THE BAYESIAN RECEIVER OPERATING CHARACTERISTIC CURVE – AN EFFECTIVE APPROACH TO EVALUATE THE IDS PERFORMANCE

Evgeniya Nikolova, Veselina Jecheva
Burgas Free University

Abstract: *The aim of intrusion detection systems is to protect the computer networks from attacks or unauthorized access to their valuable resources. For that purpose they have to perform real-time processing of current system activity and make a decision whether the pattern is normal or intrusive. The present paper is an extension to our previous work, which presented an anomaly-based IDS model, based on some decoding algorithms. It introduces the evaluation of the proposed methodology, using the Bayesian Receiver Operating Characteristics Curve.*

Key words: *database, www programming, ODBC, JDBC database connections.*

1. Introduction

Information security is an important issue in the contemporary ubiquitous network environments. Intrusion detection systems (IDS) monitor the events occurring in a network or a single host and analyse them in order to detect attempts to security violations. The two main categories of IDS principles are signature detection (misuse detection) and anomaly detection [1].

Signature detection attempts to identify events that misuse a system and is performed by creating models of known intrusions. Current events are compared against preliminarily described intrusion models to make a detection decision. This method is good in discovering known attacks, but cannot detect new intrusions with high degree of probability. Anomaly-based IDS, on the contrary, create a model of normal system use and look for activity that does not conform [6]. Any significant deviation is labeled as an attack, since it does not fit the defined model.

The concept of IDS quality is critical, but not clearly defined yet. There are many factors to consider when evaluating IDSs: speed, cost, effectiveness, ease-of-use, scalability, interoperability, etc. Since the main task of IDS is to recognize whether an intrusion attempt is present or absent, the detector’s performance could be considered as the most important feature of any IDS [5]. Classification accuracy IDSs deals with such fundamental problems as how to compare two or more IDSs, how to evaluate the performance of IDS, and how to determine the best configuration of the IDS [2].

Regardless of the underlying methodology, an IDS performance can be treated as a binary classification problem and therefore described by its receiver operating characteristic

(*ROC*) curve [4]. It is a graph of the detection probability versus false alarm rate. It takes into account both false positives, when the IDS flags normal activity as anomalous and false negatives, when the IDS reports an attack as normal activity in error. The *ROC* curve could be applied in order to analyse the tradeoff between the two types of errors. A methodology with perfect discrimination has a *ROC* plot that passes through the upper left corner, consequently the closer the *ROC* plot is to the upper left corner, the higher the overall accuracy of the test [10].

In the case of intrusion detection however, the set of normal activity data contains many more patterns than the set, which contains the intrusion data. Under this setting, many classifiers tend to concentrate on the large classes and disregard the ones with small number of patterns. As a result, IDSs with *ROC* curves achieving "good" operating points still produce a large amount of false alarms in real environments.

The present paper considers detection effectiveness as a general metric to compare IDSs. More details about the applied intrusion detection methodology, which is based on the Junction Tree algorithm and the conducted simulation experiments and their results, are presented in our previous works [7, 8]. The IDSs performance evaluation is based on the Bayesian Receiver Operating Characteristic Curve [3]. *B-ROC* provides a better way to evaluate and compare classifiers in the case of class imbalances, compared to the classical *ROC* curves. *B-ROC*s can be used for comparing classifiers without any assumptions of misclassification costs.

2. The Bayesian Receiver Operating Characteristic Curve

Let I denotes whether a given observation x was generated by an intrusion (represented by I=1) or not (denoted as NI=0). Also let A denotes whether the output of an IDS is an alarm (denoted by A=1) or not (denoted by NA=0). An IDS can then be defined as an algorithm that receives a continuous data stream $X=\{x_1,x_2, . . . ,\}$ and classifies each input x_j as being either a normal event or an attack, i.e. IDS: $X \rightarrow A, NA$. The system can be in one of two states or conditions: either with an intrusion present (I) or with no intrusion present (NI). The prior probability of an intrusion is called p=P[I=1]. The IDS reports either an intrusion alarm (A) or no alarm (NA). The parameters of the IDS's *ROC* curve are: the probability of an alarm given an intrusion, the detection probability, $p_d = P(A|I)=1–\beta$ (or the probability of no alarm given an intrusion, $P(NA|I)=\beta)$, and the probability of an alarm given no intrusion, the false alarm probability, $p_{fa} = P(A|NI)=\alpha$. Thus, α and β are the probabilities of the two types of reporting errors.

An IDS's receiver operating characteristic (*ROC*) curve describes the relationship between the two operating parameters of the IDS, its probability of detection, p_d =1–β, and its false alarm probability, p_{fa} =α. That is, the *ROC* curve displays the 1–β provided by the IDS at a given α. It also displays the α provided by the IDS at a given 1–β. The *ROC* curve thus summarizes the performance of the IDS. A plot lying above and to the left of another plot indicates greater observed accuracy. If on a graph containing *ROC* curves, a single curve lies outside of every other curve, then it dominates the others completely. It is possible for an *ROC* curve to dominate partially, meaning over a certain region of the *ROC* space but not over the entire graph.

