Shodan Visualized

Vincent J Ercolani
University of Arizona
Department of Management
Information Systems
Tucson, AZ 85721, USA
vercolani@email.arizona.edu

Mark W Patton
University of Arizona
Department of Management
Information Systems
Tucson, AZ 85721, USA
mpatton@email.arizona.edu

Hsinchun Chen
University of Arizona
Department of Management
Information Systems
Tucson, AZ 85721, USA
hsinchun@email.arizona.edu

[1]***Abstract*— The purpose of this paper is to discuss how using Gephi to visualize the open ports at IP addresses in Shodan may provide a means of identifying SCADA devices. Visualizations were created using both IP addresses and open ports as nodes. Modularity, centralities, and layout were used to enhance the visualizations. From these visualizations we hope to gather a better understanding of what devices are on the network.**

I. Introduction and Background

Over the last few years there has been growing interest in understanding what is running on the internet. The growth of internet-enabled devices for personal, business, and industrial use has increased interest in the security of the devices themselves as well as the data. Shodan is an internet portal that has been scanning the internet to find open ports on IP addresses and try to determine what services are running on said ports. We propose using the data from Shodan to try to determine the type(s) of devices running at each IP address. We focus on finding Supervisory Control and Data Acquisition (SCADA) and Industrial Control System (ICS) devices due to their critical role in infrastructure support. Further, by looking at the services, ports, and headers returned from Shodan, we would like to be able to deduce who owns the devices, where they are from, and through additional work, determine if the devices are vulnerable to attack.

Previous work on the Shodan database has shown that devices are indexed within 3 weeks of coming online [2]. Other researchers have looked at the lifetime patterns across industries [3]. No one, however, has visualized the patterns of ports open on a device and tried to identify what the device is. In this study we have used network patterns to try to understand a device's purpose.

Early Allen-Bradley SCADA devices using the PLC-5 Communication protocol communicate on port 2222, whereas more modern SCADA devices using the ControlLogix-based EtherNet/IP communications communicate using TCP and UDP on port 44818. Shodan has a plc5 module for scanning port 2222 and these results are in our downloaded Shodan datasets.

II. Methodology

Daily downloads of the Shodan scans have been parsed and inserted into a MySQL Database. The dataset is derived from the set of all scans from the 1st through the 14th of September 2015. We have extracted all IP addresses that are identified as being in Iran and have at least one of the common SCADA ports open in Shodan. We then created a node file of IP addresses in the dataset and all ports open on those IPs. An edge file was created of all the connections between IP addresses and open ports. Ports were labeled SCADA and non-SCADA and included the Shodan Module used to communicate with the port.

This dataset was imported into Gephi to create visualizations of the data. Modularity, average degree, network diameter, and density calculations were run on the directed graph that was created. We then ran a Force Atlas on the graph to get the visuals seen in this paper.

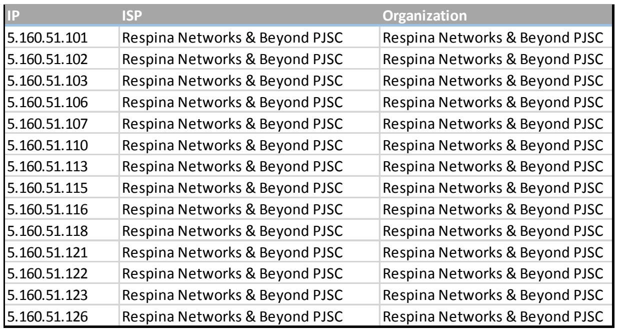
III. Main Results

The first coloration of the data (Figure 1) shows the IP addresses, ports and SCADA ports that are in the network. The large green dot on the right is port 2222 using the plc5 Shodan module that has a large cluster around it. Other ports on the left are shown to be the only port that was found on an IP addresses. We also note a significant number of IP addresses with port 7457 open, which is used for the CPE WAN Management Protocol (CWMP), a protocol typically used for automated configuration of non-SCADA devices connected to the internet including modems, routers, gateways, set-top boxes, and VoIP-phones.

When we color the graph by the Modularity of the network, as seen in Figure 2, we can see three main clusters. These clusters are:

- Olive: PLC devices with an https interface on port 443
- Teal: PLC devices with an http interface on port 80
- Red: PLC ports that are open that we suspect may be honey pots or external IP addresses for many NAT'd devices

