https://gscjournal.com/IJASDR

International Journal of Agribusiness and Sustainable Development Research
Vol. 1, No. 1, January 2024,
pp. 38–45 || EISSN XXXX-XXXX

Data Access Control In High-Performance Computing: Preventing Unauthorized Access To Sensitive Data In Shared Clusters

Mfon O. Esang
Dept. of Computer Science Federal Polytechnic, Ukana
mfonesang@yahoo.com

Engr. Imaobong O. Akpan
Dept. of Mechanical Eng. Federal Polytechnic, Ukana
imaobongokpongette@gmail.com

Tope G. Jimoh
Dept. of Computer Science Federal Polytechnic, Ukan
jimohtope2012@gmail.com

Habeeb Ramoh Ajibola
Dept. of Computer Science Federal Polytechnic, Ukana

Engr Ekerette Dan
Dept. of Mechanical Eng. Federal Polytechnic, Ukana

Received: 2023 15, Nov
Accepted: 2023 18, Dec
Published: 2024 19, Jan

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract. High-Performance Computing (HPC) clusters are increasingly relied upon for processing large datasets, many of which contain sensitive information critical to research, industry, and government applications. Ensuring that access to such data is tightly controlled paramount to prevent unauthorized access, data breaches, and privacy violations. This paper explores various access control mechanisms and policies designed to secure sensitive data in shared HPC clusters. This paper also discusses the challenges, Control policy, and emerging technologies in the field of data access control for HPC, providing preference result for researchers, system administrators, and organizations operating HPC environments.

Key words: Data Access, Machine learning, Control Mechanisms, HPC clusters, Access Control Policies, High-Performance Computing, Emerging Technologies

1.INTRODUCTION

High-Performance Computing (HPC) clusters are essential tools for scientific, engineering, and data-intensive applications. Cloud HPC are computer clusters located in the cloud which addresses complex computational requirements, support applications with significant processing time requirements, or require processing of significant amounts of data (C. Vecchiola, S. Pandey and R. Buyya, 2009). HPC in many cases do not use virtualization, but rather an architecture based around docker containers and the unique method that the technology has in sharing resources (M. Eldred, et al, 2015). The advantages of cloud HPC is that they offer the potential to enhance innovation, as users can benefit from the emerging HPC ecosystem that is industry diagnostic and one that continually upgrades. This allows for organisations to quickly try new scientific approaches, with the elasticity of scaling up and down the infrastructure as required (M. Eldred and R. Mcavey, 2015). However, (M. Eldred, et al 2015) the suitability of Cloud solutions for complex HPC projects is less well understood, particularly the decision-making processes covering the practicalities, addressing the needs of key stakeholders' groups and commercial imperatives.

As their usage grows, so does the need to protect sensitive data from unauthorized access. It is well recognized that on high-performance computing (HPC) systems, the compute has become increasingly cheaper as compared to the storage and I/O (Foster, et al 2017). This architectural trend has continued to motivate and drive the HPC community to look for alternative solutions that allow data analysis to be done efficiently, rather than post-processing scientific data at persistent storage (Dan Huang, 2022) which expose the data to intruder. Data breaches can have devastating consequences, ranging from the loss of proprietary information to violations of privacy regulations. This paper delves into the critical aspect of data access control in HPC environments, highlighting the importance of implementing effective access control mechanisms and policies.

2. **Literature Review**

Challenges in Data Access Control for HPC:

According to (Al-Jody, T., 2021) there has been average progress done in researching security challenges in HPC portals. Considering that an attacker might attempt to alter multiple vectors such as firewall, system files and services, data access control and so on. Securing data in HPC clusters presents unique challenges. These challenges include:

a. Multi-user Environments: High performance computing is best achieved by parallelism. Parallelism (using parallel computers) is one of the best ways to overcome the speed bottleneck of single processors. Many computer systems supporting high performance computing have emerged like massively parallel processing systems (MPPs), symmetric multiprocessors (SMP), Distributed Systems, Cluster, and Grid. Their taxonomy is based on how their Processor, memory and interconnect are laid out (U. Chugh and A. Chugh, 2010). Cluster is a type of distributed processing system which consists of a collection of interconnected stand-alone computers working together as a single integrated computing resource. HPC clusters are often shared by multiple users and projects, necessitating access

control that can differentiate between users and their specific needs. Since the user can access from any place, they use the node machine to access their files and compute work. HPC (Gerndt, M.and Kranzlmüller, D., 2006) aims for receiving help to do complex and large computations in an environment where work is performed by many communicating computers on a single task which leads to an increased rank of security.