Area under the *ROC* curve (*AUC*) is a measure of the overall performance of a diagnostic test and is interpreted as the average value of sensitivity for all possible values of specificity. It can take on any value between 0 and 1, since both the x and y axes have values ranging from 0 to 1. The closer *AUC* is to 1, the better the performance of the test, and a test with an *AUC* value of 1 is one that is perfectly accurate.

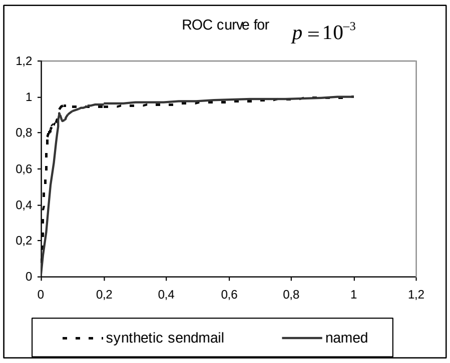

Figure 1. *ROC* curves of synthetic sendmail and named

In Figure 1, although the *AUC* of the two tests have the same value of 0,94, the two curves dominate over a select region of the *ROC* graph. This is a typical problem with the comparison of classifiers by using *ROC*. Since in general case it is difficult to determine misclassification costs, we can get a better comparison of two classifiers without them, using the Bayesian receiver operating characteristic (*B-ROC*) curve.

The *Bayesian detection rate* or *Positive Predictive Value* (*PPV*) is the posterior probability of intrusion given that the IDS fired an alarm, i.e. the probability that a pattern is intrusive when restricted to those patterns which test positive. The *PPV* could be computed as follows:

$$PPV = \frac{p.p_d}{p.p_d + (1-p).p_{fa}},$$

where *p* is a priori probability, p_d - the probability of detection, p_{fa} - the probability of false alarm. Its value is maximized when the false alarm rate of detector goes to zero, even if the detection rate also tends to zero. Therefore we exchange the *PPV* with the *Negative Predictive Value* (*NPV*):

$$NPV = \frac{1-p \;.\; 1-p_{fa}}{p.\; 1-p_d \;+ (1-p).\; 1-p_{fa}}.$$

It represents the patterns with negative test, which are not anomalous, i.e. are results of normal system activity. The *NPV* assesses the reliability of negative test performance.

The *B-ROC* curve is a method of graphically demonstrating the relationship between and *PPV*,$P_d$where we use for x-axis 1-*PPV* – the Bayesian false alarm rate. This curve is a well defined continuous and represents non-decreasing function. *B-ROC* curves get a better comparison between the two classifiers without the assumption of any misclassification costs as might be seen in Figure 2:

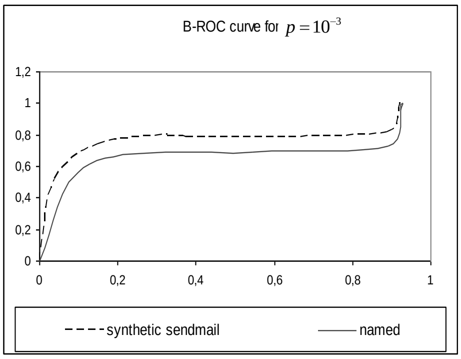

Figure 2. *B-ROC* curves of synthetic sendmail and named

3. The B-ROC curve for comparison of classifiers

In a series of previous works the authors examined the application of some decoding algorithms and techniques in anomaly-based IDS, since this recognition problem was considered as a decoding problem. The proposed methodology consists of two stages – the first contains the HMM creation and its adjustment using the gradient method, and the second one includes the intrusion recognition itself. The authors applied the well-known techniques as the Bahl-Cocke-Jelinek-Raviv (BCJR or the MAP) decoding algorithm, the max-log-MAP algorithm and the junction tree algorithm (JTA) during the recognition stage. The obtained results are presented and discussed in [9].

Since the intrusion detection is a binary classification (normal or intrusive activity), many statistical methods for the results evaluation were applied – *FPR*, *FNR*, sensitivity, specificity, accuracy, *ROC* curves, etc.

A *ROC* curve is a frequently applied non-parametric approach for binary classification method evaluation. In a *ROC* curve each p_d value can be plotted against its corresponding p_{fa} value to create the diagram for the examined processes for IDS based on BCJR algorithm (see Figure 3). The points in the upper left corner of the *ROC* space, which is produces by the proposed methodology for the processes synthetic ftp, named and xlock are (0,09; 0,97), (0,05; 0,98), (0,03; 0,99) respectively.

