



S7commTrace: A High Interactive Honeypot for Industrial Control System Based on S7 Protocol

Feng Xiao^{1(✉)}, Enhong Chen¹, and Qiang Xu²

¹ Anhui Province Key Laboratory of Big Data Analysis and Application,
School of Computer Science and Technology,
University of Science and Technology of China, Hefei, China

xiaof686@mail.ustc.edu.cn, cheneh@ustc.edu.cn

² Electronic Engineering Institute of Hefei, Hefei, China
yfnm126@126.com

Abstract. Intensively happened cyber-attacks against industrial control system pose a serious threat to the critical national infrastructure. It is significant to capture the detection and the attacking data for industrial control system by means of honeypot technology, as it provides the ability of situation awareness to reveal potential attackers and their motivations before a fatal attack happens. We develop a high interactive honeypot for industrial control system-S7commTrace, based on Siemens' S7 protocol. S7commTrace supports more function codes and sub-function codes in protocol simulation, and improves the depth of interaction with the attacker to induce more high-level attacks effectively. A series of comparative experiments is carried out between S7commTrace and Conpot, by deploying these two kinds of honeypots under the same circumstance in four countries. Data captured by these two kinds of honeypots is analyzed respectively in four dimensions, which are query results in Shodan, count of data and valid data, coverage of function code and diversity of source IP address. Experiment results show that S7commTrace has better performance over Conpot.

Keywords: Industrial control system · Honeypot · S7 · Conpot

1 Introduction

With the development of “Industry 4.0” in the world, more and more industrial control systems access to the Internet, which improves the production efficiency greatly. At the same time, the security threats in cyberspace begin to penetrate into industrial control systems. Stuxnet worm was disclosed in June 2010 for the first time, which is the first worm attacking the energy infrastructure [1, 2]. In 2014 the hackers attacked a steel plant in Germany through manipulating and destroying the control system, so that the blast furnace could not be closed properly, resulting in a huge loss [3]. On December 23, 2015, the Ukrainian power

network suffered a hacker attack, which was the first successful attack to the power grid, resulting in hundreds of thousands users suffering power blackout for hours [4]. On June 12, 2017, the security vendor ESET disclosed an industrial control network attack weapons named as win32/Industroyer, which implemented malicious attacks on power substation system. It can directly control the circuit breaker to power the substation off [5].

Industrial control systems are highly interconnected and interdependent with the critical national infrastructure [6]. Once a cyberspace security incident occurs in the industrial control systems, it has a significant impact on the country's political and economic and other aspects. Therefore, different from the traditional security strategy in Internet, security incidents in industrial control system should not be deal with after its occurrence. As we all know, every cyberspace attack was preceded by a probe to host(s) or network [7]. So it is critical for the industrial control system to build the ability of situation awareness by capturing the detection and attacking data passively and to reveal potential attackers and their motivations before a fatal attack happens.

Based on Siemens' S7 communications protocol, we develop S7commTrace which is a kind of high interactive honeypot for industrial control system. Furthermore, we deploy S7commTrace and Conpot under the same circumstance in four countries. According to the comparative experiments on the captured data by these two kinds of honeypots in the following 20 days and the searching results in Shodan after 30 days later, S7commTrace shows better performance than Conpot.

2 Related Work

Honeypot is a kind of security resource that is used to attract the attacker for illegal application without any business utility [8]. Honeypot technology is a method to set some of the hosts, network services or information as a bait, to induce attackers, so that the behavior of attack can be captured and analyzed [9]. Honeypots can be used to better understand the landscape of where these attacks are originating [10]. Venkat Pothamsetty and Matthew Franz of the Cisco Critical Infrastructure Assurance Group (CIAG) released the first PLC honeypot in 2004. They used Honeyd architecture to simulate the Modbus industrial control protocol [11]. Rist et al. [12] released Conpot, which was a open source low interactive honeypot for industrial control system in May 2013. Conpot stopped updating in November 2015. Although it supports up to seven kinds of protocols (S7, Modbus, BACnet, HTTP, Kamstrup, SNMP, IPMI), all of them are low interactive and only support a small number of function codes.

