

Technological process monitoring system on the basis of artificial intelligence technology

M.I. Arpishkin¹, A.M. Vulfin¹, V.I. Vasilyev¹, A.V. Nikonov¹

¹Ufa State Aviation Technical University, K. Marks St. 12, Ufa, Russia, 450077

Abstract. The relevance of the work is due to the widespread use of network telecommunications in the automated process control system and the high level of danger of replacing, distorting or losing accumulated data on the process progress as a result of the attacker's influence. Objective: increasing the security of measurement results from unauthorized modification in the databases of information systems of an industrial enterprise by improving process monitoring system based on the intellectual analysis of technological time series. A structural scheme of process monitoring as part of an information protection system in a segment of an automated process control system network has been developed. The algorithm of intellectual analysis of technological time series in the task of detecting violation of the integrity of data on the process progress due to their unauthorized modification was proposed. Evaluated the effectiveness of the proposed solution on field data.

1. Introduction

Modern automated control systems (ACS) manage complex and dangerous technological processes, failure in which can lead to accidents at work or man-made disasters. This significantly increases the cost of the risks of violating information security, since the realization of threats can lead to harm to people's life and health, environmental damage, and financial and reputational losses [1].

Current priority in ensuring the information security of the automated process control system is ensuring the integrity and availability of configuration and control information and information on the parameters of the technological process.

One of the key decisions in the field of information security is a means of correlation and monitoring. When it comes to the technology segment of enterprises where round-the-clock monitoring of industrial systems is necessary, in addition to the above functionality, these solutions can and should become one of the components of an integrated situational center, where a single point of tracking is formed, both technical systems performance and process control, and monitoring, and enterprise information security management.

Currently, there are a variety of tools and instruments that ensure information security in the composition of the automated process control systems (ACPS) and automated control systems at various levels of protection [2]:

1. At the perimeter level of the corporate network (network monitoring systems (IDS / IPS, DLP), protocol analyzers, firewalls);
2. At the level of systems (antivirus, physical protection);
3. At the level of communication channels (cryptographic tools);
4. At the level of application software (authorization tools, mandatory management, selective management, role-based management, auditing).

Existing tools have several disadvantages:

1. The analysis of parameters of information flows and computational processes is carried out without taking into account the semantics of the protected data;
2. Data protection is not ensured in the case when the user legally has access to them, and make unauthorized modification;
3. Do not provide data protection in the case when an attacker gets full access to the ACS network;
4. Information protection systems (IPS) are universal for information systems, and do not take into account the specifics of the data and the subject area in which the information system operates.

The object of research is the information security of data, which are the results of measurements of parameters of technological objects, from the threat of unauthorized modification of information.

The subject of research is the algorithms for detecting data integrity violations in the databases of information systems of an industrial enterprise.

The goal of the study is to increase the security of results of measurement from unauthorized modification in the databases of information systems of an industrial enterprise by improving the technological process monitoring system based on the intellectual analysis of technological time series.

To achieve this goal, the following tasks were formulated:

- Development of a structural scheme of technological process monitoring as part of an information protection system in a segment of a network of ACPS;
- Development of algorithms for intellectual analysis of technological time series in the task of detecting violations of the integrity of data about the technological process in the form of their unauthorized modification;
- Implementation of the proposed analysis algorithms in the form of a software module of the information protection system in the ACPS network segment and assessment of the effectiveness of proposed solution on full-scale data.

2. Analysis and development of a process monitoring system as a component of intrusion detection system

Taking into account the requirements for continuity of technological processes, it is necessary to use passive intrusion detection systems, which will analyze network traffic without interfering with data transfer processes. Also, development of intrusion detection system components capable of detecting distortions and falsification of data transmitted in industrial networks is relevant.

Monitoring system – hardware and software complex that performs continuous measurement of environmental parameters and technological processes at controlled facilities, records events that occur, warns about unacceptable deviations of parameters, signals emergency situations, collects and archives data, generates reports. Process monitoring system, implemented as a component of an intrusion detection system (IDS), implies continuous monitoring of technological process parameters to identify significant deviations from “normal behavior”, which in turn will indicate possible malicious intentions. This approach is the development of the concept of Data Centric Security, which implies the security of the data itself. To determine deviations from the “normal behavior” of the TP, a system model will be used, which is the concept of Fault Detection and Identification [3].

The main problems of IDS used in industrial automation systems are [4]:

- False system alarms;
- The inconsistency between the intelligence of the algorithm (required computational and time resources) of intrusion detection and system performance;
- The contradiction between ease of administration and detection accuracy;

Examples of the use of intelligent methods for processing technological process monitoring data as part of IDS shown in Table 1.

