Internet-scale Probing of CPS: Inference, Characterization and Orchestration Analysis

Claude Fachkha[1,2], Elias Bou-Harb[3], Anastasis Keliris[1], Nasir Memon[1], and Mustaque Ahamad[1,4]

[1]New York University (NYU) and NYU Abu Dhabi
[2]University of Dubai
[3]Florida Atlantic University
[4]Georgia Institute of Technology

***Abstract*—Although the security of Cyber-Physical Systems (CPS) has been recently receiving significant attention from the research community, undoubtedly, there still exists a substantial lack of a comprehensive and a holistic understanding of attackers' malicious strategies, aims and intentions. To this end, this paper uniquely exploits passive monitoring and analysis of a newly deployed network telescope IP address space in a first attempt ever to build broad notions of real CPS maliciousness. Specifically, we approach this problem by inferring, investigating, characterizing and reporting large-scale probing activities that specifically target more than 20 diverse, heavily employed CPS protocols. To permit such analysis, we initially devise and evaluate a novel probabilistic model that aims at filtering noise that is embedded in network telescope traffic. Subsequently, we generate amalgamated statistics, inferences and insights characterizing such inferred scanning activities in terms of their probe types, the distribution of their sources and their packets' headers, among numerous others, in addition to examining and visualizing the co-occurrence patterns of such events. Further, we propose and empirically evaluate an innovative hybrid approach rooted in time-series analysis and context triggered piecewise hashing to infer, characterize and cluster orchestrated and well-coordinated probing activities targeting CPS protocols, which are generated from Internet-scale unsolicited sources.**

Our analysis and evaluations, which draw upon extensive network telescope data observed over a recent one month period, demonstrate a staggering 33 thousand probes towards ample of CPS protocols, the lack of interest in UDP-based CPS services, and the prevalence of probes towards the ICCP and Modbus protocols. Additionally, we infer a considerable 74% of CPS probes that were persistent throughout the entire analyzed period targeting prominent protocols such as DNP3 and BACnet. Further, we uncover close to 9 thousand large-scale, stealthy, previously undocumented orchestrated probing events targeting a number of such CPS protocols. We validate the various outcomes through cross-validations against publicly available threat repositories. We concur that the devised approaches, techniques, and methods provide a solid first step towards better comprehending real CPS unsolicited objectives and intents.

NDSS '17, 26 February - 1 March 2017, San Diego, CA, USA

http://dx.doi.org/10.14722/ndss.2017.23149

I. Introduction

Critical infrastructure systems are indispensable to the broader health, safety, security, and economic well-being of modern society and governments. In recent years, many of these systems have been undergoing large-scale transformations with the infusion of new "smart" cyber-based technologies to improve their efficiency and reliability. These transitions are being driven by continual advances and cost-efficiencies in areas such as integrated networking, information processing, sensing, and actuation. Hence increasingly, physical infrastructure devices and systems are being tasked to co-exist and seamlessly operate in cyber-based environments. Indeed, tightly coupled systems that exhibit this level of integrated intelligence are often referred to as Cyber-Physical Systems (CPS) [1].

Nowadays, CPS can be found in significantly diverse industries, including, but not limited to, aerospace, automotive, energy, healthcare and manufacturing. Undeniably, the development and adoption of such CPS will generate unique opportunities for economic growth and improvement of quality of life [2]. For instance, in the transportation sector, CPS would be rendered by interactive traffic control systems that aim at creating the notion of zero-fatality highways through automated accident prevention and congestion reduction [3]. Further, in the healthcare sector, CPS would be perceived by wearable and implantable sensors for cost-effective healthcare as well as timely disease diagnosis and prevention [4]. While CPS endeavor great opportunities, the complexity which arises from the fusion of computational systems with physical processes indeed hinders their utmost embracing [5]. Particularly, within the context of security, these integrated systems yield substantial challenges as new vulnerabilities manifest themselves, leading to attack models that are fundamentally new, and predominantly hard to infer, characterize, attribute, and analyze [6]. In turn, these gaps pose immense risks to the physical integrity and operation of critical infrastructures.

Indeed, historical events confirm that industrial control systems have long been the target of disruptive cyber attacks. A few examples include the exploitation of a security flow in the control system of Diesel Generators at Idaho National Laboratories [7] and the prominent Stuxnet malware, which targeted a critical uranium enriching facility, triggering immense plant damage and even endangering human life [8]. Moreover, in March 2016, the U.S. Industrial Control Systems-Computer Emergency Response Team discovered an ongoing malware-orchestrated campaign targeting critical infrastructure. The same campaign was inferred to be responsible for the massive power outage that struck Ukraine in December 2015 [9]. In the same context, given the rapid transformation of industrial

control systems towards CPS-based setups, attacks are indeed anticipated to increase in frequency, sophistication and target diversity. In fact, the latter trend was even recently confirmed by the U.S. Department of Homeland Security (DHS), as they reported thousands of highly-tailored and specifically engineered CPS attacks targeting diverse sectors [10].
While a plethora of research efforts, from both, the control and cyber perspectives have been dedicated to tackling the security of CPS (please refer to Section II), there still exists a significant gap, which is rendered by the lack of properly comprehending and accurately characterizing malicious attackers' capabilities, intents and aims, when targeting such systems. This is largely due to the lack of real malicious empirical data that can be captured, inferred, and analyzed from within the boundaries of operational CPS realms [11, 12]. Thus, without having access to such critical information, it is practically infeasible to elaborate effective security approaches which aim at inferring, attributing or mitigating tangible CPS attacks. Indeed, the goal which endeavors to capture notions of "true maliciousness" in the context of CPS is significantly challenging, due to many factors, including, (1) the lack of complete maturity and the scarcity of elaborative technical details related to CPS [13], (2) the significant diversity of such types of systems which exist in numerous sectors, and (3) logistic and privacy constraints which are often strictly enforced by CPS owners and operators. Therefore, it is evident that auxiliary cyber threat intelligence approaches are required in order to contribute to better grasping the notions of CPS maliciousness. While instrumenting malicious payloads inferred from CPS vulnerabilities [14] and concepts related to CPS honeypots have emerged and have been investigated [15], this paper takes a complementary yet a unique step towards this goal. To this end, we offer a first comprehensive analysis of probing activities that specifically target ample of CPS communication and control protocols by exclusively monitoring and characterizing network traffic targeting a newly deployed network telescope IP address space (i.e., unsolicited Internet traffic targeting routable, allocated yet unused Internet Protocol (IP) addresses) [16]. In this context, we study various dimensions related to such misdemeanors and examine the occurrence and orchestration patterns of such abused services. Indeed, the presented work is innovative and transformative in its capacity to design, implement and evaluate automated approaches that aim at disclosing real CPS attackers' strategies, by passively inferring, characterizing, and correlating CPS probing events. In summary, we frame the paper's contributions in the following three threads:

• Proposing a formal preprocessing probabilistic model that aims at filtering noise (i.e., misconfiguration traffic) that is embedded in darknet data to prepare it for effective use. The model is advantageous as it does not rely on arbitrary cut-off thresholds, provides different likelihood models to distinguish between misconfiguration and other darknet traffic, and is independent from the nature of the source of the traffic. Further, the proposed model neatly captures the natural behavior of misconfiguration traffic as it targets the darknet. To the best of our knowledge, the presented model presents a first attempt ever to systematically fingerprint and thus filter-out darknet misconfiguration traffic.
• Inferring, characterizing and executing multidimensional investigation of probing activities targeting more than 20 CPS protocols by passively monitoring 7 /24 network telescope spaces. To this end, we study their overall trends, abuse per protocol, probes' co-occurrences, source countries and employed protocols, among various others. Additionally, we propose an innovative hybrid approach rooted in temporal analysis and context triggered piecewise hashing to infer, characterize and report on previously undocumented orchestrated and well-coordinated probing activities targeting a number of CPS protocols.
• Validating the proposed models, methods and approaches by experimenting with 50 GB of darknet data, in addition to relying on corroborations against publicly available threat repositories.

The road-map of this paper is organized as follows. In the next section, we discuss related works in terms of CPS security approaches, probing analysis and traffic measurements. In Section III, we elaborate on our proposed approaches, methods and techniques. The corresponding evaluations, inferences and validations are presented in Section IV. We provide a discussion in Section V, while Section VI summarizes this paper and pinpoints an area that paves the way for future work.

II. Related Work

In this section, we review the literature by providing two distinct taxonomies in the context of CPS security approaches from both, the physical/control perspective as well as from the cyber security perspective. We further extend this section by elaborating on probing analysis and traffic measurement studies. The aim is to shed the light on the state-of-the-art of those research areas, in addition to pinpointing various research gaps, which this paper intends to address and/or examine to pave the way for future work.

