Improving ICS security through Honeynets and Machine Learning techniques

Obieda Ananbeh (oananbeh@oakland.edu)
Oakland University
Rund Alomari
Oakland University
Austin Daniell
Oakland University

Research Article

Keywords: Industrial control system; Cyber security; Honeypot, internet of things(IoT), Cyber-Physical Systems (CPS), Machine Learning (ML)

Posted Date: February 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1333285/v1

Improving ICS security through Honeynets and Machine Learning techniques

Obieda Ananbeh
Department of Computer Science and Engineering
Oakland University
Rochester, MI 48309
oananbeh@oakland.edu

Rund Alomari
Department of Computer Science and Engineering
Oakland University
Rochester, MI 48309
rundalomari@oakland.edu

Austin Daniell
Department of Computer Science and Engineering
Oakland University
Rochester, MI 48309
adaniell@oakland.edu

Abstract **The internet of things(IoT), the Industrial Internet of Things (IIoT), and Cyber-Physical Systems (CPS) can be seen everywhere, Home applications, Buildings, Cars, Space Industry, Military, Health Care, and in many other fields. On the other hand, they become an easier target for attackers, due to many reasons including the limitation of hardware, so from that point, companies start working to build a secure systems by keep themselves updated about their system threats and vulnerabilities, and also by studying how the attackers can gets into their system, how they act, what is the attack flow, and also the identity of the attackers by trapping and tricking them into believing that they have got access to the actual system or assets . And that's what it's called a Honeypot. [1] As the technology keeps changing and becomes more powerful, so do the attackers, and for that reason companies should use new techniques to enhance Honeypots efficacy by making it undetectable by cybercriminals, more usable and make use of the information that the honeypots gather in a more efficient way. Moreover, Machine Learning (ML) techniques are able to provide intelligence to IoT, IIoT, and ICS systems and networks, and enhance its ability to deal with various security problems, hence, in this research, we are developing a new solution that improves the architecture of SCADA (An ICS System) by adding CamouflageNet Honeynet into it and ML techniques, in order to defend and acquisition system security performance.**

Keywords: Industrial control system; Cyber security; Honeypot, *internet of things(IoT), Cyber-Physical Systems (CPS), Machine Learning (ML)*

I. INTRODUCTION ABOUT IOT,IIOT AND ICS AND THE SECURITY ISSUES

Internet of things (IoT) is a concept describing billions of connected devices or objects that can sense, transmit, process the data from the surrounding environment. Industrial-Internet of Things (IIoT) is an application of (IoT) which is used for industrial purposes [2]. Cyber-Physical Systems (CPS) is a connected physical and engineered system that transforms how users interact with the physical objects around them. [3] for example, the Industrial Control System (ICS). Furthermore, Technology has advanced rapidly in many domains as a result of ICT convergence (Information and Communications Technologies). One of those technologies is the ICS (Industrial Control System), which is a concept that incorporates a variety of control systems. Also included in the ICS is SCADA (Supervisory Control And Data Acquisition). It's a system for controlling water, oil, and gas distribution, pipelines, ships, trucks, rail systems, and wastewater collection systems that gathers data and sends it to distant field control stations and monitors. [6]

As the Internet and CPS have developed, an increasing number of industrial control equipment are being connected to the Internet for remote control and monitoring. Furthermore, the security of industrial protocols is rarely taken into account. Researchers focused on securing ICS infrastructures and preventing them from being subverted by attackers in the face of Internet threats against ICS devices and systems. [8-12]

Honeypot is a strong effective way to study attacker behavior. Honeypot could record all activities of attackers and scanners by impersonating ordinary workstations or devices, which could subsequently be used for further analysis.

There are many open-source honeypot systems, which have been proposed and widely used to analyses ICS attack behaviors, on the other hand, have some restrictions.[13–15, 17, 18] , Honeypot is only designed to collect access traces from outsiders, hence it has no data analysis capabilities.[19]

To address the above - mentioned problem, we propose a SCADA architecture that includes Honeynet and machine learning approaches for capturing and analyzing suspicious SCADA behavior from the Internet.