Serbanescu et al. [13] analyzed the attractiveness of the industrial equipment exposed in the public network to the attacker and the behavior of the attacker by setting low interactive ICS honeypot on a large scale. Jicha et al. [10] deployed 12 Conpot SCADA Honeypot on AWS to evaluate the attractiveness and behavior in detail, by analyzing the scan results of NMAP and SHODAN. Buza et al. [14] divided the honeypot into three categories according

to the complexity: low interaction, high interaction and hybrid. They summarized the development of related project about honeypot for industrial control system since 2004, including: CIAG SCADA HoneyNet Project, Honeyd, Digital Bond SCADA HoneyNet, Conpot Project. After analyzing the advantages and disadvantages of these projects, they designed and developed the CrysPLC honeypot (CryPLH), which supported Http, Https, SNMP, and Step 7.

Search engine for Internet-accessed devices, which is different from the traditional content search engine, probes the Internet network equipment information, stores the results in the database, and provides web and API query interface. Commonly used search engines for Internet- accessed devices are Shodan [15] Censys [16,17] and ZoomEye [18]. Shodan uses the industrial control protocol directly to crawl the industrial control equipment on the Internet, and visualized the location and other information of them [19]. It is not only convenient for network security practitioners, but also facilitates the attacker to locate victims. Furthermore it may expose the existence of honeypots. Bodenheim et al. [20] deployed four Allen-Bradley ControlLogix 1756-L61 PLCs on the Internet to check Shodan's capabilities and found that four PLCs were all indexed by Shodan within 19 days. Subsequently, he proposed a solution to reduce the risk of exposure in Shodan by transforming the web service banner.

In summary, previous studies mainly focused on low interactive honeypots for industrial control system. Conpot is one of the most famous and advanced honeypot in recent years. Although it supports various industrial control protocols, Conpot is easy to be recognized as honeypot by cyberspace search engine for its characteristic of low interaction. CryPLH tries to improve the performance on interaction, but it still lacks of the capability of support real industrial control protocols. As we know, a deliberate and fatal network attack always starts with detections to target. If there is no response to the initial requests, attackers will abort their further operations. Therefore, it is quite necessary to develop a kind of high interactive honeypot based on industrial control protocols to capture detection and attacking data with good quality, while reducing the risk of being marked by cyberspace search engine.

3 Honeypot Based on S7 Protocol

S7 communications protocol is a Siemens proprietary protocol [21] running on programmable logic controllers (PLCs) of Siemens S7-300, 400, 1200 and 1500 series. It is suitable for either Ethernet, PROFIBUS or MPI networks. Because the objects of this study are those industrial control systems being accessed to the Internet, we only discuss the TCP-based S7 communications protocol in Ethernet networks. As shown in Fig. 1, S7 communications protocol packets are packed by COTP protocol, and then packed by TPKT protocol package for TCP connection. As shown in Fig. 2, the communication procedure of S7 protocol is divided into three stages. The first stage is to establish COTP connection, the second stage is to S7 communication setup, and the third stage is to exchange the request and the response for function code.

