

Next Generation Cyber Defense with Big Data Insights

¹ Navya vd S M, ² Mss. Pooja kn

¹ Student, Department of MCA, BIET, Davangere, India

² Assistant Professor, Department of MCA, BIET, Davangere, India

ABSTRACT

The rapid growth of digital technologies has significantly increased cyber threats across networks and organizations. Traditional security systems are unable to detect advanced and large-scale attacks effectively. Big Data analytics provides powerful tools to process massive volumes of security logs in real time. This paper proposes a next-generation cyber defense model using big data insights for proactive threat detection. Machine learning algorithms are integrated to improve accuracy and reduce false positives. The proposed system enhances security intelligence, scalability, and response efficiency.

Keywords: *bigdataanalytics, cyber security, machinelearning, threat intelligence*

I. INTRODUCTION

With the expansion of cloud computing, IoT, and online services, cyber-attacks are becoming more sophisticated. Organizations generate huge amounts of security data daily from firewalls, servers, and applications. Traditional rule-based systems cannot efficiently analyze such large datasets. Big Data technologies like Hadoop and Spark enable distributed processing of security data. Integrating machine learning with big data helps in identifying hidden attack patterns. Therefore, next-generation cyber defense systems are essential for modern digital environments.

II. LITERATURE REVIEW

Previous research highlights the use of signature-based intrusion detection systems for threat monitoring. Researchers have applied machine learning models such as Decision Trees and Neural Networks for anomaly detection. Studies also show the effectiveness of Hadoop-based distributed frameworks in processing large-scale security logs. Recent works focus on real-time threat intelligence using streaming analytics. However, integration of scalable big data platforms with adaptive AI models is still evolving.

III. EXISTING SYSTEM

The existing cyber defense systems mainly rely on firewalls, antivirus software, and traditional IDS. These systems are rule-based and depend on known attack signatures. They cannot effectively detect zero-day attacks or unknown threats. Limited scalability restricts their performance with growing data volumes. Manual monitoring increases response time. Hence, traditional systems lack real-time intelligence and predictive capabilities.

Limitations:

1. High computational cost and infrastructure requirements.
2. Data privacy and security concerns during large-scale processing.
3. False positives in anomaly detection models.

III. PROPOSED SYSTEM

The proposed system uses Big Data frameworks like Hadoop and Spark for distributed data processing. Security logs are collected from multiple sources and stored in a centralized big data environment. Machine learning algorithms such as Random Forest and Deep Learning models are applied for anomaly detection. Real-time analytics enables quick identification of suspicious activities. The system provides automated alerts and threat prediction for proactive defense.

Advantages

1. REAL-TIME THREAT DETECTION
2. Improved accuracy with machine learning.
3. Scalability and high performance
4. Proactive threat intelligence.
5. Faster incident response.
6. Cost efficiency in the long run.
7. Enhanced data visibility.

System Architecture

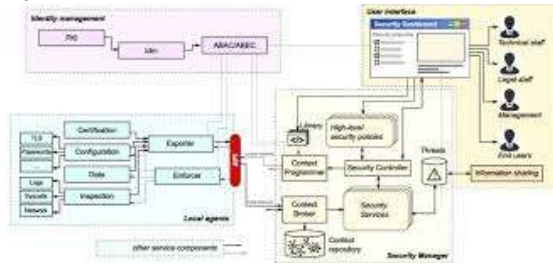


Fig 1. System Architecture

V. METHODOLOGY

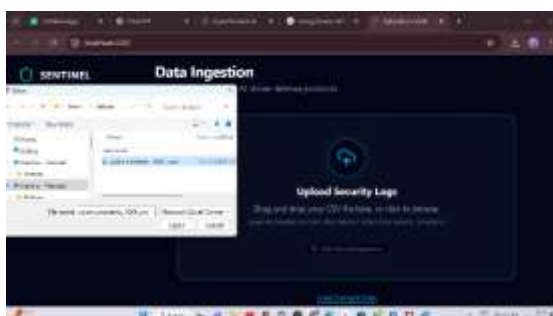
MODULES DESCRIPTION:

The initiation and implementation of the next generation cyber defense with big data insights will follow the agile development model. The scrum framework will be used predominantly in the project. The scrum model will work efficiently as requirements may vary, and such projects will have demanding requirements.

Agile execution framework: Iterative functional increments. The deployment of the next-generation cyber defense platform is governed by an agile evolutionary model. The project lifecycle is decomposed into time-boxed development cycles (sprints) typically spanning fourteen to twenty-eight days.

Dashboard and visualization design: Components of the user interface and user experience will be implemented throughout the project. User playback will play a crucial role in the design of the dashboards to ensure clarity and optimal presentation of the results.

VI. RESULT



Data Ingestion is the process of collecting, importing, and transferring data from external sources into a system for storage and analysis. In an AI-driven cybersecurity system, data ingestion is the first and most critical stage. It enables the system to gather security logs and prepare them for threat detection and analysis.

Data ingestion is the foundation of an AI-driven cybersecurity system. It ensures that security logs are collected, validated, processed, and transformed into a structured format suitable for intelligent threat detection.

Without effective data ingestion, advanced AI models cannot function properly in next-generation cyber defense systems.

VII. CONCLUSION

Next-generation cyber defense using big data insights provides an intelligent and scalable solution to modern cyber threats. By combining distributed computing and machine learning, the system enhances real-time threat detection. It overcomes the limitations of traditional rule-based security systems. The proposed framework ensures better accuracy, faster response, and improved security management. Future work can focus on integrating AI-based predictive analytics and blockchain for enhanced security.

REFERENCES

1. Whitfield Diffie and Martin Hellman, "New Directions in Cryptography," IEEE Transactions on Information Theory.
2. Matei Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing."
3. National Institute of Standards and Technology (NIST), Cybersecurity Framework.
4. IEEE, Research Papers on Big Data Security Analytics