Internet of Things and Cloud Computing
2024, Vol. 12, No. 2, pp. 17-27
https://doi.org/10.11648/j.iotcc.20241202.11

Research Article

Assessing the Effectiveness of Multi-Factor Authentication in Cloud-Based Big Data Environments

Saroj Mali*

School of Computer Science and Engineering, Central South University, Changsha, China

Abstract

There is increasing popularity of Big data and cloud computing in recent years, and it is offering both individuals and businesses a number of advantages. But as data volume and complexity rise, data security and privacy have become a serious problem. In order to safeguard sensitive data stored in the cloud from sophisticated cyberattacks, it is crucial to have strong security measures in place. Although multi-factor authentication (MFA) has gained popularity as a security mechanism, Because of the lack of in depth analysis of its efficacy in large data systems based in the cloud is not fully known. In order to determine if MFA is effective in large data environments based on the cloud, this study will examine how well it can defend against different types of cyberattacks. The study will analyze the benefits and drawbacks of MFA in this situation as well as the trade-offs that must be made between security and usability when putting this security measure into place. This study aims to evaluate the efficacy of MFA in cloud-based big data environments in order to offer insightful recommendations for the most effective ways to secure sensitive data in the cloud.

Keywords

Cloud Computing, Big Data, Big Data Security, Multi Factor Authentication

1. Introduction

Data breaches are terrible occurrences for both businesses and customers. Information that should not be made public or stolen [1]. Because of the enormous amount of data that may be taken or altered, cloud-based data breaches can be extremely damaging. In the era of Big Data, cloud security is particularly crucial [2]. Data stored in the cloud is beneficial for a variety of reasons. Trade secrets, commercial connections, and intellectual property may be of interest to rival companies [3]. Access to Big Data that is hosted in the cloud may be desirable for activists who want to reveal "secret" or potentially embarrassing material. Or perhaps someone is just trying to extort you. The threat of hackers accessing files and systems stored in the cloud has grown significantly for businesses [4]. Although data breaches are not specific to Cloud storage, they nevertheless rank as a top concern for users of the service [5]. In addition to being exposed to the same dangers as a traditional corporate network, the Cloud can also be attacked using shared resources, Cloud personnel, and outside partners of the Cloud provider [4]. Cloud service companies are highly desirable targets since they have access to Big Data. The results of a data breach, regrettably, are not always restricted to a loss of privacy. Organizations are required to uphold specific standards designed to safeguard sensitive data from unauthorized use in numerous jurisdic-

*Corresponding author: malisaroj@csu.edu.cn (Saroj Mali)

Received: 25 July 2023; **Accepted:** 24 July 2024; **Published:** 6 August 2024

tions throughout the world. Companies may be subject to fines and civil litigation [6]. A criminal investigation could result in charges. The costs of the inquiry and damage restoration come next.

MFA is widely used and effective in traditional computing contexts, however it is unclear how well it works in cloud-based big data environments. Big data environments built on the cloud provide special security problems since they need to be protected from a larger variety of cyberattacks, including phishing, malware, and password attempts. Additionally, the implementation of MFA in these settings necessitates a security and usability trade-off that may have an impact on overall effectiveness and user adoption [7]. Consequently, the purpose of this study is to evaluate the efficacy of MFA in cloud-based big data environments by determining how well it can defend users from different types of online threats. The study will analyze the benefits and drawbacks of MFA in this situation as well as the trade-offs that must be made between security and usability when putting this security measure into place. This study aims to evaluate the efficacy of MFA in cloud-based big data environments in order to offer insightful recommendations for the most effective ways to secure sensitive data in the cloud.

1.1. Big Data and Cloud Computing

It's important to make a clear distinction between "Big Data" and "Cloud Computing" before discussing how the two work together. Although they are technically distinct terms, they are frequently used in conjunction in literature because of how well they work together. Big Data simply refers to extremely huge data sets that are produced by numerous programs. It can refer to any of a wide range of data kinds, and the data sets are typically too big to browse through or query on a standard computer [8].

Cloud computing is the term for processing anything in the "cloud," including big data analytics. The term "cloud" only refers to a collection of powerful servers from a single provider. They frequently have substantially faster viewing and querying speeds than a typical PC.

In essence, "Big Data" refers to the sizable sets of data gathered, and "Cloud Computing" refers to the system that remotely receives this data and carries out any specified operations on it. Big Data is also frequently produced by extensive, network-based systems. It may be in a conventional format or a non-conventional one. If the data isn't in a standard format, machine learning and artificial intelligence from the cloud computing provider might also be utilized to standardize the data. From there, the data may be accessed and used in a variety of ways thanks to the Cloud Computing platform. It can, for instance, be searched for, updated, and used for upcoming insights.

Big Data processing may be done in real-time thanks to this cloud architecture. It can instantly comprehend massive "blasts" of data from powerful systems. Another connection between big data and cloud computing is that big data analytics now take a fraction of the time they once did thanks to the power of the cloud [9]. Any big data project will require a sizable volume of data, some of which may contain sensitive or personally identifiable information. The recommended methods for big data security in cloud computing are varied. Every company that keeps data on file needs to be cautious about problems that can affect the following:

1.2. Data Security

Is the data set safe from theft or eavesdropping? What happens to the company if any information is disclosed to unauthorized parties [10]?

1.3. Data Integrity

Is the data set secure against erasure or modification? What happens to the business if any data is altered and a big data analysis yields an unexpected or erroneous result [11]?

1.4. Data Processing

Is the computing infrastructure secure enough to protect the data set? Is it possible to hack into any aspect of the computing infrastructure to potentially reveal data or outcomes [12]?

