Analysis of Big Data Security Schemes for Detection and Prevention from Intruder Attacks in Cloud Computing

Amit Chaturvedi, PhD
Govt. Engg. College, Ajmer

Fayaz Ahmad Lone
Ph.D. Scholar,
Mewar Univ.Chittorgarh

ABSTRACT
Big Data Security is a major paradigm for research on cloud networks. Big Data (BD), with their potential to ascertain valued insights for enhanced decision-making process, have recently attracted substantial interest from both academics and practitioners. Big Data Analytics (BDA) is increasingly becoming a trending practice that many organizations are adopting with the purpose of constructing valuable information from BD. The analytics process, including the deployment and use of BDA tools, is seen by organizations as a tool to improve operational efficiency though it has strategic potential, drive new revenue streams and gain competitive advantages over business rivals. In this paper, our aim to present the various application of IDS in cloud computing.

General Terms
Intruder, security, Big Data, Cloud, Attacks, IDS.

Keywords
Multi-tenancy, cloud, Big Data, Security, Intruder, Attack, scalability.

1. INTRODUCTION
Big Data is defined as a collection of huge size of data sets with different types so that it becomes difficult to process by using traditional data processing algorithms and platforms. Recently the number of data provisions has increased, such as social networks, sensor networks, high throughput instruments, satellite and streaming machines and these environments produce huge size of data. Big data used in many applications like health care, education, natural resources, social networking and so on.

In order to secure big data, techniques such as logging, encryption, and honeypot detection must be necessary. In many organizations, the deployment of big data security framework is very attractive and useful. Big data analytics can be used to detect and prevent the malicious intruders and advanced threats [1]. Big data security in the cloud computing is essential due to the following issues such as: 1) To protect and prevent huge size of confidential business, government, or regulatory data from malicious intruders and advanced threats, 2) Lack of awareness and standards about how cloud service providers securely maintaining the huge disk space and erase existing big data, 3) Lack of standards about auditing and reporting of big data in public cloud, 4) Users who does not even work for the organization (malicious intruders), but may have full control and visibility into history of organization data (big data) [4].

Cloud computing has captured significant portion of the competitive market today. Many organizations make use of cloud services. Although cloud computing services is growing and gaining popularity, the fear about the usage of cloud services is still an open issue. Various issues deterring adoption are identified in the literature; one of the major issues is security. Security risks in the area of cloud computing has attracted attention since its beginning. New protocols and tools are always in demand to enhance and assess the security strength of a cloud computing service or service provider.

To analyze and measure a particular service based on its security properties is a challenge. Cloud computing can defined as five attributes such as Massive Scalability, Multi-tenancy (Shared Resources), Elasticity, Pay as You go and Self-Provisioning of resources. Cloud computing enables user to access the remote servers hosted on the internet to store and process the data. Service models of cloud is classified into three types such as SaaS, PaaS, Iaas and different deployment models are classified into Private, Public, and Hybrid. Due to the high availability of cloud to all end users, cloud computing faces more security challenges. These challenges are classified into two broad categories as security issues faced by cloud providers and security issues faced by Customers.

Conventional standalone IDSs are susceptible to cooperative attacks, so they're unsuitable for collaborative environments (such as a cloud computing environment). To defend against this type of attack, collaborative intrusion detection systems (CIDSs) correlate suspicious evidence between different IDSs to improve the intrusion detection efficiency. Unlike conventional standalone IDSs, a CIDS shares traffic information with the IDSs located at a local network's entry points.

We can organize IDSs within a CIDS in a decentralized or hierarchical manner over a large network. These IDSs communicate directly with each other or with a central coordinator, according to the applied mode of organization.

In a decentralized CIDS, each IDS can generate a complete attack diagram of the network by aggregating network information received from other IDSs in the CIDS. Detection of malicious attempts is undertaken locally at each IDS. In a hierarchical CIDS, a coordinator is a central point responsible for information aggregation. The central coordinator, which analyzes the aggregated information, generates a complete attack diagram of the network.

