ISSN: 2633-4828 Vol. 6 No.3, September, 2024

International Journal of Applied Engineering & Technology

ENABLING SECURE DATA SHARING IN BIG DATA ECOSYSTEMS THROUGH ADVANCED ACCESS CONTROL MODELS

Priyam Vaghasia and Dhruvitkumar Patel
Mondrian collection, Staten Island Performing Provider System
priyamvaghasia57@gmail.com and pateldhruvit2407@gmail.com

ABSTRACT

Safe big data sharing is indispensable for the optimisation of big data and safety and confidentiality of the information. Proliferation of data has resulted with the increase in gathered data at a very high rate along with requirement of real time processing / analysis, hence there is need for enhanced access control models solving problems related to data security and access in complex and distributed environments. This work focuses on how optimum access control measures can be adopted for boosting security in big data environments. These models include fine-grained access controls, role-based access control and attribute-based access control, which are comprehensive in nature and flexible enough to begin with to address the dynamism which is already associated with big data. The paper discusses the problems that arise with traditional access control models that typically lack scalability, flexibility and heterogeneity of big data technologies. The following models incorporate context-awareness and users' characteristics to guarantee the privacy of the data while using it and sharing in a more efficient and effective manner. Also, it describes the importance of using encryption and secure multi-party computing to improve the protection of the data. Through the application of these advanced models, the organizations shall ensure that the Big Data environment is safe and reliable, and thus, supporting the sharing and collaboration over Big Data insights and knowledge in a very effective and flexible way while at the same time protecting the data and knowledge from misuse. This work adds up to the existing literature on big data security that seeks to provide a framework by which secure big data sharing can be made possible, thereby enhancing efficiency in sharing of big data.

Keywords: Big Data Security, Access Control Models, Data Sharing, Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC).

1. INTRODUCTION

One of the most recent phenomenon in organizations is the big data which has brought about change to how organizations manage information. Today, in healthcare, banking and other sectors, the capability to process big data in real time can be viewed as a valuable asset, which can improve businesses and bring new ideas and competitive advantages. But with big data, the sheer volume and distributed access create acute security issues when it comes to data sharing and data access. As organizations continue to incorporate big data ecosystems into their business operations protecting confidential data has become more significant.

Conventional access control methods that are developed for small and centralized type of data systems fail in the case of big data. These systems are able to handle the volume, velocity and the variety of data hence there could be potential gaps that could be exploited by cyber criminals. To redress all these challenges, enhanced access control models have been proposed to enhance depth and comprehensiveness of big data security models. The development of these models is then examined in this paper, with special emphasis on how they handle the dynamic nature of the big data while keeping data sharing as secure and optimized as possible.

1.1 Challenges in Traditional Access Control Systems

Discretionary Access Control or DAC, and Mandatory Access Control or MAC are two traditional access control models that have been developed with the perception of a more or less static data environment and clear-cut user roles. These systems become challenging especially in the aspects of scalability and flexibility in the context of big data. Generally, with the increasing volumes of data moving across the systems and the complexity of users' needs, primitive access control mechanisms may not offer satisfactory protection.

This is due to the fact that one of the major considerations can be management of access rights for a large number of users in the geographically dispersed environments. Mainframe traditional models involve role and group based access control, although not best suited for complex and evolving big data structure. Since data is created and disseminated day in day out across various platforms, access control data accuraten ess and update is a challenge.

Yet another problem concerns the absence of context information in common access control frameworks. Big data assets may be valuable, which may also be sensitive when used in certain context and at certain point in time. These models do not allow flexibility in the setting of access control and this is a weakness because an adversary can take advantage of the existing loop holes.

1.2 Advanced Access Control Models for Big Data

To overcome with the limitations of conventional access control models, several new models has created which are RBAC and ABAC. These models provide finer and more controllable ways of controlling the access rights in the big data context, thus helping organizations to enhance protection of costly information while permitting extensive sharing.

