A Comparative Study on the Security Issues in Distributed Systems – A focus on the Security Implementation

Darius Bwalya
Directorate of Information Cmmunication and Technology
The Copperbelt University
Kitwe, Zambia
darius.bwalya@cbu.ac.zm

Victor Kasonde
Directorate of Information Cmmunication and Technology
The Copperbelt University
Kitwe, Zambia
vkasonde@cbu.ac.zm

***Abstract*— This study aimed to assess and compare the security challenges encountered by different distributed systems. The investigation focused on four frequently used distributed systems and examined the underlying technologies, security concerns, and suggested resolutions for these issues. The study culminated in a comprehensive overview of the security challenges and their potential solutions.**

Keywords—distributed systems,grid computing, security issues, reliability, performance.

I. Introduction

Since the beginning, various explanations of Distributed Systems have been put forth. One definition, as stated by [1], describes it as "a system where hardware or software components located on networked computers communicate and coordinate their actions solely through message passing." Tanenbaum, on the other hand, defines it as "a collection of independent computers that appear to users as a single computer." Based on Tanenbaum's definition, it can be deduced that a distributed system is primarily concerned with the software system rather than the hardware components that are utilized to construct it [2]. By integrating fore going definitions, it can be inferred that a distributed system refers to an application that collaborates with dispersed software and hardware components, to control various processes that run on autonomous computers connected via a communication network. This coordination enables all hardware and software elements to work cohesively towards a shared objective, by executing a series of interrelated tasks.

Ref [3] notes that there is a common misconception among many people that a distributed system and a network of computers are interchangeable terms. However, although there are similarities between the two, they have different meanings. A computer network is a collection of independent computers that communicate and interact with each other [4]. When using a computer network, a user is aware that they are utilizing various resources located on different computers. This is because a computer network does not conceal the fact that there are multiple computers involved. On the contrary, a distributed system creates the illusion that the user is utilizing a single, high-capacity autonomous computer with ample resources. The fact that multiple independent computers are cooperating to achieve a common goal is concealed from the user by the distributed system's middleware. This software operates on the computers, chooses appropriate machines, and delegates tasks without requiring any direct involvement from the user [2] [5].

II. Features of the Distributed Systems

The objective of a distributed system is to obscure the existence of numerous independent computers by functioning as an overlay network over physical network components. It is perceived as a unified entity that offers the user with various necessary services. On the other hand, a network serves as a means of linking entities, such as computers and devices, allowing for message exchange via standardized protocols between these clearly identifiable entities [6]. The key characteristics of a distributed system are as follows:

A. Distribution Focus

The utilization of a distribution-centric approach facilitates inherent decentralization of entities, including information, people, and systems. For instance, various individuals could generate and maintain different sets of information. Furthermore, information may be produced, stored, manipulated, and consumed by different applications or systems, which may or may not have knowledge of the other entities within the system.

B. Functional Separation

Functional separation is determined by the functionality, capabilities, and purposes provided by each entity within the system.

C. Reliability

Long-term and large-scale data storage and backup (replication) across various distributed locations.

D. Scalability

A distributed system has the ability to expand its resources to increase performance or availability.

E. Economy

In a distributed system, interconnected entities share resources to reduce ownership costs. These features enable the system's diverse entities to function concurrently and autonomously. Actions in a task are carried out independently and are managed through message exchanges in well-defined stages [7]. Moreover, the component entities are diverse, and failures are contained. Overall, there is no single process or entity that possesses complete knowledge of the system's state.

