A SVM Model based on Network Traffic Prediction for Detecting Anomalies

Bruno L. Dalmazo
dalmazo@dei.uc.pt
João P. Vilela
jpvilela@dei.uc.pt
Marilia Curado
marilia@dei.uc.pt

CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal

Abstract

Cloud computing is a natural evolution of distributed computing combined with service-oriented architecture. However, its broad adoption has been hampered due to the lack of security mechanisms. Facing this issue, this work aims to propose a new approach for detecting anomalies in the cloud network traffic. The anomaly detection mechanism works on the basis of a Support Vector Machine (SVM) model for binary classification. The key point to improve the accuracy of the SVM model, in the cloud context, is the set of features. In light of this, we present the Poisson Moving Average predictor as the feature extraction approach that is able to cope with the vast amount of information generated over time. We evaluate the performance of our mechanism and compare it against similar studies in the literature, resorting to a real case validation study.

1 Introduction

In the virtual environment, online threats are constantly evolving. Furthermore, cloud computing introduces significant new paths of attack. Distributed Denial of Service (DDoS) is a well-known type of attack that disrupts online operations. Usually, the assault is performed by hundreds (or thousands) of requests for service and it has to be detected before it breaks down the server. Due to the high number of simultaneous requests, this attack generates an anomalous behaviour in the network traffic. Nevertheless, the elastic and scalable nature of the cloud environments is also apt to undergo sudden changes [1], making it even harder to detect which parts of the incoming traffic are from vandalism or legitimate usage.

Several techniques have been already proposed to perform anomaly detection in the cloud environment, such as fuzzy logic, artificial neural network and decision tree classifier. Also, different types of network traffic information are used to detect anomalies, such as protocols behaviour, CPU utilization and user logs. However, there is an apparent deficiency in detecting anomalies with low rates of false alarms. In particular, these techniques require extensive tuning for improving the sensitivity and achieving satisfactory results. There is also no consensus about the best set of features which should be monitored in the network. Moreover, attackers have been able to study the mechanisms behind these techniques and adapt the attack behaviour to evade identification. In this context, the literature lacks mechanisms able to improve the accuracy of the anomaly detection for cloud while keeping a low false detection rates.

To overcome these gaps, a new approach to detect anomalies in a cloud environment is proposed. Our proposal resorts to traffic prediction to generate features that represent the expected proper behaviour of the network traffic. This information is then used jointly with a Support Vector Machine (SVM) model that is fed with the features extracted from network traffic prediction. The combination of these two tools represents a novel and effective approach for detecting anomalous events in the cloud environment. The forecast is performed by a statistical method based on a Poisson process, that has shown itself suitable for dynamic environments such as cloud computing [3]. SVM is already known as one of the best learning algorithms for binary classification [4]. Binary classification meets the goal of this proposal, since we aim to identify the anomalies inside the normal network traffic.

The remainder of the paper is organized as follows. Section 2 covers some of the most prominent related work. Section 3 describes the proposed solution and the methodology used for this paper, whilst Section 4 presents the evaluation and discusses the results. Section 5 concludes with some final remarks and prospective directions for future research.

2 Related Work

This section presents the latest research findings on SVM models applied in intrusion detection system context. Mulay *et al.* [7] presented an IDS that combines SVM and decision trees to build a multi-class SVM. This model can classify the network traffic in normal or abnormal. Horng *et al.* [6] proposed a Network Intrusion Detection System on the basis of Support Vector Machine with features selected by a hierarchical clustering algorithm. The DARPA dataset was used to evaluate the proposed IDS.

Shon and Moon [8] presented a hybrid machine learning approach to detect anomalies in the network traffic. This model is a blending between supervised and unsupervised SVM model. Besides, they use a Genetic Algorithm for extracting more appropriate packet fields (protocol, source port, IP, TTL). Chen *et al.* [2] did a comparative study between Artificial Neural Network (ANN) and Support Vector Machine to predict attacks on the basis of frequency-based encoding techniques to select the features. The results have shown that both approaches are able to detect anomalies in the network traffic, but SVM outperforms ANN.

