A Characterization of Cybersecurity Posture from Network Telescope Data

Zhenxin Zhan\textsuperscript{1}, Maochao Xu\textsuperscript{2}, and Shouhuai Xu\textsuperscript{1}

\textsuperscript{1} Department of Computer Science, University of Texas at San Antonio
jankins.ics@gmail.com, shxu@cs.utsa.edu
\textsuperscript{2} Department of Mathematics, Illinois State University
mxu2@ilstu.edu

Abstract. Data-driven understanding of cybersecurity posture is an important problem that has not been adequately explored. In this paper, we analyze some real data collected by CAIDA’s network telescope during the month of March 2013. We propose to formalize the concept of cybersecurity posture from the perspectives of three kinds of time series: the number of victims (i.e., telescope IP addresses that are attacked), the number of attackers that are observed by the telescope, and the number of attacks that are observed by the telescope. Characterizing cybersecurity posture therefore becomes investigating the phenomena and statistical properties exhibited by these time series, and explaining their cybersecurity meanings. For example, we propose the concept of sweep-time, and show that sweep-time should be modeled by stochastic process, rather than random variable. We report that the number of attackers (and attacks) from a certain country dominates the total number of attackers (and attacks) that are observed by the telescope. We also show that substantially smaller network telescopes might not be as useful as a large telescope.

keywords: Cybersecurity data analytics, cybersecurity posture, network telescope, network blackhole, darknet, cyber attack sweep-time, time series data

1 Introduction

Network telescope \textsuperscript{26} (aka blackhole \textsuperscript{105}, darknet \textsuperscript{3}, or network sink \textsuperscript{38}, possibly with some variations) is a useful instrument for monitoring unused, routeable IP address space. Since there are no legitimate services associated to these unused IP addresses, traffic targeting them is often caused by attacks. This allows researchers to use telescope-collected data (together with other kinds of data) to study, for example, worm propagation \textsuperscript{25,23,14}, denial-of-service (DOS) attacks \textsuperscript{24,17}, and stealth botnet scan \textsuperscript{12}. Despite that telescope data can contain unsolicited — but not necessarily malicious — traffic that can be caused by misconfigurations or by Internet background radiation \textsuperscript{28,15,35}, analyzing telescope data can lead to better understanding of cybersecurity posture, an important problem that has yet to be investigated.

Our Contributions. In this paper, we empirically characterize cybersecurity posture based on a dataset collected by CAIDA’s /8 network telescope (i.e., $2^{24}$ IP addresses) during the month of March 2013. We make the following contributions. First, we propose to characterize cybersecurity posture by considering three time series: the number
of victims, the number of attackers, and the number of attacks. To the best of our knowledge, this is the first formal definition of cybersecurity posture. **Second**, we define the notion of sweep-time, namely the time it takes for most telescope IP addresses to be attacked at least once. We find that sweep-time cannot be described by a probabilistic distribution, despite that a proper subset of the large sweep-times follows the power-law distribution. We show that an appropriate stochastic process can instead describe the sweep-time. This means that when incorporating sweep-time in theoretical cybersecurity models, it cannot always be treated as a random variable and may need to be treated as a stochastic process. **Third**, we find that the total number of attackers that are observed by the network telescope is dominated by the number of attackers from a certain country $X$. Moreover, we observe that both the number of attackers from country $X$ and the total number of attackers exhibit a strong periodicity. Although we cannot precisely pin down the root cause of this dominance and periodicity phenomenon, it does suggest that thoroughly examining the traffic between country $X$ and the rest of the Internet may significantly improve cybersecurity. **Fourth**, we investigate whether or not substantially smaller network telescopes would give approximately the same statistics that would be offered by a single, large network telescope. This question is interesting on its own and, if answered affirmatively, could lead to more cost-effective operation of network telescopes. Unfortunately, our analysis shows that substantially smaller telescopes might not be as useful a single, large telescope (of $2^{24}$ IP addresses).

**Related Work.** One approach to understanding cybersecurity posture is to analyze network telescope data. Studies based on telescope data can be classified into two categories. The first category analyzes telescope data alone, and the present study falls into this category. These studies include the characterization of Internet background radiation [28,36], the characterization of scan activities [1], and the characterization of backscatter for estimating global DOS activities [24,17]. However, we analyze cybersecurity posture, especially with regard to attacks that are likely caused by malicious worm, virus and bot activities. This explains why we exclude the backscatter data (which is filtered as noise in the present paper). The second category of studies analyzes telescope data together with other kinds of relevant data. These studies include the use of telescope data and network-based intrusion detection and firewall logs to analyze Internet intrusion activities [39], the use of out-of-band information to help analyze worm propagation [25,23,4], and the use of active interactions with remote IP addresses to filter misconfiguration-caused traffic [28]. There are also studies that are somewhat related to ours, including the identification of one-way traffic from data where two-way traffic is well understood [20,17,33,15].

The other approach to understanding cybersecurity posture is to analyze data collected by honeynet-like systems (e.g., [29,5,21,6,40]). Unlike network telescopes, these systems can interact with remote computers and therefore allow for richer analysis, including the automated generation of attack signatures [18,37].

To the best of our knowledge, we are the first to formally define cybersecurity posture via three time series: the number of victims, the number of attackers, and the number of attacks.

---

3 We were fortunate to see the real, rather than anonymized, attacker IP addresses, which allowed us to aggregate the attackers based on their country code. Our study was approved by IRB.
The rest of the paper is organized as follows. Section 2 describes the data and defines cybersecurity posture. Section 3 briefly reviews some statistical preliminaries. Section 4 defines and analyzes the sweep-time. Section 5 investigates the dominance and periodicity phenomenon exhibited by the number of attackers. Section 6 investigates whether substantially smaller network telescopes would be sufficient or not. Section 7 discusses the limitations of the present study. Section 8 concludes the paper.

