Proceedings of the 21st International Conference on Global Research and Education (Inter-Academia 2024)

The Model of Ongoing Diagnosis of Process Faults and Detection of Cybernetic Attacks for a Steam Line

M. Syfert[a,*], P. Wnuk[b], A. Sztyber-Betley[a] and M. Pobocha[b]

[a] *Faculty of Mechatronics, Institute of Automatic Control and Robotics, Warsaw University of Technology, św. A. Boboli 8, 02-525 Warsaw, Poland*
[b] *Enerbit Sp z o.o., ul. Czerniakowska 26b, 00-714 Warsaw, Poland*

Doi: 10.12693/APhysPolA.146.438 *e-mail: michal.syfert@pw.edu.pl

The paper presents a model of a system for ongoing monitoring and diagnosis of process faults and detection of cybernetic attacks for a section of the boiler steam line. The model consists of the project of an algorithm and a prototype implementation of the system. The implemented algorithm integrates, in one coherent approach, the tasks of monitoring process faults based on parametric partial models and detecting cybernetic attacks based on dedicated checks for control system loops based on a selected set of performance indices. The important model feature is the use of data on malfunctions and detected intrusions obtained from the industrial control system and intrusion detection system during the reconciliation stage and final diagnosis formulation. The operation of the system is presented in the example of a pilot application for a simulator of a section of a steam line. Based on the presented summary of the test results it can be stated that the proposed solution allows for obtaining high detectability and isolability of process faults and their distinguishability from cyber attacks. One of the advantages of the proposed solution is the indirect use of knowledge about phenomena occurring in the process in the task of detecting cybernetic attacks. The structure of the proposed model allows for its relatively easy adaptation to other technological processes.

topics: on-line diagnostics, process diagnostics, cybernetic attacks, steam line

1. Introduction

Industrial processes, often extremely complex and carried out in an increasingly automated way, are susceptible to many factors that affect both the quality of the process and the safety of installations, people, and the environment. Critical elements affecting the process are, among others, process faults (of technological components, measuring devices, and actuators), errors in the operation of the *industrial control system* (ICS), operator errors, or factors related to hostile external influence, i.e., cybernetic attacks [4].

For many years, one could observe the development of various types of algorithms and diagnostic systems dealing with automatic detection and isolation of process faults. A slightly different problem, for which other algorithms and systems are proposed, are issues related to monitoring the quality of control loops and the impact of operators' activities on these systems. In large-scale industrial processes, operators and engineers are responsible for the supervision of a considerable number of control loops [5]. In response to this challenge, the field of algorithms for control loop performance monitoring is developing. One of the first proposed solutions was the Harris index [6]. Currently, there are numerous indices for evaluating the tuning and quality of control loop operation [7, 8]. The concept of utilising these indices to detect and isolate cyberattacks was proposed in [9, 10]. The third area, addressed in this work, is the detection of cybernetic attacks. This issue has been particularly strongly discussed in recent years, which is related to both the growing global threat and the growing digitization of production processes. Due to the overriding need to protect the process and the potential similar effects of various types of the above-mentioned causes on proces operation, it seems interesting to develop algorithms and systems that integrate the possibility of monitoring all these impact factors in a coherent solution.

Additionally, to achieve greater detection precision and make it more possible to differentiate between process faults and cybernetic attack, integration with diagnostic features of the modern industrial control system (ICS), as well as with the *intrusion detection system* (IDS), can be considered. ICS offer extensive diagnostic capabilities, with particular focus on self-diagnosis and diagnostics of intelligent devices. A significant portion of these capabilities are designed for offline operation (e.g., automatic verification of the configuration correctness

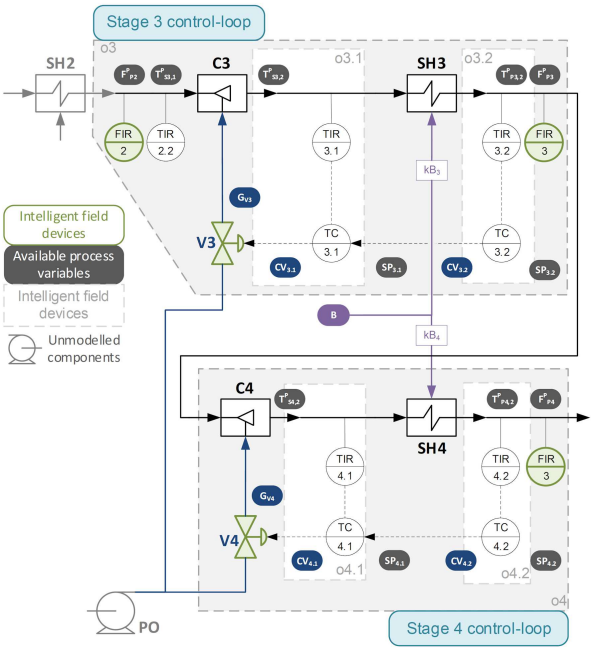

Fig. 1. The P&ID diagram of simulated process: components, control-loops, and available process variables.