A. CPS Security: Control-theoretic Approaches

The analysis of CPS security from a control-theoretic perspective has undoubtedly received considerable attention. Table I provides a brief taxonomy, due to space limitations, highlighting some fundamental and representative works in this area. In a nutshell, this taxonomy captures the modeled systems, whether or not noise has been considered in the approach, the analyzed attack model and its corresponding detection scheme. Such research works consider the system dynamics from a physical point of view to perform their analysis. For instance, in the power grid context, Liu et al. [17] investigated false data injection attacks by inserting arbitrary errors into sensor measurements. The authors analyzed two attack scenarios, where the attacker is either constrained to some specific meters, or limited in the resources required to compromise meters. For each scenario, algebraic conditions are derived to validate the existence of stealthy attack vectors, which do not yield any change to the residue. In an alternative work, Pasqualetti et al. [19] analyzed attacks on sensors and actuators by considering a generic continuous-time control system. In particular, the authors mathematically characterized certain conditions that provided the probability of detecting such attacks, given a set of known vulnerabilities. The authors further introduced the notion of attack detectability by designing centralized and distributed filters rooted in arithmetic logic of descriptor systems. In the area of distributed control systems security, Pajic et al. [21] analyzed the impact of malicious nodes in the context of a wireless control network. The

Type of System	Noise	Attack Models	Defense Mechanisms	Reference
Power Grid	✓	False data injection on sensors	Residue detector	[17]
Power Grid	✓	False data injection on sensors	Residue detector	[18]
Control System	-	Attacks on sensors & actuators	Detection filters	[19]
Control System	-	Attacks on sensors & actuators	Optimization decoders	[6]
Control System	✓	Replay attack	χ^2 detector	[20]
Wireless Network	-	State attacks	Output estimator	[21]
Distributed Network	-	State attacks	Combinatorial estimator	[22]
Sensor Network	✓	Dynamic False data injection	Residue detector	[23]

TABLE I: A brief taxonomy of CPS security approaches from a control-theoretic perspective

authors designed and assessed the effectiveness of a detector based on an approach that aims at estimating sensor outputs. Alternatively, Mo et al. [23] considered a data injection attack on a noisy wireless sensor network. The attack was modeled as a constrained optimal control problem in which the Kalman filter was used to perform state estimation, while a failure detector was leveraged to detect anomalies in the system. In addition to the above, Teixeira et al. [24] have introduced and modeled a combination of different attack scenarios such as false data injections, replay, and zero-dynamics' attacks, where adversarial activities attempt to cause damage to the controlled system while remaining stealthy. To this end, active detection methods have been proposed to infer related attacks through analyzing and manipulating the system dynamics. A few of those methods, include, a physical watermarking scheme to authenticate the nominal behavior of a control system [25] and a moving target approach [26] to detect integrity attacks.

Indeed, the rationale behind the aforementioned substantial control-theoretic CPS security contributions is based upon existing models that precisely describe the underlying physical phenomena, which enables the prediction of future behavior and, more importantly, unforeseen deviations from it. To this end, we argue that such approaches (1) do not provide any concrete evidence that such deviations are in fact originated from *malicious* entities, (2) depict attackers' models in a highly-theoretic manner, which do not necessarily reflect the behavior of real CPS attacks and (3) provide experimentation and evaluations that were executed in emulated or simulated CPS environments, without much endeavors being dedicated to real-world applications.

B. CPS Security: Cyber Security Approaches

Complementary to the above, the cyber security research community has also offered various approaches in an attempt to tackle numerous security aspects of CPS. Such approaches characteristically put less emphasis on the control system dynamics by essentially focusing on the cyber (i.e., communication networks, protocols, data, etc.) perspective. We classify a number of such fundamental approaches into four core categories as summarized in Table II and we subsequently discuss only a few of them, due to space limitations. In the context of modeling CPS protocols, Yoon et al. [27] proposed the use of message sequences derived from CPS communication traffic to capture legitimate plant behavior. To accomplish the latter task, the authors employed a dynamic Bayesian network and a probabilistic suffix tree as the underlying predictive model. Executed evaluations using synthetic data demonstrated that the proposed approach is able to accurately model normal traffic, flag certain deviations, and reduce the false positive rate. From another perspective, several research works investigated secure approaches for CPS software and memory resources. For instance, McLaughlin et al. [28] proposed an approach to verify safety-critical code executed on programmable controllers. The devised approach initially checks such code against a set of physically safe measures and subsequently present case studies of abuse in case of any inferred inconsistencies. In this context, the authors introduced the notion of temporal execution graph, which illustrates the consequences of a certain untrusted executed code. The proposed approach was validated in terms of its capability to enforce certain common safety properties by means of experimentation in an emulated environment. Several other research initiatives exploited CPS process variables for anomaly detection. For example, Hadžiosmanović et al. [11] extracted process variables from a CPS plant to build predictability models. By leveraging simple regression models, the authors alerted CPS plant operators of any deviation in the expected parameters as an indicator of an ongoing attack. From a data analytics perspective, Almalawi et al. [29] presented a machine learning approach to infer CPS attacks. By employing an unsupervised clustering mechanism based on the k-means algorithm, the proposed approach aims at distinguishing between consistent and inconsistent CPS observations. Simulations were conducted to validate the effectiveness of the devised approach. Within the same category of research works but from an industrial/operational perspective, the security community supporting the open source intrusion detection system Snort [30] has also offered and contributed to various CPS detection rules [31]. The latter aim at inferring unauthorized requests, malformed packets and rarely used and suspicious CPS protocol commands.

While the surveyed research works offer significant contributions, nevertheless, we can extract (1) the general inadequacy of research attempts to systematically combine or at least diminish the gap between cyber and control capabilities for securing CPS, (2) the lack of empirical data related to tangible malicious CPS attacks and strategies that are generated from real unsolicited attackers, which could realistically affect the stability and security of CPS, (3) the deficiency of CPS security approaches in providing, both, attribution evidence and threat severity metrics and (4) the lack of such approaches in providing means for CPS resiliency in the physical realm during or immediately after an attack. Our presented work that falls within this category of research works aims at

Analysis Perspective	Highlights	References
Protocol Vulnerabilities	Modeling CPS protocols to detect anomalies	[27, 32–36]
PLC Software	Verifying PLC code and memory to prevent violations	[28, 37–40]
Process Variables	Predicting CPS process behavior to detect anomalies	[11, 14, 41, 42]
Data Analytics	Data-driven approaches to infer CPS cyber attacks	[29, 31, 43–45]

TABLE II: A brief classification of CPS security approaches from a cyber security perspective

contributing to point (2) by providing a first thorough look in terms of insights and inferences related to CPS attackers' reconnaissance strategies, by investigating unsolicited darknet data.

C. Probing Analysis

In the context of inferring probing events, Li et al. [46] considered large spikes of unique source counts as probing events. The authors extracted those events from network telescope traffic using time series analysis; they first automatically identified and extracted the rough boundaries of events and then manually refined the event starting and ending times. At this point, they used manual analysis and visualization techniques to extract the event. In an alternate work, Jin et al. [47] considered any incoming flow that touches any temporary dark (grey) IP address as potentially suspicious. The authors narrowed down the flows with sustained suspicious activities and investigated whether certain source or destination ports are repeatedly used in those activities. Using these ports, the authors separated probing activities of an outside host from other traffic that is generated from the same host. In the area of analyzing probing events, the authors of [47, 48] studied probing activities towards a large campus network using netflow data. Their goal was to infer the probing strategies of scanners and thereby assess the harmfulness of their actions. They introduced the notion of gray IP space, developed techniques to identify potential scanners, and subsequently studied their scanning behaviors. In another work, the authors of [46, 49] presented an analysis that drew upon extensive honeynet data to explore the prevalence of different types of scanning. Additionally, they designed mathematical and observational schemes to extrapolate the global properties of scanning events including total population and target scope. In the context of probing measurement studies, Benoit et al. [50] presented the world's first Web census while Heidemann et al. [51] were among the first to survey edge hosts in the visible Internet. Further, Pryadkin et al. [52] offered an empirical evaluation of IP address space occupancy whereas Cui and Stolfo [53] presented a quantitative analysis of the insecurity of embedded network devices obtained from a wide-area scan. Additionally, Durumeric et al. [54] investigated darknet traffic to analyze the current practices of Internet-wide scanning. They generated cyber threat intelligence related to sources of scanning activity and probed services, among others. The authors also elaborated on some defensive mechanisms and provided several insightful recommendations when executing such activities for research purposes. Furthermore, Dainotti et al. [55] presented a pioneering measurement and analysis study of a 12-day Internet-wide probing campaign targeting VoIP (SIP) servers, while an "anonymous" presented and published online [56] what they dubbed as the Carna botnet. The author exploited poorly protected Internet devices, developed and distributed a custom binary, to generate one of the largest and most comprehensive IPv4 census ever.

In this work, we extend previous research contributions to identify a new threat vector; unsolicited sources employing Internet-scale scans in an attempt to fingerprint numerous CPS resources. To this end, we devise and implement innovative methods and techniques, and apply them on network telescope data to perform a comprehensive measurement and analysis of a broad list of CPS communication and control protocols.

D. Network Telescope: Measurements & Analysis

The idea of monitoring unused IP addresses for security purposes was first brought to light in the early 1990s by Bellovin for AT&T's Bell Labs Internet-connected computers [57, 58]. However, the topic of telescope data analysis received further attention after year 2000 [59]. Since then, the focus of network telescope studies has shifted several times, closely following the volatile nature of new threat actors. For instance, some of the important contributions that demonstrate the evolution of telescope research include the discovery of the relationship between backscatter traffic and DDoS attacks in 2001 [60], worm propagation analysis between 2003 and 2005 [61, 62], the use of time series and data mining techniques on telescope traffic in 2008 [63], the monitoring of large-scale cyber events through telescope in 2012 [64], and more recently, the study of amplification attacks using telescope sensors in 2013 and 2014 [65, 66].

In contrast, this work proposes and evaluates a formal probabilistic preprocessing model for network telescope traffic in an effort to fingerprint and filter out misconfiguration traffic. We believe that this proposed model is of significant value, given its postulated highly applicable nature in the field of Internet measurements.