b. Performance Overheads: The paper (Chen et al., 2017) scrutinised the traditional barrel/bucket theory and proposed a new barrel theory. The theory consists of multiple barrels on the inside of the other with each of the barrels representing the following: "data security, system security, access control, application security, network security and physical security". Traditional access control mechanisms may introduce performance overhead in HPC workloads, requiring a balance between security and performance. Traditional IT security solutions, including network and host-based intrusion detection, access controls, and software verification work about as well in HPC as traditional IT (often not very), or worse, due to constraints in HPC environments. This traditional host-based security mechanisms, such as those leveraging system call data via audited, as well as certain types of network security mechanisms, like network firewalls and firewalls doing deep packet inspection, may be antithetical to the needs of the system being protected (Sean P., 2017).
c. Data Movement: One of the major security challenges foreseen is represented by the complex tracking of data movements where Pre-processing phase to prepare the data for research analysis (this may include partial or total data transfer towards the internal scratch area). Then Job processing on the computing facility, generating both intermediate and final data components and finally, Post processing phase to derive scientific results from the processed data: this typically includes the creation of a metadata catalogue allowing to index and quickly recover scientific data (Dirk P., 2021). Parallel and distributed file systems used in HPC environments are not yet fully able to account and log internal data movements. Data access control must account for data movement within the cluster, between nodes, and to external storage systems.

Access Control Mechanisms:

Users should be granted different levels of privileges to the system resources by access control. Through this mechanism, the administrator can secure the system and prevent the leaking of sensitive information. A strong access control mechanism and careful management of users' accesses are needed for HPC systems (Sushil J., 2020). Weak access control in HPC systems can be exploited by existing attack strategies, such as Replay Attacks, Non-Control-Data Attacks, and Privilege Escalation Attacks. This results in broken confidentiality and integrity of the system. We explore various access control mechanisms suitable for HPC environments, including:

a. Role-Based Access Control (RBAC): The RBAC mechanism forms has enhanced security framework for Grids and Clouds that will allow for interoperability between technologies in the two domains. The mechanisms proposed an importance because the current lack of software tools and security standards in accessing distributed HPC systems and transporting Large Data Sets that can add immensely to overheads in data processing or data integration times. RBAC model allows dynamic user-role assignments and role-permission assignments

with respect to changes in environmental context such as user location, security of the network through which access is made or the state of resources. While addressing the important issue of scalability, RBAC does not address issues of information interaction and semantics of organization workflows. However, by virtue of being modular, RBAC allows more sophisticated access control models to be layered on it. It Assign roles and permissions to users based on their responsibilities and tasks.

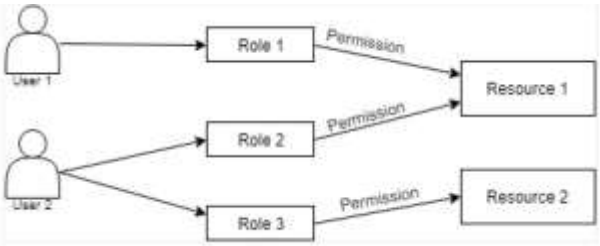

Figure 1: RBAC Sturcture

b. Attribute Based Access Control (ABAC): Unlike RBAC, the user controls (access permissions) are ensured with attributes and not roles. Thus, in the usage of Attribute-Based Access Control (ABAC), the user attributes play an important role. These attributes can access the general characteristics of the user such as age, height, personal characteristics, and can be altered again according to this information. The determination of the attributes is performed according to the topics. As the attributes are related to the information entered, they can be grouped into the following three categories: subject attributes (such as personal information), Resource attributes (information about the outcomes), and environment attributes (information about the environment) (Shen, H. B., and Hong, F., 2006). In this manner, the complexities that are rising in the growing and complicated systems are solved. Similar to RBAC, ABAC calls the permissions and limitations to action when they are needed.

c.

