Project of Russian Science Foundation No 21-71-20078

Analytical processing of large arrays of heterogeneous data in the interests of state assessment, decision support and incident investigation to ensure cybersecurity of critical infrastructures

Description of work performed and scientific results obtained in 2022

1. Methods, models, techniques and algorithms for real-time attack detection based on simulation and graph-based modeling have been developed. The developed attack detection method is based on a combination of analytical, simulation and graph-based modeling, machine learning, artificial neural networks and big data technology. The method assumes the possibility of parallelizing the procedures for detecting attacks with division into various scenarios, types of attacks, procedures for constructing and processing graphs. The model-algorithmic part of the method specifies the procedures used to describe the analyzed system and its characteristics, construct state graphs, determine transitions between states, and classify and predict future states.

2. Methods, models, techniques and algorithms for real-time detection of anomalous activity and violations of criteria and security policies based on the analytical processing of large arrays of heterogeneous data on cybersecurity events have been developed. Detection methods provide for the implementation of four stages: preprocessing of input data, classification of an object represented by a vector or matrix of attributes, formation of a decision on the status of an object, and formation of an explanation for the decisions made at the previous stage. A model-algorithmic part of the methods has been developed, which includes a formal model for detecting anomalous activity and violations of criteria and security policies based on the analytical processing of large arrays of heterogeneous data, a data model from heterogeneous sources, in particular, from technological processes of cyber-physical objects and security policies, and input data preprocessing algorithms depending on the type of their source. A formal model for explanations of decisions has been developed, as well as algorithms for their formation, taking into account the requirements for the time of their generation.

3. Methods, models, techniques and algorithms for the operational assessment of the security of information, telecommunications and other critically important resources based on the analytical processing of large arrays of heterogeneous data have been developed. As a basis for the developed models and methods, analytical expressions are used to quantify security risks, high security risk paths (taking into account existing vulnerabilities) and the maximum possible risk, taking into account the maximum possible vulnerabilities. Algorithms for operational security assessment include algorithms for building a network model, building a risk graph, identifying risk paths, assessing the probability and impact of risk paths, assessing network risks, and choosing a high-risk path. The developed methodology for the operational assessment of resource security is focused on the joint implementation of the above algorithms. The following metrics are calculated: high-risk path for a specific host; a high-risk path for a particular attacker; a high-risk path for a specific entry point; a high-risk path for the entire resource pool.

4. Methods, models, techniques and algorithms for operational analysis and management of information security risks have been developed based on the analytical

processing of large arrays of heterogeneous data on cybersecurity events in the interests of assessing the state, supporting decision-making and investigating incidents. The methods differ in calculating the level of risk based on data on attacks and anomalies detected during the joint analysis of network traffic and event logs, and using the obtained estimates to predict the development of cyber attacks, taking into account the stage of the attack, to which the detected attacking action belongs. The developed model-algorithmic part of the approach includes a model of the analyzed cyber-physical system, an attack model, a model of damage propagation in the analyzed system, as well as methods for their construction based on the analysis of event log data and network traffic, and algorithms for calculating the level of risk.

5. Methods, models, techniques and algorithms for the operational visualization of large arrays of heterogeneous data on cybersecurity events have been developed in the interests of assessing the state, decision support and investigating incidents. The method of online visualization of large arrays of heterogeneous data is based on solving the problems described at the first stage of the project, which arise when trying to visualize large volumes of security data. When rendering, the goal is to reduce the number of objects displayed on the screen using methods for aggregating the displayed dimensions and objects, as well as approximating objects. The decision to use individual algorithms and models is made based on the amount of data, the number of measurements and objects. The decision is made by the system operator experimentally. It is proposed to use algorithms for aggregating measurements (PCA, UMAP, multicollinearity analysis), objects (K-means, spectral, search for communities based on modularity and label propagation), approximations (KDE, hexagonal lattices) and visualization models (bar graphs, scatter plots, trilinear coordinates, line graphs, counters, parallel coordinates, force graphs, tree maps, radial graphs, and matrices).

6. Methods, models, techniques and decision-making algorithms have been developed for the protection of information, telecommunications and other critical resources based on the analytical processing of large arrays of heterogeneous data on cybersecurity events in the interests of assessing the state, supporting decision-making and investigating incidents. Decision-making methods differ in the allocation of decision-making levels depending on the available time, the possibility of automatic implementation of protection measures and the access rights necessary for the implementation of protection measures. In addition, proactive and reactive decision making is provided depending on the stage of the attack. The developed algorithms for choosing the optimal set of protective measures use the results of attack and anomaly detection, as well as risk analysis. The modeling-algorithmic part of the methods includes a protective measure model and a hierarchical decision-making model, as well as a methodology and algorithm for choosing protective measures based on solving multicriteria optimization problems.

7. In addition to the plan, the architecture and functional structure of the system for analytical processing of large arrays of heterogeneous data on cybersecurity events was developed. In the functional structure of the system, service components are distinguished, such as a message broker, a stream broker, online and long-term storage components, a management and computing management component. Flow control algorithms for the analysis of large arrays of heterogeneous data are investigated. Testing of existing information technologies for Big Data processing was carried out. A comparative analysis of Big Data analysis frameworks, the security models supported by them, flow control algorithms and methods, as well as the possibilities of developing a system for managing parallel computing tasks is carried out. A prototype has been implemented to evaluate the possibility of launching

these technologies on domestic operating systems. Experimental stands "Smart Home", "Smart Transport" and "Office" have been previously developed and tested.

Research results are published in 20 articles indexed in WoS and Scopus (including 5 Q1 articles), in one article indexed in RSCI, and 21 articles and abstracts indexed in RSCI.

During the implementation of the project, exclusive rights to the results of intellectual activity (RIA) were obtained: 1 patent for an invention, 7 certificates of state registration of computer programs and 1 certificate of state registration of a database.

Members of the team participated in testing the results at 13 Russian and international conferences and seminars.

URL: http://comsec.spb.ru/ru/projects/
URL: http://comsec.spb.ru/en/projects/