Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P532-cd

Creating a testbed for cyber security assessment of Industrial 4.0 factory infrastructure

Per-Arne Jørgensen

Risk & Security Department, Institute for Energy Technology / Østfold University College, Norway.
E-mail: per.arne.jorgensen@ife.no

Stine Aurora Mikkelsplass

Risk & Security Department, Institute for Energy Technology / Østfold University College, Norway.
E-mail: stine.mikkelsplass@ife.no

Addressing cyber security in Industry 4.0 is challenging, as it requires a holistic view of the perspectives of people, processes and technology. To understand how industrial control systems (ICS) are affected by cyberattacks, we must first understand how systems behave during normal operation. The emergence of Industry 4.0, and the upcoming Industry 5.0, results in industries deliberately connecting both new and legacy operational technology (OT) to the internet, i.e., information technology (IT). This interconnection between IT and OT is motivated by gaining data insights to increase efficiency and economic gain, as well as the opportunity to centralise both security monitoring and control over the factory floor. This convergence of IT and OT environments makes OT systems susceptible to external attacks. To obtain realistic insights into the introduced vulnerabilities, real systems should be exposed to cyber security event conditions in hardware in the loop environments. We are using an Industry 4.0 training factory for this purpose from the German company Fishertechnik. The setup of the ICS testbed was comprised of a training and learning environment that through learning can provide comprehends Industry 4.0 applications and demonstration. The Industry 4.0 factory environment is controlled by a real SIMATIC S7-1500 programmable logic controller (PLC) from SIEMENS. The components building up this out-of-the-box setup consist of different factory modules that replicate real components. In this paper, we present how we have established an ICS testbed, including the challenges experienced from aligning best practices, architecture designs and guidelines for network communication and integrating different agents (data beats) for data collection. We will also discuss the use of a SIEM solution, called Elasticstack, for data collection to provide these insights for further exploration of methods for anomaly detection and knowledge building.

Keywords: Industry 4.0, IACS, ICS, IT/OT, cybersecurity, monitoring, Elastic Stack

1. Introduction

The fourth industrial revolution (Industry 4.0) connects industrial systems and technology solutions in new ways, often through the use of the Internet. These connections could potentially introduce new threats and vulnerabilities, making cyber security issues relevant for systems that traditionally were not designed or hardened to be used for this purpose. In this paper, we report on the results of a project conducted at the Institute for Energy Technology (IFE) in Norway. To better understand an industrial testbed we set up an Elastic Stack with the purpose to explore available data for insights and explore the domain of operational technology with a focus on the concepts of Industry 4.0 and Industrial automation and control systems (IACS). During this project, we aimed to set up and integrate an operational system (OT) in an Elastic Stack environment (IT) at IFE's Cybersecurity Center housing the equipment. We used and evaluate the capabilities in Elastic Stack for data logging, analysis, and visualisation. As there were no "ready-made" solutions available for data collection between OT and IT environments, we had to design and set up a testbed infrastructure for this particular purpose.

To support holistic cyber resilience IFE's Cybersecurity Center brings together researchers and engineers from multiple disciplines and domains, like security engineering, safety risk management, human factors, process engineering and simulation. By utilising hardware-in-the-loop simulations to provide realistic settings to better under-

stand how an ICS system worked during an attack, this environment can test different settings and attack vectors like active penetration testing of a product or systems to find their weaknesses.

This paper is organised as follows: In the background chapter, we describe the need for using real systems by setting up an ICS testbed. Followed by an overview of different state-of-the-art infrastructure setups and describing the challenges with IT and OT to address the research question. In the discussion chapter, we present the approach and experiences integrating the testbed with Elastic Stack, followed by a conclusion and future work.

