoring sensor locations allows them to avoid detection. Table VII summarizes these related works. Abu Rajab et al. [78] highlight an evasive attack that detects passive monitoring systems such as darknet. By sampling the IP address space in a coordinated manner, the authors show that detection and evasion of monitors is possible. Using this technique, attackers can identify active hosts on the network and hence proceed with their attacks. The proposed methodology can overtake the entire vulnerable population within seconds. In a similar work, Sinha et al. [79] elaborate that monitoring sensor configurations are easy for attackers to discover. The authors discuss that manually building a monitoring system is usually a large and difficult task to handle. As such, the authors propose an automated technique for sensor configuration. They further argue that networks with consistent nodes and proportional representation are more efficient in detecting attacks and are more resistant to detection. Using random sampling and profiling, the authors propose a technique for automated configuration of sensors. More on identifying monitors' locations, Shinoda et al. [80] propose several algorithms that are designed to detect listening sensors on the Internet. Consequently, they propose an approach to enhance the sensor setup and deployment. In a similar work, Bethencourt et al. [81] demonstrate the use of probing to detect sensors' locations on systems that publicly report security results. This probe response technique, which can target darknet sensors, shows how to locate monitors. With limited capabilities, the simulation results of this technique illustrate the power of determining the sensors' identity within one week. The authors further target the anonymized schemes used by network administrators and discuss some potential countermeasures based on the sensors' characteristics. Finally, Cooke et al. [82] propose the Dark Oracle, which is an architecture that aims to uncover dark addresses. The authors validated the effectiveness of their work, which uses internal as well as external routing and host setup information for automatic discovery. The proposed methodology uncovered almost 80,000 unique source IPs compared to 4,000 with a traditional $/24$ darknet. The authors further demonstrated the capability of the Dark Oracle by shedding light on local attacks.

E. Data handling Techniques

Deploying darknet requires handling data, which includes processing, storing and sharing its traffic. Darknet may receive a large amount of unsolicited network traffic. Processing such information at the sensor level and sharing it with investigators and researchers may therefore require several development steps. Table VIII summarizes the data handling research papers.

Publications	Approach/Technique	Tool/Project
[83]	Resource-Aware Multi-Format Data Storage	IMS
[84]	Graphical Processor	Custom

TABLE VIII: Data Handling Research Papers - Summary

In this category, Cooke et al. [83] propose a resource-aware multi-format data storage of security information with the aim to simultaneously save various security information. The proposed architecture consists of a set of algorithms for storing various formats of data. Furthermore, a darknet-based prototype is built based on numerous sources of data and the results show reasonable short- and long-term outputs. Moreover, Nottingham et al. [84] suggest graphical processors to accelerate darknet data analysis using big data. They further discuss the construction, the performance and the limitations of the packet filtering approach, which employs multi-match capabilities to differentiate between packets. The aim is to build a fully programmable virtual machine with massive parallel classification, data mining and data transformation capabilities to provide complex security filtering, indexing and manipulation functions.

F. Projects

The outcome of deploying darknet sensors is to build a functional platform, an operational project or center to monitor the cyberspace. We list below the publicly known centers and projects that use darknet as a source of their data. Table IX summarizes large-scale darknet monitoring projects. Our classification is based on three groups, namely, large-scale, small-scale and unclassified projects.

The first group in this area lists large-scale darknet projects. For instance, the network telescopes project [13, 105] is a system proposed by researchers at the Center for Applied Internet Data Analysis (CAIDA). The intent of this project is to monitor pandemic and epidemic cyber incidents through the unused address space. Moore et al. [19] propose the network telescope as an efficient and effective darknet traffic monitoring system by using sensors and virtual machines. The network telescope project can monitor large chunks of unused address space. The University of California, a main contributor to this data, deploys network telescopes to monitor a single unused Internet address space of $/8$ block. The latter represents around $\frac{1}{256}^{th}$ of the overall IPv4 address space of the Internet. Collected data includes Domain Name System (DNS) data, topology traces, round-trip time, and routing data. This passive information contains insights about large-scale security events such as malware (mostly Internet worms) and DDoS. Another project is the Active Threat Level Analysis System (ATLAS) [86], the Internet's first globally scoped threat analysis network. Under the direction of Arbor Networks, this network monitoring system collectively analyzes the data traversing disparate darknet to visualize malicious activities on the Internet. Arbor is among the unique operators positioned to provide enterprise and service provider-specific intelligence related to malicious

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Publications	Approach/Technique	Contribution	Tool/Project
[80]	Marking Algorithm	Detecting Listening Sensors Online	Custom
[81]	Probe Response Attack	Proposing a Technique to Detect Sensors	Custom
[79]	Automatic Profiling - Random Sampling	Discussing Sensor Identification and Configuration	Custom
[82]	Perspective-Aware Address Discovery	Proposing a Model to Uncover Dark Addresses	Dark Oracle
[78]	Sampling	Highlighting an Evasive Attack that Identifies Sensors	Dshield

TABLE VII: Sensor Identification Research Papers - Summary

Project	Stewardship	Description	Objectives
UCSD Network Telescope	CAIDA	Passive traffic monitoring system built on a globally routed /8 network	Monitoring of DDoS, Internet worms and viruses, scanning - Data sharing
ATLAS	Arbor	The world's first and largest globally scoped threat analysis network	Providing global threat intelligence for DDoS and advanced threats
The Darknet Project	Team Cymru	Internet security research and insight	Monitoring compromised machines from malware
IMS	University of Michigan	A distributed global Internet threat monitoring system of /8 network	Measuring, characterizing, and tracking threats
PREDICT	RTI	Protected repository for the defense of infrastructure against cyber threats	Investigating spatial and longitudinal darknet data
NICTER	NICT	A large-scale network incident analysis system	Visualizing and analyzing network attacks
WOMBAT & Leurre.com	EURECOM	Worldwide observatory of malicious behavior and attacks threat	Studying cyber attacks and threats
Internet Storm Center & DShield	SANS	A global cooperative cyber threat and Internet security monitor and alert system	Monitoring the level of malicious activity on the Internet

TABLE IX: Large-Scale Darknet Projects - Summary

activities such as exploits, phishing, malware and botnet. In addition, the Darknet Project [18] is deployed by the Team Cymru Community as a passive Internet threat monitoring system. Its main purpose is to set a platform to collect packets susceptible to be sent by malware. This darknet is deployed to host flow collectors, backscatter detectors, packet sniffers and IDSs. Team Cymru aims to increase awareness about threats and enhance mitigation against malware. In addition to monitoring darknet, the authors provide a guideline to set up a darknet. Another large-scale project is the Internet Motion Sensor (IMS) [72], a distributed globally scoped Internet threat monitoring system. The IMS project has the ability to monitor dark IP space. It uses 28 unused IP blocks, ranging in size from /25 to /8 network address blocks. The IMS is based on a distributed blackhole network with a lightweight responder, a payload signature and a caching mechanism. These capabilities are used to generate new insights about worms, DDoS, and scan activities [94]. Furthermore, the Protected Repository for the Defense of Infrastructure against Cyber Threats (PREDICT) project investigates spatial and longitudinal darknet data. The authors aim to describe some of the large-scale spatial and longitudinal darknet information. Another large-scale project is the Network Incident analysis Center for Tactical Emergency Response (NICTER) [87, 88], which is a large-scale network incident analysis system that mainly monitors darknet. It represents a system that is capable of capturing and analyzing malware executable. The identification of malware propagation is the primary purpose of this project. The NICTER project is composed of four components: macro analysis system, micro analysis system, network and malware enchaining system, and the incident handling system. Additionally, the Worldwide Observatory on Malicious Behavior and Attack Threats (WOMBAT) [69, 85] center aims at providing new artifacts to understand emerging threats. The project WOMBAT is used to collect raw data and analyze it in order to identify different threat phenomena. The authors claim that the latter can discover trends of attacks by understanding the behavior of threats. With this in mind, the designers develops mechanisms for automatically collecting and analyzing malware [106]. WOMBAT has a number of features. Its main feature is to improve data acquisition technologies. The project further shares information with its partners, including SGNET [107], *Leurre.com* [89], Argos [108], Nepenthes [109], NoAH project[97], and SANS Internet Storm Center (ISC) [93, 110] which uses the DShield as firewall [111]. Moreover, The *Leurre.com* Project [70, 90], a part of the WOMBAT project, has a purpose of collecting Internet threats using worldwide distributed sensors [89]. The terms used in the context of this project include platform architecture, logs collection, data uploading mechanism and data enrichment mechanism. Furthermore, the Billy Goat project [91] is a specialized darknet traffic monitoring system deployed by IBM and its customer networks. It is used for worm detection. This project differs from other monitoring systems as it focuses on specific attacks and dynamic characteristics of worms. By taking advantage of worm propagation strategies, Billy Goat monitors unused IP address spaces that are randomly scanned by worms. Finally, the Honeynet project [92] is a dedicated system to investigate cyber attacks and develop open source security techniques to mitigate Internet threats. This project provides tools to build darknet sensors.

The second group in this category are small-scale projects. For example, Antonatos et al. [98] propose HoneyHome, a part of the NoAH project [97], a platform for monitoring unused IP addresses and ports for large-scale security events extraction. This low-cost system is based on installing sensors on regular users to monitor these unused IP addresses and ports. Since regular users come and go, it is difficult for attackers to detect these unstable sensors. Moreover, ARAKIS [96], one of the initial data sources for WOMBAT, is developed by NASK and operated by CERT Polska. The latter project is a nationwide near real-time NIDS that generates early notifications and warnings about security events. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

system consists of a central database in addition to distributed monitors, which collect and correlate security information through low-interaction honeypots and other detection systems including darknet. Finally, Daedalus [112], which is based on the NICTER project, is designed to capture cyber attacks in near real time fashion.

Last but not least, it is worthy to mention some other unclassified projects that use passive monitoring such as SWITCH [99], the National Police Agency of Japan [100], the Internet Scan Data Acquisition System (ISDAS) runs by Japan CERT Coordination Center [101], the Research and Education Networking Information Sharing and Analysis Center (REN-ISAC) [113], the IUCC/IDC Internet Telescope [102] in Israel, the Simwood Darknet [114] and many other academic systems such as the Darknet Mesh Project [115] at Oxford University as well as Rhodes University Network Telescope [116].

Summary of Findings in Section IV:

We have discussed several key elements in the darknet deployment section, namely, architecture, darknet variants, online placement and identification of sensors, data handling, and projects. From what has been discussed in the deployment part of Section IV, we can conclude the following points:

- In order to deploy darknet, several elements must be taken into consideration, such as the study of exact storage and network requirements, the knowledge of deployment techniques, as well as sensor placement and identification.
- Compared with other trap-based monitoring systems, darknet is considered as a practical and easy to implement tool in passive monitoring the cyber space. Darknet setup can be developed using basic routing techniques and can be monitored through IDSs.
- IDSs are the most used systems in darknet development and Honeyd is probably the most practical tool to implement darknet sensors.
- One of the major challenges in deploying darknet is to avoid the adversary's discovery of the sensor location. Several techniques are used to identify the location of sensors such as the sampling of IP addresses.
- Mobile darknet is a new trend that has a promising future in passive monitoring research. The future deployment will include mobile-based VoIP darknet.
- Darknet variants are not commonly used in literature. These variants can be more efficient than darknet to monitor cyber attacks; however, their implementation could be more complex.
- Darknet projects monitor various cyber threat activities and are distributed in one third of the global Internet.
- Various types of darknet projects exist. Some large-scale projects are coupled with interactive trap-based monitors to enhance network monitoring.
- CAIDA is one of the few Internet monitoring research groups, which provides darknet-based backscatter data for researchers.
- Despite the existence of some collaborative darknet projects (e.g., PREDICT), more darknet resources and information sharing must emerge to infer and attribute large-scale cyber activities. Dealing with a worldwide darknet information exchange is a capability that requires collaboration and trust; however, this collaboration raises security policies and privacy concerns.

V. Darknet Analysis

This section provides an overview of the contributions in the area of darknet data analysis and measurement. These topics are divided into three main areas: analyzing and measuring darknet data, threats, and worldwide events. Figure 10 depicts the taxonomy of research efforts in the analysis and measurement of passive monitoring systems.

A. Data

In this section, we provide a taxonomy of the research works related to darknet data. This includes profiling darknet traffic, filtering and classification of its data as well as reviewing its backscatter and misconfiguration traffic.

1) Data Profiling: Data profiling encompasses the research works that focus on the characterization of darknet data to generate statistics and insights. Table X provides an overview of the summarized research works. These contributions leverage several techniques such as packet filtering, routing, and time series.

For instance, Irwin [117] explores data across five different darknet sensors. The author discusses the differences as well as the similarities among the analysis of the five sensors and presents two case studies related to two vulnerabilities on Microsoft Windows systems. Furthermore, Pang et al. [54] present a study of the broad characteristics of darknet. The authors develop filtering techniques and active responders to use in their monitoring process. They analyze both the characteristics of completely unsolicited traffic (passive analysis) and the details of traffic elicited by their active responses (activities analysis). Moreover, Shimoda et al. [119] propose a system to improve passive darknet monitoring. The proposed approach leverages active hosts with no effect on legitimate connections. This light-weight multi-dimensional IP/port analysis system enables TCP ports monitoring. In this context, Ford et al. [118] create the first IPv6 darknet. The aim of this work is to compare between IPv6 and IPv4 darknet. The results showed that traffic targeting IPv6 darknet is minimal. Furthermore, Dainotti et al. [120] infer the evolution of Internet infrastructure. Instead of using active probing techniques, this technique leverages darknet traffic monitoring to provide some insights on the utilization of the Internet. The investigation touches the limited visibility of a unique observation point as well as the existence of IP spoofed addresses in data that can fake analysis results. The authors propose new techniques to remove spoofed packets and compare their results with methods that use active scans. Oberheide et al. [123] introduce the concept of dark DNS, which is based on the analysis of DNS queries found on darknet addresses. They also profile the DNS dark data collected from their sensor and discuss the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

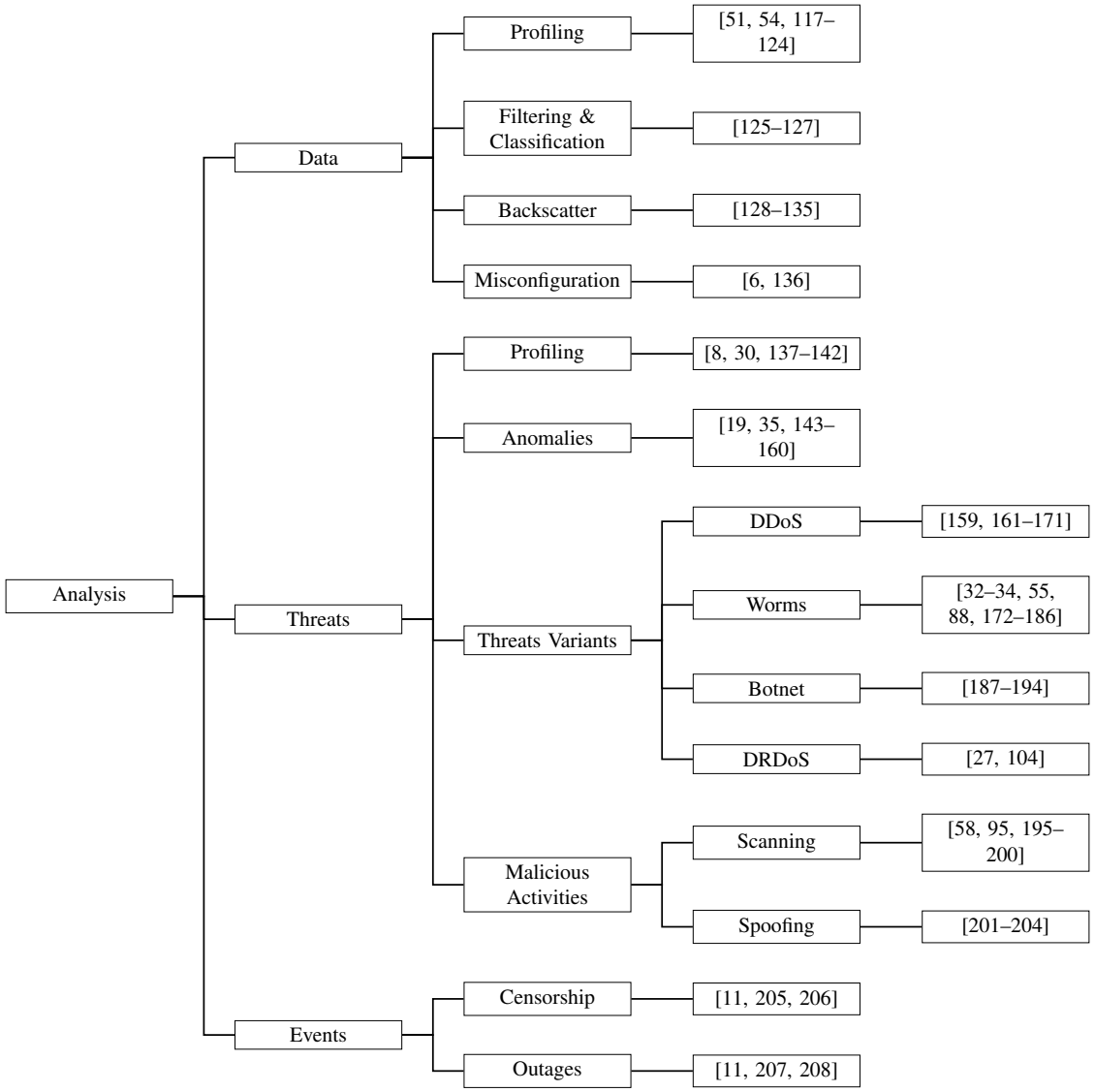


Fig. 10: Analysis Research Taxonomy - Overview

implications of evading sensor through DNS reconnaissance. They further stress on the defense aspect using proactive measures when deploying darknet sensors through delegating reverse DNS authority in a proper manner. At the end, they introduce honeydns, which complements low-interactive and darknet sensors by providing DNS trap services. Finally, Czyz et al. [124] report a large study of IPv6 darknet data. Through the analysis of five large $/12$ network address space, the authors highlight the nature of the traffic and compare it with IPv4 data. The researchers also provide various case studies to show notable properties while analyzing darknet IPv4 and IPv6 data.

