World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:1, 2016

Anomaly Detection with ANN and SVM for Telemedicine Networks

Edward Guillén, Jeisson Sánchez, Carlos Omar Ramos

***Abstract*—**In recent years, a wide variety of applications are developed with Support Vector Machines -SVM- methods and Artificial Neural Networks -ANN-. In general, these methods depend on intrusion knowledge databases such as KDD99, ISCX, and CAIDA among others. New classes of detectors are generated by machine learning techniques, trained and tested over network databases. Thereafter, detectors are employed to detect anomalies in network communication scenarios according to user's connections behavior. The first detector based on training dataset is deployed in different real-world networks with mobile and non-mobile devices to analyze the performance and accuracy over static detection. The vulnerabilities are based on previous work in telemedicine apps that were developed on the research group. This paper presents the differences on detections results between some network scenarios by applying traditional detectors deployed with artificial neural networks and support vector machines.

***Keywords*—**Anomaly detection, back-propagation neural networks, network intrusion detection systems, support vector machines.

I. Introduction

DUE to evolution of new technologies the number of security threats over communication networks has increased. Recent malware attacks increases research activity in computer security developing different elements such as Intrusion Detection System -IDS. IDS are components located in a host or networks to monitor data connections for generate alerts based on policy intrusions and malicious events. IDSs are an effective tool to detect zero-day attack compared to firewall, antivirus, and other security tools [1]. According to network properties and location, IDS are usually classified into two types, Host Intrusion Detection System –HIDS and Network Intrusion Detection System –NIDS [2], [3]. HIDS are susceptible to malware attacks by its poor visibility of network state, thus NIDS performance to monitor the network is higher than HIDS [4]. Related works in IDS field defined two main categories based on detection approaches: anomaly detection and misuse detection [1]-[5]. Anomaly detection has a wide range of applications such as natural sciences, medicine, and data security, among others. The anomaly detection approach in computer security is commonly carried-out by machine learning techniques, as Support Vector Machines -SVM and Artificial Neural Networks -ANN. SVM and ANN are classification methods in supervised machine learning discipline; they are represented in Fig. 1 [6]. SVM has additional advantages over some other techniques. Some of these advantages are higher accuracy on attack detection, and convergence time.

Edward Guillén, Jeisson Sánchez, and Carlos O. Ramos are with the Military University Nueva Granada Bogotá, Colombia (e-mail: edward.guillen@unimilitar.edu.co, gissic@unimilitar.edu.co, carlos.ramos@unimilitar.edu.co).

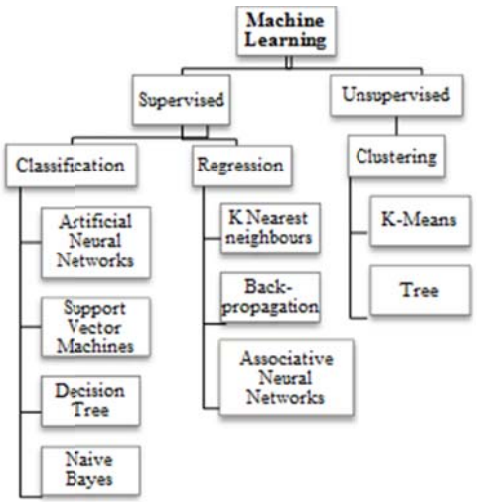


Fig. 1 Intrusion Detection System Taxonomy

Recently, telemedicine apps are developed to monitor health of patients by taking advantage of modern sensors and connectivity in smartphones and IoT –Internet of Things– appliances. Nevertheless, a wide number of medical apps have been affected by multiple vulnerabilities as it is explained in [7]. To diminish the risk of such vulnerabilities it is necessary to improve detection and prevention systems according to standards of computing security and to ensure the user information. Machine learning techniques are effective solutions in order to detect and prevent vulnerabilities on mobile apps.