and connections conducted passively or actively, analysis of system logs, generation and analysis of self-diagnostic reports) and therefore cannot be utilised in an online diagnostic system. There is also a group of information generated by the ICS that is available online. These are the results of diagnostic tests conducted on an ongoing basis (online) at various levels of the system, such as: intelligent field devices, field networks, integrated controllers, system buses, network devices, servers, operator and engineering workstations. The results of diagnostic checks are usually available via the *application programming interface* (API) and take the form of events, easily convertible into diagnostic signals. A typical IDS allows for generation of events (treated as diagnostic signals) upon detection of various events, such as:

(i) appearance of a packet with specific characteristics (*transmission control protocol* (TCP) port or protocol) sent to selected devices,

(ii) *(distributed)* denial of service (DoS/DDoS) attacks,

(iii) attacks with a known signature on all or selected devices,

(iv) appearance of a command within the selected protocol for a specific device.

In this article, a comprehensive approach to the detection and isolation of process faults, anomaly detection in control loops and detection of cybernetic attack is considered. The described model includes both the inference algorithm and the prototype implementation of system modules that implement

TABLE I

List of considered and simulated process faults.

Notation	Description
$f^{\mathrm{P}}_{T.S31}$, $f^{\mathrm{P}}_{F.P4}$, $f^{\mathrm{P}}_{T.S32}$, $f^{\mathrm{P}}_{T.P32}$, $f^{\mathrm{P}}_{T.S42}$, $f^{\mathrm{P}}_{T.P42}$, $f^{\mathrm{P}}_{G.V3}$, $f^{\mathrm{P}}_{G.V4}$, $f^{\mathrm{P}}_{F.P2}$, $f^{\mathrm{P}}_{F.P3}$	measurement path # fault
$f^{\mathrm{P}}_{CV.31}$, $f^{\mathrm{P}}_{CV.41}$	control value path fault of #.# control-loop
f^{P}_{V3}, f^{P}_{V4}	actuator # fault
f^{P}_{SH3}, f^{P}_{SH4}	process component # fault

this inference. The test process used, for which the model was applied, is a simulator developed for research purposes, encompassing a fragment of the steam line of a power unit. The simulator has the capability to model both correct operation and the introduction of process faults and selected cybernetic attack scenarios. The reasoning is conducted based on a set of various types of residuals based on *partial parametric models* (PPM), developed by control loop monitoring algorithms, or derived from external ICS/IDS.

Firstly, Sect. 2 presents the process for which the algorithm and diagnostic system was developed and the simulator that was used in the research and testing. Section 3 discusses the structure of the proposed inference algorithm. In Sect. 4 and 5 the prototype implementation of the system and the results obtained are briefly discussed. The last section contains conclusions and a summary.

2. Monitoring object

2.1. The process — selected sections of steam line

For the purposes of the research, a fragment of the steam line of an exemplary power unit was developed and implemented as a simulator. Simulator covers stages (iii) and (iv), each one consists of: a cooler, together with an actuator; a superheater; a fragment of the pipeline connecting the above units and a cascade control loop of the steam temperature at the outlet of the superheater (the auxiliary quantity is controlled by the steam temperature downstream of the superheater). The simulator is a simplified transmittance model with coefficients set on the basis of data from the actual technological process. The P&ID diagram of the process is shown in Fig. 1. It was assumed that the following parameters are available: temperatures (T), SP and CV values of the controllers, actuator position (G), and the quantity symbolizing the fuel stream fed to the boiler (B).

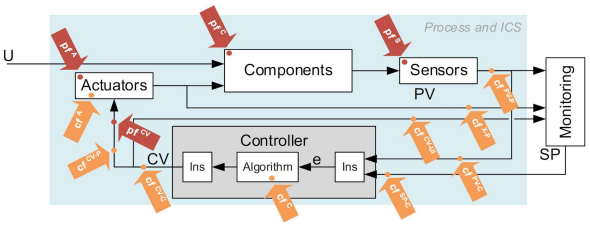

Fig. 2. Faults in the process — the places of influence of process faults (pf) and cyber faults (cf).

