Review on Anomaly Detection In Cloud Computing

Ragini[1], Mahesh Kumar[2]

[1]M.Tech. Student ,Computer Science & Engineering

Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India

[2]*Associate Professor*, Computer Science & Engineering

Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India

[1]raginipreetpanghal@gmail.com; [2]maheshmalkani@gmail.com

Abstract*— *In research checking of a given sequence of tokens for presence of constituents of some pattern is made to find anomaly in cloud environment. Proposed algorithm would be integrated to cloud environment to find anomalies. Contrast to pattern recognition; match usually has to be exact. Detecting anomaly behaviors is one of most challenging tasks for Information Systems (IS) administrators. Anomaly behavior is defined as any behavior from either inside or outside of organization's information system that deviates from normal; this includes insider attacks as well as any behavior that threatens confidentiality, integrity & availability of organization's information systems. Anomaly detection is applicable in a variety of domains, such as intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, & detecting Eco-system disturbances.

Keywords:-pattern recognition, cloud computing,anomaly detection,clustering.

I.INTRODUCTION

Anomaly detection is identified of items, events or observations that do not conform to an expected pattern or other items in a data set. Usually anomalous items would translate to some kind of difficulties like bank fraud, a structural defect, medical problems or errors in a text. Anomalies are also referred to as outliers, novelties, noise, deviations & exceptions.

In particular, in context of abuse & network intrusion detection, interesting objects are often not rare objects, but unexpected bursts in activity. Pattern did not adhere to generally statistical of an outlier as a rare object, & outlier detection methods would fail on such data, unless it has been aggregated appropriately. Instead, a cluster analysis algorithm may be able to detect micro clusters formed by these patterns.

II.CLOUD COMPUTING

Cloud computing had been describe various scenarios in computing resource is delivered as a service over a network connection. Cloud computing is therefore a type of computing that relies on sharing a pool of physical and/or virtual resources, rather than deploying local or personal hardware & software. It's some synonymous within computing as users are able to tap in to a supply of computing resource than manage equipment need to generate in same such as a consumer tapping throw national electricity supply, in place of running their own generator.

This is a definitions characteristic that differentiates it form other computing models where resource is delivered in blocks usually with fixed capacities and high costs. The cloud computing user normally paid only resource and avoids inefficiencies to expenditure of any unused capacity. Cloud hosting clients get best of both worlds. When there is more demand placed on servers, power can be automatically increased to similar that demand without this need to be give you on a permanent basis.

This is akin to a power access & then only pay to use afterwards. Unlike dedicated servers, cloud servers could be run on a hyper visor.

III. LITERATURE REVIEW

Mohammadjafar Esmaeili (2011) ***Stream Data Mining & Anomaly Detection*** *Detecting anomaly behaviors is one of most challenging tasks for Information Systems (IS) administrators. anomaly behavior is defined as any behavior from either inside or outside of organization's information system that deviates from normal; this includes insider attacks as well as any behavior that threatens confidentiality, integrity & availability of organization's information systems. One of strategies to detect an anomalous behavior is to create a clustering or classification model by utilizing data mining methodologies. models could be generated from previous historical data or it could be based on current data. Although these models could identify normal & abnormal behavior, they couldn't satisfy growing demand for better information security. primary drawback of using these methods are a high rate of false positive; model becomes outdated & there is high demand to maintain models' integrity; & they have low response rate.*

Sushil Kumar (2012) "Anomaly Detection in Network using Data mining Techniques" *As network dramatically extended security considered as major issue in networks. There are many methods to increase network security at moment such as encryption, VPN, firewall etc. but all of these are too static to give an effective protection against attack & counter attack. We use data mining algorithm & apply it to anomaly detection problem. Aim of research to use data mining techniques including classification tree & support vector machines for anomaly detection. result of experiments shows that algorithm C4.5 has greater capability than SVM in detecting network anomaly & false alarm rate by using 1999 KDD cup data. They are using network to detecting attacking is an important uses in network systems, in research we used two data mining techniques namely C4.5 and SVM to detect anomaly in network. Experiments result show, C4.5 algorithm have better result than the SVM in both detection & false alarm rate in our dataset.*

Harshna (2013) "*Data mining techniques of Intrusion Detection"*

In Information Security, intrusion detection is act of detecting actions that attempt to compromise integrity, confidentiality, or availability of a resource. Intrusion detection doesn't, in general, include prevention of intrusions. In research concentrating on data mining techniques that are being used for such purposes. Advantages & disadvantages of these techniques have been discussed in research.

Mohsen Kakavand (2014) *"**A Survey of Anomaly Detection Using Data Mining Methods for Hypertext Transfer Protocol Web Services"***

In contrast to traditional Intrusion Detection Systems (IDSs), data mining anomaly detection methods/techniques has been widely used in domain of network traffic data for intrusion detection & cyber threat. Data mining is widely recognized as popular & important intelligent & automatic tools to assist humans in big data security analysis & anomaly detection over IDSs. We discuss our review in data mining anomaly detection methods of HTTP web services. Today, many online careers & actions including online shopping & banking are running into web-services.

