ISSN 0146-4116, Automatic Control and Computer Sciences, 2015, Vol. 49, No. 8, pp. 766–769. © Allerton Press, Inc., 2015.
Original Russian Text © V.V. Platonov, P.O. Semenov, 2015, published in Problemy Informatsionnoi Bezopasnosti. Komp'yuternye Sistemy.

Using Data-Mining Methods to Detect Network Attacks

V. V. Platonov and P. O. Semenov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: Vladimir.platonov@ibks.ftk.spbstu.ru; Semenovpo@gmail.com

Received May 8, 2015

Abstract—This paper investigates the possibility of using various data-mining methods for network attack detection. A modular architecture of an intrusion detection system has been designed that enables classification of network packets in several support vector machines.

Keywords: intrusion detection system, data-mining methods, support vector machine, dimensionality reduction, principal component analysis

DOI: 10.3103/S0146411615080131

INTRODUCTION

This paper describes the architecture of a modular intrusion detection system based on several data-mining methods. In this architecture, data-mining methods are used for classification, clustering, and optimization. The classification methods categorize a sequence of network packets as either a set of attacks or as normal traffic. The dimensionality reduction methods construct an optimal feature space based on network traffic parameters to detect specific sets of attacks. The clustering methods divide various attacks into classes for a particular attack-detection module [1]. Using fuzzy logic allows reducing the number of errors first and second types in the process of interaction among attack-detection modules.

The designed architecture is based on the parallelization of the detection process, i.e., a number of attack-detection modules simultaneously run in the system in which each module is responsible for detecting a particular group of attacks (from a single attack to the set of all attacks sutable within a joint list of the traffic parameters being analyzed). Figure 1 shows a simplified scheme of a single detection module, which includes the following:

—extraction of relevant training information from network dumps;

—training the principal component analysis (PCA) on the extracted dataset;

—translating the initial data into a lower-dimensional space according to the obtained transformation rules;

—training the support vector machine (SVM) on the transformed data [2];

—testing the obtained SVM model;

—automatic adjustment of PCA and SVM parameters.

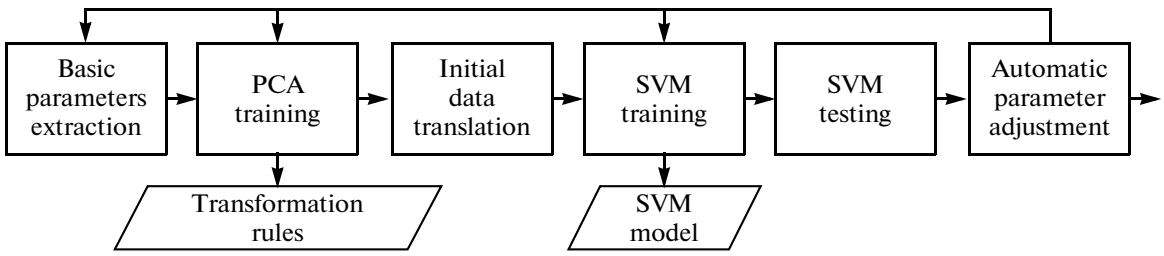

Fig. 1. Simplified scheme of the attack-detection module.

1. EXTRACTION OF BASIC TRAFFIC PARAMETERS

In the block of basic parameters extraction, the following (basic) parameters are extracted from the network traffic:

—network layer header parameters (IP and ICMP packets);

—transport layer header parameters (TCP and UDP packets);

—statistical and time parameters of the TCP session.

Both multibyte parameters and individual bytes are taken into account, e.g., the IP address can be represented by one, two, or four parameters. The output of this block is a set of vectors. Each vector contains a label, i.e., attack (−1) or normal traffic (+1), and extracted parameters.

2. REDUCING THE DIMENSIONALITY

In the dimensionality reduction block, a dimensionality reduction method is applied to the extracted vectors to construct a new feature space [3–5]. In this work, for the dimensionality reduction, we use the PCA. For this method, it is important what kind of matrix is analyzed. Here, the following matrices are used: correlation matrix, covariance matrix, and sums of squares and cross products matrix.

For the automatic selection of a sufficient number of new parameters and the filtration of less important basic parameters, the following two threshold values are used:

—ξ is the minimum importance of a new parameter at which the parameter is usable (ξ is defined as the percentage from the maximum importance);

—δ is the minimum value of the coefficient in the transformation matrix of basic parameters (if, for the selected new parameters, all coefficients at a certain basic parameter are below δ, then this basic parameter is eliminated).

