Implications of Theoretic Derivations on Empirical Passive Measurements for Effective Cyber Threat Intelligence Generation

Morteza Safaei Pour and Elias Bou-Harb
Cyber Threat Intelligence Laboratory
College of Engineering & Computer Science
Florida Atlantic University, Florida, USA
(msafaeipour2017, ebouharb)@fau.edu

***Abstract*—Cyber space continues to be threatened by various debilitating attacks. In this context, executing passive measurements by analyzing Internet-scale, one-way darknet traffic has proven to be an effective approach to shed the light on Internet-wide maliciousness. While typically such measurements are solely conducted from the empirical perspective on already deployed darknet IP spaces using off-the-shelf Intrusion Detection Systems (IDS), their multidimensional theoretical foundations, relations and implications continue to be obscured. In this paper, we take a first step towards comprehending the relation between attackers' behaviors, the width of the darknet vantage points, the probability of detection and the minimum detection time. We perform stochastic modeling, derivation, validation, inter-correlation and analysis of such parameters to provide numerous insightful inferences, such as the most effective IDS and the most suitable darknet IP space, given various attackers' activities in the presence of detection time/probability constraints. One of the outcomes suggests that the widely-deployed Bro IDS is ideal for inferring slow, stealthy probing activities by leveraging passive measurements. Further, the results do not recommend deploying the Snort IDS when the available darknet IP space is relatively small, which is a typical scenario when darknets are operated and employed on organizational sub-networks. We concur that the generated derivations and mathematical relations put forward a first-of-a-kind formal and an accurate characterization of darknet-centric notions, which possess significant implications on Internet and passive measurements. This is especially factual with the advent of evolving paradigms such as IPv6 deployments and the proliferation of highly-distributed, orchestrated, large-scale and stealthy probing botnets.**

***Index Terms*—Probing activities, Stochastic analysis, Darknet traffic, Data analytics**

I. Introduction

Cyber space has radically altered our every day life and have impacted a large number of its crucial aspects. This is clearly realized nowadays with the large-scale adoption of the Internet-of-Things (IoT) paradigm [1], the modernization of Cyber-Physical Systems (CPS) [2] and the continuous rise and utilization of digital currencies [3], to name a few. Nevertheless, the increasing dependence on cyber space continues to make organizations and Internet-wide services highly vulnerable to targeted threats and exploitations. In an attempt to thwart such malicious attempts, typically, Intrusion Detection Systems (IDS) are often configured, deployed and managed. Complementary, in recent years, security operators and researchers have become increasingly interested in passive monitoring of unused Internet address spaces, which is often known as darknets or network telescopes [4]. A darknet is a collection of routable, allocated yet unused Internet Protocol (IP) addresses. These IP addresses have no interaction with other hosts and only passively gather packets without generating any replies. Since these unused address blocks contain no legitimate hosts, the received packets are characteristically unsolicited and are often the results of Internet-scale probing activities [5], backscattered packets from victims of denial of service attacks [6] or misconfiguration traffic [7].

As noted above, one of the most prevalent darknet traffic type is related to probing activities. Such activities are indeed a first step and an enabler of a large number of cyber attacks. For instance, autonomous spreading worms [8] employ probing to fingerprint other vulnerable hosts to infect them. Botmasters, orchestrating large-scale botnets [9], adopt probing activities to identify and add more bots to their campaigns [10]. Very recently, the IoT-centric malware Mirai was inferred to be generating a momentous amount of probing activities in an attempt to exploit Internet-facing IoT cameras and video recorders [11]. To this end, promptly detecting such probing activities often aids in preventing actual attacks from occurring or at least contributes in limiting the expansion of botnets. In this context, a darknet has recurrently proven its capability to infer probing activities by analyzing incoming packets to unused IP addresses [12].

While a plethora of research contributions have been conducted on passive detection methods and the practical implementations of darknets [4], nevertheless, to the best of the authors' knowledge, the research effort which endeavors to theoretically derive and analyze darknet-specific notions in the context of vantage points, IDS operating on such darknet IP spaces, and attackers' behaviors, among various others,

have never been attempted before. Indeed, the lack of such formal understanding hinders the optimized deployments and usage of the dakrnet IP space in a given network subnet. Further, without such formal analysis, one can not determine the best IDS to leverage, given a certain attacker's behavior and the available network resources. Additionally, given the proliferation of evolving cyber events such as large-scale, stealthy probing botnets [13], one ought to possess a formal grasp of the available passive measurement strategies and inference mechanisms coupled with their implications in order to select the most suitable approach to employ against these ever-evolving phenomena. Additionally, with the continuous deployment of IPv6, one needs to comprehend the implications of passive measurements in such deployment settings, given an operated IDS and certain requirements on detection time and probability.