TABLE I. Clustered IPs

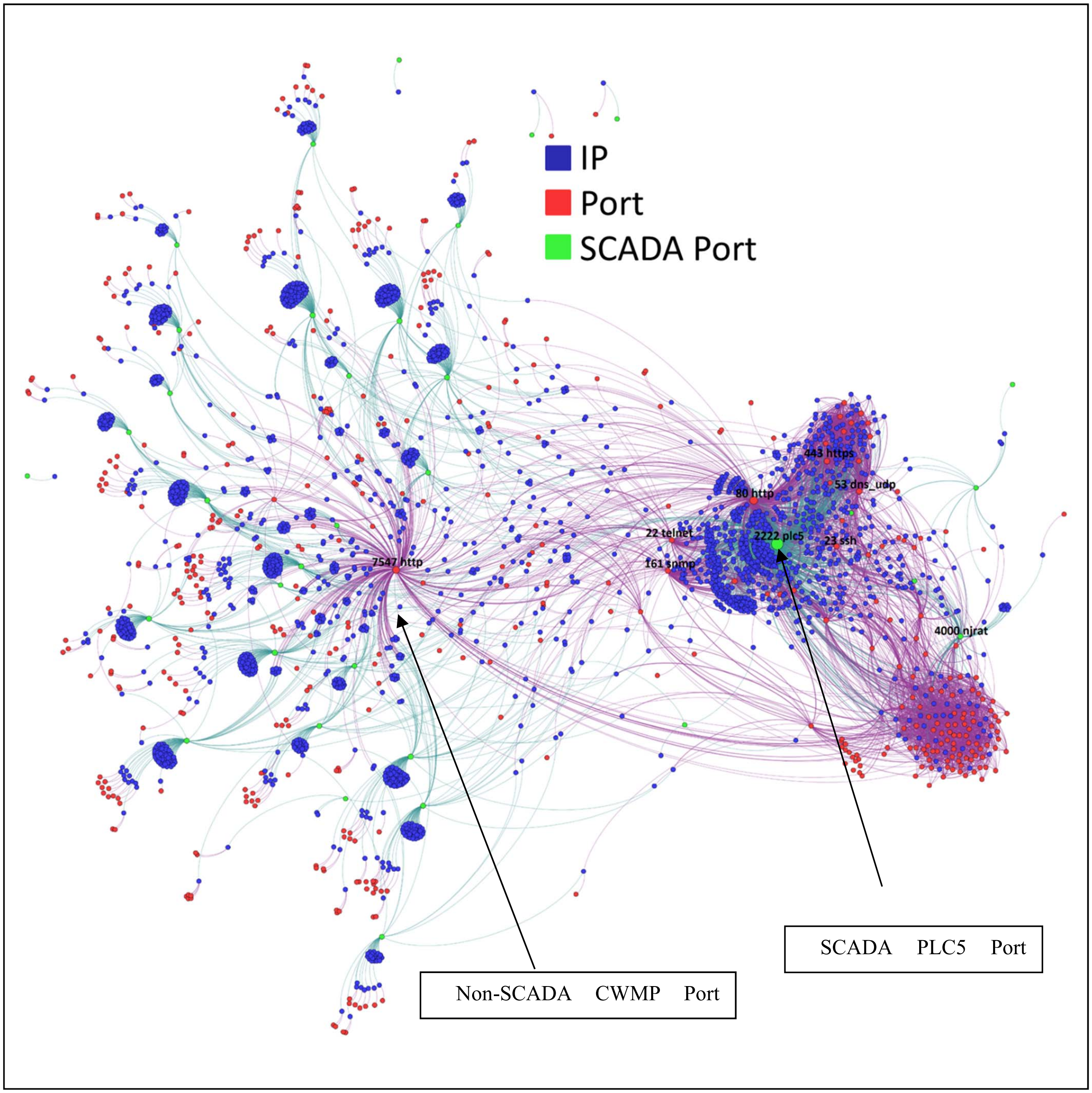


IP	ISP	Organization
5.160.51.101	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.102	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.103	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.106	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.107	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.110	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.113	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.115	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.116	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.118	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.121	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.122	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.123	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC
5.160.51.126	Respina Networks & Beyond PJSC	Respina Networks & Beyond PJSC

This material is based upon work supported by the U.S. National Science Foundation under Grant No. DUE-1303362 and SES-1314631.