Several studies have employed ANN techniques in computer security for detecting new threats such as zero-day attacks [5], [8], [9]. ANN commonly has five main processes: create the network, configure the weights and biases, training, validation, and testing network processes. In Artificial Neural Networks, the training process has a high computational cost but the classification process is highly efficient. In this case, the neural network is represented by a feed-forward model. Multiclass processes are useful in ANN for modeling multiple features and generate a simple target. In 2014, Devaraju and Ramakrishnan compared the performance with multiple classifiers based on different types of neural networks for IDS. Feed-Forward Neural Network -FFNN was tested against Denial-of-Service -DoS, Probe, Remote-to-Local -R2L and User-to-Root -U2R attacks. With this approach, the efficiency

for detecting Probe attacks was high. They are widely used in pattern recognition [10].

Instead of traditional process of machine learning techniques, SVM compute the features through kernels. The kernels are powerful elements to process data and solve the complexity of multiclass classification. Therefore, the accuracy (low false positives) on profiles generated is more efficient on SVM. Macek et al. [11] worked with SVM to reduce the false negative rates detecting U2R and R2L attacks with accuracy changes. SVM as machine learning technique is very effective against common attacks [11]. Hasan et al. in 2014, worked with SVM and Random Forest- RF improving the common KDD99 dataset to train and test machine learning methods. SVM as technique in machine learning was effective with probes attacks and other type of attacks [12].

The results are according to detection rate from SVM and ANN techniques testing stage, when the detectors are tested in different network scenarios.

The paper is divided in five sections. In Section II, the processing methods and machine learning techniques are explained. In Section III, the results are defined by statistics correlations based in Root Mean Squared Error measure and comparing the differences between each detector. In Section IV an analysis according to the targets and results are discussed to science contribution. Finally, in Section V a conclusion about the difference of training data in each detector is presented.

II. Processing Methods and Machine Learning Techniques for Intrusion Detection System

A. Data Collection and Feature-Selection Method

Datasets were collected with Spleen application from network connections [13], [14]. Data has similar features from Knowledge Discovery Databases, such as KDD99 [15]. Datasets acquired by Spleen has additional attributes and labeled to normal or abnormal behavior. Labels are filtered according to protocols, ports, IP address among others network features. The best representative features were processed by Principal Component Analysis –PCA method. PCA delete the redundant information based on the null fields and low feature variation between each connection. PCA as unsupervised method is effective for non-linear of a high dimensional data [16].

Features selected are based on Basic and Behavior Change Detector -BCD attributes. The information collected by Spleen application is based on connection or packet features. Basic features are represented by information collected through connections. The host traffic features are based on different states of the data sent between client-server as on Table I.

The BCD features selected are the average of the payload size represented in bytes, host-client count in the last 255 connections, connections from the client to the current host and States S0 and others connections to the current host. Machine learning techniques are multi-class process trained with 8 input variables selected.

TABLE I
Information Collected between Client-Server Based on Three Hand-Shake Process

State	Description
SF	Successful Connection
S0	An initial Synchronization packet was send but there is not response from the server
S1	Three-hand shake process successful but there is not packets activity
S2	Successful connection, client close the connection
S3	Successful connection, server close the connection
RSTO	Reset connection from source
RSTR	Reset connection from destination
REJ	Reject connection, a synchronization packet send a RST packet
OTH	Other

B. Feed-Forward Neural Network as Machine Learning Technique

Presently a multiple input nonlinear neural networks are deployed in different applications and similar stages: training, validation and testing stage. Feed-Forward Neural Network - FFNN such as supervised model, it is trained based on weights and biases to optimize the results on testing. According to taxonomy of neural network architectures, FFNN are classified into three types: single-layer perceptron, multilayer perceptron, and radial basis function networks [17]. The data classification is performed by prediction over different epochs during short processing times [1]. FFNN algorithm is Multilayer Perceptron -MLP. This algorithm is composed by two phases, forward phase and computational phase of an error signal as represented in (1). In the equation, d_i is the desired response and y_i represents the real output. The MLP algorithm learns whereas is trained; changing weights after the data is processed. Learning method occurs when there is a difference between a minimum errors compared to expected output [10].