The organization's security, business governance, and regulatory compliance postures may be impacted by all of these factors. Every cloud user must take steps to ensure that every large data deployment is appropriately secured, configured, and delivered because public cloud security is a shared responsibility.

2. Big Data in the Cloud

Scalability, fault tolerance, and availability are required for the storage and processing of large volumes of data [13].

All of these are provided by cloud computing via hardware virtualization. Big data and cloud computing are thus complementary ideas since the cloud makes big data accessible, scalable, and fault-tolerant. Big data is viewed by businesses as a valuable commercial prospect. As a result, numerous new businesses have begun to concentrate on providing Big Data as a Service (BDaaS) or DataBase as a Service, including Cloudera, Hortonworks, Teradata, and many others (DBaaS) [14]. Customers can access big data on demand through companies like Google, IBM, Amazon, and Microsoft.

For a long time, researchers have researched security and privacy challenges in distributed computing systems. Cloud security, however, differs from typical distributed systems security in several ways. This has to do with the fundamental characteristics of clouds.

Multi-tenancy: The concept of multi-tenancy is the first important problem. By definition, a cloud is a multi-tenant model [15]. This implies that numerous (perhaps unrelated) users will be sharing the same physical hardware and resources in a cloud at any given time. Numerous innovative attacks have been made possible by this resource pooling.

Trust inequity: The uneven trust relationship between the cloud service provider and the customers/users is another reason why cloud security is challenging. The clouds of today behave like large, opaque boxes, preventing consumers from seeing how they are built or function. Therefore, cloud consumers must have complete faith in the cloud provider. Additionally, cloud companies lack any motivation to guarantee security for their customers [16].

Insider risks and a global reach: The primary threat to the majority of distributed systems is the need to protect them from outside intruders [17]. As a result, considerable effort is put into keeping harmful attackers outside the system perimeter. In contrast, the attackers in a cloud can be within the system. They only have to pay to use the cloud resources. The majority of clouds allow anyone with a working payment card access. By using this, attackers can access a cloud without actually breaking any laws or even the usage guidelines set forth by the cloud provider. This system access has expanded the vulnerability of user data and cloud-based apps. Additionally, because clouds are global, any attacker with access to a cloud can target a victim anywhere in the world. Because cloud resources are shared, there is frequently a chance of collateral harm when other users who share the same resources as a victim also experience the effects of an attack.

3. Big Data Security Overview

Three data stages that are not all present in the network require the use of security mechanisms.

The first is data ingress, the second is stored data, and the third is data output, which refers to the information that is sent to apps and reports.

Stage 1: Data Sources.

Many different sources and data types produce big data. CRM or ERM data, transactional and database data, and enormous amounts of unstructured data like email or social media posts are all examples of user-generated data. You also have access to the vast world of machine-generated data, which includes logs and sensors. You must protect this data when it is being transferred from the sources to the platform [18, 19].

Stage 2: Stored Data.

It takes sophisticated security toolkits, such as encryption at rest, strong user authentication, intrusion protection, and planning, to protect stored data. Additionally, you will need to use a distributed cluster infrastructure with plenty of servers and nodes to operate your security toolkits. Additionally, log files and analytics tools must be protected by your security mechanisms when they are being used inside the platform.

Stage 3: Output Data.

The entire reason for the complexity and expense of the big data platform is being able to run meaningful analytics across massive data volumes and different types of data. These analytics output results to applications, reports, and dashboards. This extremely valuable intelligence makes for a rich target for intrusion, and it is critical to encrypt output as well as ingress. Also, secure compliance at this stage: make certain that results going out to end-users do not contain regulated data.

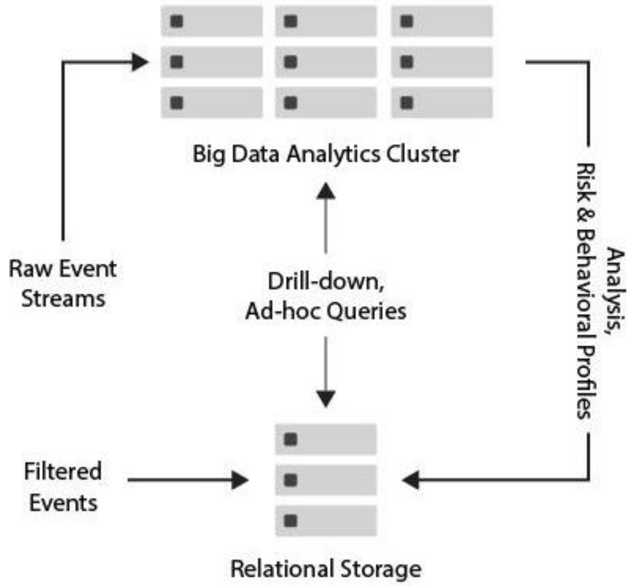
Big Data Analytics Cluster
Raw Event Streams
Drill-down, Ad-hoc Queries
Analysis, Risk & Behavioral Profiles
Filtered Events
Relational Storage

Figure 1. Big Data Analytics Cluster.

4. Security Concerns in the Big Data

Because cloud computing integrates so many different technologies, including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control, and memory management, it has several security issues.

As a result, cloud computing security challenges apply to these systems and technologies. For instance, it is crucial that the network connecting the systems in a cloud be safe. Algorithms for managing memory and allocating resources must also be secure. Particular industries, like telecom, web marketing and advertising, retail and financial services, and some government activities, are most severely impacted by big data challenges. In many industries, the data explosion will make life difficult. Companies that can adapt well and have the ability to evaluate such data explosions will have a significant advantage over competing businesses [20-22].