Furthermore, time series analysis techniques are also used to profile passive monitoring data. For example, Fukuda et al. [121] discuss the temporal and spatial correlations among piecewise unwanted traffic. The aim of their techniques is to determine whether they can estimate statistical properties of global unwanted traffic behavior from smaller darknet address blocks. They found that the fluctuation of darknet traffic was close to random compared to normal traffic. Moreover, the authors demonstrated that the TCP SYN traffic time series had a strong spatial correlation. On the contrary, for TCP SYNACK and UDP traffic time series, Fukuda et al. [122] confirmed that in this case they were less correlated. The authors stress the need for a more sophisticated classification of the UDP unwanted traffic. They further investigated the macroscopic behavior of unwanted traffic collected using a $/18$ darknet over one year period. In order to measure the complexity in network traffic, Riihijarvi et al. [209] study different entropy metrics. The generated metrics provide a better understanding of the traffic and help finding a new way to characterize the data. Moreover, the proposed technique uncovered structures on different traffic measurements and timescales. These authors extend their work to propose the use of multi-scale entropy analysis to characterize network traffic and spectrum usage. They showed that this technique can quantify complexity and predictability of analyzed traffic in widely various timescales. The results further showed that

Publications	Approach/Technique	Contribution	Tool/Project
[117]	Comparative Study - Packet Filtering	Exploring Data Across Five Different Darknet Sensors	TENET
[54]	Packet Filtering - Routing	Characterizing Darknet Data	iSink
[118]	IPv6 Packet Analysis - Routing	Creating the First IPv6 Darknet	Custom
[119]	Multi-Dimensional IP/Port Analysis	Proposing a System to Improve Darknet Monitoring	Custom
[121, 122]	Time Series - Principle Component Analysis	Discussing the Temporal and Spatial Correlations in Data	Custom
[120]	Packet Filtering	Inferring the Evolution of Internet Infrastructure	CAIDA
[51]	Time Series - Spatial Analysis	Discussing Topics Related to Darknet	Custom
[123]	DNS Analysis	Introducing the Concept of Dark DNS	honeydns
[124]	IPv6 Packet Analysis - Routing	Studying IPv6 Darknet Data	Custom
[209, 210]	Time Series	Studying Different Entropy Metrics	Custom - CAIDA

TABLE X: Profiling Research Papers - Summary

different entropy structures exist for different traffic traces such as time series and commonly-used traffic [210]. Last but not least, Wustrow et al. [51] discuss topics related to darknet. They pinpointed the rapid growth of Internet pollution that was out spacing the growth of productive network traffic. Furthermore, they noticed trends toward increasing SYN and decreasing SYN-ACK traffic. In addition, they examine several case studies in Internet address pollution and offer specific suggestions for filtering them.

2) Data Filtering & Classification: This part includes the classification and filtering approaches of darknet data. These techniques are summarized in Table XI. Glatz et al. [125], for instance, analyzed a dataset that captured a significant amount of traffic to shed light on the composition of darknet towards large networks. The approach is based on a one-way traffic classifier. The authors found that such traffic constitutes the majority of all traffic in terms of flow and can be primarily attributed to malicious causes; however, it has declined since 2004 due to the relative decrease of scan traffic. Moreover, Wang et al. [126] propose a novel approach to filter darknet traffic. Their technique is based on relative uncertainty theory and is independent of configurations or building databases. The authors assume that data coming from regular users is relatively certain and not random. Furthermore, Cowie and Irwin [127] discuss the difficulties in generating training traffic for Artificial Intelligence (AI) analysis. The authors mention the problem in accurately labeling known incidents from darknet. Other factors related to this issue include the originality of data and the time involved. To address this problem, they work on two techniques, namely manual identification and automated generation. The first counts on heuristics for finding network incidents whereas the second considers building simulated data sets. They were able to construct a sample of an AI system out of this marked dataset.

3) Backscatter Data: Backscattered data is the reply packets sent to the darknet. Several factors can produce such scenario, such as DDoS victims replying to spoofed IP addresses and misconfiguration. Table XII summarizes the works that leverage backscattered traffic to generate cyber insights. Under this category, the research contributions employ several techniques such as mathematical models, network routing, packet filtering, and visualization.

For instance, Peng et al. [128] propose a Reconstruction based on Semantically Enhanced Counting Bloom Filter (RSECBF) algorithm to reveal the distribution of the main elements from semantically enhanced Counting Bloom Filter's hash space. The proposed algorithm deploys a specific technique, which directly selects certain bits from the primary string. The authors confirm the homogeneous hash strings and show the efficiency of the algorithm using real backscatter traces. Moreover, Rahmani et al. [129] study entropy-based anomaly detection through backscatter from darknet data. In particular, the authors try to understand the detection strength of using joint entropy analysis of many data distributions. The authors found statistical correlation between time series of IP flow number and collective traffic size. The approach was tested on backscattered data and led to more effective and accurate DDoS detection techniques. Moreover, Choras and Saganowski [131, 132] leverage backscatter data to propose an anomaly detection technique for recognizing malicious traffic. Using the correlation of parameters from different layers, the authors were able to detect unknown attacks with a low amount of false positives. The authors correlated signal-based and statistical features to enhance their technique. The proposed framework uses, for the first time, the Matching Pursuit (MP) algorithm [211] on network traffic. They found superior results to other IDSs that work on discrete wavelet transform. Similarly, Fadlullah et al. [130] detected anomalies through strategically distributed Monitoring Stubs (MSs). This work was able to categorize encrypted protocols. The MSs are designed to extract features and build normal behaviors. Based on deviations in traffic, the technique can differentiate between suspicious and benign traffic. After the detection process, MSs notify victims to trace-back the source of the attack and take necessary action.

He and Parameswaran [134] leverage backscatter data to propose a novel anomaly detection system based on multiple connections. The approach is considered faster than previous anomaly detection mechanisms. This Multiple Connection based Anomaly Detection (MCAD) system relies on the concept that attackers use similar connections to execute an attack. Hence, the algorithm tests for similarities within connections, and if the value is above a certain threshold, the connections are flagged as malicious. Over one million connections of backscatter traffic were tested in this work. MCAD was able to identify fifteen forms of connections, in which fourteen were fully detected while only one was detected with 66% accuracy. In addition, in order to detect congestion online, the authors in [135] propose a mechanism to detect congestion in network traffic by analyzing passive and aggregation links. The technique is based on delays in TCP data. The approach was tested on backscatter data and proved to be efficient. This technique is considered dynamic with fast detection capability.

Publications	Approach/Technique	Contribution	Tool/Project
[126]	Relative Uncertainty Theory	Filtering Darknet Data	Custom
[125]	Classification Scheme	Composition of Darknet Data	Custom
[127]	Hybrid	Manual Identification and Automated Generation of Data	Custom

TABLE XI: Filtering & Classification Research Papers - Summary

Publications	Approach/Technique	Contribution	Tool/Project
[135]	TCP Delay	Detecting Network Congestion	CAIDA
[131, 132]	Matching Pursuit (MP) Algorithm	Detecting Malicious Traffic	CAIDA
[128]	Bloom Filter	Enhancing Counting Bloom Filters Hash Space	RSECBF - CAIDA
[130]	Distributed Stubs	Detecting Anomalies	DTRAB - CAIDA
[129]	Entropy-Based	Studying Entropy-based Anomaly Detection	CAIDA
[134]	Multi Connections	Detecting Anomalies	MCAD - CAIDA
[133]	Clustering	Extracting Traffic Signatures	PISA

TABLE XII: Backscatter Research Papers - Summary

Publications	Approach/Technique	Contribution	Tool/Project
[6]	Probing - Routing	Reachability Analysis	Arbor Networks
[136]	Packet Filtering - Routing	Studying Trends of Network Misconfiguration	CAIDA

TABLE XIII: Misconfiguration Research Papers - Summary

Publications	Approach/Technique	Contribution	Tool/Project
[30]	Broad-Spectrum	First to Investigate Trap-based Monitoring Systems	Custom
[69]	Packet Filtering	Representing the Infrastructure of Data Gathering	honeyd/WOMBAT
[137]	Time Series - Statistics - Power Spectrum	Studying the Statistical Properties of Darknet	Custom
[138]	Statistics	Presenting the Leurre.com Project	Leurre.com
[139]	Graph-Based - Statistics	Proposing a Distributed Monitoring System	Custom
[140]	Time Series	Predicting Anomalies	DCMA
[141]	Change-Point - Data Mining	Analyzing Suspicious Behaviors	NICTER
[142]	Comparative Study and Correlation	Studying the Selection of Sources of Information	CAIDA/ATLAS/SANS/DShield

TABLE XIV: Threat Profiling Research Papers - Summary

In an attempt to fingerprint malicious attacks, Chhabra et al. [133] propose PISA, a packet imprint in security attacks algorithm, for automatic extraction of traffic signatures. PISA has the capability to cluster flows based on similarity in packet information and generate signatures from clusters. This tool was tested on two weeks of backscatter data encompassing 100 million packets. The results inferred about 1744 signatures related to several malware including the Blaster worm.

4) Data Misconfiguration: This section lists research works that leverage darknet to infer misconfiguration, errors and data management in network communications. Relevant contributions are shown in Table XIII. For instance, Francois et al. [136] demonstrate that darknet is a powerful tool in analyzing malicious network activities as well as network management. The authors present trends of network misconfiguration using darknet analysis. In practice, the results illustrated that deployed networks suffer from well-known errors and faulty configuration. Furthermore, Labovitz et al. [6] present a large study on the one-sided differences in Internet service provider reachability. The authors focus on darknet and the range of topology reachable to some providers but unreachable through one or more competitor networks. They present both active and passive measurements of differences between service providers' reachability. The goal is to discover the level to which commercial strategies, peering disputes, network failures, misconfiguration, and various malicious acts can lead to a partitioning of Internet topology. The results showed that the Internet was indeed partitioned and that darknet existed in a large amount (5% of Internet addresses). Moreover, the authors found that some prefixes were reachable only to specific providers. In addition, 70% of hosts responded to reachability tests and the majority of these devices were cable/ISDN pools and US military hosts.

B. Threats

One of the major elements in passive network monitoring is the extraction of insights on suspicious activities and threats on the Internet. Recall Figure 10, this section includes the research contributions in the following areas: threat profiling, anomalies, threats' variants and malicious activities.

1) Threat Profiling: Threat profiling includes the characterization (profiling) of darknet threats. Table XIV lists the threat profiling papers. Several techniques are used to profile darknet threats such as time series, statistics, and network routing.

Various researchers utilize time series and statistical methods to profile darknet threats. For instance, Harder et al. [137] study the statistical properties of class C darknet addresses for over three months. The authors found that the majority of the traffic is based on few IP sources and destination addresses. The study included a demonstration using many visualization techniques to represent darknet data and showed severe attacks such as DDoS and scanning activities. Using

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

different techniques, such as power spectrum, inter-arrival time of packets and detrended fluctuation analysis of this data, the authors found small signs of long-range dependency within the analyzed traffic. Francois et al. [139] leverage statistical techniques to propose a distributed system that monitors threats using centrality of a graph and its time evolution. Furthermore, Holz [138] presents the leurre.com project and discusses the importance of collecting data from different locations and generating results based on correlation engines. The author highlights insights in terms of finding attack patterns and pinpointing root-causes of threats. Ohta et al. [140] uses time series analysis to study the possibility of predicting anomalous packets' behaviors by observing a small darknet address space. The researchers propose distributed cooperative monitoring architecture (DCMA) technique, which aims to probe and detect anomalous packets. The authors calculate the correlation strength of anomalous packets, observe the correlation strength when changing the sub-observation's size, and note the dependency of the correlation strength.

We further list contributions that utilize hybrid and custom techniques to characterize darknet threats. For example, Bellovin et al. [30] are among the first to investigate trap programs that search for attacks. Their work can also be the primary motivation that triggered the idea of darknet monitoring. A variety of pokes were found during their analysis. The authors believe that these attacks occured on many online sites without the administrators' knowledge. In this work, they also provide important security information to security operators on how the attackers were operating [8]. In addition, Inoue et al. [141] utilize NICTER to propose a novel method to analyze suspicious behaviors. The latter technique is based on the malware's external behavior. Their experiment is executed in a safe environment using virtual machines. Moreover, Dacier et al. [69] leverage WOMBAT to represent the infrastructure of data gathering. This project is based on an extended version of honeyd with SGNET [212]. In this experiment, the authors were able to observe the evolution an army of zombies. Their approach is found to be efficient to use for multidimensional analysis of events. The authors also shared some insights found when collecting malware (including zero-day) and described different stages of attacks. Finally, Berthier et al. [142] present a large and comparative study to help security operators in selecting sources of information. By comparing three different sources of security information including darknet dataset, the authors correlated attacks among different sources of data having various granularity.

2) Anomalies: This section provides a summary of the darknet-based research that targets the detection and mitigation of anomalies. Table XV denotes these research publications. The major techniques are based on IDS, mining, clustering, and time series.

First, several researchers leverage IDS systems to detect anomalies. For instance, Yegneswaran et al. [143] present a broad, empirical analysis of Internet intrusion activity using a large set of Network-based IDS, firewall logs and darknet data. Their breakdown of scan types showed not only a large amount of worm propagation but also a substantial amount of scanning activities. To gain insights into the global nature of intrusions, the authors use their data to project the activity across the Internet. They also present a theoretic evaluation on the potential of using data shared between networks as a foundation for a distributed intrusion detection infrastructure. Furthermore, Karthick et al. [144] use probability to describe an adaptive network-based IDS with a two-stage architecture. The first stage includes a probabilistic classifier whereas the second uses a Hidden Markov Model to narrow down the attack source IPs. The proposed hybrid model was tested and showed good performance in detecting intrusions. For the purpose of providing situational awareness, Barford et al. [145] explore techniques that can be used to integrate trap-based monitoring data into daily monitoring systems. These techniques are based on IDS system and other statistical methods. The authors also discuss techniques that can detect whether an attack is purposely or incidentally targeting a victim as part of a larger attack. The analysis showed prevalence of different scanning techniques and useful information on trends, uniformity, coordination, and darknet-avoidance.

Second, several authors utilize mining and clustering techniques to learn about anomalies. For example, Inoue et al. [146] leverage resources from Nicter to propose a novel application of large-scale darknet monitoring in live networks. The technique investigates packets transmitted from inside networks instead of outside. In addition, Thonnard and Dacier [147] aim to generate insights on the method of operation of new emerging attack phenomena. To accomplish this goal, they have presented a multi-dimensional knowledge discovery and data (KDD) mining method. This technique extracts meaningful nuggets of knowledge and synthesizes those pieces of knowledge at different dimensional levels to create a concept that can best describe real-world phenomena. Furthermore, Thonnard et al. [148] present an analysis framework for discovering groups of attack traces having similar patterns. The authors extract knowledge of darknet data by discovering attack patterns via attack trace similarity, rather than a rigid signature. The results of their clustering method enabled the identification of activities of several worms and botnet in the collected traffic. In a similar work [149], the authors introduce a general analysis method to address the complex problem related to attack attribution. Their approach is based on a mixture of knowledge discovery and a fuzzy decision making process. By applying their technique on darknet attack traces, the researchers showed how to attribute and identify large-scale orchestrated cyber campaigns. Finally, in our work [151], we have characterized darknet data and investigated darknet threats. The aim is to study threats that are found on darknet and prioritize their severities. We further explored the inter-correlation of these threats by conducting association rule mining studies to generate association rules. Our technique extracts clusters of co-occurring malicious activities targeting certain victims. This contribution proved that some threats found on darknet are correlated. Furthermore, our technique provided intelligence about patterns within threats and allowed the interpretation of attack scenarios.