$$e_i = d_i - y_i \tag{1}$$

MLP network has multiple layers that represent connections between the neurons of each layer to the next layer. MLP architecture usually has three layers: Input layer, hidden layer and output layer. Resulting network is approximate to any nonlinear function. Then the total error is estimated by total output neurons as in (2), where *m* is the number of neurons on the output node [18].

$$e_T = \frac{1}{2}\sum_{i=1}^{m}(d_i - y_i)^2 \tag{2}$$

The output layer is based on targets: normal behavior or anomaly behavior. The algorithm is with usual sigmoid function as activation function and data was normalized to 0 and 1 values. The normalization is to generate an optimization in training stage. The ANN computational model is shown in Fig. 2.

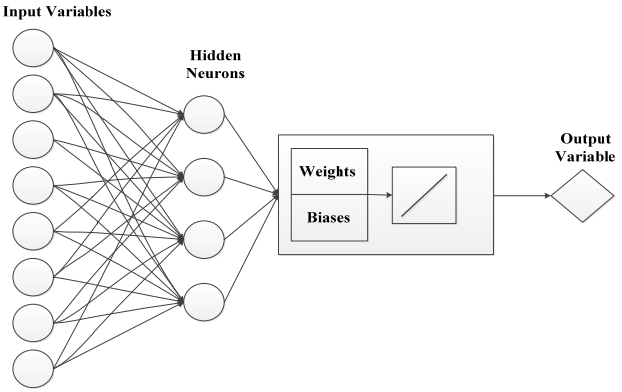


Fig. 2 Architecture of a multi-layer perceptron network

C. Support Vector Machines as Machine Learning Technique

Despite SVM is a solution in multiple computer security applications; however, it is not a stable technique for IDS non-linear classification [19]. The learning algorithm is based on Stefan Rueping's work and is shown in (3). L is the number of samples divided by the number of vectors. [20], [21]. The algorithm is well suited for regression and classification methods.

$$\Psi(w,\xi,\xi^*) = \frac{1}{2}(w^T w) + C(\sum_{i\in I}\xi_i + L\sum_{i\in S}\xi_i) \quad (3)$$

Multi-class data is processed by polynomial kernels from SVM approaches. The polynomial kernel is shown in (4) with d as the SVM kernel parameter degree. Polynomial kernel has a correct performance with normalized data in training stage [21].

$$k(x,y) = (x * y + 1)^{d} \quad (4)$$

Convergence epsilon is an optimizer parameter selected to specify the accuracy on Karush-Kuhn-Tucker conditions. Karush-Kuhn-Tucker as mathematical conditions to optimize a solution can solve problems for machine learners [22]-[24]. Support Vector Machines architecture was organized such as in Fig. 3 to train, validate, and test each network data.

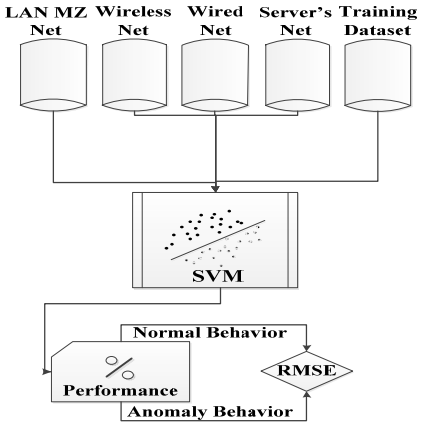


Fig. 3 SVM structure tested with four real-network datasets

D. Network Scenarios

Each network scenario has a number of connections based on the user´s connections and request to services. Three way handshake process represents a frame of real-data transmission in a network connection and it is obtained with Spleen as it is shown in Fig. 4. Spleen is an application for monitoring network connections by record behavioral features in the segment. The connections are based on mobile and non-mobile devices and the network architecture is based on Telemedicine apps running onto mobile devices.