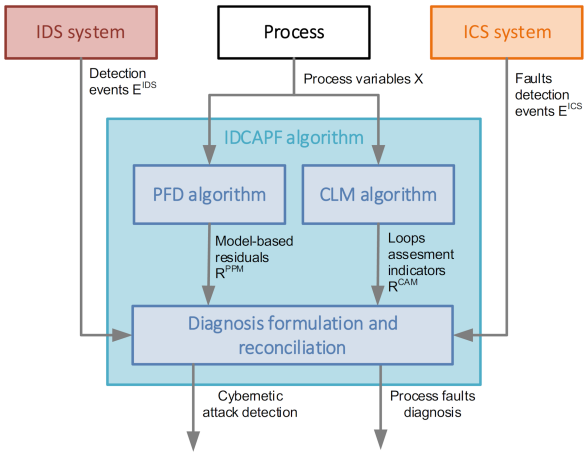

Fig. 3. General structure of on-line diagnosis of process faults and cybernetic attacks based of different sources of information.

Finally, one can define the set of control loops O

$$O = \{o_i\}, \quad i = 1, \ldots, N^O, \tag{1}$$

where $N^O = 4$ is the number of loops considered. For each loop o_i there are associated components and signals

$$o_i = \langle SP_i, PV_i, CV_i, a_i, X_i \rangle, \tag{2}$$

where a_i is actuator, SP_i — set point, PV_i — process value (controlled), CV_i — control value, X_i — set of additional process variables describing the operation of the control loop.

2.2. Process and cybernetic faults

The simulator also allows to introduce process faults and, so-called, cyber faults. The places of their influence are symbolically shown in Fig. 2. Basically, each cybernetic attack is carried out in accordance with a designed scenario consisting of elementary interactions on individual system elements and signals in communication channels, called cyber faults [11]. They are analogous to classic process faults considered in the *fault detection and isolation* (FDI) environment, for which diagnostic algorithms are designed.

The set of considered process faults F^{P}

$$F^{\mathrm{P}} = \{f_k^{\mathrm{P}}\}, \quad k = 1, \ldots, N^{\mathrm{F}}, \tag{3}$$

covers faults of all measurement and signal paths, as well as actuators, and an exemplary failure of a technological component. The list of process faults is presented in Table I.

Cyber faults form a set F^{C}. In the described solution, individual cyber faults are not indicated in the diagnosis, a general cybernetic attack detection signal is generated. For this reason, these faults are not discussed in more detail in the study.

3. Monitoring algorithms

3.1. General system structure

The general structure of the *Integrated Diagnostics of Cybernetic Attack and Process Faults* (IDCAPF) inference algorithm is presented in Fig. 3. The algorithm performs three main stages, carried out by separate modules:

(i) *Process Fault Detection* (PFD) algorithm — detection of process faults F^{P} on the basis of a set of designed parametric partial models (PPM) reconstructing selected process variables based on the analysis of available process signals. A set of residuals R^{PPM} is generated at the output.

(ii) *Control Loop Monitoring* (CLM) algorithm — analysis of operation of control loops based on available measurement signals and dedicated algorithms. A set of residuals R^{CLM} is generated in the output.

(iii) Diagnosis formulation and reconciliation — module of inference in the field of isolation of process faults and detection of cybernetic attack. At this stage, additional information from IDS and ICS is taken into account. The output is a classic diagnosis for process faults and signals of detecting a cybernetic attack, with a possible indication of the attacked loop.

3.2. Detection based on PPM

In industrial practice, full, analytical models of supervised installations are not very common. Developing such a model is expensive and time-consuming. It is easier to rely on partial parametric models obtained through identification. Partial models typically allow the reconstruction of a single process variable and the generation of one residual based on it

$$R^{\mathrm{PPM}} = \{r_p^{\mathrm{PPM}}\}, \quad p = 1, \ldots, N^{\mathrm{M}}, \tag{4}$$

$$r_p^{\mathrm{PPM}} = \hat{x}_p^{\mathrm{PPM}} - x_i, \tag{5}$$

$$\hat{x}_p^{\mathrm{PPM}} = f_p(x_n, \ldots, x_m), \tag{6}$$

TABLE II

List of model-based residuals.