2. Background

Data has become a valuable asset in the later years Bhageshpur (2019), as available computing- and processing power has increased in the last decade. With more processing power, more extensive data sets can be handled in near real-time with more complex analysis for better data-driven decision-making. Big data technologies can utilise high-performance computing and distributed storage systems to manage different types of data. Industry 4.0 is made possible by combining the worlds of OT and IT, connecting digital and physical technologies such as artificial intelligence, IoT, robotics and cloud computing, to name a few. With the introduction of IT and new connections into the OT domain, the risk to open legacy systems to a vast amount of advanced cyber threats increases. The challenges remain in the technology usage where new meets old when mixing legacy industrial protocols (OT) and Internet protocols (IT). Gartner defines Operational technology (OT) as the ”hardware and software that detects or causes a change, through the direct monitoring and/or control of industrial equipment, assets, processes and events” Gartner (2023). This gap of interoperability between these two domains is often called the ”IT/OT convergence” challenge. The differences between IT and OT become a challenge when we consider the availability aspects of an OT system where such systems were designed with high availability and reliability in mind. Another challenge with mission-critical systems is when there is a need to plan for maintenance or troubleshooting tasks, it is costly to have a downtime of systems in operation. According to Lekidis the ”use of computers and IT networks to improve visibility and remote maintenance introduces security risks that came along with the use of IT technologies though were not considered at all.” Lekidis (2023).

With the emerging use of the Internet of Things (IoT) and the integration of physical machines with networked sensors and software, the lines between IT and OT are diffuse. Traditionally, operational technology (OT) has been considered separately from IT, although one will find that industrial solutions mostly connect OT and IT systems. Cyber security has been considered for IT, while OT had a strong focus on safety. OT environments include physical processes that can lead to dangerous situations with unwanted events, and safety had to be a priority to keep people and processes out of harm’s way. Industrial Control Systems (ICS) security is in many ways different from traditional IT security where the focus is on protecting the information assets with measures of confidentiality, integrity and availability. However, when considering the protection of OT systems it’s the opposite, where the main objectives were availability with consequences of impacting physical processes and safety. This would require a more holistic approach to the people, process and technology perspectives. To better understand the gap between IT and OT, the most significant difference can be shown in terms of an attack outcome, where an attack on IT could lead to data breach and theft. The impact of an attack on the OT side could affect the physical world like people, the environment or the physical thing or asset.

3. State of the art infrastructure setups

Cyber ranges is an emerging technology setup used to explore IT and OT systems, under different attacks, behaviour and conditions in a controlled and interactive lab environment Taylor (2023). The term testbed is used to describe platforms for experimental research, product development and prototype testing across a range of different

disciplines. However, in this project, it is defined as "a controllable cyber environment that enables experimentation" Edgar and Manz (2017a). In contrast to a Sandbox[a], which is often used in software development. Testbeds consist of both hardware and software components and are often developed for a specific purpose and designed for a particular test case. An essential feature of a testbed is that it can be designed to test a subset or a subsystem of a more extensive system. This is especially important regarding large and complex systems because it allows researchers to look into more specific areas of a system that would be both costly and challenging to test in a full-scale scenario Arntzen et al. (2019). Establishing a testbed with an infrastructure for collecting available data is valuable for competence development and training purposes. What type of data could we expect from an IACS system? Before starting to collect data from the factory, we must understand the industry process and investigate what types of different industrial protocols were in use to build knowledge for what type of data is available for monitoring purposes.

4. Challenges and research questions

Cyber threats are challenges that may impact and disrupt businesses. Facing these challenges requires resilient network architectures and methods for detecting and responding to cyber-attacks. This is becoming a crucial task in every business to protect valuable assets. With the emerging development of connected things and devices, the need of having data available is vital to providing new business insights. Integrating new sensors and actuators into existing infrastructures introduces new risks, e.g. security and data breaches and lack of control, especially when dealing with operational technology.

Operational data is a valuable asset for factory owners even with a small data set, as insight into only one particular operation can lead to valuable information for the performance and the process as a whole. However, with larger sets of data, new possibilities emerge. The value of larger data sets can be found more valuable, through the use of analytical software tools utilising machine learning and Big data. These tools enable businesses to sense-make a vast amount of data, opening the possibility of finding previously unknown connections and extracting valuable insights.

The insights into the difference between IT and OT rely on the intersection of understanding how an OT system is "connected" and behaves during normal operation. Data from both sides must be gathered and consolidated. The unique challenge in collecting these data in a secure and controlled manner, will if not done correctly, have an impact on both the IT and OT side. According to Paes et al. (2020) the IT/OT convergence is defined as follows: *"IT/OT convergence is the integration of IT systems applied to data-centric computing with OT systems used to monitor events, processes, and devices and make adjustments in enterprise and industrial operations. IT is composed of those hardware and software system technologies that allow for corresponding information processing. OT is supported by physical devices, i.e., switches, sensors, power distribution networks, valves, motors, and software that allow for control and monitoring of a plant and its associated equipment."* Paes et al. (2020).