Role-Based Access Control (RBAC) works in the principle, that permissions of users are tied to the specific roles in the organization chart. This model is particularly appropriate where the user roles are generic and don't change very frequently. Nonetheless RBAC may face challenges in environments where users may require to have an access to another role or to type of data in regular basis. To counter this, ABAC proposes the idea of attributes – these are factors that can relate to the user, the environment and the data. This means it is easier to implement much smaller and more specific forms of access control which are also reactive to the state of affairs on the ground. Figure 1 shows how RBAC function.

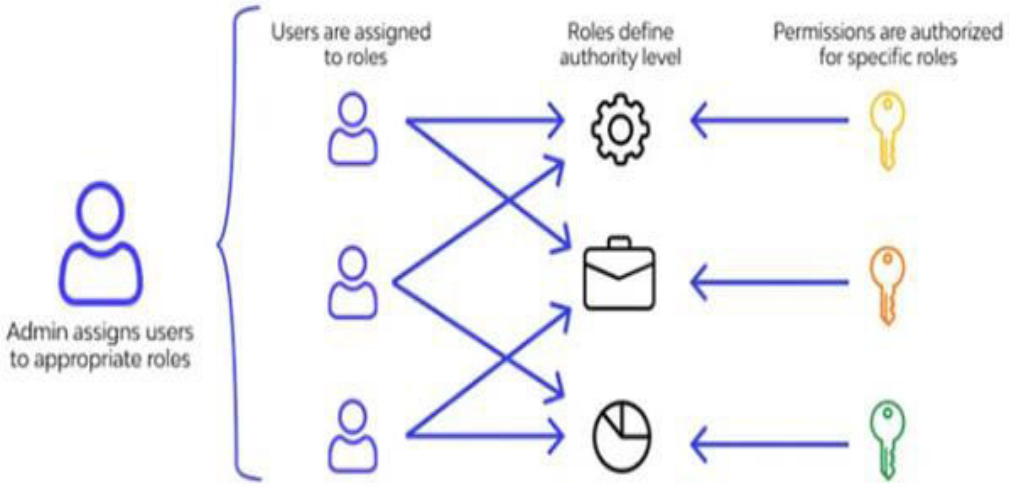

Figure 1: Framework for Role-Based Access Control (RBAC)

Given this characteristic, it can be stated that ABAC is especially efficient in the big data context as the conditions when data is required can be quite diverse. Through attributes it becomes possible to have dynamic security policies that reflect contexts and have higher effectiveness compared to simple roles as the usage of data would not be affected significantly. This flexibility is then instrumental in guaranteeing that access control will be able to address the liquid nature of big data. The figure 2 illustrates how ABAC works.

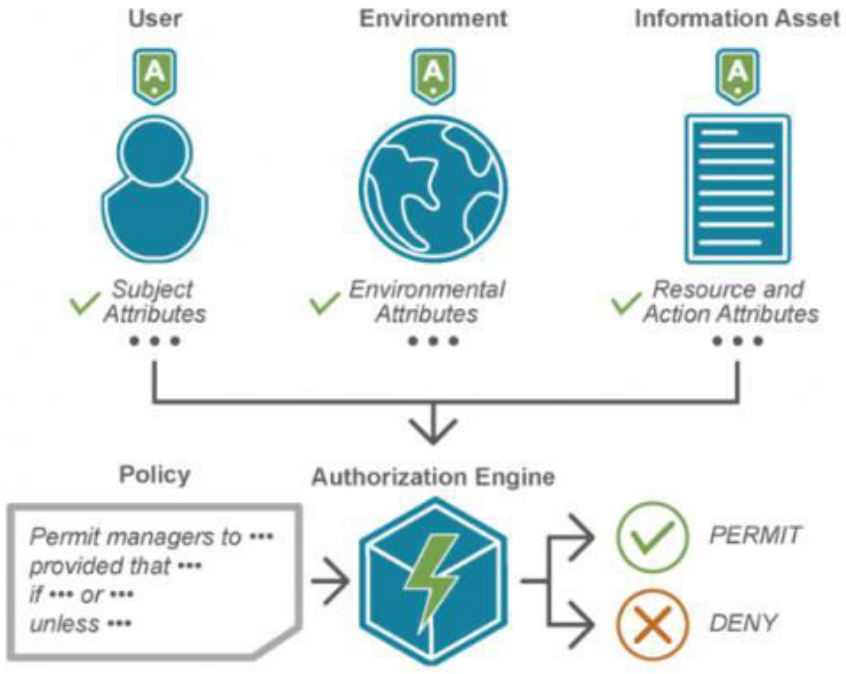

Figure 2: Framework for Attribute-Based Access Control (ABAC)

