WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A New Methodology for Storing Consistent indistinct Geo-spatial Data in Big Data

***V.Subhasree*[#1], *S.Yogalakshmi*[#2]**

[#1] M.Sc, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai-606603 .

[#2] Assistant professor, Department of Computer Science, Kamban College of Arts and Science for Women, Tiruvannamalai- 606603

ABSTRACT

Due to the high volume and velocity of big data, it is an effective option to store big data in the cloud, because the cloud has capabilities of storing big data and processing high volume of user access requests. Attribute-Based Encryption (ABE) is a promising technique to ensure the end-to-end security of big data in the cloud. However, the policy updating has always been a challenging issue when ABE is used to construct access control schemes. A trivial implementation is to let data owners retrieve the data and re-encrypt it under the new access policy, and then send it back to the cloud. This method incurs a high communication overhead and heavy computation burden on data owners. In this paper, we propose a novel scheme that enabling efficient access control with dynamic policy updating for big data in the cloud. We focus on developing an outsourced policy updating method for ABE systems. Our method can avoid the transmission of encrypted data and minimize the computation work of data owners, by making use of the previously encrypted data with old access policies. Moreover, we also design policy updating algorithms for different types of access policies. The analysis shows that our scheme is correct, complete, secure and efficient.

KEYWORDS:ABE,Encryption , Clustering

INTRODUCTION:

Virtualized infrastructure in cloud computing has become an attractive target for cyber attackers to launch advanced attacks. This paper proposes a novel big data based security analytics approach to detecting advanced attacks in virtualized infrastructures. Network logs as well as user application logs collected periodically from the guest virtual machines (VMs) are stored in the Hadoop Distributed File System (HDFS). Then, extraction of attack features is performed through graph-based event correlation and Map Reduce parser based identification of potential attack paths. Next, determination of attack presence is performed through two-step machine learning, namely logistic regression is applied to calculate attack's conditional probabilities with respect to the attributes, and belief propagation is applied to calculate the belief in existence of an attack based on them. Experiments are conducted to evaluate the proposed approach using well-known malware as well as in comparison with existing security techniques for virtualized infrastructure. The results show that our proposed approach is effective in detecting attacks with minimal performance overhead.

2.Literature survey

[1]. Konrad Grochowski As the functional complexity of the malicious software increases, their analyses faces new problems. The paper presents these aspects in the context of automatic analyses of Internet threats observed with the HoneyPot technology. The problems were identified based on the experience gained from the analyses of exploits and malware using the dedicated infrastructure deployed in the network of the Institute of Computer Science at Warsaw University of Technology. They are discussed on the background of the real-life case of a recent worm targeting Network Attached Storage (NAS) devices vulnerability. The paper describes the methodology and data analysis supporting systems as well as the concept of general and custom HoneyPots used in the research.

[2]. **Xiaolei Wang**As the dominator of the Smartphone operating system market, consequently android has attracted the attention of s malware authors and researcher alike. The number of types of android malware is increasing rapidly regardless of the considerable number of proposed malware analysis systems. In this paper, by taking advantages of low false-positive rate of misuse detection and the ability of anomaly detection to detect zero-day malware, we propose a novel hybrid detection system based on a new open-source framework CuckooDroid, which enables the use of Cuckoo Sandbox's features to analyze Android malware through dynamic and static analysis. Our proposed system mainly consists of two parts: anomaly detection engine performing abnormal apps detection through dynamic analysis; signature detection engine performing known malware detection and classification with the combination of static and dynamic analysis. We evaluate our system using 5560 malware samples and 6000 benign samples. Experiments show that our anomaly detection engine with dynamic analysis is capable of detecting zero-day malware with a low false negative rate (1.16 %) and acceptable false positive rate (1.30 %); it is worth noting that our signature detection engine with hybrid analysis can accurately classify malware samples with an average positive rate 98.94 %. Considering the intensive computing resources required by the static and dynamic analysis, our

proposed detection system should be deployed off-device, such as in the Cloud. The app store markets and the ordinary users can access our detection system for malware detection through cloud service

[3].**Pushp inderkaur chouhan**While virtualisation can provide many benefits to a networks infrastructure, securing the virtualised environment is a big challenge. The security of a fully virtualised solution is dependent on the security of each of its underlying components, such as the hypervisor, guest operating systems and storage.This paper presents a single security service running on the hypervisor that could potentially work to provide security service to all virtual machines running on the system. This paper presents a hypervisor hosted framework which performs specialised security tasks for all underlying virtual machines to protect against any malicious attacks by passively analysing the network traffic of VMs. This framework has been implemented using Xen Server and has been evaluated by detecting a Zeus Server setup and infected clients, distributed over a number of virtual machines. This framework is capable of detecting and identifying all infected VMs with no false positive or false negative detection.

