Reducing the Dimension of Input Data for IDS by Using Match Analysis

Sergey Erokhin[1], Boris Borisenko [2], Aleksander Fadeev [3]
Moscow Technical University of Communications and Informatics,
Moscow, Russia
[1]esd@mtuci.ru, [2]fepem@yandex.ru, [3]aleksandr-sml@mail.ru

Abstract—**Network attack detection is currently one of the most pressing problems in the secure use of networks. Network-based intrusion detection systems based on signature rules are unable to detect new types of attacks. Thus, the urgent task is to quickly classify network traffic to detect network attacks. This paper analyzes the possibility of downgrading the data vector in intrusion detection systems. The proposed approach will allow more efficient use of system resources and select the most appropriate (efficient) features.**

I. Introduction

Today, computers are increasingly becoming victims of attacks. Attacks are exposed not only corporate devices and local networks of large companies, but also computers of ordinary users. Attacks can be carried out both for the purpose of stealing personal data, especially financial data, and out of simple curiosity and entertainment, for example, by novice hackers. Also common causes of attacks are personal dislike to the owners of resources and competition. In the latter case, they are conducted by order and for a fee. Methods and types of attacks are very numerous, and every year they become more complex and cunning.

Intrusion detection systems (IDS) are used to detect the facts of unauthorized access to the system, as well as other situations that may violate the security of the information system [1]-[3].

The IDS usually includes:

- a sensor system designed to collect security events related to the system being protected;
- a subsystem of analysis designed to detect suspicious actions and attacks based on sensor data;
- a data warehouse required to accumulate primary information about events and analysis results;
- a subsystem for management and presentation of the analysis results, which allows setting up an intrusion detection system and monitoring the detection of the condition of the system to be protected.

The most complex intelligent part of IDS is the analysis unit. The input data for the analysis unit is the information received from the sensors. Most modern systems of this kind use a combination of several mathematical methods for analysis.

To improve the accuracy of prediction and detection of attacks, the IDS should collect various information about the operation of the protected system, and store and process a large amount of data. First of all, observations of all final (exhaustible) system resources: the number of connections for different periods of time, amount of transmitted traffic, free memory, processor load. The second type of data for observation and analysis is user behavior, which is usually recorded in special log files. The results of the analysis should be used for decision making the mode of operation of the system, in particular the need to filters incoming traffic.

However, using a filtering system with an incorrectly triggered filter in the absence of an attack leads to system performance degradation. In addition, quite often the creation of an effective protection system faces a lack of computing power.

Thus, the task of optimizing the resources spent on the security system arises maintaining the system of protection against network attacks at a high level.

The solution to this problem is to minimize the resources spent on maintaining information security at moments when the attacker's activity does not manifest itself or is insignificant. For this purpose, the IDS must use dynamic methods to quickly detect and prevent security threats. In other words, the IDS must use a mathematical model, which allows selecting the necessary set of protection means at any moment of time to ensure reliable protection, while requiring a minimum amount of resources.

There are the following mathematical methods to analyze and optimize the information protection system:

- methods based on the theory of fuzzy sets [4];
- methods based on the use of neural networks [5];
- methods of expert systems [6];
- methods of mathematical statistics [7];
- methods based on the use of Petri networks [8], [9];
- the mathematical apparatus of the theory of random processes;
- methods based on the use of the theory of automata;
- mathematical apparatus of game theory.

Statistical methods of intrusion detection use a well-established apparatus of mathematical statistics to the behavior of the subjects of the analyzed system. At first, statistical models are formed for all subjects. The components of such a model can be various parameters, for example, the total traffic in the system, the number of refusals in service, the ratio of incoming traffic to outgoing traffic, the number of unique requests to the system, etc. Any deviation of the used profile from the reference one is considered a security breach. The following disadvantages of this approach may be noted. First, intrusion detection systems based on statistical methods are, in most cases, insensitive to the order of events in the protected system: in some situations, the same events in the protected system are the same. Depending on their order may be characteristic of abnormal or normal activities. Secondly, in some cases It can be difficult to set threshold values of monitored characteristics for identification of abnormal activity. Low the threshold leads to a false alarm, and overvaluation leads to skipping intrusions. In addition, the attacker often uses individual approaches for each protection system, which makes the use of statistical methods are less effective.

