Proceedings of the 9th Annual World Conference of the
Society for Industrial and Systems Engineering,
2020 SISE Virtual Conference
September 17-18, 2020

Validation of Decisions of a Multilayer Perceptron Learning Algorithm for the Identification of Net Attacks with the Aid of Bayesian Classifiers

Y Fang[1], M Kempf[2], and A Sauer[3]

[1]Ernst-Abbe-Hochschule Jena
University of Applied Sciences
Carl-Zeiß-Promenade 2
Jena, Germany

[2]Department of Sustainable Production and Quality Management
Fraunhofer Institute for Manufacturing Engineering and Automation (IPA)
Stuttgart, Germany

[3]Institute for Energy Efficiency in Production (EEP)
Stuttgart, Germany

Corresponding author's Email: Michael.Kempf@ipa.fraunhofer.de

Author Note: Yuqi Fang is a student of computer science and is specialized in learning algorithms and Artificial Neural Networks. Michael Kempf is engaged in quality management as well as product development. His research focuses on statistical methods in the fields of reliability engineering, process optimization and risk minimization. Alexander Sauer is executive director of the EEP and is concerned with the realization of innovative solutions for sustainable production in the industrial sectors of the future in Germany as a technology base, promoting specialist expertise and interdisciplinary collaboration among staff as the driving force behind qualitative growth.

Abstract: An intrusion detection system (IDS) is a software application that monitors the network for potential malicious attacks against a single computer or a computer network. A multilayer perceptron (MLP) learning algorithm is used detect such attacks and identifies the kind of attack like *WebAttack*, *DoS* or *BruteForce*. A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN), which consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Since ANNs belong to the so called black box algorithms, it is useful to validate its results. In this paper a method is presented to validate the decisions of the MLP algorithm concerning the type of net attack with the help of Bayesian Classifiers. Particularly the Naïve Bayesian Classifier and the Tree Augmented Naïve (TAN) Bayesian Classifier are used for this task. It will be shown that these classifiers are capable to satisfactorily validate the decisions of the MLP algorithm. This will be accomplished with aid of real datasets from the Canadian Institute for Cybersecurity along with appropriate metrics to evaluate Machine Learning algorithms.

Keywords: Intrusion Detection System IDS, Bayesian Classifiers, Neural Networks

1. Introduction

The trend in modern machinery and equipment together with the concept of Industry 4.0 is the ongoing automation of traditional manufacturing and industrial practices, using modern smart technology. Large-scale machine-to-machine communication and the internet of things are integrated for increased automation, improved communication and self-monitoring, and production of smart machines that can analyze and diagnose issues without the need for human intervention. These boundary conditions make it essential to deploy effective and efficient intrusion detection systems to avoid attacks that might cause malfunctions or major damage of machinery and equipment.

2. Background

2.1 Intrusion detection systems

With the development in the communication among computer systems, security threats have emerged in the process of information flow. Regarding the necessity to guarantee that the concerned data should not be misused or lost, researchers have made great effort to develop a more reliable computer network. Considering the substantial increase of various cyber-attacks and viruses, however, it is worthwhile to turn to extra vigorous techniques besides internal defenders or authorization structures built in the software. Thus, a great number of cybersecurity techniques have been presented in the last decades like firewalls, cryptography or intrusion detection systems (IDS). Of all the security techniques applied in cyber systems, IDS has obtained notable achievements in distinguishing sophisticated and dynamic attacks.

2.2 Bayesian Networks and Bayesian Classifiers

Probabilistic graphical models (Darwiche, 2014) are a framework of statistical models for encoding probability distributions where a graphical structure encodes a set of conditional dependence and independence relations over a set of random variables representing a problem domain. Bayesian networks are directed acyclic graphs (DAG) where the nodes represent events (random variables) with a finite set of states and the arrows stand for dependencies between any pair of nodes in the network. These networks allow bidirectional reasoning, namely from cause to effect as well as from effect to cause.

More technically, a Bayesian network is a pair (*G*,***P***), where *G*=(***V***,**E**) is a directed acyclic graph over a set of random variables ***V*** and ***E*** is a set of directed edges that represent probabilistic relationships between variables in ***V*** (Pearl, 1988). ***P*** is a set of conditional probability distributions (CPDs) that quantify the strength of the relations induced by ***E***. Specifically, ***P*** contains for each *V* in ***V***, the CPD *P(V|pa(V))*, where *pa(V)* is the set of parent variables of *V* in *G*.

Such a Bayesian network supports both diagnostic and prognostic reasoning by computing the posterior probability *P(H|e)* of an unobservable hypothesis *H* given observed evidence $e = \{e_1, .., e_m\}$, where each e_j is the observed state of the variables $E = \{E_1, .., E_m\}$ (Kjaerulff, 2014).

Bayesian classifiers are Bayesian networks with a relative simple structure. In a naïve Bayesian classifier there is a class variable with finite number of states as well as a finite number of descendants, the attributes. In Tree Augmented (TAN) Bayesian classifier there also links between pairs of attributes.

