Operational Security Analysis and Challenge for IoT Solutions

Yuan Gao[1] Xinxin Lou[2]

Abstract: The marketing engagement of Internet of Things (IoT) shows a wide vista together with Industry 4.0 regarding modern manufacturing and services. However, the evolution of technologies and rising regulation concerns regarding security and privacy are bring challenges to IoT solutions. On one side, the security analysis of IoT solutions has to consider the security posture in a much wider scope including both edge and cloud sides even across global geo-locations. On the other side, new regulation requirements demand a full tracking of data access. In addition, authorizations should be evaluated explicitly and can be revoked any time for maximizing data protection. Both challenges can be solved by implementing a novel security model targeting those requirements while zero trust model is a good candidate. Thus in this paper, we compared the most commonly used perimeter security model and the zero trust model under the circumstance for modern IoT solutions. Furthermore, from the regulation perspective, the concepts of zero trust model are analyzed to show its compliance with regulation requirements. For easing the discussion of IoT solutions, a general IoT architecture is proposed and relevant zero trust model implementations are described. Especially, the zero trust model relevant security controls are highlighted as a guidance for the design of IoT solutions. As the conclusion, we propose a general implementation of zero trust model within the context of IoT solution to solve the challenges facing by the industry.

Keywords: Operational Security Model; Zero Trust Model; Cloud Security; Edge Computing; IEC 62443; Industry 4.0; GDPR; IoT; IIoT

1 Introduction

Along with the development of cloud computing and wireless technology, Internet of Things (IoT) is achieving the biggest ever market engagement. According to the GSMA Intelligence projects, the market cap of IoT will be more than 1 trillion until 2025 [GS18]. Apparently, the ubiquitous existence of IoT devices and the heavy back and forth data traffic are arising new security concerns and requirements. In addition, the deployment of cloud technology uses infrastructure sharing to enable the resource elasticity and to reduce the cost whenever applicable. Furthermore, the virtual factory concept in Industry 4.0 allows the temporary combination of services across different geographical locations [Rü15]. The above mentioned reasons are eliminating the security boundaries of IoT solutions especially the network perimeter. Meanwhile, the tremendous number of IoT devices and their limited security capacity require dynamic while robust security in architecture designs. At last, the

[1] Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, yuan.gao@ovgu.de
[2] Bielefeld University, xlou@techfak.uni-bielefeld.de

doi:10.18420/inf2020_30

new regulation requirements on personal identifiable information (PII) expose IoT solutions to the compliance risks: e.g. General Data Protection Rules (GDPR), Payment Card Industry Data Security Standard (PCI-DSS) and Health Insurance Portability and Accountability Act (HIPPA).

For handling the complex IoT solution architecture as well as fulfilling the compliance requirements, zero trust model is being introduced into modern system architectures. Zero trust model will assume no trust for any personnel, entities or services involved within a system process. Thus, every service request and reply should be under continuous monitoring and will trigger alarms for possible compromising. The operational 3-domains security model proposed in previous work considered a flexible representation of system security architecture as well as continuous security monitoring [GI2019]. Thus, it is suitable for further adaptions and extensions to handle the above mentioned IoT challenges. Within the 3-domain security model, we will address the required features for implementing an IoT solution in line with the zero trust model. In addition, analysis on chosen IoT scenarios implemented with zero trust model can show the compliance of this architecture regarding different security regulation requirements.

The rest of this paper is organized as follows: firstly, Section 2 describes a general IoT solution architecture especially defines the edge and cloud locations as two major focuses. Secondly, in Section 3 we discuss related works and relevant security regulation requirements. Then Section 4 discusses the challenges faced by the perimeter security model. Section 5 introduces the zero trust model within the IoT solution context. It is compared with perimeter security model and demonstrates how the regulation compliance can be met by implementing this model. In addition, Section 6 highlights the different security concerns between IoT and IIoT. Finally, the conclusion is summarized in Section 7 together with the discussion regarding future works.