R	Based on model structure	Model type
$r^{\mathrm{PPM}}_{T.S32.1}$	$\hat{T^{\mathrm{P}}_{S32}} = f_1\left(T^{\mathrm{P}}_{S32,k-1}, T^{\mathrm{P}}_{S31,k-1}, T^{\mathrm{P}}_{S31,k-3}, GV_{3,k-1}\right)$	neural, ARX
$r^{\mathrm{PPM}}_{T.P32.2}$	$\hat{T^{\mathrm{P}}_{P32}} = f_2\left(T^{\mathrm{P}}_{P32,k-1}, T^{\mathrm{P}}_{S32,k-1}, T^{\mathrm{P}}_{S32,k-8}, F^{\mathrm{P}}_{P3,k-8}\right)$	neural, ARX
$r^{\mathrm{PPM}}_{T.S42.3}$	$\hat{T^{\mathrm{P}}_{S42}} = f_3\left(T^{\mathrm{P}}_{P32,k-1}, T^{\mathrm{P}}_{P32,k-6}, GV_{4,k-1}, GV_{4,k-6}\right)$	neural, MA
$r^{\mathrm{PPM}}_{T.P42.4}$	$\hat{T^{\mathrm{P}}_{P42}} = f_4\left(T^{\mathrm{P}}_{S42}, B, F^{\mathrm{P}}_{P4}\right)$	linear, MA
$r^{\mathrm{PPM}}_{G.V3.5}$	$\hat{GV_3} = f_5\left(CV_{321}\right)$	linear, MA
$r^{\mathrm{PPM}}_{G.V4.6}$	$\hat{GV_4} = f_6\left(CV_{41}\right)$	linear, MA
$r^{\mathrm{PPM}}_{T.S32.7}$	$\hat{T^{\mathrm{P}}_{S32}} = f_7\left(T^{\mathrm{P}}_{S32,k-1}, T^{\mathrm{P}}_{S31,k-1}, T^{\mathrm{P}}_{S31,k-3}, CV_{31,k-1}, CV_{31,k-3}\right)$	neural, ARX
$r^{\mathrm{PPM}}_{T.S42.8}$	$\hat{T^{\mathrm{P}}_{S42}} = f_8\left(T^{\mathrm{P}}_{P32,k-1}, T^{\mathrm{P}}_{P32,k-6}, CV_{41,k-1}, CV_{41,k-7}\right)$	neural, MA
$r^{\mathrm{PPM}}_{F.P3.9}$	$\hat{F^{\mathrm{P}}_{P3}} = f_9\left(F^{\mathrm{P}}_{P2,k-1}, GV_{3,k-1}\right)$	neural, MA
$r^{\mathrm{PPM}}_{F.P4.10}$	$\hat{F^{\mathrm{P}}_{P4}} = f_{10}\left(F^{\mathrm{P}}_{P3,k-1}, GV_{4,k-1}\right)$	neural, MA

where r_p^{PPM} is residual generated by the model of the process variable x_i, N^{M} is numer of PPM models, $\hat{x}_p^{\mathrm{PPM}}$ is process variable reconstructed based on the model.

To build the models, knowledge of the technological process and available archival process data sets are used. Many different modelling techniques are described in the literature. Due to the simplicity of description, time and stability of calculations, and the possibility of extrapolation, *autoregressive model with exogenous input* (ARX) and linear *moving average* (MA) models without an autoregressive term were used. Since the modelled part of the process is not fully linear, a neural model in the form of a *multilayer perceptron* (MLP) network with external feedback was also prepared for each of the reconstructed process values. The list of the models obtained and selected for application is presented in Table II.

3.3. Dedicated algorithms for control loops monitoring

Loop performance indices are used to calculate the residuals R^{CLM}, the same set of residuals is for each control loop, thus

$$R^{\mathrm{CLM}} = \bigcup_{i=1}^{N^O} R_i^{\mathrm{CLM}}, \tag{7}$$

$$R_i^{\mathrm{CLM}} = \{r_{i,c}^{\mathrm{CLM}}\}, \quad c = 1, \ldots, N^{\mathrm{C}}, \tag{8}$$

where $r_{i,c}^{\mathrm{CLM}}$ is c-th residuum for i-th control loop, N^{C} — number of loop performance assessment indicators selected for analysis, R_i^{CLM} — set of residuals based on loop performance indicators for i-th control loop.

TABLE III

List of residuals for i-th control loops monitoring.

Name	Description based on ...
$r_{i,me2}^{\mathrm{CLM}} = \overline{e_i^2} - \mathrm{const_i}$	mean square of control error
$r_{i,cv.var}^{\mathrm{CLM}} = \sigma_{CV_i}^2 - \mathrm{const_i}$	control signal variance
$r_{i,cv.sat}^{\mathrm{CLM}} = \eta_{sat_i}$	controller saturation
$r_{i,n.pred.e}^{\mathrm{CLM}} = \hat{CV_i} - CV_i$	prediction error of the neural network model of the controller
$r_{i,kp}^{\mathrm{CLM}} = \hat{k}_{p_i} - k_{p_i}$	reconstructed proportional gain of the controller
$r_{i,Ti}^{\mathrm{CLM}} = \hat{T_{i_i}} - T_{i_i}$	reconstructed time constant of the integral action

Below, a brief description of the selected performance indices is given. The residuals employed are listed in Table III. Further details can be found in the works [9, 10].

