DOI: 10.15514/ISPRAS–2020–32(5)–10

Heterogeneous Data Aggregation and Normalization in Information Security Monitoring and Intrusion Detection Systems of Large-scale Industrial CPS

M.A. Poltavtseva, ORCID: 0000-0001-9659-1244 <poltavtseva@ibks.spbstu.ru>
Peter the Great St. Petersburg Polytechnic University,
29, Politechnicheskaya st., Saint Petersburg, 195251, Russia

Abstract. Monitoring of industrial cyber-physical systems (CPS) is an ongoing process necessary to ensure their security. The effectiveness of information security monitoring depends on the quality and speed of collection, processing, and analyzing of heterogeneous CPS data. Today, there are many methods of analysis for solving security problems of distributed industrial CPS. These methods have different requirements for the input data characteristics, but there are common features in them due to the subject area. The work is devoted to preliminary data processing for the security monitoring of industrial CPS in modern conditions. The general architecture defines the use of aggregation and normalization methods for data preprocessing. The work includes the issue from the requirements for the preprocessing system, the specifics of the subject area, to the general architecture and specific methods of multidimensional data aggregation.

Keywords: security monitoring; CPS; data processing; data aggregation; normalization; distributed systems; streaming data processing; security methods; traffic analysis; data analysis; hierarchical aggregation

For citation: Poltavtseva M.A. Heterogeneous Data Aggregation and Normalization in Information Security Monitoring and Intrusion Detection Systems of Large-scale Industrial CPS. Trudy ISP RAN/Proc. ISP RAS, vol. 32, issue 5, 2020, pp. 131-142. DOI: 10.15514/ISPRAS-2020-32(5)-10

Acknowledgements. The reported study was funded by Russian Ministry of Science (information security), project number 2/2020

Агрегация и нормализация гетерогенных данных в системах мониторинга информационной безопасности и обнаружения вторжений крупномасштабных промышленных КФС

М.А. Полтавцева, ORCID: 0000-0001-9659-1244 <poltavtseva@ibks.spbstu.ru>
Санкт-Петербургский политехнический университет Петра Великого,
195251, Россия, г.Санкт-Петербург, ул.Политехническая, дом 29

Аннотация. Мониторинг информационной безопасности промышленных киберфизических систем (КФС) является постоянным процессом, необходимым для обеспечения их безопасности. Эффективность мониторинга зависит от качества и скорости сбора, обработки и анализа гетерогенных данных КФС. Сегодня существует много методов анализа для решения задач безопасности распределенных промышленных киберфизических систем. У этих методов разные требования к характеристикам входных данных, но есть общие особенности, обусловленные предметной областью. Работа посвящена предварительной обработке данных при мониторинге безопасности промышленных КФС в современных условиях. Задача рассматривается от требований к системе предварительной обработки данных и особенностей предметной области, до специфических методов агрегации, основанных на PCr – связях между структурами данных. Разработана архитектура обработки данных сочетающая методы агрегации и нормализации информации в системе мониторинга.

Ключевые слова: мониторинг безопасности; КФС; обработка данных; агрегация данных; нормализация, распределенные системы; потоковая обработка данных; методы обеспечения безопасности; анализ трафика; анализ данных; иерархическая агрегация

Для цитирования: Полтавцева М.А. Агрегация и нормализация гетерогенных данных в системах мониторинга информационной безопасности и обнаружения вторжений крупномасштабных промышленных КФС. Труды ИСП РАН, том 32, вып. 5, 2020 г., стр. 131-142 (на английском языке). DOI: 10.15514/ISPRAS–2020–32(5)–10

Благодарности. Исследование выполнено при финансовой поддержке Минобрнауки России (грант ИБ) в рамках научного проекта № 2/2020

1. Introduction

The modern economy requires not only stability and efficiency of enterprises, but also a high degree of modern information technologies application. In developed complexes, they cover all areas of activity, from basic actions and control of physical processes to the corporate and business segments. Intelligent computerized devices that control physical processes are commonly referred to as cyber physical objects. A set of such devices that operate in a distributed network with shared information management is called a cyber-physical system.

Despite the separation of corporate and industrial networks [1], attacks on the physical process through the corporate segment and in opposite directions remain highly relevant [2]. At the same time, the number of attacks on industrial systems and networks is constantly increasing. These are both mass and targeted (APT – advanced persistent threat) attacks. The variety of attacks, the attacker's tools, and their goals is increasing. Threats to the confidentiality of business and industrial data are not only relevant for industrial networks. It is also important to maintain the correctness of the physical process.