[4].**A.P.Boedihardjo**Security is becoming a critical part of organizational information systems. Intrusion detection system (IDS) is an important detection that is used as a countermeasure to preserve data integrity and system availability from attacks. Data mining is being used to clean, classify, and examine large amount of network data to correlate common infringement for intrusion detection. The main reason for using data mining techniques for intrusion detection systems is due to the enormous volume of existing and newly appearing network data that require processing. The amount of data accumulated each day by a network is huge. Several data mining techniques such as clustering, classification, and association rules are proving to be useful for gathering different knowledge for intrusion detection. This paper presents the idea of applying data mining techniques to intrusion detection systems to maximize the effectiveness in identifying attacks, thereby helping the users to construct more secure information systems.

[5].Béla GengeModern Networked Critical Infrastructures (NCI), involving cyber and physical systems, are exposed to intelligent cyber attacks targeting the stable operation of these systems. In order to ensure anomaly awareness, the observed data can be used in accordance with data mining techniques to develop Intrusion Detection Systems (IDS) or Anomaly Detection Systems (ADS). There is an increase in the volume of sensor data generated by both cyber and physical sensors, so there is a need to apply Big Data technologies for real-time analysis of large data sets. In this paper, we propose a clustering based approach for detecting cyber attacks that cause anomalies in NCI. Various clustering techniques are explored to choose the most suitable for clustering the time-series data features, thus classifying the states and potential cyber attacks to the physical system. The Hadoop implementation of MapReduce paradigm is used to provide a suitable processing environment for large datasets. A case study on a NCI consisting of multiple gas compressor stations is presented.

3.PROPOSED SYSTEM:

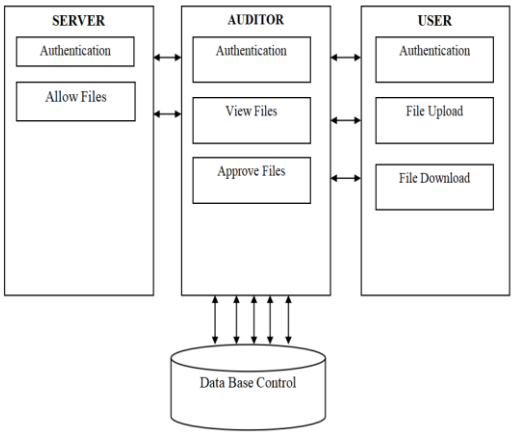
Attribute-Based Encryption (ABE) has emerged as a promising technique to ensure the end-to-end data security in cloud storage system. The grand challenge of outsourcing the policy updating to the cloud is to guarantee the following requirements:

1) **Correctness**: Users who possess sufficient attributes should still be able to decrypt the data encrypted under new access policy by running the original decryption algorithm.

2) **Completeness**: The policy updating method should be able to update any type of access policy.

3) **Security**: The policy updating should not break the security of the access control system or introduce any new security problems.

ARCHITECTURE DIAGRAM



4. METHODOLOGY:

MODULES DESCRIPTION

1.Authentication Module

This module contains all the information about the authenticated user. User without his username and password can't enter into the login if he is only the authenticated user then he can enter to his login.Authentication is the process of verifying the identity of a user by obtaining some sort of credentials and using those credentials to verify the user's identity. If the credentials are valid, the authorization process starts. Authentication process always proceeds to Authorization process. Administrators assume these responsibilities as volunteers who go through a community review process. They are not acting as users. They are never required to use their tools, and must never use them to gain an advantage in a dispute whose need the access for their database in secured way of organization. Administrators should not be confused incoming user registration time and login time

2.Security Policies Module

Attribute-Based Encryption (ABE) has emerged as a promising technique to ensure the end-to-end data security in cloud storage system. It allows data owners to define the access policy and encrypt the data under the policy, such that only users whose attributes satisfying the access policies can decrypt the data. When more and more organizations and enterprises outsource the data into the cloud, the policy updating becomes a significant issue as the data access policies may be dynamically and frequently changed by data owners.

3.Data Auditor

Every authority is independent with each other and is responsible for managing attributes of users in its domain. It also generates a secret/public key pair for each attributes in its domain, and generates a secret key for each user according to their attributes.