E. CPS Traffic Analysis

CPS network traffic monitoring and analysis can be divided in two main categories, namely, interactive monitoring and passive monitoring. On one hand, honeypots are an example of low- to high-interactive trap-based monitoring systems [67]. The first CPS honeypot, known as the SCADA HoneyNet Project, was designed and deployed in 2004 by Cisco Systems [68]. Digital Bond, a company that specializes in CPS cybersecurity, deployed two SCADA honeypots in 2006 [69]. The release of Conpot in 2013 has greatly facilitated the deployment and management of CPS honeypots [15]. While such honeypots provide an effective mechanism to generate real CPS attack models, they indeed suffer from two drawbacks. First, improper deployment of honeypots may introduce security risks (i.e., malware escaping the honeypot sandbox and propagating to the production network). Second, honeypots are

only effective if they are not detected; there exist substantial evidence that honeypots can be relatively easily fingerprinted [70, 71]. In terms of passive analysis, such methods include the study of network telescope traffic to generate statistics and trends related to various inferred CPS misdemeanors. The first limited reported network telescope study which addressed the security of CPS protocols was conducted in 2008 by Team Cymru [72]. Their report included coarse statistics on scans targeting commonly used CPS protocols such as DNP3, Modbus and Rockwell-encap.
In this work, in addition to providing a thorough measurement and analysis study of probing traffic targeting ample of CPS protocols, we further propose, evaluate and validate a novel approach to infer and report orchestrated, stealthy and previously undocumented probing activities targeting a number of CPS protocols.

III. Proposed Approach

In this section, we elaborate on the devised models, approaches and methods that aim at (1) cleansing darknet data to prepare it for effective use, (2) inferring and characterizing CPS probing traffic and (3) identifying orchestrated CPS probing campaigns.

A. Darknet Preprocessing Model

Although darknet data mostly contains malicious packets originating from probes, backscattered packets from victims of distributed denial of service attacks and malware propagation attempts, among others, it might also include what is dubbed as misconfiguration traffic. The latter non-malicious packets might be caused by network/routing or hardware/software faults that were erroneously directed towards a darknet. Such traffic can also be an artifact of an improper configuration when deploying a darknet. Indeed, misconfiguration traffic "pollutes" darknet data as such traffic can not be exploited for cyber threat intelligence. Further, misconfiguration traffic makes it harder for cyber threat intelligence algorithms to operate correctly on darknet data, which often yields to numerous undesirable false positives and false negatives. Another drawback of the existence of misconfiguration traffic within darknet data, is that it wastes valuable storage resources.

Therefore, in this section, we elaborate on the proposed probabilistic model that is particularly tailored towards the goal of preprocessing darknet data by fingerprinting and thus filtering out misconfiguration traffic.

In a nutshell, the model formulates and computes two metrics that aim at capturing the natural and the characteristic behavior of misconfiguration flows as they target the darknet IP space. The model initially estimates the "rareness of access"; the degree to which access to a given darknet IP address is unusual. The model further considers the "scope of access"; the number of distinct darknet IP addresses that a given remote source has accessed. Subsequently, the joint probability is formulated, computed and compared. If the probability of the source generating a misconfiguration is higher than that of the source being malicious (or unsolicited), then the source is deemed as one that is generating misconfiguration traffic, subsequently flagged, and its corresponding generated darknet flows are filtered out. The above two metrics are elaborated next.

Let $D = \{d_1, d_2, d_3, \cdots\}$ represent the set of darknet IP addresses and D_i a subset of those accessed by source s_i. First, the model captures how unusual the accessed destinations are. The idea behind this metric stems from the fact that misconfigured sources access destinations that have been accessed by few other sources [73]. Thus, the model estimates the distribution of a darknet IP d_i being accessed by such a source as

$$P_{misc}(d_i) = \frac{n_s(d_i)}{\sum_{\forall d_j \in D} n_s(d_j)}, \quad (1)$$

where $n_s(d_i)$ is the number of sources that have accessed d_i. In contrast, a malicious darknet source will access a destination at random. Typically, defining a suitable probability distribution to model the randomness of a malicious source targeting a specific darknet destination is quite tedious; often a simplistic assumption is applied to solve this issue. In this context, a very recent work by Durumeric et al. [54] has demonstrated that darknet sources will probe their targets following a Gaussian distribution [1]. By adopting that assumption, one can model the probability of a darknet destination accessed by a malicious source as

$$P_{mal}(d_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad (2)$$

where σ is the standard deviation, μ is the mean, σ^2 is the variance and x is the location of the darknet destination following the distribution. Recall that equations (1) and (2) allow the model to initially capture how unusual the accessed destinations are. However, further, the model considers how many darknet destinations have been accessed by a given source. The latter will be subsequently described.

Given a set of D_i, darknet destinations accessed by a specific source s_i, the model eventually aims at measuring two probability distributions, namely, $P_{misc}(D_i)$ and $P_{mal}(D_i)$. The former being the probability that D_i has been generated by a misconfigured source while the latter is the probability that D_i has been generated by a malicious darknet source.

Let $D_1 = \{d_{i1}, d_{i2}, d_{i3}\}$ be those darknet addresses accessed by s_1. The model captures the probability $P(D_1)$ of the source generating $\{d_{i1}, d_{i2}, d_{i3}\}$ as the probability of s_1 accessing this specific combination of destinations knowing that it targeted three destinations multiplied by the probability of s_1 accessing any three destinations. The latter could be generalized and formalized as

$$P(D_i) = P(D_i = \{d_{i1}, d_{i2}, \cdots, d_{in}\} \mid |D_i| = n) \times P(|D_i| = n). \quad (3)$$

For both, a misconfigured and a malicious source, the first term of equation (3) could be modeled as

$$P_{misc}(D_i = \{d_{i1}, d_{i2}, \cdots\} \mid |D_i|) = \frac{1}{K} \prod_{\forall d_j \in D_i} P_{misc}(d_i) \quad (4)$$

[1] In the presence of Network Address Translation (NAT), different IP addresses that are simultaneously probing and generating misconfiguration traffic would cause the distribution to be non-Gaussian. This might lead to falsely attributing probing traffic as misconfiguration. While we can not deny or validate this scenario, future work could investigate the empirical distribution of such phenomena to filter it out.

$$P_{mal}(D_i = \{d_{i1}, d_{i2}, \cdots\} \mid |D_i|) = \frac{1}{K} \prod_{\forall d_j \in D_i} P_{mal}(d_i) \quad (5)$$

where K, a normalization constant which is solely employed to allow the probabilities to sum to 1, could be defined as

$$K = \frac{|D|!}{n!(|D| - n)!} \times \frac{1}{|D|^n}. \quad (6)$$

Please note that K is a typical normalization constant that is often employed in Bayesian probability [74]. Further, n represents all the sources in the data set, while $|D|$, as previously mentioned, represents the darknet IP space.

The likelihood that a source will target a certain number of darknet destinations (i.e., the second term of equation (3)) depends on whether the source is malicious or misconfigured. Characteristically, misconfigured sources access one or few destinations while malicious sources access a larger pool of destinations. We have modeled such distributions as

$$P_{misc}(|D_i|) = \frac{1}{(e-1)|D_i|!} \quad (7)$$

$$P_{mal}(|D_i|) = \frac{1}{|D|} \quad (8)$$

where the term $(e-1)$ in equation (7) guarantees that the distribution will sum to 1. It is noteworthy to mention that equation (7) ensures that the probability will significantly decrease as the number of targeted destinations increases. In contrast, equation (8) captures a malicious darknet source accessing a random number of darknet addresses.

By combining the above equations, we can model the probability of a source being misconfigured or malicious, given a set of darknet destination addresses, as

$$P_{misc}(D_i) = \frac{1}{K(e-1)|D_i|!} \prod_{\forall d_j \in D_i} P_{misc}(d_i) \quad (9)$$

$$P_{mal}(D_i) = \frac{1}{K|D|} \prod_{\forall d_j \in D_i} P_{mal}(d_i). \quad (10)$$

It is imperative to note that equations (9) and (10) provide two distinct likelihood models to distinguish between misconfiguration and other malicious darknet traffic. This permits the simplified and systematic post-processing of the latter two types of darknet traffic. Moreover, as the model generalizes and formalizes the concepts of misconfiguration and other malicious darknet traffic, the proposed model does not make any assumptions related to the nature of the sources of those types of traffic. For example, the model is agnostic to whether the malicious (or unsolicited) traffic is generated by a worm or a probing tool, or whether the misconfiguration is caused by a malfunctioning Internet router or an invalid connection request.

Algorithm 1 Inferring misconfiguration flows using the probabilistic model

1: **Input:** Darknet Flows, *DarkFlows*
2: **Output:** Flag, *MiscFlag*, indicating that the DarkFlow is originating from a misconfigured source
3:
4: **for** $DarkFlows$ **do**
5: *MiscFlag* $\leftarrow 0$
6: *i* $\leftarrow$ DarkFlows.*getUniqueSources()*
7: Amalgamate $DarkFlows_i$ originating from a specific source s_i
8: Update $s_i(D_i)$
9: Compute $P_{misc}(D_i), P_{mal}(D_i)$
10: **if** $P_{misc}(D_i) > P_{mal}(D_i)$ **then**
11: *MiscFlag* $\leftarrow 1$
12: **end if**
13: **end for**

To effectively employ the proposed darknet preprocessing model, we present Algorithm 1, which provides a simplistic yet effective mechanism to infer misconfigured sources by employing the model. It is worthy to note that step 9 of the algorithm (i.e., the computation of $P_{misc}(D_i)$ and $P_{mal}(D_i)$) is easily accomplished in practice by computing the negative log-likelihoods,

$$\begin{aligned} L_{misc}(D_i) &= -lnP_{misc}(D_i) \\ L_{mal}(D_i) &= -lnP_{mal}(D_i). \end{aligned} \quad (11)$$

Thus, Algorithm 1 deems a source and its corresponding flows as misconfiguration traffic if $L_{mal}(D_i) - L_{misc}(D_i) > 0$.

B. CPS Probing Inference & Characterization

To infer CPS probing activities after preprocessing darknet data by exploiting the previously proposed model, we present Algorithm 2, which exploits both packet header information and flow-based parameters.