Protection of industrial systems is impossible without an effective solution to the monitoring problem, in order to detect intrusions into the CPS, register anomalies of processes taking into account possible malicious impact and distortion of recorded data, analyze threats and investigate security incidents [3]. Monitoring of information security of industrial CPS should not only allow solving security problems, but also do it in a timely manner [4, 5]. To do this, the data on which decisions are made must be processed and provided to the analysis tools with sufficient completeness and speed.

The work is devoted to preliminary data processing in industrial cyber-physical systems security monitoring to ensure data completeness and timeliness, taking into account the main mathematical analysis methods used in solving problems of intrusion detection and anomalies of controlled processes.

2. Related Works

Security monitoring for industrial cyber-physical systems are based on various types of input data. These are network traffic [6-8] and physical node data [9, 10]. Today, there are a large number of works aimed at analyzing network traffic in monitoring systems to solve various security problems [11-13]. Recently, this field has been dominated by works using streaming information processing technologies [14] or (earlier) combining streaming and batch technologies [6, 15]. In the field of data preprocessing methods, these works use traffic aggregation by parameters with the extraction of indicators without additional approaches to improve efficiency [15]. The authors of many works, for example [16, 17], focus on working with high-level features, adapting the processing system to the specific analysis method they use. However, the General list of methods used in industrial CPS information security monitoring is quite large [8, 9, 18-22]. Each method has its own advantages

and disadvantages. Therefore, the greatest efficiency can be achieved by using a combination of analysis methods.

Let's look at modern methods of aggregation and normalization of streaming data. Data aggregation in the cloud based on various semantic features is considered, for example, in [23]. However, semantic grouping does not imply a reduction in dimension, and the algorithm described by the authors is focused on batch, rather than streaming data processing technologies. Mathematical methods such as the principal component method [24] and the eigenvector method [25] reduce the data dimension, but their results are not used as input parameters for most analysis methods and are computationally complex. Specific aggregation methods applicable to security monitoring task data include time series aggregation [26] and aggregation of network traffic statistics [27]. However, the method from [26] creates additional calculations and is not very applicable for processing stream data. The method in [27] is the most widely used in the problems under consideration today.

Next, the paper considers the construction of stages of aggregation and normalization of streaming data for monitoring the safety of industrial CPS. The main approach is to systematize the requirements for pre-processing data from analysis methods. Data normalization is given in accordance with the approach [28] and taking into account the specifics of the cyber-physical systems. Data aggregation is based on the approach [29] for network traffic aggregation. The work also includes an approach to multidimensional data aggregation based on hierarchical relationships and the requirements of multidimensional analysis.

The novelty of the work consists in creating a generalized data processing scheme, adapting existing methods of normalization and hierarchical data aggregation, as well as developing a method of multidimensional data aggregation for monitoring information security of cyber-physical systems.

3. Data Processing Procedure in Information Security Monitoring Systems

3.1. Data Preprocessing in the CPS

Information security monitoring includes several stages. This includes collecting or registering data, pre-processing data (or preliminary data processing) including normalization and aggregation of values, generating output structures for appropriate mathematical methods, applying analysis methods, and making decisions. The general monitoring scheme is shown in fig. 1.

Preliminary data processing is an intermediate step from data collection to applying mathematical analysis methods to it. It includes a number of subtasks and links an instance of the source system (the security object) with generalized methods for solving security problems.

Data preprocessing depends on two factors. First, it depends on the types of data sources and the input data itself. Second, the data consumer is responsible for the analysis tasks. Since data preprocessing depends on the input of the monitoring system, it depends on the security object, or at least its type. Industrial systems are differing from the point of view of information security monitoring from corporate networks. Features of industrial FSCS are:

- heterogeneous data sources;
- resource restrictions;
- data loss during transmission.

Heterogeneous data sources in industrial CPS are represented by two main types of sources. These types are due to the nature of cyber-physical systems. They are:

- network traffic;
- data of physical process sensors.

The combination of these data is used to detect anomalies by various methods in industrial cyber-physical systems [8-10, 13]. The heterogeneity of the physical process sensors is an additional challenge. The same physical indicators that are comparable from an analytical point of view can be presented not only in different formats, but also in different scales (for example, the Celsius and Fahrenheit scales for temperature).

