**A logo for a university

Description automatically generated**

**MSc Networks and Systems Security**

**Assignment**

**Access Control for Securing Big Data**

**NAME: Vahin Valliyur Sankar**

**STUDENT ID: 32121535**

**LECTURER: Abel Yeboah-Ofori**

**DATE: 12.12.2023**

Contents

[Introduction: 3](#_Toc153737723)

[Big Data Attacks and Threat Analysis 3](#_Toc153737724)

[1. SQL injection: 4](#_Toc153737725)

[2. Distributed Denial-of-Service (DDoS) Attacks: 4](#_Toc153737726)

[3. Malware: 4](#_Toc153737727)

[4. The financial and reputational risks: 5](#_Toc153737728)

[Access Control Challenges for Securing Big Data: 5](#_Toc153737729)

[1. Scalability: 5](#_Toc153737730)

[2. Diversity of Data Sources: 6](#_Toc153737731)

[3. Dynamic Data: 6](#_Toc153737732)

[4. Drawbacks of Traditional Access Control Methods: 6](#_Toc153737733)

[5. Illustrative Examples of Access Control Challenges: 7](#_Toc153737734)

[Securing Big Data: A Critical Analysis of Existing Approaches: 7](#_Toc153737735)

[1. Encryption: 7](#_Toc153737736)

[2. Authentication and Authorization: 8](#_Toc153737737)

[3. Machine Learning and AI in Big Data Security: 8](#_Toc153737738)

[4. Strengths and Weaknesses of Existing Approaches: 9](#_Toc153737739)

[Conclusion: 9](#_Toc153737740)

[Reference list 10](#_Toc153737741)

# Introduction:

Big data is the combination of structured, semi-structured, and unstructured data that businesses collect. It is a vast resource that powers machine learning, advanced analytics, and predictive modelling (Botelho & Bigelow, 2022). The growing popularity of big data technologies has led to an increase in threats and attacks targeting these large datasets. Cybercriminals utilise complex techniques such as advanced persistent threats (APTs), SQL injection, and distributed denial-of-service (DDoS) to take advantage of weaknesses in big data ecosystems.

Successful big data attacks can have far-reaching consequences, including significant financial losses and permanent damage to one's reputation. A major instance happened in India in March 2018, when the world's largest biometric database leaked over a billion people's personal data due to a massive data breach. An Aadhaar holder's name, unique 12-digit identity number, and even private information like bank account details, images, thumbprints, and retinal scans were made public by the breach, which originated from a state-owned utility company's system (Tunggal, 2023).

Considering the possible consequences of malicious activity and unauthorised access, protecting large amounts of data is not just wise but necessary. Apart from the obvious financial consequences, big data breaches expose personal information, increasing the likelihood of identity theft for individuals and seriously damaging an organization's brand. Protecting the privacy and accuracy of big data is essential for maintaining legal compliance, maintaining trust, and advancing data-driven technologies.

This essay will examine the many aspects of big data attacks, examine the special difficulties associated with access control, and evaluate current methods for protecting this extremely valuable asset in the sections that follow.

# Big Data Attacks and Threat Analysis

Big data is a prime target for a variety of cyberattacks because it is an abundant supply of information. When creating effective security measures, it is essential to understand the common types of attacks. Figure 1 shows a structured categorization of big data environment attacks.

A diagram of a computer network

Description automatically generated

*Hierarchy diagram of cyber-attacks on Big data Environment (M V, et al., 2018)*

## 1. SQL injection:

SQL injection, a prevalent attack vector, allows malicious actors to manipulate databases by injecting malicious SQL queries into input fields. A perfect illustration of the catastrophic consequences of SQL injection can be found in the 2017 Equifax data breach. Equifax showed in a breach disclosure notice that over 145 million US customers' personal information was compromised by criminal hackers using the Apache Struts web application vulnerability CVE-2017-5638. In addition to the breach, and this was made possible by a failure to take precautions against a known vulnerability, Equifax suffered significant financial losses as well as irreversible harm to its reputation. The impact of these attacks is highlighted by Equifax's admission that "limited personal information" belonging to residents of Canada and the United Kingdom had also been compromised (Hedley and Jacobs, 2017).