Upon training the PCA on the initial data, a list of new parameters is created that represents the new parameters via initial parameters and specifies the importance of the former. Based on the importance, different methods of reducing the dimensionality imply different objects; in particular, for the PCA, the importance is eigenvalues of the analyzing matrix. New parameters are represented as a linear combination of basic parameters and, to calculate new parameters, a matrix of linear representation coefficients is constructed.

The PCA is characterized by the need to scale the input data. In the detection system, upon extracting basic parameters, the following three scaling techniques are used:

—The maximum and minimum values for each basic parameter are found and this parameter is translated into the interval $[-1; +1]$ by the formula $y = 2(x - min)/(max - min) - 1$, where x is the basic parameter and y is the scaled parameter.

—The largest possible value of the basic parameter is taken as the maximum value (for example, the maximum lifetime of the IP packet is $2^8 = 256$) and all basic parameters are divided by their maximums.

—The maximum and minimum values for all basic parameters are found and all basic parameters are translated by the formula $y = 2(x - min)/(max - min) - 1$.

For different types of data (various protocols and specific attacks), the best results are obtained by different data-scaling techniques.

The output of this block is the vectors containing attack labels and lists of new parameters.

3. CLASSIFICATION OF NETWORK PACKETS

The classification block is responsible for training and testing the SVM, as well as selecting the best parameters of the SVM kernel.

Before training the system, the scaling by the first technique (Section 2) is performed, since the SVM is also sensitive to the scale of data. For the SVM, the selection of the kernel is essential. The following four types of the SVM kernel are the most popular:

—linear, i.e., u'*v;

—polynomial, i.e., (gamma*u'*v + coef0)^degree;

—radial basis function, i.e., exp(−γ *|u − v|^2);

—sigmoid, i.e., tanh(gamma*u'*v + coef0).

In addition to kernel selection and kernel parameter adjustment, the effectiveness of the SVM also depends on the C parameter, which controls the error tolerance and the position of the hyperplane. This

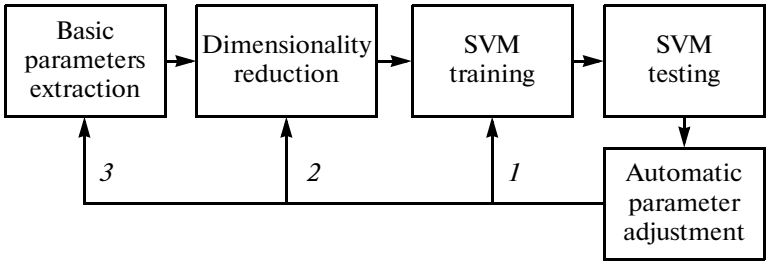

Fig. 2. Process of automatic parameter adjustment.

parameter allows one to adjust the ratio between the maximum width of the separator and the minimum total error.

SVM parameters are selected by constructing a search grid for the radial basis kernel according to the parameters C and γ.

The training procedure yields a SVM model that is used to classify the vectors.

To test the SVM model, the following characteristics are evaluated: the number of correctly detected attacks (TP), the number of false positives (FP), the number of correctly classified normal packets (TN), the number of missed attacks (FN), and the percentage of correctly classified vectors.

4. AUTOMATIC ADJUSTMENT OF PARAMETERS

To investigate the applicability of various dimensionality reduction methods to the problem under study, the automatic adjustment of parameters is required to find the best ones in terms of attack detection. An algorithm for the automatic adjustment of SVM and PCA parameters is embedded into the architecture of the intrusion detection system. Figure 2 shows a simplified scheme of this algorithm.

In the block of automatic parameter adjustment, the decision on modifying the parameters, algorithms, and properties of the other blocks (Fig. 2) is made based on the results obtained by the previous blocks of the chain. The table shows all automatically adjusted parameters for the basic modules of the system.

The purpose of this block is to achieve the highest percentage of correctly classified packets and the highest speed of the system. The process of parameter adjustment consists of three nested loops (Fig. 2, see *1*–*3*).

Upon receiving the best result of SVM testing (minimum FP, minimum FN, and minimum number of support vectors in the SVM model), the adjustment block modifies the parameters of the dimensionality reduction block, and the same cycle of operations in the SVM module is repeated. The automatic adjustment of the PCA (dimensionality reduction block) is reduced to selecting the analyzed matrix and to modifying the thresholds δ and ξ.