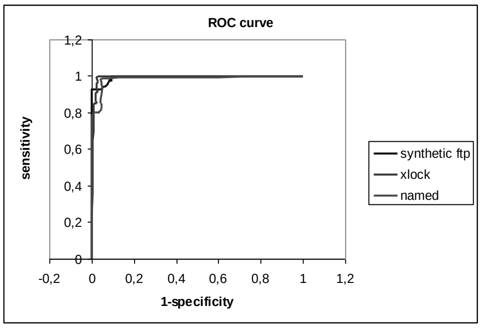

Figure 3. The *ROC* curve for IDS based on BCJR algorithm

In order to evaluate the effectiveness of IDS based on the Max-log MAP algorithm the *PPV*, *NPV* and accuracy were calculated for the processes named, synthetic ftp and xlock, which results are presented in Table 1:

synthetic ftp	named	xlock
PPV=4,9%	PPV=7,5%	PPV=7,9%
NPV=99,5%	NPV=99,4%	NPV=99,4%
Accuracy=87,8%	Accuracy=89,8%	Accuracy=91,1%

Table 1. The *PPV*, *NPV* and accuracy for the examined processes

Considering the obtained predictive values, we see in Table 1 that all *PPV* are between 4,9% and 8,7%, while all *NPV* are between 99,3% and 99,5%. Since the positive predictive values refer to the chance that a positive test result will be correct, the obtained results show that the proposed method correctly classifies the patterns with high degree of probability. On the other hand, negative predictive value is concerned only with negative test results. From the Table 1 we see that the proposed methodology produces results with excellent negative predictive values. The both predictive values depend on the prevalence of the intrusions, since they depend on the number of true positives and false negatives and true negatives and false positives, respectively. Since the accuracy values for all processes belong to the interval (87%, 92%), we can conclude that the proposed methodology produces precise and reliable detection results.

Figure 4 contains the *ROC* curves for the examined processes, obtained by JTA. The points in upper left corner of the *ROC* space for the processes synthetic sendmail, synthetic ftp, named and xlock are (0,06; 0,94), (0,05; 0,99), (0,55; 0,95) and (0,25; 0,85) respectively. As the point (0, 1) denotes the perfect detection, the proposed methodology produces reliable and qualitative results while distinguishing the normal activity from abnormal one.

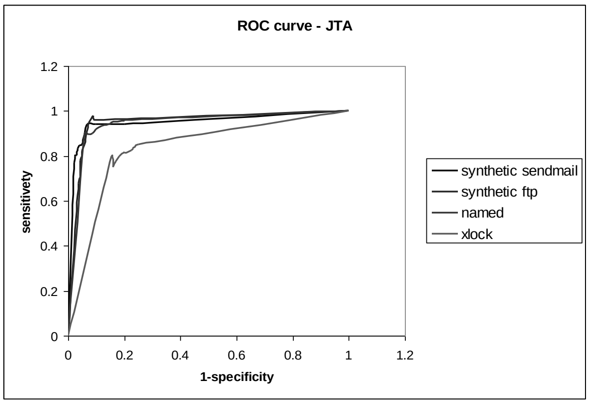

Figure 4. *ROC* curve for IDS based on JTA

It is clear from figures 3 and 4 that the comparison between results, obtained for the different processes when the corresponding decoding algorithm was applied, is not a trivial task, since the algorithms produce results with very small deviations.

In order to compare the effectiveness of IDS based on these algorithms we plot the *B-ROCs*, which results are presented in Figures 5-7:

