energies

MDPI

Article

Integrated Approach to Diagnostics of Failures and Cyber-Attacks in Industrial Control Systems

Michał Syfert, Andrzej Ordys *, Jan Maciej Kościelny, Paweł Wnuk, Jakub Możaryn and Krzysztof Kukiełka

Institute of Automatic Control and Robotics, Faculty of Mechatronics, Warsaw University of Technology, ul. Św. A. Boboli 8, 02-525 Warsaw, Poland
* Correspondence: andrzej.ordys@pw.edu.pl

Abstract: This paper is concerned with the issue of the diagnostics of process faults and the detection of cyber-attacks in industrial control systems. This problem is of significant importance to energy production and distribution, which, being part of critical infrastructure, is usually equipped with process diagnostics and, at the same time, is often subject to cyber-attacks. A commonly used approach would be to separate the two types of anomalies. The detection of process faults would be handled by a control team, often with a help of dedicated diagnostic tools, whereas the detection of cyber-attacks would be handled by an information technology team. In this article, it is postulated here that the two can be usefully merged together into one, comprehensive, anomaly detection system. For this purpose, firstly, the main types of cyber-attacks and the main methods of detecting cyber-attacks are being reviewed. Subsequently, in the analogy to "process fault"—a term well established in process diagnostics—the term "cyber-fault" is introduced. Within this context a cyber-attack is considered as a vector containing a number of cyber-faults. Next, it is explained how methods used in process diagnostics for fault detection and isolation can be applied to the detection of cyber-attacks and, in some cases, also to isolation of the components of such attacks, i.e., cyber-faults. A laboratory stand and a simulator have been developed to test the proposed approach. Some test results are presented, demonstrating that, similarly to equipment/process faults, residua can be established and cyber-faults can be identified based on the mismatch between the real data from the system and the outputs of the simulation model.

Keywords: failure detection; cyber-attack detection; cyber-attacks isolation

Citation: Syfert, M.; Ordys, A.; Kościelny, J.M.; Wnuk, P.; Możaryn, J.; Kukiełka, K. Integrated Approach to Diagnostics of Failures and Cyber-Attacks in Industrial Control Systems. *Energies* **2022**, *15*, 6212. https://doi.org/10.3390/en15176212

Academic Editor: Oscar Barambones

Received: 24 June 2022
Accepted: 18 August 2022
Published: 26 August 2022

1. Introduction

Each industrial installation is at risk of damage or destruction. The reasons may differ, ranging from spontaneous failures, errors and damage to control systems, incorrect actions of operators, and sabotage actions or cyber-attacks.

In the case of faults, the advanced techniques of automatic fault diagnosis have been developed for years. This is due to the disadvantages of commonly used alarm systems, which are the simplest diagnostic systems, as well as the advantages of advanced diagnostics, which contribute to increasing the efficiency and safety of processes. In the case of critical facilities, a properly functioning diagnostic system may prevent accidents that may threaten human health or life, as well as the natural environment.

Fault detection can be carried out using different types of models: analytical [1–3], neural [4–6], fuzzy [7], and statistical [8,9]. A among them, techniques based on partial parametric models identified in the fault-free state deserve special attention. In such cases, there are two options for obtaining knowledge about the impact of faults on residuals: using expert knowledge [1–3] or learning [5]. The learning technique requires the acquisition of experimental data not only for the normal state but also for all faulty states that are to be recognized. In practice, this is often impossible, because some faults are very rare.

The fault isolation methods that use expert knowledge to design the relationship between diagnostic signals resulting from the assessment of residuals and faults are of fundamental importance and will be addressed in this paper.

Currently, in contemporary control systems, the areas of the operational technology and the information technology are considered. The first of these, Operational Technology (OT) is hardware and software that detects or causes a change, through the direct monitoring and/or control of industrial equipment, assets, processes, and events. The second area, Information Technology (IT), is the use of computers to create, process, store, retrieve, and exchange all kinds of data and information in order to increase a company's revenue by optimization of business operations. It should be emphasized that OT has become established to demonstrate the technological and functional differences from traditional IT systems and Industrial Control Systems (ICS).

As one can see, there is a clear difference between the IT and OT areas. IT deals with information, its flow and administration, while OT manages the operation of physical processes and the machines used to implement them. Before the Internet era, the boundary between IT and OT was noticeable. Traditionally, OT systems have not been connected to an external network or digital technology. However, recently, these boundaries began to blur. Companies began the process of implementing solutions in their networks aimed at automation improvements by adding smart devices to increase production efficiency and data availability.

Thus, IT and OT systems currently operate in parallel and monitor and regulate the course of industrial processes through data exchange. The technologies being developed and implemented have the potential to enable connectivity between every device in the office and industrial workplace in order to increase the availability of OT components while collecting and analyzing data about them. Currently, two-way data exchange between IT and OT networks is called IT–OT convergence [10].

The convergence of OT and IT technologies, in addition to emerging opportunities, brings many challenges. Large data flows, also outside company networks, result in many threats related to the security of their operations. Therefore, understanding these risks and the impact of industrial network security incidents is one of the industry sector's main challenges. Production plants and all industrial devices communicate in real-time, and safety and security become a priority in this case. In this situation, standard solutions are not enough to meet the industry's requirements, and it is necessary to develop a new generation of diagnostic systems that take into account threats from cyber-attacks. Therefore, this article proposes implementing such a system for the OT area.

Important and current research directions are the development of Cyber-Physical Systems (CPS) [11,12] and the industrial transformation—Industry 4.0 [13]—in which CPS, next to cloud computing technology and big data, are a key element of the new strategy for industrial production. New technological solutions have many advantages but also contribute to the emergence of new threats [11,14].

The openness of the systems, new communication possibilities, including Industrial Internet of Things (IIoT) and wireless technologies, and the trend of integration of the communication of individual system elements lead to the growing threat of cyber-attacks [11,14–16]. Examples of such attacks are given below:

(a) One of the most famous examples of an attack on a critical installation was Stuxnet [17–19]. It was aimed at industrial control system—OT. The creators of malware had detailed knowledge of the target. Not only about a structure of industrial installation but also about the type of installed equipment. The Stuxnet was not an isolate case. There was a series of OT based cyber-attacks [19]: Havex (2013), Irongate (2014), BlackEnergy 3 (2015), Crashoverride (2016), Trisis/Triton (2017).

(b) The Dragonfly group utilized Havex malware that targeted many industrial control systems with emphasis on the energy, pharmaceutical, defense, and petrochemical industries. Researchers estimates that malware was put on over 2000 sites [20]. That malware used for its attack the Open Platform Communications (OPC) protocol, which

complied with the accepted standard, so its implementation did not differ between industrial plants. The operation of the HAVEX malware was to map industrial hardware and devices on the ICS network. The campaign created by the Dragonfly group focused on espionage activities. There was no physical damage to industrial installations in the course of the attacks. However, the data obtained during the attacks could be used to plan future attacks.

(c) During the Irongate attack, a German steel mill was successfully hacked [21]. Hackers gained access to the plant's network and then blocked the possibility of turning off the furnace. This led to damage to many components of the industrial plant and to the eventual shutdown of production.

(d) BlackEnergy 3 and CRASHOVERRIDE malware have caused power outages in Ukraine [22,23]. The first attack on the power grid took place on 23 December 2015. It caused the power to remain off for 6 h and caused difficulties in accessing the network for 225,000 customers. Malicious software made it possible to take control of SCADA systems and to shut down power substations. In 2016, a CRASHOVERRIDE attack based on the same mechanism led to an hourly power cut in Kiev.

(e) In 2017, the SIS system of a petrochemical plant in Saudi Arabia was attacked with the TRISIS malware [22,24]. It was a very powerful and advanced software designed to attack a specific system of security instruments (SIS). The attackers reprogrammed the SIS system driver in order to trigger an emergency and, as a result, to shut down the installation.

Threats related to damages and human errors or intentional external and internal attacks are particularly dangerous for critical infrastructures, such as power plants, water supply systems, power grids, chemical industry, etc. [16]. Despite various reasons, the effects of serious damage and attacks may be the same: e.g., fire, explosion, environmental contamination, the destruction of the installation, or process halt. Therefore, one of the main aims of this work is to demonstrate the need to develop an integrated approach to ensuring the safety of ICS. It should include [25,26]:

- technical safety—the problem of preventing serious industrial failures caused by the unreliability of technological installation components (e.g., pipeline cracks), damage to control system components, and human errors;
- security—the issue of protection against intentional unfriendly attacks from the outside (e.g., hacking attacks on control systems) and sabotage actions carried out from the inside.

For software systems, there are some works related to combined treatment of both safety and security issues—a good overview of actual state of research can be found in [27].

The postulate of a comprehensive security strategy is formulated in the works [25,28], with a special attention to safety and security co-assurance [29], but in practice, it has not been fully implemented so far, as evidenced by:

- separate standards for functional safety and information protection in ICS, including separate Safety Integrity Level (SIL) and Evaluation Assurance Level (EAL) specified in these standards;
- separate departments related to safety and security in industrial enterprises;
- Independent Intrusion Detection System (IDS) and Fault Detection and Isolation (FDI) diagnostics.

The development of methodology and tools for an integrated approach to detecting and isolating anomalies (process damage and cyber-attacks) has a high practical rank. It is a current challenge for scientific research. The importance of these issues is particularly high in the case of critical installations with the risk of a serious industrial failure, such as energy production and distribution.

The main contribution of this study is to show that by using partial models of the diagnosed system for a normal state of operation, and expert knowledge in the form of Fault Isolation System (FIS) containing the relationship between faults and the values of

diagnostic signals, it is possible to detect and, to some extent, also isolate various anomalies. It has been noticed that the structure of a cyber-attack, determined by its scenario, is complex. In order to enable a uniform approach to the diagnosis of various anomalies, elementary influences of a cyber-attack were distinguished, describing them as cyber-faults. A cyber-attack consists of a collection of cyber-faults. This approach makes it possible to use a uniform description of the relationship between anomalies and their symptoms.

The layout of the work is as follows: The second section presents the research problem, indicating potential threats to the safety of industrial processes. Section 3 is devoted to an overview of techniques for detecting cyber-attacks. Section 4 proposes an integrated solution for faults and cyber-attacks detection. Section 5 is devoted to the implementation of the proposed solution for an example object, Section 6 indicates further research directions. The article ends with Section 7, which is a summary of the work.

2. Problem Formulation, Research Method

The proper functioning of ICS is exposed to two types of dangers:

- Hazards related to damage and human errors not related to cyber-attacks,
- Threats related to destructive targeted activities, such as cyber-attacks and sabotage actions.

Therefore, ensuring an increase in ICS security in the event of threats related to damage to technological equipment, measurement, and automation devices and human errors and destructive targeted actions, such as cyber-attacks and sabotage actions is an important problem. The current solutions are not satisfactory. They relate to specific threats, and there are no holistic solutions that comprehensively address security issues.

The structure of the security systems for both hazards and threats is layered. The standard EN 61511 for functional safety for the process industry defines the Layers of Protection (LoP). The Defense in Depth strategy developed by the National Security Agency (NSA) at the US Department of Defense proposes a hybrid, multi-level approach to security. It consists of six steps: Security Plan, Network Separation, Perimeter Protection, Network Segmentation, Device Hardening, and Monitoring and Update. In the case of physical attacks and cyber-attacks, the security layers are referred to as Ring of Protection—RoP. The principle in the layered approach is: the more security layers, the higher the level of risk reduction. However, there is no integrated approach to different types of threats.

