Big Data Security with Access Control Model and Honeypot in Cloud Computing

Jahanara Akhtar
Associate Professor
Department of Computer Science & Engineering
Dhaka International University, Dhaka, Bangladesh

Md. Tahzib-Ul-Islam
Assistant Professor
Department of Computer Science & Engineering
Dhaka International University, Dhaka, Bangladesh

Md. Habibullah Belali
Assistant Professor
Department of Computer Science & Engineering
Dhaka International University, Dhaka, Bangladesh

Saiful Islam
4Department of Computer Science & Engineering
Dhaka University of Engineering (DUET)
Gazipur, Bangladesh

ABSTRACT
Big data typically coins with large volume of data and various enterprises are involving day to day in cloud environment. Nowadays, Cloud facility adoption has enlarged. With the acceptance of cloud facility, numerous of the enterprises are expending to store and process Large Data in cloud. Safety methods provided by the facility providers might not be sufficient to safe the data in the cloud. Enterprise as well as users are suffering with proper security aspect to store, retrieve and process big data in cloud environment. An access control model with honeypot is presented in this paper. Access control model deals with various parameters of authentication, log etc. Various link and areas are included as honeypot to catch hackers or unauthorized users.

Keywords
Cloud Computing, Big Data, Honeypot, Cloud data Security, Access Control Model.

1. INTRODUCTION
The term big data devised newly and becomes popular due to its competences for storage (external the size of the typical database), recovery, study, and to harvest valuable outcomes. It is a formless, huge size, endlessly rising on a real-time foundation, and hard to process. Technology fluctuations effect the classification changes of such data completed a period. Therefore, it's meaning fluctuations time to time and organization to organization (creation it hard to have a faultless definition). Every organization has conferred attracted in the ordering of such formless data. As the technology reformed from computers to hand devices, the processing became a large problematic. Due to this cause, cloud requirement occurs. Cloud can stock a big size of data, multipart divisions, and generation of client output.

Cloud necessity for vast data must consider its size, a diverse type of data from many sources, the speed of its movement (arriving and departing), possible value if properly ordered and handled, and secrecy. Cloud computing is a model with infinite on-demand facilities. It can virtualize hardware and software properties, high processing power, storage, and pay-per-usage. Besides, it transfers price calculation tasks to the earner and reduces the excessive setup of computing facilities at minor initiatives. It has reachable normal resources and becomes computing power as needed. Cloud services deliver infrastructure, software, and platform as services.

Cloud computing permits for the loading and processing of substantial formless size of continuously produced data, resource availability, and burden tolerance over its numerous hardware and software services. Several enterprises like Nokia, RedBus, Google, IBM, Amazon, and Microsoft offer clients with on-demand services. The large corporate decided to migrate to Hadoop distributed file system (HDFS) that incorporates data into one area and uses appealed algorithms to become correct results to clients. The benefit of using Hadoop is inexpensive storage compared to traditional databases. Now, HDFS helps Nokia, RedBus, Google, and other corporations to achieve their requirements. The facility supports these corporations to think on their trades rather than on technical particulars and supplies.

Big Data technology solves many problems irrespective of volume, velocity, and source of generation. It is a constantly changing technology, and many industries, customers, and government agencies are involved in its usage and management. Further, if the data is in a cloud environment, the user access level management, privacy policies, user accountability and service provider accountability comes into the main duty of cloud administration.

Large Data technology resolves various problems regardless of size, speed, and source of generation. It is a continually altering technology, and several trades, patrons, and government organizations are convoluted in its practice and administration. More, if the data is in a cloud location, the manipulator entrée level organization, secrecy rules, user responsibility and service supplier liability originates into the main charge of cloud management.

Due to this purpose, we requisite to form safety rules, entree privileges, safe storage, and repossession systems. Due to nonstop rising of data, regulator is mandatory to save the valued data. So, it is needed to apply data governance rules like structural performs, working performs, and relational performs.

Disaster retrieval (in the incident of hazardous misfortunes counting floods, earthquakes, fire, and accidental loss of data) for valued data is a obligation. The large businesses describe a set of actions for a disaster retrieval strategy to re-establish the data. In addition to safety rules, disaster retrieval is strongly suggested (fault-tolerance depends on disaster retrieval). Other problems comprise the safe handover of data to the cloud, joining high-performance computing, and data management. Large data in the cloud has several study and real-world challenges. Storing the data using encryption

method takes additional time. Standardization of actions is mandatory to reduce the impact of heterogeneous data. Data governess, recovery plans, quality of services for safe handover of data, and petaflop computing are some of the problems with operation.