The enterprise in which polyethylene terephthalate is performed is classified as hazardous production facilities in accordance with Federal Law No. 116-FL. It is proposed to create a system for monitoring

the process of polyethylene terephthalate production based on the mathematical model of the process discussed in [9, 10, 11, 12].

Table 1. Examples of the use of intelligent methods for processing technological process monitoring data as part of IDS.

Example of the system	Intelligent analysis method	Attack Features
Water management system [5]	Three-level neural network with feedbacks. Accuracy of detecting attacks from 8.2 to 45%	Signs of an attack were extracted from the Modbus protocol traffic; Network cyber attacks: replay attacks, MITM, DoS; The counterfeit attacks of the following parameters were synthesized: negative water level, water level higher / lower than the maximum possible; water level above / below the maximum allowed; random level
Power management system [6]	Immune Algorithms (CLONALG, AIRS2Parallel) Accuracy of detecting attacks: 80.1 - 99.7%	Synthesized attacks of falsification of data flow parameters in the engineering network nodes
Power management system [7]	N-gram method	System load (power) for 6 power network buses — rare errors were made to the data, including change of sign, degree, numbers of the observed value
Power management system [8]	Evaluation of the prediction of system states.	Data: voltage and phase measured by sensors are becoming fake

To build the model, technological signals are used that carry the signs of events occurring at the object. Signal analysis requires identification of fragments associated with individual events and the study of relevant events by segmentation of the initial technological time series and subsequent analysis of the corresponding signal wave using characteristics such as amplitude, waveform (morphology), duration, intervals between events, energy distribution, frequency content, etc. Consequently, the detection of events is one of the most important goals of analyzing technological signals when solving the problem of diagnosing an object and identifying abnormal conditions associated with the actions of an attacker. Thus, the development of a model of analysis of technological signals is the task of identifying discrete epochs of a technological signal and relating them to events and incidents characterizing the state of objects [13-16].

The main functional connections between the decision maker on the process, the process monitoring system and the components of the automated control system are shown in Figure 1.

The initial data for the analysis are the following parameters:

- Pressure at the pump inlet p_{BC} ;
- rotor speed n ;
- The current I on the motor pump from the sensor I_1 .

Method of measuring the fluid viscosity is the following (1):

$$\mu(t) = A \cdot \frac{I}{n} + B \cdot p_{BC} + C, \tag{1}$$

where A, B, C – constant coefficients; p_{BC} – inlet pressure, Pa; n – rotor speed, rpm; I – current on the motor pump, A.

The characteristic viscosity μ of the controlled fluid is calculated at the measured temperature on the measured values of the operating parameters n, p_{BC}, I . Viscosity control is performed continuously in

dynamic mode. Measured values from all sensors are fed to the dispatching system for collecting and managing the process.

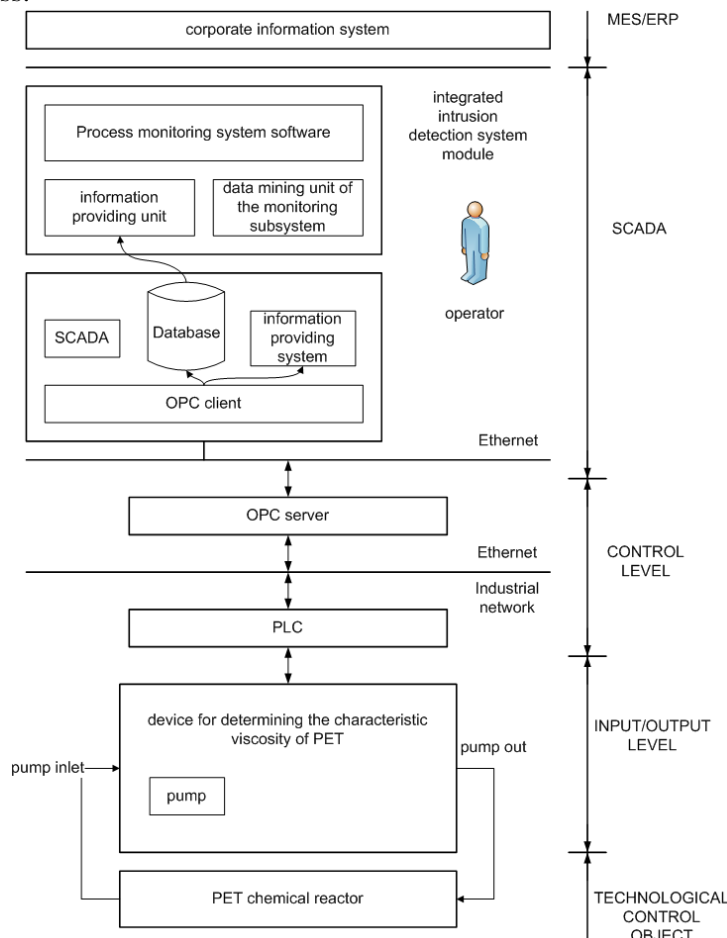


Figure 1. Structural scheme of the interaction of the process monitoring system with the components of the process control system.