1.3 Implementing Fine-Grained Access Controls

This is one of the critical requirements of the protection of big data architectures for it facilitates fine-grained access control that is necessary for the overall security of such environments. Fine grained controls are different from traditional models that give general access policies and let an organization set precise access rules down to the level of the single data element. Such level of sensitivity is important to shield sensitive information especially in businesses where data protection is paramount.

Chief among the motivation and advantages of fine grained access control is the prevention of the violation of the principle of least privilege. This also helps minimize the likely hood of a third party getting in an accessing private information hence reduces the repercussions of such leakages. It has to be this way because in big data scenario the data is distributed across multiple platforms and many users are potentially going to interact with it.

For organizations to practice fine grained access control effectively they need to incorporate technologies such as tagging of data and encryption. Datetime tagging involves adding metadata with the data component with policies that can be used to regulate access through data attributes like sensitivity, user group or environment among others. Encryption also makes the use additional layer of security whereby even if the wrong parties gain access to the data, they will be unable to decipher it without keys.

One of the most complex and still open issues of the access control in the distributed big data environments is related to the consistency and scalability of the implemented solutions. In distributed environment, the data is located in different areas and each of these areas will have its own security protocols of access. When implemented in isolation, these policies can be very effective at enforcing access controls throughout the ecosystem, but co-ordinating them in order to provide consistency of enforcement across the entire ecosystem can often be a time-consuming and resource intensive process.

Another problem lies in extending the access control polices such that any node in a distributed system can be protected in a uniform manner. Lack of coherence in the implementation of policies may result in a loss of security since the users can always find a way around the restrictions through the other nodes that offer less secure access to data. On this, it is important for organizations to develop a centralized policy management system in the distributed architecture because it should be capable to enforce the access control over all nodes.

Another major issue in distributed big data environment is scalability. The problem is that with the growing number of users and the amount of data, the management of access controls becomes a highly complicated process. To have security without losing performance, organizations must implement large- scale access control models that will work well with large number of data and users. Leveraging cloud based access control solution may be necessary to address access control challenges in highly volatile and dynamic big data environment.

2. REVIEW OF WORKS

The significance of big data can be explained as follows, it has revolutionized the way organizations store, manage and protect large volumes of information. With the increase of data volumes, there arises issues as to its management and protection especially in distributed and heterogeneous platforms. Thus, such subjects as progressed access control models and big data security measures have gained significance to provide confidentiality, integrity, and availability of information and at the same time facilitate the organization's aims to leverage the value of their data resources.

As the years go by, the worth on the big data is growing, and many researchers and industry specialists have attempted to discover the best ways to protect the big data. These have brought out the shortcoming of the earlier access control models and the need to come up with better and more efficient ones. This literature review summarizes current research on big data security, some recent works on access control models and various research trends in managing big data focusing on the existing state-of-art and future trends.

2.1 Growth and Impact of Big Data

The advancement of big data has been phenomenal and many felt that the overall data capacity of the world would be 44 zettabytes by 20120 (DataIQ News, 2014) [1]. Data being produced by organizations today is massive and has made it a requirement for new ways to be developed in order to manage and secure data. Unfortunately, as Moore (2014) indicated, with the advances in the use of big data, there is a need for optimization of data centers that are vital for governments hence the importance of proper handling of data [3]. The above advancements re-emphasize the need for security mechanisms that are elastic in order to address the issues that characterize big data systems.

In addition, Miller (2014) pointed out the use of big data approach to human generated data, coupled with the need for secure methods for data sharing and processing [5]. The utilization of advanced analytics together with big data pose an increased need for strong access control measures that will help organizations to secure their data yet allow them to realize the full potential of their data resources.