The protection of the digital part of the control system is achieved mainly by using demilitarized zones (DMZ), firewalls, data encryption, virtual private networks VPN (IPsec), network segmentation, identity verification, access authorization, and password management, among others. An example solution is given in [30]. The detection of cyber-attacks is carried out by monitoring network traffic, which does not allow the detection of all anomalies [31]. The solutions used in IT systems do not guarantee that an attack cannot get into the control system, and thus do not guarantee the elimination of its destructive influence on the controlled process [32].

Safety Instrumented Systems (SIS), which implement the algorithms of interlocks and automatic protections, are of key importance in the structure of process security. However, the operation of these systems stops all or part of the process, resulting in economic losses. Therefore, it is advisable to use solutions that can guarantee the elimination of threats at an early stage and thus prevent the SIS systems from operating and the process from being shut down. The proposed solution to the problem is the early detection of various anomalies and their accurate isolation [14,31]. Many studies have shown that detection methods are sensitive to the types of anomalies mentioned above.

However, the following research and technical problems emerge:

- Is it possible to develop the signatures for cyber-attacks?
- Is it possible to develop an inference method that allows the isolation (differentiation) of these various anomalies?
- Is it possible to integrate detecting and recognizing cyber-threats and process faults?

Solving these problems is of high practical importance because effective security measures can only be taken when the threats are known. Therefore, for an effective response to emerging hazards and threats, the early detection of anomalies and their complete isolation are necessary.

This study shows that:

- the use of model-based detection methods is an effective method of detecting various types of anomalies, both process damages and cyber-attacks;
- it is possible, in some cases, to develop signatures for cyber-attacks;
- the use of diagnostic inference methods makes it possible to distinguish between damage and cyber-attacks, as long as their signatures are known and are different;
- it is possible to integrate the tasks of diagnosing process damages and cyber-attacks into the diagnostic system.

A case study was used as the research method. The possibilities of solving the formulated research problems were analyzed on the example of the system of faults and attacks diagnostics for a set of tanks with the control system and the SIS system.

It was assumed that in the case of industrial processes, and especially critical objects, it is impossible to obtain training data for object states with existing damages and cyber-attacks. It eliminates the usefulness of all anomaly detection and isolation approaches that use such data, e.g., classification methods to identify anomalies. Therefore, for the detection of anomalies, fuzzy partial models representing the normal state of the diagnosed object were used. Such models have already been used for damage detection. In this study, their usefulness for detecting various anomalies was examined. Their signatures were developed for the specified damage and cyber-attacks, the detection and distinguishing of these anomalies were analyzed, and simulation and laboratory tests were carried out.

Note that there is an extensive literature related to security of data, i.e., how to prevent eavesdropping/stealing of information; sometimes called a "passive attack" (e.g., [33,34]). The method proposed in this paper relies on the deviations of process variables from their normal operation routine. Hence, it can only be applied if an attack is causing such changes, i.e., it is an "active attack", rather than a "passive attack". It could be argued that the IT domain is concerned with data security, whereas the OT domain is more concerned with equipment security [14]. Hence, IT methods may be better tuned towards "passive attacks". Nevertheless, it should be stressed that, in some circumstances, the solution used in IT to detect and prevent cyber-attacks may alone be sufficient. For instance, to detect so called replay attacks an authenticated encryption with freshness protection is often applied successfully. Nevertheless, as the defense methods became more advanced, the cyber-attacks also became more sophisticated. In this work we postulate that the knowledge of the process characteristics and observing deviations from these characteristics provides additional information, which may be rather difficult to acquire by an attacker and may therefore, combined with other methods, improve the chances of detecting an attack. The methods proposed in this article to detect attacks are also considered as the last layer of security that can be used when the attacker gets through the classic security characteristic for IT systems. The use of such algorithms in no way excludes the possibility and need to apply security in the IT layer.

3. Main Directions of Research towards Recognition of Cyber-Attacks

Fault diagnostics methods have been developed for about 40 years now. Hence, the knowledge in this field is already well-established. Many monographs (e.g., [35–40]), review articles (e.g., [41–43]) and a lot of works on specific issues have been written. Various diagnostic systems for industrial processes have been developed (e.g., [38,44–46]).

Research on the development of methods for diagnosing cyber-attacks started much later, and therefore it is less advanced. The proof of this is the lack of books that would comprehensively cover the issues of detection and recognition of cyber-attacks. However, synthetic and review works [27,32,47–52] on this subject and many articles on specific issues have been published.

The problem of detecting cyber threats can be considered already at the stage of systems analysis. In [53] the new approach, STPASec was proposed. STPASec is an extension of the System-Theoretic Process Analysis (STPA) [54] with security considerations. In described approach, systems are modelled as hierarchical structures. Higher layers are responsible for control processes, via so called actors placed at lower layers. In the opposite direction, sensors (placed on lower layers) send feedback to the controllers. The list of scenarios which can lead to losses is usually a result of STPASec analysis. Wide overview of this technique can be found in [55].

The papers [35,50,56] present the classifications of cyber-attacks in network control systems (NCS) and in CPS. The attack vector and the three-dimensional attack space have been defined, the axes of which are: disruption resources, disclosure resources, and system knowledge. In this space, the place of the distinguished types of attack has been defined: denial of service (DoS) attack, replay attack, bias injection attack (false data injection attacks), zero dynamic attack, and covert attack. Models of particular types of attacks have been given. The approach proposed in the above works, which focuses on control issues, allows the assessment of the effects of attacks and the analysis of methods of their detection, as well as the design of attack-resistant structures. The publication [47] has additionally defined the "impact space" in which the above types of attacks were located. In the review [32] only three types of attacks are considered: DoS attack, replay attack, and deception attack. Deception attacks are also known as false data-injection attacks or malicious attacks. It should be emphasized that it is usually not necessary to recognize the type of attack. It is important to detect the existence of an attack in order to take appropriate security measures [57].

In the field of cyber-attacks' diagnostics, the following research directions can be distinguished:

- Detection based on traffic supervision in industrial networks;
- Detection based on the analysis of process data and indicators characterizing the operation of control circuits;
- Isolation (identification) of cyber-attacks.

Techniques for detecting intrusion prior to cyber-attacks are mainly divided into two categories: based on abuses (signatures) and anomalies. Techniques based on signatures require a precise definition of the functioning of the system, i.e., its model, in the event of a given type of attack. Anomaly-based attack detection techniques use the definition of normal system functioning and detect deviations from normal behavior as deliberate attacks or unintentional errors [58]. This approach is analogous to that in the case of fault diagnostics. To detect specific states with faults, process models representing this faulty state could also be used. However, this approach is rarely applied. The dominant solution is anomaly detection with the use of normal state models. The deviations of residues from the zero value are symptoms of faults.

In [58], three types of anomaly detection approaches were distinguished:

- detection based on network traffic exploration;
- detection based on the analysis of network protocols;
- detection based on the analysis of process data.

Attacks on ICS often cause unusual network traffic or violate network protocol specifications. Moreover, they influence the functioning of the control systems, and thus the operation of the controlled process. The first two solutions are related to the supervision of network traffic. This approach, even when it comes to industrial networks, is derived from methods used in IT systems. Specialized IDS tools dedicated to OT are already offered on the market. Machine learning based approaches are still being explored, such as the hybrid system for distinguishing between attacks with known signatures, attacks with unknown signatures, and normal network traffic described in [59]. The third solution, on the other hand, is specific to control systems. It comprises the early detection of discrepancies between the current and reference functioning, represented by models characterizing the

normal state of the controlled process [35–38]. The detection scheme presented in [58,60] is identical to the fault detection scheme used for a long time in FDI methods.

Detection methods, especially those derived from approaches known in IT networks, have already reached a high level of advancement, as measured by a large number of publications and the offered IDS systems using these techniques [61]. The solutions derived from the approaches developed in the automation environment, which include intrusion detection based on the analysis of process data and the supervision of the control loop, have appeared much later and are less well documented. On the other hand, there are very few works on the isolation (identification) of cyber-attacks and their differentiation from damage that may cause similar symptoms. The state of research in all of the above-mentioned areas is described below.

3.1. Detection Based on Traffic Supervision in Industrial Networks

3.1.1. Detection by Exploring Network Traffic

Under normal conditions the control systems are characterized by stable network traffic patterns. Fluctuations in network traffic typically indicate a change in ICS state, which enables intrusion detection based on network traffic exploration. These methods use data mining technologies to identify unusual system behavior. Commonly used motion mining techniques include: supervised clustering [42], partially supervised clustering [43], mixed Gauss model [44], neural network [45,62], fuzzy logic [46,47], single-class support vector machine [48], the multiclass support vector machine [49], and deep learning [50].

3.1.2. Detection Based on the Analysis of Network Protocols

The protocol specifications define the packet formats and communication modes allowed by the protocol. This method of attack detection consists in checking whether the transmission packets in the industrial control network violate the specifications of the industrial protocol [58]. Detection is based on rules defined on the basis of protocol specifications.

3.2. Detection Based on the Analysis of Process Data and Indicators Characterizing the Operation of Control Circuits

The values of the industrial process variables should be consistent with the relationships describing the physical phenomena in the process. Damage and attacks cause changes to the functioning of the process. The symptoms of these changes can be detected on the basis of the observation of the discrepancy between the normal behavior of the process and the disturbed one. Thus, for symptom generation, models are needed that represent either the normal state or models in states with existing damage or attacks.

As in the case of fault diagnostics, one can distinguish between passive and active approaches to the detection of cyber-attacks. Passive methods are based on operating signals, while active methods require the introduction of an appropriate input signal and an analysis of its influence on the output signals of control systems. Active approaches were presented in [63–65]. In this paper, only passive approaches will be used.

In [32], the authors point out that attack detection plays a key role in maintaining performance of CPS. Classifying attack detection methods for industrial CPSs, they distinguish the following approaches:

- Bayesian detection with binary hypothesis;
- Weighted least square approaches;
- χ^2—detector based on Kalman filters;
- Quasi-FDI techniques.

In the opinion of the authors of this paper, the methods which evaluate the integrity of process data collected from measuring devices, actuators, and controllers will have the greatest practical significance. For this evaluation, methods developed on the basis of fault diagnostics, based on models linking process variables, can be used. Anomaly detection is based on the generation of residues, which are the difference between the value of the

process variable measured and calculated on the basis of the model representing the normal state of the process [58,60].

The advantage of the above approach is the possibility of using various models for the detection of anomalies: analytical (non-linear and linear) as well as models created on the basis of measurement data, e.g., neural, fuzzy, additive, etc. In the research on detection of cyber-attacks, a prevailing approach is linear state-space models [50,51,58,60]. For example, in [51,66], linear observers are used; in [67,68], Kalman filters; and in [69], a nonlinear interval observer. The use of neural models for the detection of anomalies was the subject of works [70,71].

For the detection of anomalies, not only quantitative but also qualitative models can be used. In the works [14,72], models in the form of rules were presented, enabling the detection of both damage and attacks that cause the feedback sign to change from negative to positive or to block the actuator.

Research on the detection of cyber-attacks usually focuses on a specific type of attack. The largest number of works is related to the detection of false data injection attacks [69,73–78]. The lower number of publications present studies on the methods of detecting replay attacks [78], covert attacks, and zero dynamics attacks [79].