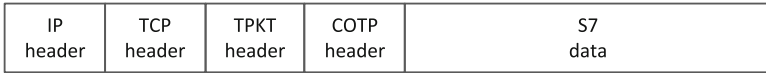


Fig. 1. Header format of S7 communication packet

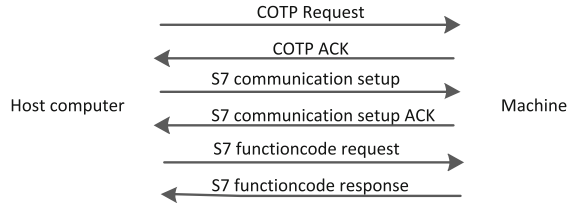


Fig. 2. Communication procedure of S7 protocol

3.1 Function Codes

The Magic flag of the S7 communications protocol is fixed to 0x32, and the following fields are S7 type, data unit ref, parameters length, data length, result info, parameters and data. In parameters field, the first byte stands for the function code of S7. Table 1 shows the optional function codes of S7. Communication Setup code is used to build a S7 connection, Read code helps the host computer to read data from PLC, Write code helps the host computer to write data to PLC. As for the codes of Request Download, Download Block, Download End, Download Start, Upload and Upload End, they are designed for downloading or uploading operations of blocks. PLC Control code covers the operations of Hot Run and Cool Run, while PLC Stop is used to turn off the device. When the function code is 0x00, it stands for system function which is used to check system settings or status. And the details are described by the 4-bits function group code and 1-byte subfunction code in the parameters field. System Functions further are divided into 7 groups, as shown in Table 2. Block function is used to read the block, and Time Function is used to check or set the device clock.

Table 1. System functions codes of S7

Code	Functions	Code	Functions	Code	Functions
0x00	System functions	0x1b	Download block	0x1f	Upload end
0x04	Read	0x1c	Download end	0x28	PLC control
0x05	Write	0x1d	Download start	0x29	PLC stop
0x1a	Request download	0x1e	Upload	0xf0	Communication setup

Table 2. System function group and corresponding subfunction

Function group code	Function	Subfunction code	Subfunction
1	Programmer commands	1	Request diag data
		2	VarTab
2	Cyclic data	1	Memory
3	Block function	1	List blocks
		2	List blocks of type
		3	Get block info
4	CPU function	1	Read SZL
		2	Message service
5	Security	1	PLC password
6	PBC BSEND/BRECV	None	None
7	Time function	1	Read clock
		2,3	Set clock
		4	Read clock (following)

3.2 S7commTrace

S7commTrace can be divided into four modules, including TCP communication module, S7 communications protocol simulation module, data storage module and user template, as shown in Fig. 3. TCP communication module is responsible for listening on TCP port 102, submitting the received data to the Protocol Simulation module, and replying to the remote peer. S7 communications protocol simulation module parses the received data according to the protocol format and obtains the valid contents at first. And then S7 communications protocol Simulation module generates the reply data referring to user template. At last, the reply data are submitted back to TCP communication module to be packaged. User template records all the user-defined information such as PLC serial number, manufacturer, and so on. The data storage module handles the request and response of data storage.

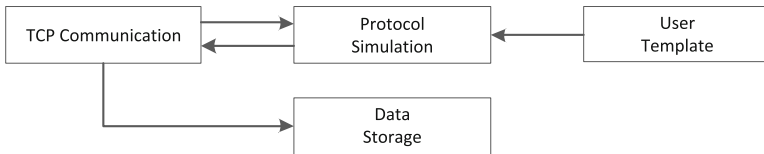


Fig. 3. Modules of S7commTrace

Cyber attacks against the S7 device are implemented based on some specific function codes, such as uploading, stopping, etc. And as we known, an

experienced attacker usually tries to check the system status list (Read SZL) or do other operations before the execution of those significant function codes. Therefore, in order to record the attacker's communication data completely and accurately, a sophisticated honeypot should have as more responses to S7 function codes as possible to induce the attacker's further operations. After setting up a S7 communication, Conpot only support the subfunction code of Read SZL and reply a fixed value of SZL ID and index. As for other function codes, Conpot has no response to them. S7commTrace makes a great improvement over Conpot by responding to all the function codes and subfunction codes listed in Tables 1 and 2.

In order to fabricate the responding data, we request and record the real responses from a S7-300 PLC device firstly. And then by means of those real data, a user defined template is made in S7commTrace. At the same time, we customize unique settings of User Template among different S7commTrace honeypots, without changing the data format.