III. Objective of the Distribute Systems

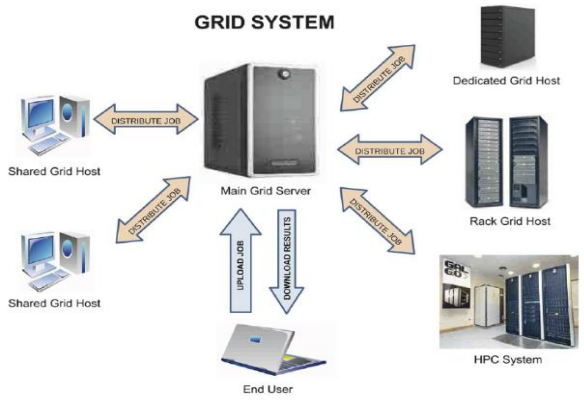
The foundation of all distributed systems are based on the objective of attaining Openness, transparency, Scalability, Performance and Reliability. Security is crucial in achieving the aforementioned objectives and demands sufficient attention as it represents a fundamental issue in distributed systems [8]. For instance, a large video file or database may be dispersed across multiple distributed storage locations. It is essential to be attentive at every stage of the design, implementation, operation, and management of distributed systems.

IV. Description of Distribution Systems and their Associated Issues

This section attempts to discuss the various distribution systems and their security related issues. The major systems considered are Grid Computing, Distributed Storage Computing[9], Cluster Computing, and Distributed Database. Below is a brief description of each:

A. Grid Computing

Rajkumar and Srikumar's (2015) defined a Grid in reference to a type of distributed computing system in which a large number of small, loosely connected computers are combined to form a virtual supercomputer. A Grid is a parallel and distributed system that can dynamically select, share, and aggregate geographically dispersed resources based on their availability, capabilities, performance, and cost to meet users' Quality of Service (QoS) requirements.



Grid computing source[21]

Grid computer systems offer multiple security mechanisms to safeguard grid resources against attacks. Middleware, which provides the common communication infrastructure and makes grid services available to applications, is a crucial system software in the grid infrastructure. Middleware plays a crucial role in ensuring security in Grid computing by facilitating a consistent security configuration at the messaging or service container level [10]. Grid authentication, based on Public Key Infrastructure (PKI), can handle various user credentials like PKI, SAML, Kerberos tickets, passwords, and more [11].

B. Distrubuted storage Computing

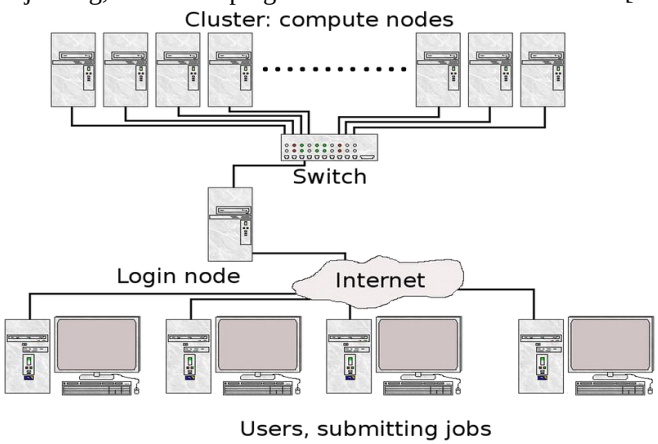
Ongoing research is focused on developing security models and threat modeling techniques to protect distributed storage systems. The critical resource in these systems is the data stored in the storage devices, which must be appropriately labeled and protected. It is also essential to ensure that any introduced protection system is backward compatible, meaning that it should protect not only the data stored after the security scheme is installed but also the data that existed before the introduction of the scheme [10].

The CIAA threat model, which encompasses Confidentiality, Integrity, Availability, and Authentication, has been proposed to address security concerns. This model categorizes the threats to a distributed storage system under each of the CIAA security pillars and proposes strategies to mitigate them.

C. Cluster Computing

In the realm of cluster computing, a set of comparable independent computers or personal computers are linked through a fast local network. These host machines operate on an identical operating system. Nevertheless, as soon as the computing clusters are made accessible to the general public or connected to networks over the Internet, they are vulnerable to a range of cybersecurity risks. The most prevalent attacks on clusters include service disruption, computation-cycle