2 Representation of Data and Definition of Cybersecurity Posture

Data Description. The data we analyze was collected between 3/1/2013 and 3/31/2013 by CAIDA’s network telescope, which is a passive monitoring system based on a globally routeable but unused /8 network (i.e., 1/256 of the entire Internet IP v4 address space) [31]. Since a network telescope passively collects unsolicited traffic, the collected traffic would contain malicious traffic that reaches the telescope (e.g., automated malware spreading), but may also contain non-malicious traffic — such as Internet background radiation (e.g., backscatter caused by the use of spoofed source IP addresses that happen to belong to the telescope) and misconfiguration-caused traffic (e.g. mistyping an IP address by a remote computer). This means that pre-processing the raw data is necessary. At a high level, we will analyze data \( D_1 \) and \( D_2 \), which are sets of flows and are obtained by applying the pre-processing procedures described below.

Data \( D_1 \). Based on CAIDA’s standard pre-processing [32], the collected IP packets are organized based on eight fields: source IP address, destination IP address, source port number, destination port number, protocol, TTL (time-to-live), TCP flags, and IP length. The flows are reassembled from the IP packets and then classified into three classes: backscatter, ICMP request and other. At a high level, backscatter traffic is identified via TCP SYN+ACK, TCP RST, while ICMP request is identified via ICMP type 0/3/4/5/11/12/14/16/18. (A similar classification method is used in [36].) Since (i) backscatter-based analysis of DOS attacks has been conducted elsewhere (e.g., [24,17]), and (ii) ICMP has been mainly used to launch DOS attacks (e.g., ping flooding and smurf or fraggle attacks [22,35,19]), we disregard the traffic corresponding to backscatter and ICMP request. Since we are more interested in analyzing cybersecurity posture corresponding to attacks that are launched through the TCP/UDP protocols, we focus on the TCP/UDP traffic in the other category mentioned above. We call the resulting data \( D_1 \), in which each TCP/UDP flow is treated as a distinct attack.

Data \( D_2 \). Although (i) \( D_1 \) already excludes the traffic corresponding to backscatter and ICMP request, and (ii) \( D_1 \) only consists of TCP/UDP flows in the other category mentioned above, \( D_1 \) may still contain flows that are caused by misconfigurations. Eliminating misconfiguration-caused flows in network telescope data is a hard problem because network telescope is passive (i.e., not interacting with remote computers [16]). Indeed, existing studies on recognizing misconfiguration-caused traffic had to use payload information (e.g., [22]), which is however beyond the reach of network telescope data. Note that recognizing misconfiguration-caused traffic is even harder than recognizing one-way traffic already (because misconfiguration can cause both one-way and two-way traffic), and that solving the latter problem already requires using extra information (such as two-way traffic [20,17,33,15]). These observations suggest that we use
some heuristics to filter probable misconfiguration-caused flows from \( D_1 \). Our examination shows that, for example, 50% (81%) attackers launched 1 attack (\( \leq 9 \) attacks, correspondingly) against the telescope during the month. We propose to extract \( D_2 \) by filtering from \( D_1 \) the flows that correspond to remote IP addresses that initiate fewer than 10 flows/attacks during the month. This heuristic method filters possibly many, if not most, misconfiguration-caused flows in \( D_1 \). Even though the ground truth (i.e., which TCP/UDP flows correspond to malicious attacks) is not known, \( D_2 \) might be closer to the ground truth than \( D_1 \).

**Data Representation.** In order to analyze the TCP/UDP flow data \( D_1 \) and \( D_2 \), we represent the flows through time series at some time resolution \( r \). We consider two time resolutions (because a higher resolution leads to more accurate statistics): hour, denoted by “\( H \),” and minute, denoted by “\( m \).” For a given time resolution of interest, the total time interval \([0, T]\) is divided into short periods \([i, i+1)\) according to time resolution \( r \in \{H, m\} \), where \( i = 0, 1, \ldots, T - 1 \), and \( T = 744 \) hours (or \( T = 4,464 \) minutes) in this case. We organize the flows into time series from three perspectives:

- the number of *victims* (i.e., network telescope IP addresses that are “hit” by remote attacking IP addresses contained in \( D_1 \) or \( D_2 \)) per time unit at time resolution \( r \),
- the number of *attackers* (i.e., the remote attacking IP addresses contained in \( D_1 \) or \( D_2 \)) per time unit at time resolution \( r \), and
- the number of *attacks* per time unit at time resolution \( r \) (i.e., TCP/UDP flows initiated from remote attacking IP addresses in \( D_1 \) or \( D_2 \) are treated as attacks).

![Fig. 1. Illustration of the attacker-victim relation during time interval \([i, i+1)\) at time resolution \( r \in \{H, r\} \) in \( D_1 \): each dot represents an IP address, a red-colored dot represents an attacking IP address (i.e., attacker), a pink-colored dot represents a victim, each arrow represents an attack (i.e., TCP/UDP flow), the number of attackers is \(|A(r; i, i+1)| = 5\), the number of victims is \(|V(r; i, i+1)| = 7\), and the number of attacks is \(y(r; i, i+1) = 9\). The same holds for data \( D_2 \).](image)

As illustrated in Figure 1, let \( V \) be CAIDA’s fixed set of telescope IP addresses, and \( A \) be the rest of IP addresses in cyberspace, where \(|A| = 2^{32} - |V|\). The major notations are (highlighted and) defined as follows:

- \( V, A \): the set of CAIDA’s network telescope IP addresses and the set of the rest IP v4 addresses, respectively.
A Characterization of Cybersecurity Posture from Network Telescope Data

3 Statistical Preliminaries

We briefly review some statistical concepts and models dealing with time series data, while referring their formal descriptions and technical details to [11][13][34][27].

Brief review of some statistical concepts. Time series can be described by statistical models, such as the AutoRegressive Integrated Moving Average (ARIMA) model and
the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model that
will be used in the paper. The ARIMA model is perhaps the most popular class of time
series models in the literature. It includes many specific models, such as random walk,
seasonal trends, stationary, and non-stationary models [11]. ARIMA models cannot
accommodate high volatilities of time series data, which however can be accommodated
by GARCH models [13]. GARCH models also can capture many phenomena, such as
dynamic dependence in variance, skewness, and heavy-tails [34].