Cyber-attacks could cause the detrimental changes of control performance indicators. Hence, the methods which are used for assessment of control loop performance can be applied to detection of not only faults but also other anomalies, such as cyber-attacks. These methods, sometimes referred to as "minimum-variance performance assessment" are mainly based on evaluation of stochastic properties of signals in the control loop.

A comprehensive report prepared by Arizona State University [80] indicates importance of considering stochastic characteristic of the process noise, as a useful indication of intrusions into the signals in control loops. Hence, the signal's parameters, such as correlation, distribution, and difference in means, can be tested to enhance the knowledge of anomalies and possible artificial injections of signals.

Article [76] proposes a technique based on the CUSUM (cumulative-sum) algorithm to defend cyber-attacks in power distribution networks. CUSUM algorithm applied here has its origins in Statistical Process Control (SPC), where it is commonly used for the on-line monitoring of the system's performance.

In [75], certain measures of distance of two probability distributions, derived from measurement variations and obtained from historical data, are explored. The assumption is that when false data are injected into the power systems, the probability distributions will deviate from those recorded previously.

The common motive in the above works is that additional level of knowledge to detect cyber-attacks can be obtained by analyzing the stochastic characteristics of signals, including the measures used for assessment of the performance of control loops in the process control system. This is the level of knowledge would be much more difficult to replicate by the attacker.

3.3. Isolation (Identification) of Cyber-Attacks

The detection of cyber-attacks based on network traffic monitoring or network protocol control allows detecting of many but not all attacks. Most IDS will be rather specific about the type of attack. If the system relies on remote attestation features, detection will also point to the exact integrity violation, e.g., the localization of the attack point. However, this is not always the case when using FDI techniques. Here, detection means that an attack has occurred but does not specify the type of attack. In many cases, this is sufficient to take appropriate protective measures. Signature-based anomaly detection techniques can accurately identify the type of attack. However, it is very difficult, in a general case, to provide exact definitions (models) for all potentially possible attack scenarios. In the case of anomaly-based detection techniques, the occurrence of a deviation from the normal state signals the occurrence of a symptom, the cause of which may be either various cyber-attacks, damage (faults), or other disturbances. Isolating these anomalies is a separate task.

Few publications deal with the isolation of anomalies. Ref. [51] characterizes fundamental monitoring limitations from system-theoretic and graph-theoretic perspectives and proposes centralized and distributed attack detection and identification monitors. Ref. [81] investigates the attack isolation and attack location problems for a CPS based on the combination of the H-infinity observer and the zonotope theory.

In [82], an algorithm is proposed enabling the correct estimation of the actual state of the process supervised by dispersed agents, such as sensor networks or power grids. The algorithm at each iteration collects information from local measurements, exchange this information with other nodes in neighborhood, and aggregates the received measurements using coordinate-wise trimmed means. Thanks to such techniques, it is possible to obtain the resilience of a multi-agent system despite the compromise of some of its nodes.

The anomaly isolation scheme (IOS—Iterative Observer Scheme) is proposed in [83]. The research concerned the Smart Grid system. The scheme is an extension of the known GOS scheme, which was used to isolate faults in individual sensors. The IOS scheme divides the set of sensors that provide measurements in the power grid system into subsets and iteratively performs subset tests. The method enables the isolation of damaged/attacked sensors. These types of anomalies are not distinguished.

In [84], the innovative secure state estimation framework was investigated which combines the IMM (Interactive Multiple Models) filter with a fuzzy-based attack isolation mechanism. A hybrid state model consisting of two behavior modes, one corresponding to the ideal scenario and one associated with the attack behavior mode are used. Three types of attacks were simulated: constant attacks; time-varying attacks; and stochastic attacks (possibly non-Gaussian).

The problem of simultaneous Cyber-Attacks and Faults Detection and Isolation (CAFDI) in CPS has not been fully analyzed in the literature [85].

In [85], a methodology based on the CPS two side filters and an Unknown Input Observer (UIO)-based detector is proposed. A linear time-invariant model is used, taking into account both the impact of damage and cyber-attacks. It is shown how to generate four residuals in order to detect an actuator cyber-attack, sensor cyber-attack, actuator damage, and sensor damage. Components' damages have not been taken into account. The conditions under which the proposed methodology can detect fraudulent attacks, such as stealth attacks, zero dynamics attacks and replay attacks that are considered undetectable [86], are characterized.

The basic method used to isolate cyber-attacks is UIO [85,87]. This method, which was developed for the needs of fault diagnostics has been presented in many publications, including [36]. This method consists of developing an observer sensitive to a specific damage/attack and at the same time making this observer independent (decoupling it) from other unknown inputs (damage/attacks and disturbances). However, the number of unknown inputs to which immunity can be obtained is limited. Therefore, this approach in industrial practice is not feasible if all possible damage and disturbances are to be taken into account, e.g., component damage level, which is often omitted in publications.

In the above works [51,81–85], failure detection is carried out on the basis of analytical models: modified Luenberger observer [51], H-infinity observers [81], Kalman Filter [83], interactive multiple model (IMM) filter [84], and UIO [85]. Linear models introduce significant limitations in the case of industrial applications because real objects are non-linear. Moreover, obtaining analytical models that take into account the impact of damage and cyber-attacks is very difficult and expensive, especially in the case of large-scale systems.

Therefore, research on the application of anomaly detection and isolation methods with the use of non-linear process models created on the basis of measured data is purposeful. The problem of adaptive error detection and isolation (FDI) for non-linear networked control systems in periodic denial of service (DoS) attacks was analyzed in [88]. A solution using a set of switching fuzzy T–S observers was proposed. The use of artificial intelligence models was also assumed in [14]. However, to record the sensitivity of the residues generated with the use of models to various anomalies, both the FSM (Fault Signature

Matrix) [35,36] and the FIS [38] used in fault diagnostics can be used. Their design does not require models that take into account the influence of anomalies, but the knowledge which resides with the designers of diagnostic systems.

4. The Concept of an Integrated Approach—A System for Diagnosis of Faults and Cyber-Attacks

In industrial practice, it is necessary to consider all possible process faults and cyber-attacks. The omission of certain types of anomalies during the analysis may generate false diagnoses in the event of their occurrence. Therefore, methods of attack detection and isolation based on process data cannot disregard the possibility of faults. Likewise, fault diagnosis systems must take into account cyber-attacks. Hence, it is purposeful to research solutions that:

- take into account the need to integrate the detection and isolation of cyber-attacks;
- use nonlinear analytical, fuzzy, neural, and other models to detect anomalies;
- use the knowledge of experts to determine the relationship between anomalies and the values of generated diagnostic signals.

This approach creates the greatest potential for industrial applications. This work proposes such an approach.

When conducting research reported in this paper, it was assumed that a good starting point would be the system based on the FIS [3].

An inference system defined in this way can be used both for the isolation of process failures and for cyber-attacks. Figure 1 shows the potential impact of both faults and cyber-attacks on an industrial installation. As can be seen—as both types of anomalies can manifest themselves with similar effects—the occurrence of discrepancies between the signals measured at individual points of the installation and their expected values obtained from the model/partial models. The only condition for the correct detection of cyber-attacks is the ability to define diagnostic signals that will be sensitive to them. Therefore, the set of diagnostic signals has been extended with new elements and the controller has been modelled. This device is rarely considered by classic industrial diagnostic systems (due to the advanced self-diagnosis usually available in devices of this type). On the other hand, an attack on the controller as such is possible and poses a real risk to the system. In order to generate the diagnostic signal, either the analytical model of the controller or the identified model can be used. Due to the fact that the analytical form of the PID controller is well known, in the example discussed in this paper an analytical model was chosen. The real and virtual controllers work in parallel, using the same deviation signal, and their outputs are compared. In the event of significant discrepancies, a diagnostic signal is generated.

The concept of a *cyber-fault* has been introduced. By analogy to spontaneous failure, a cyber-fault is the term used to define a malfunction of a separate component of a supervised process, caused by changes in its software or in data transmission. The cyber-fault should be distinguished from a *cyber-attack*. Typically, one cyber-attack can cause multiple cyber-faults. Cyber-fault is detected by the FIS in the same way as a process fault, i.e., each cyber-fault has an assigned signature, showing the values of diagnostic signals at the time of its occurrence.

Distinguishing between cyber-fault and process fault is a problem that requires more attention. In the classic FIS approach, two failures are distinguishable if they have different signatures; however, single fault scenarios are usually considered. In the case of cyber-attack, multiple cyber-fault scenarios must be usually considered. In such a case, defining the pattern values of diagnostic signals for particular cyber-attack scenarios is not a trivial task. As part of further research, FIS extensions are being proposed, taking into account the multivalent assessment of symptoms, the time of symptom appearance, or the sequence of their occurrence. Each of these techniques potentially makes it possible to distinguish between cyber-faults and process faults. For example, let us consider the sequence of symptoms. In the case of a physical fault in the system components, such a sequence is often determined by the physical phenomena occurring in the process—e.g., the clogging

of the liquid flow at the inlet of a tank, usually a symptom related to the flow model—will appear first, and after some time, a second symptom—a symptom related to the level model in the tank—will appear. In the case of a cyberattack, there is no such limitation—if the measurements are distorted, both symptoms can occur at the same time. However, the exact determination of the possibility of distinguishing cyber-attacks from physical damage using these techniques requires further research.

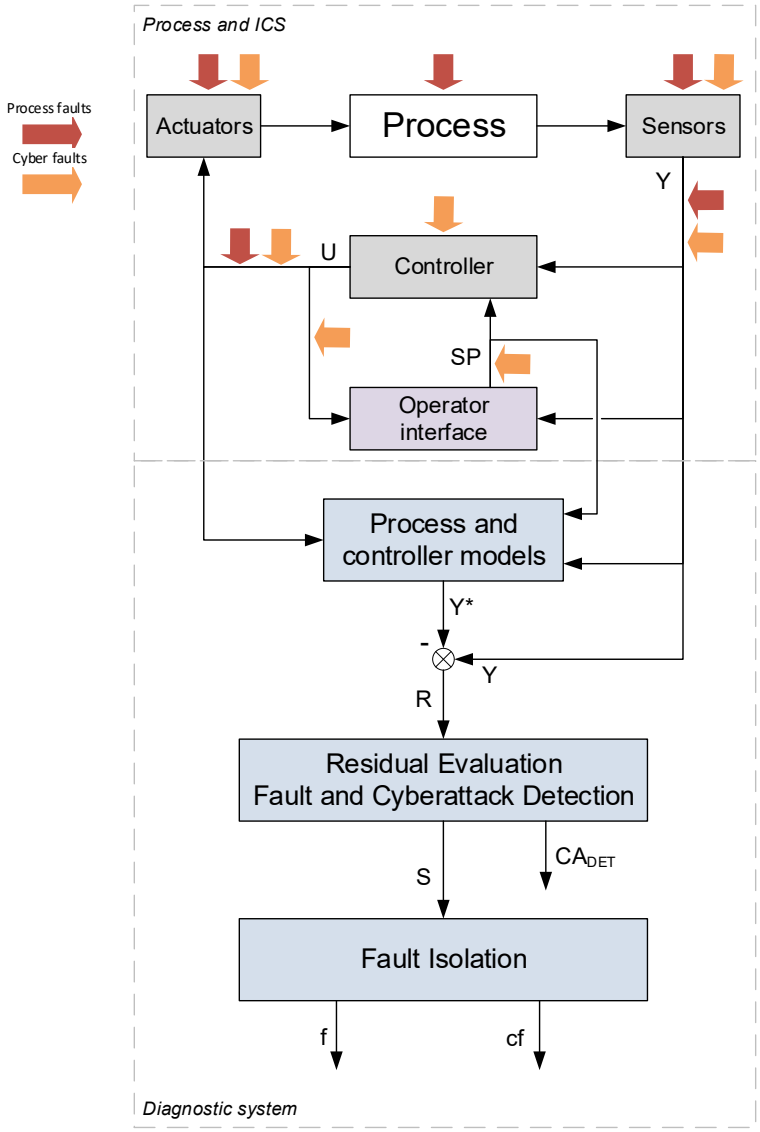

Figure 1. Faults and cyber-attacks and their influences on industrial systems.