The amount of information that the system collects is a compromise between cost and efficiency [10], [11]. A system that records every action in detail can seriously lose its performance and require too much disk space. For example, it may take hundreds of gigabytes of disk memory per day to collect full registration information about network packets in a Fast Ethernet channel.

There is no sense in controlling the system without further analysis of the obtained information. An extremely important feature of IDS is the way the accumulated data are analyzed. Due to the increasing complexity and number of attacks, there is a need to improve existing IDSs. One of the main tasks is to classify network traffic.

The task of network traffic classification is to obtain certain characteristics of network traffic at the input, and at the output - the type of class to which it belongs [12]. These packets and different frequency characteristics can be used as input characteristics. As output data, there can be application-specific identifiers, which are the sources of traffic generation. It is also possible to classify by identifiers of traffic type. Classification of traffic is necessary to identify applications and protocols, whose data are transmitted over the network. There are two classic methods of network traffic analysis: signature and heuristic. Signature methods describe each attack by a specific model or signature. As a model or signature you can accept a character string, a semantic expression in a special language, a formal mathematical model, etc. The essence of this method is to use special databases with attack signatures to find matches. The main advantage is the speed and accuracy of search for attack signatures. The disadvantage is that new attacks, whose signatures are unknown, will be impossible to determine. In the heuristic analysis, the code of the object being checked is investigated and indirectly determines whether the object is malicious. The advantage of this method is the ability to detect unknown attacks. At present, the trend in IDS development is approaches based on machine learning methods, as well as on hybrid systems, which include several methods.

Methods based on machine learning have the following positive properties:

- self-learning and adaptability - the ability to automatically adapt to dynamically changing content;
- autonomy - independence from external knowledge bases and experts.
- the main disadvantages of such methods are:
- need to have a training kit;
- increased risk of false positive errors;
- the need for stable retraining when changing the characteristics of known protocols and the emergence of new ones.

The first stage in the development of attack detection systems is the choice of the machine learning method. The second is to search for or create a dataset, which will be used to train the system.

II. Estimation of the impact of the signs on the end result

Each of the datasets has its own set of attributes, which is involved in the process of learning the adaptive algorithm. A large number of attributes leads to high time costs of data processing, large amounts of memory required for information storage, as well as the need to collect a large number of precedents for confident recovery of hidden dependencies in a significantly multidimensional space.

As a method used to reduce the number of features in an object of inquiry we will use the mechanism of multiple analysis of matches due to the large number of applications in the field of research and evaluation of characteristic parameters and objects [13]. The choice of this method is conditioned by the presence of characteristic features allowing to consider many vectors of inquiries in nominal scales, as well as by the presence of the mechanism of vector dimension reduction by transformation of quantitative characteristics into qualitative ones.

A. *Correspondence analysis method*

Conformance analysis refers to methods of preliminary (exploratory data analysis, EDA). Preliminary data analysis - analysis of basic data properties, finding common regularities, distributions and anomalies in them, building initial models. It is at this stage that you need to get an idea of the data, understand which methods are better to apply and, more importantly, what results can be expected. Most data pre-analysis methods are graphical in nature. The reason for this strong reliance on graphics is that, by its very nature, the main role of data pre-analysis is to provide an unbiased study. Combined with natural image recognition capabilities, graphics provide unprecedented capabilities for this task.

The specific graphical methods used in data pre-analysis are often quite simple and consist of the following methods [14]:

- building raw data (data tracing, histograms, probability charts, etc.);
- construction of simple statistical data (average value charts, standard deviation charts, diagrams, box chart);
- placement of charts in such a way that the analyst can use his natural ability to recognize images.

In addition to graphical methods, in the preliminary data analysis typical quantitative methods (gradient analysis, dimensional reduction method (multidimensional scaling, nonlinear dimensional reduction, main components method, independent components method) are used. The main tools of preliminary analysis are study of probability distributions of variables, construction and analysis of correlation matrices, factor and discriminant analysis, multidimensional scaling [15].

Purposes of preliminary data analysis [16], [17]:

- maximize understanding of the data set;
- identification of the main structures;
- identification of the most important variables;
- detection of deviations and anomalies;
- checking of initial assumptions and hypotheses;
- development of the initial (initial) model;
- determination of optimal parameters.