1. LAN MZ Network

Subnetwork with management data traffic with a VLAN topology. Militarized zone is a special segment of the local network with E-mail server, Web Server, DHCP server among others. Lan MZ dataset has approximately 111.000 connections and typical services as HTTPS, HTTP, WINDOWS DS, RRAC.

2. Server's Network

Is a scenario with DNS, Streaming, DHCP, Web servers and a special firewall configuration. Servers network dataset has approximately 100.100 connections.

3. Wireless Network

It is the largest network of use and coverage, Wireless dataset has approximately 105.000 connections and typical services as e-mail, streaming, data, and audio transmission.

4. Wired Network

It is a network in a private environment with World Wide Web traffic and different request services. Wireless dataset has approximately 35.000 connections and 20 target host.

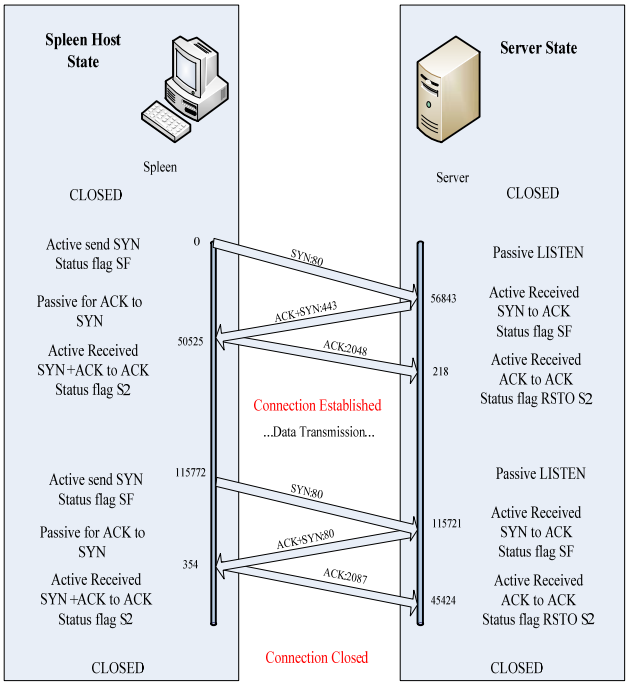


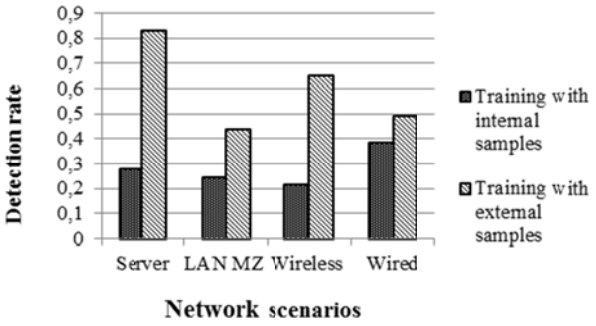
Fig. 4 Representation of real-data transmission in a network connection through a three-handshake process

III. Accuracy and Detection Rates

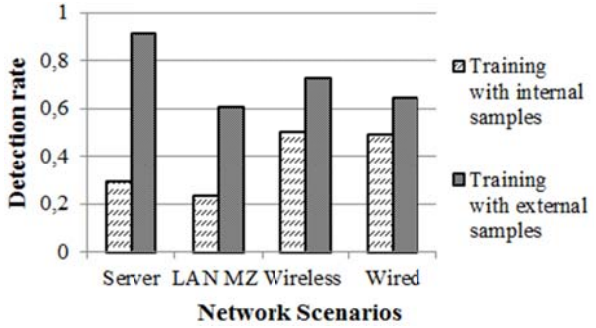
The attacks were labeled as numerical input in each machine learning technique. A normal behavior is labeled as value of zero and attack is labeled with one value. Attacks type was probes. In Fig. 5 is illustrated the Root Mean Squared Error –RMSE results in ANN and SVM techniques. RMSE is shown in (5). The difference between *A* and *F* is the estimated parameter and the result [17].