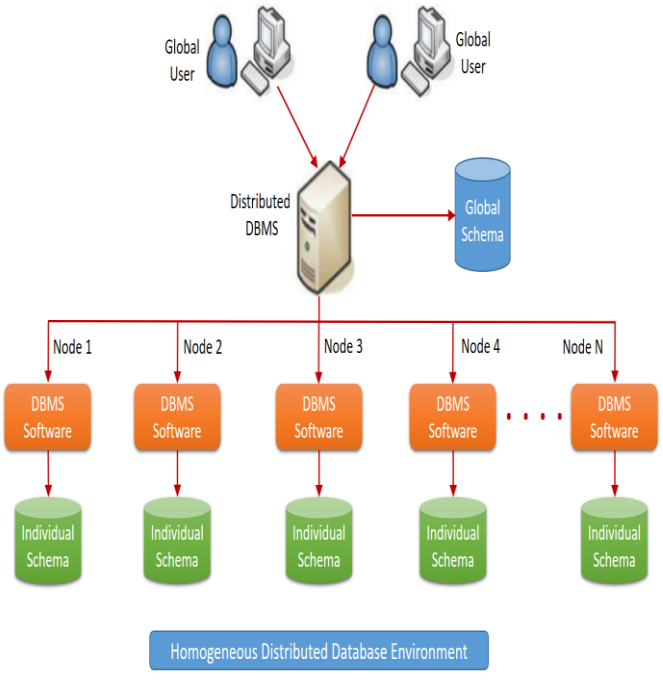
hijacking, and snooping on inter-node communication[13].



Cluster Computing source [22]

To safeguard clusters against cyber attacks, several security measures are utilized including identification, authentication, authorization, integrity verification, and confidentiality. These mechanisms are primarily aimed at preventing unauthorized access and meeting the security demands of applications. To this end, Xie and Qin have introduced two resource allocation strategies, TAPADS (which incorporates deadline and security constraints) and SHARP (which caters to security and heterogeneity concerns) to enable parallel applications to operate on computing clusters while satisfying security requirements and fulfilling execution timelines[9].

D. Distributed Database



The security of distributed database management systems is more susceptible to threats than centralized database systems. With the emergence of various database models, such as the temporal, object-relational, and object-oriented database models, implementing security for distributed database systems has become a more complex challenge.

To enhance the security of distributed database systems, a multilevel secure database system assigns security levels to each transaction and data. Unlike in a regular security model, where all the data and users belong to the same security level, a multilevel secure database system ensures that a transaction is cleared by the security level assigned to it, and the classification level of data is determined by its classification level. Zubi[20] proposed a design for distributed database systems that incorporates multi-level access control, confidentiality, reliability, integrity, and recovery mechanisms with the aim of enhancing accessibility, scalability, and flexibility of various types of data.

METHODOLOGY

Comparative Analysis

The comparative analysis refers to a technique used to contrast similar items with one another, identifying both their shared features and differences.

E. Comparative analysis is a versatile tool used in various fields and disciplines to gain a better understanding of the similarities and differences inherent in products.

F. Employing comparative analysis can aid businesses in making informed decisions regarding critical issues.

G. One important way that it is used is when it is applied to scientific data. Scientific data definition pertains to information that has been collected through scientific research to be used for a specific purpose.

The approach was to review from various studies the four(4) types of distributed systems. To what extent they each achieved the security features of the distributed system. Namely

A. Security
B. Performance
C. Reliability
D. Scalability
E. Transparency
F. Openness

A deeper comprehension of what each type is, its merits and demerits are from literature. Then we sought to find what the most common emphasis and concern was.

V. COMPARISONS OF DISTRIBUTED SYSTEMS

TABLE I. TABULAR COMPARISONS OF THE DISTRIBUTED SYSTEMS

Feature	*Cluster Computing [14]*	*Grid Computing[15]*	*Distributed Storage Systems [16]*	*Distributed Databases [16]*
Security	Good	Good	Excellent	Excellent
Performance	Excellent	Excellent	Good	Excellent
Reliability	Excellent	Good	Excellent	Excellent
Scalability	Not very scalable	Scalable	Scalable	Scalable
Transparency	Not so good	Good	Good	Excellent
Openness	Good	Good	Good	Good

VI. Critic of the Security Issues in Distributed Computing

The above discussion on distributed systems highlights the increased importance of security when systems are distributed across multiple geographic areas. Each type of distributed system has its own unique security requirements, but they all share common security needs, such as the Confidentiality, Integrity, and Availability (CIA) triad, which forms the foundation of any security implementation.

