INTRUSION DETECTION AND ANOMALY DETECTION SYSTEM USING SEQUENTIAL PATTERN MINING

Shruti Karde[1], Mettu Govind Rao[2], Rajesh Bhise[3]

[1]*Computer Engineering Department, MES, PHCET, Rasayni Tal: Khalapur, Raigad, India*
[2]*Computer Engineering Department, MES, PHCET, Rasayni Tal: Khalapur, Raigad, India*
[3]*Computer Engineering Department, MES, PHCET, Rasayni Tal: Khalapur, Raigad, India*

Abstract

Nowadays the security methods from password protected access up to firewalls which are used to secure the data as well as the networks from attackers. Several times these types of security methods are not enough to protect data. We can consider the use of Intrusion Detection Systems (IDS) is the one way to secure the data on critical systems. Most of the research work is going on the effectiveness and exactness of the intrusion detection, but these attempts are for the detection of the intrusions at the operating system and network level only. It is unable to detect the unexpected behavior of systems due to malicious transactions in databases. The method used for spotting any interferes on the information in the form of database known as database intrusion detection. It relies on enlisting the execution of a transaction. After that, if the recognized pattern is aside from those regular patterns actual is considered as an intrusion. But the identified problem with this process is that the accuracy algorithm which is used may not identify entire patterns. This type of challenges can affect in two ways. 1) Missing of the database with regular patterns. 2) The detection process neglects some new patterns. Therefore we proposed sequential data mining method by using new Modified Apriori Algorithm. The algorithm upturns the accurateness and rate of pattern detection by the process. The Apriori algorithm with modifications is used in the proposed model.

Keywords — *Anomaly Detection, Modified Apriori Algorithm, Misuse detection, Sequential Pattern Mining*

----------***----------

1. INTRODUCTION

Data mining has created a great courtesy because of its enhanced features for storing the huge amount of data and a looming requirement for fetching beneficial information from them. Along with this, it consists of the use of intellectual data analysis tools to find undiscovered, legal patterns, and correlation in big data sets. A vast amount of summarizing practices such as fraud detection, marketing, and scientific breakthrough, etc. experience data mining approaches for their efficient working.

At present, research started considering the importance of data mining skills in the computer security mainly in the backbreaking difficulty of intrusion detection. The set of activities that try to mess up the reliability, privacy or convenience of a system is known as Intrusion. Intrusion detection is the process of finding significant measures arising in a computer system and analyzing them for the possible presence of an intrusion. This system is considered as a following method of security when all the prevention techniques are compromised, and intrusions become a part of the system. Two types of known attacks are there namely inside and outside. Within inside attack, the intruder has all the privileges to use the application or the system so that it does malicious activities. Within outside attack, the intruder does not have proper privileges to access the system. Detecting inside attack is typically more challenging compared to outside attack.

2. LITERATURE SURVEY

2.1 Existing System

Currently, used intrusion detection method is based on signature-based method [9]. In detail awareness about the signature of formerly known attacks is needed for this method. The overseen events corresponding with the signature are used to detect intrusion. The various inventory databases are extracting the features and then comparing them with a set of attack signature provided by the human expert for intrusion detection. The user updates the signature database manually for each new type of intrusion that is detected. The boundaries with such method are that they cannot detect the novel attack that has a no previous signature. It also wants more training data for the associated technique, and there are many chances that they can produce many false alarms. Just because of all the problems mentioned above ultimately causes more demand for intrusion detection techniques which are based on data mining. Anomaly and Misuse detection are the two types of intrusion detection technique.

2.2 Data Mining Methods for Intrusion Detection

1. Misuse detection or Signature based method

The majority of intrusion detection systems (IDS) are signature-based. They operate in like manner as a virus scanner, by examining for a well-known identity or signature for each particular intrusion happening. Signature-based IDS is only well-nigh its database of stored signatures.

2. Anomaly or Profile based method

The anomaly detection inclines on an outline or profile of ordinary user actions. It inspected the user's current session and compared it with the profile representing his standard action. All the actual attacks can be detected using this detection method.

