Who's Scanning Our Smart Grid? Empirical Study on Honeypot Data

Daisuke Mashima*, Yuan Li*, and Binbin Chen†*
*Advanced Digital Sciences Center
†Singapore University of Technology and Design
{daisuke.m, yuan.li}@adsc-create.edu.sg, binbin_chen@sutd.edu.sg

***Abstract*—In order to implement and fine-tune cyber defense mechanisms, it is crucial to know who are the potential enemies and what tactics they are using. In the general cyber security area, honeypot, a decoy system intended to attract cyber attackers, is considered as an effective measure to collect such threat intelligence. However, publication analysing such data is scarce, especially in industrial control systems and smart grid domain. In this paper, we discuss our findings based on the empirical study with 6-month network traces collected in low-interaction smart grid honeypot systems deployed in geographically different regions on Amazon cloud platform. In particular, we discuss actual attack patterns observed as well as insights from the data-driven study on access/attack patterns, correlations among different locations, and dynamics in access sources, some of which are considered effective when configuring security mechanisms such as firewall and intrusion detection systems.**

I. Introduction

Smart grid has a very broad attack surface (i.e., entry points for attackers) owing to the nature of physically distributed systems. Besides, with increasing integration of renewable generations and distributed energy resources, which are often maintained by different organizations, the security landscape is increasingly becoming complicated. Under such circumstances, it is no longer feasible to enforce equal level of security policies throughout the infrastructure, and a single loophole in the infrastructure (e.g., missing or inappropriate security configuration of a device connected to the public network) could eventually impact the entire power grid infrastructure. Unfortunately, such a risk is not unrealistic, and a large number of industrial control system (ICS) devices, which seem to be related to the power grid system, can be found on an online search engine like Shodan [1], which collects information of Internet-connected devices. For instance, as of April, 2019, over 1,200 devices that support IEC 60870-5-104 as well as nearly 500 devices that support DNP3 protocol, both of which are representative protocols used for remote control and monitoring of modernized power grid systems, are indexed. If we include other protocols that are used in broader ICS, such as Modbus, the number is even larger.

The crucial first step to secure our critical infrastructure is to learn who is accessing and scanning such exposed devices. in the general cybersecurity domain, a technology called honeypot has been long explored and utilized. Honeypots are, in short, dummy, or decoy, systems or devices that are intended to attract attackers, and one of the purposes is to collect intelligence about attackers. In this direction, we set up a honeypot system that imitates such Internet-exposed ICS devices, which opens network ports that are typically used in a power grid context, such as IEC 60870-5-104, IEC 61850, and DNP3 [2] among others. We deployed such honeypots in instances that belong to different geographic locations on Amazon cloud platform (AWS), and collected network traces for over 6 months (from September, 2017 to March, 2018).

While we admit that there are several ways to further improve the realism of our honeypot deployments, for instance in terms of IP addresses and imitation of characteristics of real ICS devices in the market, we have collected sizable amount of accesses that are specifically targeting such ICS-related ports. In this paper, we discuss the dataset we collected, which is made available for interested researchers[1], as well as our findings based on the data. In particular, we present real-world attack/access attempts targeting smart grid devices and difference and similarity in observed patterns among honeypot instances with different geographic locations. This paper can be seen as a case study to demonstrate what types of analysis can be made based on the captured network traffic data in order to derive actionable intelligence for better protecting our critical infrastructure.

The rest of the paper is organized as follows. We first discuss related work in Section II. Section III elaborates our honeypot deployment and configurations. Discussion on the collected network traces is made in Section IV. Finally, we conclude in Section V.

II. Related Work

There are some open-source honeypot implementations for ICS including smart grid, such as [3], [4], [5]. However, such well-known implementations can be easily fingerprinted by a system like Honeyscore by Shodan [6], and therefore may dispel attackers before they access the honeypot systems. Thus, in this study we intentionally avoided using such implementation and instead chose to implement simple, but our own, system.