Having identified the aforementioned research gap, we frame the contributions of this paper as follows:

- Formalizing the operations of three, widely-deployed detection mechanisms and open source IDS by focusing on their probing detection modus operandi when operated on the darknet IP space.
- For each of the formalized detection approaches, we perform stochastic modeling, derivation and validation of their detection probabilities, their minimum detection time and the minimum number of required darknet IP addresses to achieve a certain detection promptness and accuracy, when conducting passive measurements.
- Executing several insightful experimentation, shedding the light on the impact of detection time, given a certain probing rate and a particular width of the darknet vantage points, while maximizing the detection probability. Moreover, we discuss several implications and provide a number of suggestions that are deduced from the proposed passive measurements formalization scheme, in the context of stealthy probing activities and IPv6 deployments.

The road-map of this paper is as follows. In the next section, we review the literature on various topics such as probing events, darknet as a means of probing detection and stochastic analysis of scanning behavior. In Section III, we formally define the considered detection systems and other required preliminaries. To this end, we also present the stochastic derivation, validation and analysis of the defined detection systems in the context of detection probability and time. In Section IV, we execute, compare and contrast several experimentation by leveraging the proposed formalization scheme. Subsequently, in Section V, we discuss the implications of some of the results on today's cyber security and Internet measurement challenges. Finally, in Section VI, we summarize the contributions of this work and pinpoint several topics that aim at paving the way for future work.

II. Related Work

Since probing activities play an important role in cyber security and Internet measurements, it has been the focus of attention in many research contributions. In [5], the authors provide an extensive survey in which they categorize the scanning topic based on their natures, strategies and approaches. In [14], the authors analyze data from a large darknet to study Internet-wide probing activities. Other research work have been dedicated to studying the impact of reducing the number of utilized darknet IP addresses (i.e., the width of darknet vantage points). For instance, in [15], the authors introduce the concept of sparse darknet, a network subnet that is sparsely populated with darknet addresses, as a way to study the impact of this reduction on its effectiveness. Alternatively, other literature approaches analyze effective sensor placement strategies such as distributed darknet IP address placement [16], considering placing such IP addresses near live hosts or analyzing the impact of special patterns of localization [17]. Leonard et al. [18] performed stochastic derivation of a number of relations in order to propose an optimal stealth distribution scanning activity based on the probability of detection. The authors undertook the attackers' perspective (and not the measurement point of view) in order to significantly minimize the probability of detection.
In contrast, we present a first attempt ever which exploits darknet-specific parameters and variables to formally comprehend the multidimensional relations between darknet vantage points, various IDS operating on such darknet IP spaces, the rate of the probing activities and the detection time/probability. The proposed formalization passive measurement scheme aims at laying the theoretic foundation for the field of Internet passive measurements for cyber security by putting forward such formalizations. We hope that this work would initiate much-needed discussions related to the implications of theoretical models on practical passive measurements in the short and long terms.

III. Formal Modeling and Stochastic Analysis

The purpose of this section is to formalize and define various kinds of probing IDS with the aim of finding relations between different parameters such as minimum number of required darknet IP addresses, minimum detection time and the probability of detection for different scanning rates. Given a subnet S, consisting of $|S|$ IP addresses, we notate the set of darknet IP addresses, randomly distributed within S and utilized in the detection process, as DIP. In this paper, following the natural behavior of large-scale probing events [13], we consider that the attacker intends to scan all IP addresses in S.
Indeed, there exists various scanning patterns such as sequential and uniform probing, which are typically employed for scanning Internet networks. In this context, we note the average scanning rate r and the average inter-probe delay $\frac{1}{r}$. As long as we assume that the darknet IP addresses are distributed uniformly in the intended network, which is the

de-facto practice [4], there is no difference between sequential and uniform probing. Thus, we consider the uniformity of the scans, which indicates that, on average, every $\frac{1}{r}$, the scanner would send a packet to an IP address in the analyzed network. Performing the stochastic modeling of the detection systems would be the first required step for comprehending and analyzing the detection probability and the relations between various passive measurement parameters. For the sake of this work, we focus on three different detection rules based on well-known, highly-deployed IDS.

Definition 1: $\rho(\tau)$ is the probability of detecting a probing activity X in less than τ time units from the start of the scan.

$$\rho(\tau) = \int_0^{\tau} Pr(alarm(t) = TRUE)dt \tag{1}$$

Definition 2: For a probing activity X with an average scanning rate r over a subnet S and given a certain Detection System (DS), the minimum detection time τ^{ϵ}_{min} is the minimum required time for DS to detect the scan with probability more than $1 - \epsilon$.

$$\tau^{\epsilon}_{min} = inf\{t \geq 0 : \rho(t) \geq 1 - \epsilon\} \tag{2}$$

A. FH Detection System

The first considered detection method is the First-Hit (FH) algorithm, which raises an alarm on the detection of the very first probe. Indeed, this represents the simplest detection system that we analyze here to specify some bounds on the parameters. After the first hit, this DS raises the alarm. Algorithm 1 shows the simplistic DS_{FH} algorithm. This method intuitively uses the lowest amount of memory and processing requirements for detection. While this approach might be effective, it undoubtedly could lead to a high false positive rate; it might identify received darknet packets caused by backscattered activities or misconfiguration as probing activities. Thus, we consider this technique and its detection time/probability as a reference model rather than a DS that can actually be operated in practice.