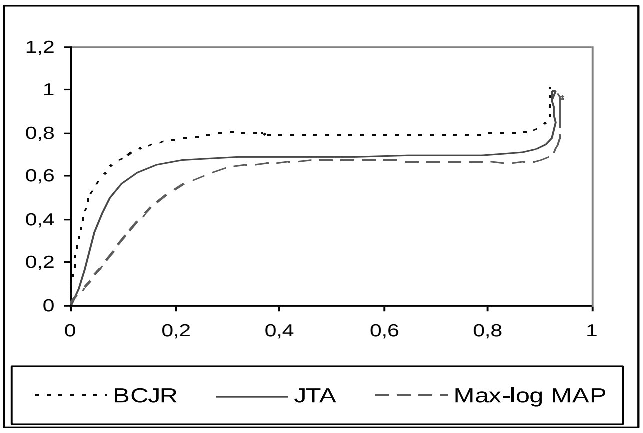

Figure 5. *B-ROC* curves for synthetic ftp

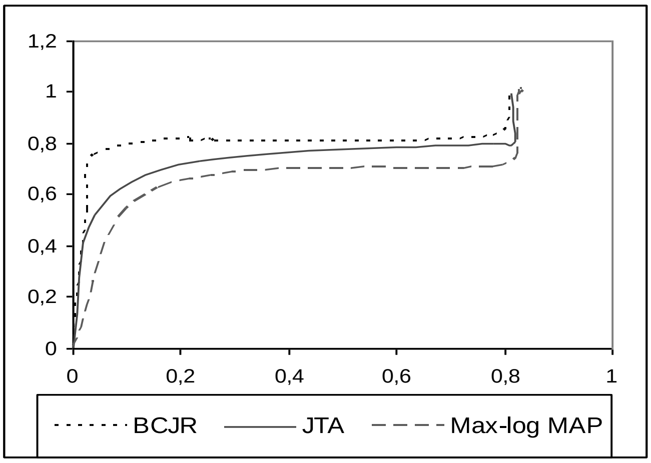

Figure 6. *B-ROC* curves for named

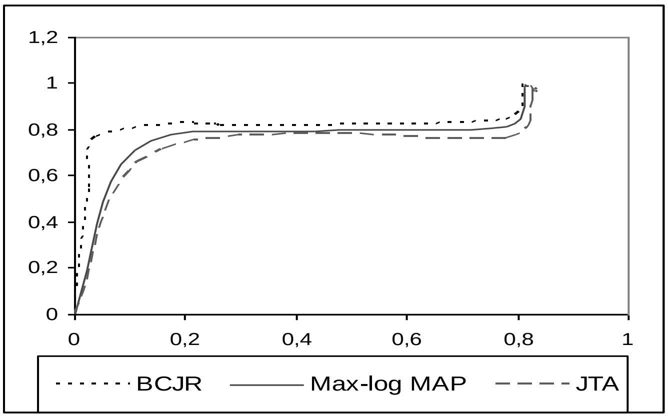

Figure 7. *B-ROC* curves for xlock

It could be seen from figures 5-7 that the methodology, based on BCJR decoding algorithm, outperforms the two other methodologies for the three examined processes. The methodology, based on JTA, produces balanced results for the processes named and synthetic ftp, and produces better classification results compared to the methodology, based on max-log-MAP algorithm. At last, the methodology, based on max-log-MAP algorithm, produces better classification results, compared to those, produced by JTA for the process xlock. We should mention that regardless of the fact, that methodology, based on BCJR decoding algorithm, outperforms the two other methodologies for the examined processes, the two other methodologies produce reliable and qualitative results while distinguishing the normal activity from abnormal one.

Conclusion

We believe that the *B-ROC* provides a better way to evaluate and compare classifiers in the case of uncertain values of *p*. When comparing two classifiers, there are cases in which by using the *B-ROC*, we do not need cost values in order to decide which classifier would be better for given values of *p*. Note also that *B-ROCs* consider parameters that are directly related to exact quantities that the operator of a classifier can measure.

References:

1. Axelsson S., “Intrusion Detection Systems: A Taxonomy and Survey,” Technical Report No 99-15, Dept. of Computer Engineering, Chalmers University of Technology, Sweden, March 2000.
2. Cardenas A. A., J. S. Baras, K. Seamon, A framework for the evaluation of intrusion detection systems, Proceedings of the 2006 IEEE Symposium on Security and Privacy, May 2006, Berkeley/Oakland, CA, ISBN: 0-7695-2574-1, pp. 15-77.
3. Cardenas A. A., J. S. Baras, B-ROC curve for the assessment of classifiers over imbalanced data sets, www.aaai.org, 2006.
4. Egan J.P., Signal detection theory and ROC-analysis, Academic Press, 1975.
5. Gaffney J.E., J.W. Ulvila, Evaluation of Intrusion Detectors: A Decision Theory Approach, The IEEE Symposium on Security and Privacy, 2001, pp. 50-61.
6. Lee W., S. J. Stolfo, “A Framework for Constructing Features and Models for Intrusion Detection Systems,” In ACM Transactions on Information and System Security, Vol. 3, No. 4, 2000, pp. 227–261.
7. Nikolova E., V. Jecheva, Anomaly Based Intrusion Detection Based on the Junction Tree Algorithm, Journal of Information Assurance and Security, Dynamic Publishers Inc., Vol. 2, Issue 3, 2007, pp.184-188.
8. Nikolova E., V. Jecheva, Some Evaluations of the Effectiveness of Anomaly Based Intrusion Detection Systems Based on the Junction Tree Algorithm, Proceedings of the 5th CITSA 2008, Orlando, Florida, June 29th - July 2nd, 2008, vol. 1, pp.115-120.
9. Nikolova E., V. Jecheva, Chapter: The Decoding Algorithms as Techniques for Creation the Anomaly Based Intrusion Detection Systems, Engineering the Computer Science and IT, IN-TECH, Vienna, Austria, 2009, ISBN 978-953-7619-32-9.
10. Zweig M.H., G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clinical Chemistry, Vol.39, Num.4, (1993), pp. 561-577.