ENHANCING NETWORK INTRUSION CLASSIfiCATION THROUGH THE KOLMOGOROV-SMIRNOV SPLITTING CRITERION

THANH-NGHI DO, PHILIPPE LENCA, AND STÉPHANE LALLICH

ABSTRACT

Our investigation aims at detecting network intrusions using decision tree algorithms. Large differences in prior class probabilities of intrusion data have been reported to hinder the performance of decision trees. We propose to replace the Shannon entropy used in tree induction algorithms with a Kolmogorov Smirnov splitting criterion which locates a Bayes optimal cutpoint of attributes. The Kolmogorov-Smirnov distance based on the cumulative distributions is not degraded by class imbalance. Numerical test results on the KDDCup99 dataset showed that our proposals are attractive to network intrusion detection tasks. The single decision tree gives best results for minority classes, cost metric and global accuracy compared with the bagged boosting of trees of the KDDCup'99 winner and classical decision tree algorithms using the Shannon entropy. In contrast to the complex model of KDDCup winner, our decision tree represents inductive rules (IF-THEN) that facilitate human interpretation.

1. INTRODUCTION

Nowadays the increasing pervasiveness of communication between computer networks and the development of the internet transform the way people live, work and play. In addition, the number of intrusions into computer systems is also growing. Therefore, security of computer networks plays a strategic role in modern computer systems. Many rule-based systems use their rule sets to detect network intrusions. Unfortunately, due to the huge volume of network traffic, coding the rules by security experts becomes difficult and time-consuming. Since machine learning techniques can build intrusion detection models adaptively, this kind of network intrusion detection has significant advantages over rule-based ones. Over the last several years, a growing number of research have applied machine learning techniques to intrusion detection.

Our contribution aims at enhancing attacks detection tasks with decision tree algorithms. Due to the class imbalance problem of network intrusion data, we propose to replace Shannon entropy [1] used in tree induction algorithms with a Kolmogorov- Smirnov criteria which locates a Bayes optimal cutpoint of attributes. The Kolmogorov-Smirnov distance based on the cumulative distributions is not degraded by class imbalance. The proposal can improve minority class prediction. The numerical test results on the KDDCup99 dataset [2] showed that our proposals are suitable for network intrusion detection tasks. The single decision tree gives better results for minority classes, cost matrix and global accuracy versus the complex model, bagged boosting of trees of the KDDCup'99 winner and classical decision tree algorithms using the Shannon entropy. In addition, our tree model is more readily interpretable than the complex model of the KDDCup'99 winner.

The remainder of this paper is organized as follows. Section 2 presents related works in the network intrusion domain. Section 3 briefly introduces decision tree using the Kolmogorov-Smirnov distance for classification. Section 4 presents numerical test results before the conclusion.

2. RELATED WORKS

Under the sponsorship of Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL), MIT Lincoln Laboratory collected the dataset for the evaluation of computer network intrusion detection systems [3]. DARPA dataset is the most popular dataset used to test and evaluate a large number of intrusion dection systems. The KDDCup99 dataset [2] is a subset of DARPA dataset prepared by Sal Stofo and Wenke Lee [4]. The data were preprocessed by extracting 41 features from the tcpdump data in the 1998 DARPA dataset. The DDCup99 dataset can be used without further time-consuming preprocessing and different intrusion detection systems can compare with each other by working on the same dataset. Therefore, researchers have carried out their experiments on the KDDCup99 dataset.

In the report [2] of KDDCup99 contest, the winning entry [5] used a mixture of bagging and boosting of decision trees [6 - 8]. The standard sampling with replacement methodology of bagging was modified to put a specific focus on the smaller but expensive-if-predicted- wrongly classes. Second-place performance was achieved by Levin [9] from LLSoft, Inc. using the tool Kernel Miner. Third-place performance was achieved by Miheev et al. [10] of the company MP13, using a version of the Fragment algorithm originally invented at the IITP (Russian Academy of Science). For constructing a decision tree, the training sample was used to build the structure of a tree (with sufficient complexity). The testing sample was used to select a sub-tree having optimal complexity. Elkan [2] showed that only nine entries scored better than 1-nearest neighbor, of which only six were statistically significantly better. Compared to 1-nearest neighbor, the main achievement of the winning entry is to recognize correctly many “remote-to-local” attacks.

