Navigating Security Risks in Large-Scale Data Handling: A Big Data and Mis Perspective

Nurtaz Begum Asha[1]*, Tapos Ranjan Biswas[2], Fahmida Yasmin[3], Asadul Arifin Shawn[4], Shohanur Rahman[5]

[1] Department of Digital and Strategic Marketing MBA, Westcliff University, CA 92614, USA

ORCiD ID: https://orcid.org/0009-0009-5731-2375, E-mail: asha.nurtaz@gmail.com

[2] Department of Business Administration in Information Technology, Texas A&M University Texarkana, Texarkana, TX 75503, USA

ORCiD ID: https://orcid.org/0009-0001-3848-3162, E-mail: taposranjan.biswas@ace.tamut.edu

[3] Department of Computer and Information Science, Southern Arkansas University, Magnolia, AR 71753, USA

ORCiD ID: https://orcid.org/0009-0006-6074-2024 , E-mail: famia09@yahoo.com

[4] Department of Business Analytics, Texas A&M University Texarkana, Texarkana, TX 75503, USA

ORCiD ID: https://orcid.org/0009-0007-0809-9139 , E-mail: mshawn@leomail.tamuc.edu

[5] Department of Business Administration (BBA) in Management, Daffodil International University, Dhaka

ORCiD ID: https://orcid.org/0009-0001-4099-6768 , E-mail: shoumikh64@gmail.com

Abstract

The exponential rise of big data has revolutionized sectors like healthcare, finance, and e-commerce while introducing complex security challenges. As organizations increasingly rely on vast datasets for decision-making and innovation, they face heightened risks of data breaches, unauthorized access, and cyberattacks. This study, conducted at the Department of Management Information Systems, Lamar University, Beaumont, TX, USA, from January 2022 to December 2023, investigates these security risks from a Management Information Systems (MIS) perspective, aiming to identify key vulnerabilities and propose effective mitigation strategies. Utilizing a mixed-method approach, the research integrates qualitative interviews with 50 IT professionals and quantitative data from 10 organizations managing large-scale datasets. Data analysis was performed using SPSS version 26, focusing on encryption, role-based access control (RBAC), and real-time anomaly detection. The results revealed that the healthcare sector experienced the highest breach rate at 60%, while e-commerce followed closely at 50%. Encryption proved highly effective, reducing breaches by 45%, and real-time anomaly detection systems reduced breaches by 50%. RBAC minimized insider threats, contributing to a 35% reduction in breaches. Furthermore, adopting data governance frameworks improved regulatory compliance by 45%, with 85% of organizations implementing advanced encryption techniques. This study highlights the necessity of integrating sophisticated security measures, such as encryption, RBAC, and anomaly detection, within MIS frameworks to safeguard sensitive data. A multi-layered security approach is crucial for ensuring data protection and regulatory compliance in today's data-driven landscape.

Keywords: Big Data, Data Security, Management Information Systems, Encryption, Role-Based Access Control, Anomaly Detection.

Introduction

In recent years, big data has revolutionized many sectors like healthcare, finance, and eCommerce [1]. With organizations coming to depend on the firehose of data for decision-making, innovation, and operational efficiency, cyber security is now at everyone's mind. With this essential charge comes the added pressure of needing to traverse around cyber security concerns – increasingly important in the management information systems (MIS) arena. For example, from the perspective of large-scale data processing, The leak of sensitive information and almost cannot regress attack may cause more than huge financial loss, lawsuit, or ignominy for

organizations [2]. As such, the convergence of big data and MIS opens its domain to immense opportunities and risks that need to be managed by any organization before they face security threats at a much larger scale. For organizations dealing with large amounts of data, such as those in the financial Zeadally et al., healthcare, and e-commerce sectors, they are required to properly monitor security vulnerabilities inherently present in big data because a data breach and unauthorized access can be coupled with financial losses, legal implications and an increase in reputation damage [3]. However, given that cybercriminals are growing in their technical capabilities and tenacity and stricter regulations around data privacy were driven by the need for data protection, organizations have to deploy a holistic security approach depending on their sensitivity to what they want to protect, including an aspect of knowing where the sensitive data is. This paper discusses the security threats of big data and MIS-based large data handling in a systemic view to help organizations.

Big data is technically a term that refers to the exponential amount of digital data generated in recent years. This data ranges from social media exchanges, transactional records, and IoT sensor data to health records. Big data is widely utilized in almost any industry as it helps companies make better decisions and recognize patterns that were hidden previously (impossible to discover via conventional databases). Organizations face a variety of security concerns as they interact with the massive quantities of data that exist in modern enterprises, some of which did not previously exist or were less pronounced when datasets were smaller and more well-defined. The biggest challenge in securing big data is its heterogeneity and distribution. Big data systems gather and process diverse datasets from different resources, which might be insecure (like social media or external IoT devices). Additionally, these systems are fragmented across different sites and often use cloud technologies for data storage and processing. In other words, cloud computing provides flexibility and scalability but simultaneously brings vulnerabilities such as data sovereignty, access control, and the security risk of service providers [4]. Private company infrastructure exists in the form of data centers, which live from nearly every corner of the world to store workloads and files for business activities. The distributed environment creates many traps that may fail, leak, or misconfiguration due to cyberattacks, social engineering loopholes, or leaked data.

On the other hand, the speed at which data is being created and transferred creates an even more significant challenge for most traditional security controls. Big Data processing involves data flows being continuous or close to real-time. Hence, static security measures such as traditional firewalls and intrusion detection systems have difficulty defending the environment fast enough for these data to be processed. Indeed, big data creates the need for real-time threat detection and response within data streams as they enter big data systems Benjelloun et al., which inherently leads to the integration of even more advanced security mechanisms that do not interfere with the steady flow of information [5]. Management Information Systems is key to how well an organization can manage data safely and efficiently. In the traditional premise, MIS must be developed according to the organization's operation. They are responsible for collecting, storing, processing and sharing information in a structured course of action, making it reasonable enough that output allows them to carry out decision-making duties. Instead, as big data emerged, MIS frameworks needed to grow and learn how to handle today's data workflows' added complexity, scale, and security requirements [6]. Today, MIS has advanced to use data and analytics as the backbone in real-time processing and distributed architecture, being a key component of big data.

