An Operational Framework for Alert Correlation using a Novel Clustering Approach

Ashara Banu Mohamed¹, Norbik Bashah Idris², Bharanidharan Shanmugum³
Universiti Teknologi Malaysia, UTM, Kuala Lumpur

¹Email: asharabanumphamed@yahoo.com ²Email: norbik@ic.utm.my ³Email: bharani@ic.utm.my

Introduction

There are two main phases in the IDS technology; detection and alert processing phases. Both stages are equally important and crucial in the success of an IDS implementation; however there are many issues in regards to these two areas that are being researched actively. In the view of latest researchers, the primary issue in IDS implementation is handling the enormous number of alerts generated by the IDS sensors. Moreover due to this obtrusive predicament, two other problems have emerged which are the difficulty in processing the alerts in a more accurate way and also the decrease in performance rate in terms of time and memory capacity while processing these alerts. Based on the above problems, the purpose of this research were finalised and decided to construct a holistic solution that is able to firstly reduce the number of alerts to be processed and at the same time produce a high quality attack scenarios that are meaningful to the administrators in a timely manner. To achieve these goals, alerts generated by IDS sensors need to be correlated and organized in an appropriate approach. Thus the significant contribution of this research is to create an integrated operational framework for alert processing that can reduce the amount of alerts to be processed and creates more meaningful attack scenarios that can be analysed.

Methods

The core feature of the proposed framework is the clustering module. The module consists of two components, firstly, it begins with the initial clustering phase and followed by cluster joining phase which uses the clusters generated by the initial stage to create attack scenarios. For the initial clustering process we will use the hash value generated in the data pre-processing level to perform the comparison and grouping off alerts. For the purpose of the pre-processing phase, a feature selection algorithm will be used to pick three main attributes from each alert; destination IP, signature id and time stamp. Each attribute will be represented by numerical values assigned via a designated converter. Nmap will be used to scan the monitored network periodically and convert any new IP address to a numeric value that will be stored and used as new attack occurs. Meanwhile the time stamp captured will be categorized into three different groups, thus the occurrence off all alerts will be grouped into three different time frames. A hash value from the three attributes is calculated mathematically into a unique hash value which is then stored as an index in a table, alerts off the same hash value will be clustered together. The hashing method is used as a search pattern matching algorithm in the medical informatics domain. The comparison feature in the algorithm is spared by the use of a hashing table. In the recent years the hashing technique is used to speed up the comparison of cDNA sequence because it is easier to compare numbers rather than character. Next is to filter clusters that are considered as false positive; this is done by referring to the asset information database and checking against the set up and configuration of the servers under attack. Clusters will be filtered automatically if the alerts are confirmed to be non-hostile or in other words the asset under attack is not vulnerable to the attempt. Meanwhile alert clusters that are considered as malicious will proceed to the next level which is the
scenario clustering component. The scenarios created are based on destination or the assets under attacked and since this is a supervised clustering algorithm, therefore it is able to create only known attack scenarios. However any new attack techniques will be updated manually on ad-hoc basis.

Dataset

For the purpose of this research two datasets will be used; data set from a division established by the Malaysian Government that monitors the government agencies and statutory bodies for any cyber activities. Next is DARPA, a globally recognized and actively used data set by many IDS researches. The proposed solution will process alerts logs produced by SNORT engine; an open source application available under GNU General Public License. Since the government uses SNORT engine to capture alerts therefore the logs gathered from this source would not require additional preparation however DARPA dataset will have to run through the snort engine to produce the required log. Data produced by SNORT engine will undergo a pre-processing phase before it could proceed to the next stage.

Results

The proposed framework will reduce the amount alerts generated through the clustering technique implemented and followed by scenario clustering module that group together alerts to form a scenario of attacks which is more meaningful than processing individual alerts. Using data collected from one of the government agencies, we are able to do a preliminary deduction manually on the output we hope to produce through the proposed framework. For the first six months of the year 2009, the amount of alerts generated were 4,392,696, we believe that we are able to reduce the alerts to 729,080, which is 83.4% less than the original amount, via the initial clustering.

These numbers will be reduced further in the scenario based clusters phase. Output from the initial clusters will be processed and clustered together to create scenarios then grouped according to a formatted list of common scenario clusters categories used by the government agencies. Instead of processing these alerts individually without understanding the pattern of attack, we propose to build scenarios based on the multiple alerts on the particular asset under attack.

Conclusion

The objective of our alert monitoring system is to protect the monitored infrastructure from any malicious activities. Therefore, in order to minimize damages or losses due to security threats, a reliable and robust monitoring operation is very much in demand. The proposed framework will be able to provide the needed capability in a timely manner. There are many monitoring application and devices available, however the proposed framework is believed to be able to reduce the amount of alerts and at the same time provide an accurate diagnosis to help the security analysts to process the alerts.

Keywords IDS, clustering, hashing technique