

ANOMALY DETECTION IN COMPUTER NETWORKS USING LINEAR SVMs

*Carolina Fortuna**, *Blaž Fortuna[#]*, *Mihael Mohorčič**

* Department of Communication Systems, Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia

[#]Department of Knowledge Technologies, Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia

Tel: +386 1 4773900; e-mail: carolina.fortuna@ijs.si

ABSTRACT

Modern computer networks are subject to various malicious attacks. Since attacks are becoming more sophisticated and networks are becoming larger there is a need for an efficient intrusion detection systems (IDSs) that can distinguish between legitimate and illegitimate traffic and be able to signal attacks in real time, before serious damages are produced. In this paper we use linear support vector machines (SVMs) for detecting abnormal traffic patterns in the KDD Cup 1999 data. The IDS system is supposed to distinguish normal traffic from intrusions and to classify the intrusions into four classes: DoS, probe, R2L and U2R. The dataset is quite unbalanced, with 79% of the traffic belonging to the DoS category, 19% is normal traffic and less than 2% constitute the other three categories. This paper studies the performance of IDSs based on linear multi-class SVMs with highest confidence (one-to-all), majority (one-to-one) and two level (one-to-all-3categ) voting on this particular dataset. The one-to-all-3categ IDS is tailored to perform well on the unbalanced dataset but it proves to be less efficient when trained on large datasets. The one-to-one IDS turns to perform the best on larger training dataset. The best performing IDS has a 90.9% intrusion detection rate, 90.7% intrusion diagnosis rate and 0.2479 average cost per test example (ACTE).

1 INTRODUCTION

A computer network can be the target of attacks both from the intra and extra domain. The critical nodes of such a network need to be monitored from a centralized location in such way as to prevent potential damages. Since computer networks increase their size and attacks are evolving continuously, it is hard for a human system engineer to efficiently combat intrusions. A machine learning software can alert in real time both known and unknown attack types. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between "bad" connections, called intrusions or attacks, and "good" normal connections [1].

The winner of the KDD Cup 1999 used a mixture of bagging and boosting taking into consideration asymmetric

error costs by minimizing the so called initial costs [2]. The competitors ranked on the second and third place used decision trees and tailored their methods on the nature of the dataset [3][4]. Recently, a three tier IDS using multi-class SVM with Gaussian kernel was used to produce better results than the KDD Cup 1999 winner. Their approach is to combine the strength of the classical signature based detection to the more recent anomaly detectors, since the first one performs well on known attacks and the second performs well on novel attacks. In their approach, they use a three tier IDS: on the first tier a black list is built that is able to filter out probe, DoS, R2L and U2R attacks based on signature. The normal and unknown attacks that pass the filters in the first tier are passed to the second tier. Here a white list is used to filter normal traffic from the unknown attacks. Finally, on the third tier, Smooth SVM [12] is used to determine the category the unknown attacks belong to [5].

In our approach we study the performance of IDSs based on linear multi-class SVM with highest confidence (one-to-all IDS) and majority voting (one-to-one IDS) respectively and propose a simple and rapid IDS that uses multi-class SVM with linear kernel and two level voting (one-to-all-3categ). In the first step of the approach, SVMs are trained on subsets of the 10% training dataset and a subset of the full dataset, according to the IDS architecture. The models built by SVM are used to classify new instances and three voting types are used to decide the final class the new instance belongs to. We expect that the one-to-all-3categ IDS performs better than the other two IDSs given the nature of the training dataset but this proves to hold only for small training datasets. On the 10% training dataset, the one-to-one IDS yields the best results, outperforming the one-to-all-3categ IDS. It seems that R2L connections are spread across the space and linear SVM is not able to build a model that can classify them accurately, especially that these instances appear in small number in the training dataset. In the testing dataset, their number increases dramatically and the misclassification cost is high. R2L connections are misclassified as normal connection most of the times.