## 2. Distributed Denial-of-Service (DDoS) Attacks:

DDoS attacks are a serious risk to big data systems because they overload targeted servers with an excessive amount of traffic, disrupting the system. The current state of DDoS attack trends and insights illuminates how this dangerous strategy is changing.

A DDoS attack on the recognised Among Us game servers occurred in March 2022, making the well-known multiplayer game unavailable for several days. The attack demonstrated the effects of DDoS events in the real world, upsetting gamers and drawing attention to the weakness of online gaming platforms (Subring and Naykodi, 2023).

## 3. Malware:

A malicious software application, additionally recognized as malware, creates a hazard to large data systems by endangering their safety and security, taking sensitive info, and penetrating networks. The cyber-attack targeting the uranium enrichment facility in Natanz used two distinctive versions of the Stuxnet virus, showcasing the tremendous capacities of malware.

The very first iteration of the Stuxnet strike was extra elaborate and refined than its subsequent variation, suggesting a greater degree of knowledge in its development. The key target was the security systems of the IR 1 centrifuges situated in Natanz, which were susceptible to failures. Despite their functional concerns, these centrifuges were mass-produced as a result of their adaptability. The malware showed a high degree of elegance in its attempt to interrupt the connected system, leading to the over-pressurization of the centrifuges.

The Natanz system was noteworthy for its in-depth design, as it was developed to resolve the restrictions of the ageing IR-1 centrifuge modern technology. Without this sophisticated system, the centrifuges would be undependable and tough to regulate, making them not sensible for real-world applications. This instance highlights the shrewd methods used by malware to make usage of susceptibilities in obsolete innovation and advanced systems created to decrease their imperfections.

The array of tools in this situation underscores the vital role of accuracy and environmentally friendly management in guaranteeing the correct functioning of IR 1 centrifuges. This incident functions as a precursor, highlighting the devastating repercussions that malware can carry systems. It emphasizes the significance for services to implement rigorous safety and security and security procedures to protect against innovative cyber threats (Van Dine, 2017).

## 4. The financial and reputational risks:

There are risks connected with cyberattacks on big data systems both in terms of finance and online credibility. As an example, the Equifax information violation acts as an instance. In this event, the personal information of over 145 million clients was endangered due to a failure to apply measures versus popular vulnerabilities like SQL shot. This not resulted in immediate economic effects yet likewise created lasting damages, to the organization's brand name. Targeted DDoS attacks versus prominent websites such as Amongst United States and Grand Theft Auto: San Andreas, together with the highly knowledgeable Stuxnet malware procedure against the uranium enrichment plant in Natanz, additionally show the range of cyber dangers.

The recent developments highlight the significance of implementing strenuous procedures to lessen cyber dangers and ensure the safety and security and protection of electronic possessions. With the enhancing dependancy on modern technology, companies should prioritize cybersecurity to protect delicate details, avoid monetary losses, and maintain their reputation.

In the following area, this essay will certainly look into the phenomenal difficulties linked with gain access to control in protecting extensive information, supplying an extensive exam of the obstacles entailed.

# Access Control Challenges for Securing Big Data:

Gain access to control is one area where big data safety and security provides unique obstacles due to the nature of large and heterogeneous datasets. Protecting the discernment and security of vital details requires an understanding of these issues and efficient alternatives.

## 1. Scalability:

Scalability is a considerable trouble in big data accessibility control, which is further complicated by the modifying eHealth landscape. Taking care of big quantities of details in big-data settings requires limited mix of access control requirements. A prime circumstances is the location of digital health documents (EHRs) in eHealth, where patient details originates from various resources such as wearables, smart sensors, and clinical imaging tools. Worrying data released by the united state Division of Health and Wellness and Individual Providers underscore the immediate requirement to secure EHR information. Especially, in between 2009 and 2017, there were over 2,181 instances of treatment information infractions, disclosing an overall of 176,709,305 medical records.

The distributed and fragmented nature of EHR data, in addition to the facility interconnections in between information proprietors and users, make it tough to enforce effective gain access to control, although file encryption offers a standard layer of safety. Because of this intricacy, a versatile and fine-grained accessibility control solution for EHR information is required to ensure that the privacy and honesty of client information are not risked by the transforming demands of scalability (Guo et al., 2019).