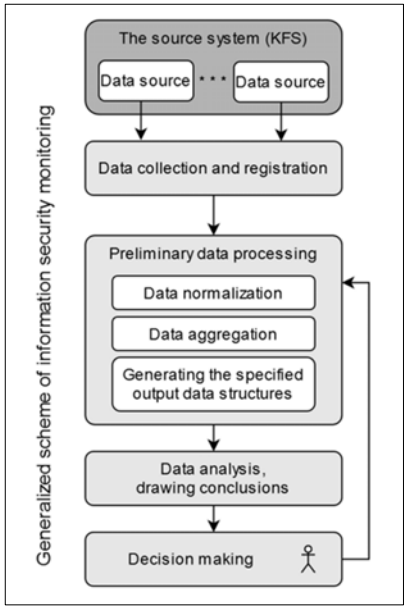

Fig. 1. Information security monitoring scheme

The security goals of a cyber-physical industrial system are achieved through the use of mathematical methods for analysis problems solving. The monitoring data consumer or analysis task has requirements for the preprocessing process:

- correspondence of output data structures to the input parameters of the analysis module;
- the reduction in data processing time.

The first requirement is mandatory for the operation of the data monitoring and analysis system. The second requirement is achieved by optimizing the data processing process. Anomaly detection methods are widely used to monitor the security of industrial cyber-physical systems [3, 4, 8-11, 18, 19, 21-23]. Their peculiarity is the approach to the analysis of network traffic and industrial processes based on time series. A feature of processes in industrial cyber-physical systems is their self-similarity [30]. Because of this, different methods perform data analysis for the same data using different time intervals. The difference between these intervals is significant (from seconds to minutes and months). Thus, the data preprocessing subsystem must convert the same set of input data into multiple time series with different aggregation periods.

Reducing data processing time is achieved due to several factors:

- excluding long operations on data;
- reducing the number of operations on data in General in the system;
- reducing the amount of data (in the system memory).

The first factor is achieved by excluding the write operation to disk during data preprocessing. The use of streaming data processing technologies eliminates long-term operations. This raises the problem of stream aggregation, which is not applicable to traditional methods of batch processing of information [31]. Reducing the number of operations is achieved by optimizing data processing structures and algorithms. It includes optimization of the order of operations (data processing scheme) and selection of optimal algorithms. Reducing the amount of data is achieved by storing in RAM only data structures that are directly used by analysis tasks. Intermediate and non-aggregated data should not take up space. This means that less data is read from memory and fewer resources

are required for the system to work. These are important factors for monitoring system efficiency and performance in large-scale distributed industrial systems.

3.2. Data Flow Processing Diagram

A large-scale distributed industrial system that monitors both network traffic and sensor data generates a large number of heterogeneous data. Achieving maximum performance of the data preprocessing subsystem requires matching the stages of normalization and aggregation of information. For monitoring the safety of large-scale industrial CPS, this agreement defines the list of initial analysis tasks. It consists in developing a data flow diagram. We introduce the following notation:

$S = \{s1, \dots, sn\}$ – data sources set, $D = \{d1, \dots, dm\}$ – set of the original pieces of data received from data sources . Mapping function $Fin: D \to S$ and its inverse define mappings between these sets, linking data fragments to sources.

$Za = \{z1, \dots zk\}$ – set of the data analysis tasks solved by a given monitoring system to achieve security goals. Each i-th problem has an input data set $Dzi = \{dzi, 1, \dots, dzi, I\}$, characterized by a structure (including the degree of aggregation).

$FA = \{a1, \dots, al\}$ – set of the data aggregation methods;

$FN = \{n1, \dots, nh\}$– set of the data normalization methods;

Then $ftr: \{D, Dzi\} \to Dzj$ – is a function for processing data and generating output sets for analysis methods, where $Ftr = \{ftr, 1, \dots, ftr, F\} = FA \cup FN$ – is a set of preprocessing functions.

The principle of a data preprocessing pipeline development is to minimize the number of data processing operations by combining the processing and data normalization stages in the most efficient way. There are two main stages for industrial CPS:

1. General stage of data preprocessing. It consists in primary aggregation of data from sources if analytical functions are applied to them separately for each stream. This stage consists of aggregation over time and does not include data normalization.
2. Local stage of data preprocessing. It consists of preparing data for joint analysis. This stage includes normalization of incoming data and repeated co-aggregation.