2 IoT Solution Architecture

As a booming technology, innovative IoT solutions are being created every day. So there is no standard or typical IoT solution architecture. Here we will use Figure 1 as an example of IoT solutions for describing the common features and components. From a simplified data flow perspective, we can go through the picture from left to right. On the most left side, IoT devices collect data from their sensors (like the thermometer on the top) or receive voice commands from nearby users (e.g. the smart TV or the speaker in the middle). IoT devices can send data to the cloud either via a relaying/edging device as the green colored symbol indicated in the picture or communicate with cloud directly, like the IoT car module or the camera in the bottom left corner. In the middle within a public cloud, collected data will be processed and stored in object storage or databases for serving queries or historical tracing. Data stored in cloud will be exposed as services. They can be accessed and consumed by other (cloud-native) services for creating various business plausible results, e.g. a historical regional temperature record. Front-end services reside in the public cloud, such as web servers and load balancers are omitted here. Finally, on the right side of this figure, useful information can be viewed on terminals, e.g. smart phones or laptops with effective UX designs like a dashboard or diagrams.

For easing the description, we may consider two IoT scenarios regarding Figure 1: smart home and smart car.

Smart Home: In this scenario, a user at home can talk to her/his smart speaker for controlling other devices at home. For example, the voice of “Open my TV” will be sent over to the cloud for recognition. The identified command for opening a TV registered in this home will be returned to the smart home router and after validation it will be further forwarded to the TV to react to this command.

Smart Car: Sensors are on a car keep sending values, e.g. location and speed to a central cloud service endpoint. At the same time, the cameras installed on top of traffic lights are transmitting pictures taken in a regular interval to the cloud for image processing. All information, collected from cars and cameras, are put together in a data processing application to build the real-time traffic loads within a specific city.

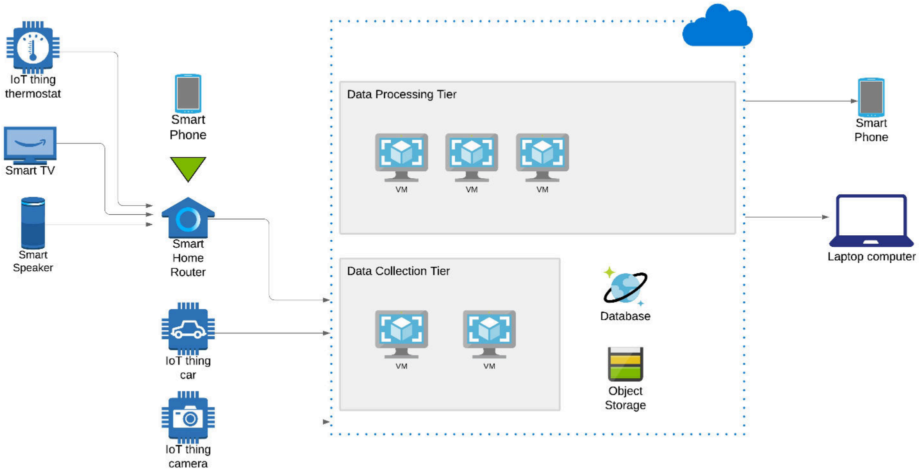

Fig. 1: A general IoT architecture.

Regarding the IoT solution architecture proposed in Figure 1, following declared important concepts can help the analysis in later chapters:

IoT Devices are devices that connect to network with specific cyber-physical functionality (e.g. sensors or actuators). Normal IoT devices can be massively produced with a low cost and have limited capacities on computation and storage. They can be equipped with only essential or little security features. However, a special kind of IoT device can have advanced network and security features, which we name as **edge device**. An edge device can relay the communication between other IoT devices and cloud services/terminals. Thus it is an ideal host for placing security controls and temporary data processing functions. Figure 1 depicts the smart home router as an edge device (indicated by the green symbol). However, an edge device does not have to be a real device but can be a virtual one. In this case, its extra network processing and security capacities can be achieved by combine several IoT devices into a mesh.

Terminals are the ending devices by which users communicate with the IoT system. Two kinds of typical terminals are smart phones and PCs. By running applications on a terminal, users can view valuable results from the IoT solution and send commands to the system to perform specific actions or modify configurations. A terminal can connect to IoT services published on cloud or communicate with devices exposed an interface hosted on the cloud. Both connections are established over internet thus communication protection mechanism is needed. In addition, terminals can access IoT devices via WIFI or Bluetooth in a small area, e.g. within a home LAN.

IoT Applications can be any application running on IoT devices, cloud resources and terminals. According to their purposes, applications can be classified into two types: *Management Application* and *Data Application*. Management applications are used for controlling the IoT system and individual devices. Comparing to this, data applications collect, process and illustrate data as the functionality of an IoT solution. For example, in the smart scenarios, applications collect values and pictures from IoT devices and they are processed and build up together in the cloud services for illustrating the real-time traffic status.In this paper, data collection and processing will be further discussed due to their tight connections to security and regulations.

