iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71583

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Analysis of Big Data Security Threats

Vandana Malik
Baba Mastnath University

Abstract: ***The exponential increase in data generation, propelled by the Internet of Things (IoT), social media, mobile technology, and cloud computing, has ushered in the era of big data. While big data enables unprecedented insights and innovations, it also exposes organizations to sophisticated security threats. This paper provides an in-depth analysis of big data security threats, examining the sources, vectors, and impacts of these threats. Furthermore, the paper reviews current mitigation strategies, highlights gaps in existing approaches, and suggests future research directions. Key topics include data lifecycle security, cloud vulnerabilities, distributed architecture risks, and the importance of governance and compliance frameworks.***

I. INTRODUCTION

Big data systems are increasingly integral to sectors such as healthcare, finance, retail, and government. These systems process petabytes of information from heterogeneous sources in real-time, enabling predictive analytics and automation. However, the scale, complexity, and speed of big data also introduce new vulnerabilities. Conventional cybersecurity models, often designed for static and smaller-scale systems, are ill-equipped to manage the dynamic nature of big data environments. Understanding these evolving threats is essential for developing adaptive and resilient security frameworks.

II. CHARACTERISTICS OF BIG DATA AND THEIR IMPACT ON SECURITY

The unique attributes of big data—commonly referred to as the 5 V's—have direct implications on security models:

1) Volume: Larger datasets present more attack vectors and storage risks.
2) Velocity: High-speed data processing requires real-time security measures.
3) Variety: Diverse data types (structured, semi-structured, unstructured) make standard security policies less effective. Veracity: The trustworthiness of data impacts the integrity of analytics.
4) Value: Sensitive and high-value data is a lucrative target for attackers.

Additional characteristics such as variability and visualization further complicate security monitoring and access control.

III. SECURITY THREATS IN BIG DATA ENVIRONMENTS

A. Data Breaches and Unauthorized Access

Big data platforms often store sensitive personal and financial information. Breaches can occur due to misconfigured cloud storage (e.g., open S3 buckets), unpatched systems, or compromised credentials. The consequences include identity theft, financial loss, reputational damage, and legal penalties.

B. Insider Threats

Insider threats are particularly difficult to detect, as they originate from trusted users. These may include deliberate sabotage, data theft, or accidental leakage. According to Verizon’s Data Breach Investigations Report (2023), over 20% of data breaches involve internal actors.

C. Distributed Denial of Service (DDoS) Attacks

Big data infrastructures, especially cloud-based ones, are susceptible to DDoS attacks that overwhelm systems with traffic, causing service disruption and data unavailability.

D. Cloud-Specific Threats

As big data is frequently deployed on public or hybrid cloud environments, it inherits cloud vulnerabilities such as:

- Shared technology exploits (e.g., hypervisor flaws)
- Insecure APIs and endpoints
- Data loss due to provider negligence or external attack

©IJRASET: All Rights are Reserved

E. Lack of Data Governance

The absence of structured policies for data ownership, classification, retention, and disposal increases the risk of regulatory violations and data misuse.

IV. BIG DATA SECURITY ACROSS THE DATA LIFECYCLE

Security should be integrated at every stage of the data lifecycle:

1) Data Generation: Ensure trusted data sources and validate data authenticity.
2) Data Acquisition: Implement secure ingestion protocols and encrypt transmission.
3) Data Storage: Apply encryption at rest, access control lists (ACLs), and regular audits.
4) Data Processing: Enforce sandboxing, secure APIs, and code review.
5) Data Analytics: Prevent data leakage via masking and secure computation.
6) Data Archival and Disposal: Use secure deletion standards and compliance checks.

V. SECURITY TECHNOLOGIES AND FRAMEWORKS

A. Encryption and Tokenization

Advanced cryptographic techniques such as homomorphic encryption, format-preserving encryption, and tokenization ensure confidentiality without sacrificing usability. Tokenization is particularly useful in compliance-heavy industries like finance and healthcare.

