A Model for Global Anomaly Detection in a Grid Environment

Pramod S. Pawar  Srinath Srinivasa
International Institute of Information Technology,
Bangalore, India 560100
sri@iiitb.ac.in

Abstract

Grids being concerned with sharing of diverse resources in a distributed environment induce security concerns. An event that occurs locally on a grid node may actually be part of a larger set of events that are collectively hazardous to the grid. Since a node only sees the local footprint of an event, it cannot know the contribution of this event at a global scale. As a result, several events that appear innocuous at any given node are ratified as safe, even though they may eventually contribute to a hazard at a global scale. To address such problems, we propose a model for anomaly detection in a grid environment by judiciously exchanging local profiles to capture global behavior. The key focus of this model is to detect anomalous behavior, which looks normal locally at any individual grid node, but when observed globally the anomalous behavior is apparent.

1 Introduction

One of the biggest challenges in current day Intrusion Detection Systems (IDS) on grid computing environments is to handle distributed attacks. Several models of distributed attacks exist, like: Distributed Denial of service (DDoS), Race condition exploitation, etc. A commonly used method to detect attacks is anomaly detection. Anomaly detection entails detecting behaviors that are in some way “abnormal” while distinct from “noise” or “novel” behaviors [1]. A common approach to anomaly detection are unsupervised learning strategies, where the underlying assumption is that most behaviors of the system are normal. Anomaly detection thus involves profiling the normal behavior of the system and detecting deviations from the normal behavior in order to flag anomalies.

In a grid computing environment, anomalies can be of two kinds: local anomalies or global anomalies. Local anomalies affect a single node in the grid and are usually detected and handled locally. Global anomalies represent anomalous behavior whose footprint goes beyond any single node and for which distributed algorithms are required for detecting or handling such behaviors. Worm attacks and distributed denial of service (DDoS) attacks are examples of global anomalies.

As the size of the grid grows, global anomaly detection becomes very challenging. Detecting global anomalies require collecting profile information from
several nodes and aggregating them on a continuous basis. This becomes infeasible as the number of nodes in the grid becomes large. Hence global anomaly detection has to resort to event-based approaches where the communication costs of both profiling and detection are manageable.

For flagging global anomalies, this usually means that some node in the grid has to start the process by suspecting that something is wrong. The trigger for this usually is the detection of a local anomaly by a grid node. Based on the nature, severity or other characteristics of the anomalous behavior the grid node suspects that the anomaly is actually on a bigger scale, and flags off a global detection algorithm. Such an approach has been successfully used for worm detection in distributed environments [2].

However, in this paper we focus on a class of anomalies that are global in nature, but which appear locally normal at every node. With the absence of any local trigger, no node can initiate a global anomaly detection process. Some examples of such global anomalies are as follows:

1. Consider different people carrying different liquids into an airplane. Each of these liquids may be safe in isolation, but their combination can be potentially dangerous. Each individual would be passed through security if the global picture is not considered.

2. An ATM machine validates a transaction by comparing the magnetic id of the card along with the PIN provided by the user. However, if a card with a given magnetic id and PIN were to be inserted in two different ATMs separated by thousands of kilometers within a matter of minutes, the behavior is clearly anomalous. However, since machine performs its validation in isolation, such anomalies are unlikely to be detected.

3. Consider the evaluation of answer papers for a nation-wide examination. If each centre has a different evaluator, then similar or identical answer sheets across the centres go undetected since each evaluator performs an integrity check on the answer-sheets in isolation.

Based on the above examples we divide the global anomaly problem into two varieties:

1. Compositional anomalies
2. Correlational anomalies

A compositional anomaly is one where the anomaly results from some form of a composition operation over local events. The first scenario in the list above is an example of a compositional anomaly.

A correlational anomaly is one that is characterized by anomalous correlations between multiple local events results. The second two scenarios in the list above are about the correlational anomaly. The focus of this paper is only on correlational anomalies.

## 2 Related Work

One of the earliest papers by Denning [3] has articulated several techniques that can be adopted for detecting intrusions. This paper proposed a model
of real time intrusion-detection expert system capable of detecting break-ins, penetrations and other forms of computer abuse. The model includes profiles for representing behavior of subjects with respect to objects in terms of metrics and statistical models, and rules for acquiring knowledge about this behavior from audit records and for detecting anomalous behavior.

AAFID [4] is a distributed intrusion detection architecture and system that proposed the use of autonomous agents for doing intrusion detection. The paper describes the AAFID architecture and the existing prototype, as well as some design and implementation experiences.