While the aforementioned ICS honeypot implementations only offer cyber-side emulation, in the recent years there are some efforts made to provide cyber-physical integrated emulation [7], [8]. Such advanced ICS honeypots are considered effective to retain attackers inside for slowing down attacks as

[1] https://www.illinois.adsc.com.sg/softgrid/honeypot/

well as for collecting intelligence based on attackers' behavior after penetrating into the system. In this study, we focus on the scanning phase, and analysis with the advanced, high-fidelity ICS honeypots will be left for our future work.

We do not find much published work analyzing network traces collected on honeypot systems, especially in the ICS domain. Fachkha et al. [9] studied network traces collected on honeypots, focusing on ICS. Their focus is mainly placed on detection and study of the Internet-scale probing activities, and also does not emphasize power grid systems. Reference [10] implemented an Internet-of-Things honeypot system called IoTPOT and discussed the analysis and findings based on the collected data. Besides the difference in the application domain, their focus is Telnet-based accesses. In this paper, we particularly focus on analysis of access trends targeting network services on smart grid devices.

III. Honeypot Configurations

In this study, we set up 5 AWS instances on different geographic locations, namely Singapore, The US (Ohio), Canada, Germany, and Brazil. We set up TCP listeners on the ports listed in Table I. As can be seen in the table, we utilized simple server programs for IEC 60870-5-104 and IEC 61850, which provide responses according to the protocol. Because we did not emulate further system/device details, we claim they are categorized as low-interaction honeypot. During the study, we checked Shodan entries [1] about our honeypot instances, and confirmed that they are not flagged as honeypot, but are registered as ICS devices.

TABLE I
ICS protocols implemented on Honeypot

Protocol	Port	Description
IEC 61850 MMS (and Siemens S7)	102	Runs simple IEC 61850 MMS server
Modbus TCP	502	Runs simple TCP listener
Niagara Fox	1911 4911	Runs simple TCP listener
EtherNet/IP (ENIP)	2222 44818	Runs simple TCP listener
IEC 60870-5-104	2404	Runs simple IEC 60870-5-104 server
DNP3	19999 20000	Runs simple TCP listener
BACnet	47808	Runs simple TCP listener

Each honeypot instance runs Wireshark network protocol analyzer [11] to capture hourly network traces, which are periodically downloaded to our local server equipped with ELK stack [12] for our analysis and visualization. For the sake of description of our tool chain, the dashborad system that imports and visualizes the downloaded network trace files is shown in Figure 1. During the processing, we estimated geographic location of each source IP address by using Maxmind's GeoLite library [13].

We ran our honeypot instances for over 6 months from September, 2017 to March, 2018, and collected approximately 6GB of ICS network traces in total. In the rest of this paper, we discuss our findings based on them.

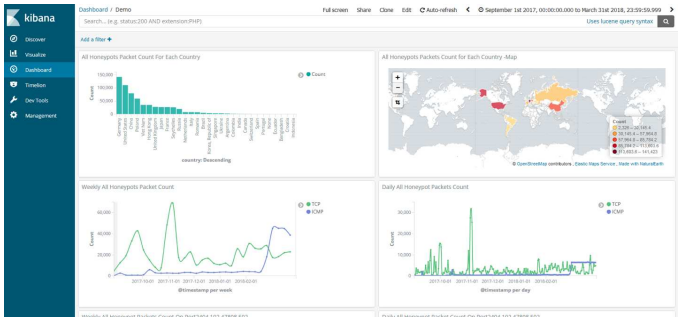

Fig. 1. Dashboard visualizing Collected Network Traces

IV. Empirical Study on Network Traces

A. Overall Trends in Access Intensity (TCP and ICMP)

Figure 2 shows the weekly packet counts throughout the data collection period, for TCP-based traffic and ICMP-based traffic. There are a few notable events to pay attention to. For instance, in terms of TCP traffic, there are 2 noticeable spikes in September and November, 2017. Based on the further inspection, the honeypot instance in Canada got a large number of access to its port 2222 (EtherNet/IP, which is one of the leading industrial protocols in the US), from a single host in the US. The second peak is associated to the event where instances in Germany (majority) and Brazil got ENIP access from hosts in Germany. Regarding ICMP traffic, we can see dramatic increase near the end of collection period. We found that a single host in Thailand started to send continuous *Ping* traffic to the Singapore instance. Although these are not sufficient to generalize, geographic proximity would be one of the factors when attackers select their targets.