Ben-Amor et al. [11] studied intrusion detection using naive Bayes and decision trees. The experimental results showed that naive Bayes, with their simple structure and despite their strong assumptions, can be very competitive and the performance difference with respect to decision trees is not significant.

Stein et al. [12] used a genetic algorithm to select a subset of input features for decision tree classifiers, with the goal of increasing the detection rate and decreasing the false alarm rate in network intrusion detection. They reported the results on the KDDCUP 99 dataset. The experiments illustrated that the resulting decision trees can have better performance than those built with whole features.

Zhang and Zulkernine [13] applied random forests algorithm [14] for network intrusion detection. They tried to make a balanced training set using down-sampling the majority classes and over-sampling the minority ones. The results showed that the proposed approach provides better performance compared to the best results from the KDDCup99 contest.

Giacinto et al. [15] combined multiple one-class classifiers using K-means algorithm [16]. Each one-class classifier is trained in order to discriminate between a specific attack and all other traffic patterns. Their method outperforms the KDDCup 1999 winner in terms of the false negatives rate and of the new attacks detection rate although the proposal performs worse in

terms of the percentage of false positives and of the overall cost. Perdisci et al. [17] also deal with attack detection tasks using an ensemble of one-class support vector classifiers [18].

Bouzida and Cuppens [19] proposed to modify the decision tree algorithm C4.5 [8] for discovering known and unknown attacks. In the KDDCup99 dataset, the different attacks presenting in the testing set but not being in the training set cannot be easily classified into their appropriate class and will be classified in the class that has a form close to theirs, generally to the normal class. Due to this problem, they introduced the following principle: A default class denoted new class is assigned to any new class that does not have a corresponding class in the training set. Therefore, if any new instance does not match any of the rules generated by the decision tree then this instance is classified as a new class instead of assigning it to the default class. However, their experiment setup (called the learning data set coherence) was inverted because they used the testing set for training and reported the results on the subset (10%) of the training data.

Bouzida and Cuppens [20] applied both neural networks and decision trees into the notion of intrusion detection. The results showed that while neural networks are highly successful in detecting known attacks, decision trees are more interesting to detect new attacks.

Xiao et al. [21] aimed at building an ensemble of support vector machines [22] to predict network intrusions. Experimental results illustrated the applicability of the approach for this kind of problems.

Engen et al. [23] proposed an evolutionary neural network, in which several evaluation functions are examined. However, when employing evaluation functions that calculate the fitness proportionally to the instances of each class, thereby avoiding the bias towards the majority classes in the data set, significantly improving true positive rates are obtained whilst maintaining a low false positive rate.

Although there are many researches for this problem over the past several years, almost existing approaches can not achieve the best results obtained by the KDDCup99 winner.

3. DECISION TREE USING THE KOLMOGOROV-SMIRNOV DISTANCE

Our investigation aims at detecting network intrusions using decision tree algorithms. The proposal is to build a single decision tree that gives best results for minority classes, cost matrix and global accuracy versus the complex model, bagged boosting of trees of the KDDCup'99 winner. We propose to replace the Shannon entropy [1] used in tree induction algorithms by the Kolmogorov-Smirnov distance. No algorithmic changes are required from the classical decision tree algorithm C4.5 [8] other than the modification of the split function (using the Kolmogorov-Smirnov metric instead of the Shannon entropy). The rest of the original decision tree methods are kept.

We first recall basic considerations on Shannon's entropy and then present briefly the Kolmogorov-Smirnov distance in the boolean case and mention the results in the general case.

3.1. Usual measures based on Shannon's entropy

Many induction tree algorithms on categorical variables use predictive association measures based on the entropy of Shannon. Let us consider a class variable Y having q modalities, $p = (p_1, \ldots, p_q)$ being the vector of frequencies of Y, and a categorial predictor (attribute) X having k modalities. The joint relative frequency of the couple (x_i, y_j) is denoted

p_{ij}, $i = 1,\ldots k$; $j = 1,\ldots q$. What is more, we denote by $h(Y) = -\sum_{j=1}^{q} p_{.j}\log_2 p_{.j}$ the a priori Shannon's entropy of Y and by $h(Y/X) = E(h(Y/X = x_i))$ the conditional expectation of the entropy of Y with respect to X.