3. Development of algorithms for intellectual analysis of technological time series in the task of detecting deviations from the normal operation mode

The task of modeling the time series of the considered technological process in general form can be formulated as follows.

Let the values of the time series be given $Y = \{y(1), y(2), \dots, y(N)\}$, where $y(t)$ – the value of the indicator of the process under investigation, registered in the t -th time step ($t = 1, 2, \dots, N$). It is required to construct estimates of the future values of the series $\hat{Y} = \{\hat{y}(N + 1), \hat{y}(N + 2), \dots, \hat{y}(N + \tau)\}$, $1 \leq \tau \leq N$, where τ is the forecast horizon [14].

The general statistical model of a numerical time series is the model of the form (2):

$$y_t = f(x_t, a) + \varepsilon_t \tag{2}$$

where y_t – observed time series; $f(x_t, a)$ – systematic component; ε_t – random component.

Controlled parameters of the technological process of production of polyethylene terephthalate are manifestations of non-stationary processes, being an example of dynamic systems:

- variable operating conditions of the equipment when the environmental parameters change dramatically (for example, switching the operating mode of the installation);
- long period of operation of components and assemblies, when the dynamic characteristics of the system change, and the same input causes different reactions of the system.

A time interval can be allocated at which the object’s parameters change insignificantly, and are considered conditionally constant. Thus, the division of the technological time series (TTS) into intervals is reduced to the construction of a change type detector of dynamics describing the state of an

object. In the operating mode, the operation of a technological object is characterized by stationarity, stability of parameter values and constancy of development over time. Most of the operating modes characteristic of an object are steady-state processes. The cumulative transient time in the network T_{per} is substantially less than the total time T of the network. The chronology of the technological object can be viewed as a temporary sequence of static modes, replaced by relatively short transients.

TTS models need to be built on a set of TTS samples, in which model parameters are assumed conditionally stationary. The size of the analysis window depends on the nature of the distribution of TTS parameters and when processing samples that hit the window, 5-50 nearby observation points are used.

The analysis uses the TTS data formed by the results of observations of the input $u(t)$ and the output $y(t)$, which is shown in Figure 2, where $u(k)$, $y(k)$ is the value of the input and output at the k -th instant of time $t = kT$; T - time sampling period; L is the size of the time window.

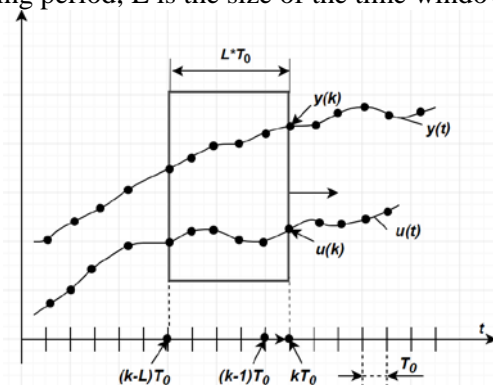


Figure 2. TTS counts falling into the sliding window $u(t)$ and $y(t)$.

Consequently, a sample can be formed containing $(L+1)$ a pair of input-output samples (3):

$$\{(u(k-L), y(k-L)); \dots; (u(k-1), y(k-1)); (u(k), y(k))\} \tag{3}$$

To obtain a nonlinear adaptive model $y=F(u)$ of the object according to the input/output data, it is proposed to use the neural network model. For this, the output of the object $y(t)$ compares with the output of the neural network (NN) $\hat{y}(t)$ with the same input effect $u(t)$, and the procedure of training consists in changing the weights of its connections in such a way that to reduce the mismatch $\varepsilon(t) = y(t) - \hat{y}(t)$ to an acceptable (fairly small) value.

Taking into account the above features of the TTS formed by the object parameters, the suitable analysis models are the NARX model (Nonlinear autoregressive with exogenous inputs), allowing to take into account nonlinear dynamics type change processes (4):

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \dots, u_t, u_{t-1}, u_{t-2}, u_{t-3} \dots) + \varepsilon_t, \tag{4}$$

where y is the desired variable, and u is an external defined variable, ε_t is white noise.