3 Related Work

The work in [KS18] reviewed the security issues and challenges in IoT. Some researchers consider combining the software-defined networking (SDN) into the IoT in order to bring security and the privacy with flexibility and scalability [KBL18]. DeCusatis et al. [De16] propose a network architecture which enables an explicit zero trust approach. This architecture is based on a steganographic overlay (with authentication tokens in the TCP packet request), and first-packet authentication. Vanickis et al. describe a policy enforcement framework to address many of open challenges for risk-based access control for zero trust networking [Va18]. Ahmed et al. [ANT2020] propose a model which provides the access control to sensitive data in zero trust model. In this model, an access control proxy is used to protect the sensitive data to implement the access control. The access control is realized by performing the analysis on access request, user type, device type, application type and data type.

To regulate the potential security threats, some privacy and security rules or standards are developed by specific organizations, e.g. the Guide to the General Data Protection Regulation (GDPR) [In18], Payment Card Industry Data Security Standard (PCI-DSS)

[PC04] and Health Insurance Portability and Accountability Act (HIPAA) [Of02]. They require PII data should be well protected. To achieve this, data access must be traceable, and protection action will be taken when the data is under a certain level of risks.

The proposed operational model here is an extension of our previous work of the 3-domains model regarding threat domain, system architecture domain and security domain [Ga19]. The work discussed in this paper focuses on the security domain.

4 Challenges of Perimeter Model

Before getting into detailed analysis regarding zero trust model, it is meaningful to declare the features of security perimeter model as a comparison. A perimeter model in security classifies a system into different areas or zones. Same security requirements are shared within one area/zone while the security level is going up for inner zones which have shorter perimeters to the system center, in other words, protection goals. The benefits of security perimeter model are quite clear. Firstly, the classification of zones with different security requirements reduces the analysis and management overhead thus increasing the security governance. Security postures can be improved by adding extra zones outside, such as demilitarized zone (DMZ). Secondly, by assuming same security requirements within one zone, traffic monitoring can focus on the communication between zones. This means investing network security appliance only on tunnels (conduits) to reduce costs. Finally, security zones are prioritized naturally by their perimeters to the center, which guides the continuous activities for security improvements. According to these reasons, the perimeter security model fits the traditional server-centric architecture well. The Zone-Conduit model introduced in IEC 62443-3-2 is an abstract implementation of perimeter model in industrial fields [IE15].

However, toward the evolution of IT technologies and new regulation requirements, the security perimeter model is not feasible anymore. On one side, the monolithic server-centered architecture is reaching its bottleneck while instead, distributed and decoupled architectures like micro-services are trending popular. Without a fixed combination of components, it is difficult to classify a stable list of security zones since the size and contained components can vary when time goes by. The virtual factory concept within Industry 4.0 is an example of this challenge. Against a virtual factory consists of dynamic available services over network on the global scale, it is almost impossible to implement a security perimeter model on it. On the other side, the authorization to access data is not checked explicitly in every access. In other words, the perimeter security model is a static model. Once an identity or a service is authorized to access the data, they obtain the access until the authorization is removed manually. For example, one employee can resign and leave the company at the same day. However, the users/passwords of the employee will stay in the company's systems for a long while until the IT-staff decide to do a cleaning. In addition, it is also a violation of the regulations of GDPR and PCI-DSS to have non-recorded data access without checking attached policies explicitly.

5 Zero Trust Model

5.1 Core Concepts

In Chapter 4 we discussed the features and challenges of the perimeter security model. In a perimeter model, a certain level of trust will be established when specific conditions are met. For example, if a server locates in the back-end network and hold the right access key, it will be authorized to access the database. However, after the authentication and authorization happened in the beginning, the database access (from the server) will not be tracked and the conditions for authorization will not be checked in a specific period. This mechanism can reduce authentication/authorization overhead effectively. However, it lacks the continuous monitoring on each access. In addition, servers will be vulnerable within a period of trust relations. For solving these challenges, zero trust model can be applied by implementing following rules:

- **Rule-1(Traced Access)**: all data access should be tracked and stored in logs.
- **Rule-2(Explicit Evaluation)**: data access authorization should be evaluated explicitly before deciding *Allow* or *Deny*.
- **Rule-3(Automatic Revocation)**: under certain conditions, the system is capable of revoking data access authorizations automatically.

