EasyChair Preprint

№ 2897

Neural Networks and Decision Trees for Intrusion Detections: Enhancing Detection Accuracy

Rachid Beghdad, Katia Bechar, Meriem Bouali and Haddadi Mohamed

March 8, 2020

Neural Networks and Decision Trees for Intrusion Detections: Enhancing Detection Accuracy

Rachid Beghdad [1], Katia Bechar [1], Meriem Bouali [1], Mohamed Haddadi[1,2]

[1] Département d'informatique, Faculté des sciences exactes, Université de Bejaia, Bejaia 06000, Algérie, rachid.beghdad@gmail.com

[2] Département des sciences commerciales, Faculté des sciences économiques, commerciales, et sciences de gestion,

Université M'hamed Bougara de Boumerdes, Avenue de l'indépendence,Boumerdes 35000, Algérie,

m.haddadi@univ-boumerdes.dz, haddadimohamed2013@gmail.com

Abstract. Many artificial intelligence methods were applied to enhance security in computer networks. These methods seem to be mainly based on neural networks and decision trees. Nevertheless, and according to literature, some of them are still suffering from some weaknesses. This is the reason why we focused in this study on the enhancement of two approaches: Iterative Dichotomiser 3 ID3 and multilayer perceptron (MLP) algorithms. The aim of this study is to use appropriate attributes of the KDD dataset, in order to obtain better detection rates. Simulations were conducted using WEKA and Tanagra tools. The results show that our contributions (ID3 and MLP) are competitive with other solutions in terms of detection accuracy.

Keywords. ID3, Decision Trees, MLP Neural Network, 10%KDD cup'99 Dataset, IDS, Experimental Comparison**.**

1 Introduction

An intrusion is a set of malicious activities to compromise computer security and computers network security [1]. In practice, there are effectively many types of IDS as mentioned in [2]. So, these IDSs are hypervisor IDS, DIDS, HIDS, HyIDS, and NIDS. In the last two decades, the well-known and widely used algorithms in detecting intrusions are decision trees and neural networks. Each of which is used in one or many fields of research, especially the classification problems and attacks detection. Therefore, there are many types for everyone. For neural networks, there are several types such as multilayer perceptron (MLP), generalized feedforward (GFF), radial basis function (RBF), self-organizing feature map (SOFM), and principal component analysis (PCA) NN as illustrated in [3]. For decision trees, there are many types like C4.5, CART, Random Tree, J48, ADTree, NBTree, Random Forest, and Hoeffding Tree as illustrated in [4]. So, these algorithms mainly use KKD cup 99 datasets [5] in order to evaluate the performance of their proposed algorithms.

According to literature, many artificial intelligence methods were applied to enhance security in computer networks. These methods seem to be mainly based on neural networks and decision trees. Nevertheless, and according to literature, some of them are still suffering from some weaknesses. This can be explained by the misuse of the KDD dataset (only 10% are used in general), but also, by using only the KDD attributes. In fact, while looking at these attributes, we can easily notice that some of them are not useful for training artificial approaches based on neural networks or decision trees. This is the reason why we focused in this study on the enhancement of two approaches: Iterative Dichotomiser 3 (ID3) and multilayer perceptron (MLP) algorithms. The aim of this study is to use appropriate attributes (14 well-chosen attributes) of the KDD dataset, in order to obtain better detection rates.

The structure of this paper is organized as follows. Section 2 outlines related works whereas Section 3 presents two suitable classifiers. Section 4, the performance of our contributions is evaluated. Section 5 gives a brief conclusion.

