ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

A Novel based Intrusion Detection System Using Apriori Algorithm

Babar Prasad H[1], Kalekar Nilesh M[2], Jadhav Nilesh S[3,] Prof. Parineeta Chate[4]

B.E. Student Dept. of Computer Engineering, Bharati Vidyapeeth Lavale, Pune, India[1,2,3]

Assistant Professor, Dept. of Computer Engineering, Bharati Vidyapeeth Lavale, Pune, India[4]

ABSTRACT: Network security is most important issue, as the fast development of the Internet. Network Intrusion Detection System (IDS), as the main security protection technique, is mostly used againstattacks by intruders. Data mining and machine learning technology has been extensively used in network intrusion detection and prevention systems by discovering user behavior patterns from the network traffic data. There are two main techniques of data mining such as sequence rulesand Association rules are used for intrusion detection. The classical Apriori algorithm with bottleneck of frequent item sets mining, we propose a Length-Decreasing Support to detect intrusion based on data mining, which is an advanced Apriori algorithm. Experiment results indicate that the proposed method is sufficient. Length decreasing is improved version of Apriori algorithm.

KEYWORDS: Intrusion Detection; Rule-based; Length-Decreasing Support; Association Rules; Data Mining;

I. INTRODUCTION

With the huge growth of electronics and information technology, computer network becomes important part of our daily lives. However, people use it as a tool for crime, and it has brought great losses and problematic to many organizations and to the world. So preventing these problem acts is on the top of our agenda. An intrusion into a computer system is any activity that violates system integrity, confidentiality, or data accessibility. To meet this challenge, Intrusion Detection System is developed to protect the availability, confidentiality and integrity of critical networked information systems. In the past IDS based on rule detection, intrusion mode is usually pre-defined by the security experts. The advantage of this approach is that the rules can be formulated to detect specific attacks in detail. Therefore, it is guaranteed to detect known attacks and produce few false alarms. However, facing with increasing and changing network data flow, it is unrealistic to discover various intrusion modes in time. New attacks can be detected in time which the system has never seen before as they deviated from normal behaviour. However, current anomaly detection schemes still suffer from a high rate of false alarms. Therefore, currently IDS have many false alarms and redundancy alarms. As a result, understanding and handling IDS alarms become more difficult. Using data mining technology, some useful knowledge can be discovered from network data, behaviour and normal behaviour rule base can be established. Then we can divide abnormal acts from much real-time data. The rule base can be updated automatically and be identified effectively. It effectively reduces the wrong alarm rate and duplication rate.

II. LITERATURE REVIEW

In allusion to the efficiency problem of the actual association rules mining algorithms in the process of disposing massive data, an improved one LRE based on one-dimensional linked list is put forward in this paper. And these tests show that LRE takes priority of Apriori and FP-Growth algorithms in implement efficiency. To resolve the large number of alerts and the high false positive rate issues, it constructs an intrusion alerts analysis system model (IAAS) with LRE applied. Finally, the validity in the aspect of reducing the number of alerts and the false positive rate has been showed by the experiments. [1]

Data Mining is introduced into the Intrusion Detection System, which overcomes the defects of traditional detection technology. The nuclear association rules algorithm applied to the intrusion detection matrix is optimized, which make it possible to reduce the Average-Case Time Complexity, improve the efficiency considerably, and make it easy to process magnanimity data. In this way, attacks will be detected promptly to achieve the goal of intrusion detection.

Finally, the mining of normal connection rules in the knowledge base of intrusion detection matrix will be accomplished. The experiment indicates that the matrix is able to generate new rules after extracting features, and also proves the validity and the feasibility of the IDS [2].