The rest of the paper is structured as follows. Section 2 describes the criteria for evaluating IDSs, Section 3 describes the dataset, Section 4 details the experiments

focusing on the data preprocessing, the machine learning method used and the three IDS architectures and discusses the results. Finally, Section 5 concludes the paper.

2 EVALUATION CRITERIA

Several metrics are used to evaluate and compare the performance of Intrusion Detection Systems (IDSs). The most basic metrics are the detection and false alarm rates. The detection rate is equal to the number of intrusions detected divided by the total number of intrusions in a data set, while the false alarm rate is equal to the number of normal instances detected as intrusions divided by the number of normal instances in a data set. False alarms are also referred to as false positives [7]. The diagnosis rate (or recall), meaning the number of correctly classified intrusions divided by the total number of intrusions, is also a relevant metric and we refer to it across this paper.

In the KDD Cup 1999 the criteria used for evaluation of the participant entries is the ACTE computed using the confusion matrix and a given cost matrix. The confusion matrix is obtained while classifying the instances in the test dataset. Each column of the confusion matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The cost matrix is given in Table 1.

	normal	Probe	DOS	U2R	R2L
normal	0	1	2	2	2
probe	1	0	2	2	2
DOS	2	1	0	2	2
U2R	3	2	2	0	2
R2L	4	2	2	2	0

Table 1 Cost matrix

From the table above, it can be noticed that the most expensive is misclassifying U2R and R2L instances as normal instances.

3 DATASET

The KDD Cup 1999 uses a version of the data on which the 1998 DARPA Intrusion Detection Evaluation Program was performed. The training dataset was acquired in a seven week time frame of monitoring the network and was processed into almost 5 million instances. The test dataset was acquired during a two week time frame and contains 311029 instances. Both training and test datasets are labeled with the name of the attack type or as being normal traffic [1]. There are 38 different attack types in training and test data together and these attack types fall into four main categories: probe, denial of service (DoS), remote to local (R2L) and user to root (U2R) [2].

The dataset is extremely unbalanced; most instances are DoS traffic (79%), while the other three attack types together make less than 2% of the instances. Around 19% of the instances correspond to normal traffic. The test dataset has different distribution than the training dataset and contains several new attacks (17 new attacks out of 38

possible attacks). Figure 1 depicts the distribution of the full training dataset, 10% of the full training dataset and of the testing dataset. It can be noticed that the normal, probe and DoS connections keep their distribution across the three datasets while the same is not valid for U2R and R2L connections. For U2R connections a slight increase in number of instances in the test dataset versus the training dataset can be noticed. U2R instances represent 0.01% of the 10% training dataset and 0.2% of the test dataset. On the other hand, the proportion of the R2L connections dramatically increases in the test dataset (5.2%) comparing to the training one (0.2%). Furthermore, the R2L connections are spread in space posing real challenge for determining an accurate model for classification.

