EasyChair Preprint

№ 129

Intrusion Detection with Probabilistic Neural Network: Comparative Analysis

Ibrahim Atay

May 15, 2018

International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES'18), May 11-13, 2018 Safranbolu, Turkey

Intrusion Detection with Probabilistic Neural Network: Comparative Analysis

İ.ATAY[1]

[1] Okan University, Istanbul/Turkey, contact@ibrahimatay.com

***Abstract* - The use of machine learning techniques has significantly increased recently. The classification of normal or abnormal situations in network traffic is successfully applied with machine learning techniques. It is possible to encounter False Positive situations during the classification process. With Probabilistic Neural Network (PNN) model, it is aimed to explore the intrusion and its types within network traffic with probabilistic distribution. Knowledge Discovery Dataset (KDD99) will be used in this study.**

***Keywords* - Probabilistic Neural Network, Intrusion Detection, KDD99.**

I. INTRODUCTION

The use of machine learning techniques has significantly increased recently. Especially, intrusion detection applications have increased with machine learning techniques. Intrusion detection systems monitor the information packages within the network and detect the normal and abnormal situations within the scope of the given rules

There are various data sets available in the literature to experience the success of intrusion detection systems. DARPA Intrusion Detection Evaluation (DARPA) and Knowledge Discovery Dataset (KDD99)[1] data sets are the most commonly used ones in the literature. Said data sets contain the traffic status of normal and abnormal situations realized in the network traffic.

Various machine learning techniques have been used to detect the intrusion within network traffic. Decision Tree, Support Vector Machines and Bayes are the most commonly used techniques in the studies [2-3]. However; it is an important factor to reach acceptable False Positive values in the techniques applied.

There are two approaches in intrusion detection systems being abuse and abnormality detection [4]. Detection of abuse involves the intrusion situations which are well-known and applied in the past. It is possible to detect such situations with the defined rules within the system. Detection of abnormality occurs with encountering unexpected situations within the network traffic. It is required to know the network traffic history in order to be able to detect the unexpected situations.

In this article, it is aimed to detect the normal and abnormal situations on the network traffic in the light of statistical values, using Probabilistic Neural Network (PNN)[5] model. The application has been developed using Visual C#. KDD99[6, 7] data set has been used in the study.

The article includes information about the data set in 2nd section, literature search about the study in 3rd section, problem to be examined in 4th section, Probabilistic Neural Network(PNN) in 5th section, the application realized in 6th section, and the conclusions of the study in 7th section.

II. DATASET

Knowledge Discovery Dataset (KDD99) is the data set used for performance comparison in intrusion detection systems. Data set includes unprocessed TCP data which was monitored for 9 weeks. Training data set contains 7-week and 5 million connections. Test data set contains 2-week and 2 million connections.

Test data set includes additional intrusion situations which are not included in training data set in order to experience the success of the intrusion situations. Training data includes 39 intrusion situations, and test includes 22 intrusion situations different from training data [1, 6]. The intrusions in data set are classified in 4 main categories. These categories are Denial of Service (DOS), Probing (Probe), Remote to Local (R2L) and User to Root (U2R).

There are 494,021 records in KDD99 data set. There are 97,277 (19.69%) normal, 391,458 (79.24%) DOS, 4,107 (0.83%) Probe, 1,126 (0.23%) R2L and 52 (0.01%) U2R intrusion connections in the records [7]. There are 41 attributes in data set and each has a label assigned as intrusion type or normal. Table 1 shows the labels and number of samples that appears within the data set.

Table 1: Class labels and the number of samples that appears [7].

Attack	Number of Samples	Minus Repeated Number of Samples	Class
Black	2,203	994	DOS
Land	21	19	DOS
Neptune	107,201	51,820	DOS
Pod	264	206	DOS
Smurf	280,790	641	DOS
Teardrop	979	918	DOS
Satan	1,589	908	PROBE
Ipsweep	1,247	651	PROBE
Nmap	231	158	PROBE
Portsweep	1,040	416	PROBE
Normal	97,277	87,831	NORMAL

Guess_passwd	53	53	R2L
ftp_write	8	8	R2L
Imap	12	12	R2L
Phf	4	4	R2L
Multihop	7	7	R2L
Warezmaster	20	20	R2L
Warezclient	1,020	1,020	R2L
Spy	2	2	R2L
Buffer_overflow	30	30	U2R
Loadmodule	9	9	U2R
Perl	3	3	U2R
Rootkit	10	10	U2R

III. RELATED WORKS

First intrusion detection system was designed by James P. Anderson[12] in 1980. Anderson designed the system so as to monitor, inspect and control all incidents occurring in computer systems or network traffic. If there is any security problem encountered within the process, a warning will be sent to the relevant personnel and unit and the measures will be taken for possible risks. Figure 1 shows the design of Anderson.