Finding prevalent patterns in large amount of data has been one of the major problems in the area of data mining. Particularly, the problem of finding frequent item set or sequential patterns in very large databases has been studied extensively over the years, and a variety of algorithms have been developed for each problem. The key feature in most of these algorithms is that they use a constant support constraint to control the inherently exponential complexity of these two problems. In general, patterns that contain only a few items will tend to be interesting if they have a high support, whereas long patterns can still be interesting even if their support is relatively small. Ideally, we want to find all the frequent patterns whose support decreases as a function of their length without having to find many uninteresting infrequent short patterns. Developing such algorithms is particularly challenging because the downward closure property of the constant support constraint cannot be used to prune short infrequent patterns. Our experimental evaluations show that both techniques, effectively exploiting the length-decreasing support constraint, are up to two orders of magnitude faster, and their runtime increases gradually as the average length of the input patterns increases. [3]

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration. The steady and amazing progress of computer hardware technology in the past three decades has led to large supplies of powerful and affordable computers, data collection equipment, and storage media. This technology provides a great boost to the database and information industry, and makes a huge number of databases and information repositories available for transaction management, information retrieval, and data analysis. [4]

III. PROPOSED SYSTEM

With the fast development of electronics and information technology, computer network plays a significant role today. However, peoples use it as a tool for crime, and it has brought great losses to many companies and countries. So preventing these acts is on the top of our agenda. An intrusion into a computer system is any activity that violates system integrity, confidentiality, or data accessibility. To recover from this, Intrusion Detection System is being designed to protect the availability, confidentiality and integrity of critical networked information systems. Association rule in mining is an effective data mining method, which extracts the database features of the relationship between the items. And the knowledge will be applied to intrusion detection. Here we propose a Length-Decreasing Support to detect intrusion based on data mining, which is an improved Apriori algorithm. The proposed method is efficient

IV. MODULES

USER

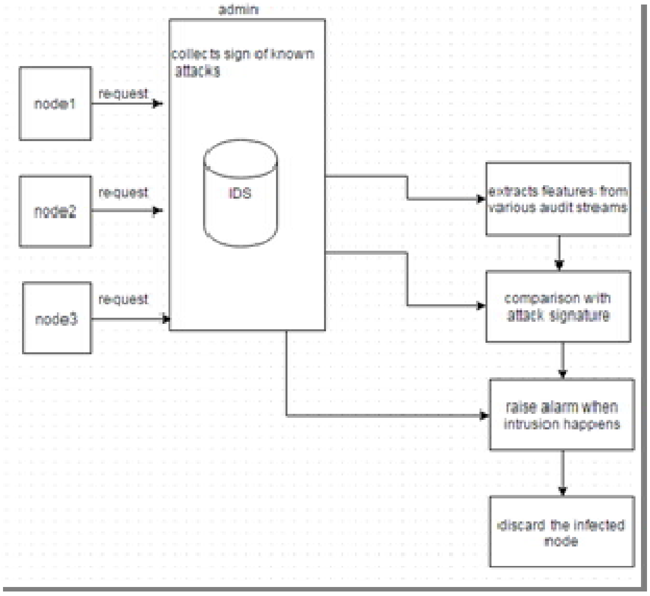
User or node sends a request to server for their requirement and wait for response

ADMIN

Admin discovers behavior pattern from network traffic data. After that he collects sign of known attacks. Input those attacks signature into IDS sign database. Extract features from various audit streams. Compare these features with attack signature. And raise alarms when possible intrusion happens.

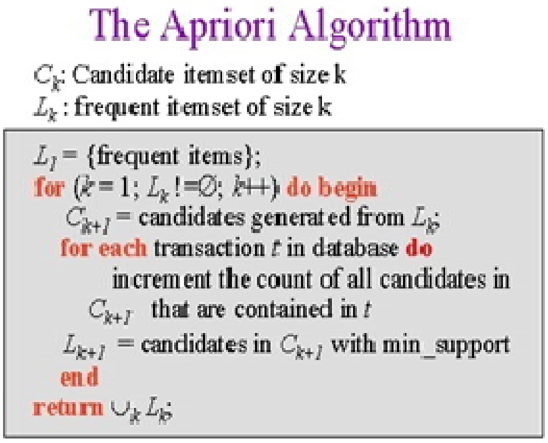
V. SYSTEM ARCHITECTURE

In the below architecture diagram no of nodes are connected. Node sends request to server. Server checks the requesting node. Firstly Admin discovers user's behavior patterns from network traffic data. Then it collects signature of known attacks after the saves them in into IDS. Extract features from various audit Streams. Compare these features with attack signature. If detects any intrusion then raise alarm and reject that Particular node.


admin
collects sign of known attacks
IDS
node1
node2
node3
request
request
request
extracts features from various audit streams
comparison with attack signature
raise alarm when intrusion happens
discard the infected node

VI. APRIORI ALGORITHM

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.