## 2. Diversity of Data Sources:

A large range of data sources, consisting of structured, semi-structured, and unstructured data, are included in big information (Google Cloud, 2023). Using standard accessibility control designs constantly across heterogeneous information kinds is challenging since they regularly assume standardised data styles and frameworks. As a result of this heterogeneity, an adaptable and versatile accessibility control system that can manage a variety of information frameworks and formats without compromising security is needed.

## 3. Dynamic Data:

High information rate-- the continuous ingestion, processing, and evaluation of information streams--.defines big information environments (Chen and Ning, 2016).The vibrant nature of big data. often causes standard gain access to control designs to fall back, resulting in old-fashioned policies. that may inadvertently allow unauthorised access to secret information. policies existing and practical while fitting changing information gain access to patterns, a dynamic. To maintain. access control system is needed.

## 4. Drawbacks of Traditional Access Control Methods:

Static customer roles and access consents are the structure blocks of conventional access control. models like Discretionary Access Control (DAC) and Role-Based Access Control (RBAC). These designs work well in standard database settings, however they are not appropriate. For big data due to their distributed and dynamic nature.

In big information, users usually require to access several services and datasets from different groups and divisions. Taking care of and transforming permissions for a great deal of customers and duties can be testing using typical RBAC designs.

Usually, traditional gain access to control designs depend on a centralised policy administration server. Nevertheless, massive implementations might encounter challenges with this web server. As data and individual numbers expand, delays in plan updates can prevent data gain access to and analytics.

Limited Granularity: Customers are frequently given wide approvals to access entire information sets. Under conventional accessibility control designs, which provide grainy accessibility control. This may lead to over-permissions and a greater chance of delicate information being accessed by unsanctioned parties.

## 5. Illustrative Examples of Access Control Challenges:

Access to a company's customer database is granted to a customer service representative. The representative's access is not cancelled even after they are fired for acting inappropriately. The agent posts private client information on social media by abusing their access (Bauer et al., 2009).

Access to a company's network is granted to a network administrator. The administrator's access is not taken away when they are fired for carelessness. The administrator attacks the company's website with a denial-of-service attack using their access.

The significance of efficient access control management is demonstrated by these instances. To ensure that only individuals with permission can access sensitive information and systems, organisations must put policies and procedures in place (adamokeeffe, 2023). When an employee leaves the company, access should be promptly revoked, and users should be informed when their privileges are altered. Organisations can aid in preventing unwanted access to their data and systems by implementing these precautions.

# Securing Big Data: A Critical Analysis of Existing Approaches:

A new era of security challenges for organisations has been brought about by the enormous growth of big data. Big data assets are frequently not sufficiently protected by traditional security measures because they were created for smaller, less complex data environments. A variety of approaches—each with pros and cons of their own—have been developed in response to deal with these critical issues.

## 1. Encryption:

The use of encryption is a basic security measure for big data protection. It ensures that sensitive information cannot be accessed or understood by unauthorised users by converting data into an unreadable format. At different stages of the data lifecycle, such as when the data is in use, in transit, or at rest, encryption can be applied (CloudFlare, 2023).

**Data at Rest Encryption**: If physical devices are misplaced or stolen, encryption makes sure that private information kept on disc or in databases is shielded from unwanted access. The encryption algorithms AES, RSA, and ECC are frequently used.

**Data in Transit Encryption:** Encryption protects data from theft and eavesdropping while it's being transferred across networks. Data encryption in transit is frequently achieved through the use of secure protocols like TLS/SSL and SSH (Information Commissioner's Office, 2023).

**Data in use Encryption**: Data can also be encrypted while it's being examined or processed on CPUs or in memory. In the process of analysis, this helps shield data from memory leaks and data breaches (Google Cloud, 2021).

## 2. Authentication and Authorization:

In big data atmospheres, verification and authorization are crucial controls for preservation. access control. While authorization establishes which resources and activities a customer is enabled.to gain access, verification validates the identity of a customer or device (OneLogin, 2023).

**Authentication:** To ensure the security of large-scale data systems and avoid unapproved gain access to, implementing robust verification approaches is critical. One such approach is multi-factor authentication (MFA), which combines numerous confirmation elements, such as passwords, security tokens, or biometric information (UTrust for Software Testing & Quality Service, 2023), to give an included layer of security.