The mean squared control error is computed as follows

$$\overline{e}^2 = \frac{1}{N}\sum\nolimits_{i=1}^{N} e^2(i), \tag{9}$$

where N is the number of samples. The control signal variance is computed as follows

$$\sigma_{CV}^2 = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left(CV\left(i\right) - \overline{CV}\right)^2, \tag{10}$$

where $\overline{CV}$ is the mean value of CV, and the controller saturation as follows

$$\eta_{sat} = \frac{1}{N}\sum\nolimits_{i=1}^{N} t_{sat}, \tag{11}$$

where t_{sat} takes the value 1 if CV is greater than 90% or lower that 10%, and 0 otherwise.

TABLE IV

List of exemplary/considered information from ICS and IDS.

Source	Notation	Description
ICS	$e^{\text{ics.sf}}_{FP2P}$	error of $FP2P$ sensor
ICS	$e^{\text{ics.sf}}_{FP3P}$	error of $FP3P$ sensor
ICS	$e^{\text{ics.sf}}_{FP4P}$	error of $FP4P$ sensor
ICS	$e^{\text{ics.af}}_{V3}$	actuator $V3$ fault
ICS	$e^{\text{ics.af}}_{V4}$	actuator $V4$ fault
IDS	$e^{\text{ids.sc}}_{CONT3}$	suspicious communication with PLC3 controller
IDS	$e^{\text{ids.sc}}_{CONT4}$	suspicious communication with PLC4 controller

Furthermore, two models, i.e., linear and neural, are utilised to predict the controller output based on the control error values and the past control signal values. The details of the models can be found in references [9, 10]. The coefficients of the linear model are employed to estimate the *proportional–integral–derivative* (PID) controller settings, specifically k_p and T_i.

3.4. Integration with data from ICS and IDS

The last step of the configuration is to take into account such events generated by ICS

$$E^{\text{ICS}} = \{e^{\text{ICS}}_c\}, \quad c = 1, \ldots, N^{\text{C}}, \tag{12}$$

where N^{C} is the number of events reported by ICS and IDS

$$E^{\text{IDS}} = \{e^{\text{IDS}}_d\}, \quad d = 1, \ldots, N^{\text{D}}, \tag{13}$$

where N^{D} is the number of events reported by IDS. The events considered in this work are presented in Table IV.

3.5. Reasoning algorithm

Finally, as a result of ongoing monitoring, a set of residuals R is obtained

$$R = \{r_j\} = R^{\text{PPM}} \cup R^{\text{CLM}}, \quad j = 1, \ldots, N^{\text{J}}. \tag{14}$$

By analysing the time series of residuals for normal operation and taking into account the available expert knowledge, a binary evaluation of residuals with dead zone was selected. Additionally, binary signals were evaluated and filtered against a defined threshold (the number of "1" values in a given time window). Finally, a set of diagnostic signals S is generated

$$S = \{s_j\} = S^{\text{PPM}} \cup S^{\text{CLM}}_i, \tag{15}$$

TABLE V

Basic diagnostics matrix $R^{\text{S.PPM-PF}}$.

	$f^{\text{P}}_{F.P2}$	$f^{\text{P}}_{T.S31}$	$f^{\text{P}}_{CV.31}$	$f^{\text{P}}_{G.V3}$	f^{P}_{V3}	$f^{\text{P}}_{T.S32}$	f^{P}_{SH3}	$f^{\text{P}}_{T.P32}$	$f^{\text{P}}_{F.P3}$	$f^{\text{P}}_{CV.41}$	$f^{\text{P}}_{G.V4}$	f^{P}_{V4}	$f^{\text{P}}_{T.S42}$	f^{P}_{SH4}	$f^{\text{P}}_{T.P42}$	$f^{\text{P}}_{F.P4}$
$s^{\text{PPM}}_{T.S32.1}$		1		()		1										
$s^{\text{PPM}}_{T.P32.2}$						1	1	1	1							
$s^{\text{PPM}}_{T.S42.3}$								1			()		1			
$s^{\text{PPM}}_{T.P42.4}$													1	()	1	1
$s^{\text{PPM}}_{G.V3.5}$			1	1	1											
$s^{\text{PPM}}_{G.V4.6}$										1	1	1				
$s^{\text{PPM}}_{T.S32.7}$		1	1		1	1										
$s^{\text{PPM}}_{T.S42.8}$								1		1			1			
$s^{\text{PPM}}_{F.P3.9}$	1			()			1		1							
$s^{\text{PPM}}_{F.P4.10}$									1		()			1		1

() — potential/possible influence

unisolable process faults

conditionally isolable process faults

TABLE VI

Impact of process faults on the ICS events ($R^{\text{E.ICS-PF}}$).