Finally, malware detection in clouds can be done via data mining approaches. Network-level, user authentication level, data level, and generic issues can all be characterized as security challenges in cloud computing systems.

Network-level: Issues with distributed nodes, distributed data, and inter-node communication are examples of issues that fall within the network level category.

Level of authentication: The difficulties that fall under this category relate to encryption/decryption methods, authentication mechanisms such as administrative privileges for nodes, authentication of apps and nodes, and logging.

Data level: Data integrity and availability-related concerns, including data protection and dispersed data, fall under this category.

Generic types: Traditional security tools and the usage of various technologies are two difficulties that fall under this category.

4.1. In Distributed Programming Frameworks, Secure Computations

Two specific security vulnerabilities are outlined in the first identified risk, which examines the security of computational components in frameworks like MapReduce. First, it is necessary to assess the reliability of the "mappers," or the computer code that separates data into its parts analyzes it, and returns key-value pairs. Secondly, data sanitization and de-identification capabilities must be included to prevent the storage or leaking of sensitive data from the platform. To achieve this, businesses utilizing sophisticated tools like MapReduce will need to apply de-identifier procedures and SELinux's Mandatory Access Controls. On the same point, businesses should find out how cloud providers are managing and resolving this issue in their settings.

4.2. Recommended Practices for Non-Relational Data Stores in Terms of Security

Due to a potential lack of capabilities in numerous crucial areas, including any actual authentication, encryption for data at rest or in transit, logging or data tagging, and classification, the adoption of No SQL and other large-scale, non-relational data storage may result in new security vulnerabilities. To enforce authentication and data integrity, organizations should think about using separate application or middleware layers.

All passwords must be encrypted, and Secure Sockets Layer/Transport Layer Security should preferably be used for all connections to the system. Make that logs are produced for each transaction involving sensitive data.

Transaction logs and safe data storage.

Organizations may store data and transaction logs in multi-tiered storage media, but they must protect against illegal access and guarantee continuity and availability. Only authenticated users and programs can access the platform thanks to policy-based private key encryption.

4.3. Screening and Validating Endpoint Input

Many endpoints may contribute data for processing and storage in a big data implementation.

Organizations need to thoroughly check each endpoint connecting to the corporate network to verify that only trustworthy endpoints are providing data and that fraudulent or malicious data is not transmitted.

Unfortunately, aside from the request to include the Trusted Platform Module chips (present in many newer endpoint devices) in the validation process where possible, the working group does not have a viable set of recommendations for addressing this risk.

Along with sound procedures for system inventory management and maintenance, host-based and mobile device security controls may be able to reduce the risk brought on by untrusted endpoints.

4.4. Monitoring of Security in Real-Time

Security analytics should be run in close to real-time together with monitoring big data platforms.

The sheer volume (and variety of formats) of data used in genuine big data implementations makes it difficult for many conventional security information and event management platforms to keep up.

Unless database and other front-end monitoring technologies are in use, there is currently very little real monitoring of Hadoop and other big data platforms. Privacy-preserving data mining and analytics that are scalable and modular.

Implementing big data might raise privacy issues related to data exposure and leakage. To help businesses deal with this issue, several security controls can be implemented. These include the use of strong encryption for data at rest, access limits to data, and separation of duty procedures and controls to reduce the effectiveness of insider assaults.

4.5. The Security That Is Data-Centric and Enforced by Cryptography

Historically, rather than securing the data itself, the common approach to data control has been to secure the systems that manage the data. But those programs and systems have repeatedly shown themselves to be weak. Strong cryptography is used to encrypt sensitive data in cloud provider environments, along with cutting-edge methods that more effectively enable key management and secure key exchange. This is especially true given that data lives in the cloud independent of any one platform.

4.6. Access Control with Granularity

Mandatory Access Control and strong authentication must be implemented to establish fine-grained access to huge data stores like NoSQL databases and the Hadoop Distributed File System. Cloud service providers should also be able to describe the kinds of access controls that are in place in their settings. New NoSQL implementations, like Apache Accumulo, can enable very fine access control to key-value pairs.

4.7. Granular Audits

Regular audits and analyses of log and event data, along with continuous monitoring, can aid in the detection of intrusions or attempted attacks within the big data environment. Logging at all layers both inside and outside the big data environment should be the main area of control in this case.

4.8. Data Attribution

In this instance, the emphasis of provenance is on the reliability and validity of the data. In big data environments, authentication, end-to-end data protection, and granular access restrictions can help to verify and validate provenance; cloud service providers should already have these policies in place to handle other challenges.

5. Big Data Security and Trends: A Guide

Two of the most significant advancements in the field of big data are the explosion of big data that drives smart technology and the growing desire of people to own and control how their personal data is used, but they are also somewhat at conflict with one another. Terabytes of data including very sensitive personal information are being gathered via IoT, AI, ML, and even databases used for customer relationship management (CRM). Big data in the form of personal information can be advantageous for companies looking to more effectively target their goods and services at their target market, but it also means that all companies and outside vendors must now be responsible for the proper use and management of personal information. As big data and its commercial use cases mature, the majority of companies work to comply with consumer data norms and regulations, yet their security shortcomings leave data open to breaches [23]. Examine some of the most well-known big data trends, significant security holes that many companies have, and suggestions for protecting big data security:

5.1. Updating Your Distributed and Cloud Security Infrastructure

Many businesses are switching to cloud and data fabric infrastructures that enable greater data storage scalability as a result of the proliferation of big data. The issue? Because cloud security features are frequently configured incorrectly and left vulnerable, traditional security principles are frequently used to construct cloud security. Consult your cloud and storage providers to learn more about their offerings, find out if a security solution is built in, and find out if they or a third party partner recommend any additional security tools.