Third, the following group of authors uses time series tech-

Publications	Approach/Technique	Contribution	Tool/Project
[146]	Packet Filtering - Routing - Data Mining	Proposing a Novel Application of Large-scale Monitoring	Nicter
[19]	Time Series - Mathematical Model	Monitoring Large-scale Security Threats	CAIDA
[156]	Opportunistic Measures	Uncovering Hidden Regions of the Internet	Custom
[157]	Automated Knowledge Discovery	Introducing Cliques	Cliques/Leurre.com
[143]	Generic IDS - Firewall	Presenting an Empirical Analysis of Internet Intrusion	Custom
[35]	Time Series - Discrete Wavelet Transform	Observing Unknown Malicious Activities	Custom
[154]	Cardinality Variation	Analysing and Detecting Cyber Attacks	Custom
[155]	Poisson Statistical Process	Detecting Malware	IMS
[147]	Knowledge Discovery - Data Mining	Studying New Emerging Attack Phenomena	Honeyd/Leurre.com
[150]	Context-Aware	Detecting and Mitigating Online Threats	Custom
[151]	IDS - Association Rule Mining	Characterizing Data and Investigating Threats	Custom
[152, 153]	Time Series - Sliding Window Cumulative Sum - Change Point	Automatic Recognizing Variations in Network Traffic	Custom
[148]	Time Series - Pattern Recognition - Clustering	Analyzing Attack Patterns	Honeyd/leurre.com
[149]	Knowledge Discovery - Fuzzy Decision Making	Proposing and New Technique for Attack Attribution	Honeyd/leurre.com
[144]	IDS - Hidden Markov	Describing an Adaptive NIDS with a Two-stage Architecture	Custom
[158]	Messaging Framework	Proposing a Framework for Real-time Analysis	Custom
[159]	Selective Pulling - Ratio-Based Algorithm	Proposing a System for Timely Business Intelligence and Decision Making	RTQ
[145]	IDS - Statistics	Exploring New Techniques to Leverage Darknet Monitoring	Honeyd
[160]	Hotspots	Defining Hotspots for Malware Detection	IMC

TABLE XV: Anomalies Detection and Mitigation Research Papers - Summary

niques. Limthong et al. [35] applied Discrete Wavelet Transform (DWT) techniques for traffic signal decomposition and observed unknown malicious acts from darknet information. In particular, the authors focus on TCP SYNs, TCP SYN/ACKs and UDP packets based on three time intervals. The purpose of this work is to show the importance of time series wavelet methods in finding insights about malicious communications on darknet. In addition, Ahmed et al. [152, 153] leverage time series techniques and the dynamic sliding window cumulative sum (CUSUM) algorithm to automatically recognize nested changes in network traffic and detect any number of these changes. This automatic detection approach can identify both the beginning and the end of abnormal changes.

Finally, several hybrid and custom techniques are used to detect and mitigate anomalies. For example, Chen et al. [154] focus on the analysis and inference of cyber attacks through a technique based on the changes in the cardinality of the attack traces. The approach develops a nonparametric error-bound scheme with the capability to detect cyber attacks through a centralized data center of multi-monitoring sensors. This scheme uses small space and constant processing time, which allow the system to operate in near real time. In addition, a statistical approach is used by Soltani et al. [155] to detect malware on darknet traffic. The authors introduce the Piecewise Poisson process Model (PPM) and check the rate of traffic to detect malware outbreaks. The researchers then implement a regression model that can be used to characterize changes in the PPM data rates. In addition, Moore et al. [19] leverage the analysis of darknet traffic to monitor large-scale security threats. They showed a trend in cyber attacks based on a period of over two years. Moreover, these authors study the relation between the detection ability and size of these sensors, profile precision in detecting duration and rate of an attack, and discuss good practices in darknet deployment. Furthermore, Casado et al. [156] propose opportunistic measures from spurious network traffic (such as darknet) to uncover hidden regions of the Internet. The authors identify such sources and demonstrate their possible use in providing efficient statistical data. In addition, Pouget et al. [157] introduce an automated knowledge discovery technique called Cliques. The Cliques methodology provides insights on how attacks occur and potentially identifies the source behind them. The authors used the proposed methodology and found useful data about similarities in the method of operation of many potentially unrelated malicious tools. In addition, Hunter et al. [158] propose a framework for real-time analysis of darknet and honeypot data. The technique uncovers several malicious behaviors. In order to collect data, the authors develop an automated reconnaissance (AR) framework that works in both passive and active modes. The authors utilize several features to identify malicious users such as OS name, targeted service, location and services operating on the adversary. Gupta et al. [159] propose a ratio threshold queries (RTQs) system that can be used for timely business intelligence and near real time decision making. For instance, the system can defend against malicious attacks on the Internet such as DDoS when the ratio of queries surpasses a certain threshold. The system further uses selective pulling techniques for inferring extra sources. In addition, Sinha et al. [150] investigate techniques in detecting and mitigating online threats via the context available in network, environment and host. The authors explain the context concept which is based on three main properties: vulnerability profile, attack surface, and usage model. The authors leverage ten years of experience to prove the efficiency of the approach in enhancing security techniques. Last but not least, Cooke et al. [160] define hotspots as the root cause of non-uniformity in self-propagating malware. In this work, the authors claim that two main factors are behind its existence, namely, the algorithmic and environmental factors. Using eleven sensors located at different addresses around the Internet, they measured the impact of these factors on the propagation of worms and bots (or zombies). Based on this idea, the authors simulated the outbreak of new threats with hotspots and demonstrated the effect of the aforementioned factors on the visibility of monitors and hence the efficiency of detecting new threats.

3) Threats Variants: We list in this section the threats that can be detected through the analysis of darknet data, namely, DDoS, worms, botnets and DRDoS.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Publications	Approach/Technique	Contribution	Tool/Project
[213, 214]	Chi-Square Statistics	Detection	CAIDA
[161]	Identifier/Location Separation	Prevention	Custom
[215]	Greedy Algorithm	Detection	Custom
[159]	Greedy Algorithm	Detection	Custom
[162]	Stream-Based Processing	Prevention and Mitigation	STONE
[163]	Adaptive and Hybrid Neuro-fuzzy	Detection	NFBoost
[164, 165]	Traffic Analysis - MIB	Detection and Mitigation	D2M2
[166, 167]	Flow-Based Algorithm - Data Mining - Time Series	Prediction	Custom
[168]	Data Correlation	Detection and Mitigation	OADS
[169]	Total Variation Distance	Detection	Custom
[170]	Flow-level - Reputation-Based	Prevention and Mitigation	TrustGuard
[171]	Change Point	Detection	Custom

TABLE XVI: DDoS Research Papers - Summary

a) DDoS: DDoS is one of the most severe cyber attacks. Denial of Service (DoS) attacks are characterized by an explicit attempt to prevent the legitimate use of a service. Table XVI summarizes darknet-based DDoS research papers. Below is an overview of these studies. Some techniques include mathematical models, network routing and packet filtering.

First, the following publications leverage mathematical and statistical models for generating DDoS insights through darknet analysis. For instance, Andrysiak et al. [215] focus on detecting DDoS threats using greedy algorithms. More specifically, the approach uses Matching Pursuit algorithms, which look into best matching projections of multidimensional dataset. Similarly, Gupta et al. [159] focus on the analysis of backscatter and MAWI data [216] to detect DDoS by means of greedy algorithms. The approach is based on several matching pursuit algorithms. In addition, Arun et al. [163] propose NFBoost, an adaptive and hybrid neuro-fuzzy approach to detect DDoS attacks. The approach combines various classifier outputs and cost strategy minimization technique for classification determination. The approach was tested on real DDoS traces and trained with publicly available dataset. Furthermore, the evaluation was based on two metrics, namely the detection accuracy and cost. In our work [166, 167], we have propose an approach to predict, within minutes, certain DDoS and their attacks features; namely, intensity/rate and size. The aim is to forecast the future short term trend of DDoS attacks. The analysis is based on darknet data and the attack traces are tested for predictability using a time series approach prior to predicting. In addition, Rahmani et al. [169] propose a two-stage DDoS detection approach based on the breaks in the connection size distribution. To achieve this goal, the authors employ a total variation distance technique to calculate the horizontal and vertical similarity between flows. The approach detects high- and low-rate DDoS attacks. Furthermore, Abouzakhar et al. [213] present a network-based system for anomaly detection using chi-square statistics. This technique is a robust multivariate anomaly detection method with minimum computation cost. The objective of this method is to reduce the limitation of intrusion detection and network forensics. In an extended work [214], the same authors developed patterns for intrusion detection based on data mining techniques and Fuzzy algorithm. This Association Rule Mining-based (ARM-based) detection technique was successfully tested on real DDoS data. They further presented an enhanced Fuzzy ARM matrix for mining and associating rules. This hybrid approach can improve the efficiency of anomaly detection.

Second, other group of researchers tackle IP filtering and network routing techniques to investigate DDoS. For instance, Luo et al. [161] apply the identifier/location separation technique, a mechanism to solve the issue of routing scalability on the Internet to prevent distributed DDoS attacks. The proposed approach hardens the security to control machines (e.g., controlling infected machines through a botnet). This approach was evaluated using real DDoS traffic and showed promising results. Furthermore, a change point technique coupled with an analysis of source IP addresses are used by Ahmed et al. [171] to detect high-rate flooding attacks. The authors use a proof of concept development of the proposed methodology and show the efficient representation of pre-onset IP addresses that can also be used for threat mitigation.

Finally, other DDoS detection and mitigation techniques are used such as Chen et al. [162] who introduce STONE, a stream-based framework designed to defend against DDoS attacks. The STONE is a hybrid and scalable system that leverages anomaly-based inference and mitigation. The system operates through continuous data streaming queries to maintain data processing. The approach is also useful in the case of flash (high speed) events and operates in a priority-based fashion. STONE is built on top of StreamCloud, which is an elastic parallel-distributed stream processing engine. Additionally, Bhatia et al. [164, 165] propose a model to detect and mitigate DDoS attacks. The model uses an MIB (Management Information Base) server load and network traffic analysis to detect DDoS attacks from various network layers. The proposed model has the capability to distinguish DDoS from flash events. Further on DDoS, Feitosa et al. [168] also propose an approach to detect and mitigate DDoS attacks. They present the specification of a new orchestration-based technique to infer and mitigate threats. The proposed approach is based on a framework that coordinates alerts and events, infers threats, and consequently chooses the ultimate action. The authors generate rules and infer attacks with a greater degree of certainty than simple anomaly detectors. Finally, Liu et al. [170] examine drawbacks of existing DDoS defense schemes and propose a credit-based defense system. This approach focuses on the diversity in size of the attack;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

the less diverse the attack flow, the smaller credit it gets. The DDoS attacks were found to accumulate less credit as they naturally have low diversity in their traffic. This mechanism was able to identify the characteristics of micro and macro DDoS attackers and victims.

b) Worms: Worms are malicious codes known to infect and propagate in a rapid manner. We list in this section the darknet-based worms related contributions. Table XVII summarizes these publications. The majority of techniques are focused on packet analysis, routing, mathematical models, statistics and time series.

Moore et al. [172] analyze the Code-Red worm. Primarily, the authors showed the spread of this worm based on its deactivation and infection. The worm infection rate peaked at 2000 hosts per minute. Subsequently, the authors geographically located and measured the population of the worm and checked affected ISPs and top level domains. Additionally, Moore et al. [32, 173] study the Slammer worm through darknet analysis. In particular, they showed how this worm selects its victims and explained the reasons behind its fast propagation. They further discussed the pitfalls of the worm's author which aid in its discovery. In addition, they executed several related measures, geographically located the worm's victims, and listed the geographic distribution of the worm. Finally, the authors highlighted the impact of the Slammer worm on Internet operations. Likewise, Berk et al. [174] present a technique to identify worm spread after a short period of time. This method detected worm spread only when 10% of the victims are infected and with a detection performance achieved with sensor covering only 1% to 2% of the monitored space. This automated system is based on ICMP unreachable messages. This proposed methodology examines worms and presents simulation results that measure the detection speed of active hosts. Also, Staniford et al. [175] investigate UDP-based worms. The authors simulate the Slammer worm, adjust its latency measurements and monitor its packet delivery rates. The results showed that 95% of 1 million vulnerable hosts can be infected in only 510 milliseconds, whereas another TCP based service can be 95% saturated within 1.3 seconds. To avoid worm containment techniques, the authors suggest that flash worms should reduce their speed and use deeper spread trees. Furthermore, the proposed approach includes defense mechanisms to detect flash worms. In addition, Shannon and Moore [33] study the Witty worm, which targets a buffer overflow flaw in many firewall products having Internet Security Systems. The authors shared a general view of the worm's spread, its victims and features. Similarly, Kumar et al. [178] analyze the propagation of the Witty worm. The authors exploit the structure of the code including its pseudo-random number generation function. Using limited darknet data, the researchers were able to mine individual packets' rate of infection prior to loss, corrected noise generated by the worm, and disclosed the worm's failure to reach all potential victims. Furthermore, these scientists explored the complete attack infection scenario tree and uncovered a target on a US military base. Furthermore, Abu Rajab et al. [179] utilize darknet data to infer the sequence of worms infection. The authors test the reliability and effectiveness of their proposed technique independent of scanning rate, vulnerable population and the sensor size. These authors measured the accuracy of this time series-based methodology, which reaches 80% after a few hundred initial infected machines. This technique further provided insights into the characteristics of the hit-list. Last but not least, Zou et al. [180, 186] investigate worms in two separate works. First, the authors propose several algorithms (e.g., Kalman filter) that effectively detect worm presence and its corresponding sensors. Second, they showed that they can predict the overall vulnerable population size of a uniform-scan such as Code Red. They further accurately estimated the infection size based on the analyzed data.

Bailey et al. [34] use a $/8$ darknet network from the IMS project to describe observations of the Blaster worm since its beginning in 2003. The authors explain how they measure its propagation, attack scenario, worm characteristics, life cycle in 2003, and persistence in 2004. Furthermore, Richardson et al. [177] examine how darknet affected the ability of global scanning worm detectors. They propose statistical models of darknet and combine them with random constant spread model of worm propagation to calculate the probability that a worm detector would be able to raise an alarm. Through their analysis, the authors concluded that global scanning worm detectors are not a viable long-term strategy for detecting worms in early stages. Additionally, Cooke et al. [176] try to understand non-uniformity in worms' behavior. Using a large darknet data rich with Blaster, Slammer and Code Red II infections, the authors analyzed and discovered three bias in malware behavior. More on worms detection, Pham et al. [181] tackle the problem of discovering multi-headed worms in the context of a larger dataset. Based on a 15 month of data, the researchers were able to confirm the existence of multi-headed worms and provided insights on worm behaviors. Kanda et al. [182] believe that worm-infected machine traffic characteristics are distinguishable from regular machines. They state that the difference in traffic between benign and malicious traffic can be classified by k-means clustering. Based on the volume of data, they also found that the proposed metric can isolate malicious traffic which has a small influence. Furthermore, Eto et al. [88] proposed an approach to understand the intentions of attackers and to have a comprehensive look of online threats. With focus on the W32.Downadup worm, the latter researchers found that 60% of their darknet attacking hosts are related to the above-mentioned malware. The authors also validated their findings with 86.18% accuracy using correlation analysis. Furthermore, Wang et al. [183, 184] estimate the temporal features of worms through simulation and analysis of darknet traffic. They propose several methods to detect the time of infection in order to rebuild the worm infection pattern. They leveraged this inference as a detection mechanism and estimate the detection error for various estimators. In addition, Irwin [55] studied worms in general and Conficker in particular. The author discussed the analysis of 16 million related darknet packets targeting port 445/tcp using a $/24$ address block. He further provided an overview and characterization of the collected data, including size and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Publications	Approach/Technique	Tool/Project	Worm
[172]	Packet Filtering - Routing	CAIDA	Code Red
[32, 173]	Packet Filtering - Routing	CAIDA	Slammer/Sapphire
[174]	ICMP Packet Analysis - Simulation	Custom	
[175]	UDP Packet Analysis - Routing - Simulation	CAIDA	Slammer
[33]	Routing - Time Series	CAIDA	Witty
[176]	Distributed Monitoring Sensors	Custom	Blaster - Slammer - Code Red II
[34]	Packet Filtering - Routing	IMS - Arbor	Blaster
[177]	Analytic Modeling - Simulation	Custom	
[178]	Packet Filtering - Routing	CAIDA & iSink	Witty
[179]	Time Series	CAIDA	Witty
[180, 186]	Kalman Filter - Simulation	CAIDA	Code Red - SQL Slammer - Blaster
[181]	Time Series - Clustering	Honeyd/Leurre.com	
[182]	Flow-Based - Clustering	Custom	Welchia-Slammer-Opasoft-Messenger
[88]	Micro & Macro Analysis	NICTER	W32.Downadup
[183, 184]	Maximum Likelihood & Regression	Custom	Code Red
[55]	Packet Filtering & Source OS	Custom	Conficker
[185]	Bloom Filter - Packet Filtering	Custom	

TABLE XVII: Worm Investigation Research Papers - Summary

Publications	Approach/Technique	Contribution	Tool/Project
[187]	Diurnal Shaping Functions	Studying Botnet Spread Dynamics	Honeyd
[188]	Time Series	Tracking Botnet Activities	Honeyd/Leurre.com
[189]	Hybrid	Designing a Hybrid P2P Botnet	Custom
[190]	Data Correlation	Outlining the Genesis and Structure of Zombies and Botnet	IMS
[191]	Cross Cluster Correlation	Presenting a Platform for Botnet Detection	BotMiner
[192]	DNS-based Blackhole	Fingerprinting Botnet Activities Using a Non-interactive Approach	Custom
[193]	Spam Sinkhole	Studying Spammers Behavior	Custom
[194]	IDS Correlation	Botnet Infection Detection	BotHunter

TABLE XVIII: Botnet Research Papers - Summary

time to live (TTL) value analysis. This work pinpointed a flaw in the Conficker scanning algorithm. Finally, the author located geographically the highly targeted victims based on region and numerical proximity. Finally, in an attempt to identify the location of worms binaries and stop their spread, Chen et al. [185] use the flexibility and high performance of network processors. The proposed inspection engine is built on top of an advanced network processor. The work includes testing and evaluating procedures to improve the performance of the proposed technique on real darknet data. The authors made the tool available for the anti-worm research community.

c) Botnet: Botnet is a platform for adversaries to generate large-scale and distributed cyber attacks. We list below the research works that leverage passive monitoring to investigate botnet activities. Table XVIII summarizes these publications. Contributions are divided into several techniques, mainly, trap-based monitors, clustering and correlation.