The fruitful placement of large data on cloud needs building a commercial case with a suitable strategic plan to use the cloud. The project must improve efficiency, extract additional important worth, continuous development, client achievement, fulfilment, loyalty, and safety. Assessment of a appropriate cloud atmosphere (private or public) and progress of a technical method are also needed. Then, address the governance, secrecy, safety, threat, and responsibility necessities. Lastly, the operational environment can be settled. The supplier encounters several challenges depending on the cloud data situation. Safety, price issue, client fulfilment, and service reliability are main matters.

2. LITERATURE REVIEW

Storage, processing, and repossession of large data in the cloud are important difficulties in existing study. Pedro et al. [1] studied the outline of current and upcoming issues. The document discusses scalability and fault tolerance of several merchants with Google, IBM, Nokia, and RedBus. The writers additional considered the safety, secrecy, integrity, disaster retrieval, and fault tolerant matters. The review of present service models, importation ideas of cloud computing, and processing of large data are explained in [2]. Elmustafa and Rashid [3] shown the review topics of large data safety in cloud computing.

Linda et al. [4] presented ecological samples of large data use in government that contains Environmental Protection Agency, Department of the Interior, Department of Energy, and Postal Services. The revision contains of the government exposed entree initiatives, central data centre alliance initiative, and implementation of obedience online. James [5] offered a roadmap to the achievement of large data analytics and applications. The story discourses the definition and description of formless data, related use cases in the cloud, possible profits, and challenges related with organizing in the cloud.

The effect of cloud computing on Healthcare studied in [6]. The learning discovers on-demand access to computing and big storage, supporting large data sets for electronic health records and the aptitude to examine and pursuing the health records. Keeso offered the ecological sustainability of big data, barriers, and prospects [7]. This study similarly contains fresh chances for partnership based teamwork, sustainability to organizations to large data labours, and developing business models.

Yan et al. [8] described the admission regulator in cloud computing. The study shows the time-based entree control in cloud computing using encryption methods. Yuhong et al. [9] inspect data privacy in cloud computing. The paper uses the trust-based assessment encryption model. In this model, the faith issue chooses the entree control of customer status. Young et al. [10] shows the safety matters in cloud computing and defined the entree control requirements, verification and ID supervision in the cloud.

Ali and Erwin [11] studied safety and secrecy matters on large data and cloud features. They decided that cloud data secrecy and security is based on the cloud owner. They also described large data safety issues and cloud safety issues. Their study inspects the safety strategy management and huge data structure and software design models. They did not propose any specific model but discussed all probable resolutions for the safety of large data in the cloud.

Marcos et al. [12] offered methods and atmospheres to convey out large data computing in the cloud. The study deliberates the picturing and /user communication, model structure, and data managing. Venkata et al. [13] observed topics in a cloud atmosphere for large data. The main attention is safety problems and imaginable solutions. Again, they discoursed MapReduce and Apache atmospheres in the cloud and required for the security.

Saranya and Kumar [14] talked the safety matters related with large data in a cloud atmosphere. They recommended some methods for the difficult business situation. The study discourses formless big data features, analytics, Hadoop construction, and real-time large data analytics. The writers did not show any specific model in the study. They clarified little ideas related to safety in a cloud environment.

Avodele et al. [15] described matters and contests for placements of large data in the cloud. They advised clarifications that are pertinent to administrations to organize the data in the cloud. The writers designated the significance of verification controls and entree controls.

Ramgovind and Smith [16] described the general safety view and possible of cloud computing. They inspected cloud intimidations and safety needs with identification and authentication, authorization, privacy, honesty, non-repudiation, and obtainability. They also provided the cloud transfer models for the private, public and mix cloud. They did not change any model in this study. Dimitrios and Dimitrios talked the cloud computing safety issues [17]. They recognized the safety needs and describe the feasible result to remove the possible dangers. To guarantee verification, integrity, and privacy, they advised a cryptography-based solution.

Gai et al. [23 - 25] suggested a cloud-based method to safe sensitive data. The model dynamically allocates the data packet to cloud resources founded on safety requests. The planned model needs the third party to review the confirmation procedure. The writers also recommended completely Homomorphic Encryption for Blend Operations (FHE-BO) model. It uses tensor laws to transmit the computations of mixture mathematics procedures over real numbers. The authors requested that the technique parallel contracts with the confrontational threats and care computations on cipher-text.