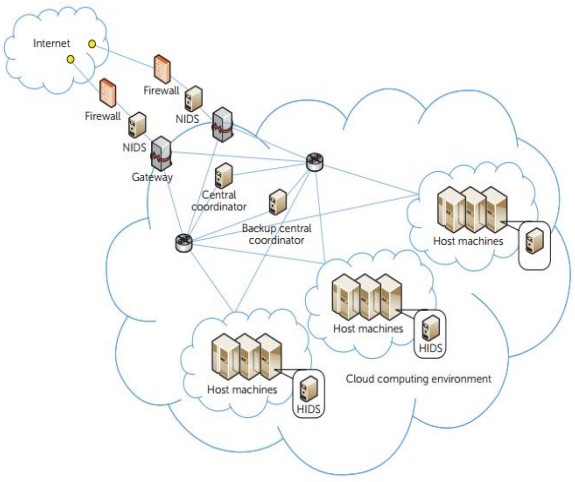

Fig 1: Framework of a collaborative intrusion detection system (CIDS).[1]

2. RELATED WORK

David Gill and II-Yeol Song, discussed the challenges and opportunities for Modeling and Management of Big Data. The term Big Data denotes huge-volume, complex, rapid growing datasets with numerous, autonomous and independent sources. In these new circumstances Big Data bring many attractive opportunities; however, good opportunities are always followed by challenges, such as modelling, new paradigms, novel architectures that require original approaches to address data complexities. The purpose of this special issue on Modeling and Management of Big Data is to discuss research and experience in modelling and to develop as well as deploy systems and techniques to deal with Big Data. A summary of the selected papers is presented, followed by a conceptual modelling proposal for Big Data. Big Data creates new requirements based on complexities in data capture, data storage, data analysis and data visualization. These concerns are discussed in detail in this study and proposals are recommended for specific areas of future research.[6]

M.M. Potel, C A Dhote, D.H. Sharma proposes Homomorphic Encryption for Security of Cloud Data. They state that Cloud computing is a broad and diverse phenomenon. Users are allowed to store large amount of data on cloud storage for future use. The various security issues related to data security, privacy, confidentiality, integrity and authentication needs to be addressed. Most of the cloud service provider stores the data in plaintext format and user need to use their own encryption algorithm to secure their data if required. The data needs to be decrypted whenever it is to be processed. This paper focuses on storing data on the cloud in the encrypted format using fully homomorphic encryption. The data is stored in DynamoDB of Amazon Web Service (AWS) public cloud. User's computation is performed on encrypted data in public cloud. When results are required they can be downloaded on client machine. In this scenario users data is never stored in plaintext on public cloud.[7]

Zhong Wang, Qian Yu, presented the survey and countermeasures of Privacy trust crisis of personal data in China in the era of Big Data. Privacy trust directly affects the personal willingness to share data and thus influences the quality and size of the data, thus affecting the development of big data technology and industry. As China is probably the largest personal data pool and vastest application market of big data, the situation of Chinese privacy trust plays a significant role. Based on the 17 most common data collection scenarios, the following aspects have been observed through 508 questionnaires and interviews of 20 samples.To start with, there is a severe privacy trust crisis in China, both in the field of enterprise services such as online shopping and social networks, etc. and in some public services like medical care and education, etc. Besides, there are also doubts about data collected by the government since individuals refuse to offer personal information or give false information as much as possible. Some people even buy two phone numbers, one is in use, while the other is not carried around or used by them, which is only bought to be offered to data collectors. Secondly, in terms of gender, females have lower trust in enterprises and social associations than males, especially in the fields of social networks and personal consumption.[8]

However, there is no obvious difference in fields of government and public services. Females possess stronger awareness but less skilled in precautions than males. Thirdly, people between the ages of 18 and 50 are more suspicious of data collected by enterprises, while age exerts little obvious influence on the credibility of data collected by the government, social associations and public services. Older people are less aware of precautions than people at other ages. In addition, from the perspective of education background, people with higher degrees possess stronger awareness of precautions and thus lower degree of trust. Therefore, it is suggested that more education on privacy consciousness should be given, and relative laws as well as regulations need improving. Besides, innovation in privacy protection technologies should be encouraged. What is more, we need to reinforce the management of the internet industry and strictly regulate personal data collection of the government.

N. Kshetri analysed Big data's impact on privacy, security and customer welfare and highlight the costs, benefits, and externalities associated with organizations' use of big data. Specifically, it investigates how various inherent characteristics of big data are related to privacy, security and consumer welfare. The relation between characteristics of big data and privacy, security and consumer welfare issues are examined from the stand points of data collection, storing, sharing and accessibility.[9]