First of all, all data access should be logged and this rule contains several layer of details. Since IoT solutions across edging and locations, the tracked data access includes both raw data collected on devices and internal processed data. From the security perspective, both the privacy of individual user and the trade secrets of IoT solution providers should be protected. In addition, logs need to be time-stamped correctly for later correlated analysis, e.g. using a security information and event management (SIEM) system. Based on this, it is possible to apply artificial intelligence (AI) for indicating possible compromises by learning from normal data access behaviors.

Secondly, for maximizing the protection of data, access request should be explicitly evaluated every time without exceptions. However, this will bring the system extra overhead while provide more secure access control. Later in Section 6 we will discuss relevant trade-offs especially for non-privacy scenarios. In general, a careful classification of data can help to design cost-effective evaluation mechanisms thus reducing the overhead.

Finally, the explicit evaluation discussed above enables the automatic authorization revocations to prevent malicious accesses. Revocation can be triggered by different ways. For example, when an abnormal access behavior is detected, the authorization can be revoked to interrupt possible on-going attacks. In the case of a false positive alarm, authorization can be requested by administrator or through user's self-services. Authorization can be automatically revoked by a timely manner as well. As an example, an administrative data access authorization should be allowed only in a reasonable time windows and will be revoked automatically when expires. This will minimize the possible attack window and reduce the administrative efforts regarding the security governance.

In the following section, implementations of the three rules will be proposed to show how zero trust model can improve the security posture of IoT solutions.

5.2 Threats and Controls

Regarding an IoT solution described in Figure 1, it is meaningless to go through all attack vectors and associated mitigation. For example, uploading crafted malicious firmware to IoT device or virus affected PC/smartphones will not be considered. Instead, only the zero trust model relevant threats will be considered. There are threats for the edge location and the cloud location. It is also possible a threat with a long kill chain can go over both locations. Compared to this, security controls are bounded to their locations. A control or a mitigation will take effect either on the edge or on the cloud with limited exceptions. One such kind of exception is customized communication protocol which will affect both edge and cloud locations. However, considering the majority of IoT solutions involving public cloud, it is very rare that a customized protocol will be used. In summary, security controls which implement zero trust model on either edge or cloud location will be discussed. The threats identified here are not part of a formal risk assessment and they can be easily performed by *insiders*. Thus, in the following paragraphs, we will discuss threats without assessing their impact and likelihood explicitly.

Edge Location

Table 1 lists three identified threats towards a general IoT system on the edge location.

Tab. 1: Threats on Edge Location.

No.	Threat	Relevant Rule(s)
1	Traffic Sniffing	rule 1
2	Rogue Device	rule 2 and 3
3	Malicious Operation/Configuration	rule 2

The first considered threat is traffic sniffing. We can use the scenario of smart home to analyze this threat. An attacker might sniff data traffic within the home network, either via cable or wireless connections. It can be achieved by compromising the single router in the network or by hijacking wireless connections between IoT devices within the network. This kind of attacks are possible since violating rule 1: *Traced Access* described in Section 4.2. In the traditional perimeter security model, a home network after a router plus a firewall towards internet is considered secure. So communications within the home network are treated as harmless without protection. To fulfil the requirement of rule 1, communications within the home network must be monitored by recording at least basic information, such as time, protocol and length of packets. In addition, communications within the home network should be encrypted for preventing manipulations which can break the integrity of traced records.

The second threat is to put a rogue device into the edge network. A rogue device can act as a normal IoT device. However, it can be controlled to perform malicious activities, such as eavesdropping its surroundings or joining DDoS attacks. This threat can be mitigated by implementing rule 2: *Explicit Evaluation* and rule 3: *Automatic Revocation*. For rule 2, in a zero trust model, a connection request must be checked whether its authorization exists and is still valid firstly. There should be no assumption that a device is *probably secure only* because it locates within the edge network. Request from an unknown device within the network should be rejected and an alarm will be triggered for attracting the administrator's attention. Furthermore, according to rule 3, regular review of existing devices should be performed. An administrator or a user has to provide a trace of the device how it is installed within the network in a given time window. If not, the device must be isolated automatically. Again in the smart home scenario, if the user only setup a speaker and a TV previously, then any packet from the unknow thermometer device should be dropped.