Authors of [26] proposed a framework model contains the data encryption, correctness, and processing. They talked over the many methods to examine cipher text and query separation (evade the untrusted server). Organized searching, trade with adjustable word sizes, penetrating encrypted index, then provision for unseen search is slice of the study. Authors also projected admission regulator model with encrypted processing data is suitable to avoid untrusted earner and mischievous handlers in the cloud

3. PROPOSED PROTOCOL

Proposed protocol includes Access Control Model and Honeypot described in 3.1 and 3.2.

3.1 Access Control Model

The access control model for cloud storage incorporates the authentication of user and users' current access level. The user token identification (UTI) is attached as soon as the user login

into the system.

UTI will compose of a set of parameters as

$UTI = \{UTI_I, UTI_2, UTI_3, UTI_4 \text{ ---------- } UTI_n\}$

Where UTI_I may be user ID, UTI_2 = date of issue, UTI_3 = date of expiration, UTI_4 =access time, UTI_5=access place, UTI_6=log entry, UTI_7=alarm status etc. When any user entered into the cloud, all parameters of UT_I will be initialized and UTI verifies the user's access limits and allows or denies appropriate file access. System will observe the activities/movements of the users and update the UTI when necessary. For example, UIT_7(alarm status) is zero initially, when any user will try to access unauthorized file, then UIT_7 will set 1 and set alarm as well as freeze this user. Again, UTI_3(date of expiration) will be checked and if this value is expired then system will set alarm (UIT_7). System will observe UTI_4(access time). If the system finds any unusual time of access, then sets UIT_7. By this method all parameters of UTI will be verified and system will set alarm for awareness of the administrator or treated the user as unauthorized. Parameter UT6 {Log Entry (UTI_I, UTI_2, UTI_3, UTI_4 ------)} will log all information of users for future uses.

3.2 Honeypot

Honeypots are mostly installed to confuse the enemies from the actual servers. Using a honeypot may provide the system chance to detect and answer to an attack (on the fake system) before the attackers are able to do any real damage. Some interesting links or areas will be deployed as honeypots in the system to catch unauthorized users or intruders or hackers. Authorized users will be informed previously not to click on this link or not move mouse over these areas.

$HP = \{HP_1, HP_2, HP_3 \text{ ---------- } HP_n\}$

Let HP_1= "Click here for more details", HP_2=" Next here" etc or some relevant sentence or area related to corresponding organization and some link and area will be shown over the screen. It is forbidden for authorized users not to click on those links or not to move the mouse over the interesting area. As it is informed to users previously, only authorized users will not click or over the mouse. But unauthorized users not know about these honeypots. If they click on links or over the mouse on those areas, they will be caught. But one problem may be here, authorized users may click or over the mouse on these areas (honeypots) wrongly. Then a OTP (one-time password) will be sent to their registered mobile or email. When this OTP will be entered then system will treat this user as authorized. Unauthorized users will not able to get or enter this OTP, so system will treat them as unauthorized and set alarm and freeze

3.3 Example of Access Control Model and Honeypot

Table 1. User Log and Action

User	Status	Result	Action
User1	Initial login	Not existing in log entry	Identified as new hacker, alarm
User2	Frequent login	Available in log	Alarm and freeze
User3	Confidential User	Unauthorized access	Alarm and freeze

Table 2. Hacker Log and Detection

User	Status	Time and access limits	Result	Action
User 1	Confidential User	Access with access limits	Confidential hacker and clicked on a link	Freeze and alarm
User 2	Frequent login	Access outside bounds and different times of the day	Verify the time and log entries and identified as inside hacker, mouse over a region	Alarm and Freeze
User 3	Confidential User	Unauthorized access	Alarm and freeze	

The tables (Table 1 and Table 2) provide the clue of log study to find the inside and outside hackers. The user access logs, authentication, and access privileges have a major role in providing the hacker data to the security administrator. Tables show status, results action of various users. According to Table 1, at first time "User1" is entered but no log entry available in history. So this user will be treated as new hacker and alarm to administrator. Again, if any other user "User2" is frequently try to login and log history is available in log, then set alarm and freeze this user. By same method, if "User3" is a confidential and unauthorized user, then set alarm and freeze this user. Table II shows hacker log and detection process. Let "Use1" is a confidential user and he is accessing with access limit. Then this user will be treated as confidential hacker as he clicked on a link (honeypot). "Freeze and set alarm" action will be taken. Again, "User2" is a frequent login user and he accessed of outside bounds and different times of the dada, then system will verify time as well as log entries history. Moreover, this user covered his mouse on a forbidden region (honeypot). The "set alarm and freeze" this user action will be taken. Again, if "User3" is new user and no entry history is available in log, then it will be treated as illegal user and set alarm to administrator.