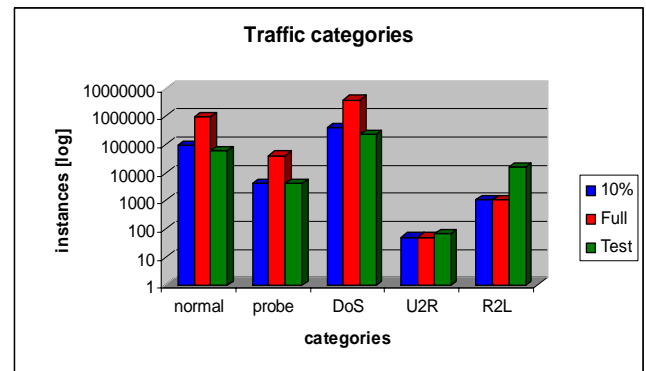


Figure 1 Traffic distribution in KDD Cup 1999 dataset

4 EXPERIMENTS

4.1 Data preprocessing

Each instance in the KDD Cup 1999 datasets contains 41 features that describe a connection. Features 1-9 stand for the basic features of a packet, 10-22 for content features, 23-31 for traffic features and 32-41 for host based features [6]. There are 7 nominal and 34 continuous features. Since SVM does not take as input nominal values, the 7 nominal features had to be transformed so that the resulting datasets had 108 features for each instance. Given the large dimension of the full dataset (around 5 million instances), we used only 10% of it (494021 instances) in most of the experiments. However, the first experiments were performed on a smaller dataset (100.000 instances) sampled from the 10% dataset in such way that the three minority classes were kept unchanged. This approach is expected to be faster (since the dimension of the training data is smaller) and build a better model for the three minority classes (as their weight in the dataset has been artificially increased).

4.2 SVM

The machine learning method used in this paper is the support vector machines (SVMs) [11]. SVMs are a set of related supervised learning methods used for classification and regression. The experiments in this paper use linear SVM as implemented in TextGarden [10].

4.3 One-to-all, one-to-one and one-to-all-3categ IDSs

The one-to-all IDS uses the 10% training dataset and preprocesses it as described in Section 4.1. After preprocessing, five training files are created. In each of the files, one attack type represents the positive class and all the other attacks represent the negative class. The SVM is trained on these five files and for each input file, it builds an output model that distinguishes between the positive class and all the other classes in the input, this is why the name one-to-all. Each connection in the test data is then fed to the models, each model decides if the connection belongs or not to a class with a certain degree of confidence. The connection is classified as belonging to the class that classified it with highest confidence. Figure 2 presents the workflow of the one-to-all IDS. The outcome of the voting is summarized in a confusion matrix and finally the average cost per text example is computed.

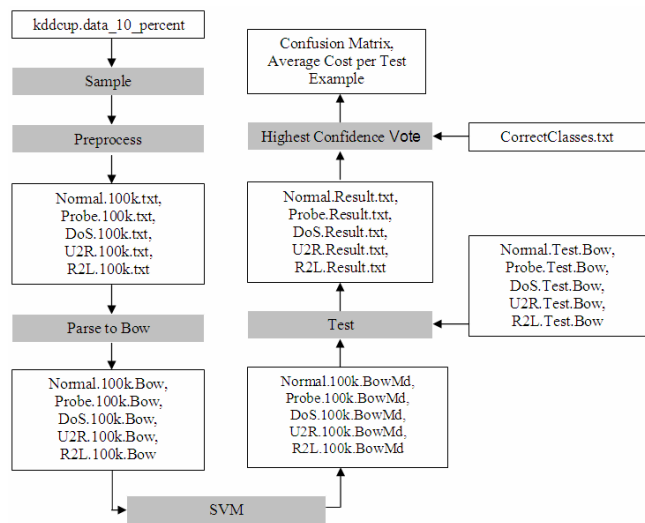


Figure 2 One-to-all IDS

The one-to-one IDS works similarly as the one-to-all IDS with two exceptions: the training files and the voting method. Each training files contains only two types of attacks: one represents the positive class and the other represents the negative class. This way 10 training files are prepared and 10 models are built. When a new connection has to be classified, each model decides for one of the two classes the connection belongs to. The connection is classified as belonging to the class to which the majority of the models assigned it to.

Figure 3 presents the workflow of the one-to-one IDS.

The third IDS tries to adapt to the nature of the training data. Given the unbalanced nature of the data, it attempts to build a better model for classifying minority classes. In order to achieve this, two sets of one-to-all training files are used. The first set is formed of two files in which the positive class is represented by normal and DoS connections respectively, and the negative class is represented by all other types of connections (one-to-all test files). The second set of training files contains only three types of connections: probe, R2L and U2R filtered from the full dataset, resulting in three one-to-all files (one-to-all-

3categ files since the “all” stands for the other two minority categories). SVM is trained on all five files and a two level voting is applied to the new instances. In the first level, the system determines if the connection belongs to any of the two majority classes, DoS or normal, based on a highest confidence voting. If the connection does not belong to any of the two classes, it goes to the second level where the system determines if it belongs to probe, R2L or U2R classes also based on a highest confidence vote. Figure 4 presents the workflow of the one-to-all-3categ IDS.