B. Protocol Specific Access Attempts

As mentioned in Table I, we ran simple server application on port 102 and port 2404. On each port, we observed a number of, but similar, access attempts that are compliant with the corresponding protocols. Figure 3 shows a series of interactions captured on port 2404. As seen in the figure, the source will establish the connection, and then send interrogation request. After receiving the response from the honeypot, it closes the connection.

Because the access pattern was identical over multiple attempts observed on multiple honeypot instances, we guess they are using the same tool for collecting information. In addition, this is considered as an example showing that there are (potential) attackers who are particularly interested in smart grid devices, because IEC 104 protocol is used only in the smart grid context.

C. Observed Attacks Targeting Smart Grid Devices

During the data collection, we observed several access attempts targeting smart grid communication protocols. In this section we discuss some notable examples.

First one is the denial-of-service (DoS) attack against port 102 (i.e., IEC 61850 MMS or Siemens S7 protocol). As seen in Figure 4, the attack strategy is a traditional SYN-flooding

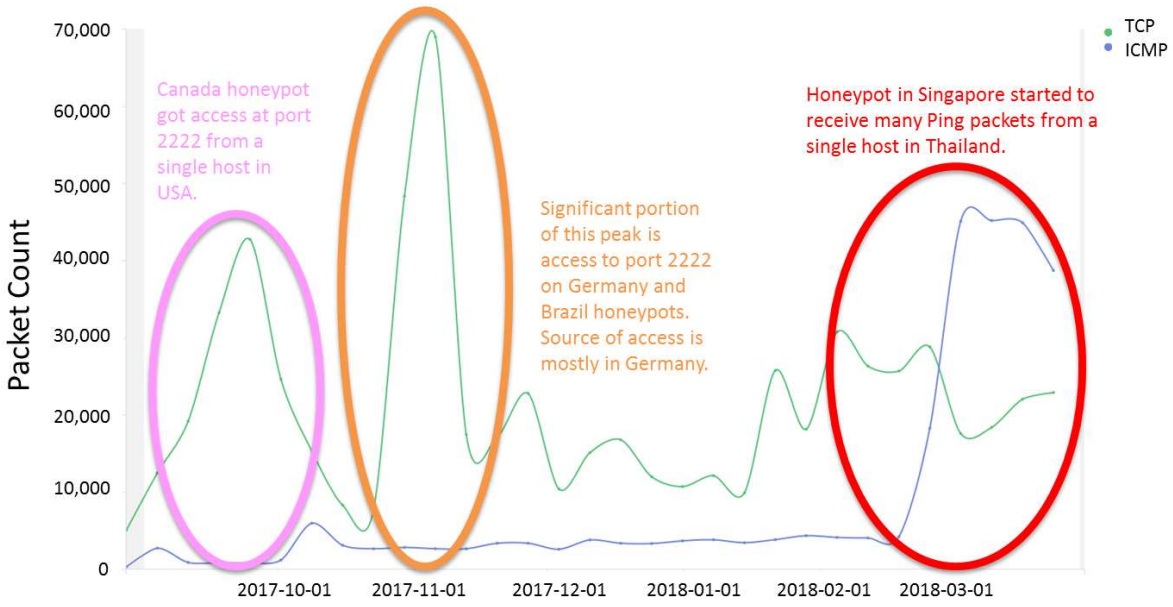

Fig. 2. Overall Trends in Traffic Intensity

