Alerts correlation by using heterogeneous sources and neural network

1 venous ghanbari 2 Mehdi Bateni

1 Department of Computer Engineering, Sheikhbahaee University, Isfahan, Iran
v.ghanbari@shbu.ac.ir

2 Department of Computer Engineering, Sheikhbahaee University, Isfahan, Iran
bateni@shbu.ac.ir

Abstract – an intrusion is determined as a set of operations trying to put a source's integration, confidentiality and accessibility at risk. By execution of particular rules, an intrusion detection system (IDS) restricts and reviews the user accessibility to computer system. The rules are defined based on expert knowledge derived by experienced responsible individuals who were constructed the attack scenarios. The system identifies all the users' infractions and operates necessarily to stop attacking the database. The problem of interruption identification over the computer security is detected. The intrusion detection system is either a software or hardware which proceeds intrusion detection automatically. The wireless networks are applied by various equipments and therefore can be easily attached comparing than cable systems. There are different ways have been offered for alarm correlation among which the ones using data mining or machine learning are popular. This essay offers a model based on congruence to increase intrusion accuracy over alarms to correlating. In this model, when data or information enters through sensor into the intrusion detection system, they are compared to alarm congruence tree and combined, where correlation is performed. The idea of this model is using allocation of alarm congruence decision tree to divide the properties complex space into simpler sub-spaces, and then assigns each space a category. The results show that this method has high efficiency over heterogeneous DARPA2000 CUP data set than previously offered methods.

Keywords: correlation, intrusion detection, neural network, decision tree

1. Introduction

An interruption is defined as set of operations trying to put a source's integration, confidentiality and accessibility at risk. By execution of particular rules, an intrusion detection system (IDS) restricts and reviews the user accessibility to computer system. The rules are defined based on expert knowledge derived by experienced responsible individuals who were constructed the attack scenarios. The system identifies all the users infractions and operate necessarily to stop attacking the database. The problem of interruption identification over the computer security is detected. The interruption detection problem in computer security is widely evaluated by Idris & Shanmugam (2005) [1]. As already discussed, detecting abuse and abnormality are both methods applied by intrusion detection system. Detecting abuse is capability to recognize the interruption based on known patterns by which IDS assures that activity is suspicious. The users and producers have most updated patterns. To abnormality detection, the system responsible defines a base line or normal case for network traffic loading and typical size of package. Abnormality recognizer reviews network segments to compare their status to normal base line and searching for abnormalities. Abuse detection is an effective way to respond attacks recognized by system. Indeed, recognizing different attacks are impossible to predicate in a
system. Furthermore, such attempt is so complicated and almost time consuming. The abnormality recognition is complementary with abuse detection. One of the problems in abnormality detection is probable wrong alarms generating by system.

The offered works

Though several methods are introduced in correlation but still there is no agreement on that what is the process or who it must to be implemented and evaluated. In particular, the available correlation ways operate only based on few aspects of correlation process such as fusion; the alarms are produced by different correlation systems to respond an attack, or recognition of multi-attacks in different frames of correlation which made other researchers to analysis these frames more precisely. Herrerias et al. [6] offered a lug correlation frame which helps analyst to searching the evidences and defense better the network. The procedure is that by normal distribution and in specific period, it takes positions from network and saves in system database. It then controls behaviorally the data circuit into the network and compares to its database if any abnormal behavior is seen.

Babak et al. [7] has explained alarm reduction by dynamic and comparative filtration, normalization, aggregation and correlation. Beckers et al. [8] have focused on security information management (SIM) and claim that compression, aggregation and correlation components play an important role in IDS logs analysis so that it can be stated that they continued to correlating. In this essay, we mostly review correlation of alarms and data base and using suggested algorithm, they increased alarm detection rate which caused better search in database to accommodate our alarm and as a result, then, correlation. Valeur et al.[1] also suggested a public frame for correlation consists of comprehensive components set. Following them, Yousef et al.[4] represented a developed alarm correlated frame for sensors and heterogeneous sources which is able to determine attack view and decrease fake alarms by increasing alarm reduction and correlation values and reducing the mistake categorization rate percentage. In above essay in some cases, unfortunately, alarm reduction rate, correlation rate and mistake categorization rate are less than 52%, 79% and more than 20%, respectively.