The block of automatic adjustment finds the best parameters of the system as follows:

—minimum FP and FN values characterize the quality of attack detection;

—the minimum number of basic parameters, new parameters, and support vectors in the SVM model characterizes the speed of the system.

Table

Extraction of basic parameters	Dimensionality reduction	SVM training
List of basic parameters	Analyzed matrix	SVM kernel
Bit capacity of parameters	Training set	Parameter C
Scaling formula	Threshold parameter δ	Kernel parameter γ
	Threshold parameter ξ	Kernel parameter *degree*
	Scaling formula	Kernel parameter *coef0*

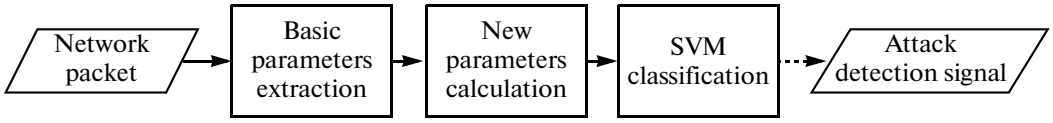

Fig. 3. Classification of network packets in the attack-detection module.

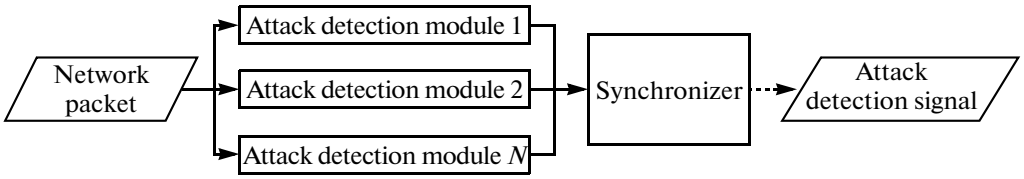

Fig. 4. Simplified scheme of the intrusion-detection system.

5. INTERACTION OF ATTACK-DETECTION MODULES

The main feature of the modular approach to attack detection is the division of all network attacks into groups for a particular detection module with detailed adjustment. Each group of attacks has its own network filter, type of packets to be analyzed (TCP, UDP, ICMP, and IP), list of basic parameters, and setting of the SVM and dimensionality reduction.

At the stage of data extraction, a separate set of vectors, i.e., a detection synchronizer, is constructed. These vectors contain the ordinal number of a packet within the attack and a list of system modules that respond to this attack. Each attack-detection module independently performs the operations of dimensionality reduction, SVM training, SVM testing, and parameter adjustment. In addition, the SVM model is constructed for the vectors of the synchronizer. Figure 3 shows a simplified scheme of traffic analysis based on an individual attack-detection module.

Once all new parameters are determined, joint lists of network filters, basic parameters, and new parameters are created. All of the parameters used in the system have a common numbering, which considerably reduces the amount of computations (basic parameters are extracted from network packets only once).

In the process of analysis, each packet is checked on a number of network filters; next, for each filter, all basic parameters are extracted and new parameters are calculated. Then, the vector of new parameters is classified on the SVM models of the corresponding attack classes. All signals from attack-detection modules are recorded into a separate vector, which is classified on the SVM model of the synchronizer. Figure 4 shows a simplified scheme of the intrusion detection system.

Due to the synchronizer, all signals from individual modules are filtered, and only a verified signal is returned. Therefore, the number of false responses of the whole system is considerably reduced and complex attacks, which include a lot of packets, can be identified as a single attack rather than a sequence of attacks that consist of individual packets.

REFERENCES

1. Aivazyan, S.A., Bukhshtaber, V.M., Enyukov, I.S., and Meshalkin, L.D., *Prikladnaya statistika: Klassifikatsiya i snizhenie razmernosti: Spravochnoe izdanie* (Applied Statistics: Classification and Dimension Reduction: Reference Guide), Moscow: Finansy i Statistika, 1989.
2. Hsu, Ch.-W., Chang, Ch.-Ch., and Lin, Ch.-J., A Practical Guide to Support Vector Classification. http://www.csie.ntu.edu.tw/~cjlin
3. Fodor, I.K., *A Survey of Dimension Reduction Techniques,* U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory, 2002.
4. van der Maaten, L.J.P., Postma, E.O., and van den Herik, H.J., *Dimensionality Reduction: A Comparative Review,* Maastricht: Maastricht University, the Netherlands.
5. Carreira-Perpiñán, M.Á., *A Review of Dimension Reduction Techniques. Technical Report CS-96-09,* 1997.

Translated by Yu. Kornienko