Heinrich C. Mayr, Martin Pinzger (Hrsg.): INFORMATIK 2016,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2016 559

SIEM: Policy-based Monitoring of SCADA Systems

Yuan Gao[1], Xin Xie[2], Mithil Parekh[1], Edita Bajramovic[3]

Abstract: Security Information and Event Management (SIEM) systems work on SCADA systems by observing and reacting to the dynamic security-related events of the target automation system. These events are created by collecting/filtering raw logs maintained by its sub-components. Preferably, logging items are attached with synchronized timestamps. Specific data of security-related event can be correlated and analyzed as security measures of the SCADA system. Possible correlation rules represent the power of SIEM system for handling security dynamics. A SIEM system can help to recognize security breach in a short time and optionally can react to the breach automatically. SIEM systems are able to monitor the system's residual risks, while continuously track the deployed security controls and measure their effectiveness. In this paper, we proposed the common requirements of a SIEM system and discussed its important enhancements within the context of SCADA systems. The SIEM system can be supported by the overall security model and designed in a model-driven manner. At last, this paper proposes a preliminary model of correlation rules.

Keywords: cybersecurity, SIEM, SCADA, log management, real-time system, network security, security modelling, security testing

1 Introduction

An information security management system (ISMS) [II13] is required by an industrial automation and control system (IACS) [Ie15]. Especially, the supervisory control and data acquisition (SCADA) system, which acts as the major type of an IACS, needs to be protected by ISMS too. For identifying as well as mitigating vulnerabilities of a SCADA system, security experts perform risk assessments according to the system's architecture and configurations. Vice versa, forensics investigation could be performed when a security breach in the daily operation is discovered, to determine how the breach happened as well as to collect the digital evidence. Both the risk assessment and the forensics investigation approach work either on a static system specification in its design phase or on a static snapshot of the system. Compared to this, the security information and event management (SIEM) system takes the responsibilities to handle the dynamic security statuses of a SCADA system. Within the run-time environment of the system, the SIEM system is able to monitor its residual risks and to track the triggered security

[1] Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, Universitätsplatz 2, 39106 Magdeburg, yuan.gao@ovgu.de
[2] Siemens AG, Digital Factory, Siemensallee 84, 76187 Karlsruhe, xin.xie@siemens.com
[3] Friedrich-Alexander University Erlangen-Nürnberg, Informatics 1 Department - IT Security Infrastructure, Martensstr. 3, 91058 Erlangen, edita.bajramovic@gmail.com

events on-the-fly. Besides [Ie15], SIEM is also addressed in the security context of Industry 4.0 [Bu16].

Run-time events of an automation system carry the system's statuses as well as their transitions in between. The logging functionality of the system will continuously record these events, typically with timestamps. One major purpose of recording these events is the diagnostic of the system. Furthermore, security-related events are part of these run-time events and are especially associated to the system's security statuses, e.g. a locked door is opened with an employee card. Either a normal operation or a security breach can trigger multiple security-related events. Furthermore, among them, the event that indicates a possible security breach is named as security event (SE) [II11]. On one side, the SIEM system collects and analyzes them (in real-time) to create security warnings/reports which support the daily work of the incident response team (IRT). On the other side, with the record of system's executions, the SIEM system can also assist security experts to perform risk assessments as well as support forensic investigations with strict constraints.