The last threat is performing malicious operations or configurations on IoT devices. Direct access to IoT devices, such as operating on device user interface is not protected due to the assumption the physical security protection is sufficient. As a security guideline for medical devices in Hospital, user interfaces should be protected by passwords and manufacturers do provide such functionality for compliance purposes. However, in reality, due to the high frequency of emergency situations, these passwords are normally disabled or set to an easily memorized value, such as *0000*. Apparently, this is a violation of rule 2. Considering our smart home example so far, it is also so annoying to input a password every time serve a cup of coffee from a coffee machine. As a plausible mitigation, innovative authentication methods like face recognition should be applied to reduce the impact while to fulfil the requirement of rule 2. In addition, implementing rule 1 to log user activities might provide a basement for detection of abnormal behaviors, which can contribute to mitigation too.

Cloud Location

On the cloud side, three threats are identified in Table 2.

Tab. 2: Threats on Cloud Location.

No.	Threat	Relevant Rule(s)
1	Insufficient Logging	rule 1
2	Hard-coded Access Key	rule 3
3	DoS on shared resources	rule 2 and 3

The first considered threat on the cloud location is insufficient logging configured for infrastructure as a service (IaaS), e.g. virtual machines (VMs). The access to data on VMs will be logged on the application level within individual VMs. Both Amazon Web

Services and Microsoft Azure provide the services to grab and backup logs from VMs for monitoring or regulation requirements. Without correct logging configuration, in case a VM is terminated due to a hardware failure under the virtualization layer, the application log content will be lost. This will cause missing data access records, which violates of rule 1 and regulation requirements, such as PCI-DSS. Thus, IoT solutions should follow the cloud best practices to enable IaaS logging and log collection functionality.

In the next we pay attention to the threat from hard-coded access keys. Access keys can be used for authenticating a user or an application for accessing data or cloud services. However, hard-coding access key in an application is a bad security practice. On one side, this violates rule 3, since it is very difficult to revoke the access key which is designed for a long retention time. Revoking an access key will invalid all encrypted data linked to this key. Furthermore, using access key cannot assign a user or an application authorization based on the least privilege principle. Holding an access key means the root access privilege. Instead, role-based access control (RBAC) should be used to provide granular access controls. It is also possible that the access key will be leaked to public not only to insiders. However, since normally cloud services are configured with firewall rules, here we consider only the access key might be misused within a trust network, e.g. a private network on cloud. Considering the smart car scenario, it is a best practice to segment service access controls regarding different cities. However, with a universal access key, no valid access control will be in place.

At last, the threat of DDoS attack on shared resources will be discussed. Dockerization is the state-of-the-art technology of computing virtualization. Applications will be packaged with dependencies and executed in individual light-weighted docker containers. Reflecting our smart car scenario, servers for different cities can be deployed in different containers thus to isolate them targeting a robust service architecture. However, as the design, several docker containers will be hosted on the same hardware. Thus, it is possible for one docker container to consume up all computing resources so that other containers will be in a status of DoS. To mitigate this threat, an implementation of rule 2 and 3 can be helpful. Within a computation windows, a container has to request computing resource before its execution (rule 2). Once allowed, the authorization to computing resource will be revoked after a given time period (rule 3). Thus, other containers can access computing resources without starving status.

6 IoT vs. IIoT

To summary the analysis of IoT solutions so far, implementations of zero trust model can effectively overcome the drawbacks of perimeter security model mentioned in Section 4. However, there is no single silver bullet for everything. Especially, according to the IEC 62443-3-2 [IE15], a perimeter-based zone-conduit model is proposed as the solution for Industrial Automation and Control System (IACS). In the context of this paper, this means towards an IACS enhanced with IIoT solution, perimeter security model is the major choice. This trade-off can be understood from two perspectives:

- **Availability vs. Confidentiality**: For IACS, the availability of system has higher priority.
- **Safety vs. Privacy**: Functional Safety is strictly regulated in industrial fields while regulation on privacy is being progressed.