First, Dagon et al. [187] study how time and location affect botnet spread dynamics. They create a diurnal propagation model that uses shaping functions to capture regional variations in online vulnerable populations. The model aims at comparing propagation rates for different botnets, prioritizing response and predicting future botnet infections. The authors found that time zones play an important role in botnet growth dynamics, and that factors such as time of release are important to short term spread rates. For data collection and validation, the authors used several tools including Tarpit [22]. The researchers demonstrated that their model is more accurate than the previous ones and that it accurately predicts botnet population growth. Furthermore, Ramachandran et al. [192] perform a counter-intelligence passive monitoring of DNS trap to infer botnets' activities without interacting with them. The authors were able to identify scanning activities performed by botmasters and suggested early bot detection techniques. Ramachandran et al. [193] further study the behavior of spammers through the analysis of a 17 month period of traffic flows to spam trap.

Second, Cooke et al. [190] outline the genesis and structure of zombies and botnet. The authors monitor command and control (C&C) and Internet Relay Chat (IRC) communication. By correlating security information from multiple sources, the authors elaborate on their botnet detection strategy. Additionally, Gu et al. [194] present a new strategy for network monitoring, which aims at inferring the infection and the coordination dialog of a successful malware infection. Through the analysis of a $/17$ unused address space, the authors introduce BotHunter as an application to track the flows between internal and external entities. Using real network traces, the authors further evaluate the effectiveness of the project in detecting a variety of real-world botnets with low false positive rates. In a similar work, Gu et al. [191] present a general botnet detection framework called BotMiner. The authors started their investigation from essential botnet properties such as bots communication with C&C servers/peers. The technique uses cross cluster correlation to identify bots that share common malicious network patterns. This clustering methodology adopts many filters which include one-way traffic extraction such as scanning activity through uncompleted communication.

Other researchers like Pham and Dacier [188] demonstrate how to track botnet armies of zombies to characterize their

Publications	Approach/Technique	Contribution	Tool/Project
[195, 196]	Time Series - Statistics - IDS	Inferring Scanning Behavior	Custom
[95]	Spectrum Analysis	Extracting Malware Feature	SPADE
[58, 198]	Packet Filtering - Routing	Analyzing a World-wide VoIP SIP Scanning Campaign	CAIDA
[197, 200]	Statistics - Time Series	Proposing a General Framework to Extract Global Botnet Events	honeyd
[199]	White Hole	Defeating Malicious Probing Activities	Dark Oracle

TABLE XIX: Scanning Research Papers - Summary

lifetime and size. First, they propose a time series technique to identify attack events in a large dataset of traces. Second, they identified long-living armies of zombies. Third, they showed the importance of selecting the observation viewpoint when trying to group such traces for analysis purposes. Last but not least, Wang et al. [189] present the design of an advanced hybrid P2P botnet. The system uses various features such as robust network connectivity, individualized encryption and control traffic dispersion, etc. To defend against such a botnet, the authors elaborate on various approaches including analysis via darknet space. In this context, the researches discover that if the darknet can capture 200 copies of peer lists, network security defenders will be able to know more than 95% of bots used in the peer-list updating procedure.

d) DRDoS: An amplification or reflection DRDoS attack is a well known practice of a DDoS, in which malicious users abuse publically reachable servers to overwhelm a victim with amplified reply traffic [49, 50]. The technique consists of an invader directing a query to an open server having the source IP spoofed to be the victims address. Subsequently, all server responses will be sent to the targeted victim. In order to have a high impact on the victim, the attackers leverage requests with large size replies, and hence increase the amplification of the attack. Moreover, in order to increase the size of the attack with little effort, attackers use botnets to synchronize an army of bots and order them to send the requests. In recent research, DRDoS [29] activities are found to abuse several applications running on top of TCP [217] and UDP [28].

In our previous work [27, 104], we have proposed a novel approach to infer and characterize large-scale DNS-based DRDoS activities through the darknet space. Complementary to the pioneer work on inferring DDoS victims using backscattered traffic [31], the proposed approach leverages DNS queries (non backscattered) that seek open DNS resolvers to execute the attack. The approach uncovered traces from the largest DRDoS attack of March 2013 against Spamhaus [59].

4) Malicious Activities: We list below darknet research contributions that focus on the investigation of malicious activities, namely, scanning and spoofing.

a) Scanning: Scanning or reconnaissance activities are the first step in a cyber attack life cycle. In a typical attack scenario, adversaries execute a scan activity to search for vulnerabilities online before launching an attack on the vulnerable victim(s). Table XIX summarizes related research publications. The techniques are mainly based on time series and statistical models, in addition to network routing and packet analysis.

The first group leverages time series and statistical models to investigate cyber activities on darknet. For instance, Bou-Harb et al. [195] attempt to infer scanning or probing activities and identify the technique used to perform such probing. The approach, which is based on various statistical and probabilistic techniques, tries to identify the machinery of the scan. The analysis is done on large darknet data and shows promising results. The same authors propose an approach to detect and cluster cyber attacks targeting corporate networks. They evaluated the approach and found promising results when compared with the mostly used NIDS (snort) [196]. Furthermore, Eto et al. [95] focus on the oscillations of the destination IP addresses of scan packets to propose the concept of malware feature extraction. They implemented and evaluated a distinct analysis method dubbed as SPADE. The technique applies a spectrum analysis methodology to realize a fundamental goal, which is to grasp the general trend of malware propagation from only scan data. Through several evaluations, the authors show that SPADE successfully extract and distinguish malware features. Additionally, Li et al. [197] propose a general framework to extract botnet global scanning events. Using honeyd, where half of the sensors are dark, the authors analyze one year of data from a large research institution to study six different botnet scanning characteristics. Based on scanning techniques, the researchers distinguish two botnet arrival/departure patterns. The authors study the scan behavior and differentiate between exponential and linear distributions. The result was affected by the randomness of scanning activities and the high range of scans, which cross the sensor IP space. In a relevant work, the authors investigate probing events of botnet. The discussed techniques are suitable for users who deploy darknet. The goal is to implement techniques that can help understand the strategy and purpose of the distributed probing events on the local network. Moreover, through the local view of sensors, the researchers designed the scheme of scanning activities, cross-validated their findings with DShield data and showed promising precision [200].

The second group leverage network routing techniques. For example, Dainotti et al. [58] present the measurement and analysis of a 12-day world-wide cyber scanning campaign targeting VoIP (SIP) servers. The discovery has occurred while analyzing some large-scale probing events [198]. Their analysis is based on their collected data using a $/8$ dark IP address block. The authors note that the SIP scanning campaign involved approximately 3 million distinct source addresses (scanning bots), generated around 20 million probes, and targeted roughly 14.5 million destinations. For illustration purposes, the authors created a world map animation of the scanning campaign. Finally, Gu et al. [199] introduce a technique to counter scanning propagation. They propose the use of several components exploiting white hole networks, which are systems that co-occupy populated

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

network segments to mislead and defeat malicious probing activities. Among these components, the authors use an address mapper that actively gathers and updates the unused IP/port segment of the network that the white hole can occupy. The white hole technique can deter, slow down and halt the spread of critical scanning malware. The authors demonstrate the effectiveness of their approach by using analytical reasoning and simulations using real trace and address distribution data. They further prove the success of their work even when applied to small darknet address blocks.

b) Spoofing: Spoofing is a technique to fake the identify of adversaries. Table XX summarizes the related works that leverage darknet data. The techniques employ packet analysis and are divided into two categories.

Publications	Approach/Technique	Tool/Project
[201]	TTL Fields & Statistics	NICT
[202]	TTL & Identification Fields	Custom
[203]	ICMP Packets - Classification	Custom
[204]	ICMP Packets	CAIDA

TABLE XX: Spoofing Investigation Research Papers - Summary

The first group of authors leverage TTL values to investigate spoofing activities. For instance, Eto et al. [201] propose an inspection method focusing on the TTL field of each packet in order to statistically extract spoofed IP packets from traffic observed by darknet. They also provide an analysis engine against network attacks. Through an empirical evaluation, the authors found that at most 1.26% of spoofed packets exist in the darknet traffic. Similarly, Ohta et al. [202] propose an approach for detecting spoofed packets using the TTL and identification field frame values. The latter approach is based on time series analysis coupled with a statistical methodology. To validate the proposed approach, the authors used two darknet samples. They claimed that their method can extract a number of plausible spoofing packets from real darknet traces.

The second group of researchers uses ICMP packets to classify and trace-back spoofing activities. For example, Bi et al. [203] characterize spoofing attacks on the Internet. They classify address spoofing into six classes based on the position of the node being spoofed. This work also presents a trace-back mechanism to identify the origin of DDoS source based on the ICMP packets found on darknet. The results showed that attackers mostly use HTTP and HTTPS on top of TCP to execute their attacks. Last but not least, Yao et al. [204] present an Internet-scale Passive IP Trace-back mechanism that allows the tracking of the origin of anonymous traffic. A developed Internet route model is sequentially used to aid in reconstructing the attack path. The researchers applied their technique to darknet data and found that the proposed mechanism can construct a trace tree from at least one intermediate router in 55.4% of the spoofing attacks, and can construct a tree from at least 10 routers in 23.4% of attacks.

C. Events

We list below the cyber events that are extracted through the monitoring of darknet. Table XXI provides a summary of these publications. The majority of the works leverage packet filtering and analysis to extract insights on events such as network outages, censorship, etc.

For instance, Quan et al. [208] propose Trinocular, an outage detection platform that uses ICMP probes which target darknet space. This system helps in understanding the reliability of edge networks and has the capability to provide precised indicator on outage period in terms of time and date. The approach leads to more accurate (fewer false conclusions) results in comparison to the best available techniques. Furthermore, Dainotti et al. [11] study darknet during two natural phenomena: country-level censorship and two recent earthquakes. For country-level outages, the authors note that these events are stunningly visible using darknet instrumentation, and in conjunction with other sources of data, this can reveal information about how censorship is being implemented over time. The authors also examine the number of distinct source IP addresses in darknet. They further study how the ratio of this number, before and after the earthquakes, varies by distance from the epicenters. It is shown that some graphs illustrated significant differences before and after the events, while other graphs showed more subtle differences. Similarly, the authors in [205] analyze episodes of disruptions caused by Internet censorship in two countries. Their analysis rely on multiple sources of large-scale data, including Internet registries files, Internet routing information, and darknet data. The authors were able to pinpoint the forms of Internet access disruptions, which were implemented in a given region over time. Among other insights, the authors detected Libya's attempts to test firewall-based blocking before executing aggressive external routing-based disconnection. The researchers claim that their methodology can be used, in an automated fashion, to detect outages or similar macroscopic events in other geographic or topological regions. Furthermore, Bailey et al. [206] leverage Internet routing, backbone traffic and darknet data to explore different infrastructure-based works to interrupt Internet services. The authors focused on the risks of this long-term Internet evolution based on several realistic events, such as WikiLeaks DDoS, China Facebook filtering, Iran elections and Egypt Internet outages. Finally, Benson et al. [207] extend their disruption of Internet connectivity analysis to study the causes of macroscopic online disruptions. The authors propose metrics for inferring loss of packets in link congestion through AS analysis. This work listed three case studies to show how the approach can be used to identify and characterize large-scale outages.

Summary of Findings in Section V:

The previous section provided a taxonomy of several elements in the analysis process of darknet traffic, namely, data, threats, and events. Profiling darknet data allowed researchers to understand the nature of its traffic. Moreover, darknet was mainly designed to infer threats and malicious activities. Therefore, the analysis of darknet data in general and the threat analysis in particular, represented the largest part of this survey. From this section, we can conclude the following:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

Publications	Approach/Technique	Contribution	Tool/Project
[208]	ICMP Probes	Proposing an Outage Detection Platform	Trinocular
[11]	Packet Filtering - Routing	Studying Darknet Events During Natural Phenomena	CAIDA
[205]	Packet Filtering - Routing	Studying Internet Censorship and Disruption	CAIDA
[206]	Packet Filtering - Routing	Exploring Internet Service Interruption	Custom
[207]	Packet Loss	Studying the Causes of Macroscopic Online Disruptions	CAIDA

TABLE XXI: Events Research Papers - Summary

- A study in 2001 shows that darknet sensors occupy 5% of the whole IPv4 address space. An up-to-date study is needed to approximate the current size of darknet.
- Various data analysis techniques are leveraging darknet traffic. The majority of studies tackle the IPv4 address space whereas less than 1% investigate IPv6 darknet.
- Packet analysis, network routing, statistics and time series techniques are the mostly used in darknet analysis.
- Filtering misconfiguration packets is still not thoroughly investigated and is still a gray area that requires more attention from the research community.
- Worms and scanning activities are the most common threats that can be found on the darknet.
- Code Red and Slammer/Sapphire are the most analyzed worms on the darknet due to their large-scale infection and propagation mechanisms.
- CAIDA data is the most widely used by researchers to investigate worms and other malicious activities.
- Denial of Service attacks are the most severe threats that are extracted from the analysis of darknet data.
- Botnet investigation is considered challenging through monitoring solely darknet traffic. The reason behind this is that darknet considers passive monitoring only. Therefore, other interactive techniques such as honeypots can be used in parallel with darknet to enhance botnet investigation.
- Nowadays, DRDoS are the largest cyber threats reaching a peak of 400 Gbps in 2014. DRDoS activities can also be measured by analyzing darknet data. Less than 1% of the research tackled this promising area of study.
- Due to the nature of darknet, which is based on passive monitoring, and the inter-activities in botnet systems, very few researchers were able to link botnet research to darknet analysis. In addition, due to amplification activities, which have risen in the past couple years, several researchers were unaware of the importance of darknet in investigating such reflection activities. As such, botnet and DRDoS activities require more attention from the darknet research community.
- Differentiating between scanning and DRDoS is still partially a difficult problem due to the fact that both leverage scan-based techniques to operate. Scanning activities probe the Internet to collect information, whereas amplification activities generate scan-based requests to redirect amplified reply traffic to victim.
- Scanning and spoofing are not threats but malicious activities that adversaries utilize to acquire information or hide identities respectively. More research has been done on scanning.
- Packet analysis is the only technique used on darknet data to investigate spoofing activities. This method includes inspecting ICMP packets and TTL values. Less than 2% of research has been done on spoofing and darknet. Therefore, spoofing is still a severe malicious activity that needs more attention from the security research community.
- Darknet can be used to check Internet policies due to certain events such as political, or geophysical, among others. For instance, the variation in the amount of darknet traffic generated, before and after a censorship policy, could allow researchers to assess the failure-to-success ratio of initiating this policy. During our study, we have found that analyzing darknet traffic upon worldwide events is the most recent.

VI. Darknet Visualization

In this section, we survey the literature and elaborate on the usage of darknet traffic in detecting malicious activities by exploiting visualization techniques and tools. The taxonomy of the visualization-based research works is shown in Figure 11. Table XXII summarizes these publications.