The following paper is organized as follows. We describe the Industrial Control Systems and Honeypot first, then look to the Related Work and then present the design and Architecture details of. Finally, we close with the conclusion and Future Work.

II. OVERVIEW OF INDUSTRIAL CONTROL SYSTEMS

Industrial Control Systems is a system or network of devices, sensors, controllers used for industrial purposes in different critical fields such as water, gas pipelines and other industrial uses. ICS include: supervisory control and data acquisition (SCADA), distributed control systems (DCS), Programmable Logic Controllers (PLC), Historian and many other systems [4]. Each system has different characteristics and uses but at the same time they could integrate with each other to achieve an overall goal [5]

A. Supervisory Control and Data Acquisition (SCADA)

SCADA is a highly distributed system [4] that collects and processes critical data to make long distance operational decisions [5].

B. Distributed Control Systems (DCS)

DCS are mostly used in process-based industries such in water and wastewater treatment, chemical and food industry. The architecture of DCS contains a Supervisory level followed by integrated sub-systems that manage the details on the local process level[5].

C. Programmable Logic Controllers (PLC)

PLC is a solid-state control system that is mainly used for industrial purposes within SCADA and DCS for long distance and big systems and it's also used as the core system in small systems [5]. It stores instructions to build a specific system function such as data and file processing [4].

III. HONEYPOT

As the applications of ICS -in particular- are increasing so do the attacks and as we mentioned earlier ICS are uses in critical industries some are managed by the federal agencies of the US and most of them are privately owned[5], hence, in order to protect theise critical systems from attackers, we can deploy different security aspects such as IDS, IPS, Firewalls, cryptography, antivirus and other tools can be applied to secure the system. But on the other hand these tools or mechanisms can't offer a chance to study how the attackers can get into the system, how they act, what is the attack flow, and also the identity of the attackers[1]. However, here comes Honeypot and Honeynet, a system that traps and tricks the attackers into believing that they have got access to the actual assets of the system while they just got into a simulated system that registers their attack patterns.

A. Honeypot classification

Honeypot is classified into 7 categories : purpose, role, level of interaction, scalability, resource level, availability of the source code, and application.

As our concern on this paper is ICS security we will focus on Honeypot classification by application

1. classification by application:

Honeypot is a tool created to be applied and used within IoT systems such as IoT candyjar honeypot and IoTPOT honeypot, or for industrial purposes with ICS such as Honeyd+ honeypot and CamouflageNet honeynet[1].

IV. RELATED WORK

A. Related Work:

As we are in the phase of studying a proposed solution that could make use of the data provided by Honeypots and apply some of ML techniques on these data, we found that there are two papers that proposed different solutions that we could build our work on top of.

Hidemasa Naruoka and others created a Honeynet (Camouflagenet) not only to record how the attackers act and what types of attacks they do, it also helps to disturb the attackers and make it harder for them to do the attack.

By understanding the three phases of the attack (information gathering, attacking time and cover up time) Camouflagenet is working on increasing the time of information gathering phase for the attackers as the attackers have to check and collect information about every object on the network before they start the attack.

CamouflageNet increases the number of the connected machines in the ICS network by adding multiple honeypots to the network and creating a dynamic network so it's harder for the attacker to know the real asset from the virtual one which cost the attackers more time to collect data so the security staff can know about the attack before it begins [6].

The Second paper by Seonggwan Ahn and others was a study on improving SCADA architecture by adding to the network a Honeypot and a protector. Honeypot is to attract the attackers by making them believe that they got to the actual system. Then the IPS (Intrusion Prevention System) got the log information from the Honeypot to determine if the user is registered on the system as a trusted entity or not. If the user is not registered, the IPS will send a command to the protector to do an ARP spoofing attack. The ARP spoofing attack changes the gateway's MAC address on the attacker ARP table to the protector's MAC address which will prevent the suspicious user from entering the SCADA network[7].