Preliminary data analysis is used when, on the one hand, a researcher has a table of multidimensional data and, on the other hand, a priori information about the physical (causal) mechanism of generation of these data is missing or incomplete. In this situation, preliminary data analysis can help the researcher to describe the data structure in a compact and understandable way, from which he can raise the question of a more detailed study of data with the help of a statistical hypothesis checker, and perhaps also make some conclusions about the causal model of the data. This step is called a confirmatory data analysis (CDA). Identifying the data structure with a preliminary data analysis can also act as a confirmatory data analysis. A number of methods of preliminary data analysis can also be seen as methods of preparing data for subsequent statistical processing without any examination of the data structure to be carried out at subsequent stages. In such a case, the preliminary data analysis phase plays the role of some stage of recoding and transforming the data (e.g. by reducing the dimensionality) into a form suitable for subsequent analysis. In any case, whatever the purpose of the preliminary data analysis methods, the main task is to move to a compact description of data with the fullest possible preservation of data [18].

The methods of preliminary data analysis are designed to generate hypotheses about the distribution and interconnections of data, after which, at the next stage, the obtained hypotheses can be tested by confirming methods.

There are two approaches to match analysis: simple correspondence analysis (CA) and multiple correspondence analysis (MCA). These approaches are descriptive, multidimensional methods to investigate associations inherent to problems with multiple solutions. The input data for simple correspondence analysis is typically a frequency matrix of one or more conjugation tables. Generally speaking, the correspondence analysis can be applied to any rectangular matrix composed of query vectors. The only limitation is non-negative numbers in the matrix cells [19]. The distinctive feature of the correspondence analysis is the ability to convert absolute values of data into nominal values and then introduce a metric.

Correspondence analysis allows you to examine both the strength and nature of relationships between categories in rows and columns of the table.

Multiple Correspondence Analysis is an extension of the simple Correspondence Analysis that allows you to analyze the relationship patterns of several categorical dependent variables.

Technically, simple and multiple correspondence analysis are closely related to canonical and multiple correlation analysis, respectively.

Canonical correlation analysis is used to describe relationships between two sets of continuous variables (vectors), while multiple correlation analysis captures relationships between more than two sets of continuous variables. In canonical correlation analysis and multiple correlation analysis, a series of linear combinations or weighted compositions of each set of variables, called canonical variations, are obtained so that they are mutually orthogonal to each other. Within a single set of linear combinations, while remaining maximally correlated with different sets of linear combinations. These correlations between variables are called canonical correlations [20].

Technically, MCA is obtained using the standard analysis of the correspondence of the indicator matrix (i.e. matrix whose elements are equal to 0 or 1) [21]. It is necessary to adjust the percentage of the explained dispersion and to adapt the interpretation of the inter-point distance correspondence analysis.

The correspondence analysis contains three basic concepts [22], [23]:

- a point in multidimensional space (profile);
- the weight (or mass) assigned to each point;
- function of distance between points (chi-square distance (χ^2).

Four derivative concepts are also considered:

- centroid (weighted average value);
- inertia;
- subspace;
- projection.

Points in multidimensional space (profiles) are multidimensional points weighted by masses, and the distances between profiles are measured using chi-square distance. Points in multidimensional space (profiles) are visualized by projecting them into a subspace of low dimensionality, which best corresponds to points in multidimensional space (profiles), and then projecting the vertices of points in multidimensional space (profiles) into a subspace as reference points for interpretation. However, there are many other ways to define and interpret the correspondence analysis, and that's why the same basic methodology was reopened many times in different contexts (dual scaling, optimal digitization, simultaneous linear regression). One of these alternative interpretations is called Optimal Scaling.

B. *Points in multidimensional space (profiles)*

The Correspondence Analysis method is used to explain the structure of relationships (matches) between categories of x_1 and x_2 variables. The categories are considered as points in some multidimensional space.

The initial profile of the object d in the correspondence analysis is understood as a vector with the representation of features in numerical values:

$$d^i = (d_1^i, d_2^i, ..., d_m^i), \text{ where } d_j^i = \{1, ..., n_j\}, n_j \in \mathrm{N}$$

The set of vectors is represented as a matrix of D(n,m) dimension nxm called the correspondence matrix, where elements d^i its string, n is the number of elements d, m is the number of features in elements d. The indicator matrix Z means the representation of matrix D in binary form. The binaryization process is described by a sequence of next steps:

a vector-string is selected where $d^i = \{d_1^i, d_2^i, ..., d_m^i\},\ d_j^i = j,\ j \in \{1, ..., n_j\}$ - some numerical value;

d_j^i is represented as a binary string $z_j^i = (0, ..., 1_j, ..., 0)$, where the vector length z_j^i is equal to the maximum value d_j^i for all vectors d.