4 Evaluation

We deploy Conpot and S7commTrace honeypots in United States (US), German (GE), China (CN) and Singapore (SG) around the global area at the same time. The deployment utilize Aliyun (US, CN, SG) and Host1Plus (GE) as virtual host with configuration of 1.5 Ghz single core CPU, 1 GB RAM and 40 GB Disk. All the operation systems of virtual hosts are Ubuntu Server 16.04 64 bits. Every virtual host installs MySQL database to store data captured by local honeypot. Furthermore, two copies of the VPS are rented in every county to make sure that Conpot and S7commTrace are deployed under the same circumstance. The experiment lasts for 60 days.

4.1 Query Results in Shodan

According to the statistics of Bodenheimer et al. [20], if a PLC device accesses to the Internet, it will be marked by Shodan up to 19 days later. Therefore, after 30 days of deployment, we search all the honeypots in Shodan. We find that all Conpot honeypots in four countries are indexed and marked as Conpot by Shodan, as shown in Fig. 4. But only one S7commTrace honeypot in Singapore is indexed by Shodan, and marked as Conpot. Considering this IP was used for Conpot deployment before the experiment, the search results may be affected. We deploy another new S7commTrace honeypot in Japan (JP), and it isn't indexed by Shodan after 30 days, as shown in Fig. 5. Table 3 makes a detailed comparison of how the two kinds of honeypots are indexed by Shodan. As we know, Shodan only detects but not indexes the real ICS device for security consideration. In the experiment, S7commTrace was not indexed by Shodan with capturing the real detection data of Shodan. This is a clear evidence which proves that S7commTrace vividly simulates a PLC and is recognized as a physical device. Therefore, the risk of exposure by cyber space search engines is greatly reduced.

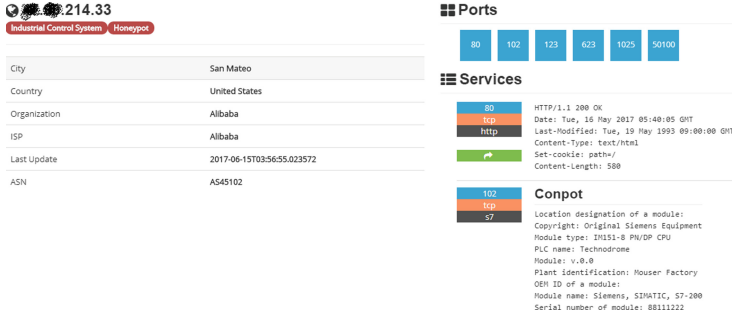


Fig. 4. Lables of Conpot in Shodan

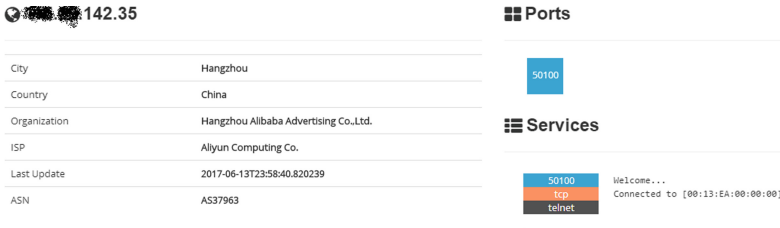


Fig. 5. Lables of Conpot in Shodan

Table 3. Conpot and S7commTrace honeypots indexed by Shodan

Locations	Conpot	S7commTrace
US	Indexed and marked as Conpot	Not being indexed
GE	Indexed and marked as Conpot	Not being indexed
CN	Indexed and marked as Conpot	Not being indexed
SG	Indexed and marked as Conpot	Indexed and marked as Conpot
JP	–	Not being indexed