Few suggested methods include a stepped analysis over alarm circuit. For example, the model proposed by Anderson et al.[9] and Valdez et al.[10],[11] shows a correlation process by two steps. The first step accumulates low level events using attack field concept. Second step converts alarms to a meta-alert using a similarity criterion in order to provide a high level security for network. These steps equal to field reconstruction components and focal detection in our model, respectively. Those steps are important but not efficient to reduce fake positives and the respective alert numbers. The mode firstly introduced by Valderz & Skeener [10], were used later by Valders et al.[12] and developed to consider the effects on overall mission by network. Their suggested method was based on a knowledge database which explains the properties related to a protected network in order to prioritized alerts and perform a simple form of potential alert review. The information corresponding to networks capitals are accumulated by using Nmap that only includes information accumulated by these specific means such as IP addresses, installed operating systems and open ports.

Another development is presented recently by Feng et al. [13] using an attack modeling language called CAML to determine an attack's preconditions or post-conditions. The preconditions or post-conditions are defined for individual attacks, and if an alert's post-condition is in accordance with alert's precondition, then alerts are related (correlated). This helps detecting complex fields of attacks without explicitly modeling the complicated scenarios.

Many other researchers also presented the same mechanisms with preconditions and post-conditions to modeling the correlation between attacks. Ning et al. [14,15] suggest a model which determines the casual relations between alerts using defaults and consequences . The linked alert graphs called hyper-alert, are constructed and then, reduced by manipulation graph techniques in same level. The method proposed by Cuppens and Miege [16] also using preconditions and post-conditions. Furthermore, it is associated with various stages such as alert clustering, alert merging and intention detection. At the first couple stages, using a similarity function, the alerts are clustered and mixed. The intention detection stage is considered in models but not implemented. An interesting way to do so is producing correlation regulations automatically. Although
interesting but this technique can produce some fake correlation regulations that in turn, by reducing numbers of alerts and increasing the reports' abstraction level, it introduces correlation of "close" or "similar" alerts as an implicit probability which raises the noise available in alert circuit.

Another example uses preconditions and post-conditions to determine casual relations between alerts is JIGSAW introduced by Templeton and Levitt [17].

The problems of existed methods

One of unsolved problems in correlation system is not introducing a suitable mechanism to decrease the fake alerts. One way to decrease fake alerts is using heterogeneous sources and correlation of the alerts produced by these sources. When we say heterogeneous sources that mean the alerts generated by correlation system including intrusion detection system based on network and the intrusion detection system based on host, operating system log files, firewall and application programs. So, researchers are focused on correlation of the alerts produced by heterogeneous sources. [4]

Suggested method

Due to high input information in correlation system, the duties of system are categorized so that it can process our data efficiently and detect abnormal factors in input data to network. As already explained in previous chapters, sometimes intrusion detection system are based on network's past operation and data which have passed network and performed attack. It can not be said that such intrusion detection system exorcises the attacks perfectly as they are designed based on past attacks because if there is an attack enters into the system in a new form, it can not defend system against such attacks. The correlation system which we suggested monitors behaviorally and at the moment the network's data and sampling it and as already said in chapter 2, there have been too much work and evaluation on this matter. For this reason, due to high sampling volume in correlation system, the system has to be consisted of parts those can evaluate data suitably and precisely. Here high accuracy means using correlated parts of our data which can increase its correct detections by correlating data.

1-1- Suggested model

Our model is as following.

First, using selection techniques the best traits are selected.

Using combination some categorization, the applied categorizations are:

- Support Vector Machine with Gaussian Kernel
- Support Vector Machine with polynomial
- Support Vector Machine with Quadratic Kernel

Decision tree

Fig.(2-3) shows above model. We will explain each procedures of model as following.
2-4 Suggested algorithm code network

```plaintext
global normalization_db, alertname_db

normalize(raw_alert) {
    alert ← new alert
    alert.alertid ← get_unique_id()
    alert.name ←
        get alertname from alertname_db using
        (raw_alert.name, raw_alert.sensortype) as key
    mappings ←
        get all m:mapping from normalization_db
        where m.sensor = raw_alert.sensor
    for each m:mapping in mappings:
        alert_attr ← m.alert_attr
        raw_attr ← m.raw_attr
        alert.alert_attr ← raw_alert.raw_attr

    pass alert to next correlation component
}
```

3-4 Data analysis in suggested method

If using a sensor to collect data in a correlation system, we need to apply a standard format to normalization because our dataset is heterogenous. We use neural method to analyze normalized data. We are able to perform logarithmic order in neural network due to its treelike structure which increases process power of estimation. It can be observed clearly in fi.3-3 that our alerts are classified by using neural network method and by congruence of our tree with other alert trees in analysis data base system, it can link them quickly and prepare system toward attacks. One of these attacks enters data system in form of web access. According what we defined in neural network, if a request is observed as web access then it may asking for shell access and then, transfers it to system; what we see in neural network tree is that the shell access probability will be reviewed by using such tree. So, it can be said that suggested algorithm is more smarter which is able to detect any kind of algorithms by these trees while raises efficiency during in detection of alert correlation factors. Classification algorithms are designed to classify a set's samples based on their properties. There are many classification algorithms which have been developed. We will evaluate some of them at the following section.
Neural network tree: decision tree arranges samples according an up to down structure.

