Identifying infected energy systems in the wild

ABSTRACT

The 2016 Mirai outbreak established an entirely new mindset in the history of large-scale Internet attacks. A plethora of Mirai-like variants have emerged in the last two years that are capable to infiltrate any type of device. In this paper we provide a 7-month retrospective analysis of Internet-connected energy systems that are infected by Mirai-like malware variants. By utilizing network measurements from several Internet vantage points, we demonstrate that a number of energy systems on a global scale were infected during the period of our observation. While past works have studied vulnerabilities and patching practises of ICS and energy systems, little information has been available on actual exploits of such vulnerabilities. Hence, we provide evidence that energy systems relying on ICS networks are often compromised by vulnerabilities in non-ICS devices (routers, servers and IoT devices) which provide foothold for lateral network attacks. Our work offers a first look in compromised energy systems by malware infections, and offers insights on the lack of proper security practices for systems that are increasingly dependent on internet services and more recently the IoT. In addition, we indicate that such systems were infected for relatively large periods, thus potentially remaining undetected by their corresponding organizational units.

1 INTRODUCTION

The rapid convergence of the IoT with ICS systems is undoubtedly a core functional block in modern energy systems. We witness a radical shift towards the administration of power infrastructures, such as wind-turbines, PV panels and residential UPS batteries, that are controlled remotely through IP-supported ICS protocols. Traditionally, ICS systems run by protocols originally designed to operate in networks with no public connectivity. Built-in security mechanisms within such protocols tend to be minimal or even nonexistent [4]. Regardless of this worrying fact, such systems nowadays are configured to enable public connectivity. Moreover, the ICS ecosystem now includes multiple Internet-facing elements, such as web and file server, IP cameras, and other IoT devices used to orchestrate critical energy systems.

The ubiquity of IoT devices has also transformed the way in which Internet-wide volumetric attacks are instrumented. The most prominent example of this new trend is the Mirai botnet [1]. Ever since the first reported incident in 2016, followed by the malware's source code being leaked, there is a growing number of Mirai-like variants targeting a plethora of vulnerable deployments. Hence, modern energy systems controlled by recently converged ICS systems are potential targets of such botnets. Potentially infected energy system may participate in coordinated large-scale DDoS attacks over a critical infrastructure or even act maliciously at a local level causing a resilience havoc. Therefore, we consider this problem as a new pragmatic challenge. We argue that this challenge needs to be addressed under a data-driven and holistic approach in order to aid the composition of future automated defense mechanisms.

As a step towards a systematic approach, in this paper we demonstrate some work in progress related to the identification and profiling of energy systems that are currently infected by Mirai-alike variants. We employ a measurements-based methodology in which we utilise Internet-wide measurements gathered by a number of different network vantage points. Our 7-month retrospective analysis considers active and omni-directional scans as well as Telnet/SSH honeypots that track Mirai-alike botnets. Earlier pieces of work (e.g., [4, 9, 11] focused on identifying vulnerabilities and on patching practises of ICS and energy systems. On the contrary, our work is the first to provide an insight related to existing compromised energy systems at a global scale. We hope that this work will bring this pragmatic challenge in to the light and will consequently alarm security experts, organisations, as well as end-users that design and manage such systems.

2 METHODOLOGY

The discovery and analysis of infected energy systems across the Internet requires the synthesis of multiple diverse network measurements. Hence, we utilise a variety of measurement sources summarized in Table 1. In this section, we discuss our data sources and the role they play in our analysis.

2.1 Detection of Malicious Hosts

The first step in our study is to compile an extensive dataset of Internet hosts that participate in malicious network activity. To this end, we combine three different sources of malicious traffic: a distributed honeypot for Mirai-like botnets, a network telescope for unsolicited traffic, and blacklists with IP addresses of spamming botnets.

2.1.1 Telnet/SSH Honeypots. Since the Mirai 2016 outbreak little has been done to assess on whether Internet-connected devices that operate on protocols explicit to the control and management of energy networks are likely to be infected a with any Mirai-like variant. More importantly, no study has been done to assess on whether any infected devices by such variants were actually patched by the organisation handling them. To assess compromised energy systems by such botnets we have deployed a low-interaction distributed honeypot. We operate 11 SSH and Telnet honeypots located in three countries: United States (3 in Las Vegas, Nevada, 1 in Minden, Nevada, 3 in Los Angeles, California), Russia (2 in Moscow), and Brazil (2 in Sao Paulo). Each honeypot is configured to log all incoming traffic, and logfiles are then aggregated and indexed using Splunk for analysis. From there we match Mirai-like fingerprint by comparing the TCP sequence with the IP addresses. In particular, Mirai bots send TCP SYN packets with the TCP initial sequence number equal to the destination IP of the targeted host [1]. Given that the TCP sequence number is a 32-bit integer, the likelihood of an identified Mirai-like fingerprint being set at random is only $\frac{1}{2^{32}}$. Based on this technique, our honeypots have detected 511,636 Mirai-like probes between 2017/02/17 – 2019/01/23.