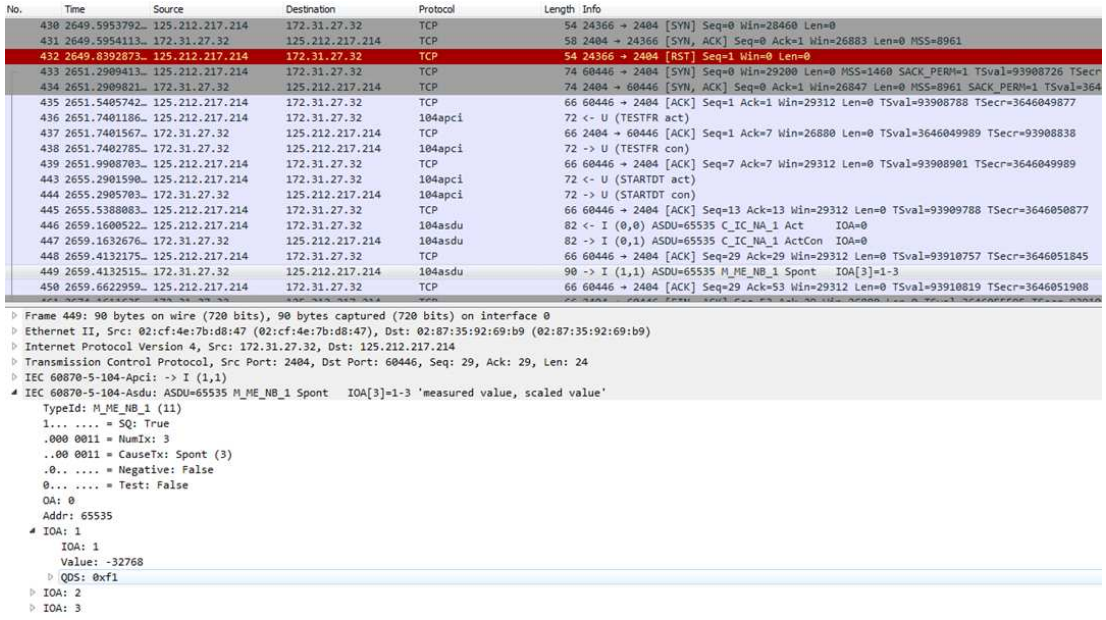

Fig. 3. Snapshot of Access Attempts for IEC 60870-5-104 Port

attack. In this attack, the source IP addresses are different but seem to belong to the same internet service provider, and therefore we assume that they are coordinated. Although the attack strategy is not novel and can be countered by firewall etc., if there is a loophole in the security configurations such DoS attack would easily affect the functionality of target ICS devices owing to its limited resources.

Another type of attack we observed is scanning against DNP3 and Modbus TCP. As seen in Figure 5, Modbus scanning is performed by sending a series of queries for exhaustive set of "Unit" identifiers. On the other hand, DNP3 scanning contains multiple query requests in a single packet, and therefore it increases in packet size (see Figure 6). Such findings can be coded into signature-based or statistics-based intrusion detection systems (e.g., those discussed in [14]) to fine-tune the attack detection for the real system.

D. Correlation among Honeypot Instances

Because we deployed honeypot instances on different geographic locations, in this section we look into similarity and difference among the patterns observed on different instances. We looked into the cross-correlation of daily and weekly packet counts.