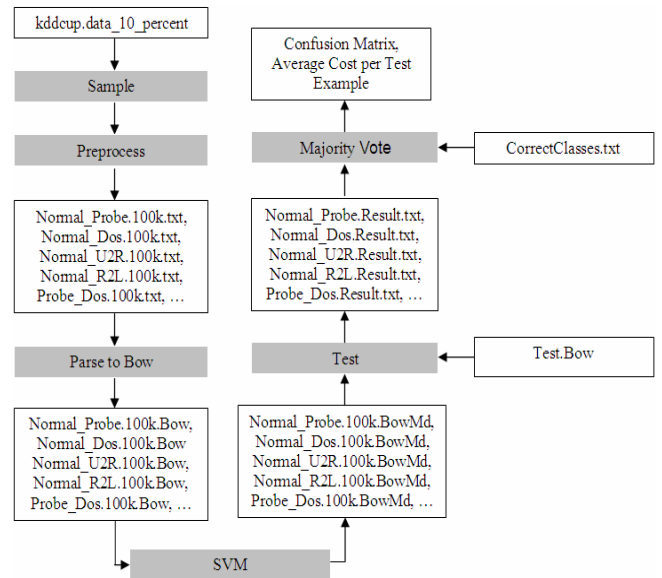


Figure 3 One-to-one IDS

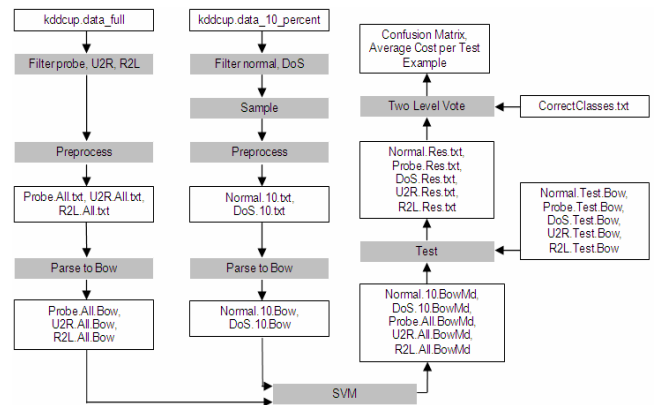


Figure 4 One-to-all-3categ IDS

4.4 Experimental Results

When dealing with such large and unbalanced datasets as the one provided for the KDD Cup 1999, an important step is to understand the data and find a suitable model for it. Our approach was to build models on a 100.000 instance dataset obtained as explained in Section 4.1 and classify the test dataset using the three IDSs described in Section 4.3. Table 2 presents the results obtained at this stage. The one-to-one IDS has the poorest ACTE, the one-to-all-

3categ IDS has the best ACTE while the results for one-to-all IDS are somewhere in between. The one-to-all IDS has a high detection rate, a good diagnosis rate but a very high false alarm rate meaning that it classifies most of the normal traffic as intrusion. This system doesn't detect probe, R2L and U2R intrusions at all. All the traffic is classified as DoS or normal, but it seems that it confuses DoS with normal quite often. This might be due to the SVM cost parameters that are not optimized for this dataset or to the nature of the dataset. The one-to-one scenario has lower false alarm rate, but has poor diagnosis performance, meaning that it detects most of the alarms, but it doesn't classify them correctly. The high ACTE seems to come from misclassifying DoS attacks (over 220.000 instances out of 311.000) for R2L attacks. Finally, the one-to-all-3categ IDS gives the best results: good ACTE, good detection and diagnosis rates and low false alarm rate. However, this result might be further improved by parameter tuning or increasing the size of the training dataset.