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[(A_t - F_t)]^2} \quad (5)$$

According to detection, rates in SVM and ANN to create detectors the accuracy was tested by classification operators and the results were false positives. The effectiveness in SVM technique to detect abnormal behaviors was better than ANN detectors; therefore, the accuracy is higher in SVM. The false positives in each artificial intelligence technique are shown in Fig. 6.



(a)



(b)

Fig. 5 (a) Root Mean Squared Error in Artificial Neural Network results with internal examples training and testing, (b) Root Mean Squared Error in Support Vector Machines results with internal examples training and testing

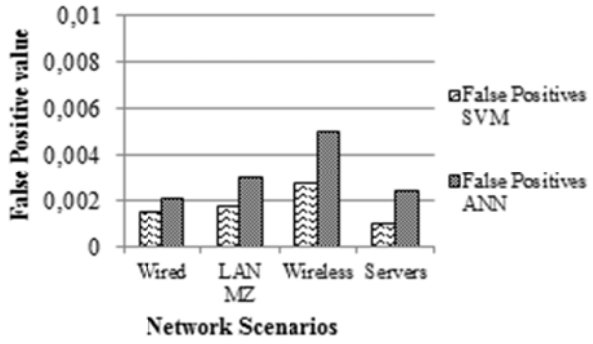


Fig. 6 False Positives rate on Support Vector Machines and Artificial Neural Networks detectors

IV. Discussion and Analysis

Different intrusion detectors are deployed by methods of machine learning every day. However, the increasing of malware attacks, insider, and outsider vulnerabilities, policy violations in companies undertake the researches based on security information to be effective. The effectiveness in a machine learning technique is according to training samples and its adaptability to a network features.

Results are based on training stage. Weights and biases are the input samples from the training dataset to each network. As expected in ANN and SVM, the best performance is when each network is trained with internal datasets samples. The difference is during testing stage, where a dataset is processed by a classifier and tested over the same data or other dataset. RSME in SVM training stage had the best performance. Therefore by results, the average difference value between training with internal and external samples was 32%, whereas a 0.05% in the same network. Multiple solutions are proposed for the dissimilarity between communication networks behavior as transfer learning. Transfer learning is still in research and does not fit properly to the changes of network features.

V. Conclusion

Detectors as result of machine learning techniques are deployed in different applications for multiple targets. Although the neural networks and Support Vector Machine are useful techniques for the adaptive learning and flexibility to network variations, it has not a reliable accuracy when the detector is tested in dissimilar networks scenarios. Anomaly detection is a convenient method of learning to detect new security threats and reduce the search space compared with misuse detection.

Old Datasets used in multiple approaches are useful to estimate new detectors and evaluate its effectiveness but not for detection in real network with dissimilar features. New data collection is necessary to detect new attacks deployed every day.

Acknowledgments

This work was possible in part with the financial support of Military University Nueva Granada with the project ING-

1772.

REFERENCES

[1] Shah, B. & Trivedi, B. H, Improving Performance of Mobile Agent Based Intrusion Detection System, in 'Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on', pp. 425-430, (2015).

[2] Hoque, M. S.; Mukit, M.; Bikas, M.; Naser, A. & others, 'An implementation of intrusion detection system using genetic algorithm', arXiv preprint arXiv: 1204. 1336. (2012).

[3] Rahman, M. & Cheung, W. M, 'A Novel Cloud Computing Security Model to Detect and Prevent DoS and DDoS Attack', International Journal of Advanced Computer Science and Applications (IJACSA) 5(6), (2014).

[4] Bhat, A. H.; Patra, S. & Jena, D., 'Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines', International Journal of Application or Innovation in Engineering & Management (IJAIEM) 2(6), 56-66, (2013).

[5] Kim, G.; Lee, S. & Kim, S., 'A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection', Expert Systems with Applications 41(4), 1690-1700, (2014).

[6] Murphy, K. P., Machine Learning: A Probabilistic Perspective, MIT press, (2012).