The data gathering was collected and analysed through a realistic Industry 4.0 scenario. We used this data and information for cyber security management and to empower operational cyber security detection. The result of this process was presented to demonstrate the capabilities of the Elastic Stack for cybersecurity management and detection in an IT/OT environment. We have developed a research question for this purpose of investigation.

RQ1: How to monitor and detect cybersecurity threats in an IACS environment using the Elastic Stack?

5. Methodology

The purpose of this project was to explore, test and evaluate the capabilities of the Elastic stack as a tool for cybersecurity management and monitoring of an IT/OT environment. The method used for

[a] A Sandbox is often an isolated virtual machine used to test program code.

this exploration follows an exploratory research method. By investigating the problem, we built a better basis for understanding the problem faced by the industry today; the problem of integrating OT systems with IT systems and the intersection between them. An exploratory research method would be the best fit for this purpose Edgar and Manz (2017b), and by using this method, we aimed to;

- gain further insight into how the Elastic Stack can be used for cybersecurity management and monitoring
- considers topics from both the aspect of operational technology and informational technology.
- gain a better insight and understanding of the what and the how of Industry 4.0 data collection

6. Testbed requirements and implementation results

The Elastic stack is the software tool chosen for this project that stores and analyses the collected data. For any test or experiment, there is a risk of making configuration changes to the testbed setup that introduces hazards to a system and influence the results of the experiment. This is the same for digital testbed setups. In cybersecurity research, testbeds are particularly useful as there are fewer risks to real systems. As cyber research is in general recent, and system solutions are generally not unique, predicting the specific behaviour of the system under cyber conditions is difficult. Hence, a testbed alleviates the need for prediction and allows for the experiment to run its course without risks. However, the testbed setup and the data-gathering mechanisms from the experiment can influence the experiment itself as well as how the experiment is analysed and understood.

How would the software stack from Elastic fit the purpose of data-gathering solutions and bridging the gap between IT and OT? The Elastic Stack[b] is built on the components from ELK, an acronym for Elasticsearch, Logstash and Kibana. These three open-source products were used to search and analyse data (Elasticsearch), process and transform data (Logstash) and visualise data (Kibana). The ELK Stack allows for storing multiple kinds of data and features integration that analyse the data with machine learning capabilities. While setting up this OT/IT integration, we focused mainly on the perspective of using IT technology to gather data from OT through the use of different data beat agents.

Network segmentation is a central part of a defence-in-depth strategy with three (3) goals: (1) to protect the network services; (2) to reduce the susceptibility of endpoint systems and applications to threats originating from the network; and (3) to protect data during transmission across the ICS testbed network. As every environment is different, we need to identify hardware and software components like network equipment, PLCs, disk drives, computers, firewalls, services, and protocols. In addition, be depict shared hardware, network paths and software present in the environment Delay (2021). The following principles apply; where lower security levels require fewer security controls, segmentation or zoning and sub-zoning provide a defence-in-depth strategy for better protection.

- Level **0** - The physical production line process in the factory
- Level **1** - The field devices with actual actuators and sensors: **1a)** Factory modules i.e the VGR, HBW and MPO, **1b)** Environment sensors i.e temperature, air quality and humidity
- Level **2** - The process control network: **2a)** The Siemens S7-1500 PLC, **2b)** Raspberry PI IoT gateway, **2c)** Wireless Access Point
- Level **3** - Supervision and human-machine-interface (HMI): **3a)** Engineering station computer with Siemens TIA portal and supported applications, **3b)** Node-Red with data flow and dashboard of factory insights and control.