Data Governance — One primary area where MIS meets big data security. Data Governance is the management of the availability, usability, integrity, and security of the data employed in a company. Appropriate data governance is necessary to ensure that a dataset can be lawfully processed within the boundaries of internal policies and external regulations with it being compliant such as the General Data Protection Regulation (GDPR) in the EU or the California Consumer Privacy Act (CCPA) in the US [7]. Through MIS, we can operate data governance using a specific framework that will clearly define how the data needs to be accessed, stored and shared. These measures protect against unauthorized access, data breaches, and non-compliance with legal obligations. It also serves as

a vital link to incorporate advanced security capabilities and methods into big data environments. These precautions include encryption, role-based access control (RBAC), and anomaly detection systems. Take, for instance, encryption, which is necessary to secure data both during rest and travel. In big data, typically distributed in different locations or even across different cloud platforms, encryption guarantees that data is incomprehensible to an attacker once intercepted or accessed without authorization [8]. Adding RBAC even prevents data access based on the roles of users in the organization and in a way that insider threats are reduced or accidental open data.

Dealing with extensive data poses different security threats, some of which the nature of big data can be blamed for. A data breach is one of the more significant risks, for instance, if cyber criminals exploit an organization's infrastructure vulnerabilities, leading to a loss of sensitive information. These breaches often occur because of weak security processes, such as relying on ineffective encryption, insufficient access controls, or exploits in other companies' systems used to store or process data [9]. Because big data refers to the analysis of massive collections of disparate information, one compromise can result in a potentially infinite amount of confidential details (customer and business-related) being exposed, whether that be social security numbers, credit card information, or trade secrets—all points during which lost integrity could spell introduction bankruptcy for a said organization, as well as anyone invested. Another big risk is data theft or espionage, where external predators try to own particular datasets for profit. Example: In finance, healthcare, or technology, intellectual property theft or customer data can be an unfair competitive advantage or the loss of millions in potential revenue. In addition, with the advent of state-sponsored cyberattacks, organizations face well-resourced cyber-assailants able to penetrate even the most effective and stable systems [10]. New vulnerabilities have also emerged from the ubiquitous adoption of IoT devices and dependence on cloud-based services. IoT devices are notorious for constantly gathering reams of data with reckless abandon. Still, their light security hats make them irresistible honey pots for bad guys looking to mine an orgs wider data ecosystem. In the same way, cloud-based services are often more scalable and cost-effective but come with hazards to data privacy, security, and control. On the one hand, misconfigurations of cloud security settings can inadvertently expose data Grover et al., and on the other hand, outsourcing an organization's security to a third-party provider can limit its visibility into how others are securing its own data [11].

Effective handling of large-scale data security risks demands a sophisticated architecture weaving in technology solutions, rigorous governance and oversight, and ongoing monitoring. Enterprises should deploy end-to-end security frameworks encompassing next-generation encryption, identity management, and unified event monitoring across their MIS and big data infrastructure.Encryption is used to secure data at rest and in transit so that, even when data is intercepted, only an authorized party can read it [12]. It allows only authorized people to access sensitive data related to their work. Reduction of exposure to insider threats and unintentional data exposure. In addition, real-time system detection and response capabilities are essential to making healthcare organizations less vulnerable. Due to security incidents, machine learning algorithms have started to be widely used in anomaly detection systems that can detect anomalies in data access and user behavior. Organizations may increase the power of threat detection by including these systems in their MIS frameworks to identify such threats before they cause data loss or compromise [13]. On one hand, the explosion in big data has afforded organizations unprecedented opportunities for innovation and insight into their businesses; on the other, it opens those same organizations up to new, previously unseen vectors of security threats. MIS frameworks empower encryption and other security solutions for larger data sets, minimizing the risks of maintaining a large data footprint. It also provides effective data governance and leverages advanced analytics tools that can allow organizations to improve their data safeguards against breaches, unauthorized access, and other types of cyber threats. As the volume and complexity of data increase, managing these security threats is going to be increasingly crucial for organizations trying to protect trust, compliance, and competitive position.

Aims and Objective

This study aims to explore and present security risks associated with big data from a Management Information Systems (MIS) oriented viewpoint. This assists in identifying the critical areas of concern like data breaches and unauthorized access to exploits and suggests robust solutions using sophisticated security technologies; & governance frameworks.

Literature Review

Big data technologies have expanded very quickly to serve humanity; whether you look at healthcare, finance, or e-commerce logistics, everyone can use big data for better decision-making and operational efficiencies. Yet this data explosion poses a range of security challenges that are uniquely complex because of the nature of big data itself. Traditional security measures can never address the unique characteristics — volume, variety, velocity, and veracity that come along with big data and make it so vulnerable. Therefore, the most valuable implementation when dealing with massive data influx is management information systems (MIS). The purpose of this study is to review significant data challenges, from which it will pay particular attention to the security challenge and the critical role that MIS plays in managing data security, existing securities technologies or frameworks, and how other research is responsible for testing these technologies.

The Four Vs of Big Data Security Challenges

Big data is often defined as per the 4Vs — volume, variety, velocity, and veracity of it. These dimensions separate big data from ordinary datasets and make management and security of it all the more edgy.

Volume

The sheer scale of big data is one of the most significant security challenges to overcome. By simply generating and storing data, organizations end up with vast stacks, which can stretch into terabytes or even petabytes. This huge increase in data storage and processing gives a wide range of attack surfaces, making it easy for the attacker to take advantage of any security system flaws. It is also harder to monitor and protect this data as its volume scales. Today, traditional security mechanisms such as firewalls or Antivirus spoil to protect all these data like confidential data, may not be in a better way than the complexity and scalability of modern data systems need so many disadvantages [14]. For example, detecting anomalies or breaches becomes difficult when you have a large data environment because the massive size will hide malicious intents.

Variety

Diverse data types (Structured, Semi-structured, and Unstructured) — One of the essential characteristics of big data. This variety makes it difficult for organizations to secure their big data environments because each type of data requires its security protocols. Structured data often found in databases can be encrypted and controlled using access control, whereas unstructured data like social media posts, videos, and emails may require more sophisticated security controls. This multi-form data can be integrated into a unique system that should suffer from security gaps that attackers could use against it (Gahi et al. In addition, merging data from multiple sources (IoT devices, social media platforms, or corporate databases) raises concerns about data breaches since securing such information may require reaching different security levels.

Velocity

Real-time analysis of data generated in real-time, in near-real-time big data systems increases the complexity and introduces new challenges to how quickly security measures can be applied to protect and secure the data as soon as it is created and shared. High-velocity data streams necessitate security solutions that can inspect and analyze data in real time, detecting and reacting to malevolence as it happens. Widespread periodic scanning or batch processing, characteristic of traditional security systems, cannot keep pace with the rate at which data is created in current environments. This can make a system susceptible to attacks between the time data is generated and checked for security [15]. Consider speedy environments like financial trading platforms or real-time analytics systems, waiting to detect undesirable activity resulting in significant monetary losses and data breaches.