Data Source	Collection Site	Collection Period	Data Volume
Active scanning	Censys	01/05/2018 -30/11/2018	147 full IPv4 scans, 6 protocols
Omni-directional scanning	Grey Noise	01/11/2017 -30/11/2018	330K log entries
Telnet/SSH honeypot	BadPackets LLC	02/19/2017 - 23/01/2019	511K IP addresses
Telnet/SSH honeypot	Grey Noise	01/01/2018-30/11/2018	260K IP addresses
IP Blacklists	UCEProtect,Spamhouse	01/02/2018-26/12/2018	347K IP addresses
Total Infected IPs			**684K**

Table 1: Data sources: network vantage points analysed to identify infected energy system devices with Mirai-like variants.

2.1.2 Unsolicited Internet Traffic Telescope. We augment the identified infected IPs by Mirai-like variants using GreyNoise, a telescope for unsolicited Internet traffic [5] . Typically botnets and malware discover potential vulnerable attack targets by conducting large-scale opportunistic scans of randomly selected IP prefixes. However, large scale scans of the IP address space are also conducted for research and survey purposes by systems such as Censys [2] and Shodan [3], while network misconfigurations may cause similar effects. This incessant non-productive traffic termed *Internet Background Radiation* may mask malicious traffic sources when studying isolated traffic snapshots [8]. However, traffic analysis across large time scales and across multiple vantage points allows the extraction of unique patterns and characteristics for malicious scanners. GreyNoise enables such an analysis through a network of constantly shifting servers in hundreds of data centers across the Internet. On a daily basis, GreyNoise aggregates approximately 2M iptable events, 1M SSH logins, 10M telnet login attempts as well as 100K HTTP requests. The aggregation is performed over GreyNoise's $50 - 100$ cloud servers in varying geographical regions. In parallel an internal tagging process is employed using passive operating system (p0f) logs for each observed record in which a scan might be considered benign or potentially malicious depending on its origin, its frequency and its visibility across the GreyNoise network. We extract bulk measurements by a centralised GreyNoise server between November 2017 and November 2018 and as we show in section 3 we utilise them to identify which IPs originate malicious scans and on which ports.

2.1.3 Real-time Blackhole Lists (RBL). Network operators use Real-time Blackhole Lists (RBLs) to track and block IP addresses that participate in malicious activities, such as spamming, click-fraud, and distribution of malware. RBLs are constructed either through *spamtraps*, unused email addresses that receive spam emails from bots that search and scrape the web for spamming targets, or by aggregating logs from firewalls and intrusion detection systems from multiple networks. We aggregate blacklisted IPs from two major RBLs, UCEProtect [7] and Spamhaus [10]. We use only the Level-1 from the UCEProtect project, which is the most conservative level, to minimize false positives. Overall our list contains 347,551 IPs between February and December 2018.

2.2 Active Scanning

After we compile the set of malicious IP addresses, we need to narrow down our results on ICS devices. it is necessary to first gather Internet-wide scans over particular ICS protocols. To fingerprint the ports and services active at the identified , we utilise network scans provided by Censys [2] performed over the full IPv4 address space in the period between May 2018 and December 2018.

Protocol	Banners	Devices
Modbus	540,329	104,187
Fox	463,478	43,167
BACnet	286,021	23,765
DNP3	8,536	1,648
Siemens S7	36,728	8,704
Telnet	74,100,017	17,001,702
Total	75,435,109	17,183,173

Table 2: Responsive banners and devices to active scans performed via Censys between May and November 2018.

Our study considers Censys scans over 5 distinct protocols used in modern energy systems, namely; Modbus, Siemens S7, DNP3, BACnet and Tridium Fox. All 5 protocols enable intra- and inter-network communication between (sub)stations, and mainly focus on aspects of control, distribution and automation.

To facilitate the management of the devices that rely on the above protocols and enable remote access, system administrators often enable the Telnet protocol on the TCP ports 23 and 2323. Therefore, our analysis also considers full IPv4 Censys scans over the Telnet protocol in those TCP ports.