**Authorization:** To limit accessibility to delicate info and sources, granular permission. plans should be developed. One popular method is role-based accessibility control (RBAC),.which disperses authorizations according to user duties.A lot more specific control is enabled.by fine-grained accessibility control (FGAC), which permits specific access to particular data things.or operations (Auth0, 2023).

## 3. Machine Learning and AI in Big Data Security:

The landscape of big data security is undergoing an extreme change, many thanks to the advent of device discovery and fabricated intelligence (AI). These cutting-edge modern technologies are giving novel methods to spot, evaluate, and reply to hazards, thus improving the total safety pose of organizations.

### Anomaly Detection:

Artificial intelligence algorithms can determine unusual patterns in historical data, indicating potential security threats, such as intrusions or breaches, by evaluating past information patterns and recognizing deviations from the standard.

### User Behaviour Analysis:

Our cutting-edge innovation may look at customer behavior to discover unique crazes that might show dangerous or deceitful activities. By taking a look at consumer styles, our experts can easily reveal irregularities such as significant adjustments in login routines or excessive attempts to get access to info. These warnings may function as early cautions of possible protection dangers, enabling our company to take speedy activity and also defend versus any type of type of damage.

Equipments outfitted with sophisticated formulas may figure out and banner rare designs in financial and also different other info flows, assisting to stay clear of illegal tasks and also lower would-be reductions. These devices play an important function in seeking and stopping monetary criminal offenses.

## 4. Strengths and Weaknesses of Existing Approaches:

Each technique of collecting considerable appropriate records possesses its own unique advantages and also weak spots.

### File encryption:

Strong information guard of security supplies a sturdy personnel against unauthorized availability as well as also offenses, however it can additionally present difficulty along with rise expenditures for unit surveillance.

### Verification and Authorization:

Accomplishing multi-factor authorization and thorough recommendation methods may improve defense as well as also individual privacy, yet they might additionally lead to boosted managerial roles as well as customer frustration.

### AI & ML:

Offer long lasting efficacy in finding and analyzing possible hazards while maintaining reliability and also staying clear of inaccurate positives through making use of and also applying cutting-edge technologies deliberately.

# ****Conclusion:****

In conclusion, this essay explored the globe of securing big data examining the various susceptibilities it experiences and the problems in getting its own massive data sources. The exam of large range information breaches, such as DDoS attacks, malware seepages and SQL shots showcased the risks encountered by companies as exhibited by reality events like the Equifax violation and the Stuxnet procedure. Applying safety and security procedures is of utmost value because of potential economic and reputational consequences. This was emphasized by Equifaxs losses and noteworthy disturbances triggered by DDoS assaults, on preferred PC gaming platforms.

Examining access control concerns within the context of large data highlighted the diverse nature of the obstacles available, consisting of the monitoring of varied data resources, making sure scalability, and adapting to changing conditions. Standard techniques to controlling gain access to were discovered to be inadequate, and adaptable methods were essential to effectively take care of access. Real-world instances highlighted the significance of revoking accessibility promptly upon an employee's separation and the consequences of failing to do so.

The important analysis of existing safety methods, including file encryption, authentication, consent, and the transformative function of equipment discovery and AI, gives an understanding of The strengths and weaknesses of each method. At the same time, file encryption offers a robust defence, its. implementation complexities need to be recognized. Authentication and consent, however. improving protection, bringing management burdens. The assimilation of AI and maker discovery. Presents awesome hazard detection capabilities yet requires attention and care. application.

In the changing world of big quantities of data, where dangers are constantly transforming and innovation is constantly proceeding, this essay stresses the continuous need for creative security techniques. Future studies ought to concentrate on enhancing current methods and producing new ways to enhance big data versus emerging threats, guaranteeing its important role in advertising improvements while maintaining safety.

# Reference list

Adamokeeffe (2023). *Best Practices for Limiting Access to Sensitive Data in the Workplace*. [online] All Med Pro. Available at: https://www.allmedpro.co.uk/best-practices-for-limiting-access-to-sensitive-data-in-the-workplace/#:~:text=It [Accessed 12 Dec. 2023].

Auth0 (2023). *Role-Based Access Control*. [online] Auth0 Docs. Available at: https://auth0.com/docs/manage-users/access-control/rbac# [Accessed 12 Dec. 2023].