3. Dependency mining or Association rule method

Association indicates the relationship between the items in a transaction. Dependency mining is the technique, in which one item has modified another item refer with this also modify. The data dependency refers to the access relationships among data items. These data dependencies are generated in the form of classification rules.

Table1. Comparison between Misuse and Anomaly Detection

Misuse Detection/Signature Based Detection	Anomaly Detection/Profile Based Detection
Advantages	**Advantages**
1. High detection rate and accuracy for known behaviors or patterns 2. Simplest and useful method 3. Low false alarm rate	1. Accuracy for unknown behaviors 2. Missing pattern rate is low 3. Detect novel and unexpected vulnerabilities
Disadvantages	**Disadvantages**
1. It can detects only known attacks 2. Needs usual updates of rules 3. No segregation between attack endeavor and successful attack 4. Missing pattern rate is high	1. Needs to be more trained or else increases false alarm rate. 2. Low detection rate and high false alarm rate 3. It can detect new attacks because intrusion detection is based on latest updates.

3. DIFFERENT DATA MINING TECHNIQUES USED FOR ANOMALY DETECTION

There are some data mining algorithms which are used for intrusion detection system. These algorithms are listed as follows:

1) Association Rule
   This is used to find out association between database attributes and their values. It is pattern identification technique. Minimum support and confidence values are required for association rule mining. [1]
2) Classification
   It is the method used for learning the different functions that calibrate data objects to subsets of the corresponding class. Classification finds the good mapping method that can predict the class for unknown data [1][10].
3) Clustering Techniques
   Clustering is the process to create the different groups of datasets according to similarity between them within asingle group or within equivalent classes. [1][2][11]
4) Decision Tree
   It is the tree which Decision tree at first builds a tree with classification. Each node of the tree denotes a binary predicate on one attribute. One branch of the tree denotes the positive and another branch denotes negative instances of predicates . It does not require prior knowledge of domain and also handles high dimensional data[11][4]
5) Genetic Algorithm
   This algorithm is used for finding the optimal solution to a particular problem by placing the rules .This algorithm set the rules arbitrarily [11]
6) K- Nearest Neighbor
   By using bulk quantity of votes from the neighbors the object classification is achieved. It is assigned to it most k nearest neighbors. If k=1 then it is assigned to its in close proximity neighbor. [2]
7) Support Vector Machine
   Support vector machine algorithm is associated with supervised learning methods. It is use is for classification prediction. [4]
8) Neural Network
   It is an adaptive system. It changes its structure according to the internal-external information which flows throughout the network. It identifies the complex nonlinear relationship between variables. [11]
9) Bayesian Methods
   It is based on rules and uses probabilities of sample class and observations. It try to find out conditional probabilities of classes by giving simplify the computation and generates great accuracy. [4]
10) Fuzzy Logic
    It is used for both anomalies as well as misuse detection. It makes use of linguistic variables and allows inexact inputs. It also gives permission for fuzzy thresholds. [11]

4. PROBLEM STATEMENT

If any acknowledged pattern or behavior apart from those regular patterns which can be of highly probable is found at the time of detection of any offence on the information level in the database is being considered as an intrusion. This process intended to place the typical executions done by a transaction. One main problem of the process is the accurateness of process of identifying frequently occurring patterns within the database. All the patterns are not detected using existing algorithms and because of this, a data mining approach comes in the picture for the detecting malicious transactions in a system. This system maintains the data dependencies between data items in the database.

Our approach is to design a system to mine recorded, or stored log of the authentic transaction performed with database and based on security rules generate the signature for legal transactions. The malicious transactions are those who are not yielding to a signature of a legal transaction. The security techniques presented with DBMS is not sufficient to identify intrusive movements. Unauthorized users can access important data by executing the malicious transaction. The intrusive detection system can be used to improve security measures. So, to efficiently detect the malicious transactions carrying some data dependencies which exist in the database we have to use proposed system.

5. PROPOSED SYSTEM

The proposed system is to be presented to recognize any new activity that can say as an intrusion on a transactional

database. The algorithm for sequential data mining with some added features is used for this purpose. The aim is to develop a system which reduces the processing time, the database scan numbers and detection of entire probable behaviors in the database.