5.2. Set Up Policies and Practices for Mobile Device Management

IoT and other mobile devices rank among the top sources and receivers of big data, but because so many of these gadgets are utilized in daily life, they also present a number of security flaws. Establish stringent guidelines for your employees' use of company data on their personal devices, and be sure to implement extra security layers to control which devices have access to important information.

5.3. Educate Users on Data Security Recommended Practices

Big data is frequently hacked as a result of an effective phishing attack or other targeted attack against an unaware employee. Set up many layers of authentication protection to restrict who may access sensitive data storage, and again, educate your staff on typical socially engineered assaults and what they look like.

Cloud-based big data environments refer to the infrastructure, tools, and services that enable the storage, processing, and analysis of large volumes of data in the cloud. These environments leverage cloud computing technologies and platforms to handle the massive scale, variety, and velocity of data generated by organizations.

The background and context of cloud-based big data environments can be understood by considering the following aspects:

1. Big Data: With the proliferation of digital devices, social media, and Internet of Things (IoT) devices, organizations generate vast amounts of data. Big data refers to datasets that are large, complex, and difficult to manage using traditional data processing and analysis techniques. Big data is characterized by the three Vs: volume, velocity, and variety.
2. Cloud Computing: Cloud computing provides on-demand access to a shared pool of computing resources, such as storage, processing power, and applications, delivered over the internet. Cloud computing offers scalability, flexibility, and cost-efficiency compared to traditional on-premises infrastructure. It allows organizations to store and process large volumes of data without investing in extensive hardware infrastructure.
3. Scalability and Elasticity: Cloud-based big data environments provide the scalability and elasticity required to handle the growing volumes of data. Organizations can easily scale up or down their resources based on demand, ensuring efficient storage and processing of data without incurring unnecessary costs.
4. Data Storage and Processing: In cloud-based big data environments, data is stored in distributed storage systems, such as cloud object storage or distributed file systems. Data processing is performed using distributed computing frameworks like Apache Hadoop or Apache

Spark. These frameworks enable parallel processing across multiple nodes, allowing for efficient processing of large datasets.

5. Data Analytics and Insights: Cloud-based big data environments provide a platform for advanced analytics and deriving insights from the data. Organizations can leverage machine learning, data mining, and predictive analytics techniques to extract valuable information, discover patterns, and make data-driven decisions.
6. Data Security and Privacy: As organizations store and process large volumes of data in the cloud, ensuring data security and privacy becomes crucial. Cloud-based big data environments need robust security measures, including authentication, access controls, encryption, and data governance practices, to protect sensitive information and comply with relevant regulations.
7. Integration with Other Technologies: Cloud-based big data environments often integrate with other technologies, such as Internet of Things (IoT) platforms, real-time data streaming systems, and data visualization tools. These integrations enable real-time data ingestion, processing, and visualization, enhancing the value and usability of big data.

Overall, the context of cloud-based big data environments lies at the intersection of big data challenges, cloud computing capabilities, scalable storage and processing, advanced analytics, data security, and integration with other technologies. This context enables organizations to leverage the power of the cloud to effectively manage and extract insights from their large volumes of data.

Authentication and security play a critical role in cloud-based big data environments due to the following reasons:

1. Data Protection: Cloud-based big data environments often store vast amounts of sensitive and valuable data. This data may include proprietary business information, customer data, financial records, and intellectual property. Authentication and security measures are essential to prevent unauthorized access, data breaches, and potential misuse of this valuable data.
2. Compliance Requirements: Many industries have specific compliance and regulatory requirements regarding data security and privacy. Organizations operating in cloud-based big data environments must adhere to these regulations, which often include stringent authentication and security measures. Compliance with these requirements helps avoid legal and financial consequences associated with non-compliance.
3. Insider Threats: Cloud-based big data environments may involve multiple users, including employees, contractors, and partners. While these users have authorized access to the system, there is a risk of insider threats, such as unauthorized data access, data leaks, or intentional data manipulation. Authentication ensures that only authorized individuals can access the data, reducing the risk of insider threats.
4. Mitigating External Threats: Cloud-based big data environments are potential targets for external threats, including hackers, cybercriminals, and malicious actors. Authentication measures, such as multi-factor authentication (MFA), help protect against unauthorized access attempts. By implementing robust authentication mechanisms, organizations can significantly reduce the risk of external threats compromising the security and integrity of their data.
5. User Accountability and Access Control: Authentication provides a means to establish user accountability and enforce access control policies. By identifying and authenticating individual users, organizations can track and audit user activities, monitor data access, and enforce role-based access controls. This ensures that data is accessed and utilized only by authorized individuals and reduces the risk of data breaches or unauthorized modifications.
6. Trust and Confidence: Authentication and security measures in cloud-based big data environments contribute to building trust and confidence among users, customers, and stakeholders. When organizations demonstrate a commitment to data security through strong authentication practices, it enhances their reputation, fosters customer trust, and encourages data sharing and collaboration.
7. Continuity and Resilience: Robust authentication and security measures are essential for maintaining the continuity and resilience of cloud-based big data environments. By implementing strong authentication mechanisms, organizations can prevent unauthorized disruptions, data loss, or unauthorized modifications. This ensures the availability, integrity, and reliability of data and services even during security incidents or attacks.