On one side, availability has the highest priority in the availability, integrity and confidentiality (AIC) triad in the context of an IACS. Especially, when the availability of a critical infrastructure, e.g. the power grid is affected, the impact will be very high. In this case, security controls that affect availability should be avoided. For example, an automatic access deny on requests from safety-critical system is not allowed. In additional, the implementation of rule 2 (Explicit Evaluation) might affect the availability for time-critical functions, such as hard real-time tasks. On the other side, the zero trust model focuses on data access which is more important for protecting privacy. If the IACS or a partial system of it is running without processing PII, implementations of zero trust model is not mandatory without regulation requirements. There are more differences between a consumer-oriented solution and an industrial solution, by which trade-offs on system architecture will be decided.

To conclusion, zero trust model do not have to be implemented completely for an IACS. However, as the trend of Industry 4.0, the border between IACS and consumer systems is blurring. For example, medical device manufacturers are under the heavy regulation requirements from both safety and data protection aspects. Thus it is meaningful to discuss factors for deciding trade-offs between a perimeter security architecture and a zero trust architecture. The first trade-off is regarding the three rules of a zero trust model. The implementation of rule 2 and rule 3 can be flexible while rule 1 should be implemented whenever possible. Due to the highest priority of availability, security controls affect system performance should be avoided. The implementation of rule 1 will collect and store all data access and operation logs without put much impact to the system performance. Then the collected logs can used for analysis to detect abnormal behaviors which might indicate an attack.

The second trade-off is regarding the two architectures (perimeter and zero trust). In some cases, an IACS can be divided into sub-systems. The security architecture of individual sub-systems can be designed following different guidelines. Especially considering one example of IIoT scenario, a great number of sensors will be installed to monitor the water quality in a water supply area. In this example, the availability of sensors is robust it can be designed with redundancy (e.g. 2-3 times of the minimum required number). Then the security architecture can be designed following the zero trust model for maximizing security. However, due to a local disaster or a regular maintenance, when number of sensors are reduced to a certain level, the security system should be switched to a mode based on perimeter model, which will maximize the availability of the monitoring system for water supply.

7 Conclusion and Future Work

In this paper, we discussed the actual challenges over the security architecture of IoT solutions. For solving the problems, we compared the perimeter security model and the zero trust model. Three rules were abstracted and discussed within the IoT solution context. In the future, a more detailed IIoT solution can be selected and analyzed using the proposed methodology. It can be expected that more refinements about the trade-offs on security architecture will be discussed. Especially, together with a formal risk assessment, the priority and cost for security mitigation should be quantified for supporting decisions.

Bibliography

[De16] DeCusatis, C.; Liengtiraphan, P.; Sager, A.; Pinelli, M.: Implementing zero trust cloud networks with transport access control and first packet authentication. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud). IEEE, pp. 5–10, 2016.

[Ga19] Gao, Y.; Zid, I. B.; Lou, X.; Parekh, M.: Operational Security Modeling and Analysis for IACS. Gesellschaft für Informatik, 2019.

[GS18] GSMA: Opportunities in the IoT: evolving roles for mobile operators, https://www.gsma.comot/wp-content/uploads/2018/09/New-Roles-for-Operators-in-the-IoT-k.pdf, [Online: accessed 02-August-2020], 2018.

[In18] intersoft consulting: General Data Protection Regulation, https://gdpr-info.eu/, [Online: accessed 02-August-2020], 2018.

[IE15] IEC: 62443 Security for Industrial automation and control systems, Part 3-2: Security risk assessment and system design, 2015.

[KBL18] Kouicem, D. E.; Bouabdallah, A.; Lakhlef, H.: Internet of things security: A top-down survey. Computer Networks 141, pp. 199–221, 2018.

[KS18] Khan, M. A.; Salah, K.: IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems 82, pp. 395–411, 2018.

[Of02] Office for Civil Rights, HHS: HIPAA Privacy Rules, tech. rep., [Online: accessed 02-August-2020], 2002.

[PC04] PCI Security Standards Council: Payment Card Industry Data Security Standard (PCI DSS), tech. rep., 2004.

[Rü15] Rüßmann, Michael; Lorenz, Markus; Gerbert, Philipp; Waldner, Manuela; Justus, Jan; Engel, Pascal; Harnisch, Michael: Industry 4.0: The future of productivity and growth in manufacturing industries. Boston Consulting Group 9, 2015.

[Va18] Vanickis, R.; Jacob, P.; Dehghanzadeh, S.; Lee, B.: Access Control Policy Enforcement for Zero-Trust-Networking. In: 2018 29th Irish Signals and Systems Conference (ISSC). IEEE, pp. 1–6, 2018.