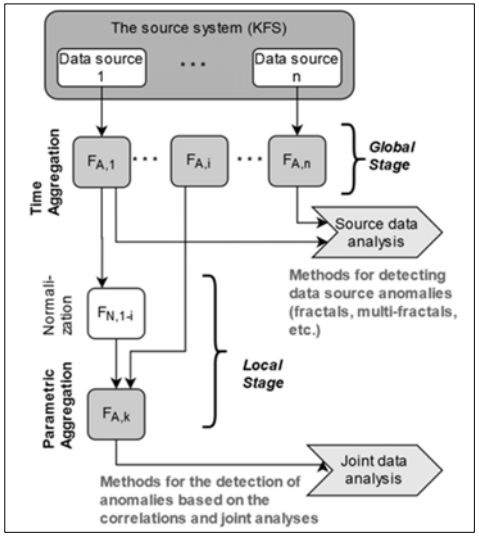

Fig. 2. Diagram of data flows and operations during preliminary data analysis in the monitoring system

Thus, the generalized data preprocessing pipeline for information security monitoring ($DataFlowTransform$) has the form: $DataFlowTransform = (FG: S \to FA) \Rightarrow (\{FL: \{\{FN, 0\} \Rightarrow FA\} \to Za\})$.

At the global FG stage, each data source is aggregated separately on the source itself, the intermediate node, or the monitoring server. The local stage consists of a sequence of normalization (if necessary) and aggregation functions. The result of each aggregation is the input data set of an analysis task. An example is shown in fig. 2.

In some systems, the data processing time optimization problem can be formulated over the given scheme, taking into account the specific characteristics of the processing tools.

3. Data Normalization

The main methods of cyber-physical industrial systems data pre-processing, in solving the security monitoring problem, are methods of aggregation (time and parametric) and data normalization. Time aggregation is also a subspecies of parametric data aggregation (the aggregation parameter is time). Despite this, we will highlight it separately, since this type of aggregation is used at the global stage and has its own characteristics. Data normalization consists of bringing heterogeneous data to a single view. It includes:

- syntactic (structural) normalization;
- semantic normalization.

Syntactic or structural normalization is the data transformation to a single view in terms of the presentation format. This is the process of matching data structures and types between incoming data and data at the input of an aggregation (analysis) method. An example of syntactic normalization is type casting. A special feature of syntactic normalization is that it does not require knowledge of the subject area and additional data about the data source.

Semantic normalization consists in bringing the values of the original data fragments to a single form, taking into account their semantics. For example, if data sources record information in different metric systems, one need to bring measurement scales and convert values. These transformations are also formalizable, but they require prior preparation based on knowledge of the subject area. For example, to aggregate temperature values together, one may need to convert indicators from the Celsius scale to the Kelvin scale, or in the opposite direction.

Stream data normalization corresponds to traditional normalization methods [28] used for industrial systems. Aggregation of the input stream additionally requires the following requirements:

- to use streaming tools or real-time modules for normalization;
- location in the memory reference metadata.

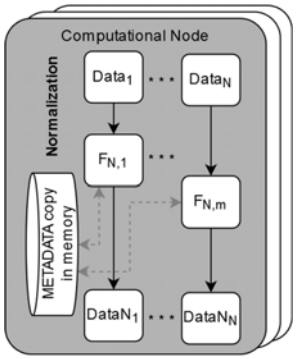

Fig.3 Diagram of data normalization process (one computational node)

In this case, an in-memory DBMS should be used to store metadata structures, or the metadata structure should be denormalized and simply placed in memory. A generalized normalization scheme for a single computing node is shown in fig. 3.

Dynamic load planning for data processing does not allow one to place only fragments of the metadata system on nodes to reduce resource requirements. However, this problem is not critical, since the share of data requiring semantic normalization in modern industrial systems is small. In turn, format normalization is implemented through automatic conversion of the data format during transmission to the local stage. This approach increases the load on individual nodes that do not require format validation. However, it significantly reduces the overall verification time in each individual case, since there is no access to metadata and calling an external conversion function. The choice between the two approaches is determined by the total number of types of transformations in a particular system.

The article goes on to reveal new methods of temporary data aggregation at both stages. Semantic and algorithmic related aggregation methods are used when it comes to a monoparametric data flow from a single source, and when a joint analysis of several data parameters is considered. When developing these methods, the features of the subject area were taken into account: the self-similarity of processes and data, the requirement for high reaction speed, and the requirement to reduce the load on the computer system.

5. Aggregation of streaming data

5.1. Global Aggregation Based on Time Series Hierarchies

The global stage of data aggregation is common for all parameters that enter the monitoring system from the CPS object. The main task of stream data aggregation at this stage is to generate time series of individual parameters with minimizing the number of operations and data stored in memory.