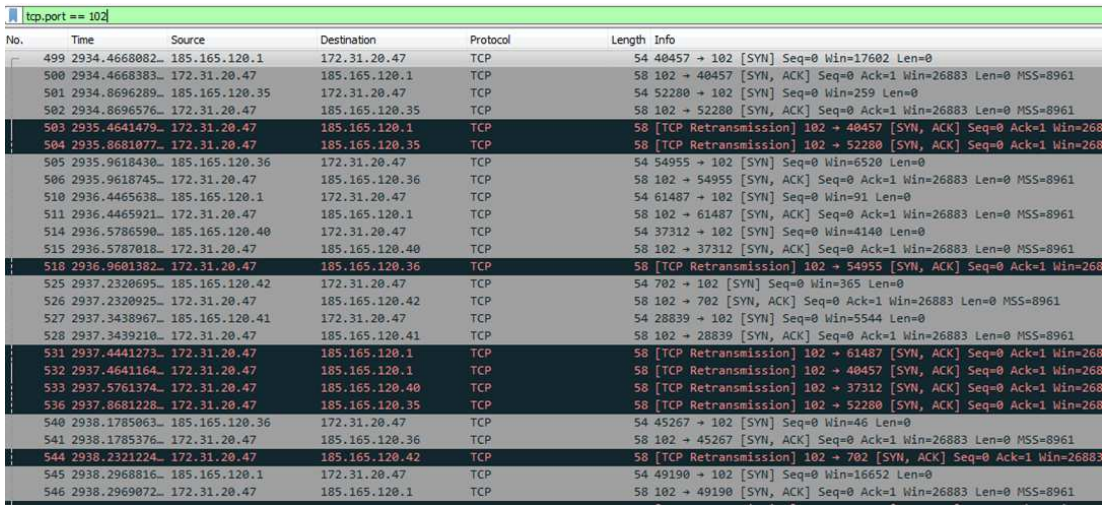

Fig. 4. Snapshot of DoS Attack (SYN-flooding) against Port 102

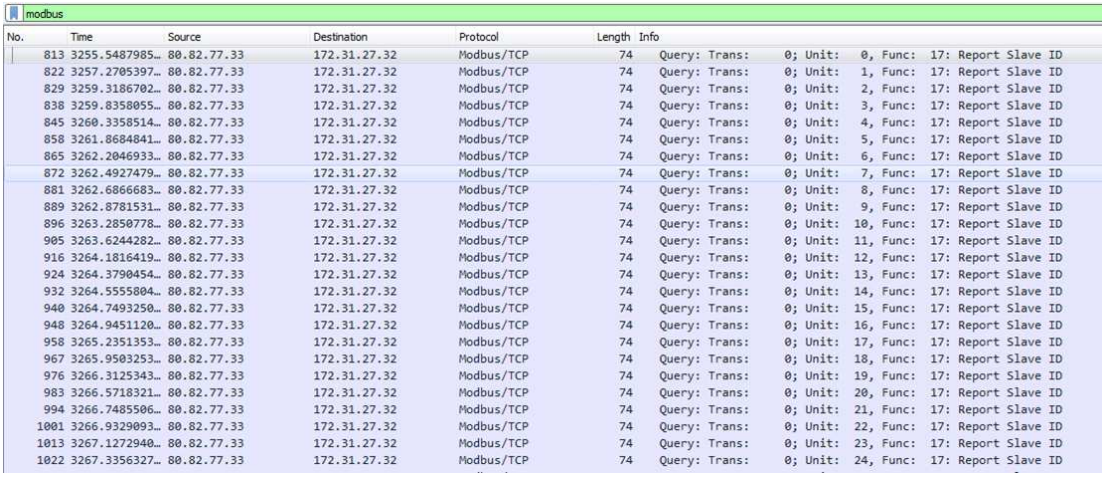

Fig. 5. Snapshot of Scanning against Modbus TCP

TABLE II
CROSS-CORRELATION OF TRAFFIC INTENSITY (DAILY)

	HP 1	HP 2	HP 3	HP 4	HP 5
HP 1	1.000	0.008	0.579	0.050	0.016
HP 2	0.008	1.000	-0.021	-0.042	**0.968**
HP 3	0.579	-0.021	1.000	-0.050	-0.019
HP 4	0.050	-0.042	-0.050	1.000	-0.055
HP 5	0.016	**0.968**	-0.019	-0.055	1.000

We found that, while the cross-correlation in trends among honeypot instances is overall not high, the Germany instance and Brazil instance showed very high correlation in both daily and weekly time-series data, 0.97 and 0.98 respectively (see HP2 and HP5 in Table II). As can be seen in the daily plots in Figure 7, the patterns, e.g., the position of peaks, are very similar. Based on our further investigation in source addresses, we found that majority of access sources are shared between the two, and are originated from server/web hosting service providers in Germany, Brazil, Russia, and Netherlands. This finding implies that the same set of attackers are probing smart grid devices in multiple different locations.

Besides, we also identified correlation with "lag". For instance, weekly access pattern observed by instance in the Canada and one in Brazil (and therefore also Germany) showed similar trends (e.g., in terms of position of significant peaks) with 1-week lag (see Figure 8). This information can be used as an advance warning to prepare for upcoming attack/probing campaign.

On the other hand, in both weekly and daily access intensity, we did not observe significant auto-correlation. Thus, it is considered that there is no periodicity in the access intensity.