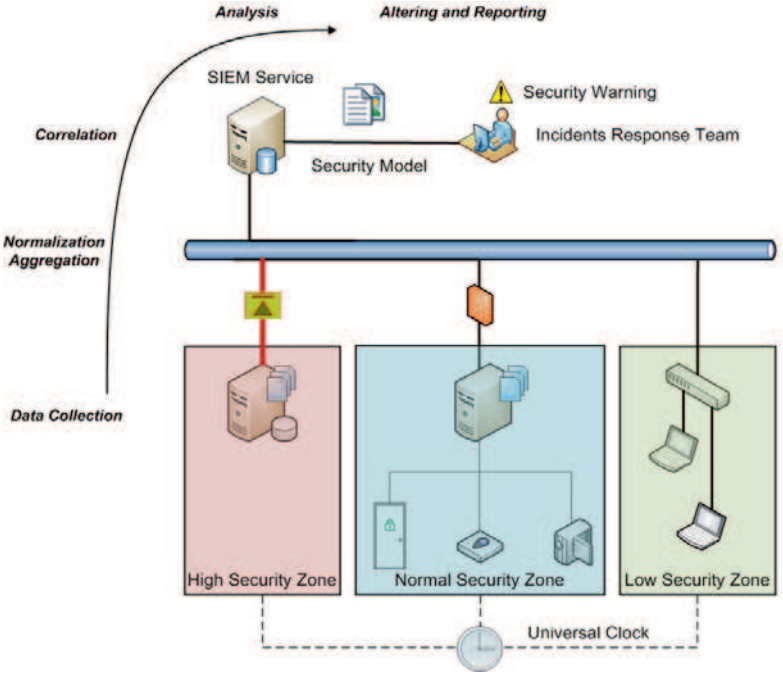

Fig. 1: A SIEM prototype and the associated data flow.

In this paper, we stick to the context of a SCADA system. However, with proper adaptions, the same design concepts of SIEM systems can be applied on a general IACS or its other sub-types, like distributed control systems (DCS) as well as compositions of Programmable Logic Controllers (PLCs) [Ni11].

2 Common Features of SIEM systems

Fig. 1 (partially based on preliminary work at ISO/IEC JTC1/SC27 WG4) illustrates a SIEM prototype and its associated data flow. Security-related events are created from logging items belong to separate components of a SCADA system. In this paper, we assumed that all logging items are already timestamped by its logging system. These event specific data will be collected to a central SIEM service for further processing. The curve in the left part of the picture shows the major steps of SIEM processes which are further explained in Section 2.4.

2.1 Timestamp of Security-related Events

For meeting the real-time impactions of SCADA systems, the timestamps attached to security-related events have to share a predefined accuracy across different sub-components. As depicted in Fig. 1, a universal clock is ideal for the synchronization task. However, the universal clock might be unavailable since it is not mandatory for a SCADA system. Normally each sub-component has its own clock. In this case, proper protocols need to be considered for the reliable time synchronization between different logging systems [St10]. Beside the time differences among multiple logging systems, the time accuracy also needs to be considered. When several events happen in the same minimum time unit (e.g. in one second), their sequence cannot be determined.

Within the same logging system, the timestamps bring the most important information to indicate the chronical sequence of events. Based on concrete implementations of logging systems, it might happen that events happened later were recorded in the log file priory to the events happened earlier. However, their sequence can still be determined according to the associated timestamps. When the execution time between two relevant events are not so important as well as these events are recorded in the log file strictly chronically, these events can be processed without timestamps.

A SIEM system can be created by introducing accurate time synchronization between sub-components in the design phase. Or with existing logging systems, the SIEM can be designed to be able handle timestamps in a fuzzy way which means inaccurate timestamping of events is allowed.

2.2 Collection of Event Specific Data

For correlating security-related events originating from different components, the relevant SIEM data, either events or raw logs, will be collected to a central service for further processing and analyzing. On one hand, the created security-related events need to retain the references to their raw log items, which is important for fulfilling the requirement of forensics investigations. On the other hand, the transfer of SIEM relevant data should neither overload the SCADA system nor breach network security controls.

More considerations about the transfer security are discussed in Section 4.2.