Le et al. [218] propose a novel approach to infer malicious network traffic based on graph theory concepts such as degree distribution, maximum degree and distance measures. The authors model the network traffic using the traffic dispersion graphs (TDG) technique. As such, they analyze the differences of TDG graphs in time series to detect malicious activities and introduce a technique to identify attack patterns. The approach was validated using real network traces. Similarly, Joslyn et al. [219] propose a new technique to facilitate and visualize large-scale data. The graph-based approach leverages network routing databases. The authors described and presented real use cases in two graph-oriented query languages. This hybrid approach presents a new class of graph-relational analysis. In another visualization contribution, Krasser et al. [220] build a network traffic visualization system capable of both real-time and forensic data analysis. They aim to complement manual and automated analysis of network traffic by applying effective information visualization techniques in order to decrease the ratio of false positives and false negatives. Using link analysis and parallel coordinate plots with time sequence animation, the authors examine various dimensions that provide insights into both legitimate and malicious network activity. To validate the system operation, they used a dataset from five large-scale botnet traffic collected using darknet. Their results indicated that the system provides the capability to rapidly scan large dataset of network traffic for malicious activity despite visual noise. Moreover, Fontugne et al. [227] propose

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

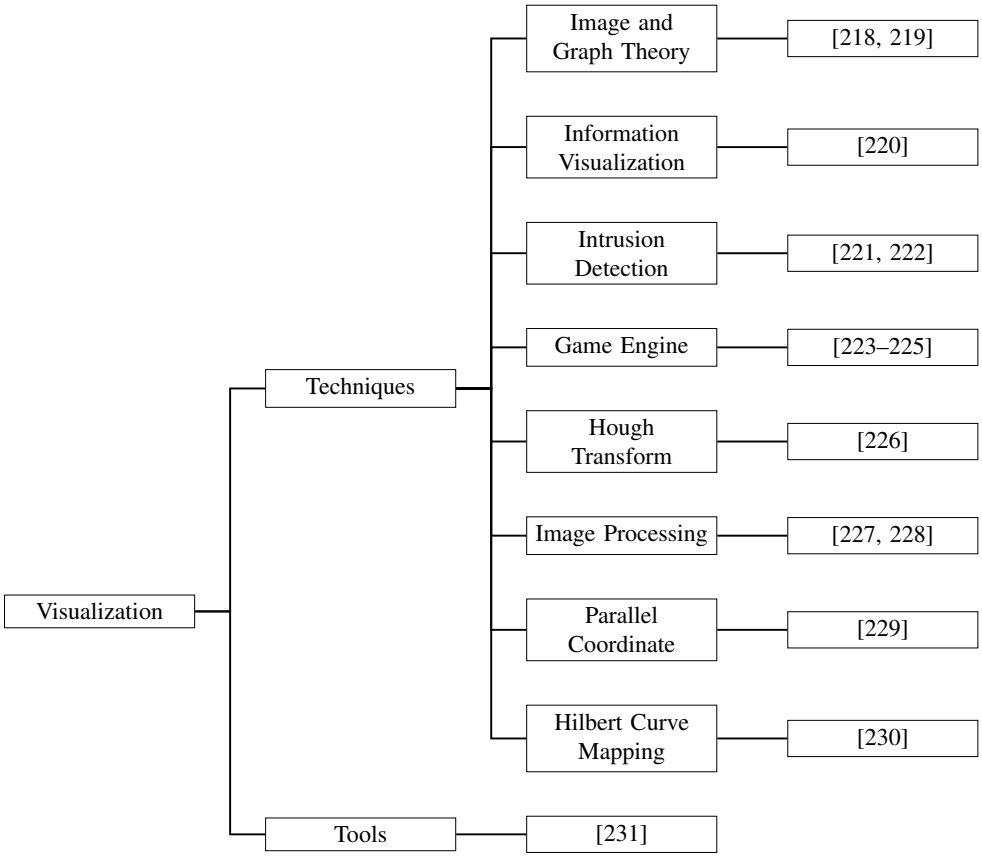


Fig. 11: Visualization Research Taxonomy - Overview

Publications	Approach/Technique	Contribution	Tool/Project
[219]	Graph-Based	Facilitating and Visualizing Large-scale Data	Custom
[218]	Graph Theory	Inferring Malicious Network Events	CAIDA
[220]	Link Analysis - Parallel Coordinate Plots	Building a Network Traffic Visualization System	Custom
[227, 228]	Pattern Recognition - Image Processing	Detecting Traffic Anomalies	Custom
[226]	Hough Transform	Proposing a Technique to Detect Scanning Activities	Custom
[223, 224]	Game Engine	Enabling Collaborative Network Control	Custom
[225]	Game Engine - Greynet	Visualizing Network Data	L3DGEWorld - OpenArena
[229]	Parallel Coordinate Information Visualization	Detecting and Visualizing Network Threats	PCAV - CAIDA
[230]	Hilbert Curve Mapping	Facilitating the Analysis of Large-scale Events	CAIDA
[221, 222]	IDS	Visualizing Various Backscattered and Scanning Traffic	InetVis

TABLE XXII: Visualization Research Papers - Summary

an approach for detecting traffic anomalies based on pattern recognition. The authors take advantage of graphical representations to break down the dimensions of network traffic. They further map network traffic data into snapshots rather than traditional time series. Moreover, these researchers identify unusual distributions in the traffic features through simple patterns. Their technique was implemented and its efficiency was demonstrated by comparing it with another statistical analysis technique. A variety of network traffic anomalies were detected by analyzing traffic from $/18$ network address space. They also propose a tool for visualizing and exploring network traffic on all temporal and spatial scales. Their tool aims to help researchers inspect traffic with basic features [228].

More on visualization techniques, Fukuda et al. [226] propose a technique to detect scanning activities in darknet traffic. They aim to estimate probing speed of change in terms of destination addresses, source ports and destination ports. Their method is based on an image processing technique applied to a two-dimensional image that represents unwanted traffic. They employ the progressive probabilistic Hough transform algorithm to detect edges in an image representing unwanted activities as lines. The authors apply their method on darknet traffic traces collected over a three-year period. They concluded that most of the scanning activities were characterized by intensive scans to a specific host. Furthermore, they found that few port scanning activities take place over a wide destination port space. Harrop and Armitage [223, 224] describe a system where a 3D game engine technology is used to enable collaborative network control. The proposed approach leverages simplistic interaction techniques by translating network events into visual activities. Their idea is to monitor a darknet network state that is represented in the 3D world by avatars spinning and jumping to visually alert network operators to a network anomaly. Subsequently, the operators can detect and shoot the alerting avatars to trigger a firewall access control list on a border router, preventing any further attacks. In a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

similar work, Parry [225] describe the L3DGEWorld project that is based on the OpenArena open source game engine platform. The approach aims to visualize network data based on the engine of a specific game. The approach describes the input interface to the L3DGEWorld server, which can be used to visualize and represent data in a real-time fashion. Moreover, the proposed approach also describes the output abstraction layer, through which data is connected from the virtual platform to the external daemon on the output interface.

Furthermore, several contributions attempt to visualize backscatter data from darknet. For instance, Choi et al. [229] build a model to detect and visualize network threats on parallel coordinates. This parallel coordinate attack visualization (PCAV) tool is able to detect zero-day attacks such as DDoS. PCAV operates based on several coordinates in a packet such as source and destination IPs, ports, and average packet length. Nine signatures were developed based on a hashing algorithm. Following the detection phase, network administrators must intuitively recognize and respond to the threat. This flow-based tool was proven to be efficient when applied to backscatter data. Furthermore, Irwin and Pilkington [230] develop a tool for facilitating the analysis of large-scale darknet traffic. In particular, the tool focuses on the analysis of data coming from different sources. The authors preserve the concept of nearness among numerical sequential IP address blocks using a Hilbert curve technique as a means of ordering dots within a visualization area. The authors also visualize the evaluation of worm spread algorithms. They further discussed the results and the importance of such tools to facilitate the analysis of big data. Riel and Irwin [221] propose InetVis, a visualization tool for darknet traffic. This tool plots TCP and UDP packets within a cube and ICMP packets within a flat plane. The authors adopt a time window to continue displaying an event after it has occurred. The researchers capture thousands of packets to test their tool. They observed numerous probing activities such as vertical and horizontal scans, step up scans (scan on the same port followed by stepping up the port range) and slow scans. Similarly, the authors in [222] demonstrate and compare InetVis with two open source NIDSs, Snort and BRO, where the advantages of the former are discussed. In this work, InetVis was re-implemented and enhanced. The authors argue that their tool is effective in visualizing various backscattered and scanning traffic while not suffering from high rates of false positives and negatives as do the other NIDSs.

It is noteworthy to mention that various tools exist on the Internet for darknet data visualization and analysis. The aim is typically to facilitate the analysis, the display, the collection and the presentation of the data. We list below tools from CAIDA [231] such as Cuttlefish for producing animated images that uncover the connection between the diurnal and geographical patters of data; GTrace for graphically trace-routing the destination; Geoplot for creating geographical images of data; LibSea for representing big directed graphs in memory and on disk; Mapnet for visualizing the infrastructure of multi-backbone providers; Otter for showing arbitrary communication information that can be presented as a group of nodes, connections or paths; Plankton for illustrating international cache topology; Plot-latlong for geographically mapping hosts; PlotPaths for displaying reverse and forward packets from one to one or one to many connections; and finally Walrus for representing large graphs in 3D.

Summary of Findings in Section VI:

In the last part of this darknet taxonomy, we have discussed several key elements in terms of darknet visualization, namely, techniques and tools. From what has been discussed in the that part of Section VI, we can conclude the following points:

- Several research works attempt to visualize darknet data by leveraging various techniques. The majority of these visualization techniques falls into two main areas, namely, generic-based and threat-based.
- Generic-based techniques aim at providing a graphical representation of usual darknet data (e.g., backscatter), whereas threat-based ones visualize darknet threats (e.g., DDoS).
- Generic-based techniques leverage mainly graph theory, whereas threat-based ones utilize mainly pattern recognition and image processing.
- Nevertheless, graph theory and game engines methods are used in both generic-based and threat-based techniques to model darknet network traffic.
- The majority (66.6%) of the visualization techniques are used to visualize threats on darknet.
- Although darknet data is similar to any network traffic, CAIDA research center is the primarily contributor to develop designated darknet visualization tools to depict large-scale events and threats.
- The visualization of darknet data is the smallest part of our darknet taxonomy.

VII. Discussion

Recall that our taxonomy split darknet research into various areas, namely, deployment, analysis and visualization. We accordingly chose the most relevant topics for discussion.

- **Deployment and Technology Development:** nowadays, technology has become a part of our daily life. Basic electronic devices such as phones, watches, and glasses have evolved into smart equipment and become easily accessible through the Internet. This new shift has obviously increased the opportunity for malicious users to abuse such services. The latter threat can have a direct impact on our lives. For instance, attackers are abusing the Internet to generate flood of Voice over IP phone calls to attack 911 emergency phone services or spam mobiles with anonymous call or SMS messages [60]. Therefore, deploying darknet that operates on phone and mobile numbers is highly needed. The latter techniques are considered significantly important and require enormous attention from the research community.
- **Analysis of IPv4 & IPv6 Darknet Data:** a major element that distinguishes IPv4 from IPv6 is the size of the address space. In a nutshell, IPv6 is designed to provide significantly more address space to handle the Internet

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/COMST.2015.2497690, IEEE Communications Surveys & Tutorials

growth in a more secure and efficient manner. The migration and integration between these two technologies have already started [232]. For instance, several techniques are being leveraged to handle this migration such as tunneling and address translation. This shift will obviously affect network monitoring systems such as darknet. As such, both defense and attack mechanisms will be affected. For instance, in regard to security, IPv6 packets might have higher encryption. The latter will make it harder for defense team to analyze and interpret suspicious traffic and easier for attackers to obfuscate. Furthermore, since IPv6 is larger in address space, this will make it harder to monitor huge amount of traffic and even for attackers to probe the large address space to look for vulnerabilities. Regardless of the aforementioned impacts, it is only a matter of time when IPv6 darknet will become more involved in the era of trap-based monitoring system. This requires an attention from the security community.

- **Visualization and Learning:** researchers have found that, in a learning environment, the majority of people need to see information before learning [233]. As such, visualization and gaming has emerged largely into technology such as social media, mobile and web services. In regard to cyber security, our vision is coherent with some of the aforementioned research works in Section VI which emphasize on building monitoring systems based on game engine and visualization techniques. Therefore, we believe that the future generation of tools and technologies in cyber security will include more visual effects and game-based services. Such technologies already exist. For instance, the LOIC [234] is a well-known network stress testing and DDoS attack tool used by malicious and benign users in a game-friendly manner. We predict that, in the upcoming years, similar technologies will become a new trend for the cyber space.
- **Visualization Era:** today's revolutionary technology is putting emphasis on visualization for the simple and friendly use of machines and information. Unsurprisingly, researchers are using game and image processing engines together with visualization techniques to simplify monitoring tasks, detecting as well as mitigating cyber threats. For instance, the NICTER project [87] is a typical scenario of such a visualization capability. The latter still shows promising future work in darknet research. In regard to darknet-specific visualization tools, they already exist; however, since darknet is like any other network traffic, the majority of network monitoring tools can be applied or tailored to visualize its traffic.
- **Cyber Capabilities:** some of the challenges today is to build cyber capabilities with the ability to provide a generic technique to automate the inference of botnet and orchestrated campaigns (e.g., DDoS and Spamming). Moreover, another challenge is to build a trusted centralized repository of darknet data that can be used for worldwide monitoring and intelligence sharing. Such a worldwide project requires a thorough understanding of the challenges behind the privacy and legal issues.
- **Cyber Awareness:** as mentioned earlier, enforcing cyber laws has escalated the intensity of attacks by 52%. Therefore, technology by itself is not the ultimate solution to mitigate and defend against cyber attacks. As such, other techniques like learning and education are needed to increase the awareness and help in applying best practices for ethically using the cyber space as a service, instead of abusing its enormous capabilities.

VIII. Related work

As previously discussed, a thin line distinguishes darknet from other trap-based monitoring systems such as IP gray space [36, 37], greynet [20], honeytokens [38] and darkports [39]. However, two main groups of surveys can be related to our work. The first group focused solely on specific technology or threat whereas the second elaborated on trap-based monitoring systems.

First, various surveys tackled the detection techniques in network traffic such as NIDS [40], threats such as DDoS [41], botnet [42, 43] and worms [44], and malicious activities such as scanning [45]. Compared to our work, this group of research focused on a specific technology or a threat only whereas ours was more comprehensive. For instance, our survey included not only a study on DDoS threats, but also provided an overview on several darknet topics that can be leveraged to infer various insights from the Internet, including threats, events, techniques and tools.

Second, in regard to surveys that tackled trap-based monitoring systems, Zhang et al. [46] were among the first to classify honeypots in 2003. They highlighted data capture and data control in honeypots. Furthermore, they provided a classification of these traps based on security and application purposes. Furthermore, Seifert et al. [47] presented a taxonomy of honeypots. The authors described a classification of honeypots based on several schemes and were able to distinguish between seven types of honeypots (e.g., low and high interactive). In 2012, Bringer et al. [48] divided honeypot research into 5 major areas: types of honeypots, analysis of data, configuration, detection of sensors, and legal and ethical issues. The main difference between the works in [46–48] and ours is the scope of the survey. This group of works focused on honeypots, including active monitoring. Complementary, our work focused solely on passive monitoring of unused IP addresses. The only work that touched darknet research is [47] by discussing darknet and comparing it to other monitoring systems (low and high interactive honeypots). Our work is more comprehensive in regard to darknet study as it covers development, data analysis, and visualization.

Therefore, our survey is more close to the second group of contributions which tackled trap-based monitoring systems. Our survey complements the aforementioned related research works. Furthermore, the realistic analysis and investigation of real data provides more understanding and hands-on investigation experience on darknet data and threat analysis. We provided a guideline to develop, analyze and visualize real cyber insights by leveraging darknet data. The extracted darknet knowledge in our work can help in building a cyber intelligence platform for Internet monitoring. We are not aware of any similar contribution.

IX. Concluding Remarks

In today's word, technology has emerged in all aspects of our lives. Regrettably, adversaries are abusing technology for their own benefits. As a result, Internet services have become a cheap tool for attackers to generate malicious activities such as infecting victims' machines, taking control, exhausting resources and stealing information. An efficient technique to observe the Internet and its threats is to monitor unused IP addresses, known as darknet. Since these addresses are unused, all traffic destined to it is considered suspicious.

In this work, we have presented a survey on darknet research for the past thirteen years. The aim is to provide a cutting edge taxonomy of the topic that simplifies the understandability of the related domains, allows classifying the research works, and highlights overlaps of various contributions. Furthermore, this survey reveals research gaps and provides a discussion on future works and directions. After defining, listing and comparing trap-based monitoring systems, we provided case studies based on the analysis of real darknet traffic. The analysis included investigation of the largest DRDoS in history. Next, we classified darknet research into three main categories: deployment of monitors, traffic analysis, and visualization of data.