Algorithm 1: First-hit detection algorithm

```
C_i(0) = 0;
alarm = FALSE;
while do
    if A packet from souce i is received then
        C_i(t) = 1;
        alarm = TRUE;
    end
end
```

Recall that there exists $|DIP|$ darknet IP addresses in the subnet S. Thus, the probability of one of these darknet IP addresses being hit by probing packets is $q = \frac{|DIP|}{|S|}$. Therefore, the effective rate λ, the scanning rate that would actually be sensed by the darknet, would be $\lambda = qr$. Now given an average scanning rate r, we can write $\rho(\tau)$ as in

$$\rho(\tau) = 1 - e^{-\lambda\tau} \tag{3}$$

Based on (2) and some mathematical operations, we can easily derive τ^{ϵ}_{min} from equation (3), as in

$$\tau^{\epsilon}_{min} = \frac{\log(\epsilon)}{-\lambda} \tag{4}$$

Further, we can infer the minimum required darknet IP addresses for specific τ^{ϵ}_{min} and ϵ, as follows.

$$\min |DIP| = \frac{|S| \ln(\epsilon)}{-r\tau^{\epsilon}_{min}} \tag{5}$$

B. DSI Detection System

The second detection method is a window-based detection technique that is based on the widely-deployed, open source Snort [19] IDS. We refer to this detection system as DSI and subsequently describe its operations. Consider a counter $C_i(t) = 0$ for each observed source IP address i. After its reset (at time t), it starts counting received packets in a time window $[t, t + \Delta_{DSI}]$. During this time window, if the counter hits the threshold α_{DSI}, DSI raises an alarm, otherwise, the counter and the time window will be re-initiated. This algorithm is clearly more complex than the FH algorithm because it requires a timer to check the window's timeout and thus memory is required for storing $C_i(t)$ for all packets arriving from different source IP addresses i. The operations of DSI is summarized in Algorithm 2.

Algorithm 2: DSI detection algorithm

```
C_i(0) = 0;
alarm = FALSE;
t_reset = 0;
while do
    if t ≤ t_reset + Δ_DSI then
        if A packet from souce i is received then
            C_i(t) = C_i(t) + 1;
            if C_i(t) ≥ α_DSI then
                alarm = TRUE;
            end
        end
    else
        t_reset = t_reset + Δ_DSI;
        C_i(t) = 0;
    end
end
```

The detection system DSI is defined with the parameter pair $(\Delta_{DSI}, \alpha_{DSI})$. Let $\tau = p\Delta_{DSI} + \nu$ where $0 \leq \nu \leq \Delta_{DSI}$, then we can compute $\rho(\tau)$, as follows.

$$\rho(\tau) = 1 - W_0^p W_1 \tag{6}$$

where

$$W_0 = e^{-\lambda\Delta} \sum_{k=0}^{\alpha-1} \frac{(\lambda\Delta)^k}{k!} = \frac{\Gamma(\alpha, \lambda\Delta)}{(\alpha - 1)!} \tag{7}$$

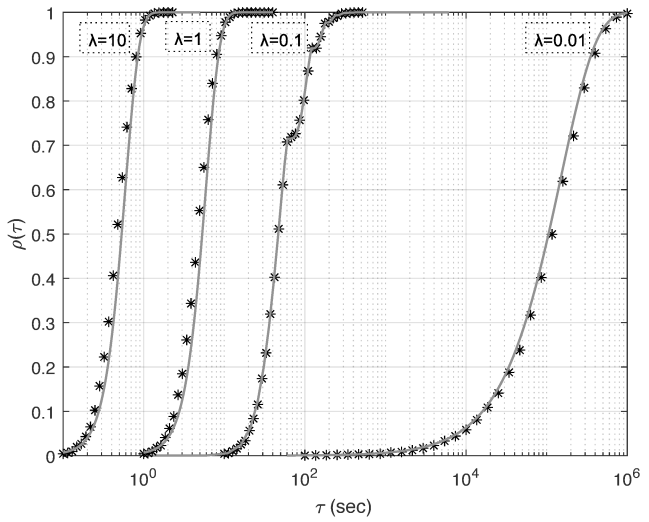

Fig. 1. Validating the accuracy of the relation derived in (6) against simulation results (marked with asterisks) for DSI.