E. Dynamics in Source IP Addresses

Lastly let us discuss the difference/similarity of observed source IP addresses over time. Table III summarizes the number of months in which each IP address is observed. As seen in the table, while majority of IP addresses only appear in 1 or 2 months, there are also IP addresses consistently observed throughout the period. Among the 54 IP addresses

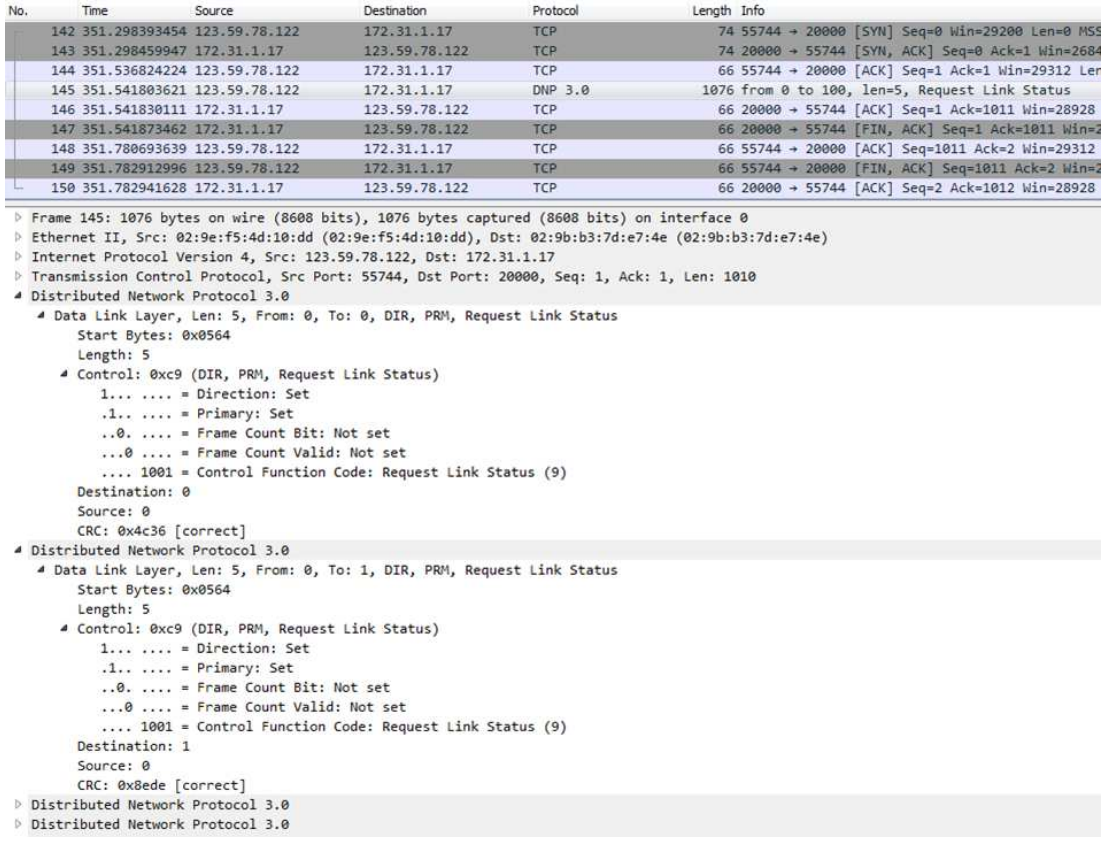

Fig. 6. Snapshot of Scanning against DNP3

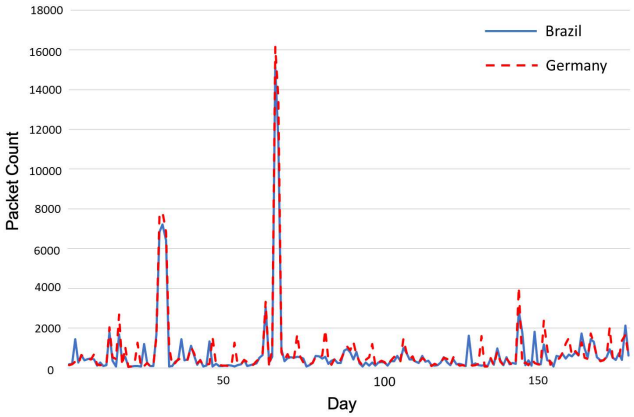

Fig. 7. Correlation in Packet Counts (Germany and Brazil)