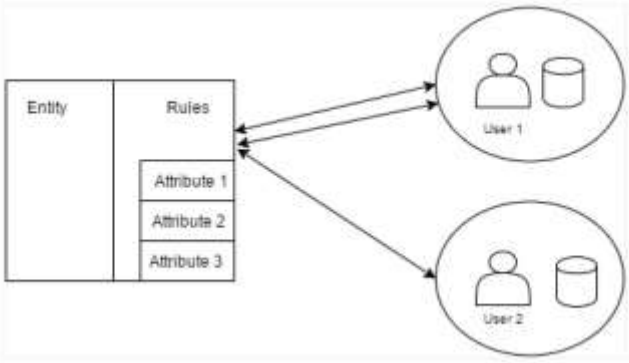

Figure 2: ABAC Structure

Mandatory Access Control (MAC): Enforcing strict access policies based on predefined security labels. Mandatory Access Control (MAC) surfaced when the experts decided that DAC is not efficient in terms of security since it is not being able to control every piece of information, which the Working model of the MAC scheme is shown in Figure 3. The decisions in MAC are not made by an owner but a central system (Khan, M. F. F., and Sakamura, K., 2015 and Msahli, M., et al., 2014). In this way, more powerful processes are performed in security models. The wholeness and privacy of the system are the

most important features after security. Thus, it can increase the security to the highest level in a whole system. Its flexibility is really low (Fan, Y., 2009). The reason for that is the system security, however, even this does not ensure the absolute privacy. It is used in government and military system improvements as a result of these features. Also, SElinux has a MAC mechanism that enables the security protocol to reduce the level of control over objects (Blanc, M., and Lalande, J. F., 2013). The most important feature that values this system above DAC is the fact that MAC distinguishes between the subject and object domains and allows the usage of permissions and limitations accordingly.

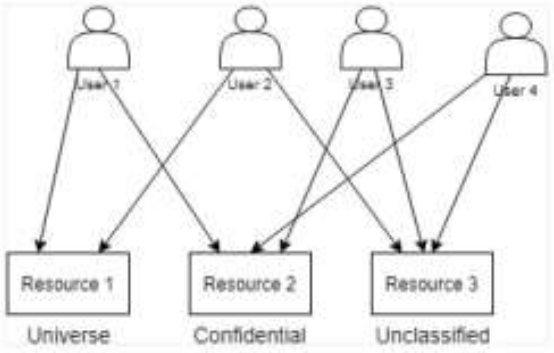

Figure 3: MAC structure

4. Methodology

Access Control Policies:

A system that controls access to services/resources made by cloud users based on authentication, authorization attributes of subjects, attributes of objects/resource as well as system attributes which conforms to policies. Each entity that is subject and object/resource is identified by its attribute (Subhash C. P., 2012). Effective access control requires well-defined policies. We discuss policy considerations such as:

a. Authentication and Authorization: To tackle the interoperability issue of distributed access control policies they propose an ontology-based access control. Interoperability between different schemes and models where a new semantic access control policy language (SACPL) is created to describe access control policies in cloud computing environments (L. Hu, (2009), S. Paraboschi, and P. Samarati, (2007) and C. Wang, (2010)). Authorization deals with the verification of an action that an entity can perform after authentication is performed successfully. In a grid, resource owners will require the ability to grant or deny access based on identity, membership of groups or virtual organizations, and other dynamic considerations. Authorization refers to a statement or combination of statements that prove a subject's right to access a resource, or to describe the act of validating that a subject has such a right. Implementing robust user authentication and authorization mechanisms to ensure only authorized users can access data.

b. Audit Trails: One of the most significant security issues is the lack of regular security audits (Rasheed, 2014). However, auditing requires personnel's familiarity with analysis tools and knowledge of the current vulnerabilities. For HPC systems, patches could hugely impact performance, as explained in section: (HPC Systems Performance and Security). The

importance of log auditing in strengthening security for HPC systems collecting logs from different services and system components is an imperative step for auditing security issues on a system (Vaarandi and Pihelgas, 2014). Logs will contain the activity of services, users and security events such as elevated commands. Maintaining detailed audit trails to track access and detect suspicious activities. Nevertheless, auditing the nodes independently ensures the most accurate policy will be obtained because each node can receive its specific policy module (M. Blanc and J.F. Lalande (2013)).