4. Development of a TTS adaptive segmentation algorithm

The algorithm is based on the work of Bodenstern and Pretorius (1977) and uses the construction of TTS models in sliding “windows”:

1. The model is built for some initial reference TTS section;
2. Results of the model prediction are compared with the remaining TTS counts coming through a sequentially sliding window moving through the TTS;
3. If the characteristics of the TTS in the reference section and in the moving window differ by more than a certain threshold, the boundary of the segment is taken, immediately after which a new reference section is taken, and the procedure is repeated;
4. Segmentation ends when a moving window reaches the end of a row [8].

To evaluate the deviation of the simulated values of the real data a parametric global method is used:

- The countdown in the current sliding window is selected, which divides the TTS into two segments;

- The empirical estimate of the discrepancy between model values and field data for each segment is calculated;
- At each point of the segment, the magnitude of the deviation of the empirical estimate from the model values is measured. Deviations for all points are calculated;
- The total residual in each segment is calculated;
- The relocation of the dividing point is performed until the total residual error is minimized.

5. Building a training set

For the training and test set of samples, the same data used to build the mathematical model and the data of the segment types are used. The training and test samples are taken in the ratio 70/30. In the analysis, TTS data are used, formed by the results of observations of the input $u(t)$ and the output $y(t)$. Therefore, a sample can be formed containing $(L+1)$ a pair of input-output samples (6):

$$\left\{ \left\{ (u(k-L), y(k-L)); \dots; (u(k-1), y(k-1)); (u(k), y(k)) \right\}, C_m \right\} \quad (5)$$

where C_m is the class of a known type of event set in correspondence for the current segment, which is used to build the next sample of the set.

6. Development of algorithms for intellectual analysis of technological time series in the task of detecting violations of the integrity of data about the technological process in the form of their unauthorized modification

The scheme of realization of the learning process of the NARX model is presented in Figure 3. A time series of process parameters U is fed to the input of the model, where $u_1(t), u_2(t), u_3(t)$ are the current, the number of rotor turns and pump suction pressure in accordance with (1), to train the NN to predict the technological process within the adaptive window and the division of the series into segments. The viscosity of the fluid $y(t)$ is fed to the output of the NN for comparing the predicted value of viscosity with the real value, if the values do not match, then a new segment is created.

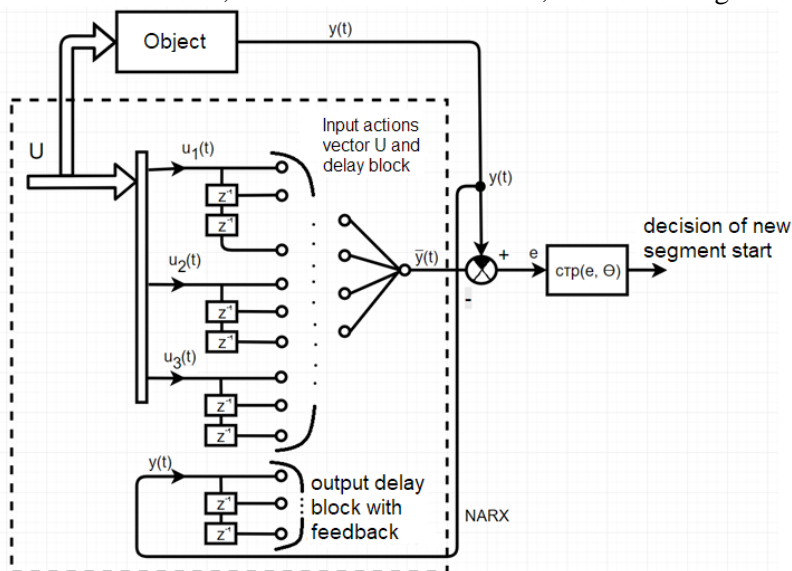


Figure 3. Scheme of the implementation of the learning process NARX model.

The operation algorithm of the process monitoring system consists of the following main steps:

1. TTS adaptive segmentation;
2. Merging segments and clustering the remaining segments by the parameters of models of the TTS selected sections;
3. Comparison of the history of the state of the technological object and TTS segments;
4. Training of the classifier analyzing the dynamics of the current window of the analyzed parameters;
5. Deciding on the type of technical condition of the technological object.

This algorithm performs the process of obtaining data on the course of technical process in the form of a TTS, segmentation of a given TTS and the merging of segments according to a similar type of TTS behavior. Next comes the training of the classifier, which analyzes the dynamics of the process and decides on the type of technological state of the technological object.

7. Development of a block diagram of a system for monitoring the technological process of polyethylene terephthalate production

A structural scheme of the process monitoring system for the production of polyethylene terephthalate as part of an intrusion detection system has been developed (Figure 4).