The key point for using the SVM model with success, in the cloud context, is finding the proper feature extraction approach able to deal with the vast amount of information generated over time. In summary, there is no traditional anomaly detection system to meet these requirements efficiently, since the cloud computing environments have their particular nature and essence. In order to deal with these limitations, in the following section we introduce a conceptual solution for detecting anomalies in the cloud network traffic, by means of a Support Vector Machine fed with features gathered from the Poisson Moving Average predictor.

3 Anomaly Detection Mechanism

The purpose of our *Anomaly Detection Mechanism* is to provide an efficient method to detect anomalies in the cloud-based network traffic. Figure 1 depicts the basis of our mechanism, by highlighting the application scenario and the main conceptual components.

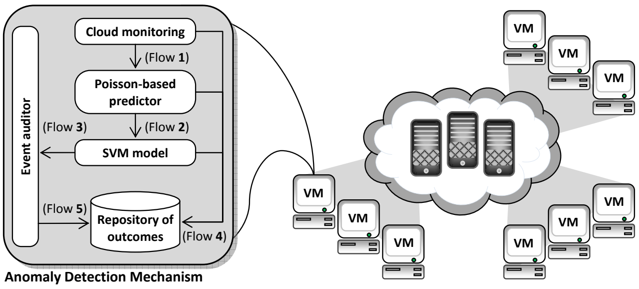

Figure 1: Application scenario and elements of the proposed mechanism

The cloud provider offers several services by the Internet, such as infrastructure, software and platform to the clients. Real-time cloud traffic data (Flow 1) is continuously being gathered from the cloud environment by the *Cloud Monitoring* module. This information is subsequently processed by the *Poisson-based Predictor* that performs prediction based on information such as the protocol type, the number of network packets and timestamp.

After that, the *SVM Model* is fed with features extracted from the predicted data (Flow 2). Then, the *SVM Model* triggers a warning to the *Event Auditor* when an anomalous behaviour is detected (Flow 3). In the meantime, the *Repository of Outcomes* component stores a detailed output regarding the historic of the Virtual Machine (VM) operation (Flow 4). Furthermore, the *Event Auditor* represents an agent placed in the VM that is able to communicate collaboratively with agents in the other VMs. This agent receives any anomalous event from the *SVM Model* and builds a message with information of all components (Flow 5) for sending alerts to other agents.

Having presented an overview of the anomaly detection mechanism, in the following subsections there will be a more detailed description of the forecasting approach for estimating network traffic on the basis of a Poisson process and the Support Vector Machine model for detecting anomalies in the cloud-based environment.

3.1 Poisson-based Predictor

Predicting the network traffic is an important instrument to support a better understanding of the network traffic behaviour. In light of this, the *Poisson-based Predictor* component represents the feature extraction approach. For, after that, it feeds the SVM model.

The predictor receives as input, the time series created in the *Cloud Monitoring*. Therefore, the predictor is able to forecast the expected value in the network traffic according to the temporal granularity of the time series. We consider prediction based on the Poisson Moving Average (PMA) because it has shown to be suitable for dynamic cloud environments [3]. At this point, the predictor will generate several features, as the outcome of this process. The set of variables used in this approach to describe the network traffic are: the time, the type of protocol, the number of packets, the predicted data and the variance between real network traffic and the predicted network traffic.

3.2 SVM Model

The Support Vector Machine (SVM) model uses a methodology for choosing the best hyperplane (among many others) that represents the largest margin between two classes, namely, normal network traffic and anomalies in this work. Then, the hyperplane is chosen such that the distance from it to the nearest support vector on each side is maximized [4].

The Support Vector Machine learning model includes two stages: training and testing. The first learns the two possible patterns of the network traffic (the normal and the anomalous behaviour). The second tests the knowledge achieved in the past stage to detect unknown anomalies. Separating data into training and testing data is an important part of validating the SVM model. By this, we can minimize the effects of data inconsistencies and better understand the characteristics of the model. Once the SVM model has been processed by using the training set, it is needed to evaluate the prediction capability against the training set. Considering the data in the testing set already contains known values for the attribute that we want to predict, it is possible to determine whether the model's suggestions are correct.

In short, the anomaly detection for the cloud network traffic based on SVM with PMA expresses a process of pattern recognition. In this process, the training data represents the standard pattern and the testing data alludes to identify such pattern. The process of identifying a particular behaviour inside of the testing data is a mapping process of the testing data in some existing pattern of the training data.