5**. Discussion**

Emerging Technologies:

We explore emerging technologies and trends that can enhance data access control in HPC, including:

a. Machine Learning and Anomaly Detection: By training a Machine Learning model to detect such behaviour, the model would be able to detect automated types of impersonation. The access control and anomaly detection methods for provenance analysis will deter some attackers, but do not eliminate the risk of attack – they merely increase the chance of attackers getting caught. Considering a Model; Bearicade, the Machine Learning models and data used to train and validate the models applied for HPC security were presented in to integrate into the Bearicade framework. Leveraging machine learning for detecting abnormal access patterns. Bearicade, which is based on a RESTful API, helps administrators and developers visualise and analyse activities over the systems, allowing machine learning deployment to aid in enhancing systems security based on the data gathered. Artificial Neural Network (ANN) could be used to analyse the data from collectors, then perform data classification, identify patterns, and detect anomalies (Al-Jody, T., 2021).
b. Blockchain for Provenance: Using blockchain technology to establish data provenance and audit trails. Use the blockchain to store the non-fungible signatures of the program and data. While this does not prohibit users from uploading fake data or algorithms or tampering with data and algorithms, we can trace the exact version used in a specific run. Control the access to provenance analysis and log the accesses for anomaly detection. Access control can be reinforced with blockchainmaintained logs and smart contracts. We can also build an anomaly detection subsystem, learning from the tamper-resistant provenance access log. The challenge is to develop an effective anomaly detection algorithm using the provenance access patterns. We can also prohibit access to the provenance data related to private components or their nearby components, which will significantly reduce the utility of provenance data. Blockchain applications are still in the embryonic stage. The cost of using current public blockchains is too high to be practical, while permissioned blockchains, such as HyperLedger, will need users to trust the management peers.

6. **Result:**

This paper provides a comprehensive overview of data access control in HPC, emphasizing its significance and the various tools and strategies available for safeguarding sensitive data. It combines theoretical insights with practical recommendations to guide researchers and practitioners in improving the security posture of shared HPC clusters.

HPC in the cloud can benefit many users who cannot own or access on-premise HPC resources. The fact that HPC systems tend to be used for very distinctive purposes, notably mathematical computations, may mean the regularity of activity within HPC systems can benefit the effectiveness of machine learning analyses on security monitoring data to detect misuse of cycles and threats to computational integrity Bblockchains and similar technologies may provide means for both monitoring the integrity of raw scientific data in HPC contexts, as well as for maintaining secure audit trails of accesses to or modifications of raw data.

7. **Conclusion:**

Data access control is a critical aspect of securing sensitive data in shared HPC clusters. Implementing access control mechanisms and policies can mitigate the risk of unauthorized access and data breaches. This paper has explored the challenges, mechanisms, policies, and emerging technologies in data access control for HPC, providing valuable insights for those responsible for managing and securing HPC environments.