First, the deployment encompassed the deployment techniques, darknet variants such as IP gray space and greynet, sensor placement and identification, data handling and projects at different scales. We realized that honeyd is the most commonly used low-interactive honeypot that has the capability to deploy darknet services. Second, the darknet analysis included traffic as well as threats and events classifications. Darknet traffic analysis was classified into general profiling, filtering and classification techniques, backscatter and misconfiguration traffic. We found that the time series techniques are mostly utilized in darknet traffic analysis. In regard to threat analysis, which formed the largest part of the survey, the classification consisted of profiling its malicious traffic, anomaly detection and mitigation techniques, threat variants (DDoS, worms, botnet, and DRDoS), as well as scanning and spoofing malicious activities. Furthermore, in terms of analysis of events, the taxonomy contained political and geophysical incidents, such as the war in Libya. Third, in regard to visualizing darknet traffic, we classified this research into tools and technologies. Finally, we have identified hot topics in this area, such as mobile darknet, IPv6 darknet, DRDoS, and event-based analysis, all which require significantly more attention from the security research community.

In a nutshell, darknet is a trap-based monitoring system running in passive mode. This technology is designed to infer Internet activities and threats. Darknet has emerged due to recent advances in three key disciplines: deployment techniques, data analysis, and visualization. In the near future, darknet is expected to remain one of the major approaches for Internet and cyber security monitoring, and distributed deployments are foreseen in various organizations and governments. Deployments will include IPv6 and VoIP mobile-based darknet traffic analysis and perhaps game-based engines. In this work, we surveyed cutting edge darknet research. Although important contributions have been achieved by researchers, much work is still needed towards improving darknet cyber monitoring.

References

[1] Charles Doyle and Alyssa Bartlett Weir. *Cybercrime: An Overview of the Federal Computer Fraud and Abuse Statute and Related Federal Criminal Laws*. Novinka Books, 2006.

[2] BBC News. Stuxnet worm 'targeted high-value Iranian assets'. http://www.bbc.co.uk/news/technology-11388018. Last accessed in July 2014.

[3] BBC News. Flame: Massive cyber-attack discovered, researchers say. http://www.bbc.com/news/technology-18238326. Last accessed in July 2014.

[4] CloudFlare. Technical Details Behind a 400 Gbps NTP Amplification DDoS Attack. http://tinyurl.com/p3exvnc. Last accessed in July 2014.

[5] Federal Bureau of Investigation. Testimony - Taking Down Botnets. http://www.fbi.gov/news/testimony/taking-down-botnets. Last accessed in July 2014.

[6] Craig Labovitz, Abha Ahuja, and Michael Bailey. Shining light on dark address space. Technical Report TR-2001-01, Arbor Netwoks, Ann Arbor, Michigan, USA, November 2001.

[7] Fotis Gagadis and Stephen D Wolthusen. *Topological Models and Effectiveness of Network Telescopes*. TechTarget, 2008.

[8] Steve Bellovin. There be dragons. In *USENIX Summer*, 1992.

[9] David Moore, Colleen Shannon, Douglas J Brown, Geoffrey M Voelker, and Stefan Savage. Inferring Internet denial-of-service activity. *ACM Transactions on Computer Systems (TOCS)*, 24(2):115–139, 2006.

[10] Michael Bailey, Evan Cooke, Farnam Jahanian, Andrew Myrick, and Sushant Sinha. Practical darknet measurement. In *40th Annual Conference on Information Sciences and Systems*, pages 1496–1501. IEEE, 2006.

[11] Alberto Dainotti, Roman Amman, Emile Aben, and Kimberly C Claffy. Extracting benefit from harm: using malware pollution to analyze the impact of political and geophysical events on the internet. *ACM SIGCOMM Computer Communication Review*, 42(1):31–39, 2012.

[12] Evan Cooke, Michael Bailey, Z Morley Mao, David Watson, Farnam Jahanian, and Danny McPherson. Toward understanding distributed blackhole placement. In *Proceedings of the 2004 ACM workshop on Rapid malcode*, pages 54–64. ACM, 2004.

[13] CAIDA: The UCSD Network Telescope. http://www.caida.org/projects/network_telescope/. Last accessed in July 2014.

[14] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing Privacy Enhancing Technologies*, pages 46–66. Springer, 2001.

[15] Chao Zhang, Prithula Dhungel, Di Wu, Zhengye Liu, and Keith W. Ross. BitTorrent Darknets. In *INFOCOM*, pages 1460–1468, 2010.

[16] Peter Biddle, Paul Engl, Marcus Peinado, and Bryan Willman. The darknet and the future of content distribution. In *Proceedings of the 2002 ACM Workshop on Digital Rights Management*, 2002.

[17] P Krishna Gummadi, Stefan Saroiu, and Steven D Gribble. A measurement study of Napster and Gnutella as examples of peer-to-peer file sharing systems. *ACM SIGCOMM Computer Communication Review*, 32(1):82–82, 2002.

[18] Team Cymru, Inc. Team Cymru Community Services: The Darknet Project. http://www.team-cymru.org/Services/darknets.html. Last accessed in July 2014.

[19] David Moore, Colleen Shannon, Geoffrey M Voelker, and Stefan Savage. *Network telescopes: Technical report*. Department of Computer Science and Engineering, University of California, San Diego, 2004.

[20] Warren Harrop and Grenville Armitage. Defining and evaluating greynets (sparse darknets). In *The IEEE Conference on Local Computer Networks*, pages 344–350. IEEE, 2005.

[21] Niels Provos. A virtual honeypot framework. In *USENIX Security Symposium*, volume 173, 2004.

[22] LaBrea: Sticky Honeypot and IDS. http://labrea.sourceforge.net/labrea-info.html. Last accessed in July 2014.

[23] Zakir Durumeric, Michael Bailey, and J Alex Halderman. An internet-wide view of internet-wide scanning. In *USENIX Security Symposium*, 2014.

[24] Alberto Dainotti, Alistair King, Ferdinando Papale, Antonio Pescapè, et al. Analysis of a/0 stealth scan from a botnet. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 1–14. ACM, 2012.

[25] Evan Cooke, Michael Bailey, Farnam Jahanian, and Richard Mortier. The dark oracle: Perspective-aware unused and unreachable address discovery. In *NSDI*, volume 6, pages 8–8, 2006.

[26] David Moore, Geoffrey Voelker, and Stefan Savage. Inferring internet denial-of-service activity. In *Proceedings of the 10th Usenix Security Symposium*, 2001.

[27] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. Fingerprinting Internet DNS Amplification DDoS activities. In *6th International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5. IEEE, 2014.

[28] Christian Rossow. Amplification hell: Revisiting network protocols for DDoS abuse. In *Symposium on Network and Distributed System Security (NDSS)*, 2014.

[29] Marc Kührer, Thomas Hupperich, Christian Rossow, and Thorsten Holz. Exit from Hell? Reducing the Impact of Amplification DDoS Attacks. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 111–125, San Diego, CA, August 2014. USENIX Association.

[30] Steven M. Bellovin. Packets Found on an Internet. *Computer Communications Review*, 23:26–31, 1993.

[31] D. Moore, G. Voelker, and S. Savage. Inferring Internet Denial-of-Service Activity. In *USENIX Security Symposium*, Washington, D.C., Aug 2001.

[32] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. The Spread of the Sapphire/Slammer Worm. Technical report, CAIDA, ICSI, Silicon Defense, UC Berkeley EECS and UC San Diego CSE, 2003.

[33] Colleen Shannon and David Moore. The spread of the witty worm. *IEEE Security & Privacy*, 2(4):46–50, 2004.

[34] Michael Bailey, Evan Cooke, Farnam Jahanian, David Watson, and Jose Nazario. The blaster worm: Then and now. *IEEE Security and Privacy*, 3(4):26–31, July 2005.

[35] Kriangkrai Limthong, Fukuda Kensuke, and Pirawat Watanapongse. Wavelet-based unwanted traffic time series analysis. In *IEEE International Conference on Computer and Electrical Engineering (ICCEE)*, pages 445–449, 2008.

[36] Yu Jin, Zhi-Li Zhang, Kuai Xu, Feng Cao, and Sambit Sahu. Identifying and tracking suspicious activities through IP gray space analysis. In *Proceedings of the 3rd annual ACM workshop on Mining network data*, pages 7–12. ACM, 2007.

[37] Yu Jin, György Simon, Kuai Xu, Zhi-Li Zhang, and Vipin Kumar. Grays anatomy: Dissecting scanning activities using IP gray space analysis. *SysML07*, 2007.

[38] Lance Spitzner. Honeytokens: The other honeypot, 2003.

[39] David Whyte, Paul C van Oorschot, and Evangelos Kranakis. Tracking darkports for network defense. In *ACSAC*, pages 161–171, 2007.

[40] Garcia-Teodoro et al. Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28(1):18–28, 2009.

[41] Tao Peng, Christopher Leckie, and Kotagiri Ramamohanarao. Survey of network-based defense mechanisms countering the DoS and DDoS problems. *ACM Computing Surveys*, 39(1), April 2007.

[42] Zhaosheng Zhu, Guohan Lu, Yan Chen, Zhi Fu, Phil Roberts, and Keesook Han. Botnet research survey. In *32nd Annual IEEE International Computer Software and Applications (COMPSAC)*, pages 967–972, 2008.

[43] Michael Bailey, Evan Cooke, Farnam Jahanian, Yunjing Xu, and Manish Karir. A survey of botnet technology and defenses. In *Proceedings of the Cybersecurity Applications & Technology Conference for Homeland Security (CATCH)*, pages 299–304, Washington, USA, 2009.

[44] Pele Li, M. Salour, and Xiao Su. A survey of Internet worm detection and containment. *IEEE Communications Surveys Tutorials*, 10(1):20–35, 2008.

[45] Monowar H. Bhuyan, D.K. Bhattacharyya, and J.K. Kalita. Surveying port scans and their detection methodologies. *The Computer Journal*, 2011.

[46] Feng Zhang, Shijie Zhou, Zhiguang Qin, and Jinde Liu. Honeypot: a supplemented active defense system for network security. In *The Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pages 231–235. IEEE, 2003.

[47] Christian Seifert, Ian Welch, and Peter Komisarczuk. Taxonomy of honeypots, 2006.

[48] Hiroshi Fujinoki Matthew L. Bringer, Christopher A. Chelmecki. A survey: Recent advances and future trends in honeypot research. *International Journal of Computer Network and Information Security (IJCNIS)*, 2012.

[49] CERT advisory. Smurf IP Denial-of-Service Attacks. http://www.cert.org/historical/advisories/CA-1998-01.cfm? Last accessed in July 2014.

[50] Vern Paxson. An Analysis of Using Reflectors for Distributed Denial-of-service Attacks. *SIGCOMM Comput. Commun. Rev.*, 31(3):38–47, July 2001.

[51] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. Internet background radiation revisited. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 62–74. ACM, 2010.

[52] E. Bou-Harb, M. Debbabi, and C. Assi. Cyber Scanning: A Comprehensive Survey. *Communications Surveys Tutorials, IEEE*, 16(3):1496–1519, March 2014.

[53] NMAP. Port Scanning Techniques. http://nmap.org/book/man-port-scanning-techniques.html. Last accessed in July 2014.

[54] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of Internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 27–40. ACM, 2004.

[55] Barry Irwin. A network telescope perspective of the Conficker outbreak. In *Information Security for South Africa (ISSA)*, pages 1–8, 2012.

[56] CAIDA. Conficker/Conflicker/Downadup as seen from the UCSD Network Telescope. http://www.caida.org/research/security/ms08-067/conficker.xml. Last accessed in March 2015.

[57] Yu Yao, Wen long Xiang, Hao Guo, Ge Yu, and Fu-Xiang Gao. Diurnal forced models for worm propagation based on conficker dataset. In *Third International Conference on Multimedia Information Networking and Security (MINES)*, pages 431–435, Nov 2011.

[58] Alberto Dainotti, Alistair King, Ferdinando Papale, Antonio Pescape, et al. Analysis of a/0 stealth scan from a botnet. In *Proceedings of the ACM conference on Internet measurement conference (IMC)*, pages 1–14. ACM, 2012.

[59] Team Cymru, Inc. The DDoS That Knocked Spamhaus Offline. http://tinyurl.com/d46gpkj. Last accessed in July 2014.

[60] Nan Jiang, Yu Jin, Ann Skudlark, and Zhi-Li Zhang. Greystar: Fast and accurate detection of SMS spam numbers in large cellular networks using gray phone space. In *Proceedings of 22nd USENIX Security Symposium*, pages 1–16, 2013.

[61] Vinod Yegneswaran, Paul Barford, and Somesh Jha. Global intrusion detection in the domino overlay system. In *NDSS*, 2004.

[62] Vinod Yegneswaran, Paul Barford, and Dave Plonka. On the Design and Use of Internet Sinks for Network Abuse Monitoring. In *proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection (RAID)*, pages 146–165, 2004.

[63] Vinod Yegneswaran, Paul Barford, and Vern Paxson. Using honeynets for internet situational awareness. In *Proceedings of the Fourth Workshop on Hot Topics in Networks (HotNets)*, 2005.

[64] Sang-soo Choi, Jungsuk Song, Seokhun Kim, and Sookyun Kim. A model of analyzing cyber threats trend and tracing potential attackers based on darknet traffic. *Security and Communication Networks*, 2013.

[65] Balachander Krishnamurthy. Mohonk: mobile honeypots to trace unwanted traffic early. In *Proceedings of the ACM SIGCOMM workshop on Network troubleshooting: research, theory and operations practice meet malfunctioning reality*, pages 277–282. ACM, 2004.

[66] Bailey et al. A hybrid honeypot architecture for scalable network monitoring. *Technical Report CSE-TR-499-04, University of Michigan*, 2004.

[67] Michael Bailey, Evan Cooke, Farnam Jahanian, Niels Provos, Karl Rosaen, and David Watson. Data reduction for the scalable automated analysis of distributed darknet traffic. In *Proceedings of the USENIX/ACM Internet Measurement Conference (IMC)*, 2005.

[68] Fabien Pouget and Thorsten Holz. A pointillist approach for comparing honeypots. In *Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)*, Vienna, AUSTRIA, 07 2005.

[69] Marc Dacier, Corrado Leita, Olivier Thonnard, Hau Van Pham, and Engin Kirda. Assessing cybercrime through the eyes of the WOMBAT. In *Cyber Situational Awareness*, pages 103–136. Springer, 2010.

[70] Fabien Pouget, Marc Dacier, and V Hau Pham. Leurre.com: on the advantages of deploying a large scale distributed honeypot platform. In *E-Crime and Computer Conference (ECCE)*. Citeseer, 2005.

[71] Peter Komisarczuk and Ian Welch. Internet sensor grid: experiences with passive and active instruments. In *Communications: Wireless in Developing Countries and Networks of the Future*, pages 132–145. Springer, 2010.

[72] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson. The Internet Motion Sensor: A Distributed Blackhole Monitoring System. In *proceedings of Network and Distributed System*

Security Symposium (NDSS, pages 1–13, San Diego, CA, 2005.

[73] Pei-Te Chen, Chi-Sung Laih, Fabien Pouget, and Marc Dacier. Comparative survey of local honeypot sensors to assist network forensics. In *First International Workshop on Systematic Approaches to Digital Forensic Engineering*, pages 120–132. IEEE, 2005.

[74] Robin Berthier and Michel Cukier. The deployment of a darknet on an organization-wide network: An empirical analysis. In *High Assurance Systems Engineering Symposium (HASE)*, pages 59–68. IEEE, 2008.

[75] Moheeb Abu Rajab, Fabian Monrose, and Andreas Terzis. On the effectiveness of distributed worm monitoring. In *Proceedings of the 14th USENIX Security Symposium*, pages 225–237, 2005.

[76] Barford et al. Toward a model for source addresses of Internet background radiation. In *Proceeding of the Passive and Active Measurement Conference*, 2006.

[77] Dean Pemberton, Peter Komisarczuk, and Ian Welch. Internet background radiation arrival density and network telescope sampling strategies. In *Australasian Telecommunication Networks and Applications Conference (ATNAC)*, pages 246–252. IEEE, 2007.

[78] Moheeb Abu Rajab, Fabian Monrose, and Andreas Terzis. Fast and evasive attacks: Highlighting the challenges ahead. In *Recent Advances in Intrusion Detection*, pages 206–225. Springer, 2006.

[79] Sushant Sinha, Michael Bailey, and Farnam Jahanian. Shedding light on the configuration of dark addresses. In *NDSS*, 2007.

[80] Yoichi Shinoda, Ko Ikai, and Motomu Itoh. Vulnerabilities of passive internet threat monitors. In *Proceedings of the 14th USENIX Security Symposium*, pages 209–224, 2005.

[81] John Bethencourt, Jason Franklin, and Mary Vernon. Mapping internet sensors with probe response attacks. In *USENIX Security*, 2005.

[82] Evan Cooke, Michael Bailey, Farnam Jahanian, and Richard Mortier. The dark oracle: Perspective-aware unused and unreachable address discovery. In *NSDI*, volume 6, pages 8–8, 2006.

[83] Evan Cooke, Andrew Myrick, David Rusek, and Farnam Jahanian. Resource-aware multi-format network security data storage. In *Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense*, pages 177–184. ACM, 2006.