2.2 Traditional Access Control Models and Their Limitations

Some of the well-known models of access control are Discretionary Access Control (DAC) and Mandatory Access Control (MAC) models that have been practiced across different domains. However, as Hu et al. noted, those models would have certain challenges with big data especially scalability and flexibility as stated by Hu et al. (2011) [6]. Due to the conceptual rather than real model, it is not convenient for the usage in such big data environments as data is stored in different places and can be used by different individuals and applications.

Bell (2014) also pointed out that existing concepts of access control in big data have also been witnessed to present myriad of issues since traditional modes of AC do not support strong security policies in loosely federated systems [4]. This is especially so since most of these models are not context-aware, making them highly vulnerable to various forms of security threats that could be engineered by bad actors.

2.3 Advancements in Access Control Models

However, the aforementioned access control models are not sufficient enough, and for this reason, the researchers have come up with the following Enhanced Models: Attribute-Based Access Control (ABAC). Hu et al. proposed big data access control model based on attribute-based access control (ABAC) and they discussed the flexibility of ABAC based on user attribute and environmental conditions in big data environment [9]. Thus, the decision point provides flexibility in the formulation of access control policies to match the contexts in which data is accessed resulting to better security.

Smith (2014) has elaborated the role of Hadoop's security model and pointed out that the newer and more enhanced access control mechanism to safeguard sensitive data within the Hadoop cluster [11]. With Hadoop still reigining as one of the popular and prominent big data processing platforms, the improvement of the necessary security measures such as role-based and attribute-based access control is valuable for making the processing safe for the data.

Hence, security risks posed by big data are manifold, which include matters of privacy, authorization, and protection of distributed systems. The security and privacy issues as outlined by Shea (2013) includes the problem of managing access control in a distributed data environment and that of data leakage [16]. Such issues call for effective security models that could cater for the needs of the big data environments.

An access control model of big data has been developed by Zeng et al. 2013, in which details allow for understanding of the rights of access with data content giving a comprehensive approach to data protection [15]. This model enriches notion of access control by including data content in the decision process – only a user with appropriate credentials can access information based on context and value.

This is the case because the technologies and trends comprising big data are constantly changing and, hence, so is the approach to its protection. In the same year, Hu and Scarfone also mentioned that the evaluation of access control systems is crucial in terms of providing information to the organizations so that they can modify their security systems [8]. Without these metrics, organisations cannot be confident that their access control systems will be capable of keeping up with the fast-pace of the BIGDATA environment.

On this note, there is significant potential with the future technological rolling out of invite innovations like the blockchain and artificial intelligence in bolstering big data security. When implemented these continued technologies can offer better solutions for the administration of the access control as well as enhanced data integrity within the distributed systems. With increased investment in big data, there will be a need to design and apply good security measures and practices to protect the stored data and uphold credibility in the undertaking of data-driven plans.

3. PROPOSED METHODOLOGY

In this methodology, the processes involved in this systematic review and analysis of the current literature on secure data sharing in big data ecosystems especially with the focus on advanced access control models are described. The purpose is to review the findings of documents, assess the short-comings of documented work, and find out possibilities for further research. Since the work is aimed at providing an extensive understanding of the subject of the research given its broad spectrum and non-experiment based approach of the study, it relies on secondary research.

To that extent, the first method undertaken in the actual approach used in the study was to perform a systematic literature review and this was accomplished by conducting searches through academic databases and general online search engines. The databases, IEEE Xplore, Google Scholar, and ScienceDirect were used to search using the keywords: "big data security," "access control models," "data sharing," "Hadoop security," and "attribute-based access control. " The search was done to contain only articles published within the last decade, in order to capture recent literature. Apart from journal articles, conference papers, technical and white papers from research organisation of comparable repute were also included where needed to get an insightful survey of the field.

Following that, there was a procedure of selection in order to find the most suitable studies for analysis. Specifically, the inclusion criteria concerned the work that was devoted to the question of access control in the big data context, investigated empirical or/and theoretical comparative evaluation of the effectiveness of certain access control models, and/or discussed trends and problems associated with security of big data. Articles that did not conform to these criteria or were not recent sources or that are considered non-representative were omitted. The selected studies were then classified into important areas including traditional access control models, modern day access control mechanisms, security issues in big data and future developments in big data security.