The structural scheme of the monitoring system of production process consists of six blocks interconnected with each other. The first block is a block for preparing TTS data, to which a time series of process parameters is fed, and data is generated for further work with them. The next block is a block for constructing a mathematical model of a technological object on the basis of the analysis of TTS using the NARX model, which allows to take into account non-linear processes of changing the type of dynamics. The third block is the block for analyzing TTS. In this block, there is an adaptive segmentation of the time series and the integration of similar segments by types using the k-means clustering method.

The fourth block is the monitoring system management module. It receives data from the first three blocks, it is the link to transfer information to the database for storing, and submitting information (training parameters) to the next block - a classifier block for identifying known types of events on an object. The task of this classifier is to learn how to classify for each type of segments the events occurring at the facility, including the detection of violation of the integrity of data on the progress of the process due to it unauthorized modification).

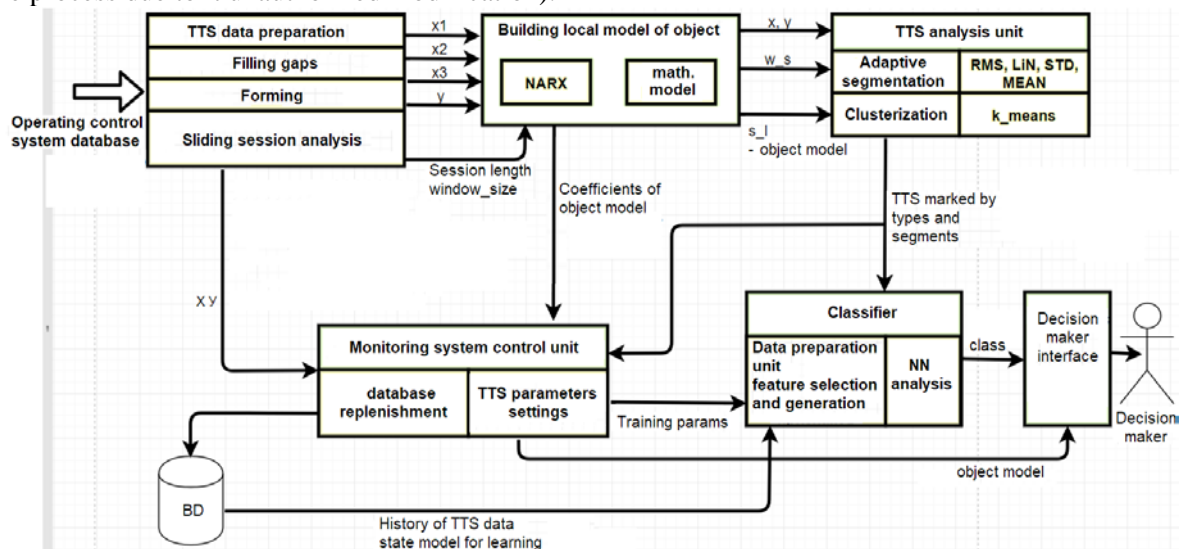


Figure 4. Structural scheme of a process monitoring system as part of an information protection system in a segment of an APCS network.

8. Conducting an experiment on full-scale data using the software implementation of the process monitoring system

Having a scheme for implementing the learning process of the NARX model for adaptive segmentation of TTS, an experiment was performed in the MATLAB application package. For the NN learning process, data were taken on the course of production process of polyethylene terephthalate for the year, namely viscosity, current, rotor speed and suction pressure: $y(t)$, $u_1(t)$, $u_2(t)$, $u_3(t)$.

Figure 5 shows in graphical form the technological time series for the input parameters $u_1(t)$ - the current strength, $u_2(t)$ - the rotor speed, $u_3(t)$ - the suction pressure and the output parameter $y(t)$, the polyethylene terephthalate viscosity.

The selection of the number of TTS segments for each of the TTS parameters is implemented according to the criterion for finding the inflection point on the graph of the dependence of the total discrepancy of the model and field data $y(t), u_1(t), u_2(t), u_3(t)$ (Figure 6).