2.3 Multilayer Perceptron (MLP) Learning Algorithm for the identification of net attacks

The MLP Learning algorithm uses datasets from the Canadian Institute for Cybersecurity, namely a collection of data called CSE-CIC-IDS2018. Figure 1 shows the number of instances in the data subsets along with the type of attack. These data also contain 78 different attributes of the network traffic, the so-called features.

Set	Collected Time	Attack type	Original		
			Total	Benign	Attack
1	Friday Afternoon	DDoS	225745	97718	128027
2	Friday Afternoon	PortScan	286467	127537	158930
3	Friday Morning	Bot	191033	189067	1966
4	Monday	-	529918	529918	0
5	Thursday Afternoon	Infiltration	288602	288566	36
6	Thursday Morning	WebAttack	170366	168186	673
		BruteForce			1507
7	Tuesday	Patator	445909	432074	13835
8	Wednesday	DoS	692703	440031	252661
		Heartbleed			11
0		All	2830743	2273097	557646

Figure 1. Number of instance in each original dataset

ISBN: 97819384961-9-6

3. Bayesian Classifiers for the validation of the decisions of the MLP Learning Algorithm for the identification of net attacks

In the following two Bayesian classifiers are presented along with their results regarding the classification of net attacks. Both approaches will be analyzed and compared with the aid of appropriate machine learning metrics.

3.1 Naïve Bayesian Classifiers

The naïve Bayesian classifier shown in figure 2 has a node called *Label*, which represents the type of attack and 78 descendants which are the attributes of the network traffic, the so-called features.

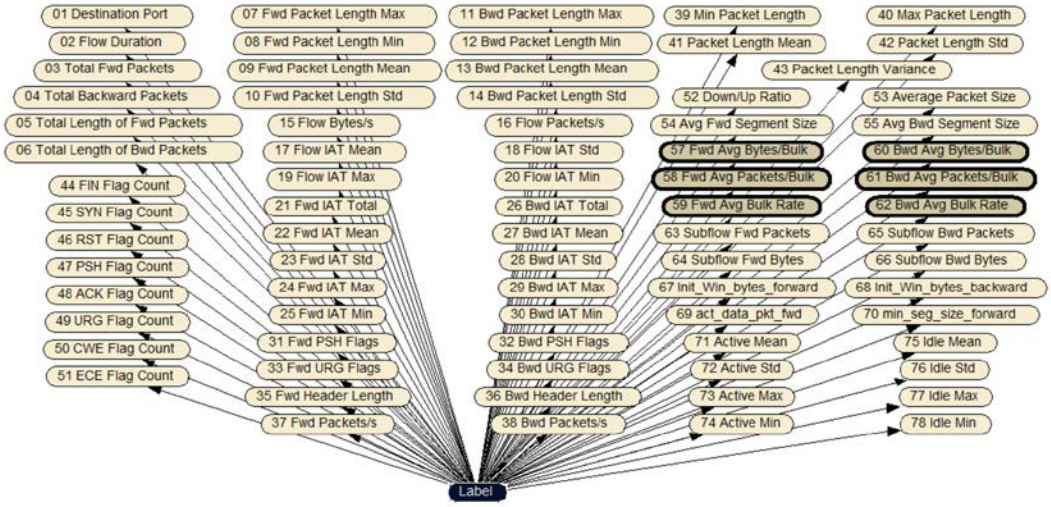

Figure 2. The naïve Bayesian classifier modeled with Netica

3.2 Tree Augmented Naïve (TAN) Bayesian Classifier

The TAN Bayesian Classifier has, in addition to the structure of the naïve Bayesian Classifier, links between some pairs of features covering the dependencies among the network traffic attributes.