978-1-5090-3865-7/16/$31.00 ©2016 IEEE

Fig. 1. IP, Port, SCADA Port Network



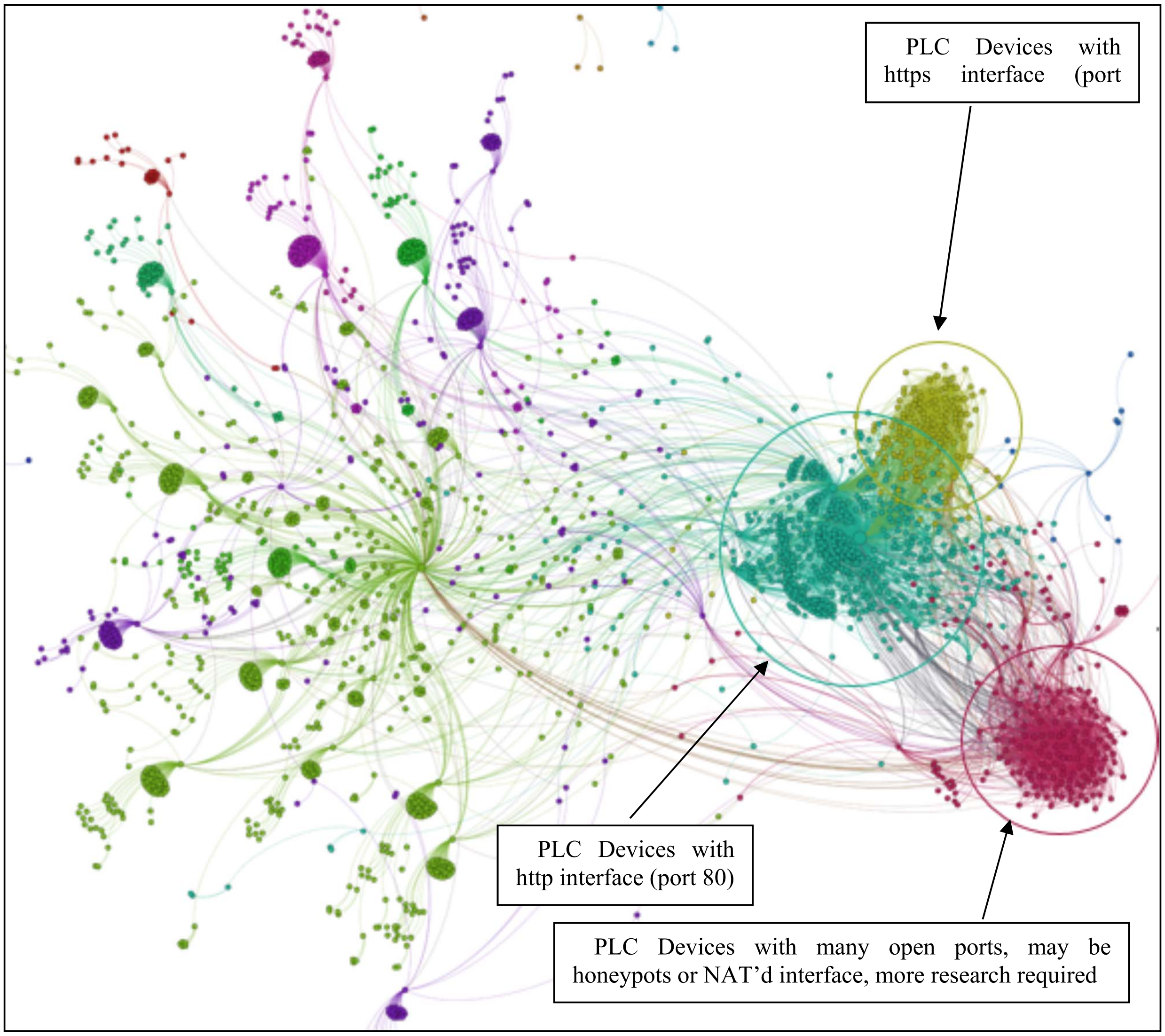
Non-SCADA CWMP Port

SCADA PLC5 Port

Further explorations need to be carried out with the IP addresses found in each of these clusters. The smaller clusters also need to be investigated further to discover their commonalities. There are many SCADA ports with only a few attached IP addresses; these need to be associated with individual manufacturers, and future research should also correlate with organization, to determine if organizations are standardizing by manufacturer and protocols which would imply standard port sets and enhance identification.

One of the most interesting findings is that the majority of the IP addresses in the red clusters (Fig. 2) belong to a single ISP in a very small range, as can be seen in Table I. Each of these IPs have many ports open using many different Shodan modules to communicate.

Fig. 2. Modularity of Network


PLC Devices with https interface (port
PLC Devices with http interface (port 80)
PLC Devices with many open ports, may be honeypots or NAT'd interface, more research required

IV. Conclusion

Initial exploration into the Shodan database has shown interesting information that we would like to look into more deeply. There is a clear differentiation between PLC devices using the secure https port 443 versus those using the insecure http port 80 which needs to be understood, as it may allow us to segregate organizations or PLC devices. Going forward, Shodan scans require verification, PLC devices should be assessed for vulnerabilities, and more analytics on other commonalities of PLC devices in each node needs to be categorized.

References

[1] Auffret, P., "SinFP, unification of active and passive operating system fingerprinting," *Journal in Computer Virology*, vol. 6, no. 3, pp. 197-205, 2010.

[2] Bodenheim, R., Butts, J., Dunlap, S., & Mullins, B. (2014). Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. *International Journal of Critical Infrastructure Protection*, 7(2), 114–123. http://doi.org/10.1016/j.ijcip.2014.03.001

[3] Genge, B., & Enăchescu, C. (2015). Non-Intrusive Historical Assessment of Internet-Facing Services in the Internet of Things, 25–36.

[4] Rasmeet S. Bali, Neeraj Kumar, Secure clustering for efficient data dissemination in vehicular cyber–physical systems, in *Future Generation Computer Systems*, Volume 56, March 2016, Pages 476-492, ISSN 0167-739X, http://dx.doi.org/10.1016/j.future.2015.09.004.