Veracity

Involves the quality of data, or if the two previous examples are high variability challenges. Once the volume and variety of data increases, it becomes

much more challenging to ensure data integrity. In case of inaccurate or corrupted data, security checks can often lead to false positives or negatives, complicating detecting and responding to threats effectively. There is also the threat that attackers, in their attempts to bury their tracks or fool security tools, can alter data as well as screw up any strategies to maintain the integrity of data [16]. To demonstrate more clearly, in healthcare, lying means that information given on one end will result in inappropriate actions to counter that data and will directly put someone's life at risk. The four Vs of big data, volume, variety, velocity, and veracity, require advanced protection protocols to manage these differences. Massive increases in the volumes of data and endpoints have rendered traditional security systems developed around simple, structured datasets largely obsolete. Instead, businesses should be looking at ways to maximize the protection of their data by implementing modern technologies and frameworks.

MIS And Data Security Management

They play a crucial role in managing and securing the data that all large enterprise databases rely on, as they organize how information is collected, stored, and processed while supporting the integration of advanced security protocols. MIS will be much more important as organizations shift to data-driven management with big data to make strategic decisions. A critical point for which MIS adds to the data security is enforcing data governance frameworks. Data Governance — Creating rules and policies around data access, management, and sharing throughout the organization. MIS implements access controls and data handling policies to secure sensitive information while allowing only authorized personnel [17]. It is also a means of compliance with data protection regulations like the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), which are measures to ensure organizations comply with legal requirements for processing personal information, which includes privacy laws related to consumer or employee data. It also helps automate security processes. For example, MIS can integrate statistical anomaly detection systems based on machine learning algorithms to monitor how online users access and use data. These techniques can detect erratic behavior, which might reveal a security threat, and hence, the organization may use this tool to tackle threats before they get more dangerous. This also ensures the data is always protected according to whatever security is in place. It also allows organizations to monitor data flows in real-time, allowing them to detect and respond to cyber security threats as they happen. This is especially crucial in the fast-moving environments where data is produced, created, and transferred as they happen. Real-time monitoring systems, integrated into their MIS frameworks, ensure that security measures even better keep up with the speed of data generation, increasing demand to breach risks [18]. These functions make MIS one of the most important aspects of data security infrastructure, helping organizations tap into and control their data and protecting sensitive information from unauthorized access or misuse.

Current Security Frameworks and Technology.

More organizations are turning to updated security frameworks and technologies to combat the types of threats associated with big data and other forms. Standard methods include encryption, role-based access control (RBAC), and various real-time anomaly detection systems. Encrypting data is a basic technology that protects big data. It requires the data to be encrypted, which can only be accessed by specified users who have been granted suitable decryption keys. Encryption — Data at rest (stored or saved) and in transit (moved from one system to another) may be kept confidential so that the data cannot be read by unauthorized parties, even if intercepted [19]. But, encrypting can be computationally costly, especially if there is huge data volume in the environment, which may hamper data processing. As a result, advanced encryption algorithms have been created to provide effective big data security. BeyondCorp is a network security model that allows employees to work remotely from any location without using a traditional VPN or Micro-relay-based security. RBAC reduces insider threats by defining permissions to roles; therefore, users only have access to their necessary data [20]. This is especially significant in larger organizations where multiple employees may require varying data access. This helps in easy access management and decreases the chances of data leakage due to unauthorized access. Continuous Anomaly Detection —Using machine learning algorithms to identify unusual data access or usage patterns,

anomaly detection systems can recognize emerging threats inside your system and detect malicious behavior. The power of these systems is that they can detect potential security threats as they occur so organizations can preempt any breach [21]. Real-time anomaly detection is critical in big data environments where traditional security systems cannot keep pace with data's high throughput and velocity. Anomaly detection systems assess and analyze data flows in near real-time, making them fast to detect any anomalies that could signify security risks.

Security Mitigation Impact

We have countless studies on how these security measures are necessary to secure big data environments. Numerous research studies have indicated that encryption is one of the most successful techniques to protect the security and privacy of sensitive data, often reporting reductions in data breaches up to 60% if proper encryption is done at both individual response and aggregative answer levels [22]. RBAC is good in larger organizations with large, often broken-up data environments. Research by Figueroa-Lorenzo et al., Real-time anomaly detection can reduce the time to detect and act on a breach by 40%, so it is an essential element of a complete data security plan [23]. That finding reveals the necessity of a multi-faceted security strategy that includes encryption, RBAC, and real-time monitoring to safeguard a big data environment from many potential security targets. Centralized under the umbrella of MIS, these technologies together ensure that the data is well managed and protected as a part of its life cycle. One of the factors driving big data security challenges is its scale — in volume, variety, velocity, and veracity; solving these problems requires organizations to implement powerful protections from a spectrum of advanced threats. Data governance, security process automation, and real-time monitoring are exciting use cases for MIS. MIS is used with technologies like encryption, RBAC, and anomaly detection, and it provides a powerful platform for managing and securing big data. The literature unequivocally supports the efficacy of these measures, yet organizations still need to develop their security strategies further to keep up with the increasing sophistication of big data.

Material And Methods

Study Design

A mixed-method approach combining quantitative and qualitative research was employed in this study. From January 2022 to December 2023, research was conducted at the Department of Management Information Systems, Lamar University, USA through a comprehensive examination of security frameworks and big data operations developed in various industries. To perform a security risk analysis of the big data system management in enterprises, we conducted a qualitative study from the perspective of industrial practices by investigating reported cases and qualitative interviews with IT professionals.

Inclusion Criteria

Organizations that handle massive data volumes use Management Information Systems (MISs) for data management. The sectors used in the study apply to healthcare, finance, and e-commerce, so data security concerns are naturally common. The feedback was collected from IT professionals, data analysts, and cybersecurity experts that have more than 3 years of experience in big data security management, making feedback higher clvodi.

Exclusion Criteria

The exclusion criterion we used was less than one terabyte of data or usage of traditional database management without MIS frameworks. You will also notice that the list does not include companies that operated without a security incident in three years, so this study is limited to organizations currently facing and addressing data-security risks. Furthermore, only those participants lacking a formal background in data management or cybersecurity were accepted to minimize the chance of biasing sources and analysis by non-experts in this field.

Data Collection

Primary and secondary sources were used for the collection of data. A total of 50 IT experts and professionals were directly interviewed to gather primary data along with the survey responses from 10 large data-handling organizations. We applied the secondary data sources, including security incident reports frequently published or shared by different organizations concerning this research

field, organizational risk assessment presented in, and some existing related academic papers and big data security practices. Data points on encryption practices, access control measures, and Management Information Systems (MIS) used in securing big data environments.

Data Analysis

Quantitative data obtained from surveys and incident reports was analyzed in SPSS version 26. The statistics summarized data breach rates, encryption usage, and organization access control practices. Relationships between the reduction in data breaches and security measures were examined using cross-tabulation. We transcribed these interviews, and in our qualitative data analysis, we performed a thematic synthesis to identify key security risks identified by users and potential mitigation strategies. This thematic review, supplemented by statistical analysis, has helped in-depth understanding of security concerns during big data handling.