[7] Cifuentes, Y.; Beltrán, L. & Ramirez, L., 'Analysis of Security Vulnerabilities for Mobile Health Applications', 2015 Seventh International Conference on Mobile Computing and Networking (ICMCN 2015).

[8] Mirza, N. A. S.; Abbas, H.; Khan, F. A. & Al Muhtadi, J., Anticipating Advanced Persistent Threat (APT) Countermeasures Using Collaborative Security Mechanisms, in 'Biometrics and Security Technologies (ISBAST), 2014 International Symposium on', pp. 129-132, (2014).

[9] Saied, A.; Overill, R. E. & Radzik, T., Artificial Neural Networks in the Detection of Known and Unknown DDoS Attacks: Proof-of-Concept' Highlights of Practical Applications of Heterogeneous Multi-Agent Systems. The PAAMS Collection', Springer, pp. 309--320, (2014).

[10] Devaraju, S. & Ramakrishnan, S., 'Performance Comparison for Intrusion Detection System Using Neural Network with KDD Dataset.', ICTACT Journal on Soft Computing 4(3), 743-752, (2014).

[11] Macek, N. & Milosavljević, M. (2014), 'Reducing U2R and R2L Category False Negative rates with support vector machines', Serbian Journal of Electrical Engineering 11(1), 175-188.

[12] Hasan, M. A. M.; Nasser, M.; Pal, B. & Ahmad, S., 'Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS)', Journal of Intelligent Learning Systems and Applications 2014, (2014).

[13] Guillén, E.; Rodriguez, J.; Páez, R.; Rodriguez, A. Detection of non-content based attacks using GA with extended KDD features. In Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, 24-26 October 2012; pp. 30–35.

[14] Guillén, E.; Rodriguez, J.; Páez, R. Evaluating Performance of an Anomaly Detection Module with Artificial Neural Network Implementation. Int. J. Comput. Inf. Syst. Control Eng. 2013, 7, 836–842.

[15] Stolfo, S. J.; Fan, W.; Lee, W.; Prodromidis, A. & Chan, P. K., Cost-based modeling for fraud and intrusion detection: Results from the JAM project, in 'DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings', pp. 130--144, (2000).

[16] Kim, B.-J. & Kim, I. K. (2005), Machine learning approach to real time intrusion detection System' AI 2005: Advances in Artificial Intelligence', Springer, pp. 153-163.

[17] Pati, J. & Shukla, K., A comparison of ARIMA, Neural Network and a Hybrid Technique for Debian Bug Number Prediction, in 'Computer and Communication Technology (ICCCT), 2014 International Conference on', pp. 47-53, (2014).

[18] Ndiaye, A.; Thiaw, L.; Sow, G. & Fall, S., 'Development of a Multilayer Perceptron (MLP) Based Neural Network Controller for Grid Connected Photovoltaic System', Int. J. Phys. Sci 9(3), 41--47, (2014).

[19] Meyer, D. & Wien, F. T., 'Support Vector Machines', The Interface to Libsvm in Package e1071, (2014).

[20] Rüping, S., Incremental Learning with Support Vector Machines, in '2013 IEEE 13th International Conference on Data Mining', pp. 641-641, (2001).

[21] Rüping, S., 'Incremental Learning with Support Vector Machines', Technical Report, Technical Report, SFB, 475: Komplexitдtsreduktion in Multivariaten Datenstrukturen, Universitдt Dortmund, (2002).

[22] Mei, S. & Zhu, X., Using Machine Teaching to Identify Optimal Training-Set Attacks on Machine Learners, in Proceedings of Association for the Advancement of Artificial Intelligence, Austin, Texas USA. January 25 –30, 2015, pp. 2871–2877, (2015).

[23] Levy, B. C., Principles of Signal Detection and Parameter Estimation, Springer Science & Business Media, (2008).

[24] Murphy, K. P., Machine Learning: A Probabilistic Perspective, MIT Press, (2012).