[84] Alastair Nottingham and Barry Irwin. Towards a GPU accelerated virtual machine for massively parallel packet classification and filtering. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pages 27–36. ACM, 2013.

[85] Stefano Zanero. Observing the tidal waves of malware: experiences from the WOMBAT project. In *Proceedings of the 2010 Second Vaagdevi International Conference on Information Technology for Real World Problems*, pages 30–35. IEEE Computer Society, 2010.

[86] Arbor Networks. ATLAS. http://atlas.arbor.net/. Last accessed in July 2014.

[87] Daisuke Inoue, Masashi Eto, Katsunari Yoshioka, Shunsuke Baba, Kazuya Suzuki, Junji Nakazato, Kazuhiro Ohtaka, and Koji Nakao. nicter: An incident analysis system toward binding network monitoring with malware analysis. In *WOMBAT Workshop on Information Security Threats Data Collection and Sharing (WISTDCS)*, pages 58–66. IEEE, 2008.

[88] Masashi Eto, Daisuke Inoue, Jungsuk Song, Junji Nakazato, Kazuhiro Ohtaka, and Koji Nakao. NICTER: a Large-Scale Network Incident Analysis System: Case Studies for Understanding Threat Landscape. In *proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, BADGERS, pages 37–45, New York, NY, USA, 2011.

[89] Corrado Leita, Van-Hau Pham, Olivier Thonnard, Eduardo Ramirez, Fabien Pouget, Engin Kirda, and Marc Dacier. The Leurre.com Project: Collecting Internet Threats Information using a Worldwide Distributed Honeynet. In *Proceedings of the 1st WOMBAT Workshop on Information Security Threat Data Exchange (WISTDE)*, pages 40–57, 2008.

[90] Fabien Pouget, Marc Dacier, Hervé Debar, and Van Hau Pham. Honeynets: Foundations for the development of early warning information systems. In *Cyberspace Security and Defense: Research Issues*, pages 231–257. Springer, 2005.

[91] James Riordan, Diego Zamboni, and Yann Duponchel. Building and deploying Billy Goat, a worm-detection system. IBM Zurich Research Laboratory, May 2006.

[92] The Honeynet Project. http://project.honeynet.org/. Last accessed in July 2014.

[93] Internet Storm Center. http://isc.sans.org/. Last accessed in July 2014.

[94] Michael Bailey, Evan Cooke, Tim Battles, and Danny McPherson. Tracking global threats with the Internet Motion Sensor. In *32nd Meeting of the North American Network Operators Group*, 2004.

[95] Masashi Eto, Kotaro Sonoda, Daisuke Inoue, Katsunari Yoshioka, and Koji Nakao. A Proposal of Malware Distinction Method Based on Scan Patterns Using Spectrum Analysis. *Neural Information Processing*, 5864:565–572, 2009.

[96] Arakis project. http://www.arakis.pl/pl. Last accessed in July 2014.

[97] NoAH project. http://www.fp6-noah.org. Last accessed in July 2014.

[98] Spiros Antonatos, Kostas Anagnostakis, and Evangelos Markatos. Honey@home: a new approach to large-scale threat monitoring. In *Proceedings of the ACM workshop on recurring malcode*, pages 38–45. ACM, 2007.

[99] SWITCH Internet Background Noise (IBN). http://www.switch.ch/security/services/IBN/. Last accessed in July 2014.

[100] Police Internet Activities Monitored. http://www.cyberpolice.go.jp/english/obs_e.html. Last accessed in July 2014.

[101] Japan cert coordination center. http://www.jpcert.or.jp/. Last accessed in July 2014.

[102] The IUCC/IDC Internet Telescope. http://noc.ilan.net.il/research/telescope/. Last accessed in July 2014.

[103] Jakub Czyz, Michael Kallitsis, Manaf Gharaibeh, Christos Papadopoulos, Michael Bailey, and Manish Karir. Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS Attacks. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 435–448. ACM, 2014.

[104] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. Inferring Distributed Reflection Denial of Service Attacks from Darknet. *Computer Communications*, 2015.

[105] Tanja Zseby et al. Workshop report: darkspace and unsolicited traffic analysis (DUST 2012). *ACM SIGCOMM Computer Communication Review*, 42(5):49–53, 2012.

[106] Paolo Milani Comparetti, Guido Salvaneschi, Engin Kirda, Clemens Kolbitsch, Christopher Kruegel, and Stefano Zanero. Identifying dormant functionality in malware programs. In *IEEE Symposium on Security and Privacy (SP)*, pages 61–76. IEEE, 2010.

[107] Corrado Leita, Marc Dacier, and Georg Wicherski. SGNET: a distributed infrastructure to handle zero-day exploits. Technical report, Eurecom, 2007.

[108] Georgios Portokalidis, Asia Slowinska, and Herbert Bos. Argos: an emulator for fingerprinting zero-day attacks for advertised honeypots with automatic signature generation. In *proceedings of EuroSys*, pages 15–27, 2006.

[109] Paul Baecher and Markus Koetter and Maximillian Dornseif and Felix Freiling. The Nepenthes Platform: An Efficient Approach to Collect Malware. In *Proceedings of the 9th International Symposium on Recent Advances in Intrusion Detection (RAID)*, pages 165–184, 2006.

[110] Van Horenbeeck, M. The SANS Internet Storm Center. In *proceedings of WOMBAT Workshop on Information Security Threats Data Collection and Sharing (WISTDCS)*, pages 17–23, 2008.

[111] DShield: Community-based collaborative firewall log correlation system. http://www.dshield.org/. Last accessed in July 2014.

[112] Daisuke Inoue, Mio Suzuki, Masashi Eto, Katsunari Yoshioka, and Koji Nakao. DAEDALUS: Novel Application of Large-Scale Darknet Monitoring for Practical Protection of Live Networks. In *Recent Advances in Intrusion Detection*, pages 381–382. Springer, 2009.

[113] Research and Education Networking Information Sharing and Analysis Center. http://www.ren-isac.net/. Last accessed in July 2014.

[114] An introduction to the simwood darknet. http://blog.simwood.com/2011/08/an-introduction-to-the-simwood-darknet/. Last accessed in July 2014.

[115] The darknet mesh project. http://projects.oucs.ox.ac.uk/darknet/. Last accessed in July 2014.

[116] Barry Vivian William Irwin. *A framework for the application of network telescope sensors in a global IP network*. PhD thesis, Rhodes University, 2011.

[117] Barry Irwin. A baseline study of potentially malicious activity across five network telescopes. In *5th International Conference on Cyber Conflict (CyCon)*, pages 1–17, 2013.

[118] Matthew Ford, Jonathan Stevens, and John Ronan. Initial Results from an IPv6 Darknet. In *Proceedings of the IEEE International Conference on Internet Surveillance and Protection (ICISP)*, pages 13–, 2006.

[119] Akihiro SHIMODA, Tatsuya MORI, and Shigeki Goto. Extended Darknet: Multi-Dimensional Internet Threat Monitoring System. *IEICE Transactions on Communications*, pages 1915–1923, 2012.

[120] Alberto Dainotti, Karyn Benson, Alistair King, Michael Kallitsis, Eduard Glatz, Xenofontas Dimitropoulos, et al. Estimating Internet address space usage through passive measurements. *ACM SIGCOMM Computer Communication Review*, pages 42–49, 2013.

[121] Kensuke Fukuda, Toshio Hirotsu, Osamu Akashi, and Toshiharu Sugawara. Correlation among piecewise unwanted traffic time series. In

IEEE Global Telecommunications Conference (GLOBECOM), pages 1–5, 2008.

[122] Kensuke Fukuda, Toshio Hirotsu, Osamu Akashi, and Toshiharu Sugawara. A PCA analysis of daily unwanted traffic. In *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, pages 377–384, 2010.

[123] Jon Oberheide, Manish Karir, and Z.Morley Mao. Characterizing Dark DNS Behavior. In Bernhard M. Hmmerli and Robin Sommer, editors, *Detection of Intrusions and Malware, and Vulnerability Assessment*, volume 4579 of *Lecture Notes in Computer Science*, pages 140–156. Springer Berlin Heidelberg, 2007.

[124] Jakub Czyz, Kyle Lady, Sam G Miller, Michael Bailey, Michael Kallitsis, and Manish Karir. Understanding IPv6 internet background radiation. In *Internet Measurement Conference (IMC)*, pages 105–118, 2013.

[125] Eduard Glatz and Xenofontas Dimitropoulos. Classifying Internet one-way traffic. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*, pages 417–418, 2012.

[126] Ruoyu Wang, Ling Zhang, and Zhen Liu. A novel method of filtering internet background radiation traffic. In *Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*, pages 371–376, 2013.

[127] Bradley Cowie and Barry Irwin. Data classification for artificial intelligence construct training to aid in network incident identification using network telescope data. In *Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*, pages 356–360, 2010.

[128] Yanbing Peng, Jian Gong, Wang Yang, and Weijiang Liu. Disclosing the element distribution of bloom filter. In *Computational Science–ICCS 2006*, pages 1022–1025. Springer, 2006.

[129] Hamza Rahmani, Nabil Sahli, and Farouk Kammoun. Joint entropy analysis model for ddos attack detection. In *IEEE Fifth International Conference on Information Assurance and Security (IAS)*, volume 2, pages 267–271, 2009.

[130] Zubair M Fadlullah, Tarik Taleb, Athanasios V Vasilakos, Mohsen Guizani, and Nei Kato. DTRAB: combating against attacks on encrypted protocols through traffic-feature analysis. *IEEE/ACM Transactions on Networking (TON)*, pages 1234–1247, 2010.

[131] Łukasz Saganowski, Tomasz Andrysiak, Michał Choraś, and Rafał Renk. Expansion of Matching Pursuit Methodology for Anomaly Detection in Computer Networks. In *Computer Recognition Systems 4*, pages 727–736. Springer, 2011.

[132] Michal Choras, Lukasz Saganowski, Rafal Renk, and Witold Holubowicz. Statistical and signal-based network traffic recognition for anomaly detection. *Expert Systems*, pages 232–245, 2012.

[133] Parminder Chhabra, Ajita John, and Huzur Saran. PISA: Automatic Extraction of Traffic Signatures. In *Fourth International Conference in Networking*, pages 730–742, 2005.

[134] Xin He and S Parameswaran. MCAD: Multiple connection based anomaly detection. In *11th IEEE Singapore International Conference on Communication Systems (ICCS)*, pages 999–1004, 2008.

[135] Khaled Dassouki, Herve Debar, Haidar Safa, and Abbas Hijazi. A TCP delay-based mechanism for detecting congestion in the Internet. In *Third International Conference on Communications and Information Technology (ICCIT)*, pages 141–145. IEEE, 2013.

[136] Jerome Francois, Olivier Festor, et al. Tracking global wide configuration errors. In *IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation*, 2006.

[137] Uli Harder, Matt W Johnson, Jeremy T Bradley, and William J Knottenbelt. Observing internet worm and virus attacks with a small network telescope. *Electronic Notes in Theoretical Computer Science*, pages 47–59, 2006.

[138] Thorsten Holz. Learning More About Attack Patterns With Honeypots. In *Proceedings of Sicherheit*, pages 30–41, 2006.

[139] Jérôme François, Radu State, and Olivier Festor. Activity Monitoring for large honeynets and network telescopes. *International Journal On Advances in Systems and Measurements*, 1(1):1–13, 2008.

[140] Masayuki OHTA, Shu SUGIMOTO, Kensuke FUKUDA, Toshio HIROTSU, Osamu AKASHI, and Toshiharu SUGAWARA. Analysis of time-series correlations of packet arrivals to darknet and their size- and location-dependencies. *Computer Software*, 28(2), 2011.

[141] Daisuke Inoue, Katsunari Yoshioka, Masashi Eto, Yuji Hoshizawa, and Koji Nakao. Malware behavior analysis in isolated miniature network for revealing malware's network activity. In *IEEE International Conference on Communications (ICC)*, pages 1715–1721, 2008.

[142] Robin Berthier, Dave Korman, Michel Cukier, Matti Hiltunen, Gregg Vesonder, and Daniel Sheleheda. On the comparison of network attack datasets: An empirical analysis. In *11th IEEE High Assurance Systems Engineering Symposium (HASE)*, pages 39–48, 2008.

[143] Vinod Yegneswaran, Paul Barford, and Johannes Ullrich. Internet intrusions: Global characteristics and prevalence. In *ACM SIGMETRICS Performance Evaluation Review*, pages 138–147, 2003.

[144] R. Rangadurai Karthick, V.P. Hattiwale, and B. Ravindran. Adaptive network intrusion detection system using a hybrid approach. In *Fourth International Conference on Communication Systems and Networks (COMSNETS)*, pages 1–7, 2012.

[145] Paul Barford, Yan Chen, Anup Goyal, Zhichun Li, Vern Paxson, and Vinod Yegneswaran. Employing Honeynets For Network Situational Awareness. In Sushil Jajodia, Peng Liu, Vipin Swarup, and Cliff Wang, editors, *Cyber Situational Awareness*, volume 46 of *Advances in Information Security*, pages 71–102. Springer US, 2010.

[146] Daisuke Inoue, Katsunari Yoshioka, Masashi Eto, Masaya Yamagata, Eisuke Nishino, Junichi Takeuchi, Kazuya Ohkouchi, and Koji Nakao. An incident analysis system nicter and its analysis engines based on data mining techniques. In *Advances in Neuro-Information Processing*, pages 579–586. Springer, 2009.

[147] Olivier Thonnard and Marc Dacier. Actionable knowledge discovery for threats intelligence support using a multi-dimensional data mining methodology. In *IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 154–163, 2008.

[148] Olivier Thonnard and Marc Dacier. A framework for attack patterns' discovery in honeynet data. *Digital Investigation*, pages S128–S139, 2008.

[149] Olivier Thonnard, Wim Mees, and Marc Dacier. Addressing the attack attribution problem using knowledge discovery and multi-criteria fuzzy decision-making. In *Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics*, pages 11–21, 2009.

[150] Sushant Sinha, Michael Bailey, and Farnam Jahanian. One size does not fit all: 10 years of applying context-aware security. In *IEEE Conference on Technologies for Homeland Security (HST)*, pages 14–21, 2009.

[151] Claude Fachkha, Elias Bou-Harb, Amine Boukhetouta, Son Dinh, Farkhund Iqbal, and Mourad Debbabi. Investigating the Dark Cyberspace: Profiling, Threat-Based Analysis and Correlation. In *IEEE Seventh International Conference on Risks and Security of Internet and Systems (CRiSIS)*, pages 1–8, 2012.

[152] Ejaz Ahmed, Andrew Clark, and George Mohay. A novel sliding window based change detection algorithm for asymmetric traffic. In *IFIP International Conference on Network and Parallel Computing (NPC)*, pages 168–175. IEEE, 2008.

[153] Ejaz Ahmed, Andrew Clark, and George Mohay. Effective change detection in large repositories of unsolicited traffic. In *Fourth International Conference on Internet Monitoring and Protection (ICIMP)*, pages 1–6. IEEE, 2009.

[154] Wenji Chen, Yang Liu, and Yong Guan. Cardinality change-based early detection of large-scale cyber-attacks. In *Proceedings IEEE INFOCOM*, pages 1788–1796, 2013.

[155] Sohraab Soltani, Syed Ali Khayam, and Hayder Radha. Detecting malware outbreaks using a statistical model of blackhole traffic. In *IEEE International Conference on Communications (ICC)*, pages 1593–1597, 2008.

[156] Martin Casado, Tal Garfinkel, Weidong Cui, Vern Paxson, and Stefan Savage. Opportunistic measurement: Extracting insight from spurious traffic. In *Proceedings of the 4th ACM Workshop on Hot Topics in Networks (Hotnets-IV)*, 2005.

[157] Andrew Clark, Marc Dacier, George Mohay, Fabien Pouget, and Jakub Zimmermann. Internet attack knowledge discovery via clusters and cliques of attack traces. *Journal of Information Assurance and Security*, 1(1):21–32, 2006.

[158] Samuel Oswald Hunter, Barry Irwin, and Etienne Stalmans. Real-time distributed malicious traffic monitoring for honeypots and network telescopes. In *Information Security for South Africa, 2013*, pages 1–9. IEEE, 2013.

[159] Rajeev Gupta, Krithi Ramamritham, and Mukesh Mohania. Ratio threshold queries over distributed data sources. *Proceedings of the VLDB Endowment*, 6(8):565–576, 2013.

[160] Evan Cooke, Zhuoqing Morley Mao, and Farnam Jahanian. Hotspots: The root causes of non-uniformity in self-propagating malware. In *IEEE International Conference on Dependable Systems and Networks (DSN)*, pages 179–188, 2006.

[161] Hongbin Luo, Yi Lin, Hongke Zhang, and Moshe Zukerman. Preventing DDoS attacks by identifier/locator separation. *IEEE Network*, pages 60–65, 2013.