Network security zones define a foundation for a balanced and layered security architecture that can

[b] https://www.elastic.co/elastic-stack/

support a range of security solutions. The attack surface of the systems within a zone can be significantly reduced by exposing a limited number of services through a zone's perimeter and implementing rigorous access controls to limit access to specific groups and users. Additionally, if a breach occurs, an attacker would have to compromise access to all the outer zones before getting to the zone where the critical data is stored, reducing the likelihood of data ex-filtration and increasing the availability of critical systems. A zone is defined as a separate network segment (VLAN or physical network). It is a logical grouping of information systems in an enterprise network where the zone class is a collection of zones with the same security rules. "Today, you have more open factory floors and supply chains. You must have granular visibility and controls, eliminating risks of unauthorised users, applications and data on the network. You also have to accept that nothing is perfect despite these controls, that threats can still get in." Zhang (2019).

6.1. *The ICS testbed*

By utilising affordable commercial-off-the-shelves (COTS) sensors and technologies based on IoT principles, it is possible to gain a range of new data about the manufacturing process. Both the PLC program and the system it selves delivered from the Fischertechnik company, as the scope of this project is to focus on the IT/OT data, not to harden the system for resiliency. Only minor modifications were performed when necessary for data gathering. The setup was a training and learning environment where we can learn and comprehend Industry 4.0 applications. The factory was controlled by a SIMATIC S7-1500 programmable logic controller (PLC) from SIEMENS. The setup consists of factory modules as follows; a storage and retrieval station consists of an input and output station with colour detection and an NFC Reader for tracking the workpieces in the factory. A vacuum gripper robot with a 3-axis 3D industrial robot arm can pick up workpieces and move them within the working space. For storing the workpieces, the high-bay warehouse is modern and effective computer-aided storage for the retrieval of goods. The multi-processing station has a kiln oven and can move the workpieces automatically through several stations that simulate different processes. The sorting section has built-in colour detection used to automatically separate workpieces of different colours. To monitor the environment in the factory, the testbed is equipped with a bosch 680 sensor (for temperature, humidity, air pressure and air quality) and a pivoting camera. In addition to the Siemens PLC controller, the factory is equipped with the Fichertechniks own ROBOTICS TXT controller, with Bluetooth, WiFi and a radio frequency (RF) module, for connecting the environmental sensor along with the pivoting camera Fischertechnik (2023). The network infrastructure for the testbed in Figure 1 describes how the hardware components and devices are interconnected between the different network segments of the testbed.

The testbed was designed specifically to answer the research question: How to monitor and detect cybersecurity threats in IACS? First, we need to know what processes to monitor, the overview and a good understanding of the components and devices in our system, along with their capabilities and features. Second, we needed to know how these components communicated and which protocols were used. The next task was to get a holistic view of the actual process, to better understand how the factory production line works in an everyday situation without interruption. The fourth task, and maybe the most important one, was to investigate where it would be beneficial to install and integrate the data beats agents for retrieving data points of interest and forward them to Elastic Stack for further analysis. Elasticsearch is, at its core, a server that can process JSON requests and provide JSON data in return. It is built on Apache Lucene, a full-text search engine software library developed in Java. Elasticsearch provides the opportunity to store, search and analyse massive volumes of data in near real-time. Elasticsearch comes with extensive representational state transfer application programming interfaces (REST APIs) for storing and searching data, using a scale-out big data platform for indexing different types of log data from various sources to aggregate

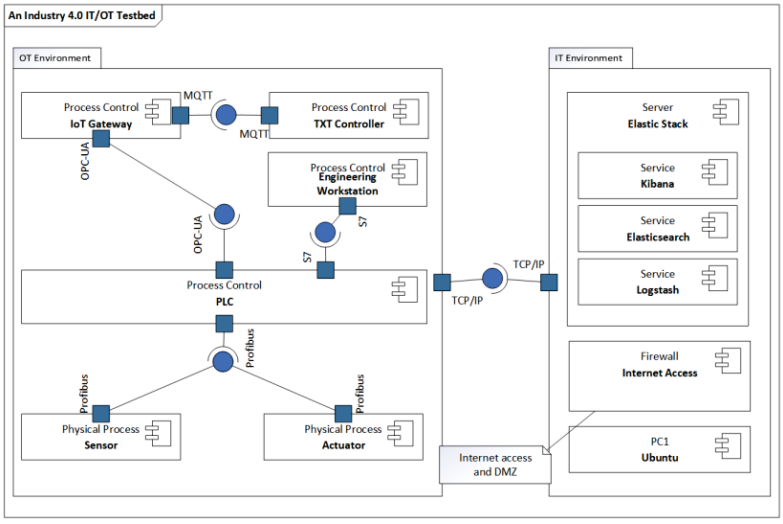

Fig. 1. UML component diagram overview of the IT/OT testbed

and compile data for new insights.