Bauer, L., Cranor, L.F., Reeder, R.W., Reiter, M.K. and Vaniea, K. (2009). Real Life Challenges in access-control Management. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*. [online] doi:https://doi.org/10.1145/1518701.1518838.

Botelho, B. and Bigelow, S.J. (2022) *What is Big Data and why is it important?, Data Management.* Available at: https://www.techtarget.com/searchdatamanagement/definition/big-data (Accessed: 27 November 2023).

Chen, J. and Ning, H. (2016). *Big Data Processing - an Overview | ScienceDirect Topics*. [online] www.sciencedirect.com. Available at: https://www.sciencedirect.com/topics/computer-science/big-data-processing.

CloudFlare (2023). What Is Encryption? | Types of Encryption | Cloudflare UK. *Cloudflare*. [online] Available at: https://www.cloudflare.com/en-gb/learning/ssl/what-is-encryption/.

Google Cloud (2023). *What Is Big Data?* [online] Google Cloud. Available at: https://cloud.google.com/learn/what-is-big-data [Accessed 12 Dec. 2023].

Google cloud (2021). *Confidential computing: Data Encryption during Processing*. [online] Google Cloud Blog. Available at: https://cloud.google.com/blog/products/identity-security/confidential-computing-data-encryption-during-processing# [Accessed 12 Dec. 2023].

Guo, H., Li, W., Nejad, M. and Shen, C.-C. (2019). *Access Control for Electronic Health Records with Hybrid Blockchain-Edge Architecture*. [online] ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/abstract/document/8946149.

Hedley, D. and Jacobs, M. (2017). The Shape of Things to come: the Equifax breach, the GDPR and open-source Security. *Computer Fraud & Security*, [online] 2017(11), pp.5–7. doi:https://doi.org/10.1016/s1361-3723(17)30080-5.

Information Commissioner's Office (2023). *Encryption and Data Transfer*. [online] ico.org.uk. Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/security/encryption/encryption-and-data-transfer/#:~:text=Encrypting%20personal%20data%20whilst%20it [Accessed 12 Dec. 2023].

onelogin (2023). *Authentication vs. Authorization: What’s the Difference? | OneLogin*. [online] www.onelogin.com. Available at: https://www.onelogin.com/learn/authentication-vs-authorization#:~:text=It%20is%20vital%20to%20note [Accessed 12 Dec. 2023].

Pratt, M.K. (2023) *Emerging cyber threats in 2023 from AI to Quantum to data poisoning*, *CSO Online*. Available at: https://www.csoonline.com/article/651125/emerging-cyber-threats-in-2023-from-ai-to-quantum-to-data-poisoning.html (Accessed: 25 November 2023).

Sudbring, A. and Naykodi, C. (2023). *2022 in review: DDoS Attack Trends and Insights*. [online] Microsoft Security Blog. Available at: https://www.microsoft.com/en-us/security/blog/2023/02/21/2022-in-review-ddos-attack-trends-and-insights/#:~:text=In%20March%202022%2C%20a%20DDoS [Accessed 12 Dec. 2023].

Suraj, M.V., Kumar Singh, N. and Singh Tomar, D. (2019). *Big Data Analytics of Cyber attacks: a Review | IEEE Conference Publication | IEEE Xplore*. [online] ieeexplore.ieee.org. Available at: https://ieeexplore.ieee.org/abstract/document/8541263/authors#authors [Accessed 4 Dec. 2023].

Tunggal, A.T. (no date) *The 72 biggest data breaches of all time [updated 2023]: Upguard*, *RSS*. Available at: https://www.upguard.com/blog/biggest-data-breaches (Accessed: 27 November 2023).

UTrust for Software Testing & Quality Services (2023). *Enhancing Big Data Security: Best Practices for Protecting Valuable Data*. [online] www.linkedin.com. Available at: https://www.linkedin.com/pulse/enhancing-big-data-security-best-practices-protecting-valuable-ulrkf [Accessed 12 Dec. 2023].

Van Dine, A. (2017). *Acknowledging the Cyber Threat to Nuclear Facilities*. [online] Center for Strategic and International Studies. Available at: https://www.jstor.org/stable/resrep23162.11 [Accessed 12 Dec. 2023].