It seems that the challenge for the proposed technique will be detecting replay-type attacks when, from the point of view of the diagnostic system, a record of the correct operation of the installation is observed and examined; therefore, no symptoms that may indicate a cyber-attack will be observed (the diagnostic system is blinded, similarly to the operator). In such a case, it is desirable to be able to use active diagnostics to impose an additional input on the control signals, the effects of which should be observable in the measured values of the process variables.

5. Research Results—Example of a Group of Tanks

5.1. Description of the Process and the Simulator

For the purpose of this research, a simulator of the two-tank system has been developed. It allows for the testing of the algorithms of the detection of the process faults and disturbances. In the proposed approach, the FDI methods can be used. Moreover, because of the simulator's ability to inject process faults, the issue of distinguishing between cyber-faults and process faults can be researched. The communication channels are also modelled, and these channels are the places where cyber-attacks can be injected.

The simulator was developed in MATLAB/Simulink (version R2021b by The MathWorks Inc.). The general block diagram of the simulator is presented in Figure 2.

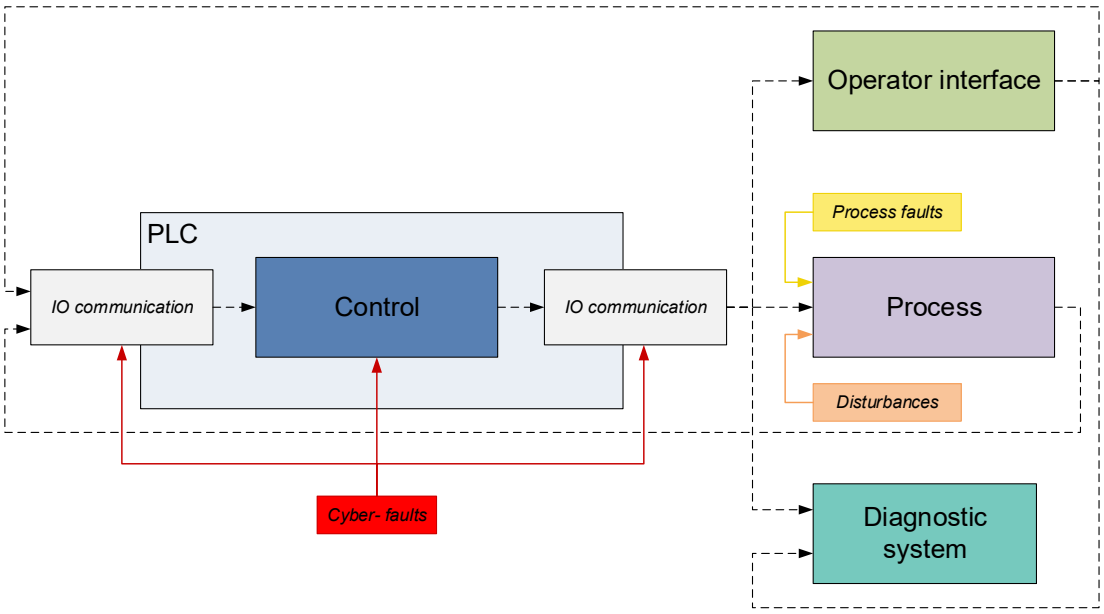

Figure 2. Block diagram of the simulator in the MATLAB/Simulink environment.

In the simulator, one can distinguish the following main subsystems:

- **Process.** A subsystem which represents components of the physical process together with measuring devices.
- **Control.** A subsystem representing the PLC controller, on which signals controlling the configuration are developed, and the control system is implemented.
- **Operator interface.** The subsystem corresponds to the HMI/SCADA system on which the operator interface is implemented.

The process part schematic diagram is given in Figure 3.

Depending on the configuration of valve states V1, V2, V3, Vd1, and Vd2, various properties of the control object can be realized (e.g., First Order Lag/Second Order Lag/Delay). The set of available process variables in the experimental stand is given in Table 1. For the purpose of this paper the configuration with two tanks and no delay in water supply line was considered.

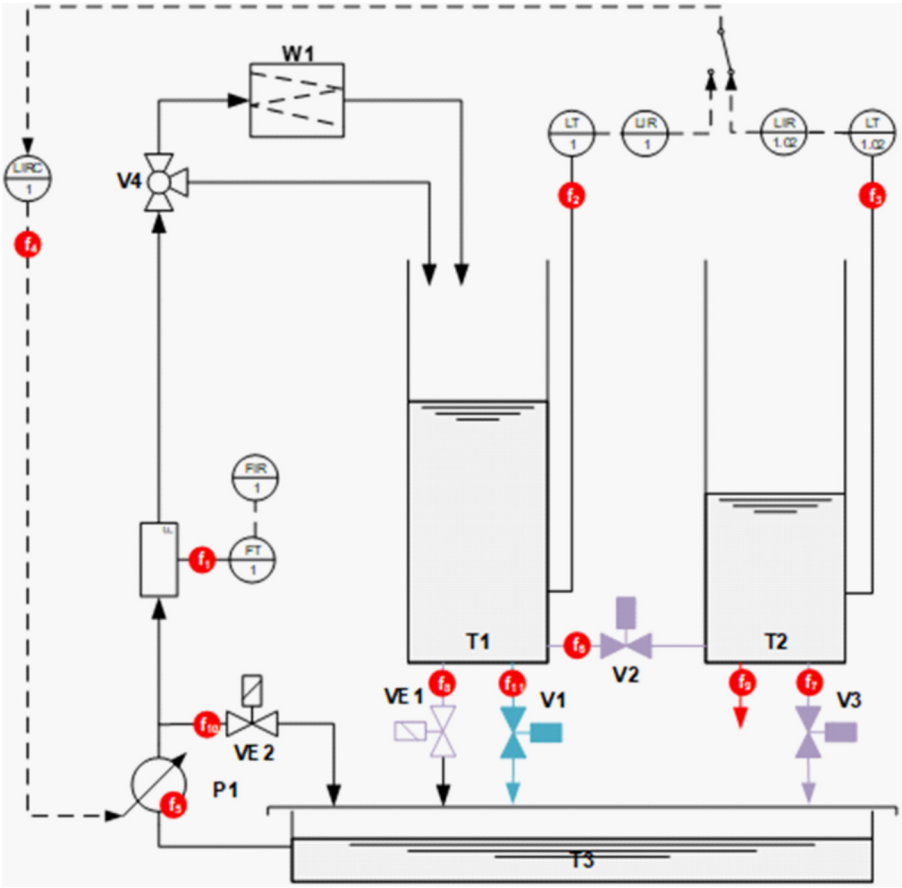

Figure 3. A schematic diagram of the laboratory test stand. Additionally, available measurements and the symbolic places of introduction of process faults are depicted.

Table 1. Set of available process variables.

Name	Description	Units	Range	Signal
SP_L	Tank level setpoint	%	$\langle 0, 100 \rangle$	-
CV_V	Control signal	%	$\langle 0, 100 \rangle$	$\langle 4, 20 \rangle$ [mA]
F_1	Water flow at the inlet to the tank I	L/min	$\langle 0, 6.5 \rangle$	$\langle 4, 20 \rangle$ [mA]
L_1	Tank level I	m	$\langle 0, 0.5 \rangle$	$\langle 4, 20 \rangle$ [mA]
L_2	Tank level II	m	$\langle 0, 0.5 \rangle$	$\langle 4, 20 \rangle$ [mA]
PV_L	Controlled level: 100 [%] = 0.374 [m] for L1 control 100 [%] = 0.374 [m] for L2 control	%	$\langle 0, 100 \rangle$	-

5.2. Faults and Cyber-Faults

Moreover, for the purpose of the FDI research, additional parametrizable blocks have been added to the simulator:

- **Disturbances.** Process disturbance generation subsystem. The simulator takes into account the possibility of simulating process disturbances in the form of fluctuations in the pump efficiency and outflow from tank II. The list of disturbances and their parameters are presented in Table 2.
- **Process faults.** Process fault generation subsystem. The list of faults and their parameters is presented in Table 3.
- **Cyber-faults.** Each cyber-attack is conducted according to the designed scenario—a specific attack method. Such a scenario defines independent sub-activities influencing individual system components and communication channels. Each activity is determined by the place of influence, its nature, and variability over time.

Table 2. The set of process faults in configuration for two tanks operation.

F	Description	Range
f_1	The fault in the measurement channel	$\langle -7.5, 7.5 \rangle$ [L/min]
f_2	The fault in the measurement channel	$\langle -100, 100 \rangle$ [%]
f_3	The fault in the measurement channel	$\langle -0.5, 0.5 \rangle$ [m]
f_4	The fault in the transmission of	$\langle -100, 100 \rangle$ [%]
f_5	The fault of the pump P (change in pump flow)	$\langle -100, 0 \rangle$ %
f_6	The obstruction in the pipe between tanks T1 and T2	$\langle 0, 100 \rangle$ %
f_7	The obstruction in the pipe out of tank T2	$\langle 0, 100 \rangle$ %
f_8	The leak from tank T1	$\langle 0, 100 \rangle$ %
f_9	The leak from tank T2	$\langle 0, 100 \rangle$ %
f_{10}	The leak from the inlet to tank T1	$\langle 0, 100 \rangle$ %

Table 3. The set of cyber-faults.

Name	Description
cf_F^C	Modification of the measurement signal F at the input to the control system
cf_F^{UI}	Modification of the F measuring signal at the input to the operator station
cf_{L1}^C	Modification of the L1 measuring signal at the input to the control system
cf_{L1}^{UI}	Modification of the L1 measuring signal at the input to the operator station
cf_{L2}^C	Modification of the L2 measuring signal at the input to the control system
cf_{L2}^{UI}	Modification of the L2 measuring signal at the input to the operator station
cf_{CV}^C	Modification of the CV control signal at the input to the process (actuator)
cf_{CV}^{UI}	Modification of the CV control signal at the input to the operator station
cf_{SP}^C	Modification of the SP set point at the input to the control system
cf_{PID}^{MODE}	Changing the controller's operating mode
cf_{PID}^{SET}	Modification of the controller settings

The vectors of setpoint values and configuration options, the control and configuration signals, and process variables are exchanged between subsystems. It is realized by blocks that symbolize specific communication standards (IO communication). The parameters of individual process components, including actuating and measuring devices, were selected experimentally to reflect the behavior of real components as faithfully as possible, including such elements as non-linearities, dynamics, and the amplitude of measurement noise.

Cyber-faults presented in Table 3 show almost all possible points of influence realized during a cyber-attack, consisting mainly in changing the values of signals transmitted between the components of the process, the control system and the operator's interface. Additionally, the possibility of influencing the control algorithm is considered—by changing the mode of operation and settings of the controller. The lower index of the cyber-fault denotes the affected signal or component, and the upper one denotes the affected signal path (C—signals to/from controller, UI—signals to/from operator interface) or the kind of performed action.