S.Akter, S.F.Wamba, A.Gunasekaran, R. Dubey, S. J.Childe discussed that How to improve firm performance using big data analytics capability and business strategy alignment.The recent interest in big data has led many companies to develop big data analytics capability(BDAC) in order to enhance firm performance (FPER). However, BDAC pays offfor some companies but not for others. It appears that very few have achieved a big impact through big data. To address this challenge, this study proposes a BDAC model drawing on the resource-based theory (RBT) and the entanglement view of sociomaterialism. The findings show BDAC as a hierarchical model, which consists of three primary dimensions (i.e.,management, technology, and talent capability) and 11 subdimensions (i.e., planning, investment, coordination, control, connectivity, compatibility, modularity, technology management knowledge, technical knowledge, business knowledge and relational knowledge). The findings from two Delphi studies and 152 online surveys of business analysts in the U.S. confirm the value of the entanglement conceptualization of the higher-order BDAC model and its impact on FPER. The results also illuminate the significant moderating impact of analytics capability–business strategy alignment on the BDAC–FPER relationship.[10]

A. Siddiqa, I.A.T. Hashem, I.Yaqoob, M.Marjani, S. Shamshirband, A. Gani, F.Nasaruddin presented a survey of big data management: Taxonomy and state-of-the-art. The rapid growth of emerging applications and the evolution of cloud computing technologies have significantly enhanced the capability to generate vast amounts of data. Thus, it has become a great challenge in this big data era to manage such voluminous amount of data. The recent advancements in big data techniques and technologies have enabled many enterprises to handle big data efficiently. However, these advances in techniques and technologies have not yet been studied in detail and a comprehensive survey of this domain is still lacking. With focus on big data management, this survey aims to investigate feasible techniques of managing big data by emphasizing on storage, pre-processing, processing and security. Moreover, the critical aspects of these techniques are analyzed by devising a taxonomy in order to identify the problems and proposals made to alleviate these problems. Furthermore, big data management techniques are also summarized. Finally, several future research directions are presented.[11]

R. Shaikh, M. Sasikumar, proposed a Trust Model for Measuring Security Strength of Cloud Computing Service. They state that Cloud computing has become a part of the competitive market today. Various cloud computing service providers are available with their services in the cloud environment. Techniques adopted by various providers to achieve security are of varying nature. To analyze and measure a particular service based on its security properties is a challenge. This paper presents such a measurement by using a trust model. A trust model measures the security strength and computes a trust value. A trust value comprises of various parameters that are necessary dimensions along which security of cloud services can be measured. CSA (Cloud Service Alliance) service challenges are used to assess security of a service and validity of the model. Adequacy of the model is also verified by evaluating trust value for existing cloud services. Trust model acts as a benchmark and ranking service to measure security in a cloud computing environment.[12]

Vennila.S, and Priyandarshini J. Presented a survey of Scalable Privacy Preservation in Big Data and discussed that Cloud computing provides flexible infrastructure and high storage capacity for Big Data applications. The MapReduce framework is most preferable for processing huge volume of unstructured data set in BigData. Increase in data volume leads to flexible and scalable privacy preservation of such dataset over the MapReduce framework is Big Data applications. A survey have been taken for the MapReduce framework based big data privacy preservation in Cloud environment. Existing approaches employ local recording anonymization for privacy preserving where data are processed for analysis, mining and sharing. The proposed work focus on Global recording anonymization for preserving data privacy over BigData using MapReduce on Cloud environment.[13]

B. H. Krishnaa, Dr. S. Kiranb, G. Muralia, R.P.K. Reddy presented the Security Issues In Service Model Of Cloud Computing Environment. Cloud computing is becoming increasingly fashionable in distributed computing environment. Processing and Data storage use cloud environment is becoming a movement universal. Software as a Service (SaaS) has on many business applications as well as in our day to day life, we can simply say that this disruptive technology. Cloud computing can be seen since Internet-based computing, in which shared resources, software, and information are made available to devices on demand. It allows resources towards leveraged on per-use basis. It diminishes cost and complexity of service providers by means of assets and operational costs. It allows users to access applications tenuously. On behalf of user, this construct directs cloud service provider to feel software updates and cost of servers etc. For both, cloud providers and consumers; availability, integrity, authenticity, confidentiality, and privacy are important concern. Infrastructure as a Service (IaaS) serves as base layer for many other release models and Platform-as-a-Service (PaaS) clouds. Security of PaaS clouds is considered from multiple perspective including access control, service continuity and privacy while protecting together the service provider and the user. Security problems of PaaS clouds are explored and classified. In this paper we are going to some major security issues of current cloud computing environments.[14]

N. Khana, A. Al-Yasirib proposed Identifying Cloud Security Threats to Strengthen Cloud Computing Adoption Framework. Cloud Computing allows firms to outsource their entire information technology (IT) process, allowing them to concentrate more on their core business to enhance their productivity and innovation in offering services to customers. It allows businesses to cut down heavy cost incurred over IT infrastructure without losing focus on customer needs. However, to a certain limit adopting Cloud computing has struggled to grow among many established and growing organizations due to several security and privacy related issues. Throughout the course of this study several interviews were conducted, with cloud developers and security experts, and the literature was reviewed. This study enabled us to understand, current and future, security and privacy challenges with cloud computing. The outcome of this study led to identification of total 18, current and future, security issues affecting several attributes of cloud computing.[15]