Cluster and grid computing share similar security concerns such as securing data in transit and ensuring accessibility to distributed resources. However, securing a cluster computer is relatively simpler due to its homogeneous nature compared to grid computing [17]. Denial of Service (DoS) attacks are one of the main threats to clusters, and scholars have proposed novel methods based on Markov chains to mitigate their impact. On the other hand, the middleware layer in grid computing provides a platform for implementing security measures across the entire system [15]. Grid systems use robust security measures based on Public Key Infrastructure (PKI) and X.509 certificates. However, authentication issues in grid computing have not been thoroughly addressed.

In addition, security measures are also put in place to prevent unauthorized access and ensure the confidentiality, integrity, and availability of the stored data. Some of the commonly used security mechanisms include access control, encryption, backup and recovery, and data validation techniques. To enhance the security of distributed storage systems, researchers have proposed different security models such as the CIAA model, threat modeling, and multilevel security models. These models provide a framework for identifying potential threats and vulnerabilities and developing strategies to mitigate them.

Compared to distributed storage systems, implementing security in distributed database systems has become more complex due to the presence of various database models[18]. Studies conducted by scholars and researchers indicate that the security of distributed databases can be enhanced by implementing multilevel security protocols that rely on military information classification and access control mechanisms [19].

VII. Discussion and Conclusion

The analysis of various scholars and results above indicate that security becomes a more critical requirement whenever systems become distributed across several multi-geographical locations. The results from literature and table 1 also reinforces the general common systems need for Confidentiality, Integrity and Availability (CIA) triad as the heart of any security implementation.

It can be inferred that the main goal of the study was to analyze and compare the security challenges encountered in different types of distributed systems. The study provided a detailed discussion of the various distributed systems and their objectives, followed by an in-depth examination of four commonly used systems, including their security concerns and solutions proposed by researchers. Furthermore, the study aimed to compare the characteristics of the different distributed systems presented and evaluate their impact on the performance of the distributed systems.

References

[1] [1] M. van Steen and A. S. Tanenbaum, “A brief introduction to distributed systems,” Computing, vol. 98, no. 10, pp. 967–1009, Oct. 2016, doi: 10.1007/s00607-016-0508-7.

[2] [2] R. Subramanian and B. D. Goodman, Eds., Peer-to-peer computing: the evolution of a disruptive technology. Hershey, PA: Idea Group Pub, 2005.

[3] [3] Y. S. Jghef and S. R. M. Zeebaree, “State of Art Survey for Significant Relations between Cloud Computing and Distributed Computing,” Nov. 2020, doi: 10.5281/ZENODO.4237005.

[4] [4] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, “Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination,” Comput. Netw., vol. 101, pp. 158–168, Jun. 2016, doi: 10.1016/j.comnet.2015.12.017.

[5] [5] Department of Computer Science, HRD Degree and P. G College, Narayanaguda, Hyderabad, Telangana, India, D. Harinath, P. Satyanarayana, and M. V. R. Murthy, “A Review on Security Issues and Attacks in Distributed Systems,” J. Adv. Inf. Technol., pp. 1–9, 2017, doi: 10.12720/jait.8.1.1-9.

[6] [6] M. T. Hammi, B. Hammi, P. Bellot, and A. Serhrouchni, “Bubbles of Trust: A decentralized blockchain-based authentication system for IoT,” Comput. Secur., vol. 78, pp. 126–142, Sep. 2018, doi: 10.1016/j.cose.2018.06.004.

[7] [7] S. G. Bhagwath and Dr. Mallikarjun Math, “Distributed Systems and Recent Innovations: Challenges Benefits and Security Issues in Distributed Systems,” Bonfring Int. J. Softw. Eng. Soft Comput., vol. 6, no. Special Issue, pp. 37–42, Oct. 2016, doi: 10.9756/BIJSESC.8239.