	$f^{\text{P}}_{F.P2}$	$f^{\text{P}}_{T.S31}$	f^{P}_{CV31}	$f^{\text{P}}_{G.V3}$	f^{P}_{V3}	$f^{\text{P}}_{T.S32}$	f^{P}_{SH3}	$f^{\text{P}}_{T.P32}$	$f^{\text{P}}_{F.P3}$	f^{P}_{CV41}	$f^{\text{P}}_{G.V4}$	f^{P}_{V4}	$f^{\text{P}}_{T.S42}$	f^{P}_{SH4}	$f^{\text{P}}_{T.P42}$	$f^{\text{P}}_{F.P4}$
$e^{\text{ics.sf}}_{FP2P}$																
$e^{\text{ics.sf}}_{FP3P}$									1							
$e^{\text{ics.sf}}_{FP4P}$																1
$e^{\text{ics.af}}_{V3}$					1											
$e^{\text{ics.af}}_{V4}$												1				

for $j = 1, \ldots N^{\text{J}}$ and $i = 1, \ldots, N^O$. In addition, for the purposes of the subsequent reasoning stages, auxiliary signals are determined

$$d^{\text{PPM.or}} = \bigvee_{s_j \in S^{\text{PPM}}_i} s_j, \tag{16}$$

$$d^{\text{CLM.or}}_i = \bigvee_{s_j \in S^{\text{CLM}}_i} s_j, \tag{17}$$

$$d^{\text{LOOPs.or}} = \bigvee_{i=1 \ldots N^O} d^{\text{CLM.or}}_i. \tag{18}$$

Proper diagnostic reasoning consists in determining the diagnosis

$$DGN = F^{\text{P}} \bigcup f_{\text{UPF}} \bigcup d^{\text{CA}} \bigcup D^{\text{CA}}, \tag{19}$$

which includes the determined factors of certainty for the presence of process faults (F^{P}), unknown process fault signal f_{UPF}, and signals for detecting cyberattacks (CA) — global (d^{CA}) and for an individual control loop

$$D^{\text{CA}} = \{d^{\text{CA}}_i\}, \quad i = 1, \ldots, N^O, \tag{20}$$

where d^{CA}_i is a signal of detection of a cybernetic attack in the i-th loop.

During the inference, several relations are used, which are determined based on expert knowledge about the process and configured algorithms, for generating diagnostic signals S^{PPM} and S^{CLM}_i.

TABLE VII
Link between IDS events and control loops ($R^{\text{E.IDS-O}}$).

	o_{31}	o_{32}	o_{41}	o_{42}
$e^{\text{ids.sc}}_{CONT3}$	1	1		
$e^{\text{ids.sc}}_{CONT4}$			1	1

The basic element is the diagnostic relationship $R^{\text{S.PPM-PF}}$ describing the relationship between process faults and observed values of diagnostic signals from the set S^{PPM}. It is presented in Table V. The individual fields contain a set of expected values of diagnostic signals corresponding to specific faults

$$V_{k,j} = R^{\text{S.PPM-PF}}\left(f_k^P, s_j\right) \subset V_j, \tag{21}$$

where V_j is a set of all possible values of the j-th diagnostic signal. Potential influence is depicted by "()" and means the possibility of appearing both "1" and "0" values (connected with OR operator) — it express some kind of uncertainty. One can observe that there are three pairs of indistinguishable faults.

According to the rules derived from the diagnostic relation $R^{\text{S.PPM-PF}}$

$$\left(s_1 \in V_{k,1}\right) \wedge \ldots \left(s_j \in V_{k,j}\right) \wedge \ldots \left(s_{N^{\text{M}}} \in V_{k,N^{\text{M}}}\right) \Rightarrow$$
$$\Rightarrow f_k^{\text{P}}, \tag{22}$$

assuming parallel reasoning [13], preliminary values of the certainty factors of process faults are calculated. They form the set F^{PRIM}

$$F^{\text{PRIM}}\left(s_j \in S^{\text{PPM}}\right) = \{f_k^{\text{P}} = \bigwedge\nolimits_j v_{i,j}\}, \tag{23}$$

where $v_{i,j} \in V_{k,j}$ and $k = 1, \ldots, N^{\text{F}}$.