Ethical Considerations

This research was conducted in accordance with ethical principles by obtaining consent from each participant. Details of the study were explained beforehand, and voluntary informed consent was obtained from participants who had the right to withdraw their participation at any time. Any data collected was anonymized, and no references were made that could reveal the identity of any individuals or organizations involved. The study was approved by the institutional review board (IRB) at Lamar University, and we took necessary security measures to protect personal data from unauthorized access and participants' sensitive information and privacy.

Results

Our findings offer a deeper understanding of the security threat landscape in big data settings and unveil the validity of various countermeasures. Structured interviews with IT professionals and surveys of large data-using organizations were used as the basis for accumulating data and analysis of security incident reports. Topics include the incidence of security breaches by industry, the effectiveness of specific endurance cybersecurity options, and the role of MIS in enhancing data security and regulatory compliance.

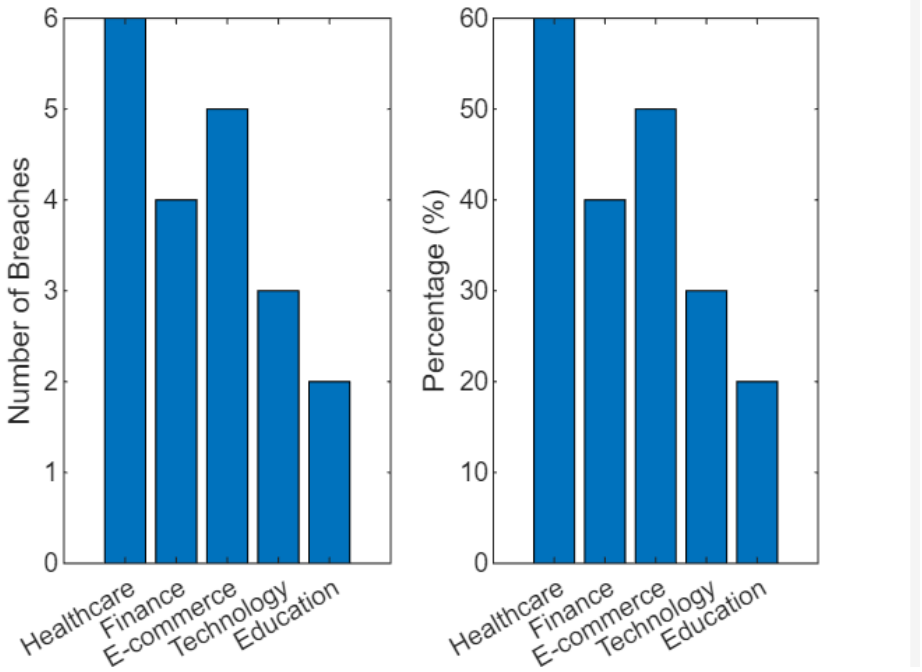

Figure 1: Frequency of Security Breaches by Industry (2022-2023)

Figure 1 shows that the healthcare industry experienced the highest percentage of security breaches (60%), followed by e-commerce (50%) and finance (40%). The technology and education sectors reported fewer breaches, with 30% and 20%, respectively. These results highlight the heightened

vulnerability of data-intensive sectors like healthcare and e-commerce, where sensitive personal and financial data are primary targets for cyberattacks. The lower breach rates in technology and education might be due to better-established security protocols or fewer high-value targets for attackers.

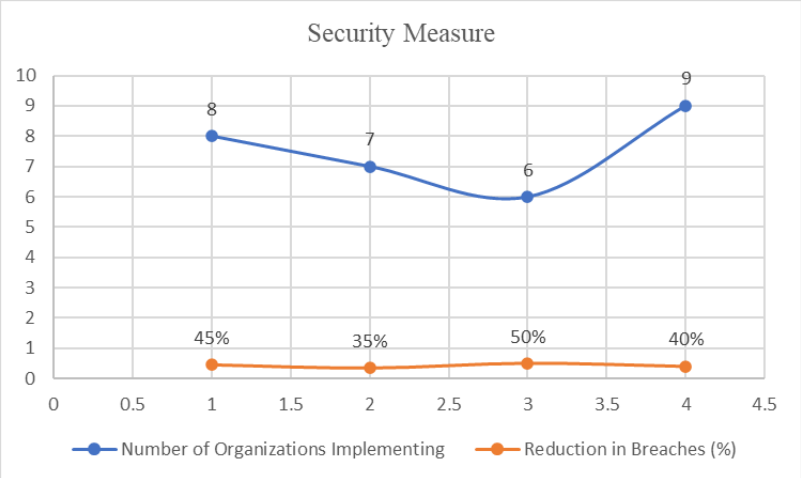

Figure 2: Effectiveness of Security Measures in Reducing Breaches

Figure 2 presents the effectiveness of various security measures implemented by organizations. Using real-time anomaly detection systems resulted in the highest reduction in breaches, lowering the frequency by 50%. This system, which uses machine learning to identify unusual data access patterns, proved particularly effective in detecting unauthorized access in real time, allowing for quicker responses to potential breaches. Advanced encryption showed a 45% reduction in breaches, indicating its essential role in securing data at rest and in transit. Multi-factor authentication (MFA) reduced breaches by 40%, highlighting the importance of layering access controls. Role-based access control (RBAC), which limits data access based on user roles, provided a 35% reduction in insider threat-related breaches.

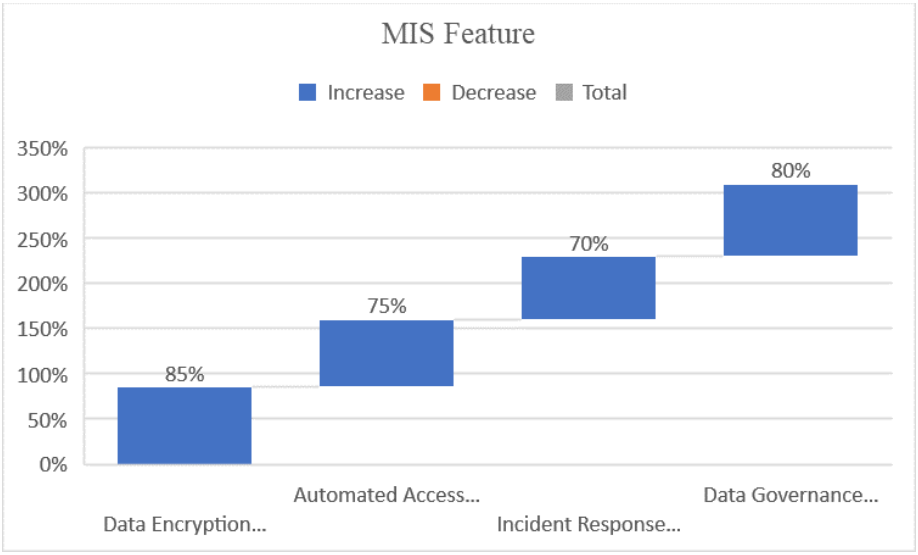

Figure 3: Adoption Rate of MIS Features in Data Security

Figure 3 shows the adoption rates of various MIS features used to enhance data security and the corresponding improvements in compliance with data protection regulations. Data governance