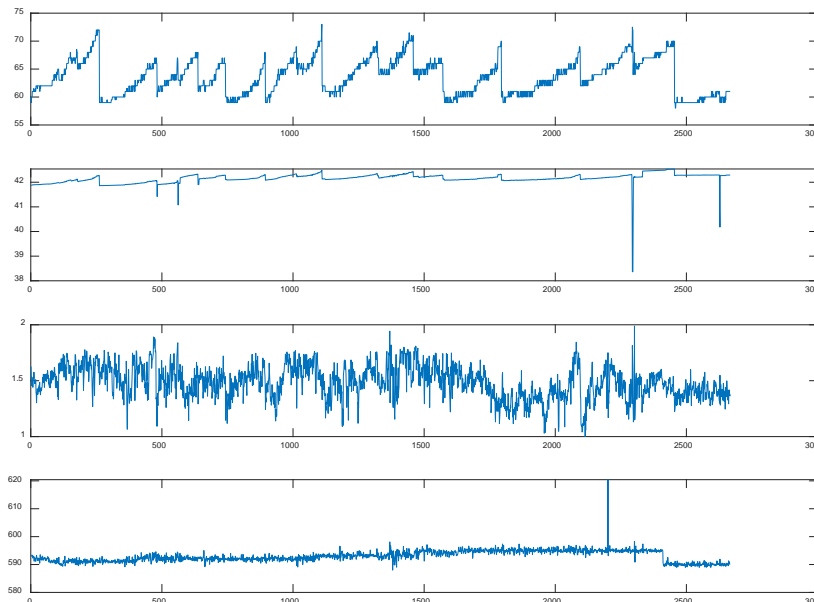


Figure 5. TTS input - $u_1(t), u_2(t), u_3(t)$ and output characteristics of the technological control object, $y(t)$.

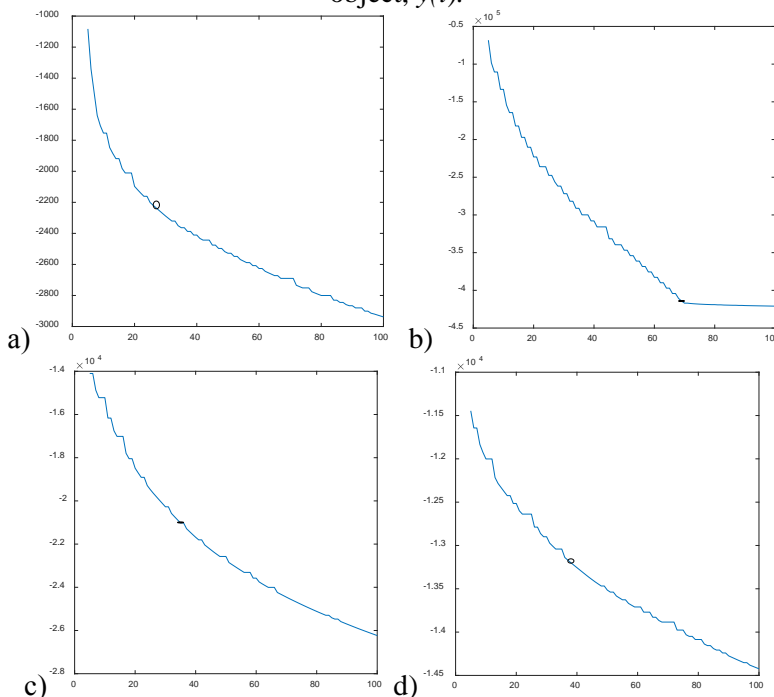


Figure 6. Dependence of the total discrepancy of the model and field data $y(t)$ on the number of segments (X axis) by the search criterion of the inflection point (knee of curve) (a) $k = 27$, b) $k = 69$, c) $k = 35$, d) $k = 38$).

To verify the correctness of finding the number of segments using the NARX model, an analysis was performed for each of the parameters. According to the results of the summary analysis, the value of the number of segments $k = 27$ is selected, which coincided with the results of the NARX model.

9. Clustering of received TTS segment

The resulting segments are combined according to “similar” types of dynamics using the implemented clustering algorithm in the MATLAB package, namely the k-means clustering method according to the Calinski-Harabasz criterion (Fig. 7) [9].

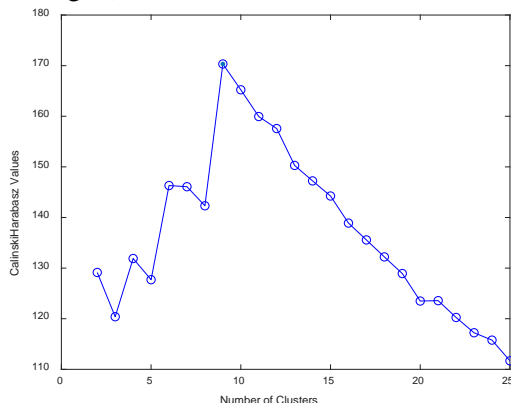


Figure 7. The optimal number of clusters for the k-means algorithm according to the Calinski-Harabasz criterion.

The total number of clusters after the merging of similar segments was 9. In Figures 8 and 9, the difference is visible before the clustering of the segments. As a result of clustering, 27 adaptive segments were divided into 9 classes of TTS dynamics, which would greatly simplify the training of a neural network to classify an event occurring at an object according to the type of TTS segment.