	One-to-all	One-to-one	One-to-all-3categ
ACTE	0.5306	1.6656	0.2641
Detection rate	99.2%	95.0%	90.3%
Diagnosis rate	91.3%	3.3%	90.1%
False alarm rate	99.6%	12.8%	1.6%

Table 2 Results for 100.000 instance training set

The next step in the approach was to tune SVM parameters in order to build more accurate models. The 10% training dataset (494021 instances) with 10 fold cross validation were used to build the models and the three resulting IDSs were then tested. The results are listed in Table 3.

	One-to-all	One-to-one	One-to-all-3categ
ACTE	0.2625	0.2479	0.2653
Detection rate	90.2%	90.9%	90.3%
Diagnosis rate	90.1%	90.7%	90.1%
False alarm rate	1.6%	2.02%	1.6%

Table 3 Results for 10% training set

The one-to-all IDS improved the overall performance as well as the detection, diagnosis and false alarm rates. Both detection and diagnosis rates are quite good and false alarm rate is low, meaning the system detects and correctly determines the class of over 90% of connections and has a small false alarm rate (1.6%). The one-to-one IDS also improved: it has the smallest ACTE and good detection and diagnosis rate. The false alarm rate is slightly higher than for the one-to-all IDS. The most unexpected result comes from the one-to-all-3categ IDS: there is no improvement in the detection, diagnosis and false alarm rates. The ACTE slightly increases, due to more expensive (see the cost matrix) misclassifications.

We can go more into detail with the analysis of the performance of the three IDSs by comparing the output confusion matrices listed in Table 4, Table 5 and Table 6. Rows represent the labels of the connections and columns represent the class attributed by the IDS. The last row displays the rate of true positives (e.g. 71.0% of the connections classified as normal are normal) and the last column displays the accuracy (e.g. 98.3% of normal traffic was classified as normal).

	normal	probe	DOS	U2R	R2L	%
normal	59611	300	678	4	0	98.3
probe	1053	2922	191	0	0	70.1
DOS	7242	22	222589	0	0	96.8
U2R	54	0	0	11	5	15.7
R2L	15959	16	2	2	368	2.2
%	71.0	89.6	99.6	64.7	98.6	

Table 4 One-to-all confusion matrix (ACTE = 0.2625)

	normal	probe	DOS	U2R	R2L	%
normal	59367	211	818	12	185	97.9
probe	901	3002	148	0	115	72.0
DOS	7047	52	222754	0	0	96.9
U2R	32	0	0	32	6	45.7
R2L	14791	11	2	11	1532	9.3
%	72.2	91.6	99.5	58.1	83.3	

Table 5 One-to-one confusion matrix (ACTE = 0.2479)

	normal	probe	DOS	U2R	R2L	%
Normal	59593	313	672	5	10	98.3
probe	767	3120	181	6	92	74.8
DOS	7113	324	222406	0	10	96.7
U2R	60	0	0	5	5	7.1
R2L	16186	11	2	1	147	0.8
%	71.1	82.8	99.6	29.4	55.6	

Table 6 One-to-one-3categ confusion matrix (ACTE = 0.2653)

It can be seen in Table 4 that the one-to-all IDS performs well on normal and DoS connections, on probe it has a rather poor performance (70.1% diagnosis) and misclassifies most of U2R (15.7% diagnosis) and R2L (2.2% diagnosis) connections. Most of the misclassified probe, U2R and R2L connections are classified as normal. The models for normal and DoS traffic are fairly accurate since they had a large set of training instances to build on.

The one-to-one IDS performs better than one-to-all IDS as can be seen in Table 5. This IDS performs significantly better than one-to-all IDS on classifying U2R and R2L connections: it classifies 45.7% of U2R connections and 9.3% of R2L connections. The R2L connections are spread in space so that linear SVM proves to be inefficient for building a good model for classifying these instances. We noticed a tradeoff: the more accurate the SVM model for classifying R2L connections, the poorest in classifying normal connections and the other way around.