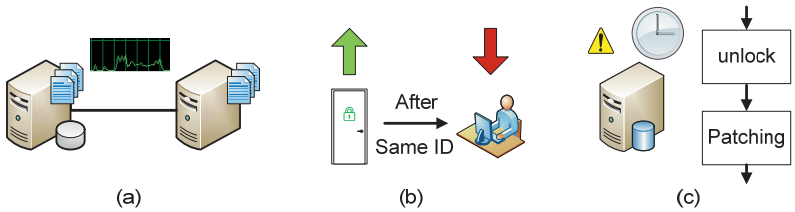

Fig. 2: Examples for correlation rules:
(a) abnormal network traffic (b) logon after leave out (c) configurations off working hour

2.3 Correlation of Security-related Events

Correlation is a key functionality of SIEM systems. Security-related events or SEs from the same component or across the considered SCADA system will be combined following predefined correlation rules. The historical correlations can back-track available security-related events in a long time to detect abnormal behaviors. For example as Fig. 2a, too much network traffic is observed or the statistic distribution of message types changes dramatically. Both of cases might indicate an attacker joined the network and manipulated the delivered messages. Furthermore, the example introduced in the preliminary work at ISO/IEC JTC1/SC27 WG4 (depicted in Fig.2b) also shows the potential of correlating events come from different components in real-time. In the example, two independent SEs of different components are taken into account:

1. The ID card of an employee was recorded by the entrance access control system that the person left the company.
2. After 30 minutes of the leaving event, a legal system logon action with the user/password belong to the same employee in the server room was logged.

Each of the two SEs complied with its relevant security controls. However, combination of the two events within a short period indicates a high probability of a security breach. An additional correlation example as illustrated in Fig. 2c could be the maintenance engineer patching the current system: when cabinet lock-monitoring system is in place, according to the flow chart, the uploading action should happen after the unlocking event of the cabinet. A reverse order or a missing of the unlocking event might indicate a compromise through network. This example can also be linked to a working schedule of the facility. When the uploading action happens after the regular working time, it is very likely an attack is ongoing.

In the correlation examples, different security systems and their associated controls as well as relevant security objects are involved. By utilizing the overall security model [Wa16] which has them already in place, correlations rules can be described by modelling the relations among them.

2.4 SIEM Process Steps

The curve in the left part of Fig. 1 shows the major steps of the SIEM processes. In this section, individual steps are introduced according to the sequence indicated by the arrow. Meanwhile, the requirements and potential risks of the SIEM system are further discussed. Especially, in parallel to these process steps, the raw logs need to be archived for later on tracing back or for possible forensic investigations.

Data Collection: The system's security-related events are collected from the raw logs of its sub-systems. These system logs and their associated logging functions are designed for either diagnosis purpose or compliance with security controls [Ie15]. Thus, the considered system logs firstly need to be examined whether they contain sufficient information as described in Section 4.1. Furthermore, during run-time of the SIEM system, log filtering is required to collect only security-relevant logging items.

Normalization and Aggregation: In this step, logging items from different sub-systems thus using various formats will be unified. The extra or duplicated information which is not required for the future steps will be reduced. Especially, in the normalization step, the associated timestamps need to be transformed into the same time system. At last in this step, security-related events will be created from the unified logging items.

Correlation: The correlation of security-related events is introduced in Section 0.

Analysis: In this step, correlation results can be linked to relevant security objects of the overall security model described in [Wa16]. With the model's support, potential security breaches (e.g. the abnormal logon action) could be discovered. Furthermore, the SIEM system will monitor the SEs associated to known vulnerabilities, thus to alert the incident response teams in advance. Meanwhile, the invoked SEs will be linked to applied security controls through the model. When a specific security control is triggered too often, then smart testing can be deployed on the associated security object to assist improving the system security.

Alerting and Reporting: The analysis results and possible subsequent system changes will be ordered according their timestamps. Alerts will be automatically sent to the IRT to acquire attention while the attached analysis report will assist the security staff to figure out the system vulnerabilities under attack and to deploy countermeasures.

Raw Log Archiving: Additionally, the raw log needs to be archived. On one hand, since the logging system could be interrupted to cause damage to log files or missing of events, log files need to be archived regularly. On the other hand, considering some sub-components like PLCs have only limited resources, the size of current log file is limited thus its content needs to be archived to a permanent storage. The archived raw logs can be utilized by the tracing back functionality or for fulfilling the regulation requirements on critical infrastructures to keep raw information for forensics investigations.