In summary, authentication and security measures are vital in cloud-based big data environments to protect sensitive data, comply with regulations, mitigate insider and external threats, enforce access controls, establish user accountability, build trust, and maintain the continuity of operations. By prioritizing authentication and security, organizations can safeguard their data and infrastructure, enabling them to leverage the full potential of cloud-based big data environments.

The purpose of conducting research on assessing the effectiveness of Multi-Factor Authentication (MFA) in cloud-based big data environments is to evaluate the security measures and controls implemented to protect sensitive data. This research aims to provide insights into the strengths and weaknesses of MFA in such environments and its ability to mitigate security risks.

The significance of assessing MFA effectiveness lies in several key aspects:

1. Security Enhancement: MFA is widely regarded as a strong security measure, but its effectiveness can vary depending on implementation factors and contextual

considerations. Assessing MFA effectiveness helps identify areas of improvement, potential vulnerabilities, and emerging threats. This research contributes to enhancing the overall security posture of cloud-based big data environments.

2. Protection of Sensitive Data: In cloud-based big data environments, organizations store vast amounts of sensitive information. Assessing the effectiveness of MFA ensures that appropriate security measures are in place to protect this data. By evaluating MFA implementation and identifying potential weaknesses, organizations can prevent unauthorized access, data breaches, and insider threats.
3. Compliance and Regulatory Requirements: Many industries have specific compliance and regulatory requirements regarding data security, privacy, and user authentication. Assessing MFA effectiveness helps organizations ensure they meet these requirements and avoid legal and financial consequences associated with non-compliance.
4. User Experience and Adoption: MFA implementation should strike a balance between security and user experience. Assessing MFA effectiveness allows for an evaluation of the user experience with different authentication factors, usability challenges, and potential barriers to adoption. Understanding these aspects can lead to the development of more user-friendly and efficient MFA solutions.
5. Continuous Improvement and Innovation: By assessing MFA effectiveness, organizations gain insights into the evolving landscape of authentication mechanisms, emerging technologies, and industry best practices. This research contributes to the ongoing improvement and innovation in MFA solutions, paving the way for stronger security measures and advancements in cloud-based big data environments.

Overall, assessing MFA effectiveness in cloud-based big data environments is of significant importance as it helps organizations enhance security, protect sensitive data, meet compliance requirements, improve user experience, and contribute to the continuous advancement of authentication practices.

6. Existing Research Approach

R. K. Banyal et al. [24] proposed a framework that incorporates multiple authentication factors, including password that the user knows, smart card that the user has and biometric data that the user has. The framework also incorporates a risk based authentication mechanism that adjusts the authentication requirements dynamically based on the perceived level of risk. The paper also implements the proposed framework in the cloud environment and evaluates the effectiveness of the framework. The evaluated result shows that the framework managed to outperform the other existing frameworks and protect against various types of attacks such as brute force attacks, phishing attacks, and man in the middle attacks.

Kurnia, S. et al. [25] conducted a systematic literature review of previous studies on authentication safety practises and identified several types of authentication practises such as password, biometrics, smart cards, one time passwords, and multi factor authentication. The authors of the paper also performed an in depth analysis on the effectiveness of each authentication practice in different scenarios and identified the strengths and weaknesses of each method. The study finds out that biometric and smart cards are more secured but not suitable for all scenarios because of their cost and complexity. One time passwords and multi-factor Authentication are also an effective authentication method that provides additional layers of security.

Y. Wang et al. [26] make in depth review on various authentication mechanisms used in cloud environments such as password-based authentication, biometric authentication, and multi-factor authentication. The paper also discusses the limitations and vulnerabilities of these mechanisms and highlights the need for more secure and efficient authentication methods in cloud based big data environments. The paper also identifies the need for secure and privacy-preserving authentication mechanisms, efficient and scalable authentication solutions for large- scale data processing, and robust authentication schemes to protect against various cyberattacks.

Existing literature on Multi-Factor Authentication (MFA) and its effectiveness in cloud-based big data environments highlights the importance of strong authentication mechanisms and the potential challenges in implementing MFA in these complex environments. Here is an overview of key findings from the literature:

1. Authentication Factors and Strength: Research emphasizes the importance of using multiple authentication factors in MFA to enhance security. Studies explore various factors such as passwords, biometrics (fingerprint, face recognition), tokens (smart cards, USB tokens), and behavioral factors (keystroke dynamics). The effectiveness of these factors depends on factors such as usability, reliability, and resistance to attacks.
2. Usability and User Experience: Usability and user experience play a crucial role in the successful adoption of MFA. Literature highlights the need for MFA solutions that are intuitive, easy to use, and do not hinder user productivity. User acceptance and satisfaction are critical factors in the effectiveness of MFA implementation.
3. Security and Vulnerabilities: Studies examine the security aspects of MFA and potential vulnerabilities. They highlight the importance of secure key management, encryption, and secure communication protocols. Additionally, research explores vulnerabilities such as credential theft, phishing attacks, and social engineering attempts targeting MFA.
4. Integration with Cloud and Big Data Technologies: Literature explores the integration of MFA with cloud

platforms and big data technologies. It emphasizes the need for seamless integration and compatibility with existing cloud infrastructure. Research also delves into the impact of MFA on performance, scalability, and usability of big data applications and services.