The one-to-all-3categ IDS performs worse than the other two IDSs in classifying R2L and U2R attacks, and

performs slightly better on classifying probe attacks. It seems indeed that linear SVM is limited in building a good model for separating normal traffic from R2L due to the spread of these connections. Even though we introduced the one-to-all-3categ IDS in order to perform better at separating the three minority classes from the two major ones (normal and DoS), it seems like the model built using SVM is not accurate enough so that this voting system proves efficient. Most of the R2L connections do not pass the first level voting, being classified as normal.

Comparing to relevant results in the literature, the IDSs studied in the paper are less accurate. The one-to-one IDS with 0.2479 ACTE would rank 8th in the KDD Cup 1999 contest. Higher accuracy can be obtained by increasing the complexity of the system. SVMs with different kernels can be used for building better models, but with this approach, classification speed would decrease [11], this is undesired in real time IDSs. Hybrid systems that combine several machine learning methods or that combine machine learning methods with the more classical ones based on signatures could be used.

4 CONCLUSIONS

In this paper we studied the performance of linear SVM in classifying normal and attack connections sniffed from a computer network. We proposed a two level voting IDS that proved to perform well on a small training set but performed relatively poor when the training dataset increased. In the context of intrusion detection in a computer network, attacks such as R2L and U2R that result in small number of traffic packets seem to pose a real challenge for detection and diagnosis. A good, simple and fast classifier that is able to detect novel attacks is hard to build. Usually simplicity and speed are traded for accuracy and machine learning methods are complemented by traditional signature based methods.

Acknowledgement

This work was supported by the Slovenian Research Agency and the IST Programme of the EC under NeOn (IST-4-027595-IP) and PASCAL (IST-2002-506778).

References

- [1] KDD Cup 1999 Task Description, <http://kdd.ics.uci.edu/databases/kddcup99/task.html>
- [2] Bernhard Pfahringer, Winning the KDD99 Classification Cup: Bagged Boosting, *ACM SIGKDD Explorations Newsletter*, Volume 1, Issue 2, p. 65-66 January 2000.
- [3] Itzhak Levin, KDD-99 Classifier Learning Contest LLSoft's Results Overview, *ACM SIGKDD Explorations Newsletter*, Volume 1, Issue 2, p. 67-75 January 2000.
- [4] Vladimir Miheev, Alexei Vopilov, Ivan Shabalin, The MP13 Approach to the KDD'99 Classifier Learning Contest, *SIGKDD Explorations Newsletter*, Volume 1, Issue 2, p76-77 January 2000.
- [5] Tsong Song Hwang, Tsung-Ju Lee, Yuh-Jye Lee, A Three-tier IDS via Data Mining Approach, *MineNet'07*, June 12, 2007, San Diego, California, USA
- [6] W. Lee. A Data Mining Framework for Constructing Features and Models for Intrusion Detection Systems. PhD thesis, Columbia University, 1999.
- [7] Computer Security and Intrusion Detection, <http://www.acm.org/crossroads/xrds11-1/csid.html>
- [8] H. Gunes Kayacik, Nur Zincir-Heywood, Malcolm I. Heywood, Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD '99 Benchmark, http://www.unb.ca/pstnet/pst2005/Shaghnessy%20Rom/Oct13/GK_FeatRelevance.ppt#256,1,Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Benchmark
- [9] Results of the KDD Cup 1999 Classifier Learning Contest, <http://www-cse.ucsd.edu/users/elkan/clresults.html>
- [10] TextGarden – Text Mining Tools, <http://kt.ijs.si/Dunja/textgarden/>
- [11] C. Cortes and V. Vapnik, Support-Vector Networks, *Machine Learning*, 20(3):273-297, September 1995.
- [12] Y.-J. Lee and O. L. Mangasarian. SSVM: A smooth support vector machine. *Computational Optimization and Applications*, 20:5–22, 2001. Data Mining Institute, University of Wisconsin, Technical Report 99-03.