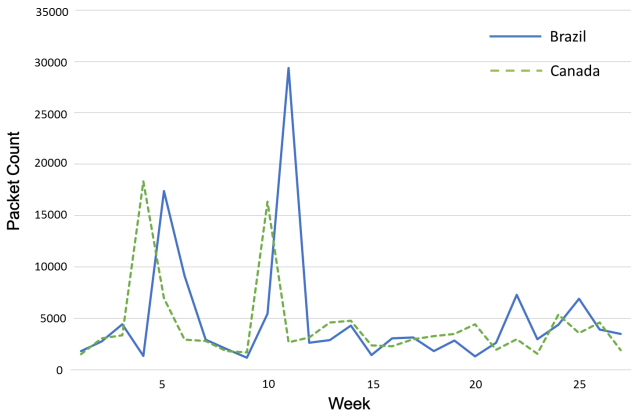

Fig. 8. Correlation in Weekly Packet Counts with Lag (Brazil and Canada)

observed in all months, 12 are from the US, followed by 10 from Australia.

TABLE III
DYNAMICS OF SOURCE IP ADDRESSES OVER MONTHS

# of Months	1	2	3	4	5	6
# of IP Addresses	6,909	601	151	99	51	54

We further looked into the appearance of IP addresses in each month. The box plot in Figure 9 shows the distribution of appearance frequency of each IP address (in terms of the number of days) for each month. From the figure, in all months, we can see that a few IP addresses are observed almost every day. We also found that all of these IP addresses belong to a cloud/hosting service provider in Japan and were attempting access to port 102 (IEC 61850 MMS or Siemens S7) and port 47808 (BACnet).

We also studied difference/similarity of source IP addresses among honeypot instances. Figure 10 shows, for each month, the fraction of source IP addresses that are observed over varying number of honeypot instances. As seen in the figure,

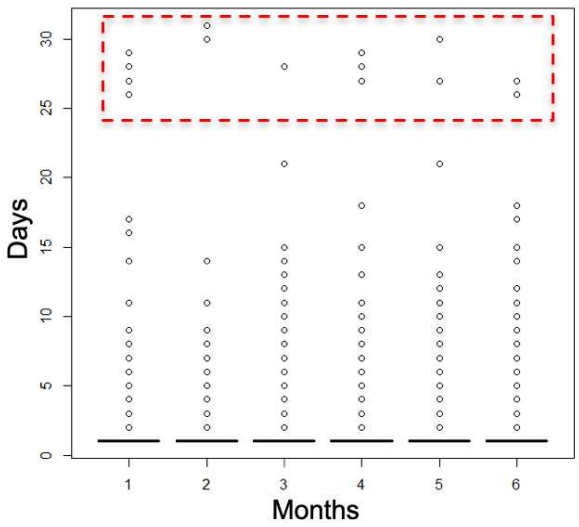

Fig. 9. Dynamics of IP Addresses in Each Month. Frequently observed IP addresses, highlighted with the red, dashed box, belong to a cloud/hosting service provider in Japan.

while majority of the IP addresses are observed only by one instance, some addresses are accessing all instances. Because such persistent IP addresses are likely to access other Internet-connected ICS devices, they should be carefully monitored or perhaps filtered by firewall in advance. The daily observation in September, 2017 is shown in Figure 11.