5.1 Modified Apriori Algorithm

To increase the accuracy of the existing Apriori algorithm, the new modified algorithm is introduced. This modified algorithm improves the performance with respect to accuracy by adding some modified conditions into it. This modification basically in two areas like the process for scanning the transaction and producing candidate sets. At the initial stage the scan of the transaction to produce the support values which can be L-1 itemset and can be applied to transactions. Then the candidate patterns should be generated in either direction not only one so that all the probable patterns can be analyzed along with Candidate set Ck. For example, suppose there are patterns like (p)(q) and these can be used as seed patterns for next candidate set then both the patterns need to be considered as candidates (p,q) and (q.p).

5.2 Pseudo code for Modified Apriori Algorithm Proposed

Step1: Initialize Ck: Candidate itemset of size k
Lk: frequent itemset of size k
SWPk: Swap frequent itemset of size k

Step 2: **L1** = {frequent items};
SWP1 = {Swap frequent items}

Step 3: for (k = 1; Lk !=ϕ; k++) do begin
Ck+1 = candidates generated from Lk;
for each transaction t in the database do increment the count of all candidates in Ck+1 that are contained in t
Lk+1 = candidates in Ck+1 with min_support
Lk=SWPk;
End Loop;

Step 4: Exit

6. IMPLEMENTATIONS

The introduced model consists of the hierarchical structure of client and server. The client should communicate with the main server through sub-server only. The introduced model works in three stages. During the first stage, it will try to detect the attacker or intruder by checking the connection between server and client. If the sub-server is not registered on the main server, it gives the message. Then in the Second Stage apply the proposed Enhanced Apriori algorithm on the transaction log of communication. To acquire the strong rules or pattern we applied this algorithm to the transaction from where we scanned the last transaction. Then those all new rules or pattern are compared with the list of suspected patterns of known signature. As transaction gets updated, it gets scan regularly.

The same algorithm is used on this at agreed fixed period. If the sent pattern is found in suspected list, the alert message gets displayed. And supposed if it is not matched then the communicator allowed to be communicating further. By providing the confidence value on the same pattern or rule sent by a client, the association rules are generated as per modified apriori algorithm.

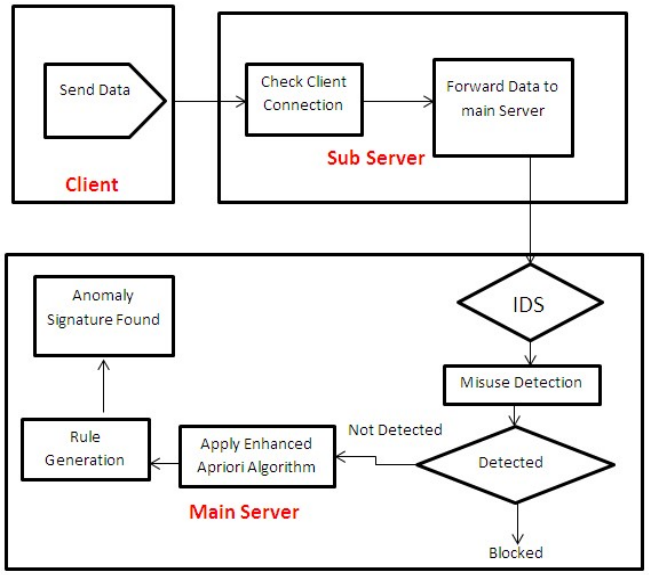

Fig1. System Architecture

If any suspected rule is found, then add it to the suspected list. When entire rules are scanned, the suspected rule is considered as anomaly signature and gets updated in the database. Finally, the action taken that the respective client is considered as suspect and all the activities for the particular suspect will be stopped. If they are intrusions, then a remedy action is to be taken against it. The trusted list of patterns should be updated if they are distinctive behaviors which are not recognized earlier so that they will not be detected again.