To reduce the high dimensionality of our datasets we only consider network scans of responsive devices running in the aforementioned protocols. Thus, our Censys queries over Google's BigQuery filter out devices that block scans or contain empty response banners where no useful information is present. Table 2 provides a per-protocol view of the number of banners as well as unique devices that were responsive to the active scans performed by Censys. ModBus and Fox were the two protocols that had the biggest numbers in terms of open banners as well as devices, while DNP3 was the protocol with the smallest number of open banners and responsive devices. In contrast with any of the of the energy systems-related protocols, we witness that Telnet, either on TCP port 23 or 2323, had a much higher number of responsive devices that have also provided 74M open banners with meaningful information. We identify energy systems that operate telnet-based services by correlating devices running on any of the conventional energy-related protocols with the responsive Telnet devices.

3 RESULTS

At a first stage, we investigate devices that operate on any of the conventional energy system protocols. We subsequently study infected systems that run conventional protocols with web services.

3.1 Tracking infected energy systems

The correlation of our Censys and GreyNoise scans with the two honeypot datasets flagged a total number of $1,604$ IP addresses

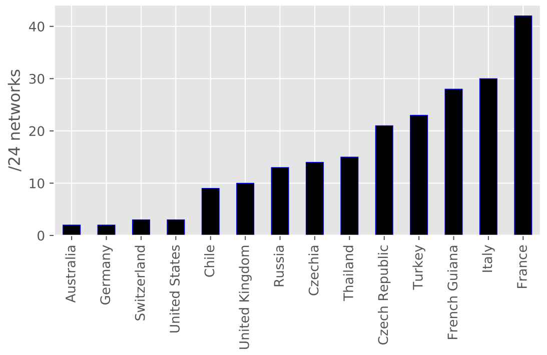


Figure 1: Global distribution - The number of /24 networks per country that contained Telnet-supported energy systems infected by Mirai-like malware strains.

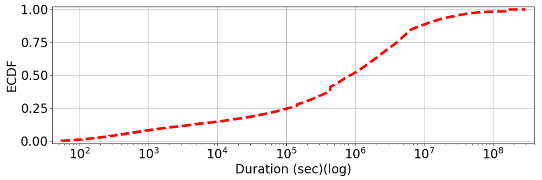


Figure 2: Difference in seconds between the first and last time an IP was captured by our honeypots.

globally that have been reported within Mirai-like botnet activities and run ICS protocols. We find that a 64.2% corresponding to 1, 304 out of all 1, 604 infected IP addresses were running the ModBus protocol. In addition, 338 infected IPs running the Tridium/Fox protocol were ranked second reaching 21.2% whereas 227 infected devices operating the BACnet protocol contributed to a 14.3%. The smallest percentage was observed by devices employing the Siemens/S7 protocol (i.e., 0.3%, 5 devices). Figure 2 shows the time length during which an IP appears to be infected based on the activity in our honeypots. More than 50% of the IPs are active for at least 11 days, indicating that the infection went undetected.

Interestingly, we did not spot any devices operating over the DNP3 protocol and this fact could potentially convey the message that systems operating under that protocol are less prone to Mirai-like variants. To understand where most of the infections are concentrated at a global scale we correlated geolocation data with each identified IP address. Table 3 depicts the top ten countries with the highest number of infected devices. As shown, the Republic of Korea reached the top place with 443 infected devices, where all of them were running the ModBus protocol. The second place is occupied by devices residing within Autonomous Systems (ASes) in the US. In contrast with most of the other countries where the majority identified devices were running ModBus, the majority of US-based infected networks was operating the Tridium/Fox protocol (48.7%). In addition, 66.4% of devices located in Canada were running on BACnet and the remaining 33.6% on Tridium/Fox, while none was operating over ModBus.

From a general viewpoint, we observe that the global distribution of infected devices was disproportional since the top two countries account for 55% of all the identified infected devices. We argue that this observation could act as an insight on questioning the cybersecurity policies employed on energy systems in both the Republic of Korea and the US.