Shannon's entropy $h(p)$, is a real positive function of $p = (p_1, \ldots, p_q)$ to $[0..1]$, verifying notably interesting properties for machine learning purposes [24]:

- $h(p)$ is invariant by permutation of the modalities of Y;

- $h(p)$ reaches its maximum $\log_2(q)$ log2(q) when the distribution of Y is uniform (each modality of Y has a frequency of $1/q$);

- $h(p)$ reaches its minimum 0 when the distribution of Y is sure (centered on one modality of Y and the others modalities being of null frequency);

- $h(p)$ is a strictly concave function.

The behavior of Shannon's entropy is illustrated in Fig. 1 in the boolean case. As example of measures based on Shannon's entropy, one can mention:

- The entropic gain $h(Y) - h(Y/X)$ [25];

- The gain-ratio $\frac{h(Y)-h(Y/X)}{h(X)}$ [26] which relates the entropic gain of X to the entropy of X, rather than to the a priori entropy of Y in order to discard the predictors having many modalities.

For more measures and details one can refer to Wehenkel [27] and Loh and Shih [28].

The particularity of these coefficients is that Shannon's entropy of a distribution reaches its maximum when this distribution is uniform. That is to say that the reference value corresponds to the uniform distribution of classes. This characteristic could be a major problem especially in case of highly imbalanced classes, or when the classification costs differ largely. It would seem more logical to evaluate $h(Y)$ and $h(Y/X = x_i)$ used in the above measures on a scale for which the reference value is centered on the independence situation i.e. on the a priori distribution of classes.

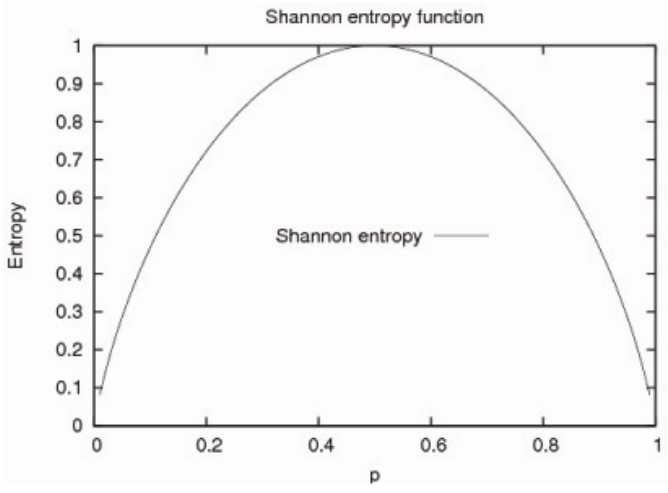

Figure 1. Shannon entropy function

3.2. Kolmogorov-Smirnov distance

The Kolmogorov-Smirnov splitting criterion has been used by Friedman [29] for a binary partition in decision rule algorithms. The Kolmogorov-Smirnov distance is to measure the separability of two distribution functions. It naturally allows separating a population into two homogeneous groups.

Let us consider the case of a class variable Y made of $q = 2$ modalities (positive and negative classes). Two probability density functions on a continuous predictor X for two classes are denoted by $f_{pos}(X)$ and $f_{neg}(X)$, respectively, as shown in figure 2.

Then an optimal cutpoint α (α = 8) is to minimise the Bayes risk of misclassification for positive and negative classes. This is accomplished through the greatest distance between the two cumulative distribution functions (denoted by $cdf_{pos}(X)$ and $f_{neg}(X)$, as shown in figure 3) that correspond to by $f_{pos}(X)$ and $f_{neg}(X)$. That maximum distance is the well-known Kolmogorov-Smirnov distance. However, these cumulative distribution functions are not known in practice, but we can consider approximations (empirical functions, denoted by $\widehat{cdf}_{pos}(X)$ and $\widehat{cdf}_{neg}(X)$)

With a continuous predictor X, a cutpoint α separates a population into two homogeneous groups as following:

1. $X \leq \alpha$ (left partition)
2. $X > \alpha$ (right partition)