5. Risk Assessment and Threat Modeling: Several studies propose risk assessment frameworks and threat modeling techniques specific to MFA in cloud-based big data environments. They aim to identify potential threats, assess risks, and prioritize mitigation efforts. These frameworks help organizations understand the effectiveness of MFA in mitigating specific risks associated with their cloud-based big data deployments.
6. Compliance and Regulations: Research examines the alignment of MFA implementations with industry-specific compliance requirements and regulations. This includes standards such as the Payment Card Industry Data Security Standard (PCI DSS) and General Data Protection Regulation (GDPR). Compliance with these standards ensures the effectiveness and legal compliance of MFA in cloud-based big data environments.
7. Performance and Scalability: The literature investigates the performance and scalability implications of MFA in cloud-based big data environments. Studies analyze the impact of MFA on response times, throughput, and resource utilization. They explore techniques such as parallelization and optimization to mitigate any potential performance bottlenecks.
8. Emerging Technologies and Innovations: Some literature explores emerging technologies and innovations in MFA for cloud-based big data environments. This includes research on biometric advancements, adaptive authentication, context-aware authentication, and machine learning-based approaches. These technologies aim to enhance the effectiveness and user experience of MFA in these environments.

Overall, the existing literature provides insights into the effectiveness of MFA in cloud-based big data environments, addressing factors such as authentication strength, usability, security vulnerabilities, compliance, performance, and emerging technologies. These findings contribute to a better understanding of MFA implementation and its impact on securing cloud-based big data environments.

Industry Best Practice: Cloud Security Alliance (CSA) Security Guidance for Critical Areas of Focus in Cloud Computing.

CSA provides industry best practices for securing cloud environments, including recommendations for implementing MFA. The guidance emphasizes the need for MFA as a crucial security measure and provides insights into factors such as identity federation, strong authentication factors, and centralized authentication services.

These studies, research papers, and industry best practices contribute to the understanding of MFA effectiveness in cloud-based big data environments. They offer insights into the strengths and weaknesses of different authentication factors, propose frameworks for implementation, and highlight key considerations for security, usability, compliance, and performance. By leveraging these resources, organizations can make informed decisions and adopt best practices when implementing MFA in their cloud-based big data environments.

While the existing research on the effectiveness of Multi-Factor Authentication (MFA) in cloud-based big data environments provides valuable insights, there are still some gaps and limitations that need to be addressed. These include:

1. Limited Focus on Big Data-Specific Considerations: Many studies primarily focus on MFA in general cloud environments without considering the unique challenges and requirements of big data environments. There is a need for research that specifically addresses the effectiveness of MFA in the context of big data, considering factors such as data volume, velocity, variety, and the distributed nature of big data systems.
2. Lack of Real-World Implementation Studies: Most research papers and studies on MFA effectiveness in cloud-based big data environments are theoretical or based on simulations. There is a need for more real-world implementation studies that evaluate MFA in actual production environments. These studies can provide practical insights into the challenges, successes, and lessons learned from implementing MFA in complex big data architectures.
3. Limited User-Centric Evaluation: While some research touches on the usability and user experience of MFA, there is a need for more user-centric evaluations. Understanding the user perception, acceptance, and satisfaction with MFA in cloud-based big data environments can help identify usability issues and inform the design of more user-friendly and efficient MFA solutions.
4. Inadequate Assessment of Emerging Technologies: The current research landscape lacks comprehensive assessments of emerging technologies and innovations in MFA for cloud-based big data environments. As new authentication methods and technologies continue to evolve, there is a need to evaluate their effectiveness, security, and scalability specifically in the context of big data environments.
5. Limited Evaluation of Risk Assessment and Threat Modeling: While some studies propose risk assessment frameworks and threat modeling techniques for MFA, there is a need for further evaluation and validation of these approaches. Research should focus on practical implementation and evaluate their effectiveness in identifying and mitigating specific risks and threats in cloud-based big data environments.
6. Insufficient Long-Term Effectiveness Analysis: Many research papers provide insights into the initial effectiveness of MFA implementations, but there is a lack of long-term analysis of their continued effectiveness. It is

essential to assess the long-term sustainability, adaptability, and resistance to evolving threats of MFA solutions in cloud-based big data environments.

Addressing these gaps and limitations in future research will contribute to a more comprehensive understanding of MFA effectiveness in cloud-based big data environments and facilitate the development of more robust and secure authentication solutions.

7. Implementation of Multi-Factor Authentication

The actual implementation of multi-factor authentication (MFA) in cloud-based large data systems falls short of the existing frameworks. Although MFA is acknowledged as an efficient security mechanism, in-depth research on its efficiency in cloud-based big data contexts is lacking. Numerous recent research on MFA in cloud computing are restricted to conventional computing environments and fail to take into consideration the special security concerns given by cloud-based big data environments, such as the requirement to defend against a greater variety of cyberattacks.

A paucity of explicit guidance exists for establishing MFA in cloud-based big data environments, despite the fact that there are numerous frameworks and recommendations available for doing so in traditional computing systems. This discrepancy emphasizes the need for additional study to create particular frameworks and best practices for implementing MFA in cloud-based big data systems and to assess how well they safeguard sensitive data.

The implementation of multifactor authentication(MFA) in cloud based big data environments involves several steps., including:

7.1. Identify CSP

Decide which CSP will be utilized to store and process massive data in the cloud first. The levels of security and MFA options offered by different CSPs vary, therefore it's crucial to pick a CSP with strong MFA capabilities.

7.2. Choose MFA Methods

Next, decide which MFA techniques will be employed. Passwords, biometric authentication, smart cards, one-time passwords, and other authentication methods may be included in this.

7.3. MFA Implementation

After these MFA techniques have been identified, their implementation is the following stage. This could entail setting up the necessary authentication process, configuring the CSP to demand MFA for user access, and testing the MFA system to make sure it is operating as intended.

7.4. Provide Training to Users

Users should receive instruction on how to utilize the MFA system. They may need to be taught about the various authentication options, how to use them, and best practices for safeguarding their authentication credentials.