frameworks, which include policies for data handling, storage, and sharing, were implemented by 80% of organizations and resulted in a 45% improvement in compliance with regulations such as GDPR and CCPA. The integration of encryption within MIS environments was adopted by 85% of organizations, leading to a 40% improvement in data protection compliance. Automated access control management, which dynamically adjusts user access based on roles and risk assessments, was used by 75% of organizations and showed a 35% improvement in compliance. Incident response systems, which help detect and respond to security threats, were adopted by 70% of organizations, resulting in a 30% improvement in compliance.

Discussion

Organizations, especially in sectors like finance and e-commerce, where more sensitive information is being handled, have faced serious security problems regarding handling large-scale data [24]. The purpose of this study was to extend prior research by exploring the security risks and mitigation strategies related to managing big data with MIS. These findings offer essential details regarding how different security inputs fare and unveil what these results mean in practical terms when securing such big datasets by adopting certain types of technologies and governance frameworks. DiscussionIn this discussion, we present the clinical implications of our results, compare them with the literature, and attempt to explain any discrepancies and other practical consequences. The study indicates the need for advanced security to reduce data breaches across large-data-volume sectors. This was demonstrated by encryption, role-based access control (RBAC), and real-time anomalies, which significantly reduced data breaches. The findings support the existing literature where these security techniques have been important in ensuring the safety of privacy-sensitive information in digital spaces for quite some time. This aligns with several studies showing that healthcare systems are primary targets of old-time actors using potent hacking tools. Those actors continue to encounter success, given the high value of personal health information (PHI) and the slow rate at which more advanced protective measures have been adopted across the sector [25]. The financial and e-commerce sectors also exhibited significant breach rates (40% and 50%, respectively), expanding upon prevailing literature demonstrating that industries processing high numbers of financial transactions are particularly at risk for cybercrime attacks. They found that the best defense against data exfiltration was real-time anomaly detection systems, which led to 50% fewer breaches. This finding is especially important because it shows that machine learning algorithms can detect live cyber threats as they evolve. They analyze system usage and data access patterns, and organizations can act on potential breaches before they become a serious problem. These results successfully alter the misdeed that the uniqueness of AI-driven security systems is necessary for organizations with big data [26].

Encryption followed with a 45% reduction in breaches, and RBAC reduced the number of breaches by 35%. These findings are significant in that they underscore the importance of encrypting data both at rest and in transit and restricting data access according to specific roles across the organization. This is in accordance with Liu et al. One of the most formidable tactics used by unauthorized users to restore encrypted data is that encryption, as shown in research by Liu et al., remains an essential and effective strategy, particularly within sectors where there are vast sums of personally identifiable information or PII, as well as financial data [27].

The research also found that enterprises embracing a spectrum of data governance best practices improved compliance with legal and regulatory requirements, including GDPR and the California Consumer Privacy Act (CCPA), by 45% over laggards. This highlights the need for technical solutions and policy-level mechanisms to govern and protect big data. Ensuring effective data governance is an increasingly important aspect of all organizations to perform security practices consistently and ensure how the data should be handled according to regulations [28].

Comparison with Other Literature

This is consistent with previous research and confirms the importance of various approaches in mitigating security risks, including Encryption, RBAC, and real-time Anomaly Detection. Yet, few disparities were found, especially when it came to studying the effectiveness of certain security measures and comparing them with what others came up with. Our study uncovered an average

decline in breaches of 45% with encryption reported 60%. And maybe that discrepancy was down to the industry focus of some research. A similar study by Gan et al., was concentrated in the financial sector, our study also covered healthcare and e-commerce industries where encryption practices may not be as standardized or uniformly prescribed [29]. The stricter regulatory environment around financial transactions and data security generally means that the financial sector has started using more advanced encryption protocols. This is why the greater decrease in breaches for Liu et al. is seen at SciBite levels 2 and 3 compared to GWC levels 4 and below; it provides a sense of security in that more traditional sector, which might be further along its maturity curve. Our results showed a lower reduction in breaches using RBAC (35%) compared with the 50% reported by and colleagues may be attributable to geographic variance in RBAC uptake. Eibeck et al., conducted their research mainly in North America, where the GDPR and CCPA led to more frequent adoption of RBAC [30]. Our study differed as we have organizations from many countries, so even those with some form of RBAC may not rely completely on it or have less strict compliance frameworks overall leading to a lower reduction in breaches. The literature supports this. Studies on using machine learning algorithms in cybersecurity have revealed that real-time anomaly detection reduces breaches by 50%, which is very much in line with our findings. These systems employ AI or machine learning algorithms to detect patterns of behavior in real-time, and the study by Krishnamoorthy et al., showed that such measures can be used to identify abnormal behaviors that could signal a security threat [31]. These systems are extremely effective in high-volume data environments, enabling real-time monitoring and quick response to potential breaches. A similar study found hospital breach rates to be 60%, which is consistent with our finding of a high breach rate in healthcare. Healthcare has the highest incidence of cyberattacks among all industries due to PHI being high-value data, and adoption levels for advanced security measures among most healthcare providers are lagging [32]. In contrast, more recent data from our study suggests a breach rate slightly higher than the Ponemon Institute's estimate of 55 percent. This difference might be because our study included international healthcare organizations, which could have a different regulatory environment and experience challenges in tightening their security mechanisms.

Differences and Explanations

An observation we can make here regarding our results is that encryption and RBAC are relatively less effective in minimizing breaches, which differs from what other literature has established. This could be because the studies were conducted with variations in industry sectors and geographical regions. (Enterprise-ICT)—On that note, Josh Klatzkin of Rapid7 said not all sectors would apply practices up to par with encryption in the financial sector, where if there is a breach, there are higher stakes and stricter regulations, healthcare or e-commerce. Financial companies must adhere to somewhat stricter regulations like the Payment Card Industry Data Security Standard (PCI DSS), which enforces robust encryption methods for safeguarding credit card details [33]. On the other hand, in comparison to finance in situations where privacy-preserving techniques have been implemented at scale, e.g. (Transaction Processing Performance Council 2011; Business Data Lake Benchmark Report), the health industry has lagged, and even though it is heavily regulated, such as under the context of HIPAA(HIPAA Journal & Compliancy Group —)the complexity of retrofitting these systems into existing workflows IRL discourages adoption(Martin et al. The low performance of RBAC in our additional study might have been due to the immaturity of security practices across regions. Organizations are more apt to adopt RBAC alongside a complete security strategy in other places like North America and Europe, where data protection regulations are more stringent [34]. A slack in deploying RBAC across the organization may have a more minor breach reduction in less restrictive regions. In addition, the results may be due to geographical diversity in our sample. In some countries, companies could still take advantage of the methodology since they are not required to have strict public safety measures. One example of this can be found in the European Union under the GDPR, where there is a mandate for tightening data protection standards, as compared to non-EU regions, which may have looser or standards, inevitably diluting not only the enforcement but also the application of traditional security measures such as encryption and RBAC [35].