B. Privacy-Preserving Computation

Techniques such as secure multiparty computation (SMC) and differential privacy allow data analysis without revealing individual data points, critical for GDPR compliance (Dwork & Roth, 2014).

C. Identity and Access Management (IAM)

IAM solutions integrate single sign-on (SSO), multi-factor authentication (MFA), and biometric verification to ensure user accountability. Integration with LDAP and Active Directory further enhances enterprise readiness.

D. Monitoring and Intrusion Detection

Behavior-based monitoring tools that leverage AI and machine learning can identify anomalies in real-time, aiding in early threat detection. Tools such as Apache Spot, Splunk, and ELK stack are commonly used for big data monitoring.

E. Blockchain-Based Security

Blockchain can provide decentralized access control, immutable audit trails, and transparent user authentication in big data systems. Pilot implementations in healthcare and finance show promise (Zyskind et al., 2015).

VI. REGULATORY COMPLIANCE AND LEGAL CHALLENGES

Organizations must align big data practices with various national and international regulations:

1) GDPR (EU): Emphasizes data minimization, user consent, and right to erasure.
2) HIPAA (US): Requires stringent protection for health data.
3) CCPA (California): Grants consumers rights over their personal information.

Non-compliance can lead to hefty fines, legal proceedings, and reputational damage. Implementing **privacy-by-design** is crucial to meeting regulatory demands.

VII. CHALLENGES IN SECURING BIG DATA

Despite advances, several challenges remain:

1) Scalability: Security solutions must scale horizontally across petabyte-level data.
2) Complexity: Multiple technologies and tools make uniform security policy enforcement difficult.
3) Lack of Skilled Professionals: There is a shortage of cybersecurity experts with big data proficiency.
4) Latency: Real-time analytics require lightweight security solutions that do not introduce significant delays.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VIII. FUTURE RESEARCH DIRECTIONS

1) Post-Quantum Cryptography: Preparing encryption systems for quantum computing capabilities.
2) Federated Learning: Allowing collaborative machine learning without sharing raw data.
3) Zero-Trust Architecture: Ensuring that no device or user is automatically trusted.
4) Explainable AI in Security: Providing interpretability to AI-based threat detection systems.