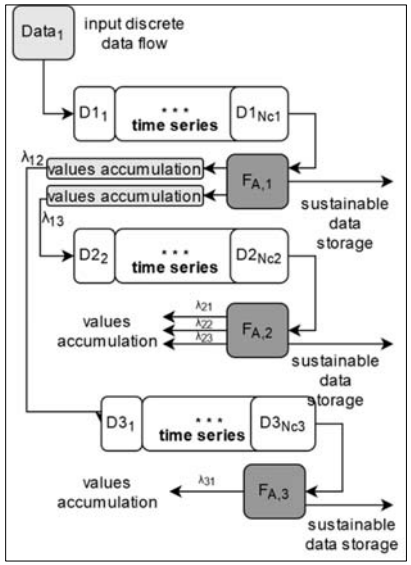

Fig. 4. Diagram of PCR Time series (Hierarchical aggregation)

To solve this problem, one can use the hierarchical aggregation based on time series, tested earlier for network traffic [30]. Aggregation of physical process sensor data over time is similar (in terms

of algorithms) to aggregation of traffic data. The approach is based on time aggregation of streaming data into related time series.

Methods for analyzing self-similar traffic and data use time series sets with different periods and a similar number of series members [8, 10, 13]. The algorithm sorts the time series in ascending order of the aggregation period. The order relation and parent – child relationship are set over time series. Time series with a common parent are located on the same level. Fig. 4 shows the flow diagram of streaming data during hierarchical aggregation in the security monitoring system.

Let's take two time series:

$W = (w_1, \dots, w_l)$ where T is the time of the aggregation window for the entire row, and dt – is the time of the aggregation window for the row element.

$W' = (w'_1, \dots, w'_i)$ where T' is the time of the aggregation window for the entire row, and dt' – is the time of the aggregation window for the row element.

A PCR (Parent-Child-Relation) relationship of the form $W \to W'$ is defined between the rows W and W' if and only if:

- $w'_1 = F_{A,i}(W) * \lambda$, where λ – the multiplier of the aggregation period, $\lambda \in N$.
- $\nexists(\lambda_i)|\lambda_i > \lambda$, λ_i – multiplier of the aggregation period for any other series.

That is, the ancestor of a given series is a series, all elements of which are aggregated to the value of the «descendant» element according to a given rule (aggregation functions $F_{A,i}(W)$ and the multiplier λ). Also, for linked $W \to W'$ series, the rule is true: $dt' = \lambda * T$.

The boundaries of the specified series are set as a set of series parameters when selecting analysis methods. Parameters are set for each time series of data:

- $dt_i \in [dt_i^{start} - \sigma, dt_i^{start} + \sigma]$, where σ – permissible error of the aggregation period of a series element, according to the analysis method;
- $Nc_i \in [Nc_i^{start} - \delta, Nc_i^{start} + \delta]$, where δ – acceptable change in the number of series elements, according to the analysis method, and Nc – is the number of series elements and Nc∈ N;
- $N_i \to min$.

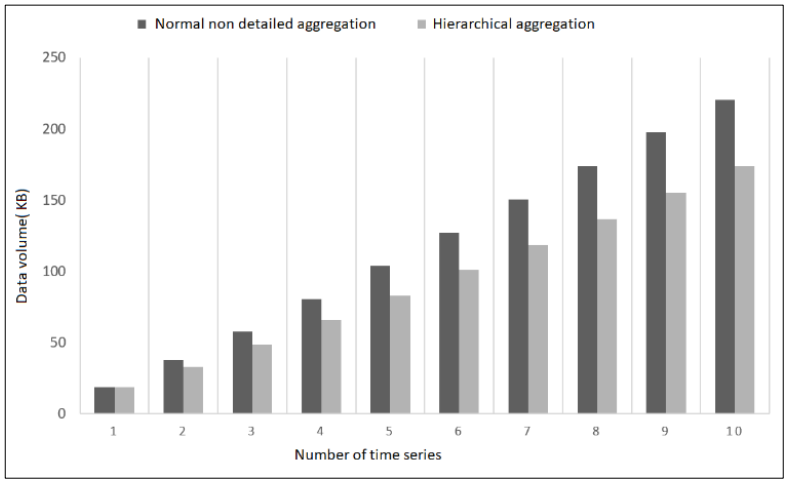

Fig. 5. Hierarchical vs Normal aggregation

Minimizing the number of elements in a series while observing other boundary conditions is a requirement to minimize computational operations during aggregation and processing of series

elements. A comparison of normal effective aggregation and adopted hierarchical aggregation is shown in fig. 5.