7. RESULTS

The implementations mentioned above were applied on the transactional log of data, and the performance was analyzed in terms of time and accurateness of the patterns. The performance was also analyzed by providing different support and confidence values. The predictable thing was that the modification tends to produce more patterns to be managed. About the time a small increase was observed which not distinguishable. Subsequently analyzing the results of the implementations, it was noticed that the increase in the accuracy by using the modified Apriori Algorithm. Along with this the supplementary fact is that it does not burden the CPU, memory as well as time.

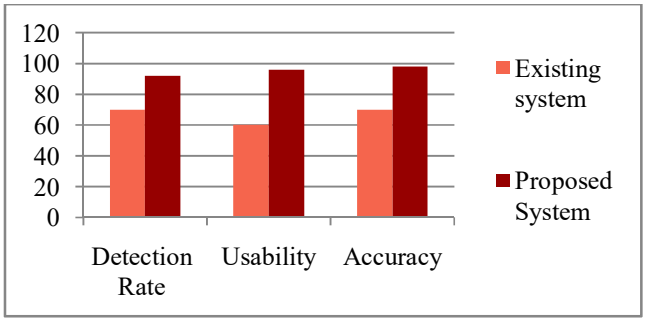

Fig 2. Comparative graph between existing and proposed system

The above graph shows comparative analysis of existing systems and the proposed system. Comparison is done by examining the factors like detection rate of the system, acuuracy of the system and the usability of the system.
Detection Rate is the percentage of attacks that system detects.

The performance of any IDS is evaluated on the basis of its accuracy and usability. A data mining based intrusion detection system founds more complex than traditional one because it requires large set of data .Sometimes it is found that it is a difficult task to handle various kinds of data. Such IDS models needs to be updated. Normally the attacks within the data must be either manually updated for signature detection model or removed for anomaly detection model.

This proposed model helps to overcome thes problem by indenfying the anomolous behaviour of the user.

As the proposed models provides more authentication communicating via intermediate server it shows that we are able to devveloped more secured maechanism than previouly mentioned ones.