4.2 Count of Data and Valid Data

In order to avoid the situation that the honeypots may not be detected at the beginning, we use the data from 31 days to 50 days after deployment. If we define an uninterrupted TCP communication connection as a session, Conpot records 17 sessions and S7commTrace records 56 sessions. Compared with Conpot honeypot, the number of session is significantly increased in every S7commTrace honeypot, as shown in Table 4. As each session contains a number of data requests, Conpot records a total of 472 data requests, while S7commTrace records a total of 1217 data requests. Compared with Conpot honeypot, the number of requests is increased to some extent in every S7commTrace honeypot, as shown in Table 5. Especially the S7commTrace honeypot in China, it records 9 times more requests than the Conpot honeypot. In fact, not all data recorded are in accordance with

the format of S7 protocol. Ignoring such data, Conpot records a total of 82 valid requests, while S7commTrace records a total of 535 valid requests. Compared with Conpot honeypot, the number of valid request is significantly increased in every S7commTrace honeypot. On the purpose of checking the quality of the data, we calculate the rate of valid request in two kinds of honeypots. The average rate of valid request in Conpot is 17.37%, while the maximum rate does not exceed 22.78%. But in S7commTrace, the average rate increases to 43.96%, while the minimum rate is not less than 36.11%. Therefore, S7commTrace not only records more requests but also records more valid request, compared with Conpot. That means data quality is great improved in S7commTrace.

Table 4. Comparison of session number between Conpot and S7commTrace

Locations	Conpot	S7commTrace	Improvement
US	3	9	200.00%
GE	6	10	66.67%
CN	3	30	900.00%
SG	5	7	40.00%
Total	17	56	229.41%

Table 5. Comparison of request number between Conpot and S7commTrace

Locations	Requests			Valid requests			Valid rate		
	Con	S7	Imp	Con	S7	Imp	Con	S7	Imp
US	126	157	24.60%	11	61	454.55%	0.0873	0.3885	345.02%
GE	141	158	12.06%	31	62	100.00%	0.2199	0.3924	78.44%
CN	79	758	859.49%	18	360	1900.00%	0.2278	0.4749	108.47%
SG	126	144	14.29%	22	52	136.36%	0.1746	0.3611	106.82%
Total	472	1217	157.84%	82	535	552.44%	0.1737	0.4396	153.08%

Note: Con, S7, and Imp is short for Conpot, S7commTrace and Improvement.

4.3 Coverage of Function Code

When we analyze the function codes used in the data captured, we find that the data of Conpot only covers the function codes of COTP Connect, Communication Setup, and Read SZL. But the List Blocks function code is included in the data of S7commTrace in addition to the above three. Furthermore, function codes are also used more frequently in the data of S7commTrace than COTP, as shown in Table 6. Read SZL is the function code to the read system status with the parameters of SZL ID and SZL Index. In the data of Conpot, there are only two kinds of parameters, like (0x0011, 0x0001) and (0x001C, 0x0001). But in the data of S7commTrace, five different kinds of parameters are found, including (0x0011, 0x0001), (0x001C, 0x0001), (0x0011, 0x0000), (0x001C, 0x0000) and (0x0131, 0x0001).

Table 6. Comparison of function codes between Conpot and S7commTrace

Functions	Conpot	S7commTrace	Improvement
COTP connect	31	142	358.06%
Communication setup	21	106	404.76%
Read SZL	30	150	400.00%
List blocks	0	10	–

4.4 Diversity of Source IP Address

Conpot records data from 14 different IP addresses, while S7commTrace records data from 43 different IP addresses, as shown in Table 7. And the total number of different IP address is 49. As shown in Table 8, 11 IP addresses appear in at least two honeypots. Meanwhile 113.225.219.220 and 113.225.210.250 are recorded only by four S7commTrace honeypots but absent in all Conpot honeypots. According to the DNS query results, 10 of the 49 IP addresses point to Shodan’s domain name with the suffix of shodan.io. 3 IP addresses point to the domain name of Electrical Engineering and Computer Science (EECS) Department of University of Michigan with the suffix of eecs.umich.edu. As we know EECS is one of the institutions which develop Censys [16, 17]. Furthermore, 2 IP addresses point to BEACONLAB’s domain name with the suffix of plcscan.org. Different with Shodan and Censys, BEACONLAB specializes in safety research and practice on industrial control systems and provides related services [22]. As for other IP address, they are resolved to be dynamic domain name or none domain name.