3 Key Features of Commercial SIEM Systems

Considering the amount of logging data that may be generated in real-time by SCADA systems, a commercial SIEM system is expected to be able to filter and search security logging information with precisely defined log file formats. Besides, the data selection process needs to cooperate with real-time system events. SPLUNK [Sp16] is one of the popular commercial solutions for the logging management of SCADA systems. Besides the ability to handle big amount of log data, it also supports different storage types such as cloud-storage, which brings more flexibility to various SCADA systems.

Other solution, the IBM QRadar also demonstrates the idea to connect the SIEM system and the IT Infrastructure under monitoring [Ib16]. It focuses on the network security while enables real-time correlation to identify high risks as well as security breaches. Since the associated firewalls as well as network monitoring systems are already policy-based, QRadar can be configured / extended for automated incidents response and the vendor announced to comply with regulatory requirements on data collection and secure reporting.

Furthermore, combining different existing tools is also an option. For example, a SIEM system could use SPLUNK to create real-time security-related events from a great amount of logs while data mining tools can be employed to build correlations from the big data set. Even different SIEM systems can be combined: for example, office IT has already SIEM from Vendor A while the SCADA system is associated with another SIEM system from Vendor B. These two SIEM systems can be combined together and share selected data between them. Furthermore, all commercial SIEM systems are able to receive threat intelligence. With this information SIEM system can always indicate the new threat. More attractively, the correlation and associated security objects could base the unified security model [Wa16] thus to implement a model-based SIEM system.

4 SIEM Framework for use in Industrial Automation

4.1 Policy on generation of Security-related Events

The security-related events created from raw logs have to fulfill the data requirements of the SIEM system. Considering different logging systems which have various non-standard data format, this is not a trivial task. In this paper, we gave the most important required information of a security event for guiding the generation process. Tab.1 lists the required attributes: *ID* is the unique identifier of the security event. *Timestamp* Attribute records when the event happened. When no universal clock is used, its specified clock reference needs to be attached. *Type* indicates which status changes happened in the system. For example, a booting event or a shutdown event has clear semantics relevant to the system. In the next, *Source Information* maintained the link from the security event to its raw log item. Finally *Invalidation Conditions* defines when

the security event turns to be invalid. The condition could be the arrival of a relevant security event or an expiration time is reached in a real-time system. For example, the close event of a door will terminate the open event of the same door. An authentication event should be valid only within a predefined amount of time.

ID	Timestamp	Type	Source Information	Invalidation Condition(s)

Tab. 1: Essential Attributes of a security-related event.

4.2 Policy for secure collection of SIEM relevant data

The collection of security-related event specific data should not bring additional risks to the SCADA system. On one hand, the traffic of SIEM data collection will consume extra network bandwidth thus it needs to be well designed to avoid impacts to the system's normal operation. On the other hand, the SIEM central service collects data from components by a read-only manner. According to [Ie13], the communications between different security zones have to be constrained regarding to their security levels. As shown in Fig. 1, communications from high and middle security zones need to be protected against leaking to lower security zones through the SIEM service. The high security zone can be protected by data-diode while the middle security zone can be protected by firewalls configured with access restrictions, which acts as a physically unidirectional security gateway.

Besides the impact on communications, the secure collection of data requires appropriate configuration management of sub-components. The agent-based solution needs to install specific software on sub-components to handle the collection. One obvious advantage is the agent can be configured to periodically push log data to the SIEM service. However, considering attaching a SIEM system to an existing SCADA system, the agentless solution requires no modification of sub-components thus more compatible. The drawback is that for the SIEM to collect log data from a sub-component, a temporary user/account for connection is required and might introduce extra risks.