The marginal sum of the i-th row (M_s) (j-column (M_r)) means the sum of the elements of the row (column).

$$M_s^i = \sum\nolimits_{k=1}^{m} d_k^i,\ \ M_r^j = \sum\nolimits_{k=1}^{n} d_j^k.$$

For binary matrix Z, the values of marginal sums are the sum of units in the corresponding rows (columns).

The total marginal sum M means the value equal to the sum of marginal sums for all rows and columns.

$$M = \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} d_j^i.$$

A standard vector in correspondence analysis means a data vector with normalized (divided by the total marginal sum) values. Hereinafter, the object vector will be understood as a standard object vector.

C. *Weight*

The weight of the standard vector (rows/columns) means the value of marginal sums by rows/columns divided by the total marginal sum. In the correspondence analysis, the values of the weights are determined by the following relations:

$$w_i = \frac{M_s^i}{M}, w_j = \frac{M_r^j}{M}.$$

The middle vector of line d^0 means a vector where features are the mean values of vectors by lines:

$$d^0 = \left\{ \frac{\sum_j d_1^j}{M}, \frac{\sum_j d_2^j}{M}, ..., \frac{\sum_j d_m^j}{M} \right\}.$$

Geometrically, the middle vector is an analogue of a point that lies in the center of the point cloud represented by other vectors. If the vector differs greatly from the mean one, the corresponding point will be far from the center and vice versa.

D. *Metriks*

In the correspondence analysis, the weighted analog of the Euclidean distance is used as a formula for calculating the metric, where the weight is the value opposite to the corresponding element of the mean vector [19]. Here, the weighting refers to the axes (measurements) of space, not to the queries themselves. In practice, such weighing is expressed in the fact that rarely met values of parameters are included in the formula with a greater weight, while more often - with a smaller weight. This equilibration is achieved by dividing each square of difference into a corresponding element of the mean vector when calculating distance:

$$\rho(d^i, d^{i'}) = \sqrt{\sum\nolimits_{j=1}^{m} \frac{(d_j^i - d_j^{i'})^2}{\left|d_j^0\right|}},$$

where ρ(d^i, $d^{i'}$) - weighted Euclidean distance between d^i and $d^{i'}$ queries, $d_j^i, d_j^{i'}$ - elements of vectors, d_j^0 - elements of the line middle vector [24].

If elements of query vectors have binary form, the following distance function is used:

$$\rho(d^i, d^{i'})_{bin} = \sum\nolimits_{j=1}^{m} \left|d_j^i - d_j^{i'}\right|.$$

For binary representation, distance - there is a coefficient of mismatch, which is the sum of the number of positions in which the elements do not match.

III. Decrease in the number of query traits

Let there be some selection of objects $X = \{x_n\}_{n=1}^{N},\ x_n \in \mathrm{R}^D$.

The task of reducing the dimensionality is to get a representation of this sample in a smaller dimensional space $t_n \in \mathrm{R}^d$, $d < D$.

Reducing the dimensionality of the data description can have many purposes:

- reduction of computational costs in data processing;
- fight against overtraining. The smaller the number of attributes, the fewer objects are required to confidently restore hidden dependencies in the data and the better the quality of recovering such dependencies;
- data compression for more efficient information storage. In this case, in addition to the X→A conversion, the reverse A→X conversion is also required;

- data visualization. Designing a sample in two-three-dimensional space allows you to graphically represent the sample;
- extraction of new features. New features obtained as a result of the X→A transformation can make a significant contribution to the subsequent solution of problems [25].

The task of reducing the number of features described above is the task of reducing dimensionality of vectors that make up anomalous queries. In the multiple analysis of matches, the task of dimensional reduction comes down to finding a hyperplane that most accurately reflects the distances between points. In fact, this problem is equivalent to the task of searching for a smaller hyperplane, which would in some sense be closer to all points simultaneously. Closeness is determined by the method of weighted least squares.

In multiple match analysis, the reduction in dimensional decomposition is achieved by decomposing the indicator matrix using the singular value decomposition method (SVD). Singular Value Decomposition (SVD) is a decomposition of a real matrix to bring it to a canonical form [26]. The singular decomposition allows one to find the orthogonal bases of different vector spaces of the decomposed matrix and calculate the rank of the current matrix.