Algorithm 3: DSII detection algorithm

```
C_i(0) = 0;
alarm = FALSE;
t_reset = 0;
while do
    if t ≤ t_reset + Δ_DSI then
        if A packet from souce i is received then
            C_i(t) = C_i(t) + 1;
            t_reset = t;
            if C_i(t) ≥ α_DSI then
                alarm = TRUE;
            end
        end
    else
        t_reset = t;
        C_i(t) = 0;
    end
end
```

and

$$W_1 = e^{-\lambda\nu} \sum_{k=0}^{\alpha-1} \frac{(\lambda\nu)^k}{k!} = \frac{\Gamma(\alpha, \lambda\nu)}{(\alpha-1)!} \tag{8}$$

where $\Gamma(\alpha, x)$ is the upper incomplete gamma function. We use the probability of events for a Poisson distribution, which is a typical distribution observed for malicious packets targeting the darknet IP space [20], to derive W_0 and W_1. The default values of $(\Delta_{DSI}, \alpha_{DSI})$ for the Snort IDS are $(60, 5)$.

We validate the accuracy of the formulation in Figure 1, which shows the derived relation in (6) against executed simulation results for DSI. For different values of λ, we can note a close to perfect accuracy, which corroborates the soundness of the derived relation.

C. DSII Detection System

The third inference method is also a window-based DS related to the well-known Paxon's Bro IDS [21]. We use DSII to refer to this detection method. In such an IDS, for each source IP address i, a counter $C_i(t)$ is created. After receiving a packet from source host i at time t, the technique will wait Δ_{DSII} time unit to receive another packet. In case a packet hit the detection system during $[t, t+\Delta_{DSII}]$, the IDS will increment $C_i(t)$; otherwise it will reset $C_i(t)$. Algorithm 3 summarizes the modus operandi embedded within DSII.

$\rho(\tau)$ for DSII can be calculated based on (9), where $p_\alpha(t)$ is the probability of $Pr(Alarm(t) = True)$ for DSII with threshold α. To this end, we compute the Probability Distribution Function (PDF) of DSII with parameter α, recursively, based on the PDF of DSII with threshold $\alpha-1$. Consequently, we can derive the CDF which in fact refers to $\rho(\tau)$.

$$p_\alpha(t) = \begin{cases} \frac{1}{A}\int_{x=0}^{\Delta_{DSII}} p_{\alpha-1}(x)\lambda e^{-\lambda(t-x)}dx, & \text{if } t \geq \Delta_{DSII} \\ \frac{1}{A}\int_{x=0}^{t} p_{\alpha-1}(x)\lambda e^{-\lambda(t-x)}dx, & \text{if } t < \Delta_{DSII} \end{cases} \tag{9}$$

where $A = (1-e^{-\lambda\Delta_{DSII}})$. Equation (9) can be shown with convolution operator as in (10). We employ Laplace Transform for calculating these recursive convolutions.

$$p_\alpha(t) = p_{\alpha-1}(t) * \frac{1}{A} g(t) = p_{\alpha-1}(t) * \frac{1}{A}\lambda e^{-\lambda t}(u(t) - u(t-\Delta)) \tag{10}$$

Therefore, $p_1(t) = \lambda e^{-\lambda t}u(t) \xrightarrow{S\ Transform} P_1(s) = \frac{\lambda}{s+\lambda}$ and $g(t) = \lambda e^{-\lambda t}(u(t) - u(t-\Delta)) \xrightarrow{S\ Transform} G(s) = \frac{\lambda}{s+\lambda}(1 - e^{-\Delta(s+\lambda)})$. We know that in S-Transform, we have the relation $f(t) * g(t) \leftrightarrow F(s)G(s)$. Thus, we can rewrite (10) as in (11).

$$\begin{aligned} P_\alpha(s) &= \frac{1}{A}P_{\alpha-1}(s)G(s) = \frac{1}{A^{\alpha-1}}P_1(s)G^{\alpha-1}(s) \\ &= \frac{1}{A^{\alpha-1}}(\frac{\lambda}{s+\lambda})^\alpha(1 - e^{-\Delta(s+\lambda)})^{\alpha-1} \\ &= \frac{1}{A^{\alpha-1}}(\frac{\lambda}{s+\lambda})^\alpha(\sum_{k=0}^{\alpha-1}(-1)^k \binom{\alpha-1}{k} e^{-k\Delta(s+\lambda)}) \end{aligned} \tag{11}$$

Inverse Laplace Transform of (11) can be calculated and the result of (12) would be the PDF of the detection at time

t. Now, we transfer the equation to the time domain, as in:

$$p_\alpha(t) = \frac{\lambda^\alpha e^{-\lambda t}}{A^{\alpha-1}(\alpha-1)!}\sum_{k=0}^{\alpha-1}(-1)^k\binom{\alpha-1}{k}(t-k\Delta)^{\alpha-1}u(t-k\Delta) \tag{12}$$