The next step was to review content of the selected studies in detail in order to extract contextual information. In the course of the review of each study, details on findings, methodologies and conclusions were obtained. The approach used meant that it focused on looking for patterns among the findings in regards to the established research questions with key interests lying in areas of agreement and disagreement, new trends or developing themes. Special emphasis was placed on the papers that offer the comparison of various access control models or present new approaches to protect big data. After that, the obtained results were discussed and connected in a logical context with the existing knowledge as a result of the prior studies, and presents the key discoveries of the existing studies in the field.

Last of all, we had to critically appraise the database of literature collected to establish their shortcomings. Evaluation was done in terms of the effectiveness of the methodologies employed in the reviewed studies, the external validity of the findings and the consideration of practical implementation issues by the studies. Consequently, the methodology outlines future avenues of research, to include following the incorporation of novel technologies like the blockchain and artificial intelligence in access control frameworks and coming up with more efficient and versatile security mechanisms for environments with growing volumes of big data. It helps to avoid such practice where literature review only presents the current state of research without adding the agenda for further research studies.

4. RESULTS

The findings of this study conducted based on the systematic literature review depict the landscape of the existing research on secure data sharing in big data context with special emphasis on the contemporary access control models. The findings are organized into five key areas: These include Traditional Access Control Models, Advanced Access Control Mechanisms, Security Challenges in Big Data Environments, Comparative Analysis of Access Control Models and Future Directions in Big Data Security.

4.1 Traditional Access Control Models

A close look at some of the traditional access control models including DAC and MAC showed some of the challenges that limit the use of these models in big data situations. However, these models have limitations in their applicability to big data ecosystem and fail to scale and adapt to the dynamic environment of big data. It is imperative to note that the various studies reviewed in this paper all pointed to the fact that introducing DAC and MAC in environments where data is constantly being produced and exchanged across various systems was proving to be quite a complex endeavour. The models that are currently being used do not have contextual awareness, and thus, were deemed not very comprehensive; in addition, their rigidity was noted to be one of the major issues that make these models less efficient in protecting big data.

4.2 Advanced Access Control Mechanisms

The review established that there is a rising trend of studies with emphasis on more effective access control models including the RBAC and ABAC. It was established that these models provided even more flexible and refined ways of managing access to the content of big data as contrasted with the more rigid sample models. ABAC in specific was observed to be flexible of adapting a number high attributes such as roles, environment, and the characteristics of data into an access control determination. This flexibility of implementing context-aware security policies can make organizations to be in a better position to protect the data as they share the data across the organization.

4.3 Security Challenges in Big Data Environments

The results also underscored the numerous security challenges associated with managing access controls in big data environments. One of the most significant challenges identified was the difficulty of maintaining consistent and effective security policies across distributed systems. The studies reviewed highlighted the risks associated with inconsistent policy enforcement, which can lead to security vulnerabilities and potential data breaches. Additionally, the complexity of managing access controls for large volumes of data and diverse user groups was a recurring theme, with many researchers emphasizing the need for scalable solutions that can handle the demands of big data environments.

4.4 Comparative Analysis of Access Control Models

Comparing various proposed models of access control allowed to identify increased interest in using advanced models and especially the ABAC model when managing the security of big data. The comparative studies that were made between the traditional and advanced models clearly depicted that the advanced models were better in performance with regard to flexibility, scalability and the ability to enforce more strict access controls. Also, context awareness and context adaptability of ABAC were being noted as its other advantages by offering the strong security policy when and where data access requirements can dramatically change. But the review also pointed out that the use of more sophisticated models entail a system design and availability of resources in order to avoid possible pitfalls on system performance.

4.5 Future Directions in Big Data Security

The review of the literature also indicated some of the research trends and future developments in big data security. Among potential directions in the development of access control systems, the potential of application of such advanced technologies as blockchain, and AI was highlighted. Blockchain provides a decentralised structure and immutability of big data which can boost up the security of distributed big data environment and, on the other hand, AI can also provide effective security measures in terms of analysing user behaviour and detecting potential threats of the big data environment in real time. According to the reviewed studies, these technologies might help to meet novel security threats arising from big data more effectively by presenting superior and more reliable ACC control solutions.