For protecting the target SCADA system, the integrity of raw log needs to be protected from tampering by indicating devices [MT15]. This will also help to prevent the manipulation or misleading of the SIEM system. The tamper resistance measurement can already be involved in the logging system. For example, in a CCTV system with long-term storage, the id of records should contain a continuously increased number. More concretely, assuming the CCTV system will record the internal vision of Room 1 every minute and save the video with the name contains the recording time (see Table 2.). A possible tampering method could be achieved by the attacker: firstly delete the record file in a previous time which contains criminal evidence. Then the system clock will be modified back to that time and the new automatically generated record will have the same name as the deleted one. At last, the system clock will be recovered to the normal status. However, as the example shown in Tab. 2, the record file name started with a continuously increased number. The file name of the new generated record starting with

250 is different with the illegally deleted one (starting with 100). Thus by checking the prefix number, the SIEM system can discover whether some records are removed illegally.

Record 1 (deleted)	Record 2 (new generated)
100_Room1_2016-06-15_12:38	250_Room1_2016-06-15_12:38

Tab. 2: An Example of Naming Convention for CCTV Records [Mo96]

Additionally, the tampering resistance can be addressed by applying correspondent storage security controls. The international standard [II15] demonstrates security controls as well as guidelines for storage security. Since in a SCADA system, the SIEM relevant raw logs are normally implemented as file-based storage, this standard is applicable for the security considerations of protecting them against tampering. For example, if the raw log is stored within the sub-component, the security controls of Direct Attached Storage (DAS) can be applied. More effectively, considering the data collection scenarios in a SIEM system, the security controls of Storage networking could address the network issues within the context of this paper. More important according to the guidelines provided by the standard, the backup of logs enable SIEM systems to detect tampering behaviors by comparing the on-line log and its backups. Meanwhile, it is also possible to recover logs after deconstruction attacks thus keep the evidence for further forensic investigations too. The security controls of Storage security services can be also applied for the consideration discussed in the next section (4.3).

4.3 Policy on Secure Evaluation and Reporting of SIEM Data

In general, the alerting/reporting of security incidents should comply with national rules as well as business domain specific collection of reports and finally towards fulfilling the requirements of BSI in Germany. In the case of that the IRT is separated from the data owner, authority processes need to be designed and agreed. Besides, regular security analytic reporting could support the daily work of IRT and empower their capabilities.

For reducing the workload of IRT, the alerting/reporting should be graded into different criticalities thus with different priorities. Meanwhile, the IRT staffs have the option to access different level of incident details when required. In some cases, the tracing back function can even navigate to the relevant raw logs. At last, correspondent to real-time systems, the announcement has to be made in time. Similar to the divisions of criticalities/priorities, SEs need to be grouped and analyzed according to their timestamps as well as time relevant invalidation constraints. The analyzing time of a group of security-related events should be predictable and ideally can be executed in different granularities which achieve a balance between execution time and analyzing accuracy.

Considering the confidentiality of the collected data, similar to the privacy-aware consideration addressed in [LMW10], the central SIEM service should try to keep data

inside and apply strong controls for data going outside which means locating in the same facility of the managed SCADA system. When a remote collection is required, transferred data need to be encrypted or obfuscated. Besides, the created security-related events as well as correlations/reports have to be controlled by the associated knowledge management system and/or asset management system. Practically, the data retention time needs to be considered and should cooperate with the requirements of forensic investigations.

4.4 Policy on Real-time Monitoring with the SIEM System

According to [Ie15], on one hand, the SIEM system should be able to monitor residual risks identified in the risk assessment approach. On the other hand, the effectiveness of security controls could be measured by the SIEM system.

For mitigating risks identified in the risk assessment, security controls will be deployed. However, in some cases, the criticality of residual risks is reduced but not eliminated. When deployed security controls are compromised too, these risks might still cause critical damage to the SCADA system. Fortunately, the SIEM system can react to knowing risks. For example, within the system abnormally increased network traffic is detected (Fig. 2). With previous risk assessment, the traffic could be with high possibility caused by an attacker who joined the network. Thus reacting to this residual risk, a scanning of the network as well as re-authentications can be performed by the SIEM system automatically or by the IRT staff manually.