The considered cyber-faults, presented in Table 3, represent the effects of the changes introduced by the attacker, not the manner in which the change was introduced. In most real-life cyber-attacks, individual cyber-faults are caused by attacks on the software, as a result of which the transmitted values of control and measurement signals will be disturbed and the operating modes and parameters of the control algorithms will be changed.

Due to a large number of possible places of impact on the process by the attacker, as well as possible different time scenarios of such an impact, the number of possible attack scenarios is very large. In the analyzed simulator, the possibility of simulating six different scenarios, characterized by different ways of influencing the process, was designed. The list of scenarios along with the symbolic designation of the attack vector for a given scenario is shown in Table 4.

Table 4. The set of considered cyber-attack scenarios.

Name	Description	Vector of the Attack
cas_1^A	Cyclic playback of the recorded history of all available measurement and control values	$[cf_{CV}^P, cf_{CV}^{UI}, cf_F^C, cf_F^{UI}, cf_{L1}^C, cf_{L1}^{UI}, cf_{L2}^C, cf_{L2}^{UI}]$
cas_1^B	Cyclic playback of the recorded history of all available measurements and the values of the control signals transmitted to the operator's interface	$[cf_{CV}^{UI}, cf_F^C, cf_F^{UI}, cf_{L1}^C, cf_{L1}^{UI}, cf_{L2}^C, cf_{L2}^{UI}]$
cas_1^C	Cyclic playback of the recorded history of the controlled variable	$[cf_{L2}^C, cf_{L2}^{UI}]$
cas_2	Gradual modification of the controlled variable	$[cf_{L2}^C, cf_{L2}^{UI}]$
cas_3^A	Changing the controller's operating mode to "reverse"	$[cf_{PID}^{MODE}]$
cas_3^B	Modification of controller settings	$[cf_{PID}^{SET}]$
cas_3^C	Changing the controller's operating mode to "manual" and setting a specific control signal	$[cf_{PID}^{MODE}]$

5.3. Algorithms of Process Diagnostics

In the considered example, the detection of cyber-attacks has been carried out with the use of algorithms prepared for the detection and isolation of process faults.

In this respect, three residues using models reconstructing individual process variables have been used. The residuals are generated based on the following scheme:

$$r_j = PV_j - PV_j^*$$

where: PV_j denotes process variable used in j-th residual, and PV_j^* denotes value of process variable reconstructed by the model.

Then, after low-pass filtering, each of the residuals is subjected to a three-valued evaluation assuming the constant lower (LIM_j^L), and upper (LIM_j^H) limits of the j-th residual variability area in the state without faults and a dead band Δ_j:

$$s_j(\mathrm{k}) = \begin{cases} 1^+ : r_j > LIM_j^H + \Delta_j \\ 1^- : r_j < LIM_j^L - \Delta_j \\ 0 : LIM_j^L \le r_j \le LIM_j^H \\ \mathrm{s_j(k-1)} : otherwise \end{cases}$$

The set of residuals together with the diagnostic matrix for the purposes of the detection and localization of process faults is shown in Table 5. Thus, the diagnostic signals take one of the three values $s_j = \{1^-, 0, 1^+\}$. Two possible signatures have been shown in the diagnostic matrix for damages to the measurement and to the control signal transmission lines, i.e., for the case when the signal value is underestimated (↓) or overestimated (↑) in relation to the real value. The diagnostic matrix shows that there are two groups of damages that are not distinguishable conditionally or unconditionally: $\langle f_2, f_6 \rangle$ and $\langle f_3, f_5, f_{10} \rangle$. These have been highlighted in the table accordingly.

Table 5. The set of considered residuals and diagnostic matrix for process fault diagnosis.

Residual/Diagnosic Signal		$f_1\downarrow$	$f_1\uparrow$	$f_2\downarrow$	$f_2\uparrow$	$f_3\downarrow$	$f_3\uparrow$	$f_4\downarrow$	$f_4\uparrow$	f_5	f_6	f_7	f_8	f_9	f_{10}
$r_1 = F_1 - f(CV_V)$	s_1	1^-	1^+	0	0	0	0	1^-	1^+	1^-	0	0	0	0	1^-
$r_2 = L_1 - f(F_1)$	s_2	1^+	1^-	1^-	1^+	0	0	0	0	0	1^+	1^+	1^-	1^-	0
$r_4 = L_2 - f(L_1)$	s_4	0	0	1^+	1^-	1^-	1^+	0	0	0	1^-	1^+	0	1^-	0

The indicators of the presence of individual process faults are determined on the basis of rules corresponding to the signatures indicated in the diagnostic matrix for that fault by

comparing the current values of diagnostic signals with those indicated in the signatures, e.g., for f_1:

$$f_1 = \begin{cases} 1 : [(s_1 = 1^-) \wedge (s_2 = 1^+) \wedge 4 = 0)] \vee [(s_1 = 1^+) \wedge (s_2 = 1^-) \wedge (s_4 = 0)] \\ 0 : otherwise \end{cases}$$

5.4. Algorithms of Cyber-Attack Detection and Isolation

For the purpose of detecting cyber-attacks, the set of residues, which was earlier prepared for detection of process faults, has been extended by the fourth residuum using the controller's model:

$$r_5 = CV - f(e)$$

where: $e = PV_L - SP_L$.

In Table 6, a theoretical analysis of the sensitivity of diagnostic tests to single cyber-fault is presented. As can be seen, most of the cyber-faults are detectable and even hypothetically distinguishable. In this analysis, it has been assumed that the diagnostic system uses in its calculations the same signal values that are sent to the operator interface.

Table 6. Theoretical matrix of sensitivities of diagnostic checks to single cyber-faults.

	cf_F^C	$cf_F^{II}\downarrow$	$cf_F^{II}\uparrow$	cf_{L1}^C	$cf_{L1}^{II}\downarrow$	$cf_{L1}^{II}\uparrow$	cf_{L2}^C	$cf_{L2}^{II}\downarrow$	$cf_{L2}^{II}\uparrow$	cf_{CV}^P	$cf_{CV}^{II}\downarrow$	$cf_{CV}^{II}\uparrow$	cf_{SP}^C	cf_{PID}^{MODE}	cf_{PID}^{SET}
s_1	0	1^-	1^+	0	0	0	0	0	0	$1^\pm$	1^+	1^-	0	0	0
s_2	0	1^+	1^-	0	1^-	1^+	0	0	0	0	0	0	0	0	0
s_4	0	0	0	0	1^+	1^-	0	1^-	1^+	0	0	0	0	0	0
s_5	0	0	0	0	0	0	$1^\pm$	0	0	0	1^-	1^+	$1^\pm$	$1^\pm$	$1^\pm$

Indicators of the presence of individual cyber-faults could be determined analogously as indicators of process faults, on the basis of rules corresponding to the signatures indicated in the sensitivity matrix depicted in Table 6.

In Table 7, the theoretical matrix of the influence of the cyber-attacks considered in the example of cyber-attacks on individual residues is shown. It can be noted that one of the scenarios is undetectable, and the others can be divided into two groups that differ in the observed symptoms. The two groups are highlighted in the table accordingly. Therefore, potentially, there is a possibility of locating (distinguishing) attacks from these two groups.

Table 7. Matrix of sensitivity of residues to factors related to the implementation of individual cyber-attacks.

	cas_1^A	cas_1^B	cas_1^C		cas_2		cas_3^A	cas_3^B	cas_3^C
s_1	0	0	0	0	0	0	0	0	0
s_2	0	0	0	0	0	0	0	0	0
s_4	0	0	1^-	1^+	1^-	1^+	0	0	0
s_5	0	$1^\pm$	0		0		$1^\pm$	$1^\pm$	$1^\pm$

General cyber-attack detection signal is generated based on all diagnostic signal values:

$$caD = (s_1 \neq 0) \vee (s_2 \neq 0) \vee (s_4 \neq 0) \vee (s_5 \neq 0).$$

The indicators of the presence of cyber-attack from one of the two groups: ($\{cas_1^B, cas_3^A, cas_3^B, cas_3^C\}$ and $\{cas_1^C, cas^2\}$) are determined analogously as the indicators

of process faults, on the basis of rules corresponding to the signatures indicated in the sensitivity matrix depicted in Table 7:

$$caIG_1 = \begin{cases} 1 : (s_1 = 0) \wedge (s_2 = 0) \wedge (s_4 = 0) \wedge [(s_5 = 1^+) \vee (s_5 = 1^-)] \\ 0 : otherwise \end{cases}.$$

$$caIG_2 = \begin{cases} 1 : (s_1 = 0) \wedge (s_2 = 0) \wedge [(s_4 = 1^+) \vee (s_4 = 1^-)] \wedge (s_5 = 0) \\ 0 : otherwise \end{cases}.$$

5.5. An Example of Detecting an Attack of Type: Replay Attacks in Open Loop—Scenario cas_1^B

In the first example, a cyber-attack was carried out according to the ca_1^B scenario, which consisted of the cyclical reproduction of the recorded history (from a period of 200 [s]) of process quantities $\left[cf_F^C, cf_F^{UI}, cf_{L1}^C, cf_{L1}^{UI}, cf_{L2}^C, cf_{L2}^{UI}\right]$ and the control signal $\left[cf_{CV}^{UI}\right]$. The reconstructed signal values were sent to both the control system and the operator interface. The cyber-attack started at 700 [s] of the simulation. The control scenario implemented in the system consisted in introducing step changes of the set values at the moments 400 [s], 800 [s], and 1000 [s]. The initial period of 100 [s] is not shown on the graph as it is related to the stabilization of the diagnostic algorithms after their activation (initialization phase).

In Figure 4, the most important signals related to the implementation of both process diagnostics and cyber-attack detection algorithms are shown.

On the charts we can observe:

(a) Only residuum r_5 shows a clear reaction to the introduced cyber-damages.
(b) Immediately after the onset of the attack, either the symptom 1^+ or 1^-—of the diagnostic signal s_5 is observed.
(c) The cyber-attack detection signal is actually active from the very beginning (caD).
(d) The possibility of an attack from group I (caIG1) is indicated practically all the time.
(e) Working in parallel, independently operating process diagnostics algorithms do not detect or indicate any of the process faults. This is due to the fact that a cyber-attack only causes the appearance of a symptom of a new diagnostic test not used in process diagnostics.
(f) In this case, it is possible to fully distinguish a cyber-attack from a process failure.

5.6. An Example of Detection of Attack Type: "False Data Injection Attacks"—Scenario cas_2

In the second example, a cyber-attack was carried out according to the ca_2 scenario, which consisted of falsifying the controlled variable $\left[cf_{L2}^C, cf_{L2}^{UI}\right]$ by gradually lowering its value (by subtracting the value increasing from 0 by 0.05 [m] over 100 [s], as well as the control signal $\left[cf_{CV}^{UI}\right]$. Both the value sent to the controller and the value sent to the operator interface were falsified. The cyber-attack started at 700 [s] of the simulation. The same control scenario was implemented in the system as in the first example (step changes in the reference value).

In Figure 5, the most important signals related to the implementation of both process diagnostics and cyber-attack detection algorithms are shown.