Fig. 2: an example of neural network's decision tree for correlation

In this tree, each sample's label (class) is placed in leafs and properties in middle nodes. Each middle node executes a test to investigate one trait of each sample and some branches are determined based on value of those traits.

Comparison and simulation

As already explained, above model is designed based on neural network by alert congruence to increase intrusion detection's efficiency in system. Also, due to its tree based structure and neural network operation, this model shows higher efficiency during operation. We will show this by tests in this chapter. For this reason, there are few tests which are arranged. First, we use Darpa2000 1999 data set which is heterogeneous. This data set is based on a server logs which is under attack and then, evaluate suggested model by using Darpa2000.

After presenting our method, we use 18285 sample of Darpa2000 data set to compare with other methods. These samples are selected by random; their properties are shown in table (3-4). As it can be observed, this subset includes any kind of attack- Probe, R2L as well as U2R but due to high diversity of Dos attacks, there are selected only 10000 samples of this kind. Also, we have chosen 3000 samples of random normal port.
Above suggested method is a biped reference has used alert congruence tree at first layer and multi-layer Perceptron neural network classification at second layer. This reference has gained Recall, Precision and F-Measure criterions to evaluate its own method. These criterions are studied in the following.

Precision: this relation is determined as below:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

Recall: this relation is same as positive sample classification value which is defined as below:

\[
\text{Recall} = \text{TPrate} = \frac{TP}{TP + FN}
\]

F-measure: this criterion has used to implement accuracy and TPrate factors in a single factor. It is defined as below:

\[
F = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{TPrate}}{\beta^2 \cdot \text{Precision} + \text{TPrate}}
\]

In order to study suggested method, we also have tested our model over this data set and have measured above criterions. Tables (4-4) and (8-4) show the results of suggested method as well as the results of presents methods.

Table (1) : properties of subset which is selected within heterogeneous Darpa2000 data set

<table>
<thead>
<tr>
<th>Connection Type</th>
<th>Training non homogene Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>2655</td>
</tr>
<tr>
<td>Dos</td>
<td>9.000</td>
</tr>
<tr>
<td>Probe</td>
<td>3990</td>
</tr>
<tr>
<td>R2L</td>
<td>1100</td>
</tr>
<tr>
<td>U2R</td>
<td>52</td>
</tr>
</tbody>
</table>
Table (2) : the results for normal samples

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Naïve Bayes</th>
<th>BPNN</th>
<th>FC-ANN</th>
<th>Our Model-Selection</th>
<th>Our Model-Combiner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>91.22</td>
<td>89.22</td>
<td>89.75</td>
<td>91.32</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>99.41</td>
<td>97.70</td>
<td>98.20</td>
<td>99.08</td>
<td>96.81</td>
<td>96.25</td>
</tr>
<tr>
<td><strong>F-value</strong></td>
<td>95.14</td>
<td>93.27</td>
<td>93.79</td>
<td>95.04</td>
<td>98.37</td>
<td>97.64</td>
</tr>
</tbody>
</table>

Table (5-4) : the results for Dos attack

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Naïve Bayes</th>
<th>BPNN</th>
<th>FC-ANN</th>
<th>Our Model-Selection</th>
<th>Our Model-Combiner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>99.84</td>
<td>99.69</td>
<td>99.79</td>
<td>99.91</td>
<td>97.73</td>
<td>98.51</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>99.41</td>
<td>97.81</td>
<td>99.10</td>
<td>98.32</td>
<td>100.0</td>
<td>88.87</td>
</tr>
<tr>
<td><strong>F-value</strong></td>
<td>96.3</td>
<td>98.8</td>
<td>97.68</td>
<td>95.45</td>
<td>98.37</td>
<td>97.64</td>
</tr>
</tbody>
</table>