7. Discussion

For the initial set-up of the factory, it was possible to follow guidance and manuals from the Fischertechnik website. However, when moving on to adapting the Structured Control Language (SCL)-program and configuring the network infrastructure, this was only possible by an exploratory approach of trial and error. The Fischertechnik factory was delivered ready to use, and no actions were taken regarding security, i.e., it came with default configurations and default or no password. Performing security-hardening of the factory was outside the scope of this project. We decided the assumption that the network surrounding the factory should be considered a black box as this was outside the project scope. This meant that it would not know the details of how malicious code got into the factory network, only that it did. The threat scenarios for the testbed were planned according to the data available from the testbed. By using the threat modelling tool STRIDE Kohnfelder and Garg (1999), software specifically created for uncovering vulnerabilities in a system, a detailed list of the possible malicious and non-malicious threats concerning safety and security was produced.

Exploring the network traffic would give us a starting point to understand protocols and application information. The use of tools for analysing network information and hooking onto the data streams was needed to gain detailed insights. Wireshark was used to analyse the network traffic on the network. From the protocol overview in Wireshark we quickly identified the OPC-UA and Profinet Realtime protocols. Wireshark provided more details on the network packet level and with further analysis we managed to find which components were communicating and on which ports. But for now, we were most interested in exploring the different protocols in use on the internal process control network to gain an overview of network communication. Another useful tool was UAExpert which gave more profound insights into the low-level data and tags published through the OPC-UA interface. This tool supported OPC-UA features like data access, alarms conditions, historical access and calling of UA methods to read out tags from the function block structure in the PLC program to find relevant data. The tool, MQTT Explorer, was used to read out all messages from the manufacturing process using the MQTT protocol. The TXT controller was the central MQTT broker in the system that was responsible for publishing messages to the Fischertechnik

cloud service for operating and controlling the factory.

One of the challenges was to understand where data is generated in the system and how we could integrate it with the system for data collection. Although the factory could utilise the MQTT protocol to collect data from the production manufacturing process, we were still lacking important operational data from the PLC. Operational data is typically metric data such as performance parameters like CPU load, memory and number of client sessions. Further exploration of possible data was needed. Using the UAExpert tool to explore the data types available by connecting this tool to the PLC's OPC-UA server. The hierarchy of data values where specified in a structure that was quite new to us to debug and understand. Each data value had its own unique address that was aliased by a tag name and we were able to interpret it through the SCL program and function blocks.

As we did not know what types of data it would be possible to collect from the testbed, it was challenging to have full control over the data-gathering aspect of the project. Following the snowball sampling Edgar and Manz (2017a) method for sampling OT data, we started sampling the most available data, and once getting these data correctly into the Elastic Stack, we moved on to similar types of data. For the IT data, it was possible to approach the issue with more of a purposive sampling method Edgar and Manz (2017a). When working towards acquiring system metrics from the PLC, we experienced some IT/OT challenges in practice. Identifying, locating and retrieving system metrics and correct parameters (like CPU, memory, etc) was more challenging than initially thought.

The beat agents all had a default setup of what data to retrieve. Through the configuration file of the Machinebeat and Filebeat agents, it was possible to specify what type of data points should be retrieved from the OPC-UA and MQTT interfaces. In practice, this meant that only the data already available through these interfaces were available. Other data could be collected by altering the SCL program, however, this required a skill set the project group did not have. When digging a little bit deeper into the configuration file of the Packetbeat agent, we were surprised to find that the agent lacked the possibility to customise and define your own protocols. The Packetbeat did not support other protocols like OPC-UA, MQTT or Siemens S7 SCL in contrast to marketing materials available from Elastic, when they state to merge into OT environments. It seems that COTS components for merging IT and OT are not yet easily available.

The Machinebeat was still in the experimental version as of writing and is stated to support both OPC-UA and MQTT protocols. The MQTT implementation with the Machinebeat did not work as expected. On the IoT Gateway, Filebeat was successfully installed and configured to forward system logs and connect to the TXT Controller grabbing all the MQTT topics and sending them to the Logstash service.