If we define the integral of the first term of (12) as in $X_0^\tau(t) = \frac{1}{A^{\alpha-1}}\int_0^\tau \frac{\lambda^\alpha}{(\alpha-1)!}t^\alpha e^{-\lambda t}u(t)dt$, then

$$\begin{aligned}\rho(\tau) = \int_0^\tau p_\alpha(t)dt &= \sum_{k=0}^{\alpha-1}(-1)^k\binom{\alpha-1}{k}e^{-k\lambda\Delta}X_0^{\tau-k\Delta}(t) \\ &= X_0^\tau(t) - e^{-\lambda\Delta}\binom{\alpha-1}{1}X_0^{\tau-\Delta}(t) + \dots\end{aligned} \tag{13}$$

Therefore, based on (13) for $\tau \leq \Delta$ only the first term is nonzero and for $\Delta < \tau \leq 2\Delta$ only the first and second term is nonzero and so on. Further, because the Δ values are usually large, the coefficient $e^{-k\lambda\Delta}$ for $k \geq 1$ is very small (for $\lambda = 0.1$ and $\Delta = 600$, $e^{-60} \approx 8.75\text{e}{-27}$). Thus, we can solely consider the first term in our formulation. Additionally, the value of $\frac{1}{A^{\alpha-1}}$ is approximately equal to 1 for $\lambda \geq 0.01$ and $\Delta = 600$. After some mathematical manipulations, we can derive the probability of detection for DSII as in (14).

$$\begin{aligned}\rho(\tau) = X_0^\tau(t) &= \int_0^\tau \frac{\lambda^\alpha}{(\alpha-1)!}t^{\alpha-1}e^{-\lambda t}u(t)dt \\ &= -\sum_{j=0}^{\alpha-1}[\frac{(\lambda t)^j e^{-\lambda t}}{j!}]\Big|_{t=0}^{\tau} \\ = 1 - e^{-\lambda\tau}(1 + \lambda\tau + \frac{(\lambda\tau)^2}{2!} + \dots + \frac{(\lambda\tau)^{\alpha-1}}{(\alpha-1)!}) &= 1 - \frac{\Gamma(\alpha,\lambda\tau)}{(\alpha-1)!}\end{aligned} \tag{14}$$

Numerical results shows that for $\lambda \geq 0.001$, the exact formulation in (13) and the approximation in (14) have similar values, demonstrating high accuracy. Additionally, Figure 2 clearly depicts that the derived equation in (14) is quite accurate in comparison with generated simulation results.

Please note that the closed form derivation of the minimum required darknet IP addresses (i.e., the minimum width of the darknet vantage point), $\min|DIP|$, for DSI and DSII will be accomplished in future work.

IV. Experimentation and Results

In this section, we execute several experimentation to comprehend (1) the impact of the probing rate on the minimum detection time related to various employed IDS operated on the darknet IP space and (2) the implication of the width of the darknet IP space on detection time. Further, motivated by real deployments of darknets, we analyze two case studies to shed the light on the implications of the discussed detection systems in contrast with their detection promptness when operated on those specific darknet IP spaces.

Figure 3 shows various values of effective rate λ in contrast

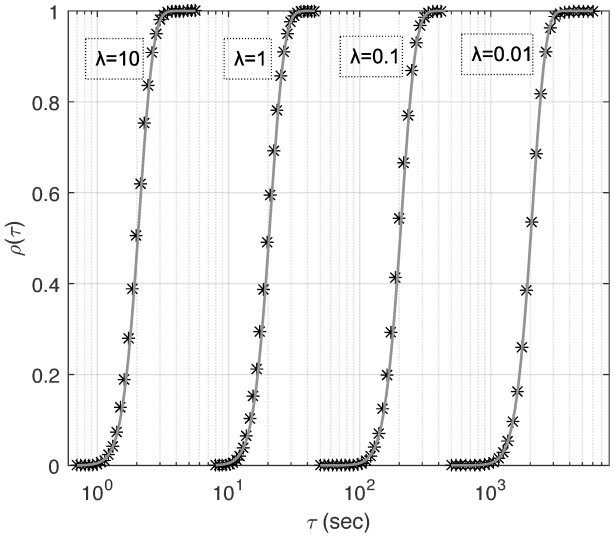

Fig. 2. Validating the accuracy of the relation derived in (14) against simulation results (marked with asterisks) for DSII.

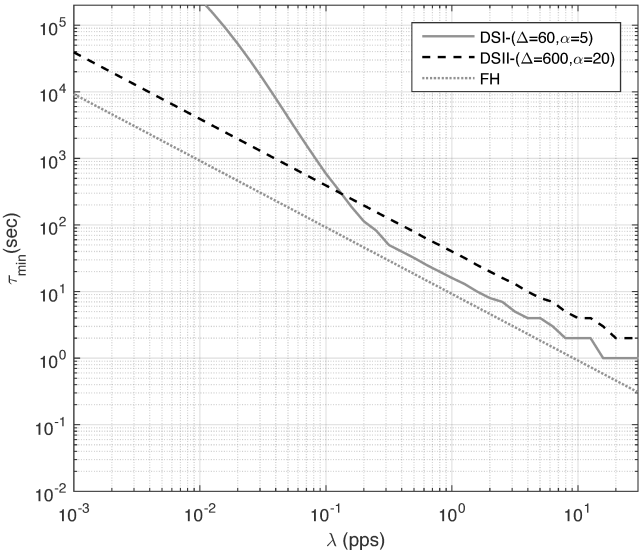

Fig. 3. Minimum detection time τ_{min} versus effective rate λ for $\epsilon = 0.0001$