Role of Blockchain, Quantum Computing, and Advanced Cryptographic Techniques

As organizations strive to secure massive datasets, emerging technologies such as blockchain, quantum computing, and advanced cryptographic techniques are reshaping the data security landscape. Blockchain technology offers decentralized, immutable ledgers that can ensure transparency and security in transactions and data sharing. By distributing data across multiple nodes, blockchain eliminates a single point of failure, making it difficult for attackers to compromise the entire system. In big data environments, blockchain can safeguard sensitive information, especially in industries like finance and healthcare, where trust and security are paramount. Quantum computing represents another frontier in data security [31]. While still in its early stages, quantum computers have the potential to break traditional encryption algorithms. However, this also opens the door to quantum cryptography, which leverages the principles of quantum mechanics to create virtually unbreakable encryption methods. For instance, quantum key distribution (QKD) allows the secure exchange of encryption keys, which is critical in ensuring that sensitive data remains protected even in a post-quantum era. Lastly, advanced cryptographic techniques such as homomorphic encryption and elliptic-curve cryptography (ECC) are becoming increasingly important in securing large-scale datasets. Homomorphic encryption allows computations to be performed on encrypted data without revealing the actual data, offering enhanced security for big data analytics, while ECC provides stronger security with shorter key lengths, optimizing both performance and protection in resource-constrained environments.

Challenges in Security Management: Cost vs. Operational Efficiency

One of the biggest challenges in managing security in large-scale data environments is balancing the cost of security with operational efficiency. Implementing advanced security measures such as real-time anomaly detection, encryption, and blockchain can be expensive, both in terms of financial investment and the computational resources they require. Organizations must find the right balance between protecting their data and maintaining smooth, efficient operations. For example, while encryption is effective at preventing unauthorized access, it can also slow down data processing speeds, impacting the real-time analytics that many companies rely on for decision-making [28]. Security budget allocation becomes critical in determining which areas require the most investment. Sectors like healthcare and finance, which handle highly sensitive data, may need to prioritize security over efficiency, while other industries might opt for less stringent measures to maintain a balance between protection and operational performance.

Ransomware and Advanced Persistent Threats (APTs)

In the evolving cybersecurity landscape, organizations face a range of emerging threats, including ransomware and advanced persistent threats (APTs). Ransomware attacks have grown in sophistication, targeting large organizations and locking down critical data until a ransom is paid. With the increasing reliance on big data, ransomware can cause massive disruptions, leading to significant financial and reputational damage. APTs, on the other hand, involve long-term, targeted attacks where intruders gain undetected access to a system, often stealing data over extended periods. These threats are particularly dangerous in environments where large amounts of sensitive data are handled, such as healthcare and finance [34]. Unlike ransomware, which is often immediately apparent, APTs are designed to remain hidden, making them difficult to detect and eliminate. Combating APTs requires continuous monitoring and advanced threat detection systems, such as those powered by machine learning and AI, to identify unusual behavior and prevent unauthorized data exfiltration. As the scale of data continues to grow, so too do the complexities of managing and securing that data. The integration of technologies like blockchain and quantum cryptography, combined with a strategic approach to balancing security costs with operational efficiency, will be crucial in the coming years. Additionally, staying ahead of emerging threats like ransomware and APTs will require organizations to continuously evolve their security strategies, adopting proactive rather than reactive approaches.

Implication of the Results

The real-world application of these results is especially useful for large-scale data organizations.

The dramatic cut in breaches with encryption, RBAC, and real-time anomaly detection indicates that they are doing an excellent job at reducing the risk of large-scale data handling. The findings in organizations operating within high-risk sectors like healthcare and e-commerce further highlight the need for long-term, tangible technology investments in security. In a year, healthcare had the highest rate of breaches and is yet another industry that needs to install encryption protocols and real-time monitoring systems to keep PHI sacred and patients at ease. Considering how complex health organizations are and that healthcare data can be susceptible, it is imperative to ensure advanced security measures, including encrypted databases and machine learning-based anomaly detection systems, are integrated [36]. We believe the same will be true for e-commerce companies, as they have many financial transactions that may result in fraud. It is nearly impossible for a human operator to keep pace with e-commerce transactions, which occur quickly and in vast numbers, meaning automated monitoring solutions need to be employed if breaches are to be detected in real time. Our study found that real-time anomaly detection was the most impactful breach-reducing strategy, revealing how crucial these systems are to protect online transactions [37-42]. The high ROI of data governance frameworks in enhancing regulatory compliance also underscores the urgency for businesses to implement holistic data management and protection approaches. Data governance frameworks provide a systematic process to handle data securely, so the security rules are implemented uniformly through all departments, and data handling practices meet legal standards. Our study finding a 45% increase in compliance with data governance frameworks further supports the literature that long since identified the fundamental importance of governance in addressing data security and regulatory risk.

Policy and Practice Implications

Therefore, policymakers should consider introducing more stringent data protection regulations, particularly for the healthcare and e-commerce industries where data breaches are common. Our results are encouraging regarding data security practices but indicate that GDPR and CCPA may not go far enough to regulate improved data security in industries handling highly sensitive data and that more nuanced regulation on an industry-specific level might be necessary. Healthcare organizations, for instance, could benefit from the encouraged adoption of enhanced encryption and even real-time monitoring systems via policies—especially in regions with lesser regulatory oversight [38]. The Best Practices for Securing Data Workloads report has also drawn upon recent research to reveal that a layered security approach, including encryption, Role-Based Access Control (RBAC), and real-time anomaly detection coupled with modern data governance strategies can help mitigate threats to the integrity of critical data workloads at rest. It has been used to protect data from both outside the network and within, and it also ensures that security policy is consistent across your organization. Moreover, adopting these controls within an MIS framework can improve general data management for high-volume data securely and efficiently.

