JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN ENHANCEMENT IN BIG DATA SECURITY AND PERFORMANCE BY USING FOG COMPUTING

[1]Ms. Monika Saini, [2]Dr. Gaurav Aggarwal,
[1]Research Scholar, [2]Research Supervisor Professor,
[1]Department of Computer Science & Engineering, [2]Department of Computer Science & Engineering
[1]Jagannath University, Bahadurgarh, India , [2]Jagannath University, Bahadurgarh, India

Abstract:

The research is focusing on providing innovative designs that would be used for security of big data. Several mechanisms could be applied to secure big data such as cryptography techniques and data encoding mechanisms. Compression of big data might improve the performance during encryption operation. Big Data analysis with cryptography helps in increasing security. Fog computing extends cloud capabilities to the network's edge, making it crucial to optimize data processing and delivery for low latency and real-time responsiveness. Strategies for improvement include data compression and aggregation to minimize network traffic, content caching for quicker delivery, and edge analytics for local processing. Load balancing, distributed computing, and edge-based machine learning can parallelize and optimize data processing tasks. Quality of Service (QoS) management ensures data prioritization, while latency-aware routing selects optimal pathways. Robust network infrastructure and well-optimized edge devices are foundational and continuous monitoring, security, and privacy measures are essential for a holistic solution that efficiently manages and processes big data in a fog environment.

Keywords:

Big data, Cryptography, Security, Quality of service, Cloud computing, Fog computing and Networks.

1. Introduction

It's impossible to keep up with the digital revolution because of privacy and security measures and conventional security. Malicious users can re-identify and link anonymous data even with advanced encryption technologies, access limitations, firewalls, and intrusion detection systems for network security. There have been several new rules proposed to deal with specific problems. A person's privacy has been threatened by big data because of issues like inference and aggregation, which allow people to be re-identified even after identifying information has been removed from the dataset. A long-standing dilemma, however, is the security triangle. A restriction on raw data analysis and modification might harm system performance and ease of use if, for example, the regulation does not enable businesses to grow. All areas of the Big Data ecosystem, from infrastructure to management to confidentiality rules to integrity and quality, must be reevaluated and further examined to ensure data security and privacy. More study is needed to fully identify and solve these problems, however, this section provides an overview of some of the most pressing difficulties. This paper examined five different aspects of large-scale data storage and processing: the framework, the infrastructure, the monitoring and auditing processes, the key management, and the security of the data. As part of the proposed project, it is hoped that the security of data sets will be improved against SQL injection, which is difficult to manage. Big data analytics is used by businesses to uncover new ways to do business while also increasing efficiency and reducing the time it takes to make critical business decisions. Security is often overlooked in open-source big data systems. Because of the dramatic surge in data use, this has become a big issue. These problems can affect both onsite and offsite systems. Aside from that, damage is done to the cloud[1]. As you deal with massive amounts of data, the most prevalent problems are described below.

- **Scale and Volume:** The sheer volume of big data being transmitted requires efficient and scalable security measures[1]. Traditional security solutions may struggle to keep up with the demands of securing large datasets.
- **Innovation and Research:** As technology evolves, new threats and vulnerabilities emerge. Research into innovative security approaches for big data transmission is crucial to stay ahead of potential risks.
- **Non-Relational Databases[2]:** Traditional relational databases are built around rows and columns. It's because of this that they're unable to deal with large numbers and different forms. Because of the inadequacies of relational databases, no SQL databases, also known as non-relational database systems, were developed. Non-relational databases have no rows or columns. Storage structures for different types of data in No SQL databases are a huge advantage. No SQL databases provide significantly more flexibility and scalability than relational databases. Performance and flexibility take precedence above security when it comes to non-SQL databases. Extra precautions must be taken by businesses that use No SQL databases to assure their safety.
- **Endpoint Vulnerabilities**: In the hands of cybercriminals[2], the data on endpoint devices may be manipulated and transferred to data lakes. Security solutions for endpoint log analysis must verify that the endpoints are legitimate.
- **Data Mining Solutions:** Many large-scale data situations need the application of data mining techniques. Data mining tools may find patterns in unstructured data. Financial and personal information are frequently combined, which creates a problem. To protect themselves from both external and internal dangers, firms are now bolstering their security measures.
- **Encryption**

Encryption solutions for big data are required when dealing with vast amounts of data, both at rest and in transit. Companies must encrypt both human and machine-generated data. No SQL databases and Hadoop[2] distributed file systems are only two examples of the many massive data storage solutions that are needed for encryption techniques.