In 1998 DARPA published an evaluation data for intrusion detection systems and the Master’s thesis by Kenndal [5] describes all the attacks that were the part of this evaluation data. It describes the simulation network used to collect this data. It focuses on different types of attacks that were developed and presents a taxonomy of computer attacks.

There are several efforts to apply statistical techniques and one such effort include application of Chi-square technique [6] for Anomaly detection. In this paper the author identifies a number of variables in a computer and network system, group them into several categories. A two-stage process is used, first to build a normal profile during training and use it to detect anomalous activities during testing.

The need for IDS within grid had been identified by the grid community [7] and Kenny and Coghlan [8] propose a solution for efficiently accessing audit data from grid but there is no mention on how to use the data to identify intrusions. Choon [9] describes a grid-based IDS architecture that consists of agents located at grid nodes responsible for collecting and sending host audit data to storage and analysis but it is a centralized solution. The Grid Intrusion Detection Architecture (GIDA) proposed by Tolba [10] solves the scalability problem by distributing the intrusion detection problem among several analysis server.

Both Choon et al. [9] and Tolba et al. [10] concentrate on the detection of anomalies in the interaction of grid users with resources which result of misuse. But they lack of unauthorized access, exploits detection. Further none of these provide typical host or network attacks. The Fang-Yie et al. [11] search for network denial of service attacks and looks only for network attacks.

The Grid base Intrusion detection system proposed by Alexandre Schulter [12] identifies that the current intrusion detection technology is limited in providing protection against the attacks that may violate the security in grids and determines the requirement to identify them. This paper proposes as distributed IDS architecture and shows how it overcomes the limitations of current technology by integrating the detection of typical host and network with the grid specific attacks and user behavior anomalies. This paper comes with their class of grid intrusions like (a) Unauthorized access, (b) Misuse (c) Grid exploits and (d) Host or network specific attacks.

Slow propagating attacks are difficult to detect which can be hidden under the veil of normal traffic. Dash et al. [2] describe a method for detecting such attacks using distributed probabilistic inferences. The main contribution of this paper is the probabilistic framework that aggregates (local) beliefs to perform network wide inferences. The local detectors(LD) detect order of magnitude slower worms at local nodes and the global detectors (GD) taking the aggregated view of LD's determine if the network as a whole is in anomalous state. Li et
al. [13] also focus on zero-day slow scanning worms. For effective intrusions the method uses the host organization based on the concept of regions and consider dependency among hosts within each region. There are 3 kinds of detector, Local detectors, regional detectors and global detectors. Local detectors which are weak in the detecting capability reside on each end hosts. Regional detectors diagnose potential problems at neighborhood level. The global detectors uses sequential hypothesis testing to determine for any intrusions at the global level. Huang e al. [14] describes the Network anomaly detection using distributed PCA (Principal Component Analysis) technique for data collected and processed over large distributed system. In this paper it considers set of local monitors each of which collect locally observed time series data stream. A central coordinator node does the global collection and makes global decision concerning to network-wide health. The focus here is to detect volume anomalies that refers to unusual traffic load levels that may be caused due to worms, DDoS attacks etc. In all the above mentioned techniques to solve the problem of global detection, a basic assumption is made that atleast one of the local detectors flags an anomaly. Also the global detector takes into account the aggregated view of the local anomaly detected to decide the global anomaly. In our work, we make no such assumption of local anomaly detection.

3 Global Anomaly Detection Model

The proposed system is based on an unsupervised learning model for global anomaly detection with the underlying assumption that most events are normal.

Based on this, the model comprises of the following: (a). building a reference global profile of all event classes in the grid that are to be observed, (b). building the detection global profile from the event instances observed, and (c). determining if there is deviation between the reference global profile and the detection global profile. The system initially works in a profiling mode to build the reference global profile of the grid event class. Once the profiling is completed for the configured amount of time, the system switches into a detection mode. In detection mode, the detection global profile is compared with the reference global profile and upon any deviations an anomaly is flagged.

3.1 Grid node and Grid events

Let

\[ E = \{e_1, e_2, e_3 \ldots e_n\} \]

be the set of event class for the entire grid. An event class is any activity of interest that takes place on one or more nodes of the grid. A user login may be an example of a event class. Other event classes can include unauthorized access, exploits etc.