S.A. Hussain, M.Fatima, A. Saeed, I. Raza, R. K. Shahzad presented a study on "Multilevel classification of security concerns in cloud computing". According to this study Threats jeopardize some basic security requirements in a cloud. These threats generally constitute privacy breach, data leakage and unauthorized data access at different cloud layers. This paper presents a novel multilevel classification model of different security attacks across different cloud services at each layer. It also identifies attack types and risk levels associated with different cloud services at these layers. The risks are ranked as low, medium and high. The intensity of these risk levels depends upon the position of cloud layers. The attacks get more severe for lower layers where infrastructure and platform are involved. The intensity of these risk levels is also associated with security requirements of data encryption, multi-tenancy, data privacy, authentication and authorization for different cloud services. The multilevel classification model leads to the provision of dynamic security contract for each cloud layer that dynamically decides about security requirements for cloud consumer and provider.[16]

C. Saadia, H. Chaouib discussed Cloud Computing Security Using IDS-AM-Clust, Honeyd, Honeywall and Honeycomb. The cloud computing security has become a basic necessity. It acquires knowledge about vulnerabilities, attacks, activities of attackers and tools to secure it. This work proposes new cloud infrastructure architecture, which combines IDS based on mobile agent sand using three types of honeypots in order to detect attacks, to study the behavior of attackers, increase the added value of Honeypot and IDS based mobile agents, solve

systems limitations intrusion detection, improve knowledge bases IDS thus increase the detection rate in our cloud environment. [17]

3. OUTCOME OF THE ANALYSIS

Cloud computing is a service-oriented architecture (SOA). The security vulnerabilities inherited from the underlying technologies (that is, virtualization, IP, APIs, and data centre) prevents organizations from adopting the cloud in many critical business applications. These vulnerabilities leave loopholes, allowing cyber intruders to exploit cloud computing services and threatening the security and privacy of big data. Various security schemes, such as encryption, authentication, access control, firewalls, intrusion detection system (IDSs), and data leak prevention systems (DLPSs), address these security issues. In this complex computing environment, however, no single scheme fits all cases. These schemes should thus be integrated and cooperate to provide a comprehensive line of defence.

IDSs aim to provide a layer of defence against malicious uses of computing systems by sensing attacks and alerting users. Because it's impossible to prevent all cyber attacks, IDSs have become essential to securing cloud computing environments.

IDSs are commonly categorized by the type of data source involved in detection. Host-based IDSs (HIDSs) detect malicious events on host machines. They handle insider attacks (which attempt to gain unauthorized privileges) and user-to-root attacks (which attempt to gain root privileges to VMs or the host). Network-based IDSs (NIDSs) monitor and flag traffic carrying malicious contents or presenting malicious patterns. This type of IDS can detect direct and indirect flooding attacks, port-scanning attacks, and so on.

Although to some extent, DLPSs can be considered a type of IDS, they're more tailored to data security. However, it's difficult to completely guarantee data security using DLPSs alone. Attackers who gain control of the host machines can modify the DLPS settings, thereby completely disclosing data to those attackers. Moreover, even though firewalls can block unwanted network traffic packets according to a predefined rule set, they can't detect sophisticated intrusive attempts such as flooding and insider attacks. IDSs, DLPSs, and firewalls are therefore not interchangeable security schemes but collaborative ones.

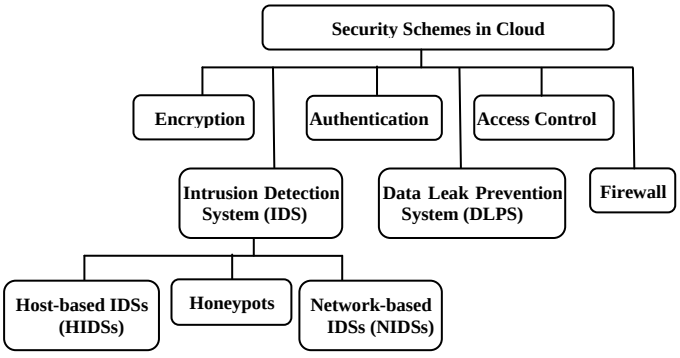

Fig 2: Hierarchy of various security schemes applied with cloud networks

4. CONCLUSION

In cloud computing, IDS is very effective to secure the crucial or important data. A router receives packets from the network and sends these packets to the firewall, which filter the packets and let passing only the authenticated packets, IDS then checks the incoming and outgoing packets and the system files, and give alerts to the administrators. Intruders mostly hit the network from remote site.