with the minimum detection time for DSI and DSII. Recall that $\lambda = qr = \frac{|DIP|}{|S|}r$, and therefore, λ is clearly dependent on the scanning rate r and the ratio of number of darknet IP addresses to the subnet size $|S|$. It is revealed from Figure 3 that for $\lambda > 0.1$, DSI outperforms DSII with respect to the minimum detection time and for $\lambda \leq 0.1$, DSII outperforms DSI. From such results, one can extract that for stealthy, low-rate probing events, the Bro IDS is more suited to perform the detection when operated on the darknet IP space. Please note that the result for the FH detection technique is solely depicted to show the lower bound for minimum detection time; a DS can not reach a lower detection time than the minimum detection time of the FH algorithm for a specific λ.

We proceed by illustrating Figure 4, which shows the

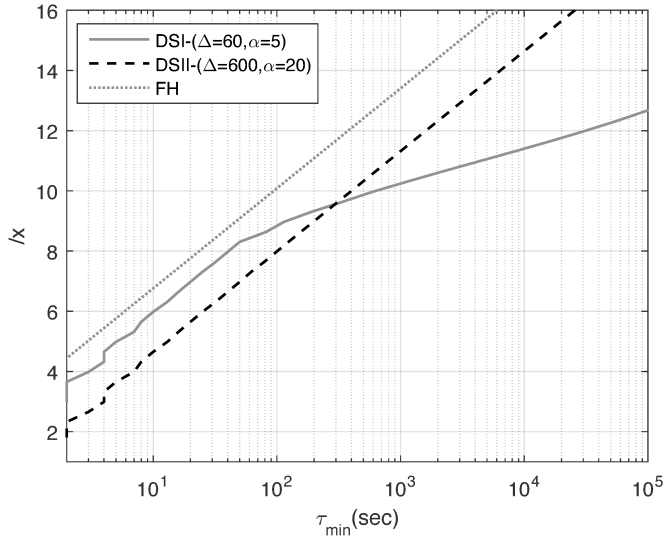

Fig. 4. Portion of darknet IP addresses deployed within a certain subnet versus minimum detection time (τ_{min}) for scanning rate $r = 100$; $\epsilon = 0.0001$