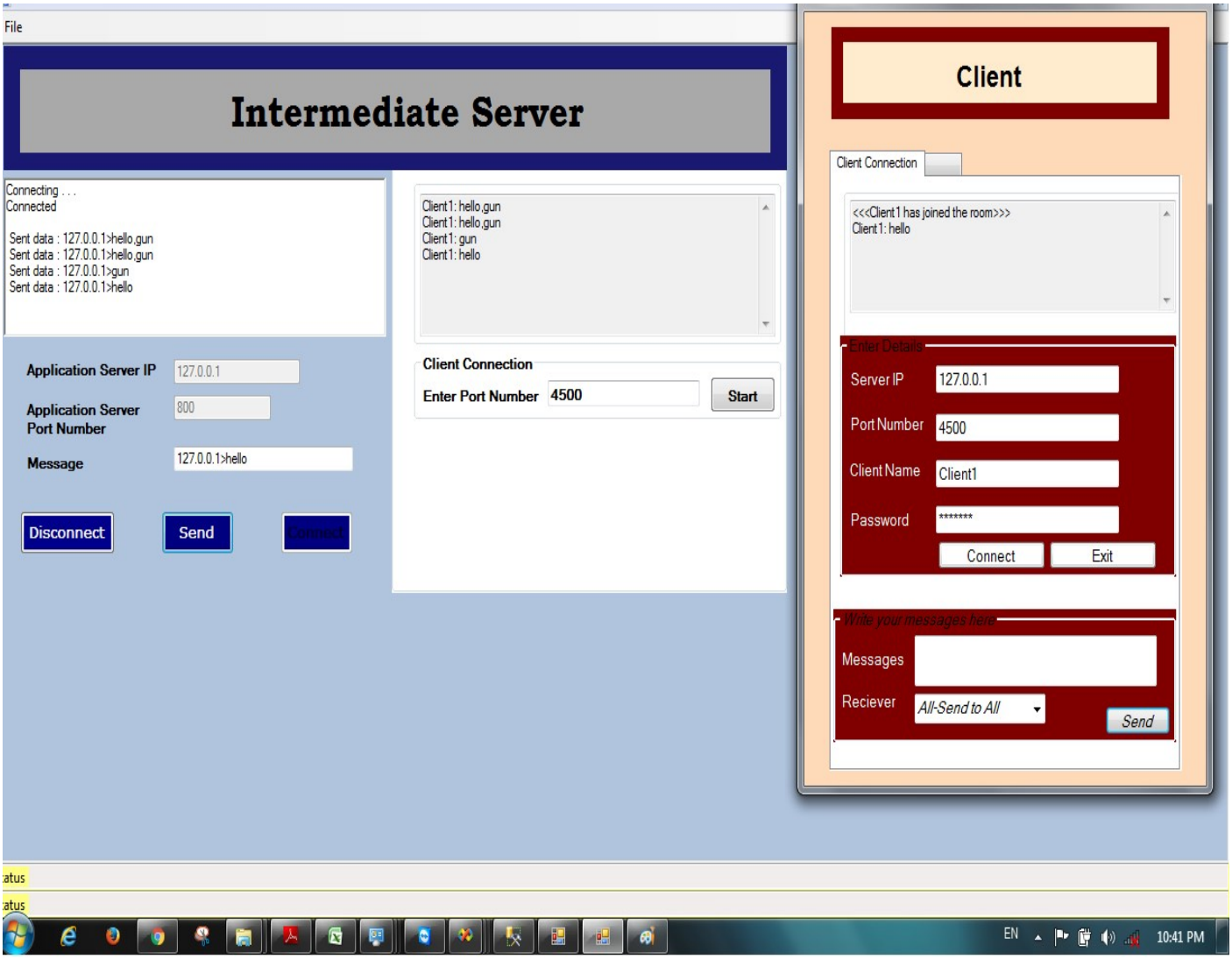

Fig 3. Communication between intermediate server and client

The above figure 3 shows the client should communicate through intermediate server only.