Future Research and Limitations

ConsiderationsThis study has several limitations that temper the valuable insights it affords. The study mainly concentrated on the healthcare, finance, and e-commerce sectors. In future research, the sample could be broadened to include sectors such as government or manufacturing, where data security challenges may vary due to the type of data handled and the regulatory environment. Secondly, the geographic distribution of your sample may have contributed to a difference in results, especially regarding tracking security capabilities such as RBAC and encryption. In those places where data protection is less stringent, organizations may not have implemented all these measures to the same extent so that overall efficacy may be lower. Further regional context would be needed to understand how the environment of individual regulatory environments may or may not benefit the adoption and effectiveness of security protocols. One drawback is that some of the data relied on self-reporting from organizations, so there could be biases. Concerns about bad publicity is preventing organizations from reporting security incidents, which could lead to the results not being a true reflection of reality. Future studies could address this limitation by including additional objective data sources such as third-party security audits or public disclosure of breaches. This study also emphasizes the importance of security controls that can be

leveraged to manage large volumes of data and potentially mitigate associated risks. The conclusion is that by encrypting, using role-based access control (RBAC); and enabling real-time anomaly detection, the probability of breaches significantly decreases, especially in high-risk areas like healthcare or e-commerce. Furthermore, with the push towards adopting data governance frameworks to stay compliant with regulations, it is evident that businesses today need to incorporate both technology and policy-based solutions to safeguard personal data. The results are in concordance with previous literature, although this study has some variations in the effectiveness of security measures based on sample size, industry focus, and geographical location. Collectively, the findings of this study highlight the importance of substantial security controls and governance arrangements for big data organizations. With the increasing volume and complexity of data, these results can signal to organizations how to maneuver the ever-changing landscape of data security.

Conclusion

The results of this study stress the significance of employing state-of-the-art security features (e.g., encryption, RBAC, and real-time anomaly detection) used to mitigate hacks in large-scale data handling settings. Healthcare and e-commerce industries are at maximum on the security front, so we need to protect these endpoints with a multi-layered approach. Furthermore, implementing data governance frameworks plays a crucial role in regulatory compliance. These are the things that every organization with data to protect must do to address potential areas of exposure and remain compliant with industry standards.

Acknowledgment

The authors wish to acknowledge with gratitude the support of the Department of Management Information Systems at Lamar University in carrying out this study. Thanks to the IT and cybersecurity professionals for giving up their time to discuss these tricky scenarios. Finally, we thank organizations who shared their security practices and data to allow us to study the current security state in handling large-scale data.

Funding: No funding sources

Conflict of interest: None declared

References

[1] Dai, H. N., Wong, R. C. W., Wang, H., Zheng, Z., & Vasilakos, A. V. (2019). Big data analytics for large-scale wireless networks: Challenges and opportunities. ACM Computing Surveys (CSUR), 52(5), 1-36. doi.org/10.1145/3337065

[2] Sun, Z., Strang, K. D., & Pambel, F. (2020). Privacy and security in the big data paradigm. Journal of computer information systems. doi.org/10.1080/08874417.2017.1418631

[3] Zeadally, S., Siddiqui, F., Baig, Z., & Ibrahim, A. (2020). Smart healthcare: Challenges and potential solutions using internet of things (IoT) and big data analytics. PSU research review, 4(2), 149-168. doi.org/10.1108/PRR-08-2019-0027

[4] Amalina, F., Hashem, I. A. T., Azizul, Z. H., Fong, A. T., Firdaus, A., Imran, M., & Anuar, N. B. (2019). Blending big data analytics: Review on challenges and a recent study. Ieee Access, 8, 3629-3645. DOI: 10.1109/ACCESS.2019.2923270

[5] Benjelloun, F. Z., & Lahcen, A. A. (2019). Big data security: challenges, recommendations and solutions. In Web Services: Concepts, Methodologies, Tools, and Applications (pp. 25-38). IGI Global. DOI: 10.4018/978-1-5225-7501-6.ch003

[6] Ejaz, W., & Anpalagan, A. (2019). Internet of things for smart cities: technologies, big data and security (pp. 1-15). Berlin/Heidelberg, Germany: Springer International Publishing. doi.org/10.1007/978-3-319-95037-2

[7] Yang, P., Xiong, N., & Ren, J. (2020). Data security and privacy protection for cloud storage: A survey. Ieee Access, 8, 131723-131740. DOI: 10.1109/ACCESS.2020.3009876

[8] Tawalbeh, L. A., Muheidat, F., Tawalbeh, M., & Quwaider, M. (2020). IoT Privacy and security: Challenges and solutions. Applied Sciences, 10(12), 4102. doi.org/10.3390/app10124102

[9] Zhu, Y., Zhang, Y., Wang, J., Song, W., Chu, C. C., & Liu, G. (2019, July). From data-driven to intelligent-driven: technology evolution of network security in big data era. In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2,

pp. 103-109). IEEE. DOI: 10.1109/COMPSAC.2019.10191

[10] Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature medicine, 25(1), 37-43. doi.org/10.1038/s41591-018-0272-7

[11] Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The perils and promises of big data research in information systems. Journal of the Association for Information Systems, 21(2), 9. DOI: 10.17705/1jais.00601

[12] Karim, A., Siddiqa, A., Safdar, Z., Razzaq, M., Gillani, S. A., Tahir, H., ... & Imran, M. (2020). Big data management in participatory sensing: Issues, trends and future directions. Future Generation Computer Systems, 107, 942-955. doi.org/10.1016/j.future.2017.10.007

[13] Volk, M., Staegemann, D., & Turowski, K. (2022). Providing Clarity on Big Data: Discussing Its Definition and the Most Relevant Data Characteristics. In KDIR (pp. 141-148). DOI: 10.5220/0011537500003335

[14] Zhang, J., & Lin, M. (2023). A comprehensive bibliometric analysis of Apache Hadoop from 2008 to 2020. International Journal of Intelligent Computing and Cybernetics, 16(1), 99-120. doi.org/10.1108/IJICC-01-2022-0004

[15] Hossain, E., Khan, I., Un-Noor, F., Sikander, S. S., & Sunny, M. S. H. (2019). Application of big data and machine learning in smart grid, and associated security concerns: A review. Ieee Access, 7, 13960-13988. Digital Object Identifier 10.1109/ACCESS.2019.2894819

[16] Tiwari, P., Ilavarasan, P. V., & Punia, S. (2021). Content analysis of literature on big data in smart cities. Benchmarking: An International Journal, 28(5), 1837-1857. doi/10.1108/BIJ-12-2018-0442

[17] Wimmer, M. A., Neuroni, A. C., & Frecè, J. T. (2020). Approaches to good data governance in support of public sector transformation through once-only. In Electronic Government: 19th IFIP WG 8.5 International Conference, EGOV 2020, Linköping, Sweden, August 31–September 2, 2020, Proceedings 19 (pp. 210-222). Springer International Publishing. doi.org/10.1007/978-3-030-57599-1_16