As a result, the number of operations with data becomes smaller. All calculations and aggregation over a data element are performed once. Data transfer from a parent with a shorter aggregation period to a child with a longer aggregation period is performed when the corresponding number of elements in the parent series is displaced. The space occupied by data also becomes smaller, since only the necessary data is stored. Storage of intermediate values is minimal.

5.2. Local Aggregation on the Basis of Related Parameters (Multidimensional Aggregation)

When solving the problem of local aggregation, there is a problem of joint analysis of parameters. There are two possible cases. The first is if the parameters of the monitoring object are generalized to a single value after normalization, which is analyzed independently (or the time series of which is analyzed independently). Then the usual hierarchical aggregation described above is used for self-similar data of the cyber-physical system. The second is if several time series of parameters are analyzed together. Then one need to solve two problems:

- consistency on time periods for time series aggregation;
- ensure fast data sharing.

One may use several approaches to solve these problems. First one is building a queue of trees. An additional hierarchy is used above the original series that sets aggregation rules from individual parameters to aggregates. Here, the hierarchy is formed in each individual element of the series. Second one is building a tree of queues. In this case, the hierarchy is set above the original time series trees. Third one is using the key graph. An additional graph (associated key graph - AKG) defines relationships between aggregated parameters, regardless of their original relationships. The graph flexibly links various parameters and queues that can be combined. The weight of the ties and set the coefficients of the transformation function data. The key graph diagram is shown in fig. 6.

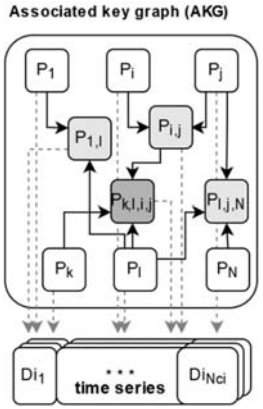

Fig.6. Diagram of associated key graph aggregation

The advantages of the key graph are:

- the possibility of movement on the graph and the changing composition of the aggregate parameters in the framework of certain relations;
- a small amount of additional data for each time window.

The drawback of the graph, as well as the previous methods, is the requirement to set relationships in advance. In other words, the parameters to be aggregated must be known in advance and explicitly

defined. Also, with increasing connectivity, the graph efficiency decreases, and the time for data processing increases. To evaluate the effectiveness, the graph parameters were determined with the growth of the number of related aggregation parameters (fig. 7) and the depth of aggregation of time series (fig. 8).

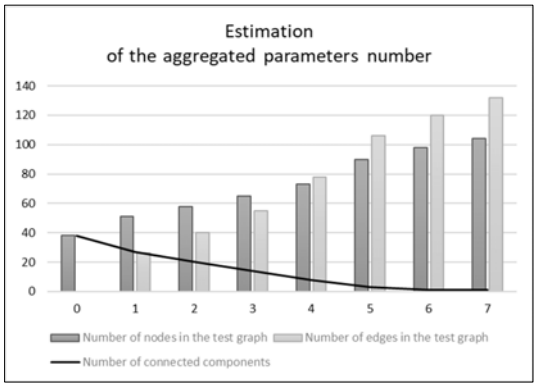

Fig.7. Evaluation of the graph effectiveness (aggregated parameters number)

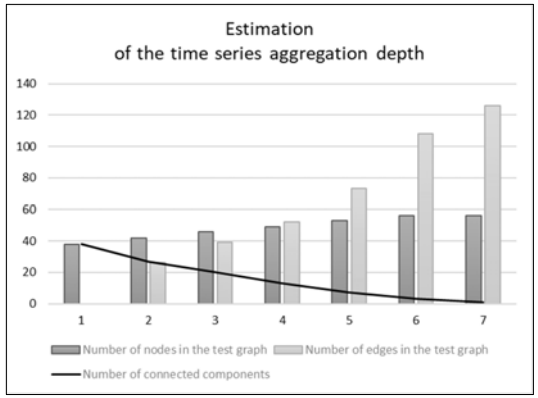

Fig. 8. Evaluation of the graph effectiveness (time series aggregation depth)

We can conclude, that this approach is most successful when the number of jointly aggregated parameters is no more than five and when the nesting depth is no more than six. This restricts the use of the method based on a connected graph. There is a need to further search for effective methods for aggregating a set of related parameters at the local aggregation stage. However, for cyber-physical systems with a small number of jointly analyzed parameters (for example, relatively homogeneous energy networks), this method can be applied.