The Apriori Algorithm

C_k: Candidate itemset of size k
L_k : frequent itemset of size k

```
$L_1$ = {frequent items};
for ($k$ = 1; $L_k$ !=∅; $k$++) do begin
    $C_{k+1}$ = candidates generated from $L_k$;
    for each transaction $t$ in database do
        increment the count of all candidates in
        $C_{k+1}$ that are contained in $t$
    $L_{k+1}$ = candidates in $C_{k+1}$ with min_support
    end
return $\cup_k L_k$;
```

VII.SIMULATION AND RESULT

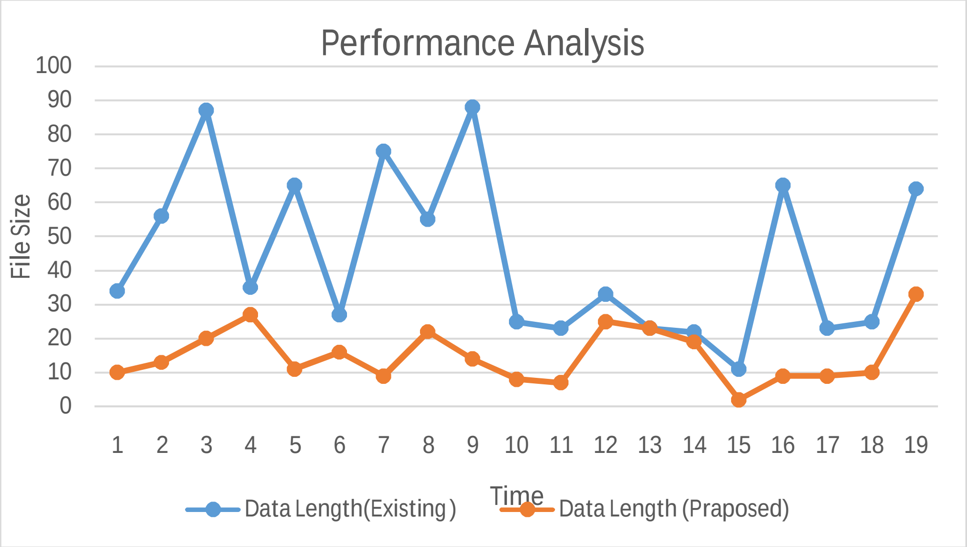
The proposed algorithm is experimented to find out the efficiency of detecting false positives andfalse negatives that are occurred in IDS. The new type of attacks can be detected effectively byalgorithm and then automatically knowledge base can be updated. The experimental result shows,that it was better to find normal packet performance, faster in terms of execution time and alsodetecting the effectiveness and accuracy of false positives and false negatives.

The input parameters are data size (The size of each audit record in database) and DataLength (Processing time of each audit record measured in seconds). Our proposed technique algorithm approach gives better performance in evaluating large size audit data when compared tothe existing techniques.