Algorithm 2 operates on darknet flows, which are defined by a series of consecutive packets sharing the same source IP address. First, each flow is scrutinized to verify if its corresponding packets contain any service ports that can effectively pinpoint to CPS darknet activities. The algorithm monitors traffic to a comprehensive list of most prominent and widely-deployed CPS communication and control protocols as summarized in Table III[2]. Such table includes, both, private (i.e., Siemens, GENe, etc.) and well-known CPS services (i.e., Modbus, ICCP, etc.).

If no CPS ports were found, the algorithm deems that specific flow as not related to CPS activities. On the contrary, if a service port was found, Algorithm 2 deems that flow as suspicious, and consequently moves forward in an attempt to assert that suspicion. Subsequently, the algorithm counts the number of packets per flow to measure the rate of the suspicious activities within a certain time window (*Tw*). If the flow packet count (*pkt_cnt*) is beyond a specific threshold, the flow is deemed as a CPS probe. To this end, we borrow the packet count threshold from [75], defined by 64 probed darknet addresses on the same port on any given day. Please note,

[2]Obtained through discussions with state and federal CPS operators in the power, water, aviation and critical manufacturing sectors.

Algorithm 2 CPS Scanning Inference Algorithm

Input: A set (*F*) of unique darknet flows (*f*),
Each flow *f* contains packet count (*pkt_cnt*) and rate (*rate*)
SP: CPS Service Port
Tw: Time window
Pth: Packet threshold
Rth: Rate threshold,
Tn: Time of packet number *n* in a flow
pkt: Packet
Output: CPS flag, *CPS_flag*

for Each *f* in *F* **do**
  while *pkt* in *f* **do**
    if *pkt.contains()* ! = *SP* **then**
      CPS_flag() ← 0
    end if
    if *pkt.contains()* = *SP* **then**
      CPS_flag() ← 1
    end if
  end while

  pkt_cnt ← 0
  T1 ← pkt_gettime()
  Tf ← *T1* + *Tw*
  while *pkt* in *f* **do**
    Tn= pkt_gettime()
    if *Tn* < *Tf* **then**
      pkt_cnt ← pkt_cnt + 1
    end if
  end while
  rate ← $\frac{pkt\_cnt}{Tw}$
  if *pkt_cnt* < *Pth* || *rate* < *Rth* **then**
    CPS_flag() ← 0
  end if
end for

CPS Communication & Control Protocols	Port Number	Type
ABB Ranger 2003	10307/10311/10364, etc.	Registered
BACnet/IP	47808	Registered
DNP/DNP3	19999/20000	Registered
Emerson/Fisher ROC Plus	4000	Registered
EtherCAT	34980	Registered
EtherNet/IP	2222/44818	Registered
FL-net Reception/Transmission	55000-55003	Dynamic/Private
Foundation Fieldbus HSE	1089/1090/1091	Registered
Foxboor/Invensys Foxboro DCS	55550	Dynamic/Private
Iconic Genesis32 GenBroker	18000	Registered
ICCP	102	Well-known
IEC-104	2404	Registered
Johnson Controls Metasys N1	11001	Registered
Modbus	502	Well-known
MQ Telemetry Transport	1883	Registered
Niagara Fox	1911/4911	Registered
OPC UA Discovery Server	3480	Registered
OSIsoft PI Server	5450	Registered
PROFINET	34962/24963/34964	Registered
Project/SCADA Node Primary Port	4592	Registered
Red Lion	789	Well-known
ROC Plus	4000	Registered
SCADA Node Ports	4592/14592	Registered
Siemens Spectrum Power TG	50001/50018/50020, etc.	Dynamic/Private
SNC GENe	62900/62911/62924, etc.	Dynamic/Private
Telvent OASyS DNA	5050/5052/5065, etc.	Registered

TABLE III: List of CPS protocols and corresponding ports

that typically, the probing engine would have also required and established a rate threshold (*Rth*). Nevertheless, we do not enforce one here, to enable the algorithm to infer very low rate, possible stealthy activities. Indeed, coupled with the analysis of the packet features, the approach embedded in Algorithm 2 would fingerprint Internet-scale CPS probing traces. From a performance perspective, when implemented "on the fly" on the darknet data stream, the algorithm can successfully process and reason about a significant 10,000 flows in 55 seconds, on average.

After inferring CPS probing activities, we characterize and profile their inferred flows to investigate their significance and prevalence, recent trends and orchestration behavior. We generate amalgamated statistics related to the categories of probes, the distribution of different types of scans within each category, the distribution of transport protocols used in the scans and the time series of various types of probes. Furthermore, we attempt to understand the behavior of these probes by studying and analyzing the similarity and co-occurrence patterns of their sources. To this end, we execute the following procedure. In order to investigate whether the same CPS scanning sources are prevalent on different time periods, we formalize the probes as a set of unordered collection of IP addresses that represent scanners observed on a daily basis. Subsequently, we define the sets, A_i, (where $i = 1, 2..., n$), where each set is indexed by the day number. We compute the similarity between two sets, A_i and A_j, represented by $S(A_i, A_j)$. The similarity has the following properties: (1) It is in the range $[0, 1]$; (2) the higher the value is, the more source probing IP addresses are shared among the sets, and thus the largest is the similarity, and vice versa; (3) it is equal to 1, if the sets are the same; and (4) it is equal to 0, if the sets do not share IP addresses or one (or both) of the sets is empty. To compute such similarity, we leverage the Jaccard index, a statistical technique used in comparing similarities and diversities among sample sets [76]. This index is defined by:

$$S(A_i, A_j) = \frac{|A_i \bigcap A_j|}{|A_i \bigcup A_j|} \tag{12}$$

where $|A_i \bigcap A_j|$ denotes the cardinality of the intersection between the sets and $|A_i \bigcup A_j|$ denotes the cardinality of the union between the sets.

C. CPS Probing Orchestration Fingerprinting

In recent years, there has been a noteworthy shift towards a new phenomenon of probing events that could be dubbed as probing campaigns. These are distinguished from previous probing incidents as (1) the population of the participating bots is several orders of magnitude larger, (2) the target scope is generally the entire IP address space, and (3) the sources adopt well-orchestrated, often botmaster-coordinated, stealth scan strategies that maximize targets' coverage while minimizing redundancy and overlap [55, 56]. In this section, we build upon our darknet preprocessing model as well as our CPS probes' inference algorithm by proposing a clustering approach to infer CPS probing campaigns. This aims at better comprehending the natures as well as the type of maliciousness of such campaigns; for instance, it could be found, through investigation, that a specific campaign is specialized in targeting particular critical infrastructure resources. Further, the proposed fingerprinting approach allows the elaboration of the actual scope and characteristics of the inferred campaigns in an effort to provide accurate measurements as well as aid in CPS situational awareness, analysis and attribution. In this context, previous works [77] suggested that coordinated unsolicited

sources within a campaign probe their targets in a similar fashion. Indeed, the proposed approach exploits this idea by automatically building notions of similar probing behavioral characteristics. To achieve this challenging task, given that we are exclusively dealing with darknet data [46], the proposed approach applies a fusion of a time-series technique in conjunction with a network forensic analysis approach between the previously inferred (independent) CPS probing flows.

1) Time-Series Analysis: As a first step, the proposed approach attempts to infer temporal similarities between the previously inferred CPS probing activities. To this end, we leverage a time-series approach rooted in Dynamic Time Warping (DTW) [78]. Motivated by its successful experimentation in diverse research areas [78, 79], the DTW technique measures the resemblance between data sequences independent of their rates. The aim is to cope with possible time deformations affiliated with time-dependent data [80]. The DTW takes two vectors defining the time series as input and produces a distance unit characterizing their temporal similarities.

2) Netflow Analysis: As a second step, the proposed approach leverages the context triggered piecewise hashing (CTPH) [81] technique using a customized developed version of ssdeep[3], which exploits netflow characteristics. The CTPH technique, which is derived from the digital forensics research field, is advantageous in comparison with typical hashing as it can provide a percentage of similarity between two traffic samples rather than producing a null value if the samples are slightly different. CTPH generates a percentage of similarity defining netflow similarities between any two given CPS probing sessions.

To this end, to infer orchestrated CPS probing campaigns, the proposed approach selects and clusters those CPS probing sessions that minimize the DTW similarity metric while maximizing the CTPH measure targeting the same CPS protocol.

IV. Results

The generated results are based upon scrutinizing 50 GB of darknet data that were collected from a newly deployed network telescope IP space during a one month period between April and May, 2016. Please note that while we do not claim that the design, deployment and management of such darknet is part of this paper's contributions, nevertheless, we have to pinpoint that this is a first of a kind cyber threat intelligence gathering project in our region. We organize this section following closely the previously proposed models and approaches.

A. Darknet Preprocessing Model

We implemented a prototype of the proposed model of Section III-A in `Java` using the `jNetPcap`[4] library. To execute the proposed model on the darknet dataset, we aggregate the connections into sessions using an approach similar to the first step algorithm by Kannan et al. [82]. We consider all those connections within T_{aggreg} of each other as part of the same session for a given pair of hosts. We used the same proposed threshold, T_{aggreg} = 100 seconds, and found that this seems to correctly group the majority of connections between any given pair of hosts. To validate the outcome of the proposed model, we compare it against the baseline; classifying misconfiguration traffic as any darknet traffic that is not scanning or backscattered [83]. The latter is a commonly employed technique, given the lack of other available formal literature approaches.