7.5. Monitor the MFA System

After it has been put into place, it is crucial to routinely check on the MFA system to make sure it is operating properly and to spot any potential security risks. This could entail putting up notifications for unsuccessful authentication attempts, keeping an eye on user access logs, and routinely evaluating security rules and procedures.

7.6. Regularly Update MFA System

Finally, it is important to regularly update the MFA system to ensure that it remains effective against emerging security threats. This may involve updating authentication methods, adding new MFA options, or implementing other security measures to strengthen security of the cloud- based big data environment.

8. Usefulness of Multi-Factor Authentication

The usefulness of multi-factor authentication in cloud-based big data environments is generally recognised, according to the many research publications, as a security strategy that lowers the danger of data breaches brought on by weak passwords. However, the degree of effectiveness may differ based on a number of variables, including how the MFA system is specifically implemented, the kinds of assaults being targeted, and how users behave and view the system.

For instance, [27] comparison of the efficiency of various MFA systems revealed that biometric-based authentication systems offered the highest level of security when compared to other MFA systems like SMS-based authentication. In contrast, a study by [28] discovered that in some circumstances likewise in fintech, SMS-based authentication was more successful than biometric-based authentication at preventing unwanted access.

User behavior and perception can also have an impact on how effective MFA systems are. Users frequently use less secure passwords when utilizing MFA systems, possibly compromising the security advantages of the system, according to a study by [29]. As a result, improving the effectiveness of MFA systems can also benefit greatly from user education and training on proper usage.

There is still a need for additional study to fill in the gaps and overcome the constraints found in the existing studies, even though the existing research offers some insights into the effectiveness of MFA in cloud-based large data systems.

Research is ongoing in the area of evaluating the efficacy of multi-factor authentication (MFA) in cloud-based big data systems. Several studies have looked at the application of MFA in various circumstances, including cloud-based settings. While some studies have examined the trade-offs between security and usability when adopting MFA in cloud-based systems, others have examined the effectiveness of MFA in avoiding data breaches caused by compromised passwords.

Balancing security and usability is one of the biggest issues when deploying MFA in cloud-based large data systems. MFA is a useful security feature, but implementing it can be difficult and annoying for users, which can lower adoption rates. As a result, research in this area tries to pinpoint the MFA tactics that work best and balance security and usability.

Research has concentrated on creating new ways and tactics to enhance MFA efficacy in order to overcome the gap between the current and used paradigm in analyzing the effectiveness of MFA in cloud-based big data environments. Using mouse movements and keyboard patterns as behavioral biometrics for supplementary authentication is one method. Others have investigated using machine learning algorithms to find unusual user behavior and possible security risks.

The importance and research value of the topic "Assessing the Effectiveness of Multi-Factor Authentication in Cloud-Based Big Data Environments" lies in the requirement to create effective security measures that strike a balance between security and usability in cloud-based environments, where data breaches can have serious repercussions. The goal of ongoing research in this field is to determine the MFA tactics and techniques that are most successful for overcoming the difficulties associated with protecting sensitive data in the cloud.

9. Conclusions

Many cloud service provider environments do big data collecting and processing in some way. Although big data platforms and controls may not be in place internally in the majority of consumer firms, it is crucial to comprehend the main dangers and threats posed to enterprise data in the cloud environment. Cloud infrastructures can be safeguarded for sophisticated commercial processes despite security concerns and difficulties. A security company can continuously update its global threat intelligence and gain better insight into threats by using big data tools to examine the enormous quantity of threat data they receive each day and correlate the many elements of an assault. Customers gain from enhanced, quicker, and more comprehensive threat protection. They prevent potential recovery expenses, damaging brand effects, and legal repercussions by lowering risk.

Finally, with the continued growth in the use of big data and cloud computing, the security and privacy of sensitive data in these settings has become a major worry. MFA, a dependable security measure, reduces the likelihood of data breaches caused by weak passwords. The value of MFA in cloud-based massive data systems is not widely acknowledged because of specific security challenges and trade-offs between security and usability.

The topic of evaluating MFA's efficacy in cloud-based big data contexts has been covered in a number of research articles, covering implementation, framework, and comparative studies. Despite the fact that these studies' findings point to the potential success of MFA in improving the security of cloud-based big data settings, additional study is still required to close the gap between the frameworks currently in use and the effectiveness of MFA.

Overall, the subject of evaluating MFA's efficacy in cloud-based big data contexts is still important and beneficial for offering perspectives and suggestions for safeguarding sensitive data in the cloud.

Author Contributions

Saroj Mali is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] J. Ruohonen, K. Hjerppe, and K. Kortesuo, "Crisis Communication in the Face of Data Breaches," *arXiv.org*, Jun. 03, 2024. https://arxiv.org/abs/2406.01744

[2] R. M and S. R. P, "Determining Intrusion Attacks Against Online Applications Using Cloud-Based Data Security," *ICST Transactions on Scalable Information Systems*, Feb. 2024, https://doi.org/10.4108/eetsis.5028

[3] Mylly, UM. Trade Secrets and the Data Act. *IIC* **55**, 368–393 (2024). https://doi.org/10.1007/s40319-024-01432-0

[4] M. Dawood, S. Tu, C. Xiao, H. Alasmary, M. Waqas, and S. U. Rehman, "Cyberattacks and Security of Cloud Computing: A Complete Guideline," *Symmetry*, vol. 15, no. 11, p. 1981, Oct. 2023, https://doi.org/10.3390/sym15111981

[5] "Data Security and Privacy Protection for Cloud Storage: A Survey," *IEEE Journals & Magazine | IEEE Xplore*, 2020. https://ieeexplore.ieee.org/abstract/document/9142202