The honeypot system is designed to lure attackers. Any attacks against the honeypot are made to seem successful, giving enough time to administrator to mobilize, log and possibly track and apprehend the attacker without exposing the production systems. Host based Intrusion Detection System (HIDS) basically examine specific host-based actions, such as what applications are being used, what files are being accessed and what information resides in the kernel logs of the Host Servers. Network based Intrusion Detection System (NIDS) basically analyse the flow of information between computers, i.e., network traffic. They essentially "sniff" the network for suspicious behaviour.

5. ACKNOWLEDGMENTS

We feel grateful to the anonymous referees for their comments and for their valuable suggestions that have helped immensely in preparing the revised manuscript.

6. REFERENCES

[1]. Zhiyuan Tan, Upasana T. Nagar, Xiangjian He, and Priyadarsi Nanda, Ren Ping Liu, Song Wang, Jiankun Hu, "Enhancing Big Data Security with Collaborative Intrusion Detection", IEEE Cloud Computing published by the IEEE computer society, 2325-6095, pp. 34-40.

[2]. Victor, N., Lopez, D., & Abawajy, J. H.. Privacy models for big data: a survey. International Journal of Big Data Intelligence, 2016 3(1), 61-75.

[3]. Lopez, D., & Gunasekaran, M. Assessment of Vaccination Strategies Using Fuzzy Multi-criteria Decision Making. In Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015) 2015: 195-208. Springer

[4]. Lopez, D., Gunasekaran, M., Murugan, B. S., Kaur, H., & Abbas, K. M. Spatial big data analytics of influenza epidemic in Vellore, India. In 2014 IEEE International Conference on Big Data (Big Data), 2014, October: 19-24. IEEE.

[5]. Thilagavathi, M., Lopez, D., & Murugan, B. S. Middleware for Preserving Privacy in Big Data. Handbook of Research on Cloud Infrastructures for Big Data Analytics, IGI Global, 2014.

[6]. David Gill and II-Yeol Song, "Modeling and Management of Big Data: Challenges and opportunities", Future Generation Computer Systems 63(2016), pp 96-99

[7]. M.M. Potel, C A Dhote, D.H. Sharma, "Homomorphic Encryption for Security of Cloud Data", Procedia Computer Science 79 (2016), pp. 175-181.

[8]. Zhong Wang, Qian Yu, "Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures", Computer Law & Security Review 31(2015), pp.782-792

[9]. N. Kshetri, "Big data's impact on privacy, security and customer welfare", Telecommunicaitons Policy 38 (2014), pp.1134-1145.

[10]. S.Akter, S.F.Wamba, A.Gunasekaran, R. Dubey, S. J.Childe, "How to improve firm performance using big data analytics capability and business strategy

alignment?", International Journal of Production Economics, 182 (2016), pp.113-131.

[11]. A. Siddiqa, I.A.T. Hashem, I.Yaqoob, M.Marjani, S. Shamshirband, A. Gani, F.Nasaruddin, "A Survey of big data management: Taxonomy and state-of-the-art", 71 (2016), pp. 151-166.

[12]. R. Shaikh, M. Sasikumar, "Trust Model for Measuring Security Strength of Cloud Computing Service", Procedia Computer Science 45 (2015), pp 380-389

[13]. Vennila.S, and Priyandarshini J. , "Scalable Privacy Preservation in Big Data A Survey", Procedia Computer Science 50 (2015), pp. 369-373

[14]. B. H. Krishnaa, Dr. S. Kiranb, G. Muralia, R.P.K. Reddy, "Security Issues In Service Model Of Cloud Computing Environment", Procedia Computer Science 87 (2016),pp 246-251

[15]. N. Khana, A. Al-Yasirib, "Identifying Cloud Security Threats to Strengthen Cloud Computing Adoption Framework", Procedia Computer Science 94 (2016), pp 485-490

[16]. S.A. Hussain, M.Fatima, A. Saeed, I. Raza, R. K. Shahzad, "Multilevel classification of security concerns in cloud computing", Applied Computing and Informatics, 2210-8327, 2016, King Saud University Published by Elsevier B.V.

[17]. C. Saadia, H. Chaouib, "Cloud Computing Security Using IDS-AM-Clust, Honeyd, Honeywall and Honeycomb", Procedia Computer Science 85 (2016), pp 433-442