- **User Access Control**

Network security relies heavily on the ability to restrict access to users. Large data systems might be devastated if access control methods are inadequate[2]. A strong policy of user control necessitates automated role-based settings and regulations. Policy-driven access control can manage several levels of user control, such as varied administrator settings, automatically.

- **Intrusion Detection and Prevention**

The scattered nature of big data makes it ideal for infiltrating organizations. An Intrusion Prevention System (IPS) that analyses network traffic can help protect large data platforms against attacks on their security. The intrusion prevention system (IPS)[3] is commonly placed behind the firewall to prevent any harm from being done by the incursion.

- **Centralized Key Management**

Preventing keys from being misused or lost is known as "key management"[3] in cryptography. The more dispersed or application-specific approach to key management is outperformed by a centralized strategy. Secure keys, audit logs and rules may all be accessed simultaneously in centralized management systems. An effective key management system is essential for firms that deal with sensitive information.

The motivation behind securing big data transmission over networks lies in protecting sensitive information, complying with regulations, ensuring business continuity, and maintaining public trust. Addressing these challenges requires a comprehensive and robust security framework that accounts for the unique characteristics of big data and the evolving threat landscape.

1.2. Fog computing

It is an architecture that extends cloud computing capabilities to the edge of the network, which can be especially valuable for applications that require low latency and real-time processing. Enhancing the performance of content in a fog environment, especially with big data, is a complex task that requires a combination of technological solutions and a thorough understanding of the specific use case and requirements. It's important to design and implement a holistic approach that addresses the unique challenges of your fog computing environment and leverages the advantages of edge computing while efficiently managing and processing big data. Enhancing the performance of content in a fog environment, particularly with big data, can be challenging due to various problems and issues. Here are some common challenges and concerns associated with this process:

Data Security[4]: Handling big data at the edge introduces security concerns, as sensitive data may be vulnerable to breaches or unauthorized access. It's essential to implement robust security measures to protect data in the fog environment.

Data Privacy[5]: Maintaining data privacy, especially when dealing with sensitive or personal information, is a significant concern. Compliance with privacy regulations like GDPR and HIPAA can be complex in a fog environment.

Data Quality: Ensuring data quality is crucial for accurate analysis and decision-making. Inconsistent or unreliable data from edge devices can lead to incorrect insights and actions.

Resource Constraints[7]: Edge devices often have limited processing power and storage capacity. Handling big data at the edge can strain these resources, potentially affecting performance and causing bottlenecks.

Scalability: The fog environment must be scalable to accommodate increasing data volumes and the addition of new edge devices.

Data Synchronization: Keeping data synchronized across edge devices and with the central cloud can be difficult. Inconsistencies in data can lead to incorrect decisions and actions.

Fault Tolerance: Fog environments must be designed to be fault-tolerant to handle device failures or network interruptions gracefully. Ensuring continuous operation in the event of failures is essential for reliability.

Interoperability: Integrating various edge devices, sensors, and platforms can be challenging due to the lack of standardized protocols and data formats. Ensuring interoperability and seamless data flow is crucial.

Latency Control: While fog computing aims to reduce latency, it's not always straightforward to achieve low latency consistently, especially in scenarios with a large number of edge devices and high data volumes.

Monitoring and Management: Monitoring and managing a distributed fog environment can be complex. Tools and strategies are needed to gain visibility into system performance, troubleshoot issues, and implement optimizations effectively.

Regulatory Compliance: Complying with industry-specific regulations and standards while handling big data in a fog environment is a critical concern, as violations can lead to legal and financial consequences[7].