[6] N. Allahrakha, "Balancing Cyber-security and Privacy: Legal and Ethical Considerations in the Digital Age," *Legal Issues in the Digital Age*, vol. 4, no. 2, pp. 78–121, Jul. 2023, https://doi.org/10.17323/10.17323/2713-2749.2023.2.78.121

[7] J. Abrera, "Data Privacy and Security in Cloud Computing: A Comprehensive Review," *Journal of Computer Science and Information Technology* -, Jul. 2024, https://doi.org/10.61424/jcsit.v1i1.58

[8] F. Aslam, "Role of Cloud Computing for Big Data," *Zenodo (CERN European Organization for Nuclear Research)*, Sep. 2023, https://doi.org/10.5281/zenodo.8311108

[9] S. Rani, P. Bhambri, and A. Kataria, “Integration of IoT, Big Data, and Cloud Computing Technologies,” in *Chapman and Hall/CRC eBooks*, 2023, pp. 1–21. https://doi.org/10.1201/9781003298335-1

[10] A. S. G. P. Cpg Jd, Dba, Mba, “Privacy and Data Security,” *Social Science Research Network*, Jan. 2023, https://doi.org/10.2139/ssrn.4566936

[11] Privacy Preservation and Secured Data Storage in Cloud Computing. 2023. https://doi.org/10.4018/979-8-3693-0593-5

[12] E. D. Knapp and J. T. Langill, Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems. 2011. [Online]. Available: http://cds.cern.ch/record/1988553

[13] Sundarakumar *et al.*, “A comprehensive study and review of tuning the performance on database scalability in big data analytics,” *Journal of Intelligent & Fuzzy Systems*, vol. 44, no. 3, pp. 5231–5255, Mar. 2023, https://doi.org/10.3233/jifs-223295

[14] V. Sakthivel, S. Nanduri, and P. Prakash, “Role of Big Data Analytics in the Cloud Applications,” in *River Publishers eBooks*, 2024, pp. 69–94. https://doi.org/10.1201/9781032630212-5

[15] “A Hierarchical Namespace Approach for Multi-Tenancy in Distributed Clouds,” *IEEE Journals & Magazine | IEEE Xplore*, 2024. https://ieeexplore.ieee.org/abstract/document/10443611

[16] S. Rizvi and I. Williams, “Analyzing transparency and malicious insiders prevention for cloud computing environment,” *Computers & Security*, vol. 137, p. 103622, Feb. 2024, https://doi.org/10.1016/j.cose.2023.103622

[17] F. Basholli, A. Daberdinİ, and A. Basholli, “Detection and prevention of intrusions into computer systems,” Mar. 22, 2023. https://publish.mersin.edu.tr/index.php/aed/article/view/941

[18] “Big Data & Cloud Computing IEEE Computer Society.” *IEEE Computer Society*, https://www.computer.org/publications/tech-news/trends/big-data-and-cloud-computing/

[19] Foote, Keith D. “Big Data and Cloud Security - DATAVERSITY.” *DATAVERSITY*, 21 July 2016, https://www.dataversity.net/big-data-cloud-security/

[20] Neves, Pedro & Schmerl, Bradley & C ámara, Javier & Bernardino, Jorge. (2016). Big Data in Cloud Computing: Features and Issues. 307-314. https://doi.org/10.5220/0005846303070314

[21] Chithik Raja, Mohamed Sinnaiya. (2015). Big Data Security Issues and Management in Cloud Computing Environment.

[22] Hasan, Ragib. (2014). Security in Big Data and Cloud Computing: Challenges, Solutions, and Open Problems. https://doi.org/10.1201/b17112-20

[23] Taylor, Christine. “What Is Big Data Security? Challenges & Solutions.” *Datamation*, Datamation, 26 Sept. 2021, https://www.datamation.com/big-data/big-data-security/

[24] R. K. Banyal, P. Jain and V. K. Jain, "Multi-factor Authentication Framework for Cloud Computing," *2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation*, Seoul, Korea (South), 2013, pp. 105-110, https://doi.org/10.1109/CIMSim.2013.25

[25] Kurnia, S., Sari, S. A., & Shalahuddin, M. (2019). A systematic literature review of the types of authentication safety practices among internet users. International Journal of Advanced Computer Science and Applications, 10 (11), 414-421.

[26] Y. Wang, R. Chen, and Z. Yan, "Authentication for Big Data Processing in Cloud Environments: State-of-the-Art and Research Challenges," IEEE Transactions on Services Computing, vol. 10, no. 2, pp. 274-284, 2017.

[27] A. H. Y. Mohammed, R. A. Dziyauddin, and L. A. Latiff, “Current Multi-factor of Authentication: Approaches, Requirements, Attacks and Challenges,” *International Journal of Advanced Computer Science and Applications/International Journal of Advanced Computer Science & Applications*, vol. 14, no. 1, Jan. 2023, https://doi.org/10.14569/ijacsa.2023.0140119

[28] H. U. Khan, M. Sohail, S. Nazir, T. Hussain, B. Shah, and F. Ali, “Role of authentication factors in Fin-tech mobile transaction security,” *Journal of Big Data*, vol. 10, no. 1, Sep. 2023, https://doi.org/10.1186/s40537-023-00807-3

[29] A. Nanda, J. J. Jeong, S. W. A. Shah, M. Nosouhi, and R. Doss, “Examining Usable Security Features and User Perceptions of Physical Authentication Devices,” *Computers & Security*, vol. 139, p. 103664, Apr. 2024, https://doi.org/10.1016/j.cose.2023.103664