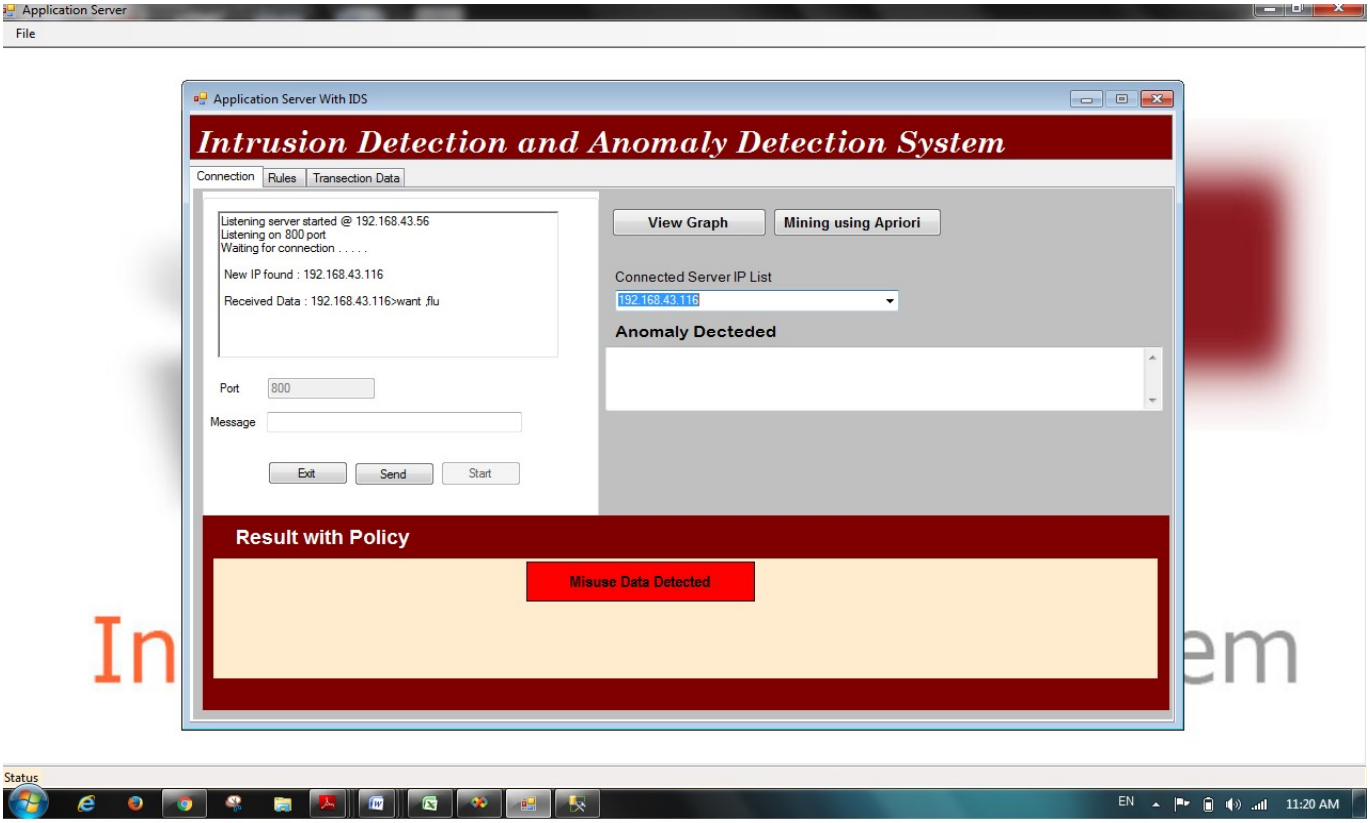

Fig 4. Misuse or Signature Based Detection

The above figure 4 shows the misuse detection within the system.