2 Related work

2.1 Decision Trees (DTs)

Decision tree algorithm has been heavily used for many years in several fields of research such as computer network security. It is one of the most important used methods for classification. Several decision tree algorithms have been proposed. One of the most used classification algorithms is ID3 [6, 7]. This well-known algorithm selects split attributes using information entropy. Another algorithm called C4.5 [8]. It is an improved version of ID3 that deals with both continuous and discrete values, and missing values. It also avoids overfitting, improves computing efficiency and performs other functions. Baik et al. [9] applied DDT (distributed decision tree) approach to intrusion detection domain. This approach integrates inductive generalization and agent-based computing. Thakur et al. [10] re-optimized ID3 and C4.5 decision tree algorithms. This proposed method provides a simple modification to the attribute selection methods of them. Lakshminarasimman et al. [11] proposed a new method for DDoS detection. This method is based on Decision. Amor et al. [12] proposed an approach using Naive Bayes (NB) and decision tree to help IDS achieve good detection rate. The experiments were conducted using KDD cup 99 datasets. Results show that the proposal achieves better results in terms of computation time. Rish et al. [13] proposed an approach called Naïve Bayes for classification. This classifier uses Monte Carlo simulations to classify different types of problems. Results show that the approach performs better than other existing approaches. AHMAN et al [14] proposed a new Learning algorithm based on adaptive IDS using DT. Experimental results, conducted using KDD cup 99 datasets, show that the approach achieves a detection rate of 98%. A hybridization is developed by PANDA et al [15], using many algorithms in order to make intelligent decisions. These classifiers are decision tree, PCA, SPegasos, SVM, END, Random Forest and Grading. Results, conducted in a variant of KDD cup 99 datasets called NSL-KDD, show that the proposed approach has 0% of false alarms and 100% of detection rate. Shah et al. [16] proposed a discriminative feature selection

and intrusion classification based on SPLR (sparse logistic regression) for IDS. Experiments conducted using NSL-KDD dataset. Results show that the proposed method has better performance compared to the other well-known techniques. Breiman et al. [17] proposed a classification and regression tree (CART). This classifier is a non- parametric decision tree which can produce classification or regression. Experimental results show that CART has low computation time in comparison to other existing approaches.

2.2 Neural Networks (NNs)

We begin with the proposal of Mukkamal et al [18], who proposed to detect intruders using NN and support vector machines (SVM). Another approach is developed by Cannady et al. [19], who used an ANN for misuse detection and anomy detection based on packet header attributes. Gupta et al [20] studied different data mining classification techniques have been tested using the 10 fold cross-validation method. These techniques like J48, kNN, FT, NB, LMT, SVM, C-RT, QUEST, MLP, ID3, Bayes Net, C4.5, CHAID, LDA, NN-RBFN, Prototype-NN, SPegasos and so on. A model called Hyperview [21] was proposed by Debar et al. It is based on two components: signature-based IDS and NN. Ghosh et al.[22] employed ANNs for misuse and anomaly detection using recent user behaviors. A framework called RBF-SOM [23] was proposed by Horeis et al. It is a combination of a RBF (radial basis function) network and SOM (self-organizing maps) based NN. A framework called NNID (neural network intrusion detector) [24] was proposed by Lin et al. It is a new IDS based on the neural network and backpropagation approach. Subarna et al. [25] proposed a novel IDS to detect threats. The proposed method is based on BPN (back- propagation neural networks) and SOM (and self-organizing map). Dias et al. [26] proposed an IDS based on artificial neural network (ANN) and the KDDCUP'99 dataset. ESMAILY et al [27] hybridized decision tree and multi-layer perceptron neural networks to classify instances. Experimental results, conducted using KDD cup 99 datasets, show that the proposed scheme identified attacks with high accuracy. A framework called FC-ANN [28] is proposed by WANG et al. It is based on ANN (Artificial Neural Networks) and fuzzy clustering. Experimental results, conducted using KDD cup 99 datasets, show that the proposed new approach performs better than other existing approaches in terms of detection precision and detection stability. A hybridization, called GN-ANN and GSPSO-ANN was proposed by Dash [29]. It is based on evolutionary algorithms such as GS (gravitational search) and PSO (particle swarm optimization). Experiments conducted using NSL-KDD dataset. Results show that the proposed approach performs better than other existing approaches.