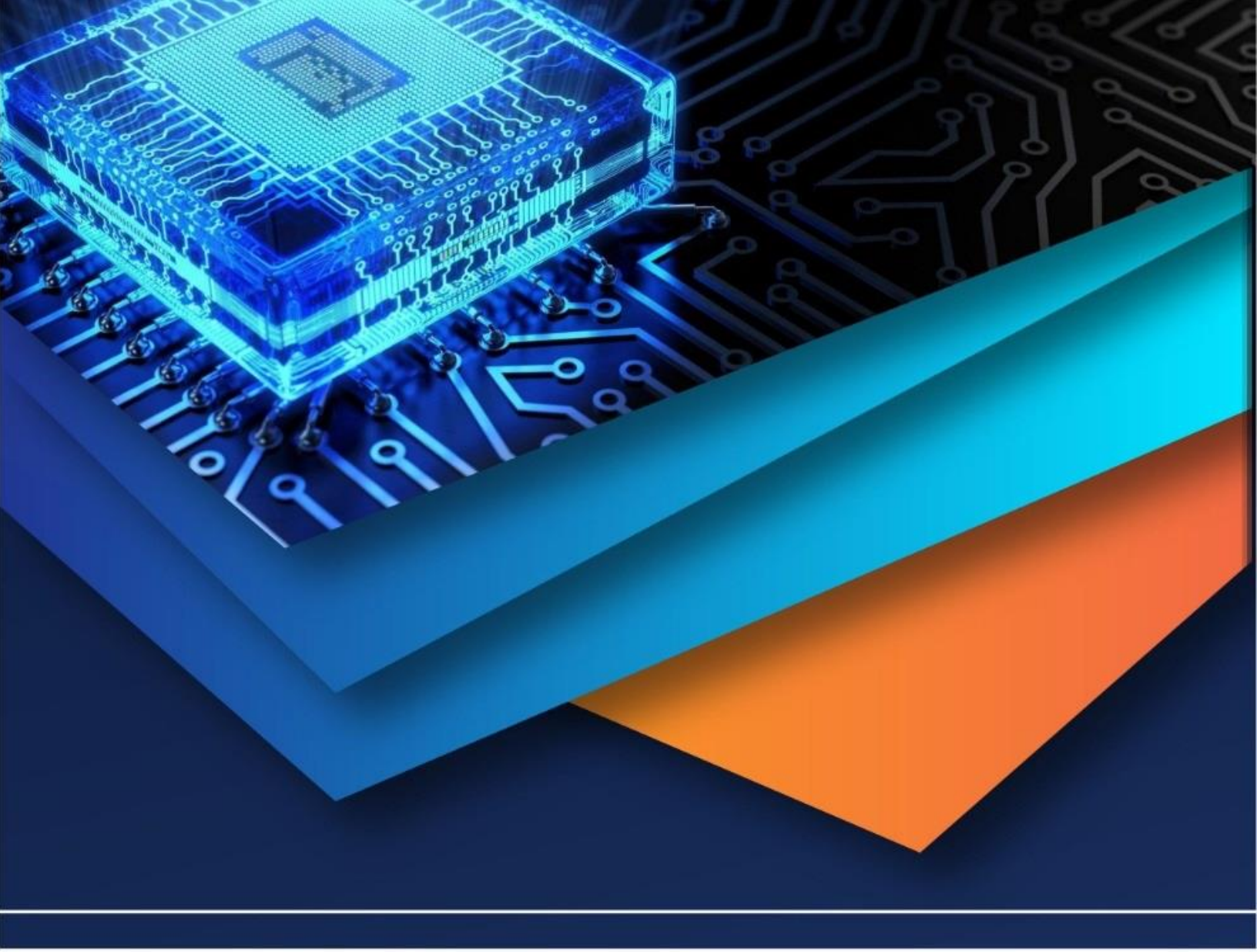
IX. CONCLUSION

Securing big data systems is no longer optional; it is a business imperative. As the data landscape evolves, so too must our security paradigms. A multi-layered, proactive approach that integrates technical defenses, user education, and compliance strategies is essential. Research and innovation must continue to anticipate and counter new threat vectors in an ever-changing digital ecosystem.

REFERENCES

[1] Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042
[2] Feng, Y., Wang, X., & Li, J. (2020). Real-time anomaly detection in big data streams using deep learning. IEEE Access, 8, 136296–136305. https://doi.org/10.1109/ACCESS.2020.3011612
[3] Gahi, Y., Guennoun, M., & El-Khatib, K. (2016). Big data security and privacy: A review. Procedia Computer Science, 83, 644–649. https://doi.org/10.1016/j.procs.2016.04.137
[4] Krebs, B. (2017). Equifax breach exposed data of 147 million people. Krebs on Security. https://krebsonsecurity.com/2017/09/equifax-breach-exposed-data-of-143-million-americans/
[5] Ponemon Institute. (2023). 2023 Cost of Insider Threats: Global Report. https://www.ponemon.org
[6] Verizon. (2023). Data Breach Investigations Report. https://www.verizon.com/business/resources/reports/dbir/
[7] Zhou, Y., Zhang, R., Xie, H., & Liu, Q. (2017). Privacy-preserving data mining on big data. Information Sciences, 379, 19–31. https://doi.org/10.1016/j.ins.2016.07.036
[8] Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. 2015 IEEE Security and Privacy Workshops (SPW), 180–184. https://doi.org/10.1109/SPW.2015.27

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)