On the charts we can observe:

(a) Residuum r_4 shows a clear and lasting reaction to the introduced cyber-fault. Residua r_1 and r_5 react in transient states when the control signal (CV) reaches extreme values.
(b) Immediately after the onset of the attack, the symptom 1^+ of the diagnostic signal s_4 is observed. Symptoms of signals s_1, s_2, and s_5 are observed in the transient states.
(c) The cyber-attack detection signal is actually active from the very beginning (caD).
(d) Finally, in steady states, the possibility of an attack from group II (caIG2) is indicated.
(e) In parallel, independently operating process diagnostics algorithms detect damage f_3, the signature of which is consistent with that of the cyber-fault cf_{L2}^{UI}.
(f) In this case, it is not possible to distinguish a cyber-attack from a process fault.

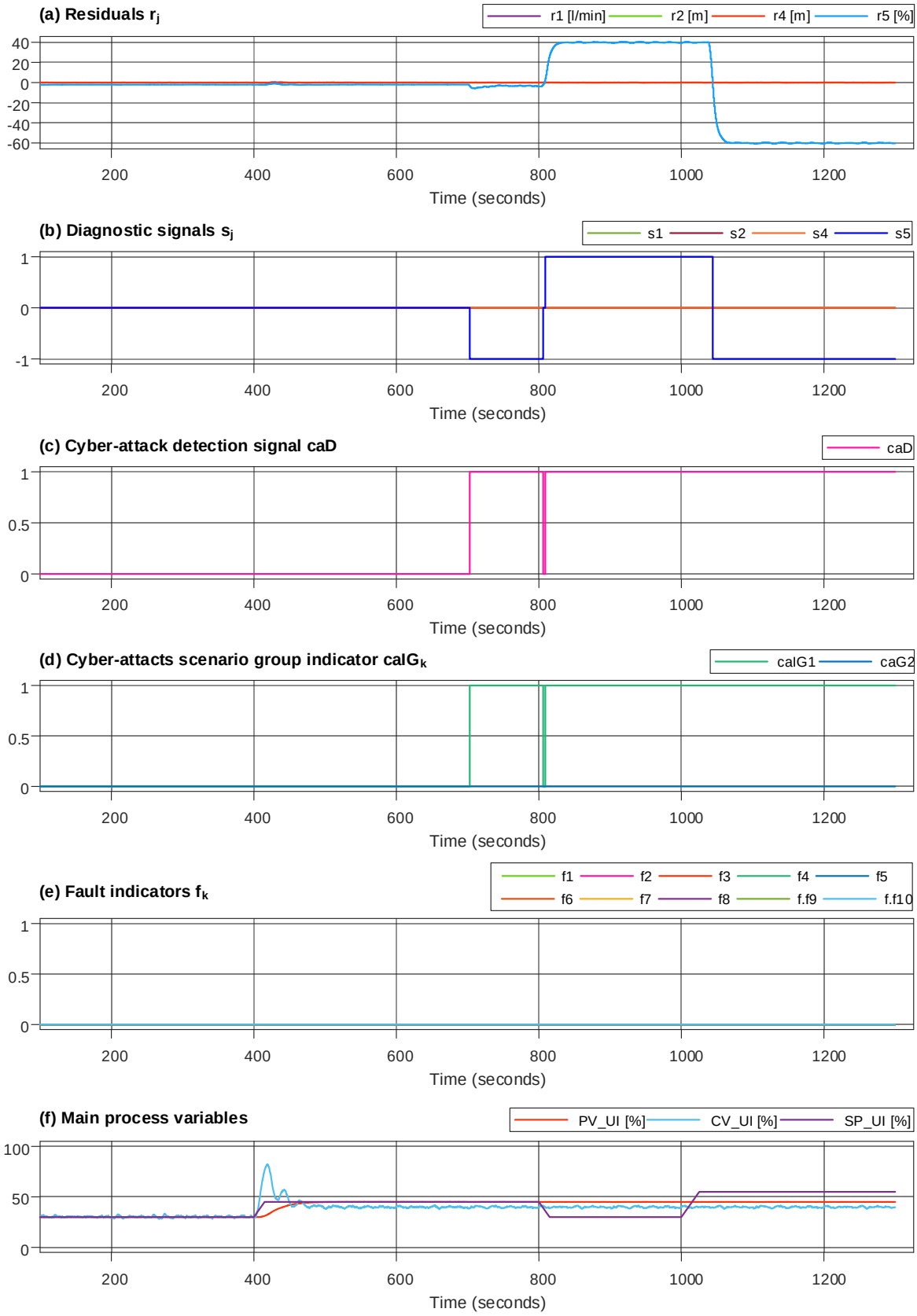

Figure 4. An example of the detection of a cyber-attack cas_1^B: (**a**) values of residuals (signals r_1, r_2 and r_4 overlap), (**b**) values of diagnostic signals (signals s_1, s_2 and s_4 overlap), (**c**) signal of cyber-attack detection, (**d**) indicators of location of groups of cyber-attack scenarios, (**e**) indicators of process failures (all signals overlap and have the value of '0'), (**f**) main process variables.

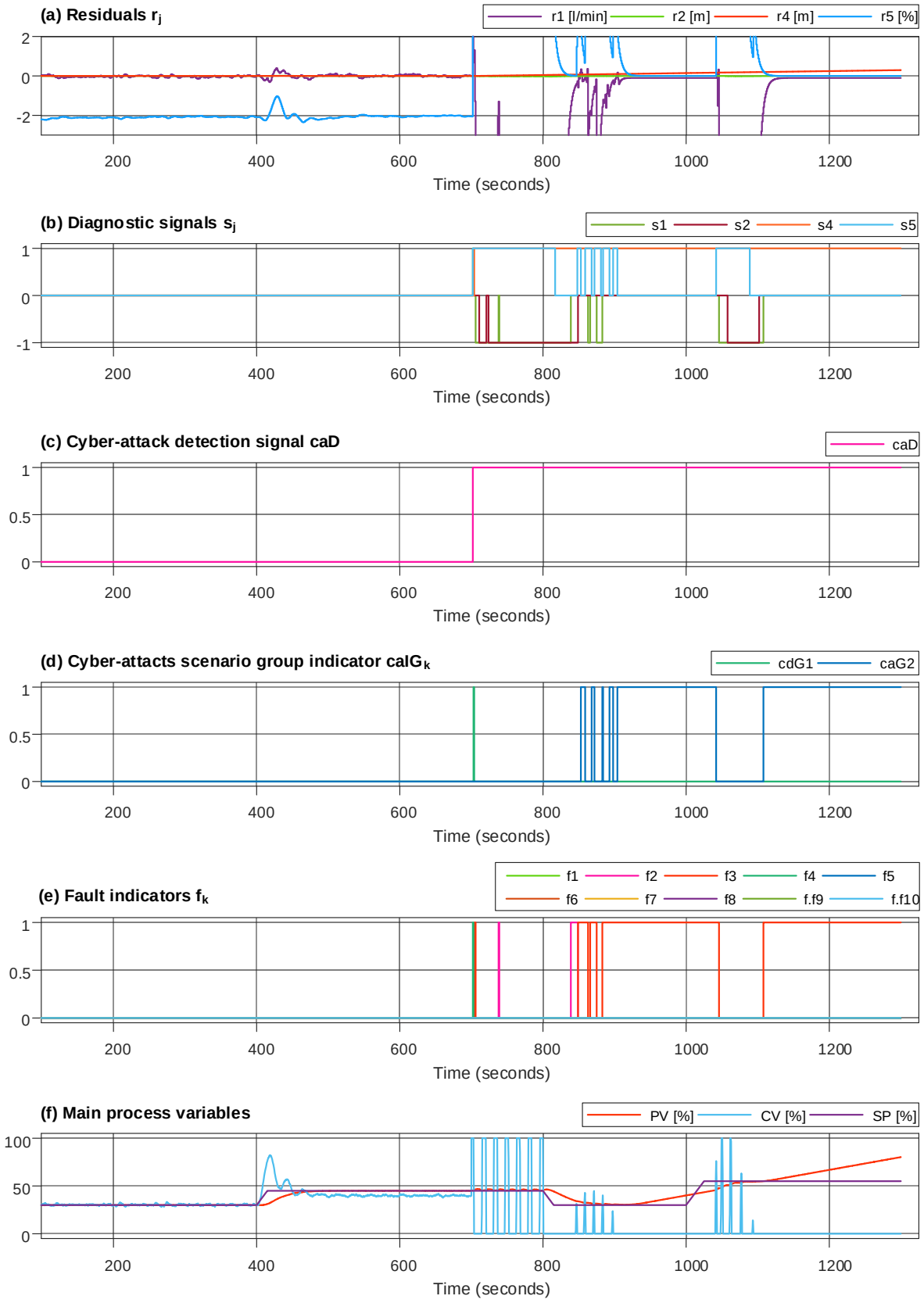

Figure 5. An example of the detection of a cyber-attack cas_2: (**a**) residual values, (**b**) values of diagnostic signals, (**c**) signal of cyber-attack detection, (**d**) indicators of location of groups of cyber-attacks scenarios, (**e**) indicators of process failures (signals f_1 and $f_5 \ldots f_{10}$ overlap and have the value of '0', signals f_4 and f_2 have the value "1" only on the beginning of cyber-attack), (**f**) main process variables.

6. Discussion—Research Problems to Be Solved

Introducing the concept of cyber-fault as the elementary interactions of a cyber-attack makes it possible to formulate an attack scenario as a vector (set) of cyber-faults. This vector is a formal record of a cyber-attack. In the future, it should be extended by the description of the cyber-fault sequences.

For each cyber-fault, as for physical faults, the signatures of the faults can be determined. Because of this, it is possible to apply a uniform approach to the detection and isolation of both faults and cyber-attacks. However, in the case of cyber-attacks, the signatures for the entire cyber-attack should also be considered and not just for cyber-faults. Such signatures, in general, can vary with time. Designing an appropriate inference mechanism taking into account this aspect is undoubtedly a challenge and requires further research.

It is much more difficult than in case of process faults to analyze the potential impact which the cyber-fault induced during an attack may have on the determined residues. When designing process diagnostics algorithms, the occurrence of single failures is often assumed. Such an assumption is justified due to the much lower probability of occurrence of multiple failures that are independent of each other. Thus, for process faults, the problem of fault influence compensation can usually be avoided. Unfortunately, when a cyber-attack occurs, the situation is much more complex. As already mentioned, virtually every cyber-attack consists of influencing a process through several different cyber-faults. Additionally, the nature of the impact may vary over time. Therefore, in the case of cyber-attacks, unlike the detection and localization of process faults, multiple failures should be considered. At the same time, the impacts of individual cyber-faults on residua also depends on mutual influence between them, which strongly depends on the nature of the cyber-attack. On the one hand, it is an element that makes conclusions very difficult, on the other hand, it may be an element that allows for the potential differentiation of a cyber-attack from process damage, assuming that the latter occur as a single event.

The use of the three-valued assessment of residues and the record of the relationship between anomalies and the values of diagnostic signals in the form of FIS creates, also in the case of the detection and isolation of cyber-attacks, the possibility of obtaining a higher distinguishability of anomalies compared to binary approaches.

The example shows that the used residues based on fuzzy partial models of the system being diagnosed are sensitive to both process faults and cyber-faults. Residuals utilizing this type of model can be used to detect anomalies not only at one operating point but also in the whole area of variability of signals that were used for model building.

The analyzed example also shows the limited use of signature-based inference for cyber-faults. Only the introduction of signatures for cyber-attacks allowed for their basic differentiation.