Table 7. Comparison of IP source between Conpot and S7commTrace

Locations	Conpot	S7commTrace	Improvement
US	4	9	125.00%
GE	7	10	42.86%
CN	4	28	600.00%
SG	6	7	16.67%
Total	14	43	207.14%

As shown in Table 9, S7commTrace records more IP addresses of Shodan and Plcscan than Conpot, while they records the same IP addresses of Censys. Therefore, compared to Conpot, S7commTrace attracts more detections from famous search engines focusing on industrial control system. Conpot records 14 IP addresses located in 6 countries and regions, while S7commTrace records 43 IP addresses located in 10 countries and regions. Figure 6 shows how these IP addresses distribute geographically. In the world wide, detections recorded by S7commTrace are more widely distributed than those by Conpot.

Table 8. IP address recorded by Conpot and S7commTrace

IP	Conpot				S7commTrace			
	US	GE	CN	SG	US	GE	CN	SG
139.162.99.243	✓	✓	✓	✓	✓	✓	✓	✓
141.212.122.145	✓	✓	✓	-	✓	✓	-	✓
113.225.219.220	-	-	-	-	✓	✓	✓	✓
113.225.210.250	-	-	-	-	✓	✓	✓	✓
80.82.77.139	✓	-	-	✓	-	-	✓	-
71.6.146.185	-	✓	✓	-	-	-	-	-
188.138.125.44	-	-	✓	-	-	-	✓	-
120.132.93.150	-	-	✓	-	-	-	✓	-
141.212.122.96	-	✓	-	-	-	✓	-	-
141.212.122.48	-	✓	-	-	-	✓	-	-
66.240.219.146	-	-	-	✓	-	-	-	✓

Table 9. Comparison of IP addresses with static suffix between Conpot and S7commTrace

Domain	Conpot	S7commTrace
shodan.io	66.240.219.146	66.240.192.138 66.240.219.146
	66.240.236.119	71.6.158.166 71.6.165.200
	71.6.146.185	80.82.77.33 80.82.77.139
	80.82.77.139	89.248.167.131 198.20.70.114
eecs.umich.edu	141.212.122.48	141.212.122.48
	141.212.122.96	141.212.122.96
	141.212.122.145	141.212.122.145
plscan.org	188.138.125.44	85.25.79.124
		188.138.125.44
		188.138.125.155

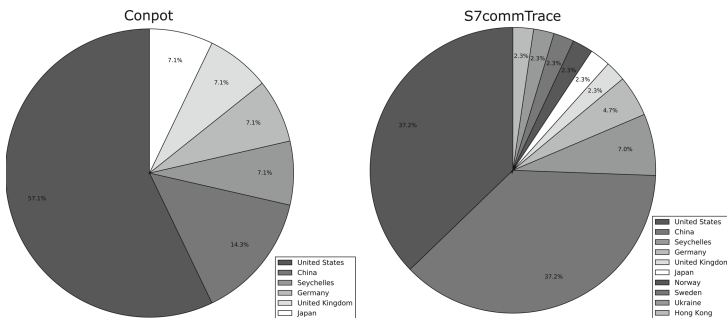


Fig. 6. Comparison of IP addresses distribution between Conpot and S7commTrace

5 Conclusions

We developed a kind of high interactive honeypot name as S7commTrace for industrial control system based on Siemens' S7 communications protocol. Through deploying Conpot and S7commTrace globally at the same time, we compared them from two dimensions: how they were indexed by cyberspace search engine, and the detection and attacking data they recorded. And thus we can draw the following in conclusion. Compared to the S7 component in Conpot, S7commTrace has the following advantages:

- S7commTrace supports more function codes and sub-function codes in protocol simulation, and improves the depth of interaction with the attacker to induce more high-level attacks effectively.
- S7commTrace has more realistic simulation of the PLC device, reduces the risk of honeypots being exposed by cyber space search engines.
- S7commTrace records more sessions and requests with higher rate of valid requests.
- S7commTrace attracts more network detections and attacks, and the recorded IP addresses are more widely distributed all around the world.