3 Our contributions

3.1 Decision tree (ID3 algorithm)

To build the decision tree, we used the ID3 algorithm as shown in Figure. 1. It is a supervised classification algorithm that is based on examples already classified in a set

of classes to determine a classification model. It consists of building a tree from its root to its leaves recursively by choosing the attribute that maximizes the information gain at each stage of the construction.

Algorithm: ID3 Algorithm
Inputs: R: a set of non-target attributes,
C: the target attribute,
S: learning data.
Output: returns a decision tree
beginning
Initialize to the empty tree;
If S is empty **then**
Return a simple node of value
End if
If S consists only of identical values for the target **then**
Return a single node of this value
End if
If R is empty **then**
Return a single node with the value of the most frequent value of values of the target attribute found in **S**
End if
D ← the attribute that has the largest Gain (**D**, **S**) among all the attributes of **R**
{dj with j = 1, 2, ..., m} ← The values of the attributes of D
{Sj with j = 1, 2, ..., m} ← The subsets of **S** respectively constituted of dj value records for attribute D
Return a tree whose root is D and the arcs are labeled by d1, d2, .., dm
and going to sub-trees ID3 (R- {D}, C, S1), ID3 (R- {D}, C, S2), .., ID3 (R- {D}, C, Sm)
End

Fig. 1 ID3 algorithm [30]

3.2 Neural network (MLP algorithm)

The Multilayer Perceptron (MLP) is a classifier that uses a supervised learning technique called backpropagation to classify instances as shown in Figure. 2. It can implement arbitrary decision limits using hidden layers. It consists of, at least, three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

Function :BackProp (examples, network) **returns** neural network
Inputs: example, each example consists of an input vector X and a vector Y output network, a multilayer network of with L layer, the weights $w_{i,j}$, and a activation function
Local Variable: error vector
// initialization of weights
For each $w_{i,j}$ **do**
$w_{i,j}$= a is a small number
For each Example (x,y) **do**
For each node i of the input layer **do**

$$a_i \leftarrow x_i$$

// calculate the activation value for each neuron
For ℓ =2 to L **do**
For each node j of the layer ℓ **do**

$$in_j \leftarrow \sum_i w_{i,j} a_i$$

$$a_j \leftarrow g(in_j)$$

// calculate the error vector
For each node j of the output layer **do**

$$e[j] \leftarrow y_j - a_2$$

// update weights
For each $w_{i,j}$ **do**

$$w_{i,j} \leftarrow w_{i,j} + a \times a_i \times e[j]$$

Until verification of a stopping criterion
Returns neural network

Fig. 2 Backpropagation algorithm [31]

4 Simulation evaluations

4.1 Simulation details

TABLE I describes the attributes used in our experiment, extracted from 10% KDD cup99 dataset (494021 instances).

Table 1. Attributes description.

N	Attributes	Description
1	protocol_type	The protocol used in the connection
2	Service	Destination network service used
3	Flag	Connection status (Normal or Error)
4	src_bytes	The number of data bytes transferred from source to destination in a single connection.
5	dst_bytes	Number of bytes of data transferred from the destination to the source in a single connection
6	Land	if the source and destination IP addresses and port numbers are equal, then this variable is set to 1, 0 otherwise
7	wrong_fragment	Total number of fake fragments in this connection
8	logged_in	Connection status: 1 if connected successfully, 0 otherwise
9	root_shell	1 if the root shell is obtained, 0 otherwise
10	Count	Number of connections to the same destination host
11	same_srv_rate	The percentage of connections that were to the same service, among the connections aggregated in Count
12	diff_srv_rate	The percentage of connections that were to different services, among the connections aggregated in Count
13	dst_host_same_ src_port_rate	The percentage of connections that were at the same source port, among the connections aggregated in dst_host_srv_count
14	Label	the label assigned to each of the examples as attack type (1) or as normal (0)

For the ID3 implementation, we selected all attributes. The first 13 attributes are inputs and the last attribute is the element to predict whereas, for the MLP implementation, we selected the attributes 4, 5, 10, 11, 12, 13, and 14 listed in TABLE I. The last attribute is the element to predict.