It has been shown in the examples that in some cases it is possible to distinguish a cyber-attack from a process fault. To achieve this, it is also necessary to use, in addition to the process models, the models of the implemented controllers, as they may be the target of an attack.

In the future, the presented method may be extended by:

- The consideration of the sequence of symptoms. The works [2,3] show that in the case of a fault, taking into account the knowledge of the sequence of symptoms leads to an increase in the discrimination of faults and thus the precision of diagnosis.
- The introduction of new techniques for assessing residues. One should take into account not only the fact of the difference between the modelled value and the measured value of the supervised process variable but also the characteristics of this difference. For example—in the case of a measuring sensor, its damage may lead to a decalibration (wrong value, but a natural measuring noise is still present), and in the case of a cyber-attack, it may lock at a specific value (no noise).
- The use of anomaly diagnostics and also of detection methods based on traffic supervision in industrial networks. In the case of network traffic analysis, most solutions also deal with a similar mechanism to technical diagnostics. In more advanced systems,

when suspicious network traffic is detected, inference is made to determine its cause. In the proposed solution, one common inference mechanism can be used for faults and cyber-attacks, then the detected network traffic anomalies will become symptoms for the inference system.

7. Summary

This article has proposed an integrated approach for the detection of faults and cyber-attacks. Whereas a commonly used approach would be to separate the IT and OT systems, consequently leading to the independent detection of faults and cyber-attacks, here it has been argued that there could be a substantial benefit from using methods related to fault detection and isolation to detect cyber-attacks. After a review of the types of cyber-attacks and the methods of detecting them, the itemization of the vector of a cyber-attack by its components—cyber-faults—has been postulated. Thanks to such an approach, residua can be determined for each cyber-fault and therefore the methods of fault detection can be used for the detection and localization of cyber-faults. The approach has been demonstrated on an example that is simple in terms of the process structure but relatively difficult in terms of the selection of residua with good sensitivity. The models used for the comparison of undisturbed system behavior with a disturbed (attacked) system have been created using fuzzy logic. The presented results are promising, but several questions are open for further research: (i) how to identify the type of cyber-attack based on the combination of detected component (cyber-faults)? (ii) how to establish conditions for distinguishability between cyber-faults and equipment faults? (iii) extend the choice of residuals by incorporating the stochastic properties of signals (some initial promising results have been obtained) and incorporate these into the anomaly diagnosis system.

Author Contributions: Conceptualization, J.M.K., A.O., M.S., P.W. and J.M.; methodology, M.S., P.W., J.M.K. and A.O.; software M.S. and P.W.; validation, M.S., K.K. and J.M.; formal analysis, M.S., J.M. and K.K.; investigation, M.S., J.M.K., A.O., P.W., J.M. and K.K.; resources, J.M. and M.S.; data curation, n/a.; writing—original draft preparation, AO.; writing—review and editing, A.O., J.M.K., M.S. and P.W.; visualization, n/a.; supervision, A.O.; project administration, A.O.; funding acquisition, A.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology within the Excellence Initiative Program—Research University (ID-UB).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kościelny, J.M.; Bartyś, M.; Sztyber, A. Diagnosing with a hybrid fuzzy-Bayesian inference approach. *Eng. Appl. Artif. Intell.* **2021**, *104*, 104345. [CrossRef]
2. Kościelny, J.M.; Syfert, M.; Wnuk, P. Diagnostic Column Reasoning Based on Multi-Valued Evaluation of Residuals and the Elementary Symptoms Sequence. *Energies* **2022**, *15*, 2614. [CrossRef]
3. Kościelny, J.M.; Syfert, M.; Wnuk, P. Diagnostic Row Reasoning Method Based on Multiple-Valued Evaluation of Residuals and Elementary Symptoms Sequence. *Energies* **2021**, *14*, 2476. [CrossRef]
4. Mur, A.; Travé-Massuyès, L.; Chanthery, E.; Pons, R.; Ribot, P. A Neural Algorithm for the Detection and Correction of Anomalies: Application to the Landing of an Airplane. *Sensors* **2022**, *22*, 2334. [CrossRef] [PubMed]
5. Romero, L.; Blesa, J.; Puig, V.; Cembrano, G. Clustering-Learning Approach to the Localization of Leaks in Water Distribution Networks. *J. Water Resour. Plan. Manag.* **2022**, *148*, 04022003. [CrossRef]
6. Pazera, M.; Buciakowski, M.; Witczak, M.; Mrugalski, M. A quadratic boundedness approach to a neural network-based simultaneous estimation of actuator and sensor faults. *Neural Comput. Appl.* **2020**, *32*, 379–389. [CrossRef]

7. Pazera, M.; Witczak, M.; Kukurowski, N.; Buciakowski, M. Towards Simultaneous Actuator and Sensor Faults Estimation for a Class of Takagi-Sugeno Fuzzy Systems: A Twin-Rotor System Application. *Sensors* **2020**, *20*, 3486. [CrossRef]
8. Jakobsson, E.; Petterson, R.; Frisk, E.; Krysande, M. Fatigue Damage Monitoring for Mining Vehicles Using Data Driven Models. *Int. J. Progn. Health Manag.* **2020**, *11*, 1–15. [CrossRef]
9. Łabęda-Grudziak, Z.M.; Lipiński, M. The identification method of the coal mill motor power model with the use of machine learning techniques. *Bull. Pol. Acad. Sci. Tech. Sci.* **2021**, *69*, e135842. [CrossRef]
10. Kamal, S.Z.; Al Mubarak, S.M.; Scodova, B.D.; Naik, P.; Flichy, P.; Coffin, G. IT and OT Convergence—Opportunities and Challenges. In Proceedings of the SPE Intelligent Energy International Conference and Exhibition, Aberdeen, Scotland, UK, 6–8 September 2016. [CrossRef]
11. EU:2020. *Emerging Technologies in Electronic Components and Systems (ECS): Oportunities Ahead*; EU Publications: Luxembourg, 2020.
12. NSF:2016. Available online: https://www.nsf.gov/news/news_summ.jsp (accessed on 1 October 2020).
13. Dastbaz, M.; Cochrane, P. *Industry 4.0 and Engineering for a Sustainable Future*; Springer: Cham, Switzerland, 2019.
14. Kościelny, J.; Syfert, M.; Ordys, A.; Wnuk, P.; Możaryn, J.; Fajdek, B.; Puig, V.; Kukiełka, K. Towards a unified approach to detection of faults and cyber-attacks in industrial installations. In Proceedings of the 2021 European Control Conference (ECC), Delft, The Netherlands, 29 June–2 July 2021; pp. 1839–1844. [CrossRef]
15. ICS-CERT. Overview of Cyber Vulnerabilities, September 2021. Available online: https://ics-cert.us-cert.gov/content/overview-cyber-vulnerabilities (accessed on 1 October 2020).
16. Béla, G.; Kiss, I.; Haller, P. A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures. *Int. J. Crit. Infrastruct. Prot.* **2015**, *10*, 3–17.
17. Chen Thomas, M.; Abu-Nimeh, S. Lessons from stuxnet. *Computer* **2011**, *44*, 91–93. [CrossRef]
18. Hagerott, M. Stuxnet and the vital role of critical infrastructure operators and engineers. *Int. J. Crit. Infrastruct. Prot.* **2014**, *7*, 244–246. [CrossRef]
19. Assenza, G.; Faramondi, L.; Oliva, G.; Setola, R. Cyber threats for operational technologies. *Int. J. Syst. Syst. Eng.* **2020**, *10*, 128–142. [CrossRef]
20. Lee, R.M.; Assante, M.J.; Conway, T. *Crashoverride: Analysis of the Threat to Electric Grid Operations*; Dragos Inc.: Houston, TX, USA, 2017.
21. Tian, J.; Tan, R.; Guan, X.; Xu, Z.; Liu, T. Moving Target Defense Approach to Detecting Stuxnet-Like Attacks. *IEEE Trans. Smart Grid* **2019**, *11*, 291–300. [CrossRef]
22. Geiger, M.; Bauer, J.; Masuch, M.; Franke, J. An Analysis of Black Energy 3, Crashoverride, and Trisis, Three Malware Approaches Targeting Operational Technology Systems. In Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 8–11 September 2020; Volume 1, pp. 1537–1543.
23. Liang, G.; Weller, S.R.; Zhao, J.; Luo, F.; Dong, Z.Y. The 2015 Ukraine Blackout: Implications for False Data Injection Attacks. *IEEE Trans. Power Syst.* **2016**, *32*, 3317–3318. [CrossRef]
24. Biffl, S.; Eckhart, M.; Lüder, A.; Weippl, E.R. (Eds.) *Security and Quality in Cyber-Physical Systems Engineering*; Springer: Berlin/Heidelberg, Germany, 2019.
25. Kosmowski, K.T. Functional safety concept for hazardous systems and new challenges. *J. Loss Prev. Process Ind.* **2006**, *19*, 298–305. [CrossRef]
26. Kosmowski, K.T.; Piesik, E.; Piesik, J.; Śliwiński, M. Integrated Functional Safety and Cybersecurity Evaluation in a Framework for Business Continuity Management. *Energies* **2022**, *15*, 3610. [CrossRef]
27. Mashkoor, A.; Egyed, A.; Wille, R.; Stock, S. Model-driven engineering of safety and security software systems: A systematic mapping study and future research directions. *J. Softw. Evol. Process* **2022**, e2457. [CrossRef]
28. Kriaa, S.; Pietre-Cambacedes, L.; Bouissou, M.; Halgand, Y. A survey of approaches combining safety and security for industrial control systems. *Reliab. Eng. Syst. Saf.* **2015**, *139*, 156–178. [CrossRef]
29. Johnson, N.; Kelly, T. An Assurance Framework for Independent Co-assurance of Safety and Security. *J. Syst. Saf.* **2018**, *54*, 32–38. [CrossRef]
30. Pfrang, S.; Meier, D. Detecting and preventing replay attacks in industrial automation networks operated with profinet IO. *J. Comput. Virol. Hacking Tech.* **2018**, *14*, 253–268. [CrossRef]
31. C´ardenas, A.A.; Amin, S.; Sinopoli, B.; Giani, A.; Perrig, A.; Sastry, S. Challenges for Securing Cyber Physical Systems. In Proceedings of the Workshop on Future Directions in Cyber-Physical Systems Security, Newark, NJ, USA, 22–24 July 2009.
32. Ding, D.; Han, Q.-L.; Xiang, Y.; Ge, X.; Zhang, X.-M. A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing* **2018**, *275*, 1674–1683. [CrossRef]
33. Lucke, M.; Lu, J.; Quevedo, D.E. Coding for secrecy in remote state estimation with an adversary. *IEEE Trans. Autom. Control* **2022**, 1. [CrossRef]
34. An, L.; Yang, G.-H. Enhancement of opacity for distributed state estimation in cyber–physical systems. *Automatica* **2021**, *136*, 110087. [CrossRef]
35. Gertler, J.J. *Fault Detection and Diagnosis in Engineering Systems*; Marcel Dekker, Inc.: New York, NY, USA; Basel, Switzerland; Hong Kong, China, 1998.
36. Chen, J.; Patton, R. *Robust Model Based Fault Diagnosis for Dynamic Systems*; Kluwer Akademic Publishers: Boston, MA, USA, 1999.