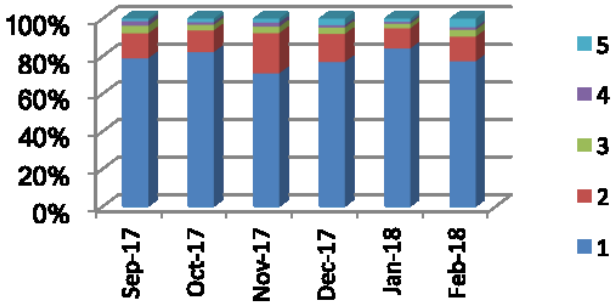

Fig. 10. Dynamics of IP Addresses among Honeypot Instances (Monthly)

V. Conclusions

In this paper, we discussed some empirical findings based on smart-grid specific network traces captured on our honeypot system. Based on the real-world network traces collected for over 6 months, our findings include: real-world attack/access attempts targeting smart grid devices, and difference and similarity in observed patterns (e.g., cross-correlation and source IP addresses) among honeypot instances with different geographic locations. Although our intention is not to derive general claims, some of the findings are considered promising to fine-tune security measures, such as firewall and intrusion detection systems. The collected network traces are available for interested researchers so that they can investigate from

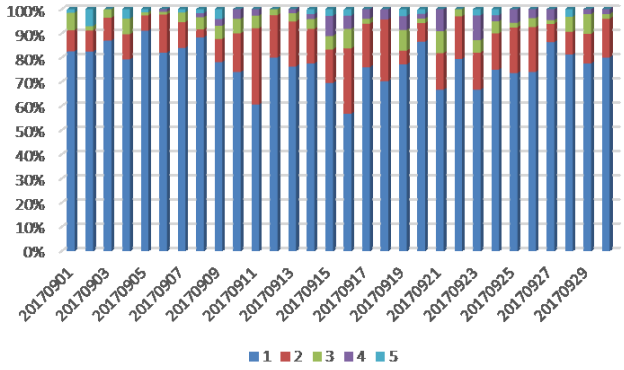

Fig. 11. Dynamics of IP Addresses among Honeypot Instances (Daily)

perspectives different from ours. In the future work, besides investigating the data from different angles, we plan to collect data using high-fidelity, smart grid honeypot systems, such as an enhanced version of [8], and analyze the data for deriving advanced threat intelligence, including attackers' movements after penetrating into the infrastructure.

Acknowledgment

This research is supported in part by the National Research Foundation, Prime Minister's Office, Singapore under the Energy Programme and administrated by the Energy Market Authority (EP Award No. NRF2017EWT-EP003-047) and is partly supported by the National Research Foundation, Prime Ministers Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] "Shodan," https://www.shodan.io/.
[2] IEC TC57, "IEC 61850-90-2 TR: Communication networks and systems for power utility automation part 90-2: Using iec 61850 for the communication between substations and control centres," *International Electro technical Commission Std*, 2015.
[3] "CONPOT ICS/SCADA honeypot," https://www.conpot.org.
[4] "Developments of the honeyd virtual honeypot," http://www.honeyd.org/.
[5] "Digital bond," http://www.digitalbond.com/tools/scada-honeynet.
[6] "Honeypot or not?" https://honeyscore.shodan.io/.
[7] D. Antonioli, A. Agrawal, and N. O. Tippenhauer, "Towards high-interaction virtual ics honeypots-in-a-box," in *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy*. ACM, 2016, pp. 13–22.
[8] D. Mashima, B. Chen, P. Gunathilaka, and E. L. Tjiong, "Towards a grid-wide, high-fidelity electrical substation honeynet," in *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2017, pp. 89–95.
[9] C. Fachkha, E. Bou-Harb, A. Keliris, N. D. Memon, and M. Ahamad, "Internet-scale probing of cps: Inference, characterization and orchestration analysis." in *NDSS*, 2017.
[10] Y. M. P. Pa, S. Suzuki, K. Yoshioka, T. Matsumoto, T. Kasama, and C. Rossow, "Iotpot: A novel honeypot for revealing current iot threats," *Journal of Information Processing*, vol. 24, no. 3, pp. 522–533, 2016.
[11] "Wireshark," https://www.wireshark.org/.
[12] "ELK Stack," https://www.elastic.co/elk-stack.
[13] "GeoLite2 Free Downloadable Databases," https://dev.maxmind.com/geoip/geoip2/geolite2/.
[14] H. C. Tan, C. Cheh, B. Chen, and D. Mashima, "Tabulating cybersecurity solutions for substations: Towards pragmatic design and planning," in *Proceedings of IEEE PES ISGT Asia 2019*. IEEE, 2019.