At each grid node, we record the instances of these event class as function of time called ‘epoch’. An epoch is an arbitrary unit of time which typically has relevance to the event class to be profiled. An epoch could typically be a day or week or month. The epoch in turn is divided into suitable time units, to form
the time intervals from $t_1$ to $t_m$. Thus an epoch 'T' with time intervals $t_1, t_2$ etc can represent a $T = \{t_1, t_2, \ldots, t_m\}$

The event signature or a local profile is the probability of occurrence of an event instance as a function of time over an epoch.

Over a period of time, we get the probability distribution of these events occurring at individual node over an epoch. The probability distribution of these events can be as shown in the Table 1.

<table>
<thead>
<tr>
<th>Event</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$\ldots$</th>
<th>$t_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>$P_{11}$</td>
<td>$P_{12}$</td>
<td>$\ldots$</td>
<td>$P_{1m}$</td>
</tr>
<tr>
<td>$e_2$</td>
<td>$P_{21}$</td>
<td>$P_{22}$</td>
<td>$\ldots$</td>
<td>$P_{2m}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\ddots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$e_n$</td>
<td>$P_{n1}$</td>
<td>$P_{n2}$</td>
<td>$\ldots$</td>
<td>$P_{nm}$</td>
</tr>
</tbody>
</table>

Table 1: Probability distribution of events at single node

Every individual node in the grid builds a local profile for the event as shown Table 1.

Each node in the grid broadcasts the local profile of all the events, to every other node in the grid. This enables a grid node with the knowledge of global event distribution using which a global profile for every event in the grid is built. An alternative to broadcast is to have hash function which maps an event class to one of the node address and the local profiles for this event class are only exchanged with this mapped node. This mapped node does the computation of the event correlation graph, finds the reference global profile and then updates all other nodes with the reference global profile of this particular event class.

Figure 1: Event Correlation graph
3.2 Event Correlation Graph

To build a global profile, we first construct a correlation graph at each node. For each event class a separate correlation graph is constructed on a grid node. The nodes of the correlation graph consist of the grid nodes and the edges of the graph represent a high degree of correlation between the local profiles at these nodes for an event class. An edge is placed between the nodes only if a high positive or negative correlation, above a certain threshold, is obtained.

Consider one such example of correlation graph depicted in Figure 1. The graph represents that there are 5 nodes in the grid. The edges represent the correlation among the nodes. The edge between the nodes $N_1$ and $N_2$ represents high correlation between nodes $N_1$ and $N_2$ for event $e_1$. Another edge between these nodes represents high correlation between $N_1$ and $N_2$ for event $e_2$. Similarly, the other edges on this graph represent high correlation of the events at neighboring systems. The Figure 2 is an alternate representation of Figure 1. Exactly the same graph will be generated at each node of the grid.

3.3 Reference Global Profile for Event Class

We now build the global profile for the event class across the grid. The Event correlation graph for only event class $e_1$ which is derived from Figure 2, shall be as shown in Figure 3.

From the graph of Figure 3, we find the maximal cliques to build the global profile for event class $e_1$. The profile for all the event classes is determined in the same way as done for the event class $e_1$. The global profile for a event class, in terms of their maximal cliques can be represented as $P_e = C_1, C_2, \ldots, C_K$ where $C_1, C_2, \ldots, C_K$ are the maximal cliques obtained for a event class from the grid.
the correlation graph. Thus for the event class \( e_1 \) the global profile \( P_{e_1} = C_1, C_2 \) as shown in Figure 4.

Similarly the global profiles for all the event classes can be obtained from the maximal cliques of the corresponding event class correlation graph. Let \( P = \{P_{e_1}, P_{e_2}, \ldots P_{e_n}\} \) represent the global profile in terms of maximal cliques, for all the event classes.

### 3.4 Global Profile Energy

We associate energy [15] value to each of the global profile which obtained in terms of maximal cliques. This energy is the measure of how well the events in the clique are correlated and quantifies the similarities between the clique node pairs. A clique having ‘n’ nodes can have max of \( n(n-1)/2 \) associations. The energy of a single clique can be represented by \( \phi \) (Number of nodes in clique). The energy model captures the global behavior of the grid event class by calculating the energy of maximal cliques for a single event class.

Thus the energy associated with maximal cliques obtained in Figure 4(a) and 4(b) can given as \( E_{C_1} = \phi(N_1, N_2, N_3) \) and \( E_{C_2} = \phi(N_3, N_4) \)

Thus for event class \( e_1 \) the global profile in terms of energy can be represented by \( PE_{e_1} = E_{C_1}, E_{C_2} \). Let \( PE = \{PE_{e_1}, PE_{e_2}, \ldots, PE_{e_n}\} \) represent the global profile of all the events in terms of their energy associated with the event.