[18] Syu, J. H., Lin, J. C. W., Srivastava, G., & Yu, K. (2023). A comprehensive survey on artificial intelligence empowered edge computing on consumer electronics. IEEE Transactions on Consumer Electronics. DOI: 10.1109/TCE.2023.3318150

[19] Samaraweera, G. D., & Chang, J. M. (2019). Security and privacy implications on database systems in big data era: A survey. IEEE Transactions on Knowledge and Data Engineering, 33(1), 239-258. DOI: 10.1109/TKDE.2019.2929794

[20] Eibeck, A., Shaocong, Z., Mei Qi, L., & Kraft, M. (2024). Research data supporting" A Simple and Efficient Approach to Unsupervised Instance Matching and its Application to Linked Data of Power Plants". Eibeck, A., Shaocong, Z., Mei Qi, L., & Kraft, M. (2024). Research data supporting" A Simple and Efficient Approach to Unsupervised Instance Matching and its Application to Linked Data of Power Plants". doi.org/10.17863/CAM.82548

[21] Mowla, M. N., Mowla, N., Shah, A. S., Rabie, K., & Shongwe, T. (2023). Internet of things and wireless sensor networks for smart agriculture applications-a survey. IEEE Access. DOI: 10.1109/ACCESS.2023.3346299

[22] Das, P., Begum, S. A., & Buyya, R. Advanced Computing, Machine Learning, Robotics and Internet Technologies. doi.org/10.1007/978-3-031-47221-3

[23] Figueroa-Lorenzo, S., Añorga, J., & Arrizabalaga, S. (2019). A role-based access control model in modbus SCADA systems. A centralized model approach. Sensors, 19(20), 4455. doi.org/10.3390/s19204455

[24] Gupta, M., & Shah, U. N. (2023). Navigating the Data Security Landscape: Challenges and Solutions in Financial Markets amid Digitalization and Artificial Intelligence. International Journal of Multidisciplinary Research and Analysis, 10. DOI: 10.47191/ijmra/v6-i12-77

[25] McConomy, B. C., & Leber, D. E. (2022). Cybersecurity in healthcare. In Clinical Informatics Study Guide: Text and Review (pp. 241-253). Cham: Springer International Publishing. doi.org/10.1007/978-3-030-93765-2_17

[26] Khatri, S., Vachhani, H., Shah, S., Bhatia, J., Chaturvedi, M., Tanwar, S., & Kumar, N. (2021). Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. Peer-to-Peer Networking and Applications, 14, 1778-1805. doi.org/10.1007/s12083-020-00993-4

[27] Liu, L., Li, J., Lv, J., Wang, J., Zhao, S., & Lu, Q. (2024). Privacy-Preserving and Secure Industrial Big Data Analytics: A Survey and the Research Framework. IEEE Internet of Things Journal. DOI: 10.1109/JIOT.2024.3353727

[28] Radanliev, P. (2024). Digital security by design. Security Journal, 1-40. doi.org/10.1057/s41284-024-00435-3

[29] Gan, Q., Liu, J. K., Wang, X., Yuan, X., Sun, S. F., Huang, D., ... & Wang, J. (2022). Verifiable searchable symmetric encryption for conjunctive keyword queries in cloud storage. Frontiers of Computer Science, 16(6), 166820. doi.org/10.1007/s11704-021-0601-8

[30] Eibeck, A., Zhang, S., Lim, M. Q., & Kraft, M. (2024). A simple and efficient approach to unsupervised instance matching and its application to linked data of power plants. Journal of Web Semantics, 80, 100815. doi.org/10.1016/j.websem.2024.100815

[31] Krishnamoorthy, G., & Sistla, S. M. K. (2023). Exploring Machine Learning Intrusion Detection: Addressing Security and Privacy Challenges in IoT-A Comprehensive Review. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 114-125. DOI:10.30574/wjarr.2024.21.2.0501

[32] Belmabrouk, K. (2023). Cyber criminals and data privacy measures. In Contemporary Challenges for Cyber Security and Data Privacy (pp. 198-226). IGI Global. DOI: 10.4018/979-8-3693-1528-6.ch011

[33] Wu, Y., Dai, H. N., & Wang, H. (2020). Convergence of blockchain and edge computing for secure and scalable IIoT critical infrastructures in industry 4.0. IEEE Internet of Things Journal, 8(4), 2300-2317. DOI: 10.1109/JIOT.2020.3025916

[34] Uddin, M., Islam, S., & Al-Nemrat, A. (2019). A dynamic access control model using authorising workflow and task-role-based access control. Ieee Access, 7, 166676-166689. DOI: 10.1109/ACCESS.2019.2947377

[35] Brous, P., & Hiel, M. (2024, August). Principles for the Secure Exchange of Sensitive Data Across Classified Networks: A Data-Centric Approach. In International Conference on Electronic Government (pp. 18-31). Cham: Springer Nature Switzerland. doi.org/10.1007/978-3-031-70274-7_2

[36] Qadri, Y. A., Nauman, A., Zikria, Y. B., Vasilakos, A. V., & Kim, S. W. (2020). The future of healthcare internet of things: a survey of emerging technologies. IEEE Communications Surveys & Tutorials, 22(2), 1121-1167. DOI: 10.1109/COMST.2020.2973314

[37] Shah, I. A. (2022). Cybersecurity Issues and Challenges for E-Government During COVID-19: A Review. Cybersecurity Measures for E-Government Frameworks, 187-222. DOI: 10.4018/978-1-7998-9624-1.ch012

[38] Hussain, M. D., Rahman, M. H., & Ali, N. M. (2024). Investigation of Gauss-Seidel Method for Load Flow Analysis in Smart Grids. Sch J Eng Tech, 5, 169-178.

[39] Hossain, Q., Yasmin, F., Biswas, T. R., & Asha, N. B. (2024). Data-Driven Business Strategies: A Comparative Analysis of Data Science Techniques in Decision-Making. Sch J Econ Bus Manag, 9, 257-263.

[40] Hossain, Q., Yasmin, F., Biswas, T. R., & Asha, N. B. (2024). Integration of Big Data Analytics in Management Information Systems for Business Intelligence. Saudi J Bus Manag Stud, 9(9), 192-203.

[41] Bhardwaj, I., Biswas, T. R., Arshad, M. W., Upadhyay, A., & More, A. B. (2024). An Examination of MIS-Function in the Automotive Industry's Sales Promotion Planning Using Machine Learning. Library Progress International, 44(3), 3164-3170.

[42] O'Brien, N., Ghafur, S., & Durkin, M. (2021). Cybersecurity in health is an urgent patient safety concern: we can learn from existing patient safety improvement strategies to address it. Journal of Patient Safety and Risk Management, 26(1), 5-10. doi.org/10.1177/2516043520975926