6. Conclusion

Data pre-processing in the security monitoring task of industrial cyber-physical systems is similar to data pre-processing tasks in other areas. However, it has such features as semantic heterogeneity of input data and heterogeneity of data sources, and limitations on computing resources. The detailed flow diagram of data flows depends on the monitoring task: the number and type of parameters collected, and the requirements of analysis methods. Using a General data flow scheme that includes two stages of aggregation (local and global) and normalization allows one to organize data

processing and decompose it across distributed computing nodes, including data sources. The general scheme is unified for this type of system.

Semantic normalization in CPS security monitoring requires the presence of processing functions directories or metadata on the handler nodes. Semantic normalization transformations in industrial systems are relatively simple and not numerous (by types of transformations). In this way, metadata can be replicated in the memory of handler nodes. However, the minimum requirements for such nodes will be higher.

Hierarchical aggregation of streaming data (at the global stage) and multidimensional aggregation at the local stage allow one to reduce the number of operations on data and reduce the amount of data for online storage. The described approaches are based on the self-similarity inherent in network traffic and industrial processes. Data analysis methods for detecting anomalies and intrusions also use this property.

These methods can be used as a basis for developing an effective system for monitoring the security of primitive cyber-physical systems. They contribute to the improvement of security systems and data streaming technology. Important further areas of research are: improving the efficiency and overcoming the disadvantages of these methods, developing methods for processing secondary data exported from other systems, and ensuring semantic correspondence between methods for solving security problems and the data processing subsystem.

References

[1]. ISA/IEC 62443 Security for Industrial Automation and Control Systems. IEC technical committee 65: Industrial-process measurement, control and automation.

[2]. APT attacks on industrial companies in Russia: a review of tactics and techniques. URL: https://www.ptsecurity.com/upload/corporate/ru-ru/analytics/apt-attacks-industry-2019-rus.pdf, accessed 10.10.2020.

[3]. Kalinin M.O. Permanent Protection of Information Systems with Method of Automated Security and Integrity Control. In Proc. of the 3rd International Conference on Security of Information and Networks, 2010. pp. 118-123.

[4]. Knapp E.D., Langill J.T. Security Monitoring of Industrial Control Systems. Lecture Notes in Computer Science, vol. 9588, 2015, pp. 351-386.

[5]. Lavrova D.S., Zaitseva E.A., Zegzhda D.P. Approach to Presenting Network Infrastructure of Cyberphysical Systems to Minimize the Cyberattack Neutralization Time. Automatic Control and Computer Sciences, vol. 53, no. 5, 2019, pp. 387–392.

[6]. Dymora P., Mazurek M. An Innovative Approach to Anomaly Detection in Communication Networks Using Multifractal Analysis. Applied Sciences, vol. 10, 2020, article no. 3277.

[7]. De La Torre Parra G., Rad P., Choo K.-K. R. Implementation of deep packet inspection in smart grids and industrial Internet of Things: Challenges and opportunities. Journal of Network and Computer Applications, vol. 135. 2019, pp. 32-46.

[8]. Kalinin M.O., Lavrova D.S., Yarmak A.V. Detection of Threats in Cyberphysical Systems Based on Deep Learning Methods Using Multidimensional Time Series. Automatic Control and Computer Sciences, vol. 52, no. 8, 2018, pp. 912–917.

[9]. Coletta A., Armando A. Security Monitoring for Industrial Control Systems. Security of Industrial Control Systems and Cyber Physical Systems. Lecture Notes in Computer Science, vol. 9588, 2015, pp. 48–62.

[10]. Zegzhda D., Lavrova D., Khushkeev A. Detection of information security breaches in distributed control systems based on values prediction of multidimensional time series. In Proc. of the International Conference on Industrial Cyber Physical Systems (ICPS), 2019, pp. 780-784.

[11]. Burska K. Oslejsek R. Visual Analytics for Network Security and Critical Infrastructures. Lecture Notes in Computer Science, vol 10356, 2017. pp. 149-152.

[12]. Kalinin M.O., Minin A.A. Security Evaluation of a Wireless Ad-Hoc Network with Dynamic Topology. Automatic Control and Computer Sciences, vol. 51, no. 8, 2017, pp. 899-901.

[13]. Lavrova D.S., Alekseev I.V., Shtyrkina A.A. Security Analysis Based on Controlling Dependences of Network Traffic Parameters by Wavelet Transformation. Automatic Control and Computer Sciences, vol. 52, no. 8, 2018, pp. 931–935.

[14]. Cejka T., Zadnik M. Preserving Relations in Parallel Flow Data Processing. Security of Networks and Services in an All-Connected World. Lecture Notes in Computer Science, vol. 10356, 2017. pp. 153-156.