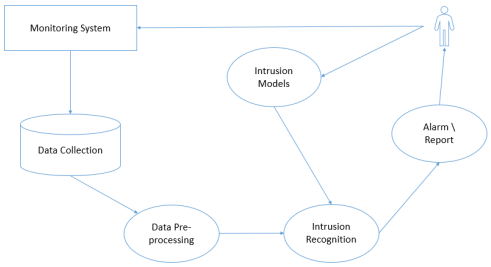

Figure 1: James P. Anderson Intrusion Detection Model.

The types of intrusion which can be applied within network traffic are described below.

- **Denial of Service (DOS):** Intrusion is made by sending connection request to a server more than it can carry, exploiting the gaps of TCP/IP protocol. The use of service by the actual users is prevented as a result of the intrusion.
- **Probing (Probe):** It is made by monitoring the active server or systems to learn IP address, open port or operating system.
- **Remote to Local (R2L)** : These are the situations in which the unauthorized access by users is encountered.
- **User to Root (U2R):** It includes the situations in which the user can make transactions beyond his/her authorization.

Various machine learning techniques have been used in the study to detect intrusion in network traffic. Decision Tree, Support Vector Machines and Bayes are the techniques which are most commonly used in the studies.

IV. PROBLEM DEFINITION

There are many traffic patterns within network traffic. Many intrusion patterns are recorded in databases thanks to past experience. However, developing computer systems brought with them the formation of different intrusion patterns.

In the intrusion model designed by James P. Anderson[8], there are network traffic monitoring, inspecting and controlling functions. However, the intrusion techniques differ with the development of computer systems. Therefore, machine learning models which can form patterns regarding new intrusion situations should be used within network traffic.

It is required to learn based on past data and to form new patterns based on new data in the process of forming new patterns in intrusion detection applications. This will cause encountering False Positive situations. However, the success of the process has increased with the use of acceptable threshold value in the study.

V. PROBABILISTIC NEURAL NETWORK (PNN)

Probabilistic Neural Network(PNN) [5] model is also known as Bayes-Parzen in the literature. PNN has the same advantages as statistical classification model. Additionally; Back-Propagation Neural Network model has advantages with its calculation power.

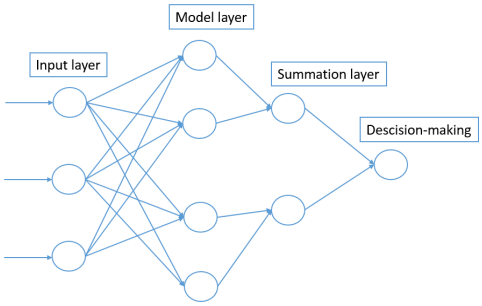

Figure 2: Layer PNN Model.

PNN can be designed with 4 layers as shown in figure 2. In the first layer of the design, the attributes are input and the weights are calculated (W). In the second layer, the connection between the input attributes and weight values should be established. Using the transfer function, vector at the size of 1xN is added to the connections established. Transfer function is shown in figure 3.

$$S_{ij} = exp\,(-\frac{||\,X_j\, - X_i||}{\sigma^2})$$

Figure 3: PNN Transfer Function [7]

Transfer function range is 0 and 1. As the weight distances between the transferred vector and model vector increase, output values are approached. In the third layer, close neural templates are formed. Number of neural and number of model is equal at this layer. Layer function is shown in Figure 4.

$$f_k(x_i) = \frac{1}{M_k} \sum_{\forall x_i \in y_k} S_{ij}$$

Figure 4: 3 Function of formation of layer template values [7]

In the fourth section, vectors are classified depending on Bayesian decision rules. Layer function is shown in figure 5.

$$P(y_i = 1\,|X_i) > \frac{L_{1,0}(X_i) - L_{0,0}(X_i)}{L_{0,1}(X_i) - L_{1,1}(X_i) + L_{1,0}(X_i) - L_{0,0}(X_i)}$$

Figure 5: Bayesian function [7]

When erroneous classification is assumed in the transaction, situation input is made on the vector. In the other case, vector distribution is applied. Therefore, correct classification cost is always higher. In the function shown in Figure 5, classification score range is defined as [1].

VI. APPLICATION

In the network intrusion model designed by James P. Anderson [12], there are monitoring, inspecting and controlling stages. As KDD99 data set is used in the study, only controlling stage will be realized. The study has been performed using Visual C# language. The application has been developed asynchronously due to the process it realizes.

PNN model has been realized with the application prepared. PNN model consists of two parts being training and test. In training part, Bayesian model is used on data, and the classification model as shown in Figure 6 is formed. In Test, the classification model as shown in Figure 7 is applied with the experience obtained in training part.