Besides the residual risks, the SIEM system can monitor the involved security controls too. Taking the same example of the increased network traffic, scanning network discovers no additional device thus the warning will be suppressed. However, the control of the network traffic is triggered too frequently, which might indicate an unknown type of security breaches or a system specific false positive constraint is triggered. In the latter case, smart testing could be deployed to determine the possible reasons.

4.5 Example of SCADA network architecture with SIEM Evaluation

As part of the concept Security by Design, the SIEM system can be integrated into the SCADA system during the design phase. In this section, a conceptual design of the SIEM system is proposed. As different security zones are shown in Fig. 1: The high security zone on the left side is protected by a data-diode which is physical unidirectional. In this case, only the traffic from the high security zone towards the SIEM service is allowed. The data-diode blocks attacks that send manipulated messages from the compromised SIEM system to the critical control systems, like sensors and actuators. Meanwhile, the data-diode can only provide limited bandwidth for collecting SIEM specific data. The middle security zone in the center of Fig. 1 contains the access control equipment, like authentication devices and CCTV monitors. This security zone is separated from the SIEM service by a configured firewall. On one side, the firewall

protects the devices in the middle security zones from tampering actions through the network. On the other side, through the firewall, high-speed connections as well as reliable protocols, like TCP/IP, are still applicable. At last, on the right side, the low security zone is connected directly with the SIEM service which provides little protection while the maximum connections speed.

Through the heterogeneous kinds of connections, security-related events and data are collected to the central SIEM service. Here after normalizations/aggregations, uniformed events can be correlated to discover security breaches or to measure security controls. The correlation rules of events are defined based on the overall security model thus enable a model-driven SIEM system. Section 5 discusses several policy-based correlation rules and their associated constraints. These rules are policy-based thus can be defined without knowing implementation details of sub-components. They are defined and parameterized in the central SIEM service complying with correspondent security controls.

The correlation results of security-related events will be automatically examined by the SIEM service according to their associated constraints. Possible security breaches or statistic info of deployed security controls will be reported to the IRT. Especially, security warnings are pushed to the IRT staffs for further reactions. The IRT staffs can monitor the automatic processing of the SIEM service as well as directly take reactions to the system for ceasing ongoing attacks.

5 Attempts for Modelling Policy-based Correlation Rules

As the example mentioned in Section 4.5, the SIEM system needs to be able to reduce the workload of IRT thus automatic processing is required. The overall security model described in [Wa16] contains both security objects and associated security controls. By linking SIEM service to this model, it is possible to correlate/analyze security-related events automatically in a model-driven way.

For example, potential security breaches in the “logon after leaving” example (Fig. 2b) could be discovered by proper predefined configurations/rules: In the example, the associated security object is the employee appeared in the two events. The security object as an employee owns several attributes: e.g. title and location. A security event might *change* or *indicate* these attributes. On one hand, the “leaving” event changes the employee’s position outside the company. On the other hand, the “logon” event indicates her/his location (near the server) inside the company. Sorting the two events according to their timestamps, the conflict of the employee’s locations will reveal the possible security breach. Furthermore, when after the “leaving” event, an extra “entering” event occurred that changed the employee’s position again inside the company before the “logon” event. Therefore, there will be no conflict on the location attribute and no warning should be sent to IRT staffs.

Further extending the same example, what does that mean when we found 500 pairs of "entering" and "leaving" events between the first "leaving" event and the "logon" event (within 30 minutes)? Similar to the example of too many login attempts (e.g. 500 times) on a workstation in a short time (e.g. 10 minutes), the high frequency indicates possible security breaches. For discovering this kind of attacks, the resource concept needs to be introduced into the overall security model. A security object has always limited resources within a given time period. In this example, we can define one person can be associated up to 5 SEs during 30 minutes. Thus the real-time monitoring of the SIEM system will announce the security staff a warning since the resource (count of 5 SEs) was consumed up in a while. Similarly the resource (number of login attempts) could also be assigned to a server. Consuming up of the resource in a very short time might indicate a denial-of-service attack.