2. Need of Research

Companies are increasingly relying on big data analytics to gain a competitive advantage. Big data architectures will allow attackers to take advantage of new methods. As a result, there are more and more serious data security issues to deal with. As a result, a comprehensive data security policy is required. It is important to remember that big data is subject to several privacy and regulatory laws[6]. Organizations and individual users frequently have no idea what happens to their data or where it is maintained. When given enough data, advanced analytics technologies may help develop new forms of security. For example, conclusions can be drawn from the combination of safety data from many systems. Today's ever-evolving cyber threats necessitate a reevaluation of security. Addressing these problems and issues requires careful planning, technology selection[7], and a holistic approach to designing and managing fog environments[8] for big data. Organizations must invest in research, development, and continuous improvement to overcome these challenges and reap the benefits of improved content performance in fog computing settings. Using the same tried-and-true procedures that have traditionally been used on other networks is not an option for ensuring the security of the large data since these approaches are no longer sufficient.. The traditional methods of encryption were not successful in preventing the loss of data. The only tier of the stack that the Tradition system protected was the application layer. The fact that there is any security at all on the session layer has been completely ignored. One channel is used for the transmission of all of the information. It would have been extremely risky if an unauthorized entity were able to decipher the large amounts of data. Because of this, cryptanalysis is able to decrypt it with relatively little effort[8]. As a result, the demand for large data security has sparked new research.

3. Methodology:

Here IP filter has been used to reject unauthenticated transmission of packets from server to client. If packet is valid then enhanced AES ENCRYPTION[8] module works. Here the process flow of data transmission has been proposed.

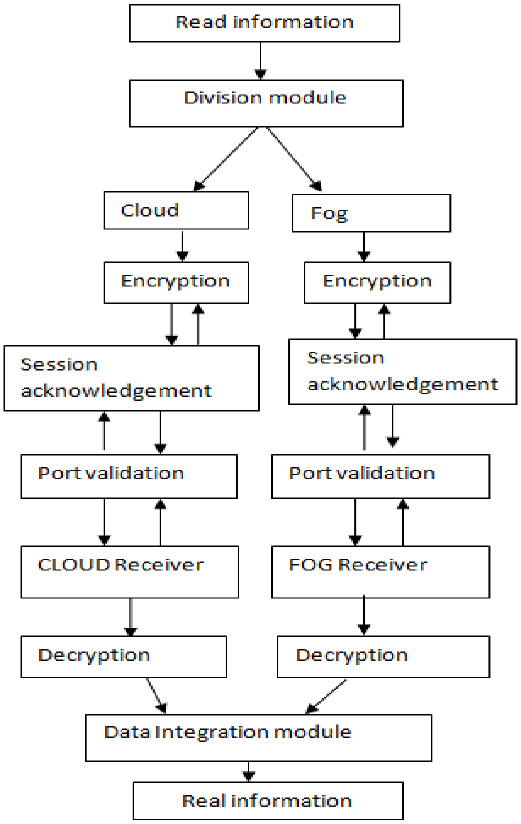

Fig 1 Proposed work

It has been explained here how the suggested work will be carried out and what the outcome will be. The FILE splitter[8] would separate the data into two separate data files, the first of which would be placed on the cloud, while the second would be placed on the fog. The name of the file as well as the authentication code is both supplied. The file is then split up and distributed in two distinct locations: one for the cloud, and the other for the fog. This increases the transmission's level of safety and dependability. File sender interface transmit data to server. User id[9], password[9], port number, IP address, and the location of the file that should be sent together with the security token and the AES CODE[15] are all entered here. The data would be sent to the server through the GUI Interface for Server[17] . Along with the security token, please enter the port number, AES CODE, and path of the file to be received below.

During the process of transmission, the contents of the file are in cypher text format. It is cypher text, which means it cannot be read. If someone were to hack into that information, however, they would not be able to comprehend[16] it in any way. The data would be sent from the cloud to the end user via the sender module. A secret authentication code would then be used to encrypt the data. The port utilized by the end user and the port on the sender's side would be identical. This is where the IP address of the end user would be supplied. The following illustration provides a visual representation of the data sender module design for the cloud. There are three boxes available for input. The first input box would receive the server's IP address, the second would receive the port number, and the third input box would obtain the authentication code to encrypt the data that is to be delivered.