The indicator of the presence of an unknown process state is determined according to the formula

$$f_{\text{UPF}} = \left(\neg \bigvee\nolimits_{k=1,\ldots,N^K} f_k^{\text{P}} \in F^{\text{PRIM}}\right) \wedge d^{\text{PPM.or}}, \tag{24}$$

expressing the detection of one or more symptoms for $s_j \in S^{\text{PPM}}$, while not indicating any of the process faults.

The general initial signal of a cybernetic attack detection is determined according to the formula

$$d^{\text{CA.PRIM}} = d^{\text{LOOPs.or}} \wedge \left(d^{\text{PPM.or}} \,\text{XNOR}\, f_{UPF}\right), \tag{25}$$

and the initial detection signals for individual control loops

$$d_i^{\text{CA.PRIM}} = d^{\text{LOOPs.or}} \wedge d_i^{\text{CAM.or}}. \tag{26}$$

The final step is to agree on the diagnosis, which takes into account possible events reported by ICS/IDS. The relationship between reported events and process faults and control loops is described by the relations $R^{\text{E.ICS-PF}}$ and $R^{\text{E.IDS-O}}$. These relations can be written in the form of a simple binary matrix. These matrices, for the case under consideration, are presented in Tables VI and VII.

The values of the process faults indicators are checked against the events reports from the ICS

$$F^{\text{P}} = \left\{f_k^{\text{P}}\right\}_{k=1,\ldots,N^K} =$$
$$\begin{cases} 1, & \exists_{e_c^{\text{ICS}}} R^{\text{E.ICS-PF}}(f,c) = 1, \\ f_k^{\text{P}} \in F^{\text{PRIM}}, & \text{otherwise}, \end{cases} \tag{27}$$

where $R^{\text{E.ICS-PF}}(f,c) = 1$ means that the c-th event determines the presence of k-th fault.

Cybernetic attack detection signals are checked against reports from IDS

$$d^{\text{CA}} = \begin{cases} 1, & \exists_{e_d^{\text{IDS}}=1}, \\ d^{\text{CA.PRIM}}, & \text{otherwise}, \end{cases} \tag{28}$$

$$D^{\text{CA}} = \left\{d_i^{\text{CA}}\right\}_{i=1,\ldots,N^O} =$$
$$\begin{cases} 1, & \exists_{e_d^{\text{IDS}}=1} R^{\text{E.IDS-O}}(i,d) = 1, \\ d^{\text{CA.PRIM}}, & \text{otherwise}, \end{cases} \tag{29}$$

where $R^{\text{E.IDS-O}}(i,d)$ means that the d-th event determines cybernetic attack on i-th control loop.

4. Prototype system implementation

The prototype implementation of the inference algorithms was made in Python, and the codes were adapted to operate in the current monitoring mode — processing successive vectors of process data appearing with a fixed sampling period.

The structure of the implementation corresponds to the structure of algorithm, with individual modules performing: determination of model outputs and calculation of coefficients for evaluating indices of control systems, initial filtering of signals, calculation of residuals, evaluation and filtering of residuals, determination of auxiliary coefficients, conducting proper diagnostic inference, determination of detection indices, and finally, reconciliation of diagnosis.

5. Exemplary results

For the final tests of IDCAPF algorithms, 8 cybernetic attack scenarios were selected, one representative of each type of attack [11], and 6 scenarios with simulation of various types of process faults — faults of the measuring paths of auxiliary and regulated quantities, the measuring path of the control signal, faults of the actuators and technological components.

For each of the selected cases, first, a simulation was carried out for the same input signal time series and process data sets covering the operation of the object within 24 h were prepared.

A detailed presentation and discussion of the results obtained are not within the scope of this work. The following is only a summary and discussion of the most important issues:

(i) A practical lack of false detections was demonstrated, as well as a good (minimum 80%) level of correct detection rate of a cybernetic attack and a very good (minimum 95%) level of isolation rate of process faults.

(ii) Detection times, both for process faults and cybernetic attacks, were usually within 5 min. In the case of three cybernetic attack scenarios, the detection time was significantly longer, which resulted from the complex nature of the attack and the dependence of the visibility of attack symptoms on the specific situation in the process; at adverse times, the effects of the attack are so small that they are imperceptible by diagnostic tests. It should be noted that this type of attack has a relatively low negative impact on the course of the technological process. Detection times could be improved in a few cases by narrowing the thresholds for evaluating residuals, but this would result in an increase in false positives.

(iii) The time of isolation of process faults were within 5 min, except for two faults, where they were 502 s and 844 s.