37. Blanke, M.; Kinnaert, M.; Lunze, J.; Staroswiecki, M. *Diagnosis and Fault-Tolerant Control*; Springer: Berlin/Heidelberg, Germany, 2004.
38. Korbicz, J.; Kościelny, J.M.; Kowalczuk, Z.; Cholewa, W. (Eds.) *Fault Diagnosis: Models, Artificial Intelligence Methods, Applications*; Springer: Berlin/Heidelberg, Germany, 2004.
39. Isermann, R. *Fault Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance*; Springer: Berlin/Heidelberg, Germany, 2006.
40. Witczak, M. *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems, from Analytical to Soft Computing Approaches*; Springer: Berlin/Heidelberg, Germany, 2007.
41. Frank, P.M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. *Automatica* **1990**, *26*, 459–474. [CrossRef]
42. Isermann, R. Model-based fault-detection and diagnosis—Status and applications. *Annu. Rev. Control.* **2005**, *29*, 71–85. [CrossRef]
43. Leonhardt, S.; Ayoubi, M. Methods of fault diagnosis. *Control Eng. Pract.* **1997**, *5*, 683–692. [CrossRef]
44. Kościelny, J.; Syfert, M.; Wnuk, P. Advanced monitoring and diagnostic system 'AMandD'. In Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Beijing, China, 29 August–1 September 2006; Volume 1, pp. 635–640. [CrossRef]
45. Korbicz, J.; Kościelny, J.M. (Eds.) *Modeling, Diagnostics and Process Control. Implementation in the DiaSter System*; Springer: Berlin/Heidelberg, Germany, 2010.
46. Natarajan, S.; Srinivasan, R. Implementation of multi agents based system for process supervisionin large-scale chemical plants. *Comput. Chem. Eng.* **2014**, *60*, 182–196. [CrossRef]
47. Sánchez, H.S.; Rotondo, D.; Escobet, T.; Puig, V.; Quevedo, J. Bibliographical review on cyber attacks from a control oriented perspective. *Annu. Rev. Control* **2019**, *48*, 103–128. [CrossRef]
48. Mahmoud Magdi, S.; Hamdan, M.M.; Baroudi, U.A. Modeling and control of cyber-physical systems subject to cyberattacks: A survey of recent advances and challenges. *Neurocomputing* **2019**, *338*, 101–115. [CrossRef]
49. Jang-Jaccard, J.; Nepal, S. A survey of emerging threats in cybersecurity. *J. Comput. Syst. Sci.* **2014**, *80*, 973–993. [CrossRef]
50. Teixeira, A.; Shames, I.; Sandberg, H.; Johansson, K.H. A secure control framework for resource-limited adversaries. *Automatica* **2015**, *51*, 135–148. [CrossRef]
51. Pasqualetti, F.; Dorfler, F.; Bullo, F. Attack Detection and Identification in Cyber-Physical Systems. *IEEE Trans. Autom. Control* **2013**, *58*, 2715–2729. [CrossRef]
52. Loukas, G. Cyber-physical attacks on industrial control systems. In *Cyber-Physical Attacks*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 105–144.
53. Young, W.; Leveson, N. Systems thinking for safety and security. In Proceedings of the ACSAC '13, New Orleans, LA, USA, 9–13 November 2013; pp. 1–8.
54. Leveson, N. A new accident model for engineering safer systems. *Saf. Sci.* **2003**, *42*, 237–270. [CrossRef]
55. Patriarca, R.; Chatzimichailidou, M.; Karanikas, N.; Di Gravio, G. The past and present of System-Theoretic Accident Model and Processes (STAMP) and its associated techniques: A scoping review. *Saf. Sci.* **2021**, *146*, 105566. [CrossRef]
56. André, T.; Pérez, D.; Sandberg, H.; Johansson, K.H. Attack models and scenarios for networked control systems. In Proceedings of the 1st International Conference on High Confidence Networked Systems, Beijing, China, 17–18 April 2012.
57. Shi, D.; Guo, Z.; Johansson, K.; Shi, L. Causality countermeasures for anomaly detection in cyber-physical systems. *IEEE Trans. Autom. Control* **2017**, *63*, 386–401. [CrossRef]
58. Hu, Y.; Li, H.; Yang, H.; Sun, Y.; Sun, L.; Wang, Z. Detecting stealthy attacks against industrial control systems based on residual skewness analysis. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 74. [CrossRef]
59. Cao, Y.; Zhang, L.; Zhao, X.; Jin, K.; Chen, Z. An Intrusion Detection Method for Industrial Control System Based on Machine Learning. *Information* **2022**, *13*, 322. [CrossRef]
60. Urbina, D.I.; Giraldo, J.A.; Cardenas, A.A.; Tippenhauer, N.O.; Valente, J.; Faisal, M.; Ruths, J.; Candell, R.; Sandberg, H. Limiting the Impact of Stealthy Attacks on Industrial Control Systems. In Proceedings of the CCS '16: 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 1092–1095. [CrossRef]
61. Mitchell, R.; Chen, I.R. A survey of intrusion detection techniques for cyber-physical systems. *ACM Comput. Surv.* **2014**, *46*, 55. [CrossRef]
62. Syfert, M.; Wnuk, P.; Kościelny, J.M. DiaSter—Intelligent system for diagnostics and automatic control support of industrial processes. *JAMRIS J. Autom. Mob. Robot. Intell. Syst.* **2011**, *5*, 41–46.
63. Trapiello, C.; Rotondo, D.; Sanchez, H.; Puig, V. Detection of replay attacks in CPSs using observer-based signature compensation. In Proceedings of the 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), Paris, France, 23–26 April 2019; pp. 1–6. [CrossRef]
64. Trapiello, C.; Puig, V. Replay attack detection using a zonotopic KF and LQ approach. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 3117–3122. [CrossRef]
65. Trapiello, C.; Puig, V. Input Design for Active Detection of Integrity Attacks using Set-based Approach. *IFAC-Pap. OnLine* **2020**, *53*, 11094–11099. [CrossRef]
66. Ao, W.; Song, Y.; Wen, C. Adaptive cyber-physical system attack detection and reconstruction with application to power systems. *IET Control Theory Appl.* **2016**, *10*, 1458–1468. [CrossRef]

67. Sinopoli, B.; Schenato, L.; Franceschetti, M.; Poolla, K.; Jordan, M.; Sastry, S. Kalman Filtering with Intermittent Observations. *IEEE Trans. Autom. Control* **2004**, *49*, 1453–1464. [CrossRef]
68. Cong, T.; Tan, R.; Ottewill, J.R.; Thornhill, N.F.; Baranowski, J. Anomaly Detection and Mode Identification in Multimode Processes Using the Field Kalman Filter. *IEEE Trans. Control Syst. Technol.* **2020**, *29*, 2192–2205. [CrossRef]
69. Wang, X.; Luo, X.; Zhang, Y.; Guan, X. Detection and Isolation of False Data Injection Attacks in Smart Grids via Nonlinear Interval Observer. *IEEE Internet Things J.* **2019**, *6*, 6498–6512. [CrossRef]
70. Abbaspour, A.; Sargolzaei, A.; Yen, K. Detection of False Data Injection Attack on Load Frequency Control in Distributed Power Systems. In Proceedings of the 2017 North American Power Symposium (NAPS), Morgantown, WV, USA, 17–19 September 2017. [CrossRef]
71. Wu, Z.; Albalawi, F.; Zhang, J.; Zhang, Z.; Durand, H.; Christofides, P.D. Detecting and Handling Cyber-Attacks in Model Predictive Control of Chemical Processes. *Mathematics* **2018**, *6*, 173. [CrossRef]
72. Kościelny, J.M.; Syfert, M.; Wnuk, P. The Idea of On-line Diagnostics as a Method of Cyberattack. In *Advanced Solutions in Diagnostics and Fault Tolerant Control*; Kościelny, J., Syfert, M., Sztyber, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 449–457.
73. Bobba, R.B.; Rogers, K.M.; Wang, Q.; Khurana, H.; Nahrstedt, K.; Overbye, T.J. Detecting false data injection attacks on dc state estimation. In Proceedings of the First Workshop on Secure Control Systems, Stockholm, Sweden, 12 April 2010; Volume 2010.
74. Yang, Q.; Yang, J.; Yu, W.; An, D.; Zhang, N.; Zhao, W. On False Data-Injection Attacks against Power System State Estimation: Modeling and Countermeasures. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *25*, 717–729. [CrossRef]
75. Chaojun, G.; Jirutitijaroen, P.; Motani, M. Detecting False Data Injection Attacks in AC State Estimation. *IEEE Trans. Smart Grid* **2015**, *6*, 2476–2483. [CrossRef]
76. Huang, Y.; Li, H.; Campbell, K.A.; Han, Z. Defending false data injection attack on smart grid network using adaptive CUSUM test. In Proceedings of the 45th Annual Conference on Information Sciences and Systems, Baltimore, MD, USA, 23–25 March 2011; pp. 1–6.
77. Kontouras, E.; Tzes, A.; Dritsas, L. Impact Analysis of a Bias Injection Cyber-Attack on a Power Plant. *IFAC-Pap. OnLine* **2017**, *50*, 11094–11099. [CrossRef]
78. Hoehn, A.; Zhang, P. Detection of replay attacks in cyber-physical systems. In Proceedings of the 2016 IEEE American Control Conference, Boston, MA, USA, 6–8 July 2016; pp. 290–295.
79. Andreas, H.; Zhang, P. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In Proceedings of the 2016 American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016.
80. Ye, N. *Cyber Signal/Noise Characteristics and Sensor Models for Early Cyber Indications and Warning*; Report by Air Force Research Laboratory/IFGB; Air Force Research Laboratory/IFGB: Hanscom, MA, USA, 2005.
81. Zhang, X.; Zhu, F.; Zhang, J.; Liu, T. Attack isolation and location for a complex network cyber-physical system via zonotope theory. *Neurocomputing* **2021**, *469*, 239–250. [CrossRef]
82. Su, L.; Shahrampour, S. Finite-Time Guarantees for Byzantine-Resilient Distributed State Estimation with Noisy Measurements. *IEEE Trans. Autom. Control* **2019**, *65*, 3758–3771. [CrossRef]
83. Manandhar, K.; Cao, X. Attacks/faults detection and isolation in the Smart Grid using Kalman Filter. In Proceedings of the 2014 23rd International Conference on Computer Communication and Networks (ICCCN), Shanghai, China, 4–7 August 2014; pp. 1–6. [CrossRef]
84. Mohammadi, A.; Yang, C.; Chen, Q. Attack Detection/Isolation via a Secure Multisensor Fusion Framework for Cyberphysical Systems. *Complexity* **2018**, *2018*, 1240149. [CrossRef]
85. Taheri, M.; Khorasani, K.; Shames, I.; Meskin, N. Cyber Attack and Machine Induced Fault Detection and Isolation Methodologies for Cyber-Physical Systems. *arXiv* **2020**, arXiv:2009.06196.
86. Zhao, Z.; Yang, Y.; Li, Y.; Liu, R. Security analysis for cyber-physical systems under undetectable attacks: A geometric approach. *Int. J. Robust Nonlinear Control* **2018**, *30*, 4359–4370. [CrossRef]
87. Sandberg, T.H.; Johansson, K.H. Networked control systems under cyber attacks with applications to power networks. In Proceedings of the 2010 American Control Conference, Baltimore, MD, USA, 30 June–2 July 2010; pp. 3690–3696.
88. Zhu, M.; Martinez, S. On the Performance Analysis of Resilient Networked Control Systems under Replay Attacks. *IEEE Trans. Autom. Control* **2013**, *59*, 804–808. [CrossRef]