Figure 1 depicts the outcome of the execution of the proposed model on the extracted sessions while Table IV summarizes the outcome of the baseline. By comparing Table IV and Figure 1, we can notice that the proposed model fingerprinted a lower percentage of misconfiguration traffic than the baseline. A semi-automated verification (i.e., using scripts and manual investigation) validated that all the sessions that the model inferred as misconfiguration traffic are true misconfiguration packets, where almost 50% of them are malformed packets while the rest are packets that targeted the darknet IP space only once. We further investigated the 4.7% darknet sessions that the baseline experiment has inferred as misconfiguration traffic and noticed that they are indeed false positives related to UDP amplification probes [66]. Thus, we can safely claim that the proposed model was accurate in distinguishing between darknet misconfiguration traffic and other malicious (or unsolicited) darknet traffic, compared to the baseline. In terms of processing performance, we were solely interested in inferring the execution time of the prototype; the time from which a darknet dataset is fed into the prototype, until the time the prototype flags the misconfiguration, filters-out such traffic and generates a new "clean" dataset. We executed the experiment on a single commodity machine running Ubuntu 16.04 LTS with an Intel Core-i7, 64-bit processor and 16 GB of RAM. The output disclosed that in order to achieve the intended tasks, the prototype approximately required, on average, 14 minutes to completely process 1 hour of darknet data. For our current tasks in hand that do not require very large measurement studies and given the accuracy and automation that is offered by the proposed model, we believe that such result is acceptable. Future work will address the performance of the proposed model by (1) dropping the `Java` implementation in favor

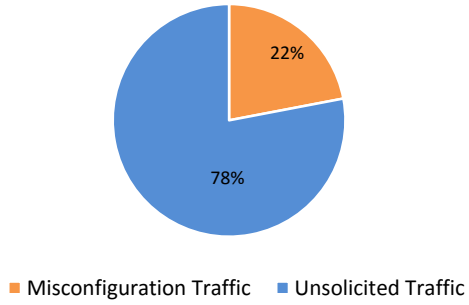

Fig. 1: Proposed Model: Distribution of darknet sessions

Scanning Traffic	Backscatter Traffic	Misconfiguration
65.1%	8.2%	26.7%

TABLE IV: Baseline: Distribution of darknet sessions

[3]ssdeep: http://ssdeep.sourceforge.net/

[4]jNetPcap: http://jnetpcap.com/?q=jnetpcap-1.4

of a C implementation that leverages the libpcap[5] library, and (2) employing multi-threading and parallel programming paradigms.

B. CPS Probing Inference & Characterization

After preprocessing the darknet dataset, we aimed at inferring and characterizing probing events targeting 120 CPS communication and control protocols covering 26 CPS services (please recall Table III). Table V provides an overview of the CPS scanning activities as inferred by the proposed algorithm. In total, we have identified 33897 CPS probing events targeting 20 CPS protocols. Figure 2 illustrates the distribution of such events during the analyzed period. In an effort to validate the occurrence of the inferred CPS probing activities, we performed the following tasks. First, we relied on third-party publicly available threat repositories provided by Cymon[6] and AbuseIPDB[7]. These repositories index Internet-scale suspicious IP addresses as reported by service providers and backbone network operators. They also identify the probable attack category. We cross-matched the inferred CPS probing events with those repositories.

	April Week3	April Week4	May Week1	May Week2
Total Scanners	7954	8871	8731	8341
Total Unique Scanners	3007	3727	3950	3731

TABLE V: Inferred CPS probing events

Our findings revealed that approximately 4.37% of scanners were indeed involved in various malicious reported activities (i.e., hacking (41.25%), portscan (31.46%), FTP/SSH brute force (13.28%), and DDoS (6.29%)). In an auxiliary attempt to validate the occurrence of the remaining scanners, we relied on DShield[8] data. By performing this, we were able to validate 88.1% of the remaining scanners, which generated 1710065 malicious activities and were involved in 151799 unique attacks, as reported by DShield. The residual 7.53% of scanning sources have never been reported in any publicly available dataset that we could find. However, our manual inspection indicated that 80% of them belong to an unsolicited CPS probing campaign (campaign B to be discussed in Section IV-C). Thus, in total, through our validation approach and manual inspection, we were able to validate the occurrence of all the inferred CPS probing sources, except 1.34%, which close to half of them were confirmed, by investigating their corresponding packets, to be related to misconfiguration traffic that were not filtered out correctly by the preprocessing model.

Our investigation revealed that 98% of the events are TCP-based, where they are rendered by vertical probing activities, in an effort to verify all running CPS services on a single host. CPS probing events that exploit the UDP were scarce (close to 2%) in our analyzed dataset and predominantly targeted BACnet (on port 47808) and Ethernet (on ports 22222/44818). In contrast, Modbus, ICCP, Niagara Fox and DNP3 were among the top abused TCP CPS services. We geo-located the probing sources of the most prominent probed services as depicted in Figure 4. The outcome demonstrates that the United States leads in terms of generating most of the probes. Additionally, we infer distributed horizontal probing events towards Germany, China, and France, and horizontal probing activities from Japan, Russia, Canada and Korea that respectively target the ICCP, Foundation Fieldbus, ROC and Modbus CPS protocols. Moreover, we infer other horizontal probing activities from Spain and Singapore, which simultaneously target the ICCP and Modbus services.

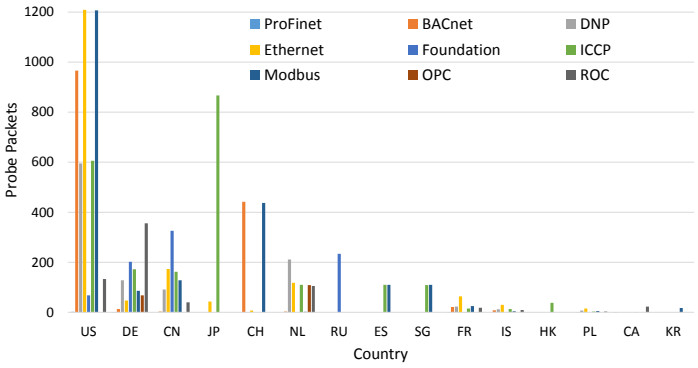

Fig. 4: CPS service scans: Top source countries

We proceed by analyzing two packet features, namely, IP Identification (*ip-id*) and source port (*src-port*), related to the inferred CPS probing events. Both features are typically used to make inferences related to the service generating the traffic [54]. On one hand, Table VI lists the top five *ip-id* values and their counts. The results revealed that the majority of the used *ip-id* values are consistent with probes generated from the Zmap probing tool, which has a default *ip-id* equal to 54321 (0xd431 in hex) [54]. As such, we can state that around 90% of the inferred CPS probing traffic are indeed generated from Zmap. On the other hand, Table VII summarizes the top employed source ports. We inferred a significant amount of probes originating from typically abused ports such as 6000 (i.e., often reported to be used by trojans). We have also noticed that the majority of the traffic have been received via specific ports within the 40k and 60k range. While analyzing Modbus communication, we have noted that around 30% of its traffic originated from source port 6706, which is the only port that consistently appeared during the entire analyzed period. We currently have no tangible explanation of such traffic, but we will be investigating their packets' details in the near future.

April Week3	April Week4	May Week1	May Week2
0xd431 (13060)	0xd431 (12632)	0xd431 (11640)	0xd431 (12849)
0x0100 (820)	0x0100 (343)	0x0100 (566)	0x0100 (530)
0x0049 (11)	0x0b1c (10)	0x843d (9)	0x0438 (13)
0x9625 (9)	0x052a (10)	0x591e (9)	0xb530 (9)
0x0ae7 (9)	0x058d (9)	0x01da (9)	0x8faf (9)

TABLE VI: Top five *ip-id* values (Probe packet count)

Consistent with Section III-B, we now investigate whether

[5]tcpdump: http://www.tcpdump.org/
[6]Cymon Open Threat Intelligence: https://cymon.io/
[7]AbuseIPDB: https://www.abuseipdb.com/
[8]DShield: https://www.dshield.org/