During the project, we explored that knowledge and experience from both domains in IT and OT could easily be underestimated by both sides. In this convergence gap, misunderstanding and shortcuts can introduce new risks and threats from IT into OT. Though we sought to gain a deeper understanding of the OT environment with the help of OT experts, still a limitation of our work was that we have been overly biased from an IT point of view.

By choosing to use an IT/OT testbed we introduced the real challenges to the integration and data collection processes. These challenges are mostly the same that would occur if such a system was installed in an actual factory, albeit on a smaller scale. The presence of actual OT components and industrial protocols in the testbed increases the validity of results gained; as data packets have to travel through an actual network infrastructure, real errors are introduced into the testbed. Depending on the type of research this can be a wanted or unwanted effect, however, for this project, it was highly advantageous for the validity of the challenges met during IT/OT integration.

8. Conclusion

In this paper we specifically address challenges with IT/OT systems and data communication and the need of integrating agents for security monitoring, and what type of competencies would be necessary to address cyber security challenges in the intersection between IT and OT.

We have developed an IT/OT testbed to explore a full-scope Industry 4.0 factory with a defined manufacturing process. However, with our experience of using IT-oriented SIEM tools like the Elastic Stack to gather and analyse data from the Industry 4.0 training factory, we can argue that the product maturity level in analysing industrial data still has considerable potential for improvement to be used for cybersecurity anomaly detection.

Much time was spent on researching and understanding the OT environment and the specific challenges that followed once the factory was in-house. The IT/OT testbed worked as intended with basic functionality. However, the availability of OT data in the IT environment was less than expected.

The main learning point was the challenge of understanding and working with OT technology and OT data with an IT background and limited experience with OT systems. Conducting this type of experimental research requires a deep understanding of both the technology and processes in the OT environment, as well as a good understanding of IT.

9. Future work

There are many avenues for future research within anomaly detection in an IACS system. The intersection between IT and OT is still a research topic to be further explored due to the ongoing digitalisation of the industry. Having a higher quality of data available for new insights and better decisions seems to be a driver but with the trade of having less security and perhaps less control. However, understanding the convergence of IT/OT with highly configurable testbeds could be a better way to explore the challenges and provide insights on how to protect, monitor and detect anomalies in an OT environment. Adopting a Zero Trust architectural model could be essential to cope with the emerging vulnerabilities and threat landscape.

Acknowledgement

We want to thank our supervisor at Østfold University College, professor Øystein Haugen, for his support and constructive feedback throughout this project. We would also like to thank senior researcher John Eidar Simensen and Dr Bjørn Axel Gran, our internal supervisors at Institute for Energy Technology for their advice and the opportunity to invest in the factory testbed.

References

Arntzen, S., Z. Wilcox, N. Lee, C. Hadfield, and J. Rae (2019). Testing innovation in the real world.

Bhageshpur, K. (2019). Council post: Data is the new oil – and that's a good thing. Section: Innovation.

Delay, J. (2021). Effective ICS cybersecurity. using the IEC 62443 standartd.

Edgar, T. W. and D. O. Manz (2017a). Chapter 13 - instrumentation. In T. W. Edgar and D. O. Manz (Eds.), *Research Methods for Cyber Security*, pp. 321–344. Syngress.

Edgar, T. W. and D. O. Manz (2017b). Part IV. experimental research methods. In T. W. Edgar and D. O. Manz (Eds.), *Research Methods for Cyber Security*, pp. 213. Syngress.

Fischertechnik (2023). Lernfabrik 24v.

Gartner (2023). Definition of operational technology (OT) - gartner information technology glossary.

Kohnfelder, L. and P. Garg (1999). The threats to our products.

Lekidis, A. (2023). Trends and security challenges from the IT/OT convergence.

Paes, R., D. C. Mazur, B. K. Venne, and J. Ostrzenski (2020). A guide to securing industrial control networks: Integrating IT and OT systems. *26*(2), 47–53. Conference Name: IEEE Industry Applications Magazine.

Taylor, H. (2023). What is a cyber range?

Zhang, F. (2019). The path to industry 4.0 is through cybersecurity.