4.Server Control Module

The cloud server stores the data of data owners and provides data access service to users. The server is also responsible for updating cipher texts from old access policies to new access policies.

5.Data Consumer

Each user is assigned with a global user identity and can freely get the cipher texts from the server. The user can decrypt the cipher text, only when the its attributes satisfy the access policy defined in the cipher text.

5 .CONCLUSION

The **Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing"** was successfully completed and then tested. Almost all the objectives of the project were meet and thereby a trial run of the project giving sample data gives good result. This system has been developed to reduce the manual work and to reduce the waste of time. This system can adopt any changes in the future also.

Thus this software is developed successfully to fulfill the objectives and satisfies the requirements of the user. And also this is efficient, reliable and secure to use and work

6. RESULTS

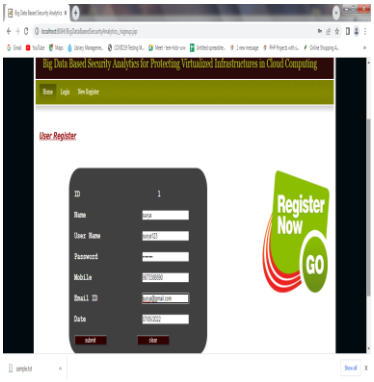


Fig 1: Input Register Acess

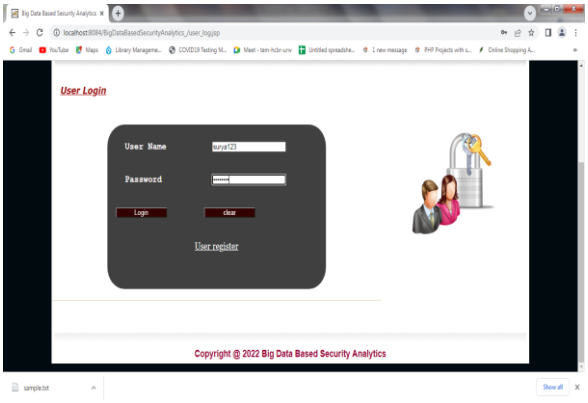


Fig 2: Input Login Acess

REFERENCES:

[1] Goyal, C. Sharma, and N. Joshi, ``An integrated approach of GIS and spatial data mining in big data,\'\' Int. J. Comput. Appl., vol. 169, no. 11, pp. 8887_8975, 2017.

[2] S. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, and T. Cheng, ``Geospatial big data handling theory and methods: A review and research challenges,\'\' ISPRS J. Photogramm. Remote Sens., vol. 115, pp. 119_133, May 2016.

[3] J. P. Mcglothlin, A. Madugula, and I. Stojic, ``The virtual enterprise data warehouse for healthcare,\'\' in Proc. 10th Int. Joint Conf. Biomed. Eng. Syst. Technol., vol. 5, pp. 469_476, Feb. 2017.

[4] C. Zhou et al., ``COVID-19: Challenges to GIS with big data,\'\' Geography Sustainability, vol. 1, no. 1, pp. 77_87, Mar. 2020.

[5] M. Lenzerini, ``Data integration: A theoretical perspective,\'\' in Proc. 21st ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst. (PODS), 2002, pp. 233_246.

[6] O. El Hajjamy, L. Alaoui, and M. Bahaj, ``Semantic integration of heterogeneous classical data sources in ontological data warehouse,\'\' in Proc. Int. Conf. Learn. Optim. Algorithms, Theory Appl., May 2018, pp. 1_8.

[7] Á. Vathy-Fogarassy and T. Hugyák, ``Uniform data access platform for SQL and NoSQL database systems,\'\' Inf. Syst., vol. 69, pp. 93_105, Sep. 2017.

[8] M. Golfarelli and S. Rizzi, ``From star schemas to big data: 20C years of data warehouse research,\'\' in A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years. Cham, Switzerland: Springer, 2018, pp. 93_107.

[9] L. Baldacci, M. Golfarelli, S. Graziani, and S. Rizzi, ``QETL: An approach to on-demand ETL from non-owned data sources,\'\' Data Knowl. Eng., vol. 112, pp. 17_37, Nov. 2017.

[10] P. Yue, X. Guo, M. Zhang, L. Jiang, and X. Zhai, ``Linked data and SDI: The case onWeb geoprocessingwork_ows,\'\' ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 245_257, Apr. 2016.