4. Result and discussion

The end-user module has been broken down into three distinct sections. You are now ready to accept data from the fog. This step opens the port for the fog and enables data capture that is sent from the fog side. The port must be shared by both parties. The data that has been transmitted from the fog side[17] may have its encryption deciphered[18] if the same common authentication code were used.

This second part is prepared to accept data from the cloud open the port for the cloud and enable data capture and transmission from the cloud side. The port must be shared by both parties. The data that was delivered from the cloud side may then have its encryption decoded since the same common authentication code would be used.

This third part of the process would combine the data that was received and then decode it based on the authentication code.

Packets	Traditional	Proposed
100	4	2
200	8	2
300	9	5
400	11	6
500	13	6
600	17	7
700	29	10
800	38	19

Table 1 Comparative Analysis of Packet Dropping In Traditional and Proposed Work

Packets	Traditional	Proposed
100	90.54%	93.48%
200	90.20%	93.55%
300	90.67%	93.11%
400	90.80%	93.76%
500	90.37%	93.43%
600	90.90%	93.60%
700	90.43%	93.21%
800	90.27%	93.80%

Table 2 Comparative Analysis of Accuracy In Traditional and Proposed Work[17]

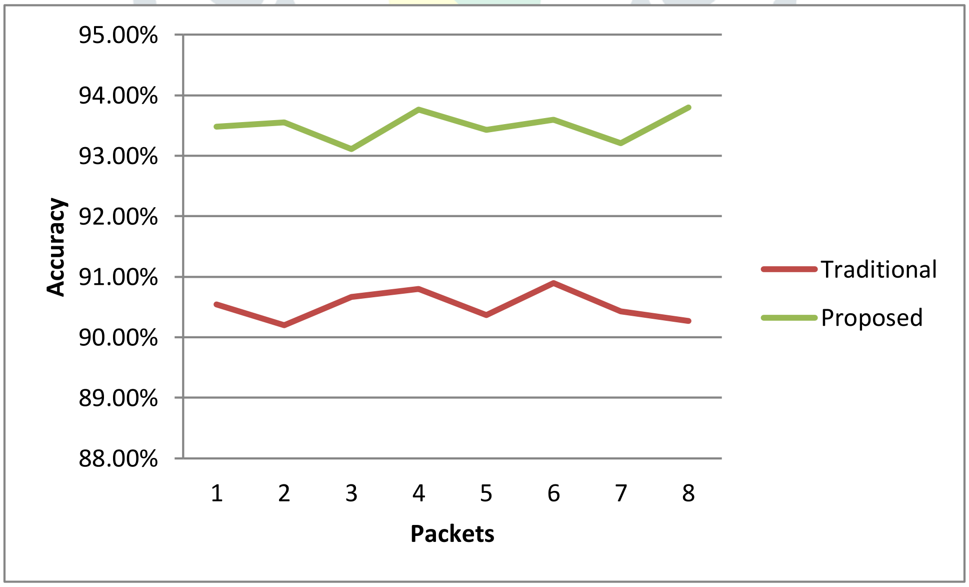

Fig 2 Comparative analysis of Accuracy in traditional and proposed work

5. Conclusion

This article offers a concise introduction to two of the most recent and exciting innovations in the realm of computing: cloud computing and fog computing. The area of research that focuses on cloud computing with the assistance of fog is still in its early stages. Because of this, there has to be a significant amount of research carried out in this field. In this study, some of the challenges and complications related to cloud computing that makes use of fog are explored. The purpose of this essay is, we hope, to shed some light on the nature of computing by providing information that readers will find helpful. To protect data at the application layer from both active and passive types of attack, a more secure approach has been devised thanks to the suggested implementation, which was described before. Comparative research has been done between the proposed strategy and the prevalent security paradigm. It has been demonstrated that there is a significantly reduced risk of packet loss when using the proposed approach as opposed to the one that is currently in use. It has been determined that traditional security measures are insufficient.