The distance $Dist(X = \alpha)$ between two empirical cumulative distribution functions $\widehat{cdf}_{pos}(X \leq \alpha)$ and $\widehat{cdf}_{neg}(X \leq \alpha)$ is:

$$Dist(X = \alpha) = |\widehat{cdf}_{pos}(X \leq \alpha) - \widehat{cdf}_{neg}(X \leq \alpha)|$$

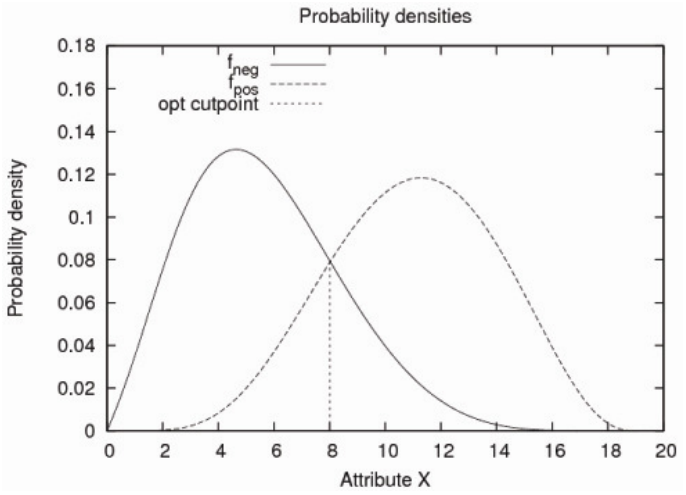

Figure 2. Probability density functions

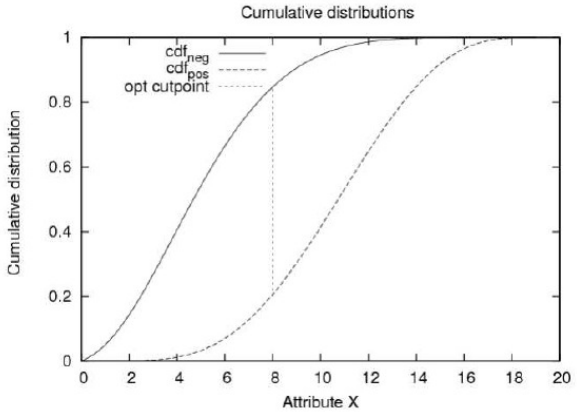

Figure 3. Cumulative distribution functions

For a discrete predictor X having the modalities $\{v, \ldots, \omega\}$, a cutpoint $\alpha = v$ splits a population into two homogeneous groups as following:

1. X = *v* (left partition);
2. X ≠ *v* (right partition).

The distance Dist(X = *v*) between two empirical functions $\widehat{cdf}_{pos}(X = v)$ and $\widehat{cdf}_{neg}(X = v)$ is:

$$Dist(X = v) = |\widehat{cdf}_{pos}(X = v) - \widehat{cdf}_{neg}(X = v)|.$$

Thus the Kolmogorov-Smirnov distance $Dist_{KS}(x*)$ between two empirical functions $\widehat{cdf}_{pos}(X)$ and $\widehat{cdf}_{neg}(X)$ is:

$$Dist_{KS}(x*) = \max_x Dist(X).$$

Let us consider an example of class imbalance using the Kolmogorov-Smirnov distance. Assume further that the data distribution on an attribute *X* with a minority class (positive) and a majority class (negative) is following table 1. Firstly, two empirical cumulative distributions of classes are calculated as shown in table 2.

Table 1. Example of class imbalance

Attribute X	1	2	3	4	5	6	7	8	9	10
#ind. of posotive (minority)	2	2	2	2	1	1	0	0	0	0
#ind. of negative (majority)	0	0	0	10	0	20	40	10	10	10

The optimal cutpoint (α = 5.5) found by the Kolmogorov-Smirnov splitting criterion (the maximum distance between the two empirical cumulative distribution functions) compromises between the minority class and the majority one[1]. Then the Kolmogorov-Smirnov splitting

[1] While the splitting criterion using the Shannon entropy cuts at α = 3.5 without respect to the minority class loss.

criterion based on the distance between two empirical cumulative distribution functions does not lose the minority class prediction in class imbalance.