Country	Infected devices	ModBus	Tridium /Fox	BACnet	Siemens S7
Republic of Korea	443	100%	0%	0%	0%
USA	437	23.3%	48.7%	28%	0%
Canada	110	0%	33.6%	66.4%	0%
Turkey	90	77.7%	22.3%	0%	0%
France	80	62.5%	0%	37.5%	0%
Italy	75	84%	10.6%	0%	5.4%
Thailand	53	100%	0%	0%	0%
Romania	42	100%	0%	0%	0%
China	38	52.6%	47.4%	0%	0%
Russia	35	100%	0%	0%	0%

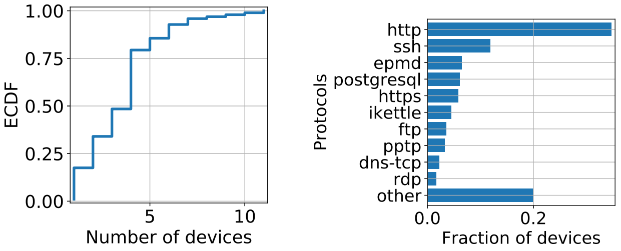
Table 3: Geographic distribution - We compare the top ten countries that harbored the most infections with respect to the 4 conventional energy system protocols.

Subsequently, we correlate Telnet scans (Table 1) with the rest of Censys scans to identify /24 network prefixes that contain devices operating both conventional energy system protocols that are also supported by Telnet-based services. Apart from typical banner metadata properties such AS number, IP address and network prefix, the correlation procedure considers geolocation coordinates in order to validate the pragmatic physical distance between the extracted set of devices. The post-processed dataset was subsequently correlated with the GreyNoise and the honeypot datasets in order to identify if any of the Telnet-supported energy systems is likely to contain infected devices.

On a global scale, we extracted a total of 215 /24 networks flagged to contain Telnet-supported infected energy systems within our observational 7-month time period. As shown in Fig. 1 the 215 /24 networks are distributed within 14 different countries. By comparing the resulted global distribution with the one obtained for conventional systems with no Telnet support (Table 3), it is evident that the rankings have changed significantly. For instance, the Republic of Korea is not present at all, meaning that most infections were within the same /24 and udner the administration of a single organization. Moreover, we witness countries that didn't appear earlier, such as the French Guiana with 28 /24 networks in which infected Telnet-supported ICS/energy systems are present.

3.2 Device Composition and Vulnerabilities

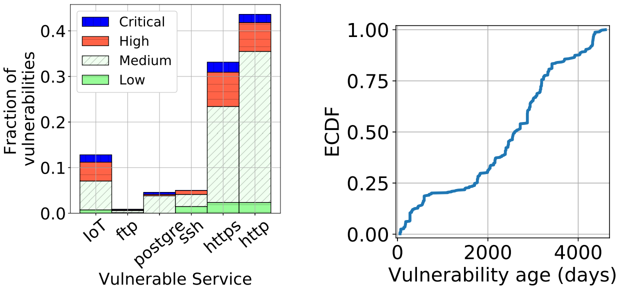
So far we have identified that each of the infected IPs has at least one ICS device, but in many cases the same IP may have multiple devices indicating the presence of Network Address Translation (NAT). To distinguish different devices behind the same IP, Censys relies on device information embedded in service banners. Figure 3a shows the number of identified devices per IP. We can see that only 20% of the IPs include only one device (which will be the identified ICS device), while the rest of the infected IPs contain a mix of two devices or more. For each of the devices behind the same NAT as the identified ICS devices we collect the service information to understand the composition of the corresponding ICS networks.



(a)

(b)

Figure 3: (a) Number of devices per infected IP – Over 30% of the ICS run behind a NAT with multple publicy exposed devices. (b) Fraction of devices per service - We observe 37 different services behind the infected IP addresses most of which enable some form of remote access.



(a)

(b)

Figure 4: (a) Fraction of disclosed vulnerabilities per service. (b) Distribution of number of days passed since vulnerability disclosure

Figure 3b plots the distribution of non-ICS devices per protocol. In total we observe 37 different protocols, with HTTP/HTTPs being the most popular, typically used to serve web-based interfaces for remote administration. Surprisingly, only 34% of the web based interfaces use HTTPS, while the rest transmit the login forms unencrypted. The banners received from these interfaces provide us further information to identify the managed ICS devices. For instance, the IP `58.64.7.60` has devices with both the modbus and HTTP protocols. By querying the HTTP server we can parse the HTML of the front-end interface which includes the banner: `<title>WAGO Ethernet Web-Based Management</title>`. Therefore, we can map the device to the actual manufacturer and product.