[162] Mar Callau-Zori, Ricardo Jiménez-Peris, Vincenzo Gulisano, Marina

Papatriantafilou, Zhang Fu, and Marta Patiño-Martínez. STONE: a stream-based DDoS defense framework. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 807–812, 2013.

[163] P. Arun Raj Kumar and S. Selvakumar. Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems. *Computer Communications*, 36(3):303 – 319, 2013.

[164] Sajal Bhatia, Desmond Schmidt, and George Mohay. Ensemble-based ddos detection and mitigation model. In *Proceedings of the Fifth International Conference on Security of Information and Networks (SIN)*, pages 79–86, 2012.

[165] Sajal Bhatia, George Mohay, Alan Tickle, and Ejaz Ahmed. Parametric differences between a real-world distributed denial-of-service attack and a flash event. In *Sixth International Conference on Availability, Reliability and Security (ARES)*, pages 210–217, 2011.

[166] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. Towards a forecasting model for distributed denial of service activities. In *Network Computing and Applications (NCA), 2013 12th IEEE International Symposium on*, pages 110–117. IEEE, 2013.

[167] Fachkha, Claude and Bou-Harb, Elias and Debbabi, Mourad. On the inference and prediction of DDoS campaigns. *Wireless Communications and Mobile Computing, Wiley*, 2014.

[168] Eduardo Feitosa, Eduardo Souto, and Djamel H. Sadok. An orchestration approach for unwanted internet traffic identification. *Computer Networks*, 56(12):2805 – 2831, 2012.

[169] Hamza Rahmani, Nabil Sahli, and Farouk Kamoun. DDoS flooding attack detection scheme based on F-divergence. *Computer Communications*, 35(11):1380 – 1391, 2012.

[170] Haiqin Liu, Yan Sun, V.C. Valgenti, and Min Sik Kim. Trustguard: A flow-level reputation-based ddos defense system. In *Consumer Communications and Networking Conference (CCNC), IEEE*, pages 287–291, 2011.

[171] Ejaz Ahmed, George Mohay, Alan Tickle, and Sajal Bhatia. Use of IP Addresses for High Rate Flooding Attack Detection. In Kai Rannenberg, Vijay Varadharajan, and Christian Weber, editors, *Security and Privacy Silver Linings in the Cloud*, volume 330 of *IFIP Advances in Information and Communication Technology*, pages 124–135. Springer Berlin Heidelberg, 2010.

[172] David Moore, Colleen Shannon, et al. Code-red: a case study on the spread and victims of an internet worm. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 273–284. ACM, 2002.

[173] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. Inside the slammer worm. *IEEE Security & Privacy*, pages 33–39, 2003.

[174] Vincent H Berk, Robert S Gray, and George Bakos. Using sensor networks and data fusion for early detection of active worms. In *AeroSense 2003*, pages 92–104. International Society for Optics and Photonics, 2003.

[175] Stuart Staniford, David Moore, Vern Paxson, and Nicholas Weaver. The top speed of flash worms. In *Proceedings of the ACM workshop on Rapid malcode*, pages 33–42, 2004.

[176] Evan Cooke, Z Morley Mao, and Farnam Jahanian. Worm hotspots: Explaining non-uniformity in worm targeting behavior. *University of Michigan TR*, 2004.

[177] David W. Richardson, Steven D. Gribble, and Edward D. Lazowska. The limits of global scanning worm detectors in the presence of background noise. In *Proceedings of the ACM workshop on Rapid malcode (WORM)*, pages 60–70, 2005.

[178] Abhishek Kumar, Vern Paxson, and Nicholas Weaver. Exploiting underlying structure for detailed reconstruction of an internet-scale event. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement (IMC)*, pages 33–33, Berkeley, CA, USA, 2005. USENIX Association.

[179] Moheeb Abu Rajab, Fabian Monrose, and Andreas Terzis. Worm evolution tracking via timing analysis. In *Proceedings of the 2005 ACM workshop on Rapid malcode*, WORM '05, pages 52–59, New York, NY, USA, 2005. ACM.

[180] Cliff C Zou, Weibo Gong, Don Towsley, and Lixin Gao. The monitoring and early detection of internet worms. *IEEE/ACM Transactions on Networking (TON)*, 13(5):961–974, 2005.

[181] Van-Hau Pham, Marc Dacier, Guillaume Urvoy-Keller, and Taoufik En-Najjary. The quest for multi-headed worms. In *Proceedings of the 5th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, pages 247–266, 2008.

[182] Yoshiki Kanda, Kensuke Fukuda, and Toshiharu Sugawara. A flow analysis for mining traffic anomalies. In *IEEE International Conference on Communications (ICC)*, pages 1–5, 2010.

[183] Qian Wang, Zesheng Chen, Kia Makki, Niki Pissinou, and Chao Chen. Inferring internet worm temporal characteristics. In *GLOBECOM*, pages 2007–2012, 2008.

[184] Qian Wang, Zesheng Chen, and Chao Chen. Darknet-based inference of internet worm temporal characteristics. *IEEE Transactions on Information Forensics and Security*, 6(4):1382–93, December 2011.

[185] Zhen Chen, Chuang Lin, Jia Ni, Dong-Hua Ruan, Bo Zheng, Yi-Xin Jiang, Xue-Hai Peng, Yang Wang, An-an Luo, Bing Zhu, et al. Anti-Worm NPU-based parallel bloom filters for TCP/IP content processing in Giga-Ethernet LAN. In *The IEEE Conference on Local Computer Networks*, pages 748–755, 2005.

[186] Cliff Changchun Zou, Lixin Gao, Weibo Gong, and Don Towsley. Monitoring and early warning for internet worms. In *Proceedings of the 10th ACM conference on Computer and communications security (CCS)*, pages 190–199, NY, USA, 2003.

[187] David Dagon, Cliff Changchun Zou, and Wenke Lee. Modeling Botnet Propagation Using Time Zones. In *NDSS*, volume 6, pages 2–13, 2006.

[188] Van-Hau Pham and Marc Dacier. Honeypot traces forensics : the observation view point matters. Technical Report EURECOM+2697, Eurecom, 2009.

[189] Ping Wang, Sherri Sparks, and Cliff Changchun Zou. An advanced hybrid peer-to-peer botnet. *IEEE Transactions on Dependable and Secure Computing*, 7(2):113–127, 2010.

[190] Evan Cooke, Farnam Jahanian, and Danny McPherson. The Zombie roundup: understanding, detecting, and disrupting botnets. In *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI)*, pages 6–6, Berkeley, CA, USA, 2005. USENIX Association.

[191] Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. Botminer: clustering analysis of network traffic for protocol- and structure-independent botnet detection. In *Proceedings of the 17th conference on Security Symposium (SS)*, pages 139–154, Berkeley, CA, USA, 2008. USENIX Association.

[192] Anirudh Ramachandran, Nick Feamster, and David Dagon. Revealing botnet membership using dnsbl counter-intelligence. *Proc. 2nd USENIX Steps to Reducing Unwanted Traffic on the Internet*, pages 49–54, 2006.

[193] Anirudh Ramachandran and Nick Feamster. Understanding the network-level behavior of spammers. In *ACM SIGCOMM Computer Communication Review*, pages 291–302. ACM, 2006.

[194] Guofei Gu, Phillip A Porras, Vinod Yegneswaran, Martin W Fong, and Wenke Lee. Bothunter: Detecting malware infection through ids-driven dialog correlation. In *USENIX Security*, volume 7, pages 1–16, 2007.

[195] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. A statistical approach for fingerprinting probing activities. In *Eighth International Conference on Availability, Reliability and Security (ARES)*, pages 21–30, 2013.

[196] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. A systematic approach for detecting and clustering distributed cyber scanning. *Computer Networks*, 57(18):3826–3839, 2013.

[197] Zhichun Li, Anup Goyal, Yan Chen, and Vern Paxson. Automating analysis of large-scale botnet probing events. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security*, pages 11–22, 2009.

[198] Alberto Dainotti, Alistair King, and Kimberly Claffy. Analysis of Internet-wide Probing using Darknets. In *Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*, 2012.

[199] Guofei Gu, Zesheng Chen, P. Porras, and Wenke Lee. Misleading and defeating importance-scanning malware propagation. In *Third International Conference on Security and Privacy in Communications Networks. SecureComm.*, pages 250–259, 2007.

[200] Zhichun Li, Anup Goyal, and Yan Chen. Honeynet-based botnet scan traffic analysis. In *Botnet Detection*, pages 25–44. Springer, 2008.

[201] Masashi Eto, Daisuke Inoue, Mio Suzuki, and Koji Nakao. A Statistical Packet Inspection for Extraction of Spoofed IP Packets on Darknet. In *Proceedings of the Joint Workshop on Information Security, Kaohsiung, Taiwan*, 2009.

[202] M. Ohta, Y. Kanda, K. Fukuda, and T. Sugawara. Analysis of Spoofed IP Traffic Using Time-to-Live and Identification Fields in IP Headers. In *IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA)*, pages 355 –361, 2011.

[203] Jun Bi, Ping Hu, and Peiguo Li. Study on Classification and Characteristics of Source Address Spoofing Attacks in the Internet. In *Proceedings of the Ninth International Conference on Networks (ICN)*, pages 226–230, Washington, DC, USA, 2010. IEEE Computer Society.

[204] Guang Yao, Jun Bi, and Zijian Zhou. Passive ip traceback: capturing

the origin of anonymous traffic through network telescopes. *SIGCOMM Comput. Commun. Rev.*, 41(4):–, August 2010.

[205] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 1–18. ACM, 2011.

[206] Michael Bailey and Craig Labovitz. Censorship and Co-option of the Internet Infrastructure. Technical Report CSE-TR-572-11, University of Michigan, Ann Arbor, MI, USA, July 2011.

[207] Karyn Benson, Alberto Dainotti, KC Claffy, and Emile Aben. Gaining insight into as-level outages through analysis of internet background radiation. In *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 447–452, 2013.

[208] Lin Quan, John Heidemann, and Yuri Pradkin. Trinocular: understanding internet reliability through adaptive probing. In *Proceedings of the ACM SIGCOMM conference on SIGCOMM*, pages 255–266. ACM, 2013.

[209] Janne Riihijarvi, Petri Mahonen, and Matthias Wellens. Metrics for characterizing complexity of network traffic. In *IEEE International Conference on Telecommunications (ICT)*, pages 1–6, 2008.

[210] J Riihijarvi, Matthias Wellens, and P Mahonen. Measuring complexity and predictability in networks with multiscale entropy analysis. In *IEEE INFOCOM*, pages 1107–1115, 2009.

[211] Mallat, Stéphane G and Zhang, Zhifeng. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[212] Corrado Leita and Ken Mermoud and Marc Dacier. ScriptGen: an Automated Script Generation Tool for honeyd. In *proceedings of ACSAC*, pages 203–214, 2005.

[213] Nasser Abouzakhar and Abu Bakar. A chi-square testing-based intrusion detection model. In *Proceedings of the 4th International Conference on Cybercrime Forensics Education & Training*, 2010.

[214] Nasser Abouzakhar, Huankai Chen, and Bruce Christianson. An Enhanced Fuzzy ARM Approach for Intrusion Detection. *IJDCF*, pages 41–61, 2011.

[215] Tomasz Andrysiak, Łukasz Saganowski, and Michał Choraś. DDoS Attacks Detection by Means of Greedy Algorithms. In *Image Processing and Communications Challenges*, pages 303–310. Springer, 2013.

[216] MAWI Working Group Traffic Archive. http://mawi.wide.ad.jp/mawi/. Last accessed in July 2014.

[217] Marc Kührer, Thomas Hupperich, Christian Rossow, and Thorsten Holz. Hell of a handshake: abusing TCP for reflective amplification DDoS attacks. In *USENIX Workshop on Offensive Technologies (WOOT)*, 2014.

[218] Do Quoc Le, Taeyoel Jeong, H. Eduardo Roman, and James Won-Ki Hong. Traffic dispersion graph based anomaly detection. In *Proceedings of the Second Symposium on Information and Communication Technology*, SoICT, pages 36–41, New York, NY, USA, 2011. ACM.

[219] Cliff Joslyn, Sutanay Choudhury, David Haglin, Bill Howe, Bill Nickless, and Bryan Olsen. Massive scale cyber traffic analysis: a driver for graph database research. In *First International Workshop on Graph Data Management Experiences and Systems*, page 3. ACM, 2013.

[220] Krasser et al. Real-time and forensic network data analysis using animated and coordinated visualization. In *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop (IAW)*, pages 42–49, 2005.

[221] Jean-Pierre van Riel and Barry Irwin. InetVis, a visual tool for network telescope traffic analysis. In *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 85–89. ACM, 2006.

[222] Barry Irwin and J-P van Riel. Using inetvis to evaluate snort and bro scan detection on a network telescope. In *VizSEC 2007*, pages 255–273. Springer, 2008.

[223] Harrop et al. Real-time collaborative network monitoring and control using 3d game engines for representation and interaction. In *Proceedings of the 3rd international workshop on Visualization for computer security (VizSEC)*, pages 31–40. ACM, 2006.

[224] Warren Harrop and Grenville Armitage. Modifying first person shooter games to perform real time network monitoring and control tasks. In *Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games (NetGames)*, New York, NY, USA, 2006. ACM.

[225] Lucas Parry. L3dgeworld 2.3 input & output specifications. *Centre for Advanced Internet Architectures, Swinburne University of Technology*, 80222:22, 2008.

[226] Kensuke Fukuda and Romain Fontugne. Estimating speed of scanning activities with a hough transform. In *IEEE International Conference on Communications (ICC)*, IEEE International Conference on Communications, 2010.

[227] Romain Fontugne, Toshio Hirotsu, and Kensuke Fukuda. An image processing approach to traffic anomaly detection. In *Proceedings of the 4th Asian Conference on Internet Engineering*, pages 17–26. ACM, 2008.

[228] Fontugne et al. A visualization tool for exploring multi-scale network traffic anomalies. In *IEEE International Symposium on Performance Evaluation of Computer & Telecommunication Systems (SPECTS)*, volume 41, pages 274–281, 2009.

[229] Hyunsang Choi, Heejo Lee, and Hyogon Kim. Fast detection and visualization of network attacks on parallel coordinates. *Computers and Security*, 28(5):276 – 288, 2009.

[230] Barry Irwin and Nick Pilkington. High level Internet scale traffic visualization using hilbert curve mapping. In *VizSEC 2007*, pages 147–158. Springer, 2008.

[231] CAIDA. Caida visualization tools. https://www.caida.org/tools/visualization/. Last accessed in September 2015.

[232] Marcelo Bagnulo, Philip Matthews, and Iljitsch van Beijnum. Stateful NAT64: Network address and protocol translation from IPv6 clients to IPv4 servers. *IETF, April*, pages 2070–1721, 2011.

[233] Richard M Felder and Linda K Silverman. Learning and teaching styles in engineering education. *Engineering education*, 78(7):674–681, 1988.

[234] LOIC. http://sourceforge.net/projects/loic/. Last accessed in July 2014.

Claude Fachkha is a cyber security scientist and R&D professional. He earned a Bachelor of Engineering in Computer and Communication from Notre Dame University in 2008, then enrolled at Concordia University Canada and completed a Master of Engineering and PhD in Electrical and Computer Engineering with high distinction. Furthermore, Claude holds a certificate in university teaching and has published several research papers in international and Tier A venues. He has has been awarded tens of bursaries and honors including international recognitions, community participation and best paper. In 2013, Claude received the prestigious government (FRQNT) award. In 2015, Claude was ranked among the top 10 researchers by NSERC Canada for his postdoctoral dossier. Claude is affiliated with NCFTA Canada, where he acts as a data scientist and secretary. His research focuses on big data analytics, data mining and cyber intelligence generation.

Mourad Debbabi is a Full Professor at the Concordia Institute for Information Systems Engineering. He holds the Concordia Research Chair Tier I in Information Systems Security. He is also the President of NCFTA Canada. He is the founder and one of the leaders of the Computer Security Laboratory at Concordia University. In the past, he was the Specification Lead of four Standard JAIN (Java Intelligent Networks) Java Specification Requests dedicated to the elaboration of standard specifications for presence and instant messaging. Dr. Debbabi holds Ph.D. and M.Sc. degrees in computer science from Paris-XI Orsay, University, France. He published 2 books and more than 230 research papers in journals and conferences on computer security, cyber forensics, privacy, cryptographic protocols, threat intelligence generation, malware analysis, reverse engineering, specification and verification of safety-critical systems, formal methods, Java security and acceleration, programming languages and type theory. He supervised to successful completion 20 Ph.D. students and more than 60 Master students. He served as a Senior Scientist at the Panasonic Information and Network Technologies Laboratory, Princeton, New Jersey, USA; Associate Professor at the Computer Science Department of Laval University, Canada; Senior Scientist at General Electric Research Center, New York, USA; Research Associate at the Computer Science Department of Stanford University, California, USA; and Permanent Researcher at the Bull Corporate Research Center, Paris, France.