Table 2. Empirical cumulative distribution

Attribute X	1	2	3	4	5	6	7	8	9	10
$\widetilde{cdf}_{pos}$(minority)	0.2	0.4	0.6	0.8	0.9	1	1	1	1	1
$\widetilde{cdf}_{neg}$(majority)	0	0	0	0.1	0.1	0.3	0.7	0.8	0.9	1

The Kolmogorov-Smirnov splitting criterion can be extended to handle more than two classes (a class variable Y having more than 2 modalities, i.e. q > 2). Friedman [29] proposed to build one tree for each class. An alternative approach is to construct a single tree with the two cumulative distribution functions for two super classes grouped by the q prior classes. Due to the problem of computational cost, we also propose to build a single tree using the method one-against-all during the splitting process. At each node having q classes, this method finds q Kolmogorov-Smirnov distances where the ith distance separates the ith class from the rest and finally it picks the greatest distance.

Table 3. Partitions obtained by the Kolmogorov-Smirnov splitting criterion

Cutpoint (α = 5.5) on attribute X	Left partition	Right partition
#ind. of positive (minority)	9	1
#ind. of negative (majority)	10	90

4. NUMERICAL TEST RESULTS

In order to evaluate the effectiveness of our proposal, we add the Kolmogorov-Smirnov distance (denoted by KS) to the free source code of decision tree algorithm C4.5 [8]. No algorithmic changes are required from the classical decision tree. All the benefits of the original decision tree methods are kept.

In this section, we summarize our experimental results for intrusion detection over the KDDCup99 datasets [2]. We first describe the experiment setup and then the classification results of the decision tree algorithm C4.5 using the Shannon entropy, the Kolmogorov-Smirnov splitting criterion. We are also interested in comparing the performance of our proposal to that of the winning entry of the KDDCup99 contest.

4.1. Experiment setup

The experimental setup uses the KDDCup99 datasets with 41 features including duration, protocol type, service, num failed logins, etc. The full training set has 4898431 connections. The 10% training set has 494021 connections. The attacks in the dataset fall into four categories: DoS (Denial of Service), R2L (unauthorized access from a remote machine), U2R (unauthorized access to root privileges), and probing. The 10% training set contains all the minority classes

(Probe, U2R and R2L) of the full training set and part of the majority classes. It is down-sampling the majority classes (Normal, DoS). Therefore, we just use the 10% training dataset to build tree models in our experiments. The task of the KDDCup99 contest was to build classifier capable of distinguishing among four kinds of intrusions and normal traffic numbered as one of five classes (c.f. table 4). The testing set with 311029 connections is used to evaluate classification models.

We start with the pre-processing step. Individual attack types are placed in the five classes using a categorization awk script [2]. We try to reduce the bias of class distribution in training set, we under-sample the Normal and DoS classes by randomly selecting 10% of connections belonging to these classes from the original dataset. We also over-sample Probe, U2R and R2L by replicating their connections. The balanced training set with 69166 connections is much smaller than the original one.

Table 4. Numbering of the attack categories

No	Class	Training set	Testing set
0	Normal	97278	60593
1	Probe	4107	4166
2	DoS	391458	229853
3	U2R	52	228
4	R2L	1126	16189

4.2. Results

A tree model with 175 nodes obtained by orginal algorithm C4.5 (using the Shannon entropy, denoted by SE) gives the classification results in table 5.

Table 5. Classification results of the decision tree C4.5 using SE

Predicted as ⇒	0 (Normal)	1 (Probe)	2 (DoS)	3 (U2R)	4 (R2L)
0 (Normal)	60111	286	81	17	98
1 (Probe)	48	3935	182	1	0
2 (DoS)	6161	239	223453	0	0
3 (U2R)	63	133	0	19	13
4 (R2L)	14945	558	3	27	656

The decision tree algorithm C4.5 using the Kolmogorov-Smirnov distance produces a single tree which has 211 nodes and achieves the prediction results in table 6.