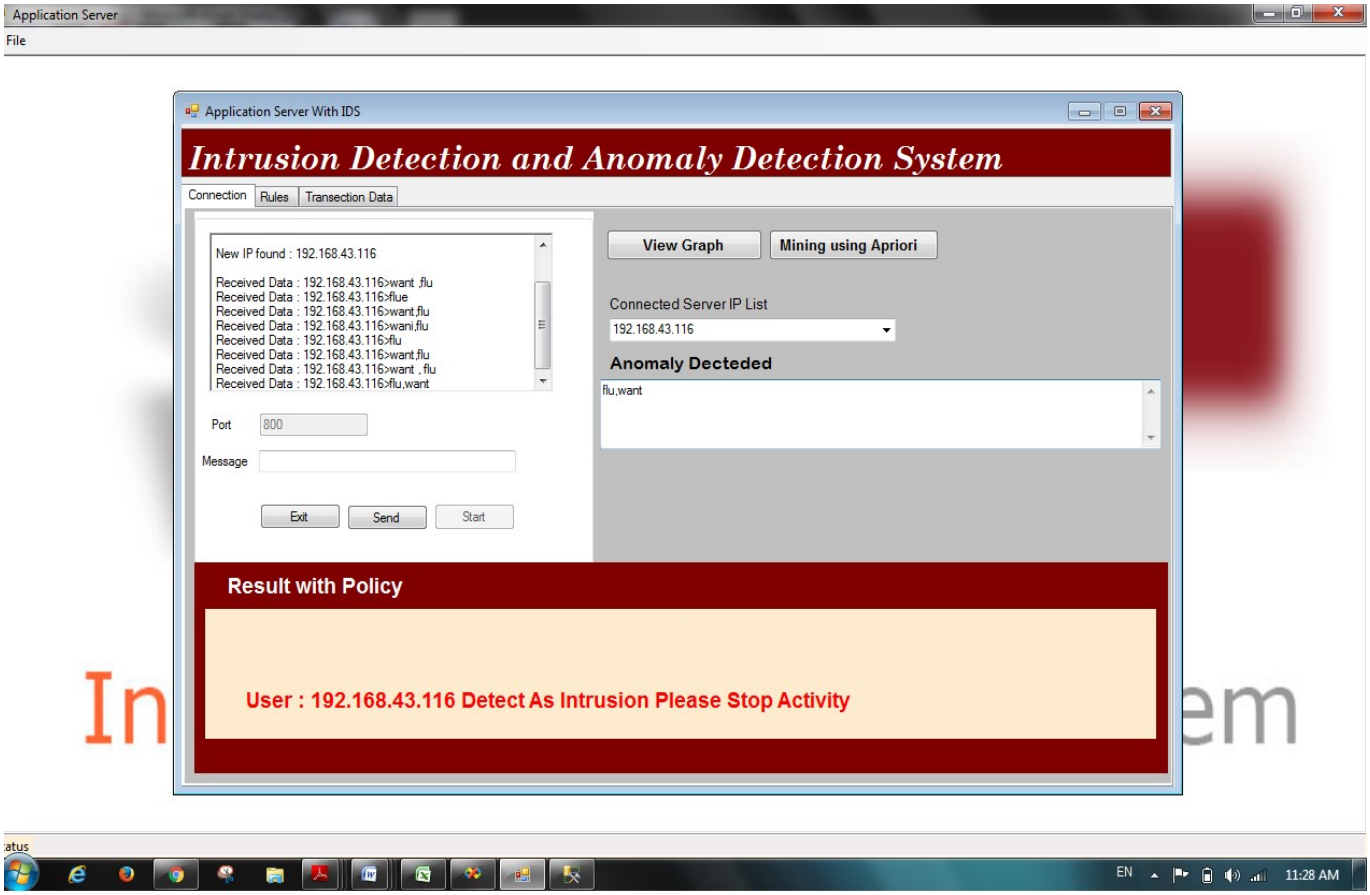

Fig 5. First policy if anomaly found

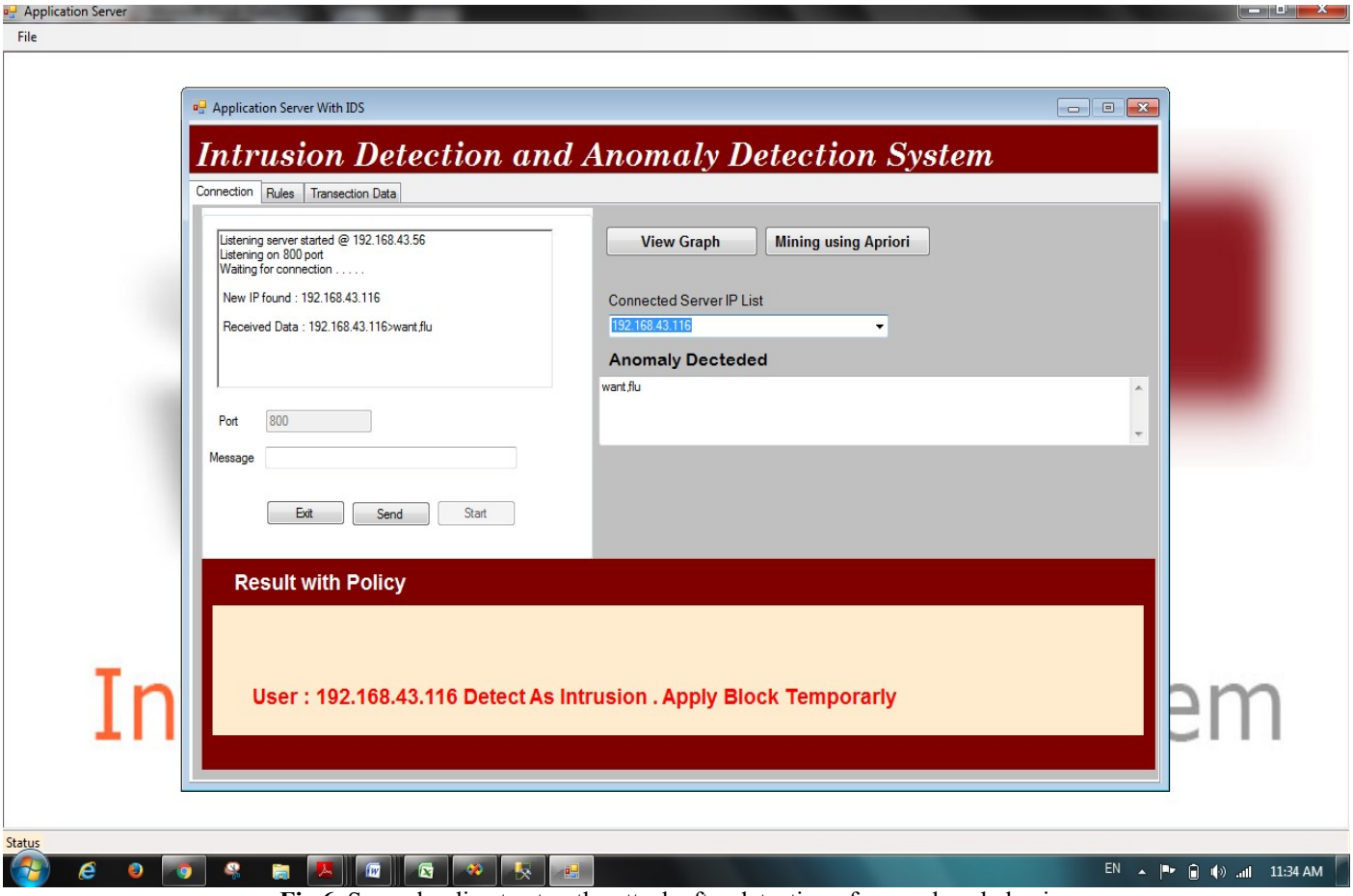

Fig 6. Second policy to stop the attack after detection of anamolous behavior

Figure 5 and 6 shows preventive actions to be taken after the anomaly detection.

8. CONCLUSION

This algorithm demonstrates that the detection rate is better than current algorithm. The revised algorithm consist of altering the method of generating new sequences by assuming these sequences in every direction and therefore leads to identify more frequent patterns that were not identified by the current algorithms.

In this intrusion-detection method, the modified algorithm is used in the database transaction log for identifying entire frequent patterns in the transactional data, and new patterns would be identified based on all previously identified frequent patterns. The primary goal behind this model is to focus generally on enhancing the correctness of the patterns identified in the database transactions. It highlights on identifying patterns which are not identified before by increasing the accuracy of pattern mining. This model increases the rate of identifying intrusions and decreases the rate of false alarms.

REFERENCES

[1] Agrawal, R. and Srikant, R. Mining Sequential Patterns. In Proceedings of the 1995 Int. Conf. Data Engineering, Taipei, Taiwan, March 1995. Pages 3 - 14.

[2] Fawcett, T. and Provost F. Adaptive Fraud Detection. Data Mining and Knowledge Discovery, pages 177-181, 1997