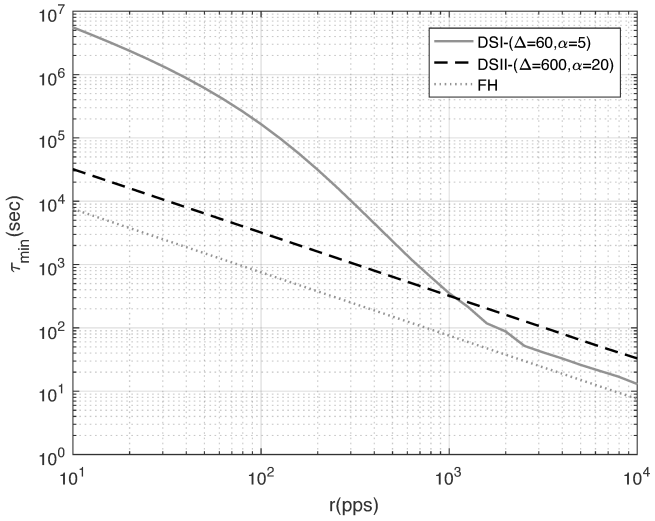

Fig. 5. A $/13$ Network Telescope; $\epsilon = 0.0001$

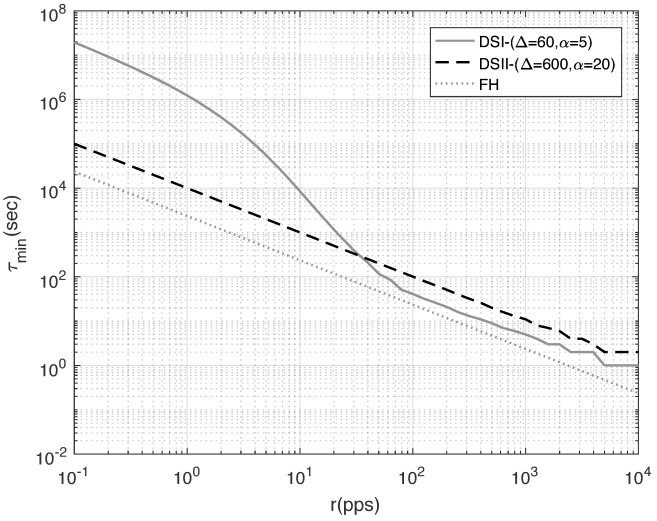

Fig. 6. A $/8$ Network Telescope; $\epsilon = 0.0001$

minimum required portion of deployed darknet IP addresses in the intended subnet in order to achieve a specific minimum detection time. We notate $/x$, which refers to the number of darknet IP addresses; $\frac{1}{2^x}$ of all the subnet address space. Therefore, $x = \log_2(\frac{|S|}{|DIP|})$ and a larger value for x indicates a lower portion of allocated darknet IP addresses. We compare DSI and DSII with their default parameters for Snort and Bro, respectively, given a fixed scanning rate $r = 100$. Figure 4 demonstrates that for $r = 100$ and $\tau_{min} < 300sec$, DSI requires less number of darknet IP addresses (thus reducing cost and management/monitoring resources) in comparison with DSII. Therefore, by employing the Snort IDS, one can achieve the same minimum detection time by utilizing a lower number of darknet IP addresses. On the other hand, for $\tau_{min} \geq 300sec$, the minimum required darknet IP addresses is far lower for DSII. For instance, consider $\tau_{min} = 10000$, then the required darknet IP space would be a $/11$ for DSI and about $/15$ for DSII. This indicates that $2^{32-11} = 2^{21}$ darknet IP addresses are required to detect a large-scale probing activity targeting the entire IPv4 address space with probability more that 0.9999 in 10000 seconds if one employs the Snort IDS, and only $2^{32-15} = 2^{17}$ darknet IP addresses would be needed if one employs the Bro IDS to achieve the same objective.

We now consider two various darknet deployments, representing two practical darknet setups that are currently deployed "in the wild". One refers to a $/8$ darknet, which resembles a large network telescope that is operated by the Center for Applied Internet Data Analysis (CAIDA)[1], while the other represents a $/13$ darknet operated by Farsight Security Information Exchange (SIE)[2]. On one hand, Figure 5 shows that for the $/13$ darknet, the execution of DSII on passive measurements leads to a lower minimum detection time in comparison with DSI (for a practical range of probing rate $1 \leq r \leq 1000$). Therefore, for a $/13$ darknet, the Bro IDS seems to be a more effective detection system, resulting in a lower detection time. On the other hand, Figure 6 shows comparative results for the $/8$ darknet. We can deduce that for this darknet setup, DSI (the Snort IDS) appears to be a more suitable choice for detection, given an average probing rate $30 \leq r \leq 10000$.

[1]http://www.caida.org/data/passive/telescope-near-real-time_dataset.xml

[2]https://archive.farsightsecurity.com/SIE_Channel_14/

V. Discussion

The implications of the proposed passive measurement formalization scheme can be discussed in the context of two topics. First, with the continuous transition from IPv4 to IPv6, the IP address space has intensively increased from 2^{32} to 2^{128}. This larger cyber space indeed requires much more efforts and resources to be monitored, measured and assessed. The darknet IP space, being one of the main sources

of Internet measurements for cyber threat intelligence, should also be adapted. To this end, the choice of the optimal required number of darknet IP addresses would be more challenging when dealing with IPv6. Consider $r = 100$, $\epsilon = 0.0001$ and $\tau_{min} = 10000$. As deduced from Section IV, the minimum required portion of darknet IP addresses for DSI is $/15$ and for DSII is $/11$, for those considered parameters. Recall that this indicates that 2^{21} darknet IP addresses are required to detect a probing activity targeting the entire IPv4 address space using the Snort IDS and 2^{17} darknet IP addresses are needed using the Bro IDS. In contrast, when dealing with IPv6, these numbers are orders of magnitude larger and the implications are even more imperative; for the Snort IDS, one requires $2^{128-11} = 2^{117}$ darknet IP addresses and for the Bro IDS, $2^{128-15} = 2^{113}$ darknet IP addresses are needed, to infer a complete scan of the IPv6 address space. Thus, for IPv4, the difference in terms of required darknet IP addresses related to various IDS types is 15×2^{17} while for IPv6, it is a momentous 15×2^{113}. One can hence note that the choice of IDS employment on passive measurements can severely affect (and amplify) the cost of the resources as well as the darknet management efforts.

Second, we ought to consider highly-distributed scans, similar to the large-scale event reported in [13]. With distributed scans, the probing activity is divided among large number of bots and as a result, the effective scanning rate that is sensed by the darknet is divided by the number of bots participating in the probing campaign. This phenomena can significantly reduce the effective rate λ. Hence, as observed in Figure 3, as λ continues to decrease, the gap between DSI and DSII increases vastly, pinpointing the importance of selecting a suitable detection methodology for combating such ever-evolving events. Nevertheless, one has to note that as seen in Figure 3 related to the minimum detection time, no current detection system is ideal for inferring such large-scale, orchestrated and distributed probing events, paving the way for more tailored detection systems to be researched, designed and implemented in the near future.