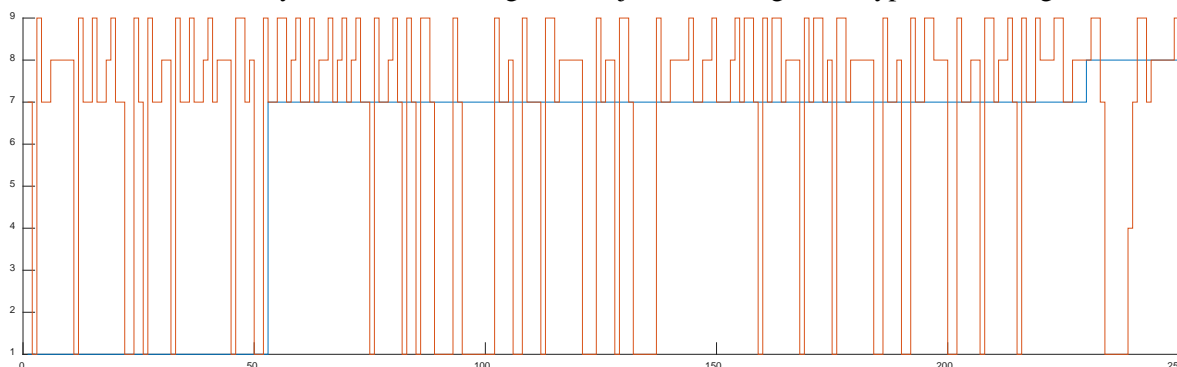


Figure 8. The first 250 samples of TTS, related to the selected clusters. The orange line - without merging neighboring, blue - taking into account the value of neighbors. X axis - reference number, Y axis - class number.

10. Construction and training of neural network classifier

To implement the classifier, which will analyze the dynamics of the process and make a decision about the type of technological state of the maintenance, a model of a multilayer perceptron was chosen. 1330 examples of 10 samples were taken for training the NN in a sliding window with overlap of 2 and 9 dynamics classes, the sample was divided in a ratio of 75 by 25. A 30 neurons was selected for the hidden layer during the experiments. 5000 epochs were chosen for training, the activation function is hyperbolic tangent. RMS error was used to estimate the network error. The conjugate gradient algorithm was used as the learning algorithm. The target learning error goal = 1e-3 was reached in 2740 iterations.

For a training sample, specificity and sensitivity are equal to:

Sensitivity = 1; Specificity = 0.998792.

Inaccuracy matrix for the training set is shown in Table 2.

Consequently, the accuracy of the classifier on the training sample was 99.79%.

For the test sample, the specificity and sensitivity are:

Sensitivity = 0.83871;
 Specificity = 0.98155.

Inaccuracy matrix for the test set is shown in Table 3.

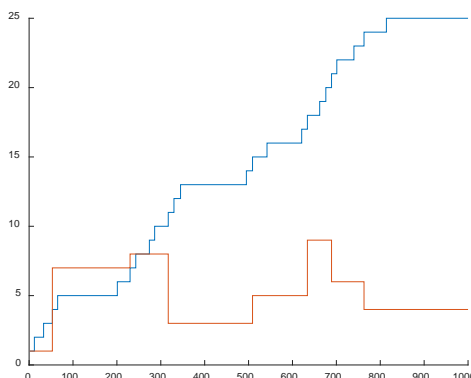


Figure 9. TTS counts related to selected clusters. The blue line is the initial numbering of the clusters (27 adaptive segments), the orange line is based on the value of the neighbors (9 types of segments). The X axis is the reference number, the Y axis is the class number.

Table 2. Inaccuracy matrix for the training set.

Class		Predicted							
		1	2	3	4	5	6	7	8
Actual	1	169	0	0	0	0	0	0	1
	2	0	1	0	0	0	0	0	0
	3	0	0	11	0	0	0	0	0
	4	0	0	0	0	0	0	0	0
	5	0	0	0	0	1	0	0	0
	6	0	0	0	0	0	303	0	0
	7	0	0	0	0	0	0	241	0
	8	0	0	0	1	0	0	0	269
	9	0	0	0	0	0	0	0	0

Table 3. Inaccuracy matrix for the test set.