A singular matrix decomposition $A_{(mxn)}$ is a representation given in the form:

$$A_{(mxn)} = U_{(mxm)} \Lambda_{(mxn)} V^T_{(nxn)},$$

where the condition is fulfilled for U and V matrices:

$$U^T_{(mxm)} U_{(mxm)} = U_{(mxm)} U^T_{(mxm)} = E,$$

$V^T_{(nxn)} V_{(nxn)} = V_{(nxn)} V^T_{(nxn)} = E,$ where E is a single matrix.

Matrix Λ - diagonal, with elements that meet the condition:

$$\lambda_1 \geq \lambda_2 \geq ... \lambda_r > \lambda_{r+1} = ... = \lambda_n = 0.$$

According to this representation, at m > n, the diagonal matrix Λ has empty rows, and at m < n, it has empty columns. Therefore, there is another economical representation [27]:

$A_{(mxn)} = U_{(mxr)} \Lambda_{(rxr)} V^T_{(rxn)},$ where r = min(m,n).

IV. SVD DECOMPOSITION IMPLEMENTATION ALGORITHM

A simple iteration algorithm of singular matrix decomposition allows a simple highly parallel implementation. Singular matrix decomposition is necessary for many data analysis tasks. In particular, the analysis of the main components is reduced to a singular decomposition of the centered data matrix [28].

The main procedure is to search for the best approximation of an arbitrary matrix X=(x_{ij}) of mxn size with a matrix of the view $b \otimes a = (b_i a_j)$, where b is an m-dimensional vector, a is an n-dimensional vector, by the method of least squares [29]:

$$F(b,a) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - b_i a_j)^2 \to \min.$$

The solution of this problem is given by consecutive iterations using explicit formulas. At a fixed vector a=(a_j), the values b=(b_i) delivering a minimum form F(b,a) are unambiguously and explicitly defined from the equations ∂F/∂bi=0:

$$\frac{\partial F}{\partial b_i} = -\sum_{j=1}^{n} (x_{ij} - b_i a_j) a_j = 0,$$

$$b_i = \frac{\sum_{j=1}^{n} x_{ij} a_j}{\sum_{j=1}^{n} a_j^2}.$$

Similarly, at a fixed vector b=(b_i), values a=(a_j) are determined:

$$a_j = \frac{\sum_{i=1}^{m} b_i x_{ij}}{\sum_{i=1}^{m} b_i^2}.$$

As an initial approximation of vector a, take a random vector of unit length, calculate vector b, then calculate vector a for this vector b, etc. Each step decreases the value of F(b,a). The stopping criterion is a small relative decrease of the value of the minimized function F(b,a) per iteration step (ΔF/F) or a small decrease of the F value itself.

As a result, for matrix X=(x_{ij}) we get the best approximation of matrix P_1 of the kind $(b^1 \otimes a^1 = (b_i^1 a_j^1)$ (here the upper index indicates the iteration number). Then, we subtract the obtained matrix P_1 from matrix X, and for the obtained matrix of evasions X_1=X - P_1 we find again the best approximation of P_2 of the same kind, etc., until, for example, the norm X_k is small enough. As a result, we get an iterative procedure of matrix X decomposition as a sum of matrices of rank 1, i.e. X = P_1 + P_2 +...+ P_q $(P_l = b^l \otimes a^l)$.

We assume σ_l = $|a^l||b^l|$ and regulate the vectors a^l, b^l: a^l:=$a^l/|a^l|$; b^l:=$b^l/|b^l|$.

The result is approximation of σ_l singular numbers and singular vectors (right - a^l and left - b^l).

The advantages of this algorithm include its exceptional simplicity and the ability to transfer it to the data with spaces, as well as weighted data almost without changes.

There are various modifications to the basic algorithm that improve accuracy and stability. For example, vectors of main components a^l at different l should be orthogonal in construction, but at a large number of iterations (large dimension, many components) small deviations from orthogonality accumulate and it may be necessary to correct a^l at each step to ensure its orthogonality to previously found main components.

For square symmetric positively defined matrices, the described algorithm turns into a method of direct iterations to find your own vectors.