4 Evaluation

We consider the DARPA dataset [5] for evaluating this proposal. Table 1 shows the comparison among several approaches that use SVM and DARPA dataset to validate the model. Regarding detection rate (DR) point of view, the best model is proposed by Chen W. *et al.* [2], but this approach showed more than 10% of false positive rate (FPR). Also, the omission of FNR results hampers a better evaluation of this approach's performance.

Another model with high DR, but low FPR, is the Soft margin SVM with Radial Basis Function (RBF) kernel. This model obtained 98.65% of DR, but at cost of high false negative rate (FNR), more than 11%. Other models presented in the Table 1 present at least one drawback: low accuracy, high FPR or high FNR.

In summary, our method on the basis of SVM and RBF kernel with features extracted from Poisson Moving Average predictor presents the best equilibrium in the results. It reaches 98.56% of detection rate and 8% of FNR. Also, our approach displays the lowest FPR among the related work, just 1.44%.

Table 1: Approaches that use SVM and DARPA dataset

Approach	Kernel	DR(%)	FPR(%)	FNR(%)
LIBSVM and PMA	RBF	98.56	1.44	8.00
Horng S. *et al.* [6]	RBF	95.72	N/A	N/A
Soft margin SVM	Inner product	90.13	10.55	4.36
Soft margin SVM	RBF	98.65	2.55	11.09
Soft margin SVM	Sigmoid	95.03	3.90	12.73
One-class SVM	Inner product	53.41	48.00	36.00
One-class SVM	RBF	94.65	20.45	44.00
Enhanced SVM [8]	Sigmoid	87.74	10.20	27.27
Chen W. *et al.* [2]	RBF	100.00	10.35	N/A

5 Conclusions and Future Work

In this paper, we have shed light on the major problem for preventing the adoption of the cloud service models: security (or lack thereof). In this context, a novel approach to detect anomalies in the cloud scenario was proposed. The anomaly detection approach relies on a distributed and collaborative mechanism that combines a Support Vector Machine model with features extracted from a Poisson Moving Average predictor.

By analysing the evaluation results, it can be seen that the anomaly detection mechanism was able to identify anomalies considering a case study with real data. Our SVM model achieved high degree of accuracy providing the best compromise in terms of detection and false alarm rates. In particular, our approach exhibits the lowest level of false positive rate and the second best false negative rate in comparison with other approaches. Prospective directions for future research also include: (i) proposing a feature selection approach using Genetic Algorithms; and (ii) proposing an unsupervised Support Vector Machine model.

References

[1] Hitesh Ballani, Paolo Costa, Thomas Karagiannis, and Antony IT Rowstron. Towards predictable datacenter networks. In *SIGCOMM*, volume 11, pages 242–253, 2011.

[2] Wun-Hwa Chen, Sheng-Hsun Hsu, and Hwang-Pin Shen. Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 32(10):2617–2634, 2005. ISSN 0305-0548.

[3] Bruno L. Dalmazo, Joao P. Vilela, and Marilia Curado. Online traffic prediction in the cloud: A dynamic window approach. In *The 2nd International Conference on Future Internet of Things and Cloud (FiCloud'2014)*, pages 9–14, Aug 2014. doi: 10.1109/FiCloud.2014.12.

[4] Naiyang Deng, Yingjie Tian, and Chunhua Zhang. *Support vector machines: optimization based theory, algorithms, and extensions*. CRC Press, 2012.

[5] J.W. Haines, L.M. Rossey, R.P. Lippmann, and R.K. Cunningham. Extending the darpa off-line intrusion detection evaluations. In *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX '01. Proceedings*, volume 1, pages 35–45 vol.1, 2001.

[6] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, and Citra Dwi Perkasa. A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert Systems with Applications*, 38(1):306 – 313, 2011. ISSN 0957-4174.

[7] Snehal A Mulay, PR Devale, and GV Garje. Intrusion detection system using support vector machine and decision tree. *International Journal of Computer Applications*, 3(3):40–43, 2010.

[8] Taeshik Shon and Jongsub Moon. A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18): 3799 – 3821, 2007. ISSN 0020-0255.