Begin

Initialization

J = 0;

Do

J ← j + 1

Normalization: $x_{jk} \leftarrow \frac{x_{jk}}{\sqrt{\sum_i^d x_{ji}^2}}$;

Learning process: $w_{jk} \leftarrow x_{jk}$;

If x ∈ w_i then $a_{ic} \leftarrow 1$;

Until j = n;

End

Figure 6: Training Algorithm

Begin

Initialization

k = 0

x = test sample

Do

$k \leftarrow k + 1; z_k \leftarrow w_k^t x$

If $a_{kc} = 1$ **then** $g_c \leftarrow g_c + e^{\frac{x_k - 1}{\sigma^2}}$;

Until k = n;

Return class ← $arg\,max(g_i(x))$;

End

Figure 7: Test Algorithm

A series of transactions has been applied before KDD99 data set. As shown in Table 1, repeated transaction records have been cleared.

Table 2: Performance Values of Machine Learning Models related to KDD99 Data Set Classification.

	DOS	**PROBE**	**R2L**	**U2R**
Bayes	%99,62[9]	%100 [9]	%99,35[14]	%99,47[14]
Support Vector Machine	%100[10]	%100[12]	%99,42[15]	%100[10]
Decision Tree	%99,98[11]	%99,66[13]	%99,70[16]	%92,5[16]
Probabilistic Neural Network(PNN)	**%99,78**	**%99,30**	**%99,48**	**%99,85**

KDD99 data set classification study was performed with the application developed. The study is shown in Table 2 with performance comparison in literature study performed with KDD99 data set.

VII. CONCLUSION

This study has been prepared in order to model a system exploring the intrusion situations within the network traffic. Within the framework of the rules determined within network traffic, the most successful intrusion detection model is seen as Decision Tree. However, the intrusion situations realized outside the defined rules are not included in performance value. This situation may cause various loss in different intrusion situations.

PNN which is discussed in the study makes classification using Neural Network and Bayes statistical model. The success of the applied classification model will increase with the growth of training set. Additionally, memorization problem of Neural Network structures has been eliminated with the use of Bayes statistical model within the model. In Table 2, PNN has been compared to the other model applied on KDD99 data set in the literature.

REFERENCES

[1] P. Aggarwal and S. Kumar, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 842–851, 2015.

[2] J. Zhao, J. Zhao, and J. Li, "Intrusion detection based on clustering genetic algorithm," *2005 Int. Conf. Mach. Learn. Cybern.*, no. August, p. 3911–3914 Vol. 6, 2005.

[3] S. Sheen and R. Rajesh, "Network intrusion detection using feature selection and Decision tree classifier," *TENCON 2008 - 2008 IEEE Reg. 10 Conf.*, pp. 1–4, 2008.

[4] D. E. Denning, "An Intrusion-Detection Model," *IEEE Trans. Softw. Eng.*, vol. SE-13, no. 2, pp. 222–232, 1987.

[5] X. Huafeng, "Probabilistic Neural Network and Its Application," *2010 Int. Conf. Comput. Des. Appliations (ICCDA 2010)*, vol. 2, no. Iccda, pp. 0–3, 2010.

[6] "KDD Cup 1999 DataSet," 1999. .

[7] A. A. Olusola, A. S. Oladele, and D. O. Abosede, "Analysis of KDD & apos ; 99 Intrusion Detection Dataset for Selection of Relevance Features Analysis of KDD ' 99 Intrusion Detection Dataset for Selection of Relevance Features," vol. I, no. January, pp. 16–23, 2016.

[8] James P. Anderson, "Computer Security Threat Monitoring and Surveillance." 1980.

[9] F. Jemili and M. Zaghdoud, "Intrusion Detection based on ' Hybrid ' Propagation in Bayesian Networks," *Intell. Secur.*, 2009.

[10] Q. Mu, Y. Chen, and Y. Zhang, "Incremental SVM algorithm to intrusion detection base on boundary areas," *2012 Int. Conf. Syst. Informatics, ICSAI 2012*, no. Icsai, pp. 198–201, 2012.

[11] V. Sharma and A. Nema, "Innovative Genetic Approach for Intrusion Detection by Using Decision Tree," *2013 Int. Conf. Commun. Syst. Netw. Technol.*, pp. 418–422, 2013.

[12] J. Wang, T. Li, and R. Ren, "Real time IDSs based on artificial bee colony-support vector machine algorithm," *3rd Int. Work. Adv. Comput. Intell. IWACI 2010*, pp. 91–96, 2010.

[13] M. Bahrololum, E. Salahi, and M. Khaleghi, "Machine Learning Techniques for Feature Reduction in Intrusion Detection Systems: A

Comparison," *Comput. Sci. Converg. Inf. Technol. 2009. ICCIT '09. Fourth Int. Conf.*, pp. 1091–1095, 2009.
[14] D. M. Farid and M. Z. Rahman, "Anomaly network intrusion detection based on improved self-adaptive Bayesian algorithm," *J. Comput.*, vol. 5, no. 1, pp. 23–31, 2010.
[15] Y. Zhang, "Application of Improved Support Vector Machines in Intrusion Detection," no. 3, pp. 0–3, 2010.
[16] A. Alazab, M. Hobbs, J. Abawajy, and M. Alazab, "Using Feature selection for intrusion detection system," pp. 296–301, 2012.