(iv) The agreed detection times of a cybernetic attack have a wider spread. This is due, as mentioned earlier, to the complex nature of some of the attack scenarios and the dependence of the visibility of attack symptoms on the specific situation in the process. A significantly long detection time (about 2 h) was achieved for three scenarios. In these cases, the system is dealing with "poorly detectable" scenarios — those that have little impact on the course of the process and are poorly distinguishable from process damage. In each of these cases, detection of the abnormal condition is relatively fast, while the agreed signal for detecting a cyberattack appears only after the IDS generates an appropriate message. Faster detection of a suspicious condition by the IDS would speed up detection based on the agreed detection signal.

(v) In all cases, the differentiation of process faults was achieved, both in relation to each other and from the diagnosis of the detection of a cybernetic attack. Cybernetic attack detection signals for the individual loops corresponded quite well to the real situation.

(vi) The use of signals from ICS and IDS made it possible to clarify the diagnosis, both in terms of distinguishing a cybernetic attack from process faults and specifying the place of the cybernetic attack introduction.

6. Conclusions

The proposed model of monitoring and diagnostic system, consisting of the reasoning algorithm and prototype system modules, has demonstrated its effectiveness through the pilot application on a steam line simulator. The system successfully integrates parametric partial models and dedicated indices for monitoring control loops performance, providing a robust framework for anomaly detection. Integration of data from ICS and IDS significantly enhances the precision of fault isolation and the differentiation between process faults and cybernetic attacks. This integration allows for a comprehensive diagnostic approach, leveraging both process and cybernetic fault data.

The proposed inference algorithm, Integrated Diagnostics of Cybernetic Attack and Process Faults (IDCAPF), effectively combines various types of residuals and diagnostic signals. This versatility ensures that the system can handle a wide range of fault scenarios, both process-related and cyber-related.

The use of partial parametric models, including ARX autoregressive linear models and neural models, proves to be a practical and efficient approach for process fault detection. Knowledge related to the relationships describing physical phenomena occurring in the process, indirectly contained in the PPM models, can be effectively used to detect not only process faults, but also the effects of cybernetic attacks.

The model's design, both the algorithm structure and the system modules, allows for scalability and adaptability to different industrial processes. By adjusting the models and diagnostic algorithms, the system can be tailored to various types of control loops and process configurations, making it a versatile tool for industrial diagnostics. Of course, the final accuracy of diagnosis of process faults and the ability to detect cybernetic attacks strongly depend on the ability to design a specific set of residuals, as well as on the available events obtained from ICS and IDS.

While the prototype system implementation for the selected process, i.e., part of steam line, has shown promising results, further implementation on target platforms is necessary to fully realize its potential. The availability of comprehensive data from ICS and IDS will be crucial during this phase, ensuring that the system can operate effectively in real-world industrial environments.

Acknowledgments

The work was partially supported by the grant no. POIR.01.01.01-00-0541/20.

References

[1] M. Blanke, J. Kinneart, M. Lonze, M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*, Springer, Berlin Heidelberg 2002.

[2] R. Isserman, *Fault-Diagnosis Systems*, Springer, Berlin Heidelberg 2005.

[3] J.M. Kościelny, M. Syfert, P. Wnuk, in: *Advanced Solutions in Diagnostics and Fault Tolerant Control, (DPS 2017)*, Eds. J. Kościelny, M. Syfert, A. Sztyber, Vol. 635, Springer, Cham 2018 p. 449.

[4] G. Béla, I. Kiss, P. Haller, *Int. J. Crit. Infrastruct. Prot.* **10**, 3 (2015).

[5] M. Bauer, A. Horch, L. Xie, M. Jelali, N. Thornhill, *J. Process Control* **38**, 1 (2016).

[6] T.J. Harris, *Can. J. Chem. Eng.* **67**, 856 (1989).

[7] A. Ordys, D. Uduehi, M.A. Johnson, *Process Control Performance Assessment*, Springer-Verlag, London 2007.

[8] M. Jelali, *Control Performance Management in Industrial Automation*, Springer-Verlag, London 2013.

[9] A. Sztyber, Z. Górecka, J.M. Kościelny, M. Syfert, in: *Intelligent and Safe Computer Systems in Control and Diagnostics*, Ed. Z. Kowalczuk, Springer International Publishing, Cham 2023, p. 100.

[10] A. Sztyber-Betley, M. Syfert, J.M. Kościelny, Z. Górecka, *Sensors* **23**, 2778 (2023).

[11] M. Syfert, A. Ordys, J. M. Kościelny, P. Wnuk, J. Możaryn, K. Kukiełka, *Energies* **15**, 6212 (2022).

[12] T. Hägglund, *Control Eng. Pract.* **13**, 1383 (2005).

[13] J. Korbicz, J.M. Kościelny, Z. Kowalczuk, W. Cholewa, *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag Berlin, Heildelberg 2004.