Table 6. Classification results of the decision tree C4.5 using KS

Predicted as ⇒	0 (Normal)	1 (Probe)	2 (DoS)	3 (U2R)	4 (R2L)
0 (Normal)	59467	257	745	26	98
1 (Probe)	47	3939	132	0	48
2 (DoS)	6334	528	222980	0	11
3 (U2R)	50	133	0	18	27
4 (R2L)	12812	76	3	28	3270

In order to the comparative study we turn back the KDDCup99 winning entry in table 7. We use the cost matrix as table 8 published in KDDCup99 to evaluate the performance of our decision tree. M_{ij} denotes the number of individuals in class i classified as class j, and C_{ij} indicates the corresponding cost in the cost matrix. Let N be the total number of the individuals. The cost metric that indicates the average damage of misclassification for each connection is computed as:

$$Cost = \frac{1}{N}\sum[M_{ij}C_{ij}].$$

Table 7. Classification results of the KDDCup99 winning entry

Predicted as ⇒	0 (Normal)	1 (Probe)	2 (DoS)	3 (U2R)	4 (R2L)
0 (Normal)	60262	243	78	4	6
1 (Probe)	511	3471	184	0	0
2 (DoS)	5299	1328	223226	0	0
3 (U2R)	168	20	0	30	10
4 (R2L)	14527	294	0	8	1360

Table 8. Cost matrix

Predicted as ⇒	0 (Normal)	1 (Probe)	2 (DoS)	3 (U2R)	4 (R2L)
0 (Normal)	0	1	2	2	2
1 (Probe)	1	0	2	2	2
2 (DoS)	2	1	0	2	2
3 (U2R)	3	2	2	0	2
4 (R2L)	4	2	2	2	0

According to the cost metric, the decision tree using the Kolmogorov-Smirnov splitting criterion achieves 0.2172 in terms of the cost that is smaller than the winning entry of KDDCup99 (cost = 0.2331) and the decision using the Shannon entropy (cost = 0.2414).

Table 9. Comparison of results obtained by C4.5 using SE, KS and Bagged boosting tree of the KDDCup99 winning entry

	C4.5-SE (%)	C4.5-KS (%)	KDDCup99 winning (%)
0:Normal	99.20	98.14	**99.45**
1:Probe	94.46	**94.55**	83.32
2:DoS	**97.22**	97.01	97.12
3:U2R	8.33	7.9	**13.16**
4:R2L	4.05	**20.20**	8.40
Overall	92.65	**93.13**	92.71
Cost	0.2414	**0.2172**	0.2331

For details of results, the tree model using the Kolmogorov-Smirnov distance gives 93.13% in terms of the overall accuracy against 92.65% of the Shannon entropy and 92.71% of the winning entry. Table 9 shows that our decision tree based on the KolmogorovSmirnov splitting criterion outperforms the winner's bagged boosting of trees and the classical C4.5 with the Shannon entropy for predicting two minority classes (1:Probe and 4:R2L). The tree using the Kolmogorov-Smirnov significantly improves the minority class prediction without penalizing too much the majority class accuracy. Furthermore, our results are obtained by one single decision tree which is simpler compared with the bagged boosting of trees of the KDDCup'99 winner. It allows extracting inductive rules (IF-THEN) that facilitate human interpretation.

5. CONCLUSION

We present a decision tree algorithm using the Kolmogorov-Smirnov distance for detecting network intrusions. In order to deal with imbalanced classes of intrusion data, we propose to use the Kolmogorov-Smirnov distance for learning induction trees instead of the Shannon entropy. A Bayes optimal cutpoint of attributes found by a Kolmogorov-Smirnov splitting criterion based on the cumulative distribution is not degraded by class imbalance. Numerical test results on the KDDCup99 dataset showed that our proposals improve network intrusion detection tasks. Our single decision tree is simple and gives better results for minority classes, cost metric and global accuracy in comparison to the bagged boosting of trees of the KDDCup'99 winner and classical decision tree algorithms using the Shannon entropy. Futhermore, the attractiveness of one tree is due to the fact that, in contrast to ensemble-based methods, a decision tree represents inductive rules (IF-THEN) that facilitate human interpretation.

We intend to provide more empirical test on a large benchmark of imbalanced datasets in the near future. Comparisons with other split functions should also be done.