[8] [8] A. Luther, R. Buyya, R. Ranjan, and S. Venugopal, "Peer-to-Peer Grid Computing and a .NET-Based Alchemi Framework," in High-Performance Computing, L. T. Yang and M. Guo, Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006, pp. 403–429. doi: 10.1002/0471732710.ch21.

[9] [9] C. Li, J. Shen, and Y. Liu, "Study on Architecture and Implementation of Port Logistics Information Service Platform Based on Cloud Computing," Int. J. Future Gener. Commun. Netw., vol. 8, no. 2, pp. 331–342, Apr. 2015, doi: 10.14257/ijfgcn.2015.8.2.27.

[10] [10] M. Firdhous, "Implementation of Security in Distributed Systems - A Comparative Study," 2012, doi: 10.48550/ARXIV.1211.2032.

[11] [11] Awad Haroon Ballaith, Faraaz Hussain, Mohammed Faizan Hussain, and Lords Insitute of Engineering and Technology, "Use of Security in Disseminated Frameworks," Int. J. Eng. Res., vol. V6, no. 02, p. IJERTV6IS020280, Feb. 2017, doi: 10.17577/IJERTV6IS020280.

[12] [12] M. Levine, K. E. Lentz, J. R. Kahn, F. E. Dorer, and L. T. Skeggs, "Partial purification of a high molecular weight renin from hog kidney," Circ. Res., vol. 38, no. 6 Suppl 2, pp. 90–94, Jun. 1976, doi: 10.1161/01.res.38.6.90.

[13] [13] D. Golightly and N. Dadashi, "The characteristics of railway service disruption: implications for disruption management," Ergonomics, vol. 60, no. 3, pp. 307–320, Mar. 2017, doi: 10.1080/00140139.2016.1173231.

[14] [14] J. Yu, J. Wu, and M. Sarwat, "GeoSpark: a cluster computing framework for processing large-scale spatial data," in Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle Washington, Nov. 2015, pp. 1–4. doi: 10.1145/2820783.2820860.

[15] [15] J. J. Cuadrado-Gallego and Y. Demchenko, Eds., The Data Science Framework: A View from the EDISON Project. Cham: Springer, 2020.

[16] [16] P. Viotti and M. Vukolić, "Consistency in Non-Transactional Distributed Storage Systems," ACM Comput. Surv., vol. 49, no. 1, pp. 1–34, Jul. 2016, doi: 10.1145/2926965.

[17] [17] M. Dawson, M. Eltayeb, and M. Omar, Eds., Security Solutions for Hyperconnectivity and the Internet of Things: IGI Global, 2017. doi: 10.4018/978-1-5225-0741-3.

[18] [18] R. Hasan, S. Myagmar, A. J. Lee, and W. Yurcik, "Toward a threat model for storage systems," in Proceedings of the 2005 ACM workshop on Storage security and survivability - StorageSS '05, Fairfax, VA, USA, 2005, p. 94. doi: 10.1145/1103780.1103795.

[19] [19] V. Kasonde and C. Chembe, "An Assessment of the Best Practices in the Implementation of Data Centers (A case of Zambia)," Zamb. ICT J., vol. 5, no. 1, pp. 8–15, Jun. 2021, doi: 10.33260/zictjournal.v5i1.92.

[20] Firdhous, Mohamed. "Implementation of security in distributed systems-a comparative study." arXiv preprint arXiv:1211.2032 (2012).

[21] Jimsgn. "Grid Computing." Accesed 30/12//2022 https://jimsgnblog.blogspot.com/2017/10/grid-computing.html April 18, 2017

[22] Google, "Cluster Computing", Accessed 30/12/2022 URL https://sites.google.com/site/cloudcclusterc/home/cluster-computing

[23] R. M. R. Pattamsetti," Distributed Computing in Java 9" Accessed 30/12/2022, O'Reilly URL https://www.oreilly.com/library/view/distributed-computing-in/9781787126992/4a10acc9-db8f-425d-9a95-f3320399663b.xhtml