In order to find accurate models for describing time series data, we need to do
model selection. There are many model selection criteria, among which the Akaike’s
Information Criterion (AIC) is widely used. This criterion is based on appropriately
balancing between goodness of fit and model complexity. It is defined in such a way
that the smaller the AIC value, the better the model [27].

Measuring the difference (or distance) between two time series. We need to measure
the difference (or distance) between two time series: \( Z_1, Z_2, \ldots \) and \( Z'_1, Z'_2, \ldots \), where
\( Z_i \geq 0 \) and \( Z'_i \geq 0 \) for \( i = 1, 2, \ldots \). This difference measure characterizes:

1. the fitting error, where the \( Z_i \) time series may represent the observed values and the
   \( Z'_i \) time series may represent the fitted values;
2. the prediction error, where the \( Z_i \) time series may correspond to the observed values
   and the \( Z'_i \) time series may correspond to the predicted values;
3. the approximation error, where the \( Z_i \) time series may describe the observed values
   and the \( Z'_i \) time series may describe the values that may be inferred (i.e., estimated
   or approximated) from other data sources (e.g., we may want to know whether or
   not the statistics derived from the data collected by a large network telescope can
   be inferred from the data collected by a much smaller network telescope).

For conciseness, we use the standard and popular measure known as Percent Mean
Absolute Deviation (PMAD) [2]. Specifically, suppose \( Z_t, Z_{t+1}, \ldots, Z_{t+\ell} \) are given
data, and \( Z'_t, Z'_{t+1}, \ldots, Z'_{t+\ell} \) are the fitted (or predicted, or approximated) data. The
overall fitting (or prediction, or approximation) error (or the PMAD value) is defined as
\[
\frac{\sum_{t+1}^{t+\ell} |Z_j - Z'_j|}{\sum_{t+1}^{t+\ell} Z_j}.
\]
The closer to 0 the PMAD value, the better the fitting (or prediction, or
approximation). We note that our analysis is not bound to the PMAD measure, and it is
straightforward to adapt our analysis to incorporate other measures of interest.

Measuring the shape similarity between two time series. Two time series may be
very different from a measure such as the PMAD mentioned above, but may be similar
to each other in their shape (perhaps after some appropriate re-alignments). Therefore,
we may need to measure such shape similarity between two time series. Dynamic Time
Warping (DTW) is a method for this purpose. Intuitively, DTW aims to align two time
series that may have the same shape and, as a result, the similarity between two time
series can be captured by the notion of warping path (aka warping function). The closer
the warping path to the diagonal, the more similar the two time series. We use the DTW
algorithm in the R software package, which implements the algorithm described in [14].
4 Characteristics of Sweep-Time

The Notion of Sweep-Time. Figure 2 describes the times series of $|V(H; i, i+1)|$ in $D_1$ and $|V'(H; i, i+1)|$ in $D_2$. Using $D_1$ as example, we make the following observations (similar observations can be made for $D_2$). First, there is a significant volatility at the 632nd hour, during which the number of victims is as low as $4,377,079 \approx 2^{22}$. Careful examination shows that the total number of attackers during the 632nd hour is very small, which would be the cause. Second, most telescope IP addresses are attacked within a single hour. For example, $15,998,907$, or 96% of $|V(H; 1, 733)|$, telescope IP addresses are attacked at least once during the first hour. Third, no victims other than $V(H; 0, 703)$ are attacked during the time interval $[704, 744]$.

![Fig. 2. Time series of number of victims: $|V(H; i, i+1)|$ in $D_1$ and $|V'(H; i, i+1)|$ in $D_2$.](image)

Since Figure 2 shows that there are a large number of victims per hour, we ask the following question: How long does it take for most telescope IP addresses to be attacked at least once? That is: How long does it take for $\tau \times |V(H; 0, T)|$ victims to be attacked at least once, where $0 < \tau < 1$? This suggests us to define the following notion of sweep-time, which is relative to the observation start time.

**Definition 2.** (sweep-time) With respect to $D_1$, the sweep-time starting at the $i$th time unit of time resolution $r$, denoted by $I_i$, is defined as:

$$\left| \bigcup_{\ell=i}^{I_i-1} V(r; \ell, \ell+1) \right| < \tau \times |V(H; 0, T)| \leq \left| \bigcup_{\ell=i}^{I_i} V(r; \ell, \ell+1) \right|.$$  

Corresponding to data $D_2$, we can define sweep-time $I'_i$ as:

$$\left| \bigcup_{\ell=i}^{I'_i-1} V'(r; \ell, \ell+1) \right| < \tau \times |V'(H; 0, T)| \leq \left| \bigcup_{\ell=i}^{I'_i} V'(r; \ell, \ell+1) \right|.$$  

By taking into consideration the observation starting time $i$, we naturally obtain two time series of sweep-time: $I_0, I_1, \ldots$ for $D_1$ and $I'_0, I'_1, \ldots$ for $D_2$. We want to characterize these two time series of sweep-time.
Characterizing Sweep-Time. Since per-hour time resolution gives a coarse estimation of sweep-time, we use per-minute time resolution for better estimation of it. Figure 3 plots the time series of sweep-time \( I_0, I_{10}, I_{20}, \ldots \) namely a sample of \( I_0, I_1, \ldots, I_{10}, \ldots, I_{20}, \ldots \) because it is too time-consuming to consider the latter entirely (time resolution: minute).

![Figure 3](image)

**Fig. 3.** Time series plots of sweep-time (y-axis) with respect to \( \tau \in \{80\%, 85\%, 90\%, 95\%, 99\%\} \), where the x-axis represents the observation starting time that is sampled at every 10 minutes. In other words, the plotted points are the sample \((0, I_0), (10, I_{10}), (20, I_{20}), \ldots\) rather than \((0, I_0), (1, I_1), \ldots\).

Figure 3 suggests that the sweep-time time series exhibit similar shape. Accordingly, we use the DTW method to characterize their similarity. Recall that DTW aims to align two time series via the notion of warping path, such that the closer the warping path to the diagonal, the more similar the two time series.