4. PROPOSED PROTOCOL

The current questions and challenges were deliberated in big data in the cloud [18] and additional related papers [8-15, 19-25]. Implementation of cloud facilities for storing Big Data is rising day by day. Still, there are several exposed safety difficulties which essential to be speaking in the cloud. By a lot of secret data has vast safety threat on Big Data files in cloud. Compromise or security breach at one of these locations lead to a huge harm for clients. After assessment of current expansions, it might be concluded that it is needed to change a faith and admission control procedure in a cloud location for big data processing. So, in the present study, an impartial purpose was planned with a set of users, related access rights, resources and return result verification. Moreover, honeypot is included as trap to catch thief or hacker and unauthorized user. The proposed system is suitable for storage, processing, and retrieval of big data in a cloud environment.

5. REFERENCES

[1] Pedro Calderira Neves, Bradley Schmerl, Jorge Bernardino, and Javier Camara., "Big Data in Cloud Computing: Features and Issues", International Conference on Internet of Things and Big Data, Jan 2016.

[2] Richard Branch, Heather Tjeerdsma, Cody Wilson, Richard Hurley, and Sabine McConnell., "Cloud Computing and Big Data: A Review of Current Service Models and Hardware Perspectives", Journal of Software Engineering and Applications, 2014, 7, 686-693.

[3] Elmustafa Sayed Ali Ahmed and Rashid A. Saeed., "A Survey of Big Data Cloud Computing Security", International Journal of Computer Science and Software Engineering (IJCSSE), vol 3, issue 1, December 2014, pp. 78-85.

[4] Linda K. Breggin, et al.(editors). "Big Data, big challenges in Evidence-based policy making", 2014, West Academic Press.

[5] James Kobielus., "Deployment in Big Data Analytics Applications to the Cloud: Road Map for Success", Cloud Standards Customer Council, Technical report, 2014.

[6] "Impact of cloud computing on Healthcare, Version 2.0", Cloud Standards Customer Council, Technical Report, February 2017. [7] Alan Keeso., "Big Data and Environmental Sustainability: A Conversation Starter", Smith School Working Paper series, working paper 14-04, 2014.

[7] Yan Zh., Hongxin Hu, Gail-Joon Ahn, Dijiang Huang, and Shanbiao Wang., "Towards Temporal Access Control in Cloud Computing", Proceedings of IEEE INFOCOM, 2012.

[8] Yuhong Liu, Jung Ryoo, and Syed Rizvi., "Ensuring Data Confidentiality in cloud Computing: An Enryption and Trust-based Solution", Proceedings of IEEE 23rd Wireless and Optimal Communication Conference (WOCC), 2014.

[9] Young-Gi Min, Hyo-Jin Shin, and Young Hwan Bang., "Cloud Computing Security Issues and Access Control Solutions", Journal of Engineering, 9, 2, 2012.

[10] Ali Gholami and Ervin Laure., "Big Data Security and Privacy Issues in the Cloud", Int. J. of Network Security & its Applications (IJNSA), vol. 8, No.1, Jan 2016, pp. 59-79.

[11] Marcos D. Asuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A. S. Netto, and Rajkumar Buyya., "Big Data Computing and Clouds: Trends and future directions", J. parallel Distributed Computing, 79- 80, 2015, pp.3 15.

[12] Venkata Narasimha Inukollu, Sailaja Arsi, and Srinivasa Rao Ravuri., "Security Issues Associated with Big Data in Cloud Computing", Int. J. of Netork Security & its Applications (IJNSA), Vol. 6, N0.3, May 2014, pp. 45-56.

[13] R. Saranya and V. P. Muthukumar., "Security issues associated with big data in cloud computing", Int. J. of multidisciplinary and development, vol. 2, issue.4, April 2015, pp. 580-585.