Based on the literature review presented herein, it is possible to pay attention to the further development of access control models as a reaction to the problems remotely connected with big data. Although the traditional models remained relevant in some situations, more progressive models like the ABAC are now considered critical in protecting the big data. The recommendation of future research directions indicates that the area has growth potential as well as future innovations especially relating to the incorporation of new technologies into the access control systems.

5. CONCLUSION

The present comprehensive review of the literature on secure data sharing in big data environment highlights an importance for the extra refined access control models that should address the novel challenges in big data environments because of its width and ever changing nature. Conventional access control instruments, though straightforward, have proven and are still relatively insufficient for equivocal data contexts. There are advanced models, which are more contextual and thus more suited to the big data security needs, such as Attribute-Based Access Control (ABAC). This involves emphasizing on the fact that as the usage of these models increases, there is an added ability to improve data protection while at the same time allowing easy access to useful data.

Future studies of access control systems can be extended in implementing the evolving technologies of the blockchain and AI. Altogether, these technologies may contribute to enhancements of the protection of and resilience of security arrangements in big data systems. However, bringing in of all these advanced models and technologies must take into consideration the strength and design of the system as well as availability and distribution of resources and most importantly the dynamic and ever-changing security threats. Hence, as

organizations are embracing big data more and more, there is always a need to have evolution of such access control mechanisms, so as to accommodate all such features as is depicted in the future.

REFERENCES

[1]. "Big data to turn 'mega' as capacity will hot 44 zettabytes by 2020,", DataIQ News, http://www.dataiq.co.uk/news/20140410/big-data-turnmega-capacity-will-hit-44-zettabytes-2020, Oct.2014.

[2]. H. Mir, "Hadoop Tutorial 1What is Hadoop," ZeroToProTraining, http://ZeroTOProTraining.com http://nusmv.irst.itc.it/.

[3]. J. Moore, "How big data is remaking the government data center," GCN, http://gcn.com/articles/2014/02/14/big-data-data-centers.aspx, Feb. 2014.

[4]. W. Bell, "The Big Data Cure," MeriTalk, http://www.meritalk.com/bigdatacure, 2014.

[5]. P. Miller, "Applying big data analytics to human-generated data," GIGAOM RESEARCH, http://research.gigaom.com/report/applyingbig-data-analytics-to-human-generated-data/, Jan.2014.

[6]. V. Hu, R. Kuhn, T. Xie, and J. Hwang, "Model Checking for Verification of Mandatory Access Control Models and Properties," International Journal of Software Engineering and Knowledge Engineering (IJSEKE) regular issue IJSEKE Vol. 21, No. 1.,2011.

[7]. Hadoop.apache.org

[8]. V. Hu and K. Scarfone, "Guidelines for Access Control System Evaluation Metrics," NIST Interagency Report 7874, Gaithersburg, MD, USA, 2012.

[9]. "The Big Data Security Gap: Protecting the Hadoop Cluster," White Paper, Zittaset, http://www.zettaset.com/wpcontent/uploads/2014/04/zettaset_wp_security_0413.pdf, 2014.

[10]. V. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone…., "Attribute Based Access Control Definition and Consideration," NIST Special Publication 800-162, Gaithersburg, MD, USA, 2013.

[11]. K.. T. Smith, "Big Data Security: The Evolution of Hadoop's Security Model," InfoQ, http://www.infoq.com/articles/HadoopSecurityModel, Aug. 2014.

[12]. "Apache Accumulo,"https://accumulo.apache.org

[13]. Hbase.apache.org

[14]. "NoSQL Databases Explained," mongoDB Inc., http://www.mongodb.com/nosql-explained, 2014.

[15]. W. Zeng, Y, Yang, B. Luo, "Access Control for Big Data using Data Content," in Proc. 2013 IEEE International Conference on Big Data, 2013.

[16]. S. Shea, "CSA top 10 big data security, privacy challenges and how to solve them," TechTarget, SearchSecurity, Nov. 2013.