Figure 4 confirms the above observation, by presenting some examples of the warping paths (the others are omitted due to space limitation). Specifically, Figures 4(a) and 4(b) show that the sweep-times with respect to \( \tau = .80 \) and \( \tau = .99 \) are almost the same in \( D_1 \) and \( D_2 \), respectively. Figures 4(c) and 4(d) show that for two thresholds, say \( \tau_1 \) and \( \tau_2 \), the smaller the \( |\tau_1 - \tau_2| \), the more similar the two respective time series of sweep-time in \( D_1 \). Figures 4(e) and 4(f) demonstrate the same phenomenon for \( D_2 \).

The above discussion suggests that the notion of sweep-time is not sensitive to spatial threshold \( \tau \). This leads us to ask: **What is the distribution of sweep-time?** However, this question makes sense only when the time series is stationary. By using an augmented Dickey-Fuller test [34], we conclude that the sample of the sweep-time time series, namely \( I_0, I_{10}, I_{20}, \ldots \) is not stationary, which means that time series \( I_0, I_1, I_2 \ldots \) is not stationary (otherwise, the sample should be stationary). Therefore, we cannot use a single distribution to characterize the sweep-time; Instead, we have to characterize the sweep-time as a stochastic process. In order to identify good time series models that can fit the sweep-time, we need to identify some statistical properties that are exhibited by sweep-time. In particular, we need to know if the sweep-time is heavy-tailed, meaning that the sweep-times greater or equal to \( x_{\text{min}} \) exhibit the power-law distribution, where \( x_{\text{min}} \) is called cut-off parameter.

Table 1 summarizes the power-law test statistics of the sweep-time with cut-off parameter \( x_{\text{min}} \). We observe that for both \( D_1 \) and \( D_2 \), all the \( \alpha \) values (i.e., the fitted
A Characterization of Cybersecurity Posture from Network Telescope Data

Fig. 4. DTW-based similarities of sweep-time time series with different threshold $\tau$.

| $\tau$ | $\alpha$ | $x_{\text{min}}$ | $KS$ | p-value | $\# \geq x_{\text{min}}$ | $\tau$ | $\alpha$ | $x_{\text{min}}$ | $KS$ | p-value | $\# \geq x_{\text{min}}$
|-------|---------|------------------|------|---------|------------------------|-------|---------|------------------|------|---------|------------------------
| 80%   | 7.89    | 78.05            | .14  | 475     | 80%                    | 8.46  | 82.05   | .19  | 391     |                          |
| 85%   | 8.46    | 94.04            | .52  | 385     | 85%                    | 8.37  | 95.04   | .36  | 379     |                          |
| 90%   | 8.89    | 118.06           | .42  | 244     | 90%                    | 9.24  | 120.05  | .39  | 237     |                          |
| 95%   | 9.52    | 148.05           | .68  | 193     | 95%                    | 12.82 | 170.04  | .99  | 72      |                          |
| 99%   | 13.67   | 215.04           | .98  | 131     | 99%                    | 15.23 | 224.04  | .99  | 94      |                          |

Table 1. Power-law test statistics of the sweep-time with respect to spatial threshold $\tau \in \{80\%, 85\%, 90\%, 95\%, 99\%\}$, where $\alpha$ is the fitted power-law exponent, $x_{\text{min}}$ is the cut-off parameter, $KS \in [0.04, 0.06]$ is the Kolmogorov-Smirnov statistic [9] for comparing the fitted power-law distribution and the data (meaning that the fitting is good) as indicated by that the $p$-values are $> 0.05$, and "$\# \geq x_{\text{min}}$" represents the number of sweep-times that are greater than or equal to $x_{\text{min}}$ (i.e., the number of sweep-times that are used for fitting).
power-law exponents) are very large. For spatial threshold $\tau = 80\%$ in $D_1$, we have $x_{\min} = 78$ minutes, meaning that the number of sweep-times that are greater than or equal to $x_{\min}$ is 475 (or 10.6\% out of 4,462). As spatial threshold $\tau$ increases, $x_{\min}$ increases and the number of sweep-times greater than or equal to $x_{\min}$ decreases. We also observe that for the same $\tau$, $D_1$ and $D_2$ have similar $x_{\min}$ values, which means that the filtered attack traffic in $D_1$ does not affect the power-law property of the data.

The above analysis suggests that in order to fit the sweep-time, we should use a model that can accommodate the power-law property. Therefore, we use the ARMA+GARCH model, where ARMA accommodates the stable sweep-times smaller than $x_{\min}$, and GARCH, with skewed student $t$-distribution, accommodates the power-law distributed sweep-times (which are greater than or equal to $x_{\min}$). Consider spatial threshold $\tau = .99$ as an example. Figures 5(a) and 5(b) plot the observed data and the fitting model for sweep-time $I_t$ (observation starting time $t$):

$$I_t - \mu_t = \phi_1(I_{t-1} - \mu_{t-1}) + \phi_2(I_{t-2} - \mu_{t-2}) + \epsilon_t,$$

where $\mu_t = \mu + \xi \sigma_t$ is the dynamic mean composed of a constant term $\mu$ and standard deviation $\sigma_t$ of the error term, $\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1} + \beta_1 \sigma_{t-1}^2$, $\mu_t = E[I_t]$, $\epsilon_t$ is the error term at time $t$, and $\sigma_t^2 = E(y_t - \mu_t)^2$ is the variance modeled via the standard GARCH(1,1) process. The fitting errors (PMAD values) are .121 and .119 for $D_1$ and $D_2$, respectively.

![Sweep-Time](image)

(a) Fitting sweep-time in $D_1$.  \hspace{1cm} (b) Fitting sweep-time in $D_2$.

**Fig. 5.** Fitting sweep-time with spatial threshold $\tau = .99$ (time resolution: minute), where black-colored dots are observed sweep-time values, red-colored dots are fitted values.

It was known that malware can infect almost all susceptible computers within a very short period of time (e.g., the Slammer worm [23]), meaning that the sweep-time with respect to a specific observation starting time is very small. This is, in a sense, re-affirmed by our study. However, for continuous attacks that are based on a bag of
attacking tools, sweep-time should be better modeled with respect to any (rather than a specific) observation starting time. We are the first to show that the sweep-time cannot be modeled by a random variable (which would make the model in question easier to analyze though). This leads to the following insight, which could guide future development of advanced cybersecurity models.