[3] Lee, W. and Stolfo S. Data Mining Approaches for Intrusion Detection. In USENIX Security Symposium, 1998.

[4] C. Y. Chung, M. Gertz and K. Levitt, "DEMIDS: A Misuse Detection System for Database Systems", In Proceedings of the Integrity and Internal Control in Information System, Pages 159-178, 1999.

[5] Lee, V. C.S., Stankovic, J. A., Son, S. H. Intrusion Detection In Real-time Database Systems via Time Signatures. In Proceedings of the Sixth IEEE Real Time Technology and Applications Symposium, 2000.

[6] Yi Hu and Brajendra Panda, A Data Mining Approach for Database Intrusion Detection, ACM Symposium on Applied computing, 2004

[7] Shubhangi S. Suryawanshi et al,, Database Intrusion Detection and Protection System Using Log Mining and Forensic Analysis (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5059-5061

[8] Er. MohitAngurala, Er. Malti Rani, Design and Develop an Intrusion Detection System Using Component Based Software Design International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 4 854 – 860

[9] Indra Jeet Rajput, Deshdeepak Shrivastava, Data Mining Based Database Intrusion Detection System: A Survey, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue4, July-August 2012, pp.1752-1755

[10] Kamini Maheshwar and Divakar Singh "A Review of Data Mining based Intrusion Detection Techniques" in International Journal or Innovation in Engineering & Management (IJAIEM) Feb 2013

[11] Harshna and NavneetKaur "Survey paper on Data Mining techniques of Intrusion Detecion " in International Journal of Science, Engineering and Technology Research (IJSETR), April 2013