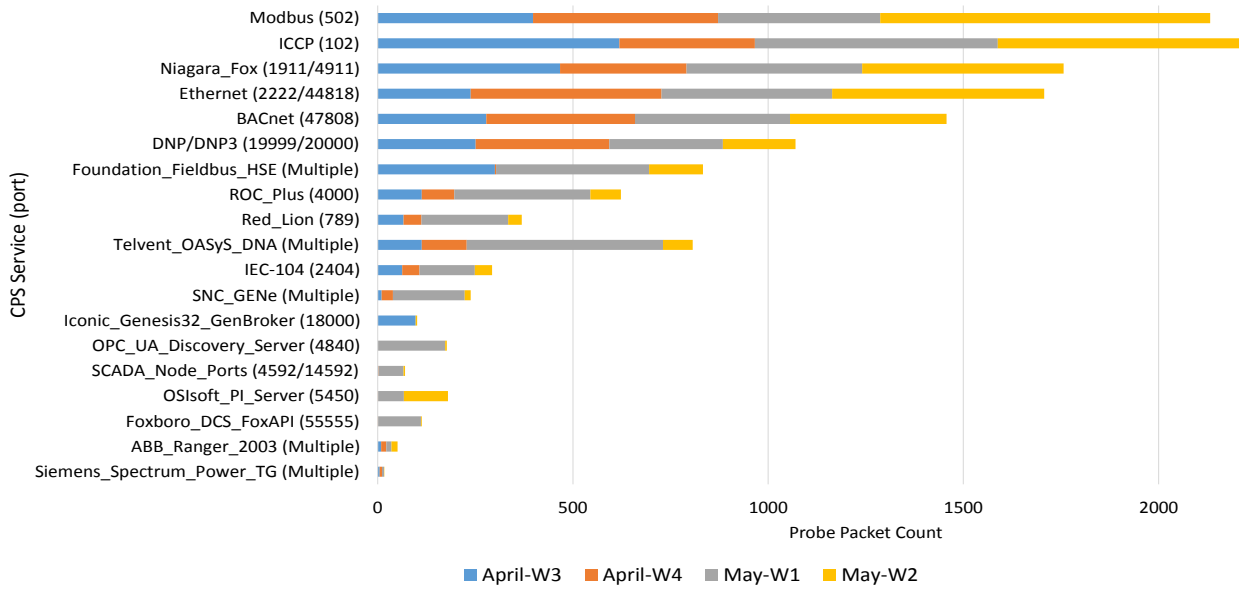

Fig. 2: TCP and UDP probed CPS services

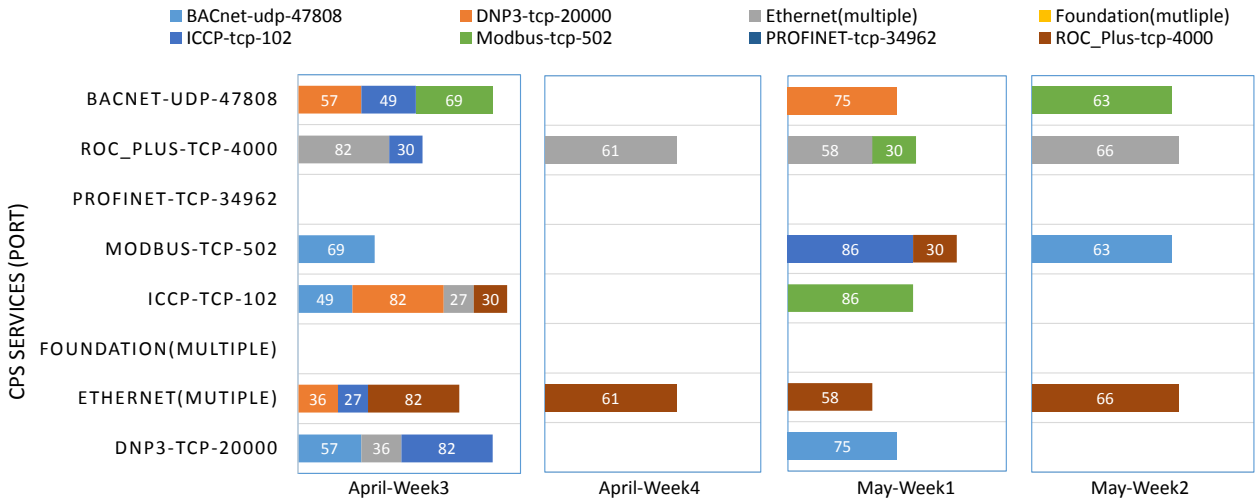

Fig. 3: The co-occurrence patterns of the inferred CPS probing events

April Week3	April Week4	May Week1	May Week2
6000 (609)	53 (535)	1048 (785)	6000 (426)
53933 (348)	43490 (356)	42880 (576)	60000 (330)
53 (315)	6000 (235)	53 (334)	53 (314)
43490 (267)	22 (214)	59651 (223)	63030 (156)
59531 (244)	1048 (146)	58017 (221)	50449 (128)

TABLE VII: Top five used/abused *src-port* (Probe packet count)

the same CPS scanning sources are prevalent on different time periods by deriving their co-occurrence patterns. Figure 3 shows the extracted patterns for the most probed CPS services, which are visualized in associated and correlated colors. Our analysis revealed one clear consistent pattern that remained active during the entire analyzed period. Such pattern is an association between sources probing the ROC PLUS protocol by abusing TCP port 4000 and those targeting the Ethernet UDP ports 22222 and 44818. In this pattern, 55 to 82 IP addresses were always persistent probing those CPS services. Another interesting finding is related to the ICCP service on TCP port 102, in which it was found to be probed with 5 other services, namely, BACnet, DNP3, Ethernet, ROC and Modbus. In these probes, 27 and 86 IP addresses concurrently shared the probing task. Additionally, we have pinpointed probing sources that exclusively targeted the Foundation Fieldbus and PROFINET services, where their probing sources did not probe any others CPS protocols. We also note that the probing sources targeting the Modbus protocol, also targeted 6 other (CPS and non-CPS) protocols 74% of the time. By focusing on the Modbus protocol, which was the most probed in our analyzed dataset, we can infer from Figure 5 that 12 Modbus scanners remained active during the entire analyzed period. Although we note such continuous and consistent activities

among Modbus scanners, the majority originated from relatively new sources, with an average of 121 new probing sources per week.

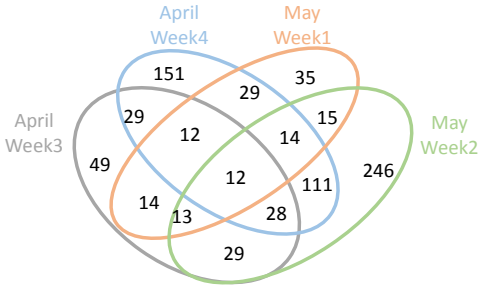

Fig. 5: Co-occurrences within Modbus sources

C. CPS Probing Orchestration Fingerprinting

Indeed, the previously inferred CPS probing events appear to originate from independent probing sources. However, consistent with Section III-C, we now execute the proposed approach to examine the existence of orchestrated CPS probing events. The proposed approach identified 9085 probing events generated from 58 campaigns. Figure 6 provides a holistic depiction of the inferred campaigns, where the nodes represent unique source IP addresses and the edges represent the existence of a concrete derived similarity based on the analyzed generated probing traffic per the proposed approach of Section III-C; one can notice the appearance of several large-scale CPS probing campaigns. In the sequel, we only elaborate on 5 of those campaigns that were shown to be of large-scale (i.e., have at least 50 sources). It is noteworthy to mention that around 60% of the inferred campaigns (including those 5 large-scale campaigns) possessed a very low portsweep probing rate, rendering them independent and/or undetectable by typical intrusion detection systems or firewall rules. In contrast, our proposed methods in terms of inference algorithm and orchestration fingerprinting, can evidently assess those seemingly-independent probes to infer their underlying coordination.

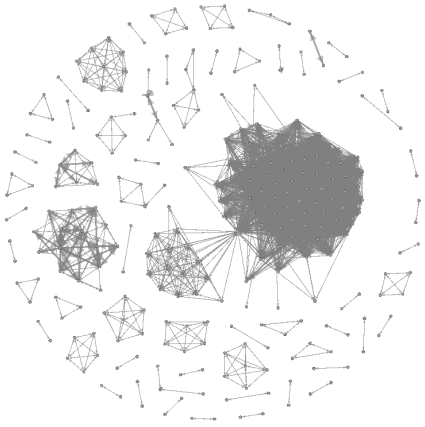

Fig. 6: A holistic illustration of the inferred orchestrated CPS probing events

The first two probing campaigns were found to be generated by an organization and an academic institution. A verification of their IP ranges and host names revealed that they are known to perform probing activities for cyber security and research purposes. Table VIII provides an overview of these unsolicited campaigns.

Reference	Source domain	Number of distinct IP addresses
A	*.edu	64
B	*.io	136
C	*.com *.de	188
D	*.cn	116
E	*.ru	54

TABLE VIII: Inferred CPS probing campaigns

We have encountered a unique probing behavior while investigating probing campaign A of Table VIII. The campaign has indeed conducted 6 operations during the analyzed one-month period. In fact, we have identified this orchestrated campaign operating in a product of 16 unique hosts, leveraging 64 (16×4) distinct IP addresses, running in parallel, from random ports, and searching for a specific list of CPS protocols in the following sequence: Modbus on TCP port 502, Niagara Fox on TCP port 1911 and BACnet on TCP port 47808. On average, for each protocol, 14 requests were sent to different dark IP addresses. It is apparent that this campaign is specifically searching for online CPS.

Concerning campaign B of Table VIII, its probing strategy is quite different from that of the first campaign. Instead of mainly targeting CPS services, this campaign probed a variety of services. In fact, the average number of services probed per unique host is 191, some of which are Modbus and BACnet. Furthermore, the campaign "recycled" 13 new hosts every week and their probes originated from random source ports. Moreover, in contrast to the first campaign, which mainly utilized TCP to probe, this second campaign leveraged more services such as UDP, NBNS, CoAP, MDNS, ISAKMP, ENIP, and QUIC. The collected information from such probes could be used by malicious entities to perform vulnerability analysis on a larger number of publicly reachable CPS services.

In contrast to the aforementioned two campaigns that were generated from unsolicited yet known sources, we now detail three inferred orchestrated campaigns that we deem as being malicious, given that their domains pointed to suspicious hosts and/or their IP addresses did not reflect any known/benign entities.

One of the largest inferred campaigns (campaign C of Table VIII) that meet such criteria originated from numerous locations in the United States and Germany. The campaign consisted of 188 distinct hosts conducting large-scale scanning in a stealthy manner. Generally, this campaign targeted each destination IP address for a maximum of 5 times. The campaign targeted Modbus 30% of the time, in addition to several other services such as CWMP, HTTP and HTTP-ALT, and HTTPS. In fact, the campaign initiated its scans against Modbus, followed by a UDP and an HTTP scan. While probing Modbus, this campaign leveraged only two source ports, 40849 and 63419, among all 188 hosts. This may serve

to indicate that such campaign is most likely running the same probing tools/techniques/malware. Due to the non-interactive nature of darknet traffic analysis, it is rather difficult to clarify the aim of this campaign's activities. However, our claim is that, given that the abused services which are tagged along with Modbus, include, CWMP, SSH and HTTP, this campaign can be dedicated to execute CPS brute force attacks. In fact, by cross-matching the campaign's IP addresses against the previously noted publicly available threat repositories, 68.7% of them were found to have been previously reported for SSH and HTTP brute force attacks.