[14] Avodele, O., Izang A. A., Kuyoro. S. O., and Osisanwo, F.Y., "Big Data and Cloud Computing Issues", Int. J. of Computer Applications (0975 – 8887), vol 133, no.12, Jan 2016, pp.14 – 19. [16] Ramgovind, S., and Smith, E., "The Management of Security in Cloud Computing", IEEE Information Security for South Africa (ISSA), 2010.

[15] Dimitrios Z., and Dimitrios, L., "Addressing cloud computing security issues", Future Generation Computer Systems 28 (2012) 583–592 [18] Dimitrios, Zissis. And Dimitrios, L., "Addressing cloud computing security issues", Future Generation Computer Systems, 28 (2012) 583–592.

[16] Dawn, x. S., David, W. and Adrian, P., "Practical Techniques for Searches on Encrypted Data", IEEE Symposium on Security and Privacy, 2000, pp. 44-55.

[17] Jangala Sasi Kiran, M. Sravanthi, K. Preethi, and M. Anusha., "Recent Issues and Challenges on Big Data in Cloud Computing", IJCCCSSST, Vol.6, issue.2, June 2015, pp. 98-102.

[18] Shehnila Z., Najeed A. K., and Mohsin Ali Memon., "Systematic Analysis of Risks in Cloud Architecture", Int. J. computer Science and Information Security (IJCSIS), vol. 14, no.11, November 2016.

[19] D Wallner, E. Harder, and R. Agee, "Key Management for Multicast: Issues and Architectures", RFC 2627, June 1999. [23] K. Gai, L. Qiu, M. Chen, H. Zhao, and M. Qiu., "SA-EAST: securityaware efficient data transmission for ITS in mobile heterogeneous cloud computing", ACM Transactions on Embedded Computing Systems, Vol. 16, No. 2, Article 60, Publication date: January 2017.

[20] K. Gai and M. Qiu., "Blend Arithmetic Operations on Tensor-based Fully Homomorphic Encryption Over Real Numbers", IEEE Transactions on Industrial Informatics, December 2017.

[21] K. Gai, M. Qiu,and H. Zhao., "Privacy-Preserving Data Encryption Strategy for Big Data in Mobile Cloud Computing", IEEE Transactions on Big Data, May 18, 2017.

[22] Y. B Reddy., "Big Data Security in Cloud Environment", 2018 4th IEEE International Conference on Big Data Security on Cloud, pp. 100-106, May 2018.

6. AUTHORS PROFILE

Mst. Jahanara Akhtar is now serving as an Associate Professor of the department of Computer Science and Engineering, Dhaka International University, Dhaka, Bangladesh. She is a Research Fellow (PhD) in the department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh. Jahanara completed B.Sc in Electronics & Computer Science and M.Sc. in Computer Science and Engineering from Jahangirnagar University. She has publications in national and international conference and journals. Her research interest includes Cryptography, Secure Wireless Sensor Network, Image Processing and Artificial Intelligence.

Md. Tahzib-Ul-Islam is currently serving as an Assistant Professor of Computer Science and Engineering department, Dhaka International University, Dhaka, Bangladesh. He is a M.Sc. student in the Institute of Information and Technology, University of Dhaka, Dhaka, Bangladesh. Tahzib completed B.Sc in Computer Science and Engineering in Department of Computer Science and Engineering from University of Dhaka. He published research papers in national and international conference and journals. His research interest includes Cryptography, Network Security, Image Processing and Cloud Computing.

Md. Habibullah Belali is currently serving as an Asistant Professor of Computer Science and Engineering department, Dhaka International University, Dhaka, Bangladesh. Habibullah completed B.Sc. in Computer Science and Engineering from University of Dhaka. He has publications in national and international conference and journals. His research interest includes Image Processing, Data Mining and Advance Database.

Saiful Islam is currently pursuing his Ph.D. in Computer Science and Engineering degree in Dhaka University of Engineering and Technology (DUET), Gazipur, Bangladesh. Saiful completed B.Sc. in CSE and M.Sc. in EEE from Dhaka University of Engineering and Technology (DUET), Gazipur, Bangladesh. He also completed M.Sc. in ICT from Bangladesh University of Engineering and Technology (BUET), Bangladesh. He has several publications in national and international conference and journals. His research interest includes Cryptography, Network Security, Wireless Sensor Networks, Cyber Physical System and Cloud Computing.