Considering the "configuration after working hour" example, after the regular working hours, the resource of allowed configuration changes will be shifted to 0. Thus any action could be counted as a configuration change, like setting parameters or patching software, will trigger a security warning.

6 Conclusions and Challenges

In this paper we discussed the common required features of a SIEM system and described its working approach by splitting it into different steps according to *the* preliminary work at ISO/IEC JTC1/SC27 WG4. Furthermore, a conceptual design of a SIEM system as well as associated network architecture is proposed. By linking the SIEM system to the overall security model of SCADA systems, a model-driven monitoring/reporting approach could be created to monitor system's residual risks while measure effectiveness of deployed security controls. Considering the context of SCADA systems, the SIEM system needs to handle security-related events from different sources without accurate universal timestamps while needs to take archiving raw logs as well as generated events into account for later on tracing back or forensic investigations.

However, creating a SIEM system for a SCADA system still faces challenges. The raw logs of system's sub-components are in different formats and probably do not share a universal clock. The correlation rules of security-related events can be complex while the system needs to react in real-time. At last, improper correlation rules and their constraints might lead to false positive alerts. These alerts will be generated especially when operation mode is changed (e.g. from normal mode to the maintenance mode) and associated rules as well as constraints are not adapted. This might result in that the IRT chooses to ignore or disable the feature.

According to the required security levels (SLs) it may be worth to spend effort in the standardization of logging format, especially for the sub-components with high SLs. This will help to avoid false positives as well as to support efficient evaluations for the most critical sub-components of a SCADA system.

Note: Some of the above described modelling-analyses are being elaborated as part of participation in the "SMARTEST" Cybersecurity Testing R&D with three German University partners, partially funded by German Ministry BMWi.

References

[Mo96] Mori, H.: The Perfect Insider, Kodansha, 1996

[LMW10] Lisovich, MA.; Mulligan, DK.; Wicker, SB.: Inferring Personal Information from Demand-Response Systems, IEEE Security & Privacy, vol. 8, no. 1, pp. 11-20, 2010

[MT15] Martyak, P.; Thow, M.: Enhancing Defense-in-Depth and Monitoring Programs to Protect Critical Digital Assets from Tampering, 9th Nuclear Plant Instrumentation and Control, Charlotte, 2015.

[St10] Stocks, W.: When Anti-virus Doesn't Cut It: Catching Malware with SIEM, 24th Large Installation System Administration, San Joes, 2010.

[Wa16] Waedt, K.; Parekh, M.; Tong, X; Gao, Y.; Ding, Y.; Xie, X.: Nuclear Safety and Risk based Cybersecurity Testing, 47th Annual Meeting on Nuclear Technology, Hamburg, 2016.

[Bu16] Bundesministerium für Wirtschaft und Energie: IT-Sicherheit für die Industrie 4.0, 2016.

[Ie13] IEC 62443-3-3: Industrial Communication Networks–Network and System Security – Part 3-3: System Security Requirements and Security Levels, 2013.

[Ie15] IEC 62443-4-2 [Draft]: Security for industrial automation and control systems – Part 4- 2: Technical security requirements for IACS components, 2015.

[II11] ISO/IEC 27035: Information technology — Security techniques — Information Security Incident management, 2015.

[II13] ISO/IEC 27002: Information technology — Security techniques — Code of Practice for Information Security Controls, 2013.

[II15] ISO/IEC 27040: Information technology — Security techniques — Storage security, 2015.

[Ni11] NIST SP 800-82: Guide to Industrial Control Systems Security, 06.2011.

[Ib16] IBM QRadar, www-03.ibm.com/software/products/en/qradar-siem, Accessed: 2016-05-30.

[Sp16] SPLUNK, www.splunk.com, Stand: 30.05.2016.