Another unique campaign (campaign D of Table VIII) originated from various cities in China. This campaign leveraged 116 IP addresses, during two non-consecutive weeks in the analyzed one-month period. The campaign targeted Modbus and BACnet, yet also focused on ports 80 and 443. After manual inspection, we noticed that many-to-one brute force HTTP and HTTPS requests are being extensively generated through different source ports. We postulate that this campaign is targeting the Human Machine Interface (HMI) of CPS.

Last but not least, we also identified a relatively short coordinated scan, campaign E of Table VIII, that remained active for a one week period. Attributed to Russia, this campaign is dedicated towards probing the Foundation Fieldbus HSE. While the campaign only leveraged 54 IP addresses, however, it has contacted almost all (98%) of our darkspace by generating traffic from random source ports within the 30k and 50k range. Such probes could identify vulnerabilities within this protocol's implementation, such as the IP multicast features on Foundation Fieldbus systems, which are typically not well protected.

V. Discussion

Our overall proposed models, approaches and techniques leverage network telescopes to infer CPS probing activities. Thus, we now present some of the assumptions that underlie our analysis, in addition to some challenges and ways to leverage the obtained results to enhance CPS security.

Attackers' IP Address Selection: Our newly deployed darknet IP address space is still at its infancy. Thus, the proposed approach is unable to monitor and infer events that do not target our sensors. This can occur when attackers use an already published hit list or test specific and known vulnerable services. Although such methods will allow scans to avoid being detected or assessed by our approaches, adversaries in general prefer to employ up-to-date and various hit lists of services to decrease their chances of being detected and to increase their chances of launching subsequent attacks. To achieve this, at least one global scan is needed to first assess the impact of the attack; a scan that would probably hit our sensors. To this end, we obviously do not claim that we did not miss any other Internet-scale CPS probing activity, however, our collaborators and us, are not aware of any worldwide reported CPS probes that were not (at least partially) inferred by our proposed methods during the analyzed period.

Incomplete view of the CPS abuse: As briefed in Section II, our approach falls under passive network traffic monitoring. As such, since we do not interact with incoming traffic, we can only observe the first communication packets related to the CPS protocols. Consequently, our approach cannot draw a holistic view of the complete CPS abuse beyond such communication attempts. This is a typical limitation though of analyzing network telescope one-way traffic.

Defense against scanning: Evidently, by using our deployed network telescope, a significant number of CPS scanning activities reached our networks, probing for Internet services including a variety of CPS-specific services. The nature and intent of these events are at best hard to be concretely verified. Therefore, it is important to have defensive mechanisms in place to protect networks and CPS realms against potentially malicious subsequent activities. The first step in defending against such malicious traffic consists of detecting the scanning activities. For instance, after a CPS scanning campaign has been identified, CPS operators can proceed with deploying solutions to protect against it. One solution in the case of known/unsolicited campaigns is to report the scans and request exclusion of a particular network address space from subsequent scans. Legitimate campaigns often have mechanisms in place for excluding networks from their scans. In the event that exclusion requests are not possible, or are not respected by the scanning entities (i.e., in the case of malicious CPS scanning activities), incoming network traffic from source IP addresses that are repeatedly involved in such scanning activities can be dropped with the use of blacklists. Such lists can be intuitively built based on darknet analysis. Ideally, deploying darknets at different locations can provide more global and accurate blacklisting information.

Indeed, the most difficult step in defending against scans and their subsequent activities is detecting the scanning activities, since deploying solutions such as reports and blacklists are relatively straightforward. Unfortunately, as [54] uncovers, the vast majority of networks do not proactively detect scans, but rather accidentally discover them during maintenance or troubleshooting. Scanning activities leave footprints in IDSs, firewall logs, webserver logs etc., that can aid in detecting them and extracting scanners' source IP addresses for consequent blacklisting. However, going through the logs can be a tedious and error-prone process that can resolve in a large number of false positive and false negative results, which can have an unwanted effect on the network's operation. To this end, we recommend the use of a network telescope within or external to CPS environments, similar to the one deployed and reported in this paper. Incoming traffic to the network telescope's dedicated IP addresses can be automatically analyzed to detect scanning activities and pinpoint the scanners' source IP addresses. A positive side effect of this approach is that the analysis of the darknet traffic can reveal other useful patterns for CPS administrators, such as misconfiguration errors, infected CPS devices, ongoing malicious campaigns, etc. For large CPS networks, interactive CPS honeypots can be complementary deployed, which can further assist in identifying the intent and nature of the incoming traffic.

Research Trends: Recent network telescope analysis focused mostly on large-scale Internet scanning activities with the goal of detecting scanners and identifying broad patterns in their scanning behavior [54]. This aspect in network telescopes research can be attributed to the emergence of highly efficient scanning tools and techniques, which can scan the entire IPv4 address space in just a few minutes. We strongly believe

that network telescope research will shift towards specialized per-protocol analysis, relatively similar to the one presented here, in an effort to generate fine-grain cyber-intelligence. Such a shift would be in line with the evolution of the threat landscape; current threat actors, particularly Advanced Persistent Threats (APTs), have become increasingly sophisticated targeting evolving paradigms (IoT, CPS, etc.) As such, we undoubtedly believe that future attacks will include APTs dedicated towards these paradigms. In this context, it would be interesting to observe how the Internet measurement, control and cyber security research communities would collaborate to leverage their capabilities to contribute to the security of such complex systems.

VI. Concluding Remarks

In a dedicated effort to capture real unsolicited and malicious notions in the realms of CPS, this paper presented a thorough investigation of CPS probing activities towards ample of CPS protocols. The latter was achieved by examining, analyzing and correlating various dimensions of significant amount of darknet data. A novel probabilistic model was presented and employed to sanitize darknet data from misconfiguration traffic. Subsequently, inference and characterization modules were devised to extract and analyze diverse CPS probing events. To this end, trends, packets' headers and co-occurrence patterns of such events, among others, were reported. Additionally, in an effort to tackle the challenging problem of inferring CPS orchestrated probing campaigns by exclusively monitoring and analyzing a network telescope IP space, we presented a hybrid approach based on time series and netflow analysis methods. The outcome disclosed more than 9 thousand orchestrated, stealthy CPS events, originating from a plethora of unsolicited and malicious campaigns. While Section II highlighted a number of research gaps that are undoubtedly worthy of being investigated, we are currently designing and deploying diverse CPS honeypots to infer tangible CPS attack models. In this context, we will be leveraging the information obtained from this work coupled with those attack models to build tailored CPS resiliency mechanisms, from the cyber as well as the control/physical perspective, to address the security of CPS in the power and critical manufacturing sectors.

Acknowledgments

The authors would like to sincerely thank all the IT personnel at New York University in Abu Dhabi and New York for their support and aid in deploying and managing the darknet monitors. The authors are also grateful to the anonymous reviewers for their insightful comments and suggestions.

References

[1] Kyoung-Dae Kim and Panganamala R Kumar. Cyber–physical systems: A perspective at the centennial. *Proceedings of the IEEE*, 100(Special Centennial Issue):1287–1308, 2012.

[2] Eric Simmon, Kyoung-Sook Kim, Eswaran Subrahmanian, Ryong Lee, Frederic De Vaulx, Yohei Murakami, Koji Zettsu, and Ram D Sriram. A vision of cyber-physical cloud computing for smart networked systems. *NIST, Aug*, 2013.

[3] Ivan Stojmenovic. Machine-to-machine communications with in-network data aggregation, processing, and actuation for large-scale cyber-physical systems. *Internet of Things Journal, IEEE*, 1(2):122–128, 2014.

[4] Y. Zhang, M. Qiu, C. W. Tsai, M. M. Hassan, and A. Alamri. Health-cps: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, PP(99):1–8, 2015.

[5] Insup Lee, Oleg Sokolsky, Sanjian Chen, John Hatcliff, Eunkyoung Jee, BaekGyu Kim, Andrew King, Margaret Mullen-Fortino, Soojin Park, Alexander Roederer, et al. Challenges and research directions in medical cyber–physical systems. *Proceedings of the IEEE*, 100(1):75–90, 2012.

[6] Hamza Fawzi, Paulo Tabuada, and Suhas Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[7] Cyber attacks mounting fast in U.S. http://www.cbsnews.com/news/cyber-attacks-mounting-fast-in-us/.

[8] Frank Kargl, Rens W van der Heijden, Hartmut Konig, Alfonso Valdes, and Marc C Dacier. Insights on the security and dependability of industrial control systems. *IEEE Security & Privacy*, 12(6):75–78, 2014.

[9] ICS-CERT: Cyber-Attack Against Ukrainian Critical Infrastructure. https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01.

[10] The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT). https://ics-cert.us-cert.gov.

[11] Dina Hadžiosmanović, Robin Sommer, Emmanuele Zambon, and Pieter H Hartel. Through the eye of the plc: semantic security monitoring for industrial processes. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 126–135. ACM, 2014.

[12] M. Caselli, E. Zambon, and F. Kargl. Sequence-aware intrusion detection in industrial control systems. In *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*, pages 13–24. ACM, 2015.

[13] Sean Peisert, Jonathan Margulies, David M Nicol, Himanshu Khurana, and Chris Sawall. Designed-in security for cyber-physical systems. *IEEE Security & Privacy*, 12(5):9–12, 2014.

[14] Stephen McLaughlin and Patrick McDaniel. Sabot: specification-based payload generation for programmable logic controllers. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 439–449. ACM, 2012.

[15] HoneyNet Project. CONPOT ICS/SCADA Honeypot. [Online]: http://conpot.org/.