The [15] energy \( (E_C) \) associated with the each maximal clique is directly proportional to the mean \( (\mu) \) of all the edge weights between the nodes of the cliques and inversely proportional to its variance \( (\sigma) \). A high mean of the indicates that the events co-occur together a lot and high variance of the edge-weights indicate that the set has some elements that do not belong this profile.

The energy of a maximal clique belonging to a profile is calculated as
3.5 Detection Global Profile

To build the detection global profile, the detection local profiles are exchanged, within the relevant reference global profile nodes and energy is calculated for this detection global profile in the same way as done for the reference global profile. Since the exchange of local profiles are only with the relevant nodes, it leads to less communication overhead compared to profiling mode where information broadcast is done to exchange the local profiles.

Every time an event instance occurs the detection global profile of the event class is built. To build a detection global profile a sliding window is used. The sliding window may have the same number of epochs as used to build the reference global profile.

3.6 Typical Scenario

Consider \( U = \{U_1, U_2, \ldots, U_n\} \) be set of users for the entire grid. Over a period of time, we profile for the probability distribution of each user for its logins on a node of a grid. Assume the user logins are checked over an epoch of a day and the epoch has intervals of duration one hour. This provides us with a 24 time intervals \( t_1 \rightarrow t_{24} \) for an epoch. At each node of the grid we shall have the matrix build, as shown in Table 2 specifying the local profile of the user login on the node for all the users.
Each node broadcasts the local profile built on its node across the network. At each node a correlation graph is constructed by finding correlation of local profile of user login between neighboring nodes. The nodes of these graph represents the nodes of the grids and the edges of the graph represent a high correlation between the local profiles of these nodes for a specific user. We maintain a threshold and a edge is placed between the node only if the positive or negative correlation is above this threshold.

Consider one such example of correlation graph depicted in Figure 5. The graph represents that there are 5 nodes in the grid. The edges represent the correlation of the user logins on the nodes. The edge between the nodes $N_1$ and $N_2$ represents high correlation for user $U_1$ login on nodes $N_1$ and $N_2$. Another edge between these nodes represents high correlation of user $U_2$ login on nodes $N_1$ and $N_2$. Similarly the other edges on these graph represents high correlation for user logins between neighbouring systems.

The correlation graph obtained in Figure 5 is used to identify a single user global profile across the entire grid with respect to login. The correlation graph for only user $U_1$ would be as depicted in Figure 6.

On the correlation graph for user $U_1$, we find the maximal cliques. The
maximal cliques are as shown in the Figure 7.
The same way we can find the maximal cliques for user $U_2$, $U_3$, $U_4$, ... $U_N$.
For user $U_1$ and $U_2$ the global behavior in terms of energy can be represented as $U_1 = \phi(N_1, N_2, N_3)$, $\phi(N_3, N_1)$ and $U_2 = \phi(N_1, N_2)$, $\phi(N_2, N_3)$, $\phi(N_1, N_4)$. Similarly the global behavior for other users can be captured by all the maximal cliques for a individual user.

3.7 Detection mode
The system would have created all the reference global profile before it switches in the Detection mode. In the detection mode, whenever an event occurs, it creates a Detection global profile for that event. Now this detection global profile is compared with the reference global profile for any deviations by calculating the Euclidean distance. If the difference is large, it shows huge deviation in the two profiles and hence an anomaly has to be flagged.

4 Model Evaluation
The proposed model will initially be cross validated by partitioning the generated synthetic data into subsets such that initial profiling can be done with one set of data and then the detection profiles can be created with the other subsets of the same data. After the cross validation of the model, we will create a module within this model which can generate events with user defined distribution and deterministic in nature. This will module will give a control to define user defined events to generate global profiles for any event class. Considering that the model has already created global reference profiles using randomly generated events, with this module can create events such that the model is suppose
Figure 7: Maximal cliques for user $U_1$ (a) maximal clique of size 3 (b) maximal clique of size 2

to generate anomalies. This module will help us to calculate the detection accuracy the model provides. We will be able to find the degree of false positives as well as false negatives. Further the model will also be evaluated for Scalability and Communication overhead.

5 Future Work

The Entire model will be simulated and initially we shall work with the synthetic data and carry out our experimentation. Later we will also work on real grid data. At present the simulator has been implemented and soon we shall soon start carrying out the experiments.

References