B. Differences from the existing work

In the first paper they deployed a number of honeypots (CamouflageNet) into the ICS network, therefore, When it detects a signal of reconnaissance activities such as an NMAP scan at the "Information Gathering Time," CamouflageNet updates its configuration. This dynamic reconfiguration makes the attackers to re-reconnaissance, wasting important "Information Gathering Time" and disrupting their concentration efforts. [6].

Because CamouflageNet is primarily designed to collect access traces from outsiders then disturbing the attackers by re-reconnaissance, it lacks data analysis capabilities. So we employed ML for CamouflageNet design.

On the other hand, if we use CamouflageNet into SCADA architecture instead of one Honeypot we could increase the intelligence of the system to defend itself depending on how the user acts, not what is registered on the database.

In nutshell, Future ICS, SCADA and CPS honeypots and honeynets, will benefit from machine learning techniques to propose smarter decoy systems that can adapt to attacker actions, discriminate known attacks from new attacks.

V. PROPOSED ARCHITECTURE

A. Background Information

As the applications of ICS -in particular- are increasing so do the attacks and as we mentioned earlier ICS are uses in critical industries[5], hence, in order to protect these critical systems from attackers, we can deploy different security aspects such as IDS, IPS, Firewalls, cryptography, antivirus and other tools can be applied to secure the system.

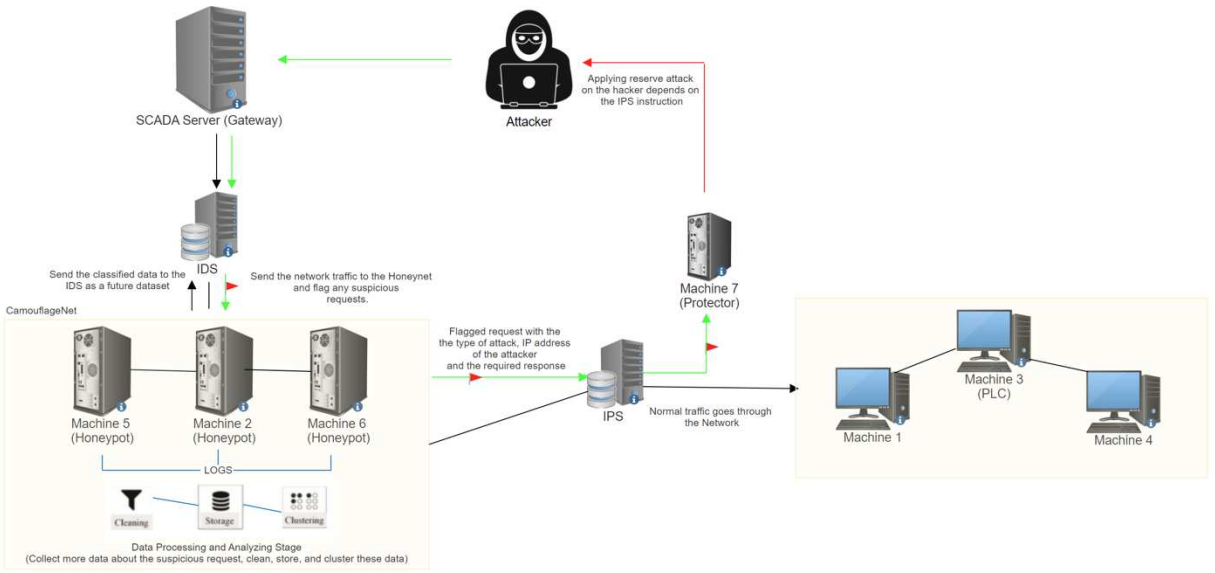

Figure 1: The Architecture

B. Proposed Architecture

We will add to ICS/ SCADA architecture Four components, CamouflageNet Honeypot[6] to study the attack and the attackers, IDS (Intrusion Detection System) and IPS (Intrusion Protection System) to detect the attack and tack action, and a Protector to carry out IPS commands. We will also add intelligence to CamouflageNet Honeypot by employing ML techniques within.