**Insight 1** When one needs to model the sweep-time (i.e., the time it takes for each IP address of a $\tau$-portion of a large network space to be attacked at least once), it should be modeled by a stochastic process rather than a random variable.

## 5 A Phenomenon Exhibited by Attacking IP Addresses

![Diagram](image)

**(a) Origins of attackers in $D_1$**

**(b) Origins or attackers in $D_2$**

Fig. 6. During the month, three countries, which are anonymized as X, Y and Z, contribute to most of the attackers in $D_1$; whereas countries X, Y and Z' ($Z' \neq Z$) contribute to most of the attackers in $D_2$.

For each attacker IP address, we can use the WHOIS service to retrieve its country code. Figures 6 plots the origins of attackers that contribute to most of the attackers (per country code). Note that the category “others” in $D_1$ include 6,894,900 attacker IP addresses (or 1.7% of the total number of attackers) whose country codes cannot be retrieved from the WHOIS service. The category “others” in $D_2$ include 10,740 attacker IP addresses (or 0.01% of the total number of attackers) whose country codes cannot be retrieved from the WHOIS service. This means that many attacker IP addresses whose country codes cannot be retrieved from the WHOIS service are filtered. Moreover, the attacker IP addresses with no country code do not have a significant impact on the result. We observe that country X contributes 30% of the attackers in $D_1$ and 76% of the attackers in $D_2$. Country Y contributes 28% of the attackers in $D_1$ and 3% of the attackers in $D_2$. This is caused by the fact that 50% attackers from country X and 98% attackers from country Y launch fewer than 10 attacks during the month, and therefore do not appear in $D_2$. This prompts us to study the relationship between two time series: the total number of attackers and the number of attackers from country X.

**The Dominance and Periodicity Phenomenon.** Figure 7 compares the times series of the total number of attackers observed by the telescope and the time series of the
number of attackers from country $X$ in $D_1$ and $D_2$, respectively. For $D_1$, Figure 7(a) shows that the total number of attackers during time interval $[455, 630]$, namely the 176 hours between the 455th hour (on March 19, 2013) and the 630th hour (on March 27, 2013), is substantially greater than its counterpart during the other hours. This is caused by the substantial increase in the number of attackers from country $Y$, despite that we do not know the root cause behind the substantial increase of attackers in country $Y$. For $D_2$, Figure 7(b) does not exhibit the same kind of substantial increase during the interval $[455, 630]$, meaning that many of the “emerging” attackers from country $Y$ are filtered (because they launched fewer than 10 attacks during the month).

Figure 7 further suggests a surprising consistency between the two time series. Specifically, when the number of attackers from country $X$ is large (small), the total number of attackers is large (small). For $D_1$, this is confirmed by Figures 8(a) and 8(b) which clearly show that the same periodicity is exhibited by the total number of attackers and by the number of attackers from country $X$. For $D_2$, this is confirmed by Figures 8(c) and 8(d) which clearly show that the same periodicity is exhibited by the total number of attackers and by the number of attackers from country $X$. We observe that the wave bases are periodic with a period of 24 hours. After looking into the time zone of country $X$, we find that the wave bases (i.e., that least number of attackers) correspond to the hour between 12:00 noon and 1 pm local time. One may speculate that this is caused by computers possibly being put into the hibernate mode (during lunch time). This may not be true because during the night hours, more computers would be put into the hibernate mode (or even powered off) and therefore even fewer attackers would be observed. However, this is not shown by the data. One perhaps more plausible explanation is that the attacking computers may be coordinated or controlled (for example) by botnets.

While we defer the detailed characterization of the phenomenon to Appendix A, we summarize the phenomenon as:

**Phenomenon 1** (The dominance and periodicity phenomenon exhibited by the number of attackers) The time series of the total number of attackers and the time series of the number of attackers from a particular country $X$ exhibit the same periodicity. Moreover, the total number of attackers is dominated by the number of attackers from country $X$. 

![Fig. 7](image-url)
6 Inferring Global Cyber Security Posture from Smaller Monitors

In this section, we explore whether it is possible to use small network telescopes to approximate bigger telescopes, from the perspectives of estimating/infering the number of victims, attackers and attacks. Answering this question is interesting on its own, and could lead to more cost-effective operations of network telescopes.

**Methodology.** We divide the /8 network telescope into $B$ equal-size blocks of IP addresses, where each block is called a small telescope. We want to know whether we can infer the number of victims (or attackers, or attacks) that are observed by the /8 telescope during time interval $[t, t+1]$ at time resolution $H$ (i.e., per hour), denoted by $Y(H; t, t+1)$, from the number of victims (or attackers, or attacks) observed by $b$ small telescopes, where $b << B$, during the same time interval, denoted by $Y_1(H; t, t+1), \ldots, Y_b(H; t, t+1)$. In other words, we want to know whether the following equation would hold:

$$Y(H; t, t+1) = c + \sum_{i=1}^{b} \phi_i Y_i(H; t, t+1),$$

where $c$ is some constant and $\phi_i$’s are coefficients. Naturally, we can use the same PMAD measure to evaluate the estimation/inference error.