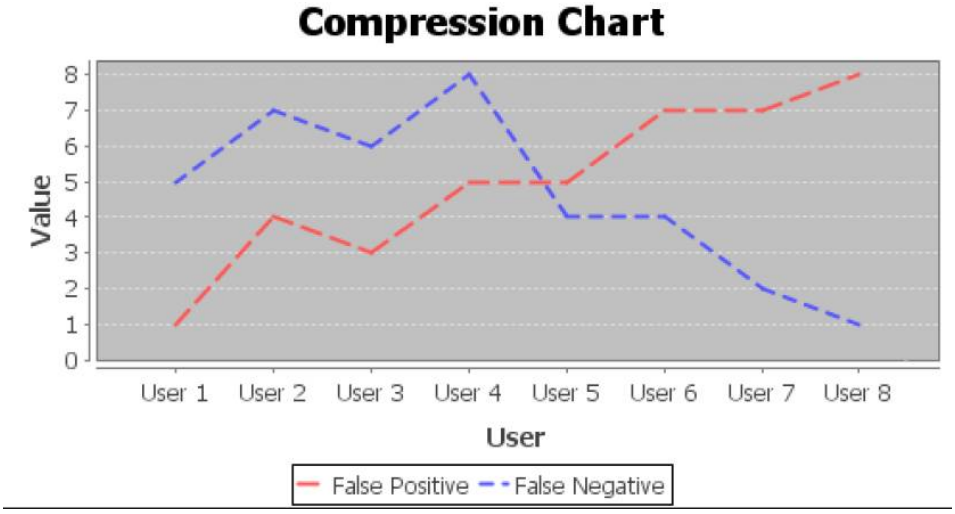
False Positive Rate = Number of false positives / Total number of abnormal labels.
False Negative Rate = Number of false negatives / Total number of normal labels.

Data Size	Data Length (Existing)	Data Length (proposed)
1876	34	10
876	56	13
4526	87	20
5435	35	27
987	65	11
1234	27	16
675	75	9
3456	55	22
4571	88	14
857	25	8
7845	23	7

This Algorithm improves the detecting speed and accuracy as a goal, and proposed a more efficient Apriori Algorithm to abnormal detecting experiment to be based on networkin this approach we show that new type of attack can be detected effectively in the system and the knowledge base can be updated automatically. Finally, the experimental results are compared and analysed. Our proposed algorithm is effectively identifying malicious attacks and also increases the efficiency. This section will show the result based on the inputs as, Data Size: The size of each audit record in database and Data Length: Processing time of each audit record measured in seconds. From the Figure, We can conclude that our proposed technique called Apriori Algorithm approach gives better performance in evaluating large size of audit data when compared to the existing techniques.



The IDS are evaluated on the basis of its accuracy, efficiency and usability. The characteristics used to evaluate the accuracy of the IDS are, Detection Rate: It is the percentage of attacks that system detects and False Positive Rate: It is the percentage of normal data that the system incorrectly determines to be intrusive.



Reasons for Supremacy over other algorithms: -

- Proposed Design will be better than existing to find normal packet performance.
- Proposed Design will be faster than existing in terms of execution time.
- Proposed Design will be smaller than existing and easy to understand and implement.
- It will do not contain complex structure, control flow will be well defined and looping Structure will be minimized. Due to the above facts it will take very less time for execution.

It is a modern Intrusion Detection with Intelligent Analysis that meets the challenges that traditional intrusion detection systems failed to meet:

- need for accuracy
- need for active responses
- need for speed and reliability
- need for usability

VIII. FUTURE WORK

In this system in future increase the security with help of encryption of the data we use the algorithms like AES 256 & Blowfish. We will also use the Wireless Networking.

IX. CONCLUSION

We present association rule-based algorithm with decreasing support. It uses the known patterns to detect the unauthorized behaviour attacks. By adding length decreasing support, it reduces the generation of a short pattern effectively and avoids ignoring the item sets with low support which is interesting in the event. This algorithm has lower consumption of time and lower false alarm rate on the detection rate.

X. ACKNOWLEDGEMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

REFERENCES

[1] Guangjun Song, Zhen long Sun, Xiaoye Li, “The Research of Association Rules Mining and Application in Intrusion Alerts Analysis”, Second International Conference on Innovative Computing, Information and Control (ICICIC 2007), pp.567, 2007.
[2] Zhan Jiuhua, “Intrusion Detection System Based on Data Mining”, First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), pp.402-405, 2008.
[3] Jiawei Han Micheline Kamber, _ Data Mining Concepts andTechniques __Second Edition.
[4] Unil Yun, _An efficient mining of weighted frequent patterns with length decreasing support constraints Knowledge-Based Systems archive Volume 21, Issue 8 (December 2008) table of contents, pp.741-752, 2008.