6 Future Work

S7commTrace only implements high interactive honeypot for S7 communications protocol. We will continue to develop high interactive honeypots for Modbus, BACnet, and Kamstrup which are already supported by Conpot. Furthermore we will focus on DNP3 and IEC104 which are not supported by Conpot. Another research work is the deployment of S7commTrace globally in 12 countries. Recently, S7commTrace has captured 42581 valid sessions, and we are committed to the analysis of fingerprint.

Acknowledgment. This study is supported by National Natural Science Foundation of China (U1605251).

References

1. Chen, T.M., Abu-Nimeh, S.: Lessons from Stuxnet. *Comput.* **44**(4), 91–93 (2011)
2. Kushner, D.: The real story of stuxnet. *IEEE Spectrum* **50**(3), 48–53 (2013)
3. Zetter, K.: A cyberattack has caused confirmed physical damage for the second time ever. <http://www.wired.com//2015//01//german-steel-mill-hack-destruction>. Accessed 8 July 2017
4. Zetter, K.: Inside the cunning, unprecedented hack of Ukraine's power grid. <https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/>. Accessed 8 July 2017
5. <https://www.eset.com/us/about/newsroom/press-releases/eset-discovers-dangerous-malware-designed-to-disrupt-industrial-control-systems/>. Accessed 8 July 2017

6. Stouffer, K., et al.: Guide to industrial control systems (ICS) security. NIST special publication vol. 800, no. 82, p. 16 (2011)
7. Hink, R.C.B., Goseva-Popstojanova, K.: Characterization of cyberattacks aimed at integrated industrial control and enterprise systems: a case study. In: IEEE International Symposium on High Assurance Systems Engineering, pp. 149–156 (2016)
8. Spitzner, L.: Honeybots: Tracking Hackers. Addison-Wesley Longman Publishing Co. Inc., Boston (2002)
9. Zhuge, J.-W., et al.: Honeybot technology research and application. Ruanjian Xuebao/J. Softw. **24**(4), 825–842 (2013)
10. Jicha, A., et al.: SCADA honeypots: an in-depth analysis of Conpot. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI)
11. Pothamsetty, V., Franz, M.: SCADA Honeynet Project: Building Honeybots for Industrial Networks. SCADA Honeynet Project, 15 July 2005
12. CONPOT ICS/SCADA Honeybot. <http://conpot.org/>. Accessed 16 July 2017
13. Serbanescu, A.V., et al.: ICS threat analysis using a large-scale honeynet. In: Proceedings of the 3rd International Symposium for ICS & SCADA Cyber Security Research. British Computer Society (2015)
14. Buza, D.I., Juhász, F., Miru, G., Félegyházi, M., Holczer, T.: CryPLH: protecting smart energy systems from targeted attacks with a PLC honeypot. In: Cuellar, J. (ed.) SmartGridSec 2014. LNCS, vol. 8448, pp. 181–192. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10329-7_12
15. Shodan. <https://www.shodan.io/>. Accessed 15 July 2017
16. Censys. <https://censys.io/>. Accessed 15 July 2017
17. Durumeric, Z., et al.: A search engine backed by internet-wide scanning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM (2015)
18. Zoomeye. <https://www.zoomeye.org/>. Accessed 16 July 2017
19. Ics-radar. <https://ics-radar.shodan.io/>. Accessed 15 July 2017
20. Bodenheim, R., et al.: Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. Int. J. Crit. Infrastruct. Protect. **7**(2), 114–123 (2014)
21. <https://wiki.wireshark.org/S7comm>. Accessed 15 July 2017
22. <http://plscan.org/blog/>. Accessed 16 July 2017