Whenever feasible, we want to consider all possible combinations of $b$ small telescopes. For $B = 16$, there are $\binom{16}{b}$ combinations; for $B = 256$, there are $\binom{256}{b}$ combinations. For $B = 256$ and $b \geq 4$, the number of combinations becomes prohibitive.
This suggests that we first cluster the $B = 256$ blocks into $b$ groups based on the DTW measure, and then sample one block from each of the $b$ groups. In a sense, this corresponds to the best-case scenario sampling because we need the prior information about the groups or clusters. If the sample statistics cannot approximate the statistics derived from the data collected by the /8 telescope, we can conclude that small telescopes are not as useful as the large telescope.

|     | $D_1$ with $B = 16$: PMAD values | $D_1$ with $B = 256$: PMAD values | $D_2$ with $B = 16$: PMAD values | $D_2$ with $B = 256$: PMAD values |
|-----|---------------------------------|-----------------------------------|---------------------------------|-----------------------------------|
| $b = 1$ | $1.0372, 0.0472, 0.0464, 0.0682, 0.0083$ | $1.0563, 0.0689, 0.0678, 0.1131, 0.0074$ | $1.0384, 0.0484, 0.0478, 0.0702, 0.0085$ | $1.0799, 0.1136, 0.1155, 0.1155, 0.0071$ |
| $b = 2$ | $0.0230, 0.0322, 0.0302, 0.0586, 0.0067$ | $0.0414, 0.0535, 0.0521, 0.1129, 0.0057$ | $0.0235, 0.0340, 0.0311, 0.0702, 0.0079$ | $0.0731, 0.1119, 0.1155, 0.1155, 0.0096$ |
| $b = 3$ | $0.0167, 0.0232, 0.0239, 0.0301, 0.0333$ | $0.0330, 0.0447, 0.0436, 0.0982, 0.0049$ | $0.0167, 0.0267, 0.0251, 0.0702, 0.0066$ | $0.0755, 0.0787, 0.0787, 0.0839, 0.0021$ |

Table 2. PMAD-based measurement of the inference error when using $b$ (out of the $B$) small telescopes to approximate the number of victims that are observed by the larger /8 telescope, where “SD” stands for standard deviation.

**Characterizing Inference Errors of Small Telescopes.** From the perspective of inferring the number of victims from small telescopes, Table 2 summarizes the inference errors, in terms of the min, mean, median and max PMAD values of all the considered combinations of sample blocks, as well as the standard deviation of the PMAD values. For $D_1$ and $B = 16$, we observe that a single small telescope (out of the 16 telescopes of size $2^{20}$ IP addresses) would give good approximation of the number of victims that would be obtained based on the larger /8 network telescope. This is because the maximum PMAD errors is .0682, namely 6.82% approximation error. For $D_1$ and $B = 256$, the mean approximation error is 6.89% for $b = 1$ (i.e., using one small telescope), and 5.35% for $b = 2$ (i.e., using two small telescopes) and 4.47% for $b = 3$ (using three small telescopes). For $D_2$ and $B = 16$, we observe a similar phenomenon as in the case of $D_1$ and $B = 16$. However, for $D_2$ and $B = 256$, the mean approximation errors is significantly larger than in the case of $D_1$ and $B = 256$, namely 11.36%, 11.19% and 7.87% for $b = 1, 2, 3$, respectively. Therefore, we can conclude that from the perspective of the number of victims, a single telescope of size $2^{20}$ IP addresses would give approximately the same result as the telescope of size $2^{24}$, and 3 randomly selected small telescope of size $2^{10}$ would give approximately the same result as the telescope of size $2^{24}$. That is, the small telescope could be used instead.

Due to space limitation, we defer to Appendix B the characterizations on inferring the number of attackers from small telescopes and on inferring the number of attacks from small telescopes. Based on these characterizations, we draw the following insight:
Insight 2  For estimating the number of victims, substantially small telescopes could be used instead. However, for estimating the number of attackers and the number of attacks, substantially small telescopes might not be sufficient.

The above discrepancy between the number of victims and the numbers of attackers/attacks is possibly caused by the following: The victims are somewhat “uniformly” attacked, but the attackers and attacks are far from “uniformly” distributed. Moreover, a single attacker that scans the large telescope’s IP address space will make it easy to estimate the number of victims from small telescopes.

7 Limitations of the Study

The present study has several limitations, which are inherent to the data but not to the methodology we use. First, our analysis treats each remote IP address as a unique attacker. This is not accurate when the remote networks using Network Address Translation (NAT), because remote attackers from the same network can be “aggregated” into a single attacker. If many networks in country X indeed use NAT, then the actual number of attackers from country X is indeed larger, although the number of attacks from country X is not affected by NAT.

Second, the characteristics presented in the paper inherently depend on the nature of network telescope. For example, $D_1$ and $D_2$ still may contain some misconfiguration-caused, non-malicious traffic. Due to the lack of interactions between network telescope and remote computers (an inherent limitation of network telescopes), it is hard to know the ground truth [16]. Therefore, better filtering methods are needed so as to make the data approximate the ground truth as closely as possible.

Third, it is possible that some attackers are aware of the network telescope and therefore can instruct their attacks to bypass it. As a consequence, the data may not faithfully reflect the cybersecurity posture.

Fourth, the data collected by the network telescope does not contain rich enough information that would allow us to conduct deeper analysis, such as analyzing the global characteristics of specific attacks. Moreover, the data is a “coarse” sample of the ground-truth cybersecurity posture because the first flow from a remote attacker may be a scan/probe activity, or a first attack attempt against a specific port.

8 Conclusion

We have studied the cybersecurity posture based on the data collected by CAIDA’s network telescope. We have found that the sweep-time should be characterized as a stochastic process rather than a random variable. We also have found that the total number of attackers (and attacks) that are observed by the network telescope is largely determined by the number of attackers from a single country. There are many interesting problems for future research, such as: How can we (more) accurately predict the time series? To what extent they are predictable?

Acknowledgement. We thank CAIDA for sharing with us the data analyzed in the paper. This work was supported in part by ARO Grant #W911NF-13-1-0141 and NSF Grant #1111925.
References