- *IDS (Intrusion Detection System)*: Check if the request on the network is a suspicious request or not and send the information to CamouflageNet for further intelligent verification.

- *CamouflageNet Honeynet*: Connected Honeypots built especially for ICS in order to distract the attackers and buy the security administrators more time to understand the attack and defend the system. It's built on top of Honeyd honeypot[6]. ML-enhanced CamouflageNet will send back detailed information for the IDS to be used as a future dataset. And send a detailed investigation report to IPS.

- ***Data Processing***: The major purpose of DP(Data Processing) is to cluster and analyze the data collected by CamouflageNet Honeypots, as well as to forecast the trend. DP creates attacker behavior models to summarize attacker activity patterns by studying historical records of attack behaviors. DP also gives you the ability to foresee how attackers will act in the future.

DP does some preliminary work, such as data cleaning, formatting, and saving the log information to the database, before diving deep into the data. The preprocessed data is then turned into a vector sequence. IP, port, protocol, and function code are all included in each vector. Within a specific time interval, a sequence of vectors is obtained. The vectors are then classified into multiple clusters using clustering methods. Each cluster could represent a different attack type.

The detailed process is described as follows:

A. Within a given time frame, DP collects data from the same IP address. Both benign network traffic and malicious commands (such as "CPU Control request STOP") are included in the data. From network flows, we extract information such as source IP, destination IP, destination port, date, packet length, and malicious command data. To represent the behaviors of network traffic intercepted by CamouflageNet Honeypots, the features are concatenated into a vector.

B. To obtain many clusters, we suggest applying the k-center clustering algorithm.

- *IPS (Intrusion Protection System)*: Defend the *system* from any suspicious requests and take action by forwarding the request to the Protector instead of forwarding it to a real asset on the network and instruct the Protector about what reverse attack should be applied.
- *Protector* : Programmable unit that does a reverse attack on the attacker depends on the Intel from IPS.

VI. CONCLUSION

In order to defend any system from attack we need to build strong architecture that can do both, understanding the new attacks and the attackers, and defend the system in front of these attacks. And to achieve these two goals we proposed an enhanced architecture for SCADA systems that could add more Intelligence to it while dealing with an attack.

Usually in SCADA systems we use IDS/IPS to defend the system from any intruders but on the other hand IDS/IPS has a high false alarm average and that is caused by the lack of analysed data. Hence, we found a need to do double verification on the requests that could be flagged by mistake by the IPS or the IDS by adding a ML-enhanced Honeypot to SCADA architecture that analyse more detailed data about any suspicious move within the system and create a future learning dataset for the IDS and an Intel data to the IPS which play a good role in instructing the protector about which action should take defending the system.

VII. FUTURE WORK

Increasing the number of objects on any system leads to increasing the amount of the data flow within it which causes some delay and a slower performance. For that reason we need to do more research on how we could enhance SCADA security without affecting data flow and response performance.

VIII. CONTRIBUTION

The first contribution of this research is proposing a method to increase the efficacy of IDS/IPS as a security measurement within SCADA systems and decrease the false alarm ratio. The second contribution of this research is proposing a method to enhance an existing ICS honeypot by deploying machine learning techniques into it and integrating it within the SCADA architecture. The Third contribution of this research is proposing a more powerful and secured architecture for SCADA systems that detect the attack, analyze the attack and the attackers and defend itself from the attacks in a more intelligent way.

1. *References*

[1] J. Franco, A. Aris, B. Canberk, and A. S. Uluagac, “A Survey of Honeypots and Honeynets for Internet of Things, Industrial Internet of Things, and Cyber-Physical Systems,” IEEE Communications Surveys and Tutorials, 2021, doi: 10.1109/COMST.2021.3106669.