REFERENCES

1. M. Eldred, A. Good and C. Adams (2015). A case study on data protection and security decisions in cloud HPC, IEEE 7th International Conference on Cloud Computing Technology and Science.
2. C. Vecchiola, S. Pandey, R. Buyya, (2009) High-Performance Cloud Computing: A View of Scientific Applications, Pervasive Systems, Algorithms, and Networks, 10th International Symposium on Pervasive Systems, and Networks.
3. M. Eldred, R. Mcavey (2015), Prepare for a Quantum Shift in Upstream Modelling, Gartner.
4. I. Foster and et al., (2017). "Computing just what you need: Online data analysis and reduction at extreme scales," in Euro-Par 2017: Parallel Processing. Springer International Publishing, 2017
5. Dan Huang, Zhenlu Qin, Qing Liu, Norbert Podhorszki, Scott Klasky, (2022), A Comprehensive Study of In-Memory Computing on Large HPC Systems, https://www.sciencedirect.com/science/article/pii/S0743731522000387
6. Gerndt, M., Kranzlmüller, D. (eds.) (2006), HPCC 2006. LNCS, vol. 4208, p. 938. Springer, Heidelberg
7. Urvashi Chugh and Amit Chugh, (2010), Security in High Performance Computing, Springer-Verlag Berlin Heidelberg, pp. 552–556
8. Al-Jody, Taha (2021) Bearicade: A Novel High-Performance Computing User and Security Management System Augmented with Machine Learning Technology. Doctoral thesis, University of Huddersfield.
9. Chen, Y., Zheng, Q., and Yang, H. (2017). A kind of university HPC platform security balance method based on the barrel theory. In Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017, pages 3708–3713. Institute of Electrical and Electronics Engineers Inc.
10. Dirk Pleitera, Sebastien Varretteb, Ezhilmathi Krishnasamyb, Enver Özdemirc, Michał Pilcd (2021). Security in an evolving European HPC Ecosystem. Partnership for Advanced Computing in Europe. Available online at www.prace-ri.eu
11. Sean Peisert (2017). Security in High Performance Computing Environments. Reviewed article on Exploring the many distinctive elements that make securing HPC systems much different than

securing traditional systems. COMMUNICATIONS OF THE ACM, vol 60, No. 9.

12. Sushil Jajodia, George Cybenko V.S. Subrahmanian, Vipin Swarup Cliff Wang, Michael Wellman (2020). Adaptive Autonomous Secure Cyber Systems. © Springer Nature Switzerland AG. ISBN 978-3-030-33431-4 ISBN 978-3-030-33432-1 (eBook) https://doi.org/10.1007/978-3-030-33432-1
13. Subhash Chandra Patel, Lokendra Singh Umrao, Ravi Shankar Singh (2012). Policy- based Access Control in Cloud Computing. Conference: International conference on Artificial Intelligent and Soft Computing. https://www.researchgate.net/publication/236578717
14. M. Blanc, J.-F. Lalande (2013) Improving Mandatory Access Control for HPC clusters. Future Generation Computer Systems 29 (2013) 876–885. www.elsevier.com/locate/fgcs
15. Al-Jody, Taha (2021) Bearicade: A Novel High-Performance Computing User and Security Management System Augmented with Machine Learning Technology. Doctoral thesis, University of Huddersfield.
16. Keke Chen (2022). Confidential High-Performance Computing in the Public Cloud. IEEE Computer Society.
17. L. Hu, S. Ying, X. Jia, and K. Zhao, "Towards an approach of semantic access control for cloud computing," LNCS Cloud Computing, vol. 5931, p. 145156, 2009. 28. S. D. Capitani, S. Foresti, S. Jajodia,
18. S. Paraboschi, and P. Samarati,"Over-encryption: Management of access control evolution on outsourced data," in Proceeding at International Conference on VeryLarge Databases, 2007. 29. S. Yu,
19. C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud omputing," in 29th IEEE International Conference on Computer Communications, 2010
20. Shen, H. B., and Hong, F. (2006). An attribute-based access control model for web services. In Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies, 2006. PDCAT'06. (pp. 74–79).
21. Khan, M. F. F., and Sakamura, K. (2015). Fine-grained access control to medical records in digital healthcare enterprises. In International Symposium on, Networks, Computers and Communications (ISNCC), 2015 (pp. 1–6). IEEE.
22. Msahli, M., Chen, X., and Serhrouchni, A. (2014). Towards a finegrained access control for cloud. In IEEE 11th International Conference on, e-Business Engineering (ICEBE), 2014 (pp. 286–291).
23. Fan, Y., Han, Z., Liu, J., and Zhao, Y. (2009). A mandatory access control model with enhanced flexibility. In International Conference on Multimedia Information Networking and Security, MINES'09. (Vol. 1, pp. 120–124). IEEE
24. Blanc, M., and Lalande, J. F. (2013). Improving mandatory access control for HPC clusters. Future Generation Computer Systems, 29(3), 876–885.