1. Mark Allman, Vern Paxson, and Jeff Terrell. A brief history of scanning. In Proc. ACM IMC'07, pages 77–82.
2. Jon Scott Armstrong. Principles of forecasting: a handbook for researchers and practitioners, volume 30. Springer, 2001.
3. M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha. Practical darknet measurement. In Proc. 2006 Annual Conference on Information Sciences and Systems, pages 1496–1501.
4. M. Bailey, E. Cooke, F. Jahanian, and D. Watson. The blaster worm: Then and now. Security Privacy, IEEE, 3(4):26–31, July 2005.
5. Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, David Watson, et al. The internet motion sensor-a distributed blackhole monitoring system. In Proc. NDSS’05.
6. P. Barford, Y. Chen, A. Goyal, Z. Li, V. Paxson, and V. Yegneswaran. Employing honeynets for network situational awareness. In Cyber Situational Awareness, pages 71–102. 2010.
7. Nevil Brownlee. One-way traffic monitoring with iatmon. In Proc. PAM’12, pages 179–188.
8. K. Claffy, H. Braun, and G. Polyzos. A parameterizable methodology for internet traffic flow profiling. IEEE Journal on Selected Areas in Communications, 13(8):1481–1494, Oct 1995.
9. Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. SIAM review, 51(4):661–703, 2009.
10. Evan Cooke, Michael Bailey, Z Morley Mao, David Watson, Farnam Jahanian, and Danny McPherson. Toward understanding distributed blackhole placement. In Proc. ACM Worm’04, pages 54–64.
11. J. Cryer and K. Chan. Time Series Analysis With Applications in R. Springer, New York, 2008.
12. A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapè. Analysis of a ”0 stealth scan from a botnet. In Proc. ACM IMC’12, pages 1–14.
13. Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, pages 987–1007, 1982.
14. Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. Journal of statistical Software, 31(7):1–24, 2009.
15. E. Glatz and X. Dimitropoulos. Classifying internet one-way traffic. In Proc. ACM IMC’12, pages 37–50.
16. F. Gringoli, L. Salgarelli, M. Dusi, N. Cascarano, F. Risso, and K. Claffy. Gt: Picking up the truth from the ground for internet traffic. SIGCOMM Comput. Commun. Rev., 39(5):12–18, 2009.
17. Alefiya Hussain, John Heidemann, and Christos Papadopoulos. A framework for classifying denial of service attacks. In Proc. ACM SIGCOMM’03, pages 99–110.
18. Christian Kreibich and Jon Crowcroft. Honeycomb: Creating intrusion detection signatures using honeypots. SIGCOMM Comput. Commun. Rev., 34(1):51–56, 2004.
19. F. Lau, S.H. Rubin, M.H. Smith, and L. Trajkovic. Distributed denial of service attacks. In Proc. 2000 IEEE International Conference on Systems, Man, and Cybernetics, volume 3, pages 2275–2280.
20. DongJin Lee and Nevil Brownlee. Passive measurement of one-way and two-way flow lifetimes. SIGCOMM Comput. Commun. Rev., 37(3):17–28, July 2007.
21. Z. Li, A. Goyal, Y. Chen, and V. Paxson. Towards situational awareness of large-scale botnet probing events. IEEE Transactions on Information Forensics and Security, 6(1):175–188, 2011.
22. Zhichun Li, A. Goyal, Yan Chen, and A. Kuzmanovic. Measurement and diagnosis of address misconfigured p2p traffic. IEEE Network, 25(3):22–28, May 2011.
23. D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the slammer worm. \textit{IEEE Security and Privacy}, 1(4):33–39, Aug 2003.
24. D. Moore, C. Shannon, D. Brown, G. Voelker, and S. Savage. Inferring internet denial-of-service activity. \textit{ACM Trans. Comput. Syst.}, 24(2):115–139, May 2006.
25. D. Moore, C. Shannon, and J. Brown. Code-Red: a case study on the spread and victims of an Internet worm. In \textit{Proc. ACM IMW’02}, pages 273–284.
26. David Moore, Colleen Shannon, Geoffrey M Voelker, and Stefan Savage. \textit{Network telescopes: Technical report}. Department of Computer Science and Engineering, University of California, San Diego, 2004.
27. John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. \textit{Applied linear statistical models}, volume 4. Irwin Chicago, 1996.
28. Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of internet background radiation. In \textit{Proc. ACM IMC’04}, pages 27–40.
29. N. Provos. A virtual honeypot framework. In \textit{Proc. USENIX Security Symposium}, pages 1–14, 2004.
30. C. Shannon and D. Moore. The spread of the witty worm. \textit{IEEE Security and Privacy}, 2(4):46–50, 2004.
31. CAIDA UCSD Network Telescope. \url{http://www.caida.org/}
32. CAIDA UCSD Network Telescope. \url{http://www.caida.org/tools/measurement/corsaro/docs/plugins.html}
33. Joanne Treurniet. A network activity classification schema and its application to scan detection. \textit{IEEE/ACM Trans. Netw.}, 19(5):1396–1404, 2011.
34. Tsay. \textit{Analysis of Financial Time Series}. Wiley, 2010.
35. N. Weiler. Honeypots for distributed denial-of-service attacks. In \textit{Proc. IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET-ICE’02)}, pages 109–114.
36. Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. Internet background radiation revisited. In \textit{Proc. ACM IMC’10}, pages 62–74.
37. V. Yegneswaran, J. Giffin, P. Barford, and S. Jha. An architecture for generating semantic aware signatures. In \textit{Proc. Usenix Security Symposium}, 2005.
38. Vinod Yegneswaran, Paul Barford, and Dave Plonka. On the design and use of internet sinks for network abuse monitoring. In \textit{Proc. Recent Advances in Intrusion Detection (RAID’04)}, pages 146–165.
39. Vinod Yegneswaran, Paul Barford, and Johannes Ullrich. Internet intrusions: Global characteristics and prevalence. In \textit{Proc. ACM SIGMETRICS’03}, pages 138–147.
40. Z. Zhan, M. Xu, and S. Xu. Characterizing honeypot-captured cyber attacks: Statistical framework and case study. \textit{IEEE Transactions on Information Forensics and Security}, 8(11):1775–1789, 2013.

A Characterization of the Dominance and Periodicity Phenomenon Exhibited by Attackers

Now we quantify the similarity between the two time series via Dynamic Time Warping (DTW), fitted model, and prediction accuracy.

Similarity based on DTW. Figure 9(a) plots the warping path between the total number of attackers in $D_1$ and the total number of attackers in $D_2$. The two time series are very similar to each other, except for the time interval $[452, 668]$ as suggested by Figures 8(a) and 8(c). Figure 9(b) plots the warping path between the two time series plotted in
Figure 7(a) namely the total number of attackers in $D_1$ and the total number of attackers from country $X$ in $D_1$. It shows that the two time series are very similar to each other except during the time interval $[455, 630]$, as suggested by Figures 8(a) and 8(b). Figure 9(c) plots the warping path between the two time series plotted in Figures 8(b) and 8(d). It shows that the two time series are almost identical to each other, and that the filtering of rarely seen attackers/attacks does not manipulate the periodic structure of the time series of the number of attackers from country $X$. Figure 9(d) plots the warping path between the two time series plotted in Figures 8(c) and 8(d), which indeed are almost identical to each other.