4.2 Software and hardware configurations

The hardware configuration is 2.53 GHz Intel® Core i5 with 4 GB memory and 118 GB hard drive. The exploiting system is Ubuntu 14.04 LTS 64 bit. So, we used Weka (Waikato Environment for Knowledge Analysis). It is a set of open- source tools for manipulating and analyzing data files [32].

4.3 Simulation results and analysis

In the following, we simulate the results of our contributions (ID3 and MLP) using 10%KDD cup99 dataset which contains 494021 instances.

- ***ID3 algorithm.*** Table 2 illustrates the results obtained using 10% KDD cup 99 datasets and WEKA tool.

Table 2. Evaluation results using ID3.

Algorithm	Correctly classified instances	Incorrectly Classified instances
Our contribution(ID3)	99.8842 %	0.1158 %

In the following, we will compare our contribution (ID3) with other existing algorithms like J48, Random Forest, Decision Tree, and Naïve Bayes as listed in Table 3. In this study, the 10% KDD cup 99 has been conducted to compute the detection accuracy.

TABLE 3. COMPARISON BETWEEN OUR CONTRIBUTION (ID3) USING ONLY 14 ATTRIBUTES AND OTHERS [11,12] USING 42 ATTRIBUTES

Algorithms	Correctly classified instances	Incorrectly Classified instances
J48 [11]	99.9415 %	0.0585 %
Random Forest [11]	96.9437 %	3.0563 %
Decision tree [12]	99.99	0.01
Naïve Bayes [12]	99.23	0.77
Our contribution (ID3)	99.8842% (only 14 attributes)	0.1158%

According to the previous table, the result of our contribution (ID3) is competitive with the work already mentioned on decision trees, especially that of J48 [11], which is very close to the result of our contribution (ID3) and whose difference is 0.1058%. The Naïve Bayes algorithm has the lowest detection accuracy of 99.23. In addition to that, note that in our experiments, we considered only 14 attributes described above in section A.

In addition to that comparison, we add supply comparison between our contribution (ID3) and other existing algorithms of top 10 algorithms in data mining [33], such as C4.5, CART, Naïve Bayes, Random tree and Bagging + CART using a well-known

tool of data mining called TANAGRA tool [34] as shown in Table 4. This comparison is done under similar attributes and values using 10% KKD Cup 99 dataset.

TABLE 4. COMPARISON BETWEEN OUR CONTRIBUTION (ID3) AND OTHER ALGORITHMS USING TANAGRA IN CASE OF 14 ATTRIBUTES

Algorithms	Correctly classified instances	Incorrectly classified instances
CART	99.7583	0.2417
Bagging + CART	99.77	0.23
C4.5	99.7749	0.2251
Naïve Bayes	96.7505	3.2495
Random tree	99.48	0.52
Our contribution (ID3)	99.8842%	0.1158%

MLP Algorithm. Table 5 illustrates the results obtained with a different number of neurons (9, 11, 13 and 14) in the hidden layer using 10% KDD cup 99 and Weka tool.

TABLE 5. COMPARISON BETWEEN OUR CONTRIBUTION (MLP) USING ONLY 14 ATTRIBUTES AND OTHERS [25, 26, 27, 28]

Algorithms	Correctly classified instances	Correctly classified instances
ANN [26]	99.9%	0.1%
BP-SOM [25]	99.95%	0.05%
FC-ANN [28]	96.71%	3.29%
BPNN [28]	96.65%	3.35%
Our contribution (MLP)	99.10%	0.9%

The result of our contribution (MLP) is also competitive with the work already cited on neural networks, especially BP-SOM [25].

ID3 and MLP Algorithm. In the following, we compare our contributions (ID3 and MLP algorithms) with other ones as listed in Table 6 using 10% KDD cup 99 datasets.