Table (6-4) : The results over Probe attack

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Naïve Bayes</th>
<th>BPNN</th>
<th>FC-ANN</th>
<th>Our Model-Selection</th>
<th>Our Model-Combiner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>99.11</td>
<td>98.95</td>
<td>99.20</td>
<td>77.41</td>
<td>99.29</td>
<td>79.68</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>88.99</td>
<td>95.37</td>
<td>96.99</td>
<td>97.65</td>
<td>99.85</td>
<td>79.85</td>
</tr>
<tr>
<td><strong>F-value</strong></td>
<td>98.88</td>
<td>97.89</td>
<td>98.44</td>
<td>93.56</td>
<td>99.57</td>
<td>79.77</td>
</tr>
</tbody>
</table>

Table (7-4) : The results of R2L attack

<table>
<thead>
<tr>
<th></th>
<th>Decision Tree</th>
<th>Naïve Bayes</th>
<th>BPNN</th>
<th>FC-ANN</th>
<th>Our Model-Selection</th>
<th>Our Model-Combiner</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
<td>99.78</td>
<td>96.56</td>
<td>97.78</td>
<td>98.68</td>
<td>100.0</td>
<td>90.0</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>80.54</td>
<td>80.35</td>
<td>85.65</td>
<td>78.88</td>
<td>87.07</td>
<td>37.21</td>
</tr>
<tr>
<td><strong>F-value</strong></td>
<td>90.01</td>
<td>91.11</td>
<td>92.10</td>
<td>91.03</td>
<td>92.80</td>
<td>45.89</td>
</tr>
</tbody>
</table>
1-5 Linear regression analysis for suggested method

One of the analysis methods is Regression. In linear regression, variable and dependent $y_i$ is a linear combination of coefficients (parameters); no need to be linear comparing independent variables. For example, below shows a simple regression analysis with N point, independent $x_i$, linear coefficient $\beta_0$ and $\beta_1$.

Right line

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, N$$

In both states, $\epsilon_i$ is error value and footnote `i number ` shows each observation (each couple $x_i$ and $y_i$). The model can be achieved with a set of these points:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i$$

Where $\epsilon_i$ implies remained; $\epsilon_i = y_i - \hat{y}_i$. The common way to result parameters is Least Squares method in which the parameters are obtained by below minimum oriented function:

$$\text{SSE} = \sum_{i=1}^{N} \epsilon_i^2$$

For simple regression, the parameters are equal to:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Where $\bar{x}$ and $\bar{y}$ are average $x$ and $y$.

As you see in below figure, in our suggested method, the blue line is regression criterion line to analyze suggested method through accurate detection regression in intrusion detection system. As we observe, blue regression blue line is located far from criteria dotted line at the beginning of data circuit entrance, then after awhile after algorithm is performed, they are closer to standard line and determine their improvement to detect accurately and increase error in network. We have used regression diagram to show improvement of criteria line.
2-5 Analyzing the efficiency of suggested method comparing theoric state

We examined the efficiency of suggested method based on neural network by placing congruence trees in network. Fig.2-4 shows that in theoric method introduced in an essay [18] in 2013, ignoring heterogeneity of data, it obtains a constant efficiency, but using suggested method and congruent data in 20th stage of algorithm procedure, we obtain theoric state quickly (best state) and this improvement value is caused by using congruence tree in neural network where we can see due to high data, optimization more quickly by increase performance duration.
Conclusion

This chapter performed some tests to examine the efficiency of suggested method. It was showed in first test that to what extent the properties reduction procedures as well as choosing accurate properties can impact model accuracy. It was also showed; the suggested model had no efficient accuracy when traits number was 19. Indeed, by using properties reduction method PCA, model accuracy was increased significantly. Also, above tests showed that using several classifications for each cluster increases efficiency rather than one classification. In second test operated on heterogeneous DARPA2000 Cup dataset, the model efficiency measured toward other produced methods and showed that in our suggested model, how much the congruence of neural tree network can be effective to increase load volume. It was also proved that when above two-phase uses selection to combine neural network trees shows comparative accuracy against Probe, R21 and U2R attacks. This shows that suggested two-layer model can be successful to classify the samples which are less than others. Another matter is studying both suggested models in this essay. There are introduced two other models in this paper whose their difference - their way to combine classification - has been explained in previous chapter. First model has used Selection for combination; it means that only one classification is selected to labeling a sample. But in second model in which Combiner is used, all the classifications are applied to labeling a sample. The results show that in intrusion detection matter, using Selection to combine the congruence trees caused to significant efficiency.
References