[2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, “Industrial internet of things: Challenges, opportunities, and directions,” IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 4724–4734, Nov. 2018, doi: 10.1109/TII.2018.2852491.

[3] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, “Cyber-Physical Systems: The Next Computing Revolution,” 2010.

[4] S. Essentials, “Industrial Control Systems.” [Online]. Available: http://www.kuleuven.be/eena

[5] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, “Guide to Industrial Control Systems (ICS) Security,” Gaithersburg, MD, Jun. 2015. doi: 10.6028/NIST.SP.800-82r2.

[6] H. Naruoka et al., “ICS Honeypot System (CamouflageNet) Based on Attacker’s Human Factors,” Procedia Manufacturing, vol. 3, pp. 1074–1081, 2015, doi: 10.1016/j.promfg.2015.07.175.

[7] Han’guk T’ongsin Hakhoe, IEEE Communications Society, Denshi Jōhō Tsūshin Gakkai (Japan). Tsūshin Sosaieti, and Institute of Electrical and Electronics Engineers, ICTC 2019 : the 10th International Conference on ICT Convergence : “ICT Convergence Leading the Autonomous Future” : October 16-18, 2019, Ramada Plaza Hotel, Jeju Island, Korea.

[8] Mo,Y., Chabukswar,R., Sinopoli,B.: DetectingintegrityattacksonSCADAsystems.IEEE Trans. Control Syst. Technol. 22(4), 1396–1407 (2014)

[9] Kleinmann, A., Wool, A.: Automatic construction of statechart-based anomaly detection models for multi-threaded SCADA via spectral analysis. In: ACM Workshop on Cyber- Physical Systems Security and Privacy. ACM, pp. 1–12 (2016)

[10] Zhou, C., Huang, S., Xiong, N., et al.: Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation. IEEE Trans. Syst. Man Cybern. Syst. 45(10), 1345–1360 (2015)

[11] TomlinJr.,L. ,Farnam,M.R.: Aclusteringapproachtoindustrialnetworkintrusiondetection [EB/OL]. http://insurehub.org/sites/default/files/reports/CyberSecurity_Final_Research_Report_LTomlin_MFarnam%20(1).pdf

[12] Goldenberg, N., Wool, A.: Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems. Int. J. Crit. Infrastruct. Prot. 6(2), 63–75 (2013)
[13] Bodenheim, R., Butts, J., Dunlap, S., et al.: Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices. Int. J. Crit. Infrastruct. Prot. 7(2), 114–123 (2014)
[14] Serbanescu, A.V., Obermeier, S., Yu, D.Y.: A flexible architecture for industrial control system honeypots. In: 2015 12th International Joint Conference on e-Business and Telecommunications (ICETE). IEEE, vol. 4, pp. 16–26 (2015)
[15] Formby, D., Srinivasan, P., Leonard, A., et al.: Who's in control of your control system? Device fingerprinting for cyber-physical systems. In: Network and Distributed System Security Symposium (NDSS) (2016)
[16] Krawetz, N.: Anti-honeypot technology. IEEE Secur. Priv. 2(1), 76–79 (2004)
[17]Bodenheim,R.C.:ImpactoftheShodancomputersearch engineoninternet-facingindustrial control system devices. Air Force Institute of Technology Wright-Patterson AFB OH Graduate School of Engineering and Management (2014)
[18] Serbanescu,A.V., Obermeier,S., Yu,D.Y.: ICSthreatanalysisusingalarge-scalehoneynet. In: Proceedings of the 3rd International Symposium for ICS & SCADA Cyber Security Research. British Computer Society, pp. 20–30 (2015)
[19] J. Cao, W. Li, J. Li, and B. Li, "DiPot: A distributed industrial honeypot system," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 10699 LNCS, pp. 300–309. doi: 10.1007/978-3-319-73830-7_30.