TABLE 6. COMPARISON BETWEEN ID3 AND MLP AND OTHERS

Algorithms	Correctly classified instances	Incorrectly classified instances
CART	99.7583%	0.2417%
Bagging + CART	99.77%	0.23%
C4.5	99.7749%	0.2251%
Naïve Bayes	96.7505%	3.2495%
Random tree	99.48%	0.52%
KNN	19.6911%	80.3089%
PNN	79.41%	20.59%

Our contribution (MLP)	99.1%	0.90%
Our contribution (ID3)	**99.8842%**	**0.1158%**

According to previous table, the result of our contributions (ID3 and MLP algorithms) are competitive with other existing algorithms of top 10 algorithms in data mining [33], such as C4.5, CART, Naïve Bayes, Random tree and Bagging + CART, KNN, and PNN, using a well-known tool of data mining called TANAGRA [34]. So, our contribution (ID3) has the highest detection accuracy compared to other ones of the same category such as C4.5 whereas our contribution (MLP) has also highest detection accuracy compared to other ones of the same category like PNN.

5 Conclusion

In this work, we enhanced two well-known algorithms, ID3, and MLP that are used for intrusion detections. The enhancement is based on the use of 14 appropriate attributes of KDD dataset for training and testing phases of the two presented solutions because some attributes are not useful for training artificial approaches. For implementing both algorithms (ID3 and MLP), we have used the Java language and WEKA tool. The simulation results show that our contribution (ID3) is competitive with some algorithms proposed in [11, 12], whereas for others such as C4.5, CART, Naïve Bayes, Random tree and Bagging + CART, it has the highest detection accuracy of 99.88 %. For MLP, it seems clearly to be competitive with a few algorithms proposed in [25, 26, 28], whereas for others like PNN and KNN, it seems clearly to have the highest detection accuracy of 99.10%.

References

1. J. M. Kizza. *Computer network security*. Springer Science & Business Media, 2005.
2. M. Haddadi, and R. Beghdad, "DoS-DDoS: Taxonomies Of Attacks, Countermeasures, And Well-Known Defense Mechanisms In Cloud Environment," *EDPACS*, *57*(5), pp. 1-26, 2018.
3. R. Beghdad, "Critical study of neural networks in detecting intrusions." *Computers & security*, *27*(5-6), pp. 168-175, 2008.
4. S. Aljawarneh, and M. B. Yassein, and M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems," *Cluster Computing*, pp. 1-17, 2017.
5. KDD Cup 1999 dada. Available http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html [Sep. 30, 2018].
6. J. R. Quinlan, "Learning efficient classification procedures and their application to chess end games," In *Machine learning*, Springer, Berlin, Heidelberg, pp. 463-482, 1983.
7. J. R. Quinlan, "Induction of decision trees," *Machine learning*, *1*(1), pp. 81-106, 1986.

8. S. L. Salzberg, "C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993." *Machine Learning*, *16*(3), pp. 235-240, 1994.

9. S. HBaik, and J. Bala, " A decision tree algorithm for distributed data mining: Towards network intrusion detection," In *International Conference on Computational Science and Its Applications*, Springer, Berlin, Heidelberg, May 2004, pp. 206-212.

10. D. Thakur,and N. Markandaiah, and D. S. Raj, "Re optimization of ID3 and C4. 5 decision tree," In *Computer and Communication Technology (ICCCT), 2010 International Conference on*. IEEE, Sep 2010, pp. 448-450.

11. S. Lakshminarasimman, and S. Ruswin, and K. Sundarakantham, " Detecting DDoS attacks using decision tree algorithm," In *Signal Processing, Communication and Networking (ICSCN), 2017 Fourth International Conference on*, IEEE, Mar 2017, pp. 1-6.

12. N. B. Amor, and S. Benferhat, and Z. Elouedi, " Naive bayes vs decision trees in intrusion detection systems," In *Proceedings of the 2004 ACM symposium on Applied computing*, ACM, Mar 2004, pp. 420-424.

13. I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22. New York: IBM, Aug 2001.

14. C. M. Rahman, and D. M. Farid, and N. Harbi, and E. Bahri, M. Z. Rahman, " Attacks classification in adaptive intrusion detection using decision tree," 2010.

