Chapter 3
ODARM: An Outlier Detection-Based Alert Reduction Model

Fu Xiao
Nanjing University, P. R. China

Xie Li
Nanjing University, P. R. China

ABSTRACT

Intrusion Detection Systems (IDSs) are widely deployed with increasing of unauthorized activities and attacks. However they often overload security managers by triggering thousands of alerts per day. And up to 99% of these alerts are false positives (i.e. alerts that are triggered incorrectly by benign events). This makes it extremely difficult for managers to correctly analyze security state and react to attacks. In this chapter the authors describe a novel system for reducing false positives in intrusion detection, which is called ODARM (an Outlier Detection-Based Alert Reduction Model). Their model based on a new data mining technique, outlier detection that needs no labeled training data, no domain knowledge and little human assistance. The main idea of their method is using frequent attribute values mined from historical alerts as the features of false positives, and then filtering false alerts by the score calculated based on these features. In order to filter alerts in real time, they also design a two-phrase framework that consists of the learning phrase and the online filtering phrase. Now they have finished the prototype implementation of our model. And through the experiments on DARPA 2000, they have proved that their model can effectively reduce false positives in IDS alerts. And on real-world dataset, their model has even higher reduction rate.

INTRODUCTION

With increasing of unauthorized activities and attacks, intrusion detection systems (IDS) have been widely deployed during the last decade, and their value as security components has been demonstrated. But this technique still has some problems. For example, practitioners and researchers have observed...
that IDS can easily trigger thousands of alerts per day, up to 99% of which are false positives (i.e. alerts that are triggered incorrectly by benign events) (Julisch & Dacier, 2002). This flood of mostly false alerts has made it very difficult for managers to analyze security state. Moreover the reactions to dangerous attacks are often delayed, because the alerts for them are hidden in huge amounts of trivial ones and are often neglected. So how to reduce false positives is an important problem needing researchers pay more attentions to.

Now one popular solution to this problem is using certain algorithms (e.g. machine learning or statistical algorithms) to identify true alerts and filter false ones from raw data. Some related methods have been proposed, such as classification-based method, root cause analysis-based method and so on, but most of these methods have three limitations. Firstly, they usually need a lot of labeled training data or domain knowledge to build their alert reduction model. However these data are often difficult to obtain. Secondly, most of them are off-line model which will delay the reaction to attacks. Thirdly, most of them can not adapt to new situations. In this chapter we proposed a novel method, which hasn’t above limitations. It is based on a new data mining technique, outlier detection. This technique has been successfully applied in many fields (e.g. fraud detection, weather prediction), but has not been used to reduce false positives.

In order to filter IDS alerts better, we have designed a special outlier detection algorithm for this field, i.e. an improved frequent pattern-based outlier detection algorithm. It assigns each alert an outlier score, which indicates how abnormal the alert is. The score is calculated based on how many frequent attribute values the alert contains. Usually the more frequent patterns an alert has, the higher its score is, and the more likely it is a false positive. In order to filter alerts in real time, we also design a two-phrase framework. In the learning phrase, we build the feature set of false positives and calculate the threshold of true alerts based on this set. Then in the online filtering phrase, we compare the outlier score of each newcome alerts with this threshold so as to determine whether it is false positive or not. And the feature set is automatically updated so as to keep its accuracy. Based on above algorithm and framework, we have built a new alert reduction system named ODARM (an Outlier Detection-Based Alert Reduction Model). And we have validated ODARM by experiments on both DARPA 2000 dataset and real-world data. The results on DARPA 2000 show that when 86% of alerts have been filtered by our model, 100% of true alerts still remain. And on real-world dataset ORARM has even higher reduction rate.

The rest of this chapter is organized as follows. The next section discusses related work. The section “Outlier Detection-Based Alert Reduction” introduces the design of our outlier detection-based alert reduction system. The next section presents our experiments and gives detailed analysis of the final result. The last section concludes the chapter and introduces the future work.

RELATED WORK

Related work towards alert reduction technique, which aims at identifying and reducing IDS alerts, is described in the first section. And then the next section reviews current outlier detection algorithms.

Alert Reduction

There is still not so much work on identifying and reducing IDS alerts. Current methods can be divided into three categories, which are described as follows:
Related Content

A Perturbation Method Based on Singular Value Decomposition and Feature Selection for Privacy Preserving Data Mining

Forecast Analysis for Sales in Large-Scale Retail Trade
Mirco Nanni and Laura Spinsanti (2010). *Data Mining in Public and Private Sectors: Organizational and Government Applications* (pp. 219-244).
[www.igi-global.com/chapter/forecast-analysis-sales-large-scale/44291?camid=4v1a](www.igi-global.com/chapter/forecast-analysis-sales-large-scale/44291?camid=4v1a)

Analytical Processing Over XML and XLink
[www.igi-global.com/article/analytical-processing-over-xml-xlink/61424?camid=4v1a](www.igi-global.com/article/analytical-processing-over-xml-xlink/61424?camid=4v1a)

Mathematical Statistical Examinations on Script Relics
Gábor Hosszú (2014). *Data Mining and Analysis in the Engineering Field* (pp. 142-158).
[www.igi-global.com/chapter/mathematical-statistical-examinations-on-script-relics/109980?camid=4v1a](www.igi-global.com/chapter/mathematical-statistical-examinations-on-script-relics/109980?camid=4v1a)