The presence of multiple services behind an infected IP means that any of the exposed devices may be the source of the infection, not necessarily the ICS device. We attempt to determine which of the devices may have been compromised by searching for documented security vulnerabilities for each of the identified services, following the approach in [11]. Through the corresponding banners we extract the service version which we use to search the National Vulnerability Database (NVD) [6] for Common Vulnerability Exposures (CVE). For the documented vulnerabilities we extract the publication date and the CVSS (Common Vulnerability Scoring System) value that indicates the severity of the vulnerability. Overall we discovered 832 unpatched vulnerabilities for the services shown Figure 4a. The IoT tag groups vulnerabilities in IoT products. The vulnerabilities are broken down based on the CVSS score, according to about 20% of the vulnerabilities have high or critical severity. Alarmingly, 75% of the vulnerabilities are older than 5 years as shown in Figure 4b. Our results provide strong evidence that the ICS networks related to particular energy systems have been compromised by vulnerabilities in non-ICS protocols. Therefore, without a more holistic approach in the security of energy systems, network components that are considered less critical create channels for lateral network attacks that can go undetected.

4 RELATED WORK

Project SHINE (SHodan INtelligence Extraction) [9] attempted to measure the ICS devices connected to the public Internet by extracting metadata from SHODAN, a search engine for embedded devices. By using a 927 keywords that were manually determined to belong to ICS products and vendor names, SHINE identified at over 500K Internet-accessible manufacturing devices between April 2012 and January 2014. The authors determined that many of the discovered devices were accidentally connected to the Internet due to misconfigurations. Mirian et al. [4] conducted a survey of the IPv4 address space to estimate the number of ICS devices exposed on the public Internet, by implementing five of the most popular ICS protocols in Zmap They discover more than 60k between December 2015 and March 2016, 69% of which are Modbus bridges. The authors then combined data from a network telescope and ICS honeypots to characterize the networks that search for vulnerable devices. While the majority of traffic originates from academic institutions and security firms, bulletproof hosting providers are also among the regular scanners. Wang et al [11] combined IP scan data from Shodan over a period of three years with public databases on vulnerability disclosures to conduct a longitudinal analysis of ICS patching practices. Their findings reveal that the majority of exposed ICS devices are not patched even 60 days after the disclosure of vulnerabilities, while the exploitability of vulnerabilities are only weakly correlated with patching deferral times.

5 CONCLUSION

In contrast to previous studies that examined potential ICS vulnerabilities, this work provides the first study assessing actual infections over energy systems. Using a combination of several Internet-wide network vantage points we identify energy systems infected by Mirai-like malware strains. We also demonstrate that security in such systems can be compromised by any type of web service and not necessarily the utilised ICS protocol. Moreover, we indicate poor patching practices on a global scale and flag patterns over particular countries that indicate the need to improve their practices on energy systems security. Alongside poor patching, we also pinpoint the weakness of intrusion detection systems to detect a range of infections since we show that infections on such systems persist over a great amount of time. Through this work-in-progress we emphasize that focus on composing holistic security mechanisms for such systems should be on all components and not explicitly on the mechanics of ICS protocols. We recognise that our results represent a subset of the reality due to limited Internet-wide visibility. Therefore our on-going work targets to significantly increase the number of network vantage points in order to enrich our Internet-wide view on infected devices operating energy systems.

REFERENCES

[1] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai Botnet. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 1093–1110. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

[2] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *22nd ACM Conference on Computer and Communications Security*.

[3] John Matherly. 2015. Complete Guide to Shodan. *Shodan, LLC (2016-02-25)* (2015).

[4] Ariana Mirian, Zane Ma, David Adrian, Matthew Tischer, Thasphon Chuenchujit, Tim Yardley, Robin Berthier, Joshua Mason, Zakir Durumeric, J Alex Halderman, et al. 2016. An Internet-wide view of ICS devices. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*. IEEE, 96–103.

[5] Andrew Morris. 2018. The Background Noise of the Internet. ShmooCon 2018.

[6] National Institute of Standards and Technology. 2019. National Vulnerability Database. https://nvd.nist.gov/

[7] UCEProtect Orga. 2019. The UCEPROTECT-Network Project. http://www.uceprotect.net

[8] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. 2004. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. ACM, 27–40.

[9] B Radvanovsky and J Brodsky. 2014. *Project SHINE (SHodan INtelligence Extraction)*. Technical Report. Tech. rep.

[10] Spamhaus. 2019. The Spamhaus Project. https://www.spamhaus.org/

[11] Brandon Wang, Xiaoye Li, Leandro P de Aguiar, Daniel S Menasche, and Zubair Shafiq. 2017. Characterizing and Modeling Patching Practices of Industrial Control Systems. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1 (2017), 18.