REFERENCES

1. Shannon C. E. - A mathematical theory of communication. Bell System Technological Journal **27** (1948) 379-423, 623-656.
2. Elkan C. - Results of the kdd'99 classifier learning. SIGKDD Explorations 1(2) (2000) 63-64.
3. Laboratory M. L. - Darpa intrusion detection evaluation http://www.ll.mit.edu/IST/ideval.
4. Lee W. - A data mining framework for constructing features and models for intrusion detection systems, 1999.
5. Pfahringer B. - Winning the kdd99 classification cup: Bagged boosting, SIGKDD Explorations **1** (2) (2000) 65-66.
6. Breiman L. - Bagging predictors, Machine Learning **24** (2) (1996) 123-140.
7. Freund Y., Schapire R. - A decision-theoretic generalization of on-line learning and an application to boosting, In: Computational Learning Theory: Proceedings of the Second European Conference, 1995, pp. 23-37.
8. Quinlan J. R. - C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993.
9. Levin I. - Kdd-99 classifier learning contest llsoft's results overview, SIGKDD Explorations **1** (2) (2000) 67-75.
10. Miheev V., Vopilov A., Shabalin I. - The mp13 approach to the kdd'99 classifer learning contest. SIGKDD Explorations **1** (2) (2000) 76-77.
11. Ben-Amor N., Benferhat S., Elouedi Z. - Naive bayes vs decision trees in intrusion detection systems, In: ACM Symposium on Applied Computing, 2004, pp. 420-424.
12. Stein G., Chen B., Wu A., Hua K. - Decision tree classifier for network intrusion detection with ga-based feature selection, In: 43rd ACM Southeast Conference, 2005, pp. 136-141
13. Zhang J., Zulkernine M. - Network intrusion detection using random forests, In: Third Annual Conference on Privacy, Security and Trust, 2005.
14. Breiman L. - Random forests, Machine Learning **45** (1) (2001) 5-32.
15. Giacinto G., Perdisci R., Roli F. - Network intrusion detection by combining one-class classifiers. In: Image Analysis and Processing, 2005, pp. 58-65.
16. MacQueen J. - Some methods for classification and analysis of multivariate observations, Berkeley Symposium on Mathematical Statistics and Probability, University of California Press (1) (1967) 281-297.
17. Perdisci R., Gu G., Lee W. - Using an ensemble of one-class svm classifiers to harden payload-based anomaly detection systems, In: the Sixth International Conference on Data Mining, 2006, pp. 488-498.
18. Scholkopf B., Platt J., Shawe-Taylor J., Smola A., Williamson R. - Estimating the support of a high-dimensional distribution, Neural Computation **13** (2001) 1447-1471
19. Bouzida Y., Cuppens F. - Detecting known and novel network intrusion, In: IFIP/SEC 2006, 21st IFIP TC-11 International Information Security Conference Karlstad University, 2006.

20. Bouzida Y., Cuppens F. - Neural networks vs. decision trees for intrusion detection, In: IEEE IST Workshop on Monitoring, Attack Detection and Mitigation, 2006.
21. Xiao H., Hong F., Zhang Z., Liao J. - Intrusion detection using ensemble of svm classifiers, In: Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007, pp. 45-49.
22. Vapnik V. - The Nature of Statistical Learning Theory, Springer-Verlag, 1995.
23. Engen V., Vincent J., Phalp K. - Enhancing network based intrusion detection for imbalanced data, International Journal of Knowledge-Based Intelligent Engineering Systems **12** (5-6) (2008) 357-367.
24. Zighed D. A., Rakotomalala R. - Graphes d'Induction - Apprentissage et Data Mining, Hermes, 2000.
25. Quinlan J. R. - Induction of decision trees, Machine Learning **1** (1) (1986) 81-106.
26. Quinlan J. R. - C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.
27. Wehenkel L. - On uncertainty measures used for decision tree induction, In: IPMU, 1996, pp. 413-418.
28. Loh W. Y., Shih Y. S. - Split selection methods for classification trees, Statistica Sinica **7** (1997) 815-840.
29. Friedman J. H. - A recursive partitioning decision rule for nonparametric classification, IEEE Transactions on Computers **26** (4) (1977) 404-408.

Received June 16, 2010

Address:

Thanh-Nghi Do,

Can Tho University, Vietnam.

Thanh-Nghi Do, Philippe Lenca,

Institut Telecom, Telecom Bretagne, UMR CNRS 3192 Lab-STICC

Université européenne de Bretagne, France.

Stéphane Lallich,

Université de Lyon, Laboratoire ERIC, Lyon 2, France.