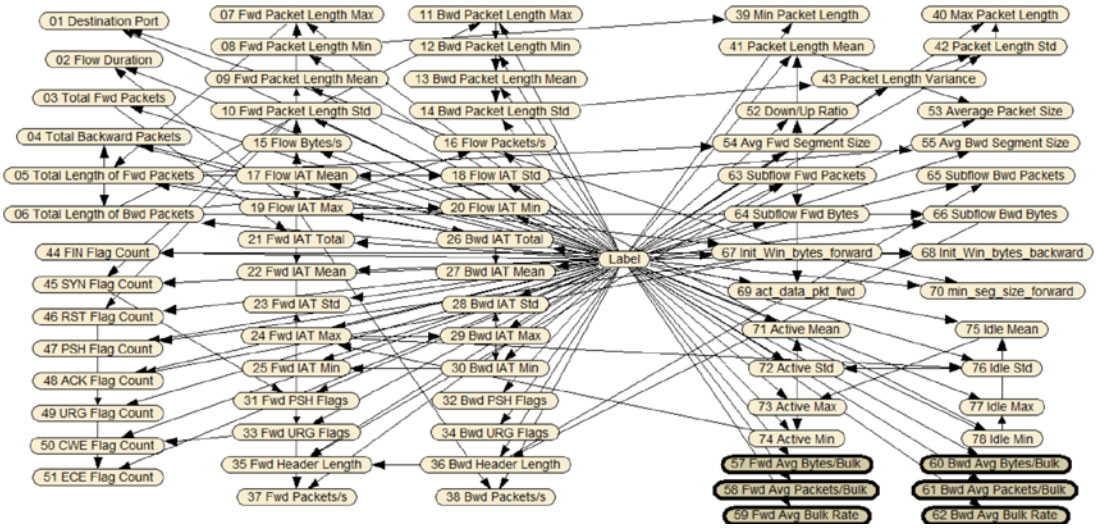

Figure 3. The Tree Augmented Naïve (TAN) Bayesian Classifier modeled with Netica

3.3 Analyzing the results

The in section 2.3 introduced data set has been divided into a training set and a test set. In figure 4 the number of instances in the training and the test set are displayed for each attack type. The training set is used to learn the Bayesian classifiers with probability values and the test set is used to evaluate the quality of the classifiers.

Set	Attack type	Training			Test		
		Total	Benign	Attack	Total	Benign	Attack
1	DDoS	203171	87946	115225	22574	9772	12802
2	PortScan	257820	114783	143037	28647	12754	15893
3	Bot	171929	170160	1769	19104	18907	197
4	-	476926	476926	0	52992	52992	0
5	Infiltration	259741	259709	32	28861	28857	4
6	WebAttack	153329	151367	606	17037	16819	67
	BruteForce			1356			151
7	Patator	401319	388867	12452	44590	43207	1383
8	DoS	623433	396028	227395	69270	44003	25266
	Heartbleed			10			1
0	All	2547668	2045786	501882	283075	227311	55764

Figure 4. Number of instances in training and test sets respectively

The confusion matrix considers a binary classification in two dimensions, *prediction* and *reality*. The 2x2 matrix reports the four variables: *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)* and *False Negative (FN)*. The left matrix in figure 5 shows the results for the naïve Bayesian classifier, the right matrix for the TAN Bayesian classifier. Figure 5 also shows that TAN classifier yields considerably better results.

		Actual	
		Benign	Attack
Predicted	Benign	218026	15736
	Attack	9285	40028

		Actual	
		Benign	Attack
Predicted	Benign	222792	4052
	Attack	4519	51712

Figure 5. Confusion Matrix for the two classifiers

Figure 6 shows the rates for *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)* and *False Negative (FN)*. The blue and red columns show the results for the naïve and TAN classifier, respectively. In column 0 the results for the entire dataset are displayed, in the rest of the columns the results for each for single dataset from figure 1 are reported. These results are more than satisfactory and again show that in almost all cases the TAN Bayesian classifier outperforms the naïve Bayesian classifier.

ISBN: 97819384961-9-6

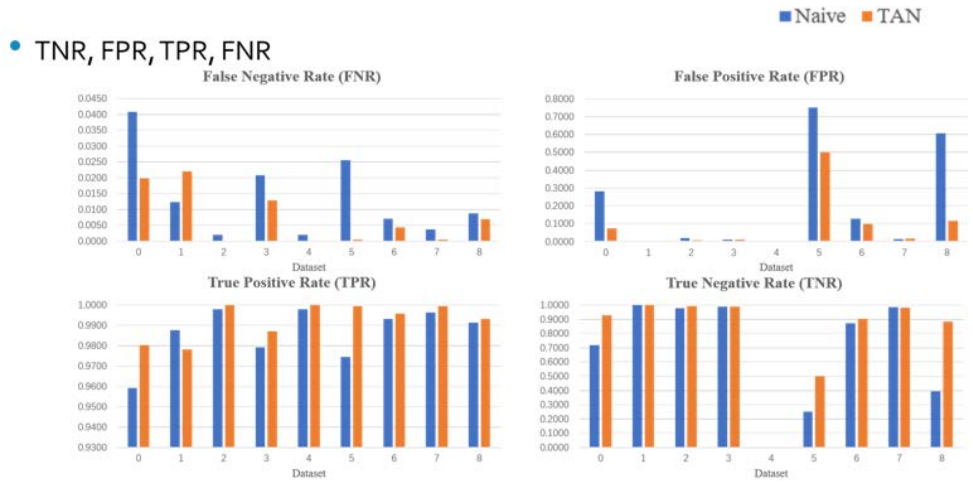

Figure 6. *TP*, *FP*, *TN* and *FN* rates for the different data sets

4. Conclusion

Regarding the network- and flow-based intrusion detection problems, naive Bayesian network classifiers and tree-augmented naive Bayesian network classifiers have been modeled for the purpose of validating and verifying the Multilayer Perceptron (MLP) machine learning model based on the dataset CSE-CIC-IDS2018. The results have proved that the classifiers are capable to satisfactorily validate the decisions of the MLP algorithm. Concerning the performance in terms of several criteria, TAN Bayesian classifiers have outperformed naive Bayesian classifiers in general, while naive classifiers were more cautious in identifying an activity as *Benign*.

5. References

Darwiche, A. (2014). *Modeling and reasoning with Bayesian networks*. Cambridge University Press, New York, USA.

Kjaerulff, U.B., Madsen A.L. (2014). *Bayesian networks and influence diagrams: A guide to construction and analysis*. Springer, New York, USA.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems – Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, California.