VI. Concluding Remarks

Motivated by the fact that passive measurements by exploiting darknet IP spaces are significantly effective in generating various cyber threat intelligence in addition to the lack of formal modeling of darknet-centric parameters, this paper presented a first formal perspective in such contexts. Several detection systems based on highly-employed methods were formalized and a number of derivations were computed and validated to shed the light on the relations between detection probability/time, scanners' rates, and the width of the darknet vantage points. Some of the outcomes suggested the practical usage of the Bro IDS for inferring low-rate probing, its effective application in smaller darknet IP spaces, given a setup that somehow tolerates a delay in detection, and its cost-reduction characteristics when implemented in IPv6 darknet deployment settings. Another outcome pinpointed the lack of effective passive detection methodologies that are capable of inferring large-scale, distributed probes in a timely and practical manner. As for future work, apart from addressing a number of current limitations as discussed throughout this paper, we are conducting various experimentation using real darknet data to better situate the formalization scheme in addition to formally investigating the impact of contemporary IoT attacks in the context of passive measurements.

References

[1] Ala Al-Fuqaha, Abdallah Khreishah, Mohsen Guizani, Ammar Rayes, and Mehdi Mohammadi. Toward better horizontal integration among iot services. *IEEE Communications Magazine*, 53(9):72–79, 2015.

[2] Elias Bou-Harb, Walter Lucia, Nicola Forti, Sean Weerakkody, Nasir Ghani, and Bruno Sinopoli. Cyber meets control: A novel federated approach for resilient cps leveraging real cyber threat intelligence. *IEEE Communications Magazine*, 55(5):198–204, 2017.

[3] Tyler Moore. The promise and perils of digital currencies. *International Journal of Critical Infrastructure Protection*, 6(3):147–149, 2013.

[4] Claude Fachkha and Mourad Debbabi. Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization. *IEEE Communications Surveys & Tutorials*, 18(2):1197–1227, 2016.

[5] M. Debbabi E. Bou-Harb and C. Assi. Cyber scanning: a comprehensive survey. *IEEE Communications Surveys & Tutorials*, 16(3):1496–1519, 2014.

[6] 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.

[7] Claude Fachkha, Elias Bou-Harb, Anastasis Keliris, Nasir Memon, and Mustaque Ahamad. Internet-scale probing of cps: Inference, characterization and orchestration analysis. In *Proceedings of NDSS*, volume 17, 2017.

[8] Binay Kumar Mishra, Santosh Kumar Srivastava, and Bimal Kumar Mishra. A quarantine model on the spreading behavior of worms in wireless sensor network. *Transaction on IoT and Cloud Computing*, 2(1):1–12, 2014.

[9] Maryam Feily, Alireza Shahrestani, and Sureswaran Ramadass. A survey of botnet and botnet detection. In *Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on*, pages 268–273. IEEE, 2009.

[10] Claude Fachkha, Elias Bou-Harb, and Mourad Debbabi. On the inference and prediction of ddos campaigns. *Wireless Communications and Mobile Computing*, 15(6):1066–1078, 2015.

[11] Manos Antonakakis et al. Understanding the mirai botnet. *Usenix Security Symposium*, 2017.

[12] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. On fingerprinting probing activities. *computers & security*, 43:35–48, 2014.

[13] A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescap. Analysis of a "/0" Stealth Scan from a Botnet. *IEEE/ACM Transactions on Networking*, 23(2):341–354, Apr 2015.

[14] Zakir Durumeric, Michael Bailey, and J Alex Halderman. An internet-wide view of internet-wide scanning. In *USENIX Security Symposium*, pages 65–78, 2014.

[15] W. Harrop and G. Armitage. Defining and evaluating greynets (sparse darknets). In *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on*, pages 344–350. IEEE, 2005.

[16] Z.M. Mao D. Watson F. Jahanian E. Cooke, M. Bailey and D. McPherson. Toward understanding distributed blackhole placement. In *Proceedings of the 2004 ACM workshop on Rapid malcode*, pages 54–64. ACM, 2004.

[17] J. Göbel and P. Trinius. Towards optimal sensor placement strategies for early warning systems. In *Sicherheit*, pages 191–204, 2010.

[18] X. Wang D. Leonard, Z. Yao and D. Loguinov. Stochastic analysis of horizontal ip scanning. In *INFOCOM, 2012 Proceedings IEEE*, pages 2077–2085. IEEE, 2012.

[19] Martin Roesch et al. Snort: Lightweight intrusion detection for networks. In *Lisa*, volume 99, pages 229–238, 1999.

[20] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. A novel cyber security capability: Inferring internet-scale infections by correlating malware and probing activities. *Computer Networks*, 94:327–343, 2016.

[21] Vern Paxson. Bro: a system for detecting network intruders in real-time. *Computer networks*, 31(23):2435–2463, 1999.