15. M. Panda, and A. Abraham, and M. R. Patra, "A hybrid intelligent approach for network intrusion detection,". *Procedia Engineering*, *30*, pp. 1-9, 2012.

16. R. A. Shah, and Y. Qian, and D. Kumar, and M. Ali, and M. B. Alvi, " Network Intrusion Detection through Discriminative Feature Selection by Using Sparse Logistic Regression," *Future Internet*, *9*(4), 81, 2017.

17. L. Breiman, and J. H. Friedman, and R. A. Olshen, and C. J. Stone, "Classification and regression trees. Belmont, CA: Wadsworth," *International Group*, *432*, pp. 151-166, 1984.

18. S. Mukkamala, and G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," In *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, IEEE, Vol. 2, 2002, pp. 1702-1707.

19. J. Cannady, "Artificial neural networks for misuse detection," In *National information systems security conference*, Vol. 26, Oct 1998.

20. S. Gupta, and D. Kumar, and A. Sharma, "Performance analysis of various data mining classification techniques on healthcare data," *International journal of computer science & Information Technology (IJCSIT)*, *3*(4), 2011, pp. 155-169.

21. H. Debar, and M. Becker, D. Siboni, " A neural network component for an intrusion detection system," In *IEEE symposium on security and privacy*, pp. 240-250, May 1992.

22. A. K. Ghosh, and A. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection," In *USENIX security symposium* ,Vol. 99, pp. 12, Aug 1999.

23. T. Horeis, "Intrusion detection with neural networks–combination of self-organizing maps and radial basis function networks for human expert integration," *Computational Intelligence Society Student Research Grants*, 2003.

24. J. Ryan, and M. J. Lin, and R. Miikkulainen, "Intrusion detection with neural networks," In *Advances in neural information processing systems* , pp. 943-949, 1998.

25. S. Shakya, and B. R. Kaphle, " Intrusion detection system using back propagation algorithm and compare its performance with self organizing map," *Journal of Advanced College of Engineering and Management*, *1*, pp. 127-138, 2016.

26. L. P. Dias, and J. J. F. Cerqueira, and K. D. R. Assis, R. C. Almeida, "Using artificial neural network in intrusion detection systems to computer networks," In *Computer Science and Electronic Engineering (CEEC), IEEE,* pp. 145-150), Sep 2017.

27. J. Esmaily, and R. Moradinezhad, and J. Ghasemi, "Intrusion detection system based on multi-layer perceptron neural networks and decision tree," In *2015 7th Conference on Information and Knowledge Technology (IKT)*,IEEE, May 2015, pp. 1-5.

28. G. Wang, and J. Hao, and J. Ma, L. Huang, " A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering,". *Expert systems with applications*, *37*(9), pp. 6225-6232, 2010.

29. T. Dash, " A study on intrusion detection using neural networks trained with evolutionary algorithms," *Soft Computing*, *21*(10), pp. 2687-2700, 2017.

30. H. Ezzikouri, and M. Fakir, "ID3 & C4.5 classification algorithms,". Available: http://www.academia.edu/33701469/Algorithmes_de_classification_ID3_and_C4.5. University of Sultan Moulay Slimane, Faculty of Sciences and Techniques, (year?). (in French)

31. Youtube, " L-layer neural network,". Available : https://www.youtube.com/watch?v=NgwvhX0xBs0&index=81&list=PL6Xpj9I5qXYGhsvMWM53ZLfwUInzvYWsm&t=0s. (in French)

32. I. H. Witten, and E. Frank, and M. A. E. Hall, and C. J. Pal, " *Data Mining," Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

33. X. Wu, and V. Kumar, and J. R. Quinlan, J. Ghosh, and Q. Yang, and H . Motoda, .and Z. H. Zhou,"Top 10 algorithms in data mining, "*Knowledge and information systems*, *14*(1), pp. 1-37, 2008

34. Tanagra. Available:

35. http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html.