Class		Predicted					
		1	2	3	4	5	6
Actual	1	52	1	0	3	0	1
	2	0	0	0	0	0	0
	3	1	0	0	0	0	0
	4	4	0	3	83	3	2
	5	1	0	0	3	77	4
	6	4	0	0	7	3	81
	7	0	0	0	0	0	0

Hence, the accuracy of the classifier on the test sample was 87.99%. Thus, for a test sample, the accuracy of the classifier turned out to be equal to 87.99%, and the probability of error of individual classifiers of types does not exceed 14%. It can be concluded that the implementation of this process monitoring system based on artificial intelligence technology will increase the degree of protection of measurement results from unauthorized modification in databases of information systems of an industrial enterprise.

11. Conclusion

As a result of the research and development carried out in the work, the following main results were obtained:

- A structural scheme of the process monitoring system was developed as part of an information protection system in a segment of the APCS system;
- An algorithm was developed for intellectual analysis of technological time series in the task of detecting violations of the integrity of data about the technological process in the form of their unauthorized modification;
- The proposed algorithm for the analysis of TTS was implemented as a software module of an information protection system in a network of APCS;
- An assessment of the effectiveness of the proposed solution on full-scale data was carried out, correctly recognizing the type of technological state of technical maintenance associated with the violation of the integrity of information on the progress of the technical process in 88% of cases.

The proposed process monitoring system, analyzing the dynamics of the technological process, allows to detect the attacker's impact on the process flow and on the components of the information system by comparing the process model and current indicators - ensuring the information security of the facility. The proposed system does not require the use of additional equipment, it is enough to install software.

12. References

- [1] APCS [Electronic resource]. – Access mode: https://ru.wikipedia.org/wiki/Автоматизированная_система_управления_технологическим_процессом (03.12.2018).
- [2] Basic processes for the protection of APCS [Electronic resource]. – Access mode: <https://www.intuit.ru/studies/courses/697/553/lecture/12442> (03.12.2018).
- [3] IS Issues of Industrial Networks [Electronic resource]. – Access mode: <https://www.iemag.ru/master-class/detail.php?ID=34562> (03.12.2018).
- [4] Kort, S. Features of intrusion detection in APSC. – Moscow: Kaspersky Lab.
- [5] Gao, W. On SCADA control system command and response injection and intrusion detection / W. Gao, T. Morris, B. Reaves, D. Richey // Proceedings of the 5th Annual Anti-Phishing Working Group eCrime Researchers Summit (eCrime). – Dallas, 2010. – P. 1-9.
- [6] Zhang, Y. Distributed intrusion detection system in a multi-layer network architecture of smart grids / Y. Zhang, L. Wang, W. Sun, R.C. Green, M. Alam // IEEE Transactions on Smart Grid, 2011. – P. 796-808.
- [7] Bigham, J. Safeguarding SCADA systems with anomaly detection / J. Bigham, D. Gamez, N. Lu // Computer Network Security. – 2003. – Vol. 2776. – P. 171-182.
- [8] He, Q. Smart grid monitoring for intrusion and fault detection with new locally optimum testing procedures / Q. He, R.S. Blum // Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 2011. – P. 3852-3855.
- [9] Federal Law of July 26, 2017 № 187-FZ “On the Security of the Critical Information Infrastructure of the Russian Federation, 2018. (in Russian).
- [10] Classification of automated systems subject to protection against unauthorized access to information, and requirements for protection of various classes in the AU [Electronic resource]. – Access mode: <http://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty/114-tekhnicheskaya-zashchita-informatsii/dokumenty/spetsialnye-normativnye-dokumenty/384-rukovodyashchij-dokument-reshenie-predsedatelya-gostekhkommisii-rossii-ot-30-marta-1992-g> (03.12.2018).
- [11] Telyashev, E.G. Determination of the characteristic viscosity of polyethylene terephthalate by controlled parameters of the pump / E.G. Telyashev, I.M. Arpishkin // World of Oil Products. The Oil Companies' Bulletin. – 2018. – Vol. 4. – P. 27-30.
- [12] Astakhov, A.A. Features of information security of industrial systems // CISA. – 2006. – Vol. 3. – P. 76-79.
- [13] The time series forecasting model for the maximum similarity sample [Electronic resource]. – Access mode: <http://www.mbureau.ru/articles/dissertaciya-model-prognozirovaniya-vremennyh-ryadov-glava-1> (03.12.2018).

- [14] Adaptive time series segmentation algorithm [Electronic resource]. – Access mode: <http://brain.bio.msu.ru/shishkin/thesis/review6a.htm> (03.12.2018).
- [15] Galkin, A.P. Protection of communication channels of enterprises and institutions from unauthorized access to information: tutorial. – Vladimir: VIGU Publisher, 2003. – 128 p.
- [16] Jingfei, Y. Power System Short-term Load Forecasting: Thesis for Phd degree. – Darmstadt, 1974.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research, research № 17-48-020095.