V. SINGULAR DECOMPOSITION OF TENSORS AND TENSOR METHOD OF THE MAIN COMPONENTS

Often a data vector has an additional structure of a rectangular table (e.g., a flat image) or even a multidimensional table, i.e. a tensor: $x_{i_1 i_2 \dots i_q}$, $1 \le i_j \le n_j$. In this case, singular decomposition is also effective. Definitions, basic formulas and algorithms are transferred almost unchanged: instead of the data matrix, we have q+1 index value $X = (x_{i_0 i_1 i_2 \dots i_q})$, where the first index i0 is the number of the data point (tensor).

The main procedure is to find the best approximation of the tensor by the $a^0_{i_0} a^1_{i_1} a^2_{i_2} \dots a^q_{i_q}$ view tensor $x_{i_0 i_1 i_2 \dots i_q}$, (where $a^0 = (a^0_{i_0})$ - m-dimensional vector (m - number of data points), $a^l = (a^l_{i_l})$ - dimensional vector n$_l$ at l>0) by the method of least squares:

$$F = \frac{1}{2}\sum_{i_0=1}^{m}\sum_{i_1=1}^{n_1}\dots\sum_{i_q=1}^{n_q}(x_{i_0 i_1 \dots i_q} - a^0_{i_0} a^1_{i_1} \dots a^q_{i_q})^2 \to \min.$$

The solution of this problem is given by consecutive iterations using explicit formulas. If all but one multiplier vectors are $a^k_{i_k}$ given, this remaining one is determined explicitly from sufficient minimum conditions.

$$a^k_{i_k} = \frac{\sum\limits_{i_0=1}^{m}\sum\limits_{i_1=1}^{n_1}\dots\sum\limits_{i_{k-1}=1}^{n_{k-1}}\sum\limits_{i_{k+1}=1}^{n_{k+1}}\dots\sum\limits_{i_q=1}^{n_q} x_{i_0 i_1 \dots i_{k-1} i_k i_{k+1} \dots i_q} a^0_{i_0} a^{k-1}_{i_{k-1}} a^{k+1}_{i_{k+1}} \dots a^q_{i_q}}{\prod\limits_{j \ne k} \left|a^j\right|^2}.$$

As an initial approximation of vectors ($a^l = (a^l_{i_l})$ l>0) (take random vectors of unit length, calculate vector a^0, then for this vector a^0 and data of vectors a^2,a^3,... calculate vector a^1, etc.). (cycling through indexes) Each step decreases the value of F(b,a). The algorithm obviously converges. We use a small relative reduction of the value of the minimized function F per cycle or a small reduction of the F value itself as a stopping criterion. Then, we subtract the obtained approximation from the tensor X $a^0_{i_0} a^1_{i_1} a^2_{i_2} \dots a^q_{i_q}$ and again look for the best approximation of the same species for the remainder, etc., until, for example, the norm of the next remainder is small enough.

This multi-component singular decomposition (the main component tensor method) is successfully applied to images, video signals and virtually any data with a table or tensor structure.

VI. JACOBI METHOD

The Jacobi method [30] can be used only for symmetric matrices and has a square convergence (convergence of about 2). The Jacobi method uses a sequence of flat rotations to diagonalize matrix A. For singular decomposition, a double-sided flat rotation of Jacoby is used:

$$A_{i+1} = J_1^T A_i J_r.$$

This transformation is called a rotation, and matrix J is called a Jacobian rotation matrix. The transformation has the following properties:

– symmetry property $a^1_{kl} = a^1_{lk}$;

– elements standing on the main diagonal remain unchanged. Only the elements in the i-th and j-th rows and in the j-m and i-th columns are changed;

– custom vectors and custom values are saved.

The Jacobian matrix is a single matrix whose four elements with indexes (p,p), (p,q), (q,p), (q,q) can be replaced with the following values:

$$J_{pp} = \cos(\theta), J_{pq} = \sin(\theta),$$
$$J_{qp} = -\sin(\theta), J_{qq} = \cos(\theta).$$