[15]. Bar A., Finamore A., Casas P., Golab L., Mellia M. Large-scale network traffic monitoring with DBStream, a system for rolling big data analysis. In Proc. of the International Conference on Big Data, 2014, pp. 165-170.

[16]. Mohapatra S.K., Sahoo P.K., Wu S.-L. Big data analytic architecture for intruder detection in heterogeneous wireless sensor networks. Journal of Network and Computer Applications, vol. 66, 2016, pp. 236-249.

[17]. Joshi M., Hassn Hadi T.A Review of Network Traffic Analysis and Prediction Techniques. arXiv preprint 1507.05722, 2015.

[18]. Fahad A., Tari Z., Khalil I., Habibb I., Alnuweiric H. Toward an efficient and scalable feature selection approach for internet traffic classification. Computer Networks, vol. 57, no. 9, 2013, pp. 2040-2057.

[19]. Trihinas D., Pallis G., Dikaiakos M. Low-Cost Adaptive Monitoring Techniques for the Internet of Things. IEEE Transactions on Services Computing, 2018, 14 p. DOI: 10.1109/TSC.2018.2808956.

[20]. Lv F., Wen Ch., Liu M. Representation learning based adaptive multimode process monitoring. Chemometrics and Intelligent Laboratory Systems, vol. 181, 2018, pp. 95-104.

[21]. Lavrova D.S., Popova, E.A., Shtyrkina, A.A. Prevention of DoS Attacks by Predicting the Values of Correlation Network Traffic Parameters. Automatic Control and Computer Sciences, vol. 53, no. 8, 2019, pp. 1065–1071.

[22]. Shang C., Yang F., Huang B., Huang D. Recursive Slow Feature Analysis for Adaptive Monitoring of Industrial Processes. IEEE Transactions on Industrial Electronics, vol. 65, no. 11, 2018, pp. 8895-8905.

[23]. Jiang Y., Yin S., Kaynak O. Data-Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond. IEEE Access, vol. 6, 2018, pp. 47374–47384.

[24]. Karthick N.G., Kalrani A.X. A Survey on Data Aggregation in Big Data and Cloud Computing. International Journal of Computer Trends and Technology (IJCTT), vol. 17, no 1, 2014, pp 28-32.

[25]. Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, vol. 2, 1901, pp. 559-572

[26]. Golub G.H., Van Loan C.F. Matrix Computations. Johns Hopkins University Press, 1996, 728 p.

[27]. Leonard M. J., Crowe K.E., Christian S.M., Jennifer Leigh Sloan Beeman, David Bruce Elsheimer, Edward Tilden. Computer-implemented systems and methods for efficient structuring of time series data. United States Patent US009244887B2, 2016.

[28]. David Anthony Hudhes, Pawan Kumar Singh. Hierarchical aggregation of select network traffic statistics. United States Patent US20200021506A1, 2020.

[29]. Poltavtseva M.A., Lavrova D.S., Pechenkin, A.I. Planning of aggregation and normalization of data from the Internet of Things for processing on a multiprocessor cluster. Automatic Control and Computer Sciences, vol. 50, no. 8, 2016, 703–711.

[30]. Poltavtseva M.A., Zegzhda P.D., Pankov I.D. The Hierarchial Data Aggregation Method in Backbone Traffic Streaming Analyzing to Ensure Digital Systems Information Security. In Proc. of the 2018 Eleventh International Conference on Management of Large-Scale System Sevelopment, 2018, pp. 1-5.

[31]. Sheluhin O., Atayero A., Garmashev A. Detection of Teletraffic Anomalies Using Multifractal Analysis. International Journal of Advancements in Computing Technology, vol. 3, no. 4, 2011, pp. 174-182.

[32]. Kleppmann M. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly Media, 2017, 640 p.

Информация об авторах / Information about authors

Мария Анатольевна ПОЛТАВЦЕВА – кандидат технических наук, доцент, доцент Института кибербезопасности и защиты информации. Сфера научных интересов: системный анализ и обработка информации, хранение данных, СУБД, Большие данные, модели данных, информационная безопасность, мониторинг информационной безопасности, поддержка принятия решений в информационной безопасности.

Maria A. POLTAVTSEVA – Candidate of Technical Sciences, associate Professor of the Institute of cybersecurity and information security since 2014. Research interests: system analysis and information processing, data engineering, DBMS, Big data, data models, information security, information security monitoring, information security decision support systems.