IV. *PATTERN MATCHING*

Most people use pattern matching in some form. Search engines on Web use pattern matching to locate information of interest. Patterns could be specific or quite general, using various wildcards that match multiple endings, words, or strings. Many databases have a similar capability, in which a character such as an asterisk is used as a wildcard.

This concept could be extended in many directions. For example, if you wanted to search for information related to "molecules", you might want to include terms such as "molecule" & "molecular". But rather than typing all possibilities, you could add an asterisk & search for "molecul" to retrieve all three possibilities at once. In Chemical Abstracts Service Molecular Database, researchers could use Markush structures attached to drawings of molecules to retrieve a base molecule, such as benzene, with any string attached, such as OH COOH & Cl. Most bioinformatics databases have similar pattern-matching capabilities. In bioinformatics, flexible pattern matching is called similarity searching. Text mining is an important step of knowledge discovery process. It is used to extract hidden information from not-structured or semi-structured data. They had been developed a crawler for getting pattern matched within a given text by using several algorithms such as:*

1. *Knuth-Morris-Pratt algorithm(KMP)*
2. *Boyer-Moore algorithm*

After a thorough study a conclusion was drawn that Knuth-Morris-Pratt was paramount choice for our work. & we would make comparative analysis between Finite automata algorithm, Knuth-Morris-Pratt algorithm(KMP), Boyer-Moore algorithm.

V. PROPOSED WORK

In this research we have to enhance performance of existing pattern matching algorithm to check anomaly by modifying them. objective of our research is to decrease time consumption during pattern matching.

Our proposed work consists of following steps:

Establishment & configuration of cloud environment.

1. We would create a function to implement KMP pattern matching using MATLAB & test it.

2. In second step we would create Booyer Moore pattern matching using Matlab & test it.

3. study of limitation of KMP & Booyer Moore pattern would be done.

4. Then we would develop a Graphic user interface environment to implement existing KMP & Booyer Moore pattern matching function & get time consumption to perform pattern matching using tic toc function in MATLAB.

5. Then we would develop a new function to implement proposed pattern matching in lesser time.

6. performance chart of existing & new algorithm would be developed.

7. Analysis of performance of proposed pattern matcher would be done using real data set.

8. This proposed algorithm would be integrated to cloud environment to find anomalies.

VI.CONCLUSIONS

In this research checking of a given sequence of tokens for presence of constituents of some pattern is made to find anomaly in cloud environment. Proposed algorithm would be integrated to cloud environment to find anomalies. Contrast to pattern recognition; match usually has to be exact. patterns generally have form of either sequences or tree structures. Uses of pattern matching include outputting locations (if any) of a pattern within a token sequence, to output some component of matched

pattern, & to substitute matching pattern with some other token sequence. Patterns are often described using regular expressions & matched using techniques such as backtracking.

Anomaly detection is applicable in a variety of domains, such as intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, & detecting Eco-system disturbances. It is often used in preprocessing to remove anomalous data from dataset. In supervised learning, removing anomalous data from dataset often results in a statistically significant increase in accuracy.

REFERENCES

1. Mohammadjafar Esmaeili (2011) "**Stream Data Mining & Anomaly Detection"** *International Journal of Computer Applications (0975 – 8887) Volume 34– No.9, November 2011* 38
2. Sushil Kumar (2012) "Anomaly Detection in Network using Data mining Techniques" **International Journal of Emerging Technology & Advanced Engineering ISSN 2250-2459, Volume 2, Issue 5, May 2012**
3. **Harshna (2013) "**Survey paper on Data Mining techniques of Intrusion Detection" ***International Journal of Science, Engineering & Technology Research (IJSETR) Volume 2, Issue 4, April 2013***
4. **Murad A. Rassam (2013)** "**Advancements of Data Anomaly Detection Research in Wireless Sensor Networks" A Survey & Open Issues**
5. **Mohsen Kakavand (2014)** "**A Survey of Anomaly Detection Using Data Mining Methods for Hypertext Transfer Protocol Web Services"**
6. Prashansa Chouhan (2015) "**A Survey: Analysis of Current Approaches in Anomaly Detection"** *International Journal of Computer Applications (0975 – 8887) Volume 111 – No 17, February 2015*
7. Crosbie M. & Spafford G., **"Applying genetic programming to intrusion detection,"** presented at AAAI Fall Symp. Series, AAAI Press, Menlo Park, CA, Tech. Rep. FS-95-01, 1995.
8. Dickerson, J. E. & J. A. Dickerson, **"Fuzzy network profiling for intrusion detection"**, In Proc. of NAFIPS 19th International Conference of North American Fuzzy Information Processing Society, Atlanta, pp. 301306. North American Fuzzy Information Processing Society (NAFIPS), July 2000.
9. Bankovic Z., Stepanovic D., Bojanic S., **"Improving Network Security using Genetic Algorithm Approach"**, Computer & Electrical Engineering, pp. 438-451, 2007.
10. Ektefa M., Memar S., **"Intrusion Detection Using Data Mining Techniques"**,IEEE Trans., 2010.