REFERENCES

[1] Kumar, R. and Goyal, R., 2019. On cloud security requirements, threats, vulnerabilities and countermeasures: A survey. *Computer Science Review*, *33*, pp.1-48.

[2] P. Sareen, "Cloud Computing : Types, Architecture, Applications, Concerns, Virtualization and Role of IT Governance in Cloud," vol. 3, no. 3, pp. 533–538, 2013.

[3]. Naveen Rishishwar, Vartika, Mr. Kapil Tomar, 2017. Big Data: Security Issues and Challenges. International Journal of Technical Research and Applications, e-ISSN: 2320-8163, Special Issue 42 (AMBALIKA), PP. 21-25.

[4] P.Pazowski , et al. "Cloud computing – a case study for the new ideal of the IS/IT implementation," in International Conference on Management, Knowledge and Learning, Zadar, Croatia, 2013, pp. 855—862

[5] M.Georgescu, et al. "The value of cloud computing in the business environment," The USV Annals of Economics and Public Administration, vol.13, no.1, pp. 222--228, 2013.

[6] B.H. Bhavani, et al. "Resource provisioning techniques in cloud computing environment: A survey," International Journal of Research in Computer and Communication Technology, vol.3, no.3, pp. 395--401, 2014.

[7] Mohamed Firdhous, et al. (2014)"Fog Computing: Will it be the Future of Cloud Computing?" ISBN: 978-1-941968-00-0 ©2014 SDIWC

[8] K.Shenoy et al. "Fog Computing Future of Cloud Computing", International Journal of Science & Research (IJSR) Volume 4 Issue 6, pp.55-56, June 2015.

[9] M.Verma, et al. "architecture for Load Balancing Techniques for Fog Computing Environment", I.J. Information Technology & Computer Science Volume 6 No.2, pp.269-274, April 2016, DOI:10.090592/IJCSC.2015.627.

[10]. Bushra Zaheer Abbasi,et al. (2017)"*Fog Computing: Security Issues, Solutions & Robust Practices",* Proceedings of 23rd International Conference on Automation & Computing, University of Hudders field,Hudders field, UK, 7-8September 2017

[11]Nabil Abubaker, et al. "Privacy-Preserving Fog Computing Paradigm" The 3rd IEEE Workshop on Security & Privacy in Cloud (SPC 2017)

[12] Jiyuan Zhou et al. (2017) "A Hierarchic Secure Cloud Storage Scheme based on Fog Computing", 2017 IEEE 15th Intl Conf on Dependable, Autonomic & Secure Computing, 15th Intl Conf on Pervasive Intelligence & Computing, 3rd Intl Conf on Big Data Intelligence & Computing & Cyber Science & Technology Congress

[13] K. Dolui, et al. "Comparison of Edge Computing Implementations: Fog Computing, Cloudlet & Mobile Edge Computing", IEEE, pp.1-6, 2017.

[14] Yunguo Guan, et al. "Data Security & Privacy in Fog Computing", 0890-8044/18/$25.00 © 2018 IEEE.

[15] Yashpal, et al. "Investigating Brute force and Timing Attack Immunity in Proposed Work", International Journal for Research Publication & Seminar Volume: 09 Issue: 04 July - September 2018.

[16] D. Bermbach et al., "A Research Perspective on Fog Computing," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10797 LNCS, no. November, pp. 198–210, 2018, doi: 10.1007/978-3-319-91764-1_16.

[17] A. Unnisa, "A Study on Review and Analysis of Cloud , Fog and Edge Computing Platforms," vol. 5, no. 11, pp. 325–331, 2020.

[18] H. Rashid Abdulqadir et al., "A Study of Moving from Cloud Computing to Fog Computing," Qubahan Acad. J., vol. 1, no. 2, pp. 60–70, 2021, doi: 10.48161/qaj.v1n2a49.

[19] M. Al Masarweh, T. Alwada'n, and W. Afandi, "Fog Computing, Cloud Computing and IoT Environment: Advanced Broker Management System," J. Sens. Actuator Networks, vol. 11, no. 4, 2022, doi: 10.3390/jsan11040084.