[16] 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.

[17] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):13, 2011.

[18] Henrik Sandberg, André Teixeira, and Karl H Johansson. On security indices for state estimators in power networks. In *First Workshop on Secure Control Systems (SCS), Stockholm, 2010*, 2010.

[19] Fabio Pasqualetti, Florian Dorfler, and Francesco Bullo.

Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[20] Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. Detecting integrity attacks on scada systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, 2014.

[21] Miroslav Pajic, Shreyas Sundaram, George J Pappas, and Rahul Mangharam. The wireless control network: A new approach for control over networks. *IEEE Transactions on Automatic Control*, 56(10):2305–2318, 2011.

[22] Shreyas Sundaram and Christoforos N Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, 56(7):1495–1508, 2011.

[23] Yilin Mo et al. False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5967–5972. IEEE, 2010.

[24] A. Teixeira, I. Shames, H. Sandberg, and K.H. Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135 – 148, 2015.

[25] Y. Mo, S. Weerakkody, and B. Sinopoli. Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *Control Systems, IEEE*, 35(1):93–109, Feb 2015.

[26] Sean Weerakkody and Bruno Sinopoli. Detecting integrity attacks on control systems using a moving target approach. In *54th IEEE Conference on Decision and Control (CDC)*, pages 5820–5826. IEEE, 2015.

[27] Man-Ki Yoon and Gabriela F Ciocarlie. Communication pattern monitoring: Improving the utility of anomaly detection for industrial control systems. In *NDSS Workshop on Security of Emerging Networking Technologies*, 2014.

[28] Stephen E McLaughlin, Saman A Zonouz, Devin J Pohly, and Patrick Drew McDaniel. A trusted safety verifier for process controller code. In *NDSS*, 2014.

[29] Abdulmohsen Almalawi, Xinghuo Yu, Zahir Tari, Adil Fahad, and Ibrahim Khalil. An unsupervised anomaly-based detection approach for integrity attacks on scada systems. *Computers & Security*, 46:94–110, 2014.

[30] Martin Roesch. Lightweight intrusion detection for networks. In *Proceedings of the 13th Conference on Systems Administration (LISA 99)*, pages 229–238, 2005.

[31] Quickdraw SCADA IDS. http://www.digitalbond.com/tools/quickdraw/.

[32] Carlo Bellettini and Julian L Rrushi. Vulnerability analysis of scada protocol binaries through detection of memory access taintedness. In *Information Assurance and Security Workshop (IAW'07) SMC*, pages 341–348. IEEE, 2007.

[33] Eric J Byres, Dan Hoffman, and Nate Kube. On shaky ground–a study of security vulnerabilities in control protocols. *Proc. 5th American Nuclear Society Int. Mtg. on Nuclear Plant Instrumentation, Controls, and HMI Technology*, 2006.

[34] Albert Treytl, Thilo Sauter, and Christian Schwaiger. Security measures for industrial fieldbus systems-state of the art and solutions for ip-based approaches. In *IEEE International Workshop on Factory Communication Systems*, pages 201–209. IEEE, 2004.

[35] Steven Cheung, Bruno Dutertre, Martin Fong, Ulf Lindqvist, Keith Skinner, and Alfonso Valdes. Using model-based intrusion detection for SCADA networks. In *Proceedings of the SCADA security scientific symposium*, volume 46, pages 1–12. Citeseer, 2007.

[36] Niv Goldenberg and Avishai Wool. Accurate modeling of modbus/tcp for intrusion detection in scada systems. *International Journal of Critical Infrastructure Protection*, 6(2):63–75, 2013.

[37] Sibin Mohan, Stanley Bak, Emiliano Betti, Heechul Yun, Lui Sha, and Marco Caccamo. S3a: secure system simplex architecture for enhanced security of cyber-physical systems. *arXiv preprint arXiv:1202.5722*, 2012.

[38] Helge Janicke, Andrew Nicholson, Stuart Webber, and Antonio Cau. Runtime-monitoring for industrial control systems. *Electronics*, 4(4):995–1017, 2015.

[39] Saman Zonouz, Julian Rrushi, and Steve McLaughlin. Detecting industrial control malware using automated PLC code analytics. *Security & Privacy, IEEE*, 12(6):40–47, 2014.

[40] Jan-Ole Malchow, Daniel Marzin, Johannes Klick, Robert Kovacs, and Volker Roth. Plc guard: A practical defense against attacks on cyber-physical systems. In *IEEE Conference on Communications and Network Security (CNS)*, pages 326–334. IEEE, 2015.

[41] Marco Caselli, Emmanuele Zambon, Jonathan Petit, and Frank Kargl. Modeling message sequences for intrusion detection in industrial control systems. In *Critical Infrastructure Protection IX*, pages 49–71. Springer, 2015.

[42] Stephen E McLaughlin. On dynamic malware payloads aimed at programmable logic controllers. In *HotSec*, 2011.

[43] Patrick Düssel, Christian Gehl, Pavel Laskov, Jens-Uwe Bußer, Christof Störmann, and Jan Kästner. Cyber-critical infrastructure protection using real-time payload-based anomaly detection. In *Critical Information Infrastructures Security*, pages 85–97. Springer, 2009.

[44] Wei Gao et al. On scada control system command and response injection and intrusion detection. In *eCrime Researchers Summit (eCrime), 2010*, pages 1–9. IEEE, 2010.

[45] Andreas Paul, Franka Schuster, and Hartmut König. Towards the protection of industrial control systems–conclusions of a vulnerability analysis of profinet io. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 160–176. Springer, 2013.

[46] Zhichun Li, A. Goyal, Yan 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.

[47] Yu Jin, Gyorgy Simon, Kuai Xu, Zhi-Li Zhang, and Vipin Kumar. Gray's anatomy: Dissecting scanning activities using IP gray space analysis. *Usenix SysML07*, 2007.

[48] 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*, MineNet '07, pages 7–12, New York, NY, USA, 2007. ACM.

[49] 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*, ASIACCS '09, pages 11–22, New York, NY, USA, 2009.

ACM.

[50] Darcy Benoit and André Trudel. World's first web census. *International Journal of Web Information Systems*, 3(4):378, 2007.

[51] John Heidemann, Yuri Pradkin, Ramesh Govindan, Christos Papadopoulos, Genevieve Bartlett, and Joseph Bannister. Census and survey of the visible internet. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, IMC '08, pages 169–182, New York, NY, USA, 2008. ACM.

[52] Y Pryadkin, R Lindell, J Bannister, and R Govindan. An empirical evaluation of IP address space occupancy. *USC/ISI Technical Report ISI-TR*, 598, 2004.

[53] Ang Cui and Salvatore J. Stolfo. A quantitative analysis of the insecurity of embedded network devices: results of a wide-area scan. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 97–106, New York, NY, USA, 2010. ACM.

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

[55] 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*, 2014.

[56] Internet census, 2012. http://internetcensus2012.bitbucket.org/paper.html.

[57] Steve Bellovin. There Be Dragons. In *USENIX Summer*, 1992.

[58] Steven M Bellovin. Packets found on an internet. *ACM SIGCOMM Computer Communication Review*, 23(3):26–31, 1993.

[59] 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.

[60] 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.

[61] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. The spread of the sapphire/slammer worm, 2003.

[62] 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.

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

[64] 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.

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

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

[67] Charlie Scott and Richard Carbone. Designing and implementing a honeypot for a scada network. *The SANS Institute Reading Room.*, 22:2016, 2014.

[68] Venkat Pothamsetty and Matthew Franz. Scada honeynet project: Building honeypots for industrial networks, 2008.

[69] Digital Bond. SCADA Honeynet. [Online]: http://www.digitalbond.com/tools/scada-honeynet/.

[70] Neal Krawetz. Anti-honeypot technology. *IEEE Security & Privacy*, 2(1):76–79, 2004.

[71] Eleazar Aguirre-Anaya, Gina Gallegos-Garcia, Nicolas S. Luna, and Luis A. V. Vargas. A new procedure to detect low interaction honeypots. *International Journal of Electrical and Computer Engineering*, 4(6):848–857, 12 2014.

[72] Team CYMRU. Who is looking for your SCADA infrastructure? [Online]: https://www.team-cymru.com/ReadingRoom/Whitepapers/2009/scada.pdf, 2008.

[73] Matthew Ford, Jonathan Stevens, and John Ronan. Initial results from an ipv6 darknet13. In *International Conference on Internet Surveillance and Protection*, pages 13–13. IEEE, 2006.

[74] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.

[75] Lukas Krämer, Johannes Krupp, Daisuke Makita, Tomomi Nishizoe, Takashi Koide, Katsunari Yoshioka, and Christian Rossow. Amppot: Monitoring and defending against amplification ddos attacks. In *Research in Attacks, Intrusions, and Defenses*, pages 615–636. Springer, 2015.

[76] Raimundo Real and Juan M Vargas. The probabilistic basis of jaccard's index of similarity. *Systematic biology*, pages 380–385, 1996.

[77] Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose, and Andreas Terzis. A multifaceted approach to understanding the botnet phenomenon. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 41–52. ACM, 2006.

[78] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49, 1978.

[79] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. A characterization of cybersecurity posture from network telescope data. In *Proceedings of the 6th international conference on trustworthy systems, Intrust*, volume 14, 2014.

[80] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[81] Jesse Kornblum. Identifying almost identical files using context triggered piecewise hashing. *Digital investigation*, 3:91–97, 2006.

[82] Jayanthkumar Kannan, Jaeyeon Jung, Vern Paxson, and Can Emre Koksal. Semi-automated discovery of application session structure. In *Proceedings of the 6th ACM SIGCOMM IMC*, pages 119–132. ACM, 2006.

[83] 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.