**Similarity based on fitted models.** Since both the time series exhibit periodicity, we use the multiplicative seasonal ARIMA model to fit the two time series in $D_1$ and $D_2$, respectively. The model parameters are: nonseasonal orders $(p, d, q)$, and seasonal orders $(P, D, Q)$, and seasonal period $s = 24$ based on the above discussion of periodicity. For model selection, the parameter sets are:

- $(p, d, q) \in [0, 5] \times \{0, 1\} \times [0, 5]$;
- $(P, D, Q) \in [0, 5] \times \{0, 1\} \times [0, 5]$.

According to the AIC criterion (briefly reviewed in Section 3), the two time series in both $D_1$ and $D_2$ prefer to the following model:

$$W_t = \phi_1 W_{t-1} + e_t + \theta_1 e_{t-1} + \Phi_1 W_{t-24} + \Phi_2 W_{t-48} +$$

$$\Theta_1 e_{t-24} + \Theta_2 e_{t-48} + \Theta_3 e_{t-96},$$

where $W_t = |A(r; t, t + 1)| - |A(r; t - 24, t - 23)|$. Table 3 summarizes the fitting results. We observe that the two fitted models in $D_1$ are similar to each other in terms of coefficients, and that the two fitted models in $D_2$ are almost identical to each other.

**Similarity based on prediction accuracy.** Table 3 summarizes the PMAD values for 1, 4, 7 and 10 hours ahead-of-time prediction of the number of attackers during the last 96 hours in both $D_1$ and $D_2$. For $D_1$, we observe that 1-hour ahead-of-time predictions for the number of attackers from country $X$ and the total number of attackers are
A Characterization of Cybersecurity Posture from Network Telescope Data

| φ_1 | θ_1 | Φ_1 | θ_2 | Φ_2 | θ_3 |
|-----|-----|-----|-----|-----|-----|
| .91 | .38 | 1.22| -.98| 2.15| 2.11|
| Coefficients | PMAD=.08 |

Fitted model of number of attackers from country X in \( D_1 \): PMAD=.06

| Coefficients | .82 | .39 | 1.22| -.99| 2.19| 2.16|

Fitted model of total number of attackers in \( D_2 \): PMAD=.08

| Coefficients | .79 | .4 | 1.21| -.99| 2.18| 2.16|

Fitted model of number of attackers from country X in \( D_2 \): PMAD=.07

| Coefficients | .79 | .4 | 1.21| -.99| 2.19| 2.16|

Table 3. Coefficients in the fitted models of the total number of attackers and the number of attackers from country X.

reasonably accurate (with PMAD value .093 and .092, or 9.3% and 9.2% prediction error, respectively), whereas the predictions for 4, 7 and 10 hours ahead-of-time are not useful. For \( D_2 \), we observe similar prediction results, namely that 1-hour ahead-of-time predictions lead to 7.5% prediction error for the total number of attackers and 9.5% prediction error for the number of attackers from country X.

| \( h = 1 \) | \( h = 4 \) | \( h = 7 \) | \( h = 10 \) |
|-----|-----|-----|-----|
| \( D_1 \): PMAD values | .092 | .244 | .333 | .404 |
| \( D_2 \): PMAD values | .075 | .156 | .180 | .177 |

Table 4. PMAD values for \( h = 1, 4, 7, 10 \) hours ahead-of-time predictions on the total number of attackers and on the number of attackers from country X, as observed by the telescope.

B Further Characterizations on the Inference Errors of Small Telescopes

Inferring the number of attackers from small telescopes. Similarly, we would like to infer the number of attackers based on small telescopes. Table 5 summarizes the inference errors in terms of the min, mean, median and max PMAD values of all the considered combinations of sample blocks, as well as the standard deviation of the PMAD values. For \( D_1 \) and \( B = 16 \), we observe that 3 small telescopes (out of the 16 telescopes of size \( 2^{20} \) IP addresses) would give good approximation of the number of attackers that would be obtained based on the network telescope of size \( 2^{24} \). This is because the maximum PMAD value is 7.34%. For \( D_2 \) and \( B = 16 \), we observe that using 4 small telescopes of size \( 2^{20} \) does not lead to good approximation. For \( B = 256 \), neither \( D_1 \) nor \( D_2 \) leads to obtain good enough approximation. These suggest that using significantly small telescopes may not lead to robust results.

Inferring the number of attacks from small telescopes. From the perspective of inferring the number of attacks, Table 6 summarizes the inference errors as in the above.
Table 5. PMAD-based measurement of the inference error when using \( b \) (out of the \( B \)) small telescopes to approximate the number of attackers that are observed by the larger /8 telescope, where “SD” stands for standard deviation.

For \( D_1 \) and \( B = 16 \), we observe that 3 small telescopes (out of the 16 telescopes of size \( 2^{20} \) IP addresses) would give good approximation of the number of attacks that would be obtained based on the larger network telescope of size \( 2^{24} \). This is because the maximum PMAD errors is 0.0761, namely 7.61% approximation error. For \( D_1 \) and \( B = 256 \), the mean approximation error is 9.58% for \( b = 5 \) (i.e., using 5 small telescopes instead), which is marginally acceptable. For \( D_2 \) and \( B = 16 \), we observe that using 4 small telescopes of size \( 2^{20} \) can lead to worst-case approximation error 8.27%. For \( D_2 \) and \( B = 256 \), we observe that using 5 small telescopes of size \( 2^{16} \) does not lead to good approximation. That is, substantially small telescope may not be as useful as the large telescope.

Table 6. PMAD-based measurement of the inference error when using \( b \) (out of the \( B \)) small telescopes to approximate the number of attacks observed by the /8 telescope, where “SD” stands for standard deviation.