Cosine and sinus depend on the rotation parameterθ . Rotations can be left-handed and right-handed, then the rotation parameter is denoted by θ$_l$ and θ$_r$, respectively. Two-handed rotation is applied to each sub matrix 2 × 2:

$$\begin{bmatrix} a'_{pp} & 0 \\ 0 & a'_{qq} \end{bmatrix} = \begin{bmatrix} c_l & s_l \\ -s_l & c_l \end{bmatrix}_p^T \begin{bmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{bmatrix} \begin{bmatrix} c_r & s_r \\ -s_r & c_r \end{bmatrix}_q^T,$$

$$\begin{bmatrix} a'_{ij} & a'_{i(j+1)} \\ a'_{(i+1)j} & a'_{(i+1)j} \end{bmatrix} = \begin{bmatrix} c_l & s_l \\ -s_l & c_l \end{bmatrix}_i^T \begin{bmatrix} a_{ij} & a_{i(j+1)} \\ a_{(i+1)j} & a_{(i+1)j} \end{bmatrix} \begin{bmatrix} c_r & s_r \\ -s_r & c_r \end{bmatrix}_j^T,$$

Where $c_l = \cos(\theta_l)$, $c_r = \cos(\theta_r)$, $s_l = \sin(\theta_l)$, $s_r = \sin(\theta_r)$, a'_{pp}, a'_{qq} - diagonal elements obtained after diagonalization. It is noticed that Jacobi's left-hand rotation» affects only the elements in p and q columns, and Jacobi's right-hand rotation affects only the elements in p and q columns.

VII. GETTING THE SIGN SPACE OF REDUCED DIMENSIONS

After calculating the reduced number of attributes, we will perform the process of clustering the abnormal query values within the resulting metric space.

The recommended clustering algorithm in multiple match analysis is Ward's hierarchical agglomeration method [31], where the objects over which the clustering process is applied are anomalous traffic records. The input to the algorithm is a set that needs to be clustered, and the output is the final set of clusters.

The output value of clustering is the elements of anomalous queries as close to the center of the received clusters as possible. The mechanism for determining the center of a cluster is to consider each cluster as a subset of a set of anomalous queries. In case this subset contains one element, it will be the center. Otherwise, the theoretical center (pseudo-element) of a given cluster is located, and the element closest to the theoretical center is taken. This element will be the center of the cluster.

The obtained cluster centers using the algorithm of inverse transformation of the singular matrix and the arising rational coefficients are brought to integer values by means of rounding operation.

Thus, signs that do not have a significant impact on the process of identifying abnormal requests are the first candidates for removal from the data set records.

VIII. CONCLUSIONS

The combination of singular tensor decomposition, main component tensor method, CA, as well as dimensional reduction algorithms are excellent tools for network traffic analysis. They allow to discard secondary features, better understand the structure of input data, and are a good addition to IDS. The speed of the considered algorithms and methods is linearly dependent on the volume of input data, which allows their use in real time systems. Thus, the CSE-CIC-IDS2018 data set (10 GB of traffic was used) using reduced input data turned out to be 5.2 times faster than its usual analogue, and the probability of detecting an attack is higher than for a usual vector.

TABLE I. PROBABILITY OF ATTACK DETECTION

	Multilayer perceptron	LSTM
Normal vector	0,82	0,88
Reduced dimensional vector	0,84	0,89

REFERENCES

[1] Basalova G.V., “Application of methods of game theory in intrusion detection systems”, *Izvestia of Tula State University. Technical sciences*, 2017.

[2] Erokhin S.D., “Managing Security of Critical Information Infrastructure”, Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Russia, 2019, pp. 1-4.

[3] Erokhin S., Petukhov A. and Pilyugin P., “Critical Information Infrastructures Security Modeling”, 24th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 2019, pp. 82-88.

[4] Medvedev N.V., Troitsky I.I. and Tsirlov V.L., “On the use of the apparatus of the theory of fuzzy sets in the analysis of information security risks”, *Bulletin of the Moscow State Technical University. N.E. Bauman. Series "Instrument Engineering"*, 2011, pp. 25-30.

[5] Frolov P.V., Chukhraev I.V. and Grishanov K.M., “Application of artificial neural networks in intrusion detection systems”, *System Administrator magazine*, №9(190), 2018.

[6] Boltunov A.I. and Krotov L.N., “Expert systems application for solving the information security problems”, *International research journal*, №9(51), 2016.

[7] Atamanov G.A. and Rogachev A.F., “Analysis of the mathematical apparatus and methods used to assess the safety of objects of informatization”, International scientific and methodological conference "Problems of modern agricultural education: content, technology, quality", Volgograd 03-07 May 2018.

[8] Yakhontov I.V., “The analysis of models of information protection systems on the base of modified petri networks”, *Scientific and practical journal "Modern science: topical problems of theory and practice." Series "Natural-technical sciences"* №3, 2012, pp. 57-65.

[9] Kurilov F.M., “Modeling of information security systems. Graph theory application”, Technical sciences: theory and practice: materials of the III Intern. scientific. conf. (Chita, April 2016). - Chita: Young Scientist Publishing House, 2016. - pp. 6-9.

[10] Anikeev, M.V., “About the possibility of using the mechanism of the hidden Markovsky models in the information intrusion detection systems (in Russian)”, Proceedings of the VI International scientific-practical conference "Information security". -Taganrog: TRTU, 2004.

[11] Gorodetsky V.I., Kotenko I.V., Karsayev O.V. and Khabarov A.V., “Multi-agent technologies of complex information protection in telecommunication systems”, ISINAS, Proceedings, St. Petersburg, 2000.

[12] Nikolskaya K.Y. and Varkentin V.V., “Features of creating data sets for training intrusion detection systems based on machine learning methods”, Science of the Present and Future VII, 2019, May 16-18, pp. 141-143.

[13] Burlakov M.E., “Application of Correspondence Analysis Method for Optimization of Attribute Combinations in Data Sets*”, PNIPU Bulletin. Electrotechnics, Information Technologies, Control Systems*, №26. Samara, 2018. pp. 7-28.

[14] Natrella M (2013) NIST/SEMATECH e-Handbook of Statistical Methods. NIST/SEMATECH. Web: https://www.itl.nist.gov/div898/handbook/eda/eda_d.htm.

[15] Peter Bruce and Andrew Bruce, “Practical statistics for professionals”, Data Science, M.: BHV-Peterburg, 2018.

[16] Matthieu Komorowski, Dominic C. Marshall, Justin D. Salciccioli, and Yves Crutain, “Secondary Analysis of Electronic Health Records”, Web: https://www.ncbi.nlm.nih.gov/books/NBK543641/#ch15.CR6.

[17] Hill T. and Lewicki P., Statistics: methods and applications: a comprehensive reference for science, industry, and data mining. StatSoft, Inc., Tulsa, 2006. Web: http://web.archive.org/web/20080209113450fw_/http://www.statsoft.com/textbook/stathome.html.

[18] Ayvazyan S.A., Bukhshtaber V.M., Anyukov I.S. and Meshalkin L.D. Applied Statistics: Classification and Reduction of Dimensionality, p.473 - M: Finance and Statistics, 1989.

[19] Nishisato, S., “Analysis of categorical data: Dual scaling and its applications”, Toronto: University of Toronto Press. 1980.

[20] Heungsun Hwang, Marc A. Tomiuk and Yoshio Takane, “Correspondence Analysis, Multiple Correspondence Analysis and Recent Developments”, Web: https://www.researchgate.net/publication/236800706

[21] Hervé Abdi and Dominique Valentin, “Multiple Correspondence Analysis”, Web: https://www.researchgate.net/publication/239542271

[22] Michael Greenacre, “Correspondence Analysis in Practice”, CRC Press, 2016

[23] Michael Greenacre, “The Use of Correspondence Analysis in the Exploration of Health Survey Data”, Web: https://www.researchgate.net/publication/228027785

[24] Clausen, S.-E., “Applied correspondence analysis: An introduction”, *Sage university papers. Series: Quantitative applications in the social sciences*, 1998, Vol. 121, pp.137-150.

[25] Reducing the data description dimension: Main component method. Web: http://www.machinelearning.ru/wiki/images/a/a4/MOTP11_5.pdf.

[26] Kirk Baker, “Singular Value Decomposition Tutorial”, Web: https://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm.

[27] J. Golub and C. Van Lowen, “Matrix Computing”, Translated from English / Edited by V.V. Voyevodin, Moscow: Mir, 1999

[28] Alan Kaylor Cline and Inderjit S. Dhillon, “Computation of the Singular Value Decomposition”, Web: https://www.cs.utexas.edu/users/inderjit/public_papers/HLA_SVD.pdf

[29] Steven L. Brunton and J. Nathan Kutz, “Data Driven Science & Engineering”, *Cambridge University Press*, 2019, pp. 15-20.

[30] Safonova A.V., “Algorithm of singular matrix decomposition”, Web: http://rts-md.com/archive/RTS_16_1/3_5%20Safonova.pdf

[31] Distances between Clustering, Hierarchical Clustering. Web: http://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf.