

Big Data Based Predictive Cyber Threat Intelligence System

Chandu KL , Dr. Vinay Patel GL

P.G. Student, Department of Master of Computer Applications, BIET, Davangere-577004.

chandulkundur@gmail.com, Davangere, Karnataka, India

Associate Professor of Master of Computer Applications, BIET, Davangere-577004.

vinaypatelhlgl@gmail.com, Davangere, Karnataka, India

ABSTRACT

The escalating sophistication and volume of cyber threats necessitate a paradigm shift from traditional signature-based defense mechanisms to a more proactive intelligent and data-driven approach. This project “Big Data based predictive cyber threat intelligence system” suggests creating a sophisticated safety system that makes use of big data. Analytics to anticipate detect and mitigate cyber-attacks in real-time. The core of this system to ingest and process vast and varied streams of data traffic system logs endpoint security solutions and external threat intelligence feeds. Deep learning models and behavioral analytics to this massive dataset the proposed system threats that would otherwise go unnoticed by conventional security tools. This data-centric approach enables the system to move beyond reactive responses. System is designed to provide security analysts with actionable intelligence through intuitive visualizations and comprehensive dashboards thereby reducing the mean time to detect (MTTD) and mean time to respond (MTTR) to security incidents. The project aims to demonstrate the feasibility and effectiveness of a big data-driven cybersecurity architecture in enhancing an organization’s resilience against the dynamic and evolving landscape of cyber threats ultimately fostering a more secure digital environment. The projected basis will not only offer superior threat detection capabilities but also provide a scalable and adaptable platform that can evolve with the ever-changing threat landscape ensuring long-term protection for critical digital assets.

KEYWORDS:

Big Data-based Predictive Cyber Threat Intelligence involves the use of advanced technologies to analyze large volumes of structured and unstructured data for

identifying and preventing cyber threats. It combines big data analytics, machine learning, and artificial intelligence to detect patterns, anomalies, and potential risks in real time. Keywords associated with this domain include cybersecurity, predictive analytics, threat detection, anomaly detection, data mining, and intrusion detection systems

INTRODUCTION

The “**Big Data–Based Predictive Cyber Threat Intelligence System**” project is a comprehensive initiative aimed at creating a robust and intelligent security framework capable of defending against the next generation of cyber threats. This particular project overcomes the constraints of the conventional security solution through the use of the transformative power of big data analytics and artificial intelligence.

“The underlying assumption in this project is that there is valuable threat intelligence embedded in an organization’s IT systems.” The system can then be described as the ‘central nervous system for security operations’ because it gives a ‘holistic view’ of the ‘cyber threat landscape’.

Essentially the project entails learning the processing and storage been gathered from various sources. Such datasets from servers and networks with high speeds and volumes and variety. However the project requires the application of effective big data tools and technologies that include the use of the Hadoop and the Spark technologies.

Once the data has been gathered the research focuses more on sophisticated analysis techniques. This involves the design ability to detect complicated malware signatures potential attack surfaces and anomalies within user interactions and system-level events.

A key part of this project involves the design and development of the intelligent threat correlation engine. In other words the purpose of this engine is to derive relationships among seemingly unrelated events

and datasets in order to identify complex multi-stage cyber attacks. Knowledge of the context & interdependencies between different security incidents offers greater insight into active threats and also suggests suitable mitigation strategies. This allows for a deeper and more precise calculation of the general security posture of an organization.

Moreover the given project emphasizes data visualization and reporting. An intuitive modern dashboard will be created that visually presents complex security information in a simple and understandable way. This will make it easy for security analysts to rapidly assess the situation recognize patterns and make informed decisions in real-time

I LITERATURE REVIEW

With the rapid growth of digital systems, cyber threats have become more complex and frequent. Traditional security approaches are mostly reactive, meaning they respond only after an attack occurs. To overcome this limitation, researchers have introduced predictive cyber threat intelligence (CTI) using big data analytics, enabling proactive detection and prevention of cyber threats.

Big data plays a key role because cybersecurity systems generate massive volumes of data from sources such as network logs, IoT devices, and user activities. These datasets are analyzed using advanced techniques like machine learning and artificial intelligence to predict future threats.

Big Data-based Predictive Cyber Threat Intelligence focuses on using large volumes of data to identify and prevent cyber threats before they occur. Traditional cybersecurity methods are reactive, but predictive approaches use historical and real-time data for proactive defense. Big data technologies help process complex and high-speed data from networks, systems, and user activities. Machine learning and artificial intelligence play a key role in analyzing patterns and detecting anomalies. These systems can predict potential cyber-attacks and improve response time. Predictive CTI is widely used in sectors like banking, healthcare, and cloud security. It enhances threat detection accuracy and reduces risks. However, challenges such as data privacy, scalability, and high computational cost still exist. Researchers are working on advanced models and secure frameworks to overcome these issues. Overall, predictive CTI with big data is transforming modern cybersecurity into a smarter and more efficient system.

EXISTING SYSTEM

The existing system in cyber threat intelligence is mainly based on traditional and reactive security approaches. These systems rely on signature-based detection methods to identify known threats. Firewalls, antivirus software, and intrusion detection systems are widely used as primary security tools. However, they can only detect attacks that have already been identified and stored in databases. Zero-day attacks and unknown threats often bypass these systems. Existing systems generate large volumes of security data, but lack efficient mechanisms to analyze it in real time. Manual analysis of logs and alerts is time-consuming and error-prone. Most systems do not utilize big data technologies for handling high-volume and high-velocity data. They also lack integration with advanced machine learning or artificial intelligence techniques. As a result, prediction of future threats is not possible in traditional systems. Security teams often respond only after damage has occurred. These systems have limited scalability when dealing with modern cyber environments like cloud and IoT. Data from different sources is often stored in silos, reducing overall visibility. Existing approaches also produce a high number of false positives. This leads to alert fatigue among security analysts. Real-time threat intelligence sharing is minimal or inefficient. Many systems lack automation in threat detection and response. The existing system is not capable of handling complex and evolving cyber threats. It mainly focuses on prevention rather than prediction. Overall, the current systems are insufficient for proactive cybersecurity needs in today's digital world.

The existing system in cyber threat intelligence mainly relies on traditional and reactive security methods. It uses tools like firewalls, antivirus software, and intrusion detection systems. These systems depend on signature-based detection to identify known threats. However, they fail to detect new or unknown attacks such as zero-day vulnerabilities. Large amounts of security data are generated, but real-time analysis is limited. Most systems require manual monitoring, which is time-consuming and error-prone.



Figure: Above figure shows automated threat analysis

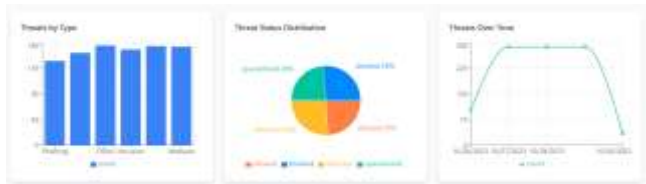


Figure: Above figure shows visualization and distribution

IV. Conclusion

The proposed Big Data-based Predictive Cyber Threat Intelligence system offers a modern and efficient solution to handle increasing cyber threats. It shifts cybersecurity from a reactive approach to a proactive and predictive model. By leveraging big data analytics, the system can process massive volumes of structured and unstructured data. It integrates machine learning and artificial intelligence to identify patterns and detect anomalies effectively. The system improves the accuracy of threat detection compared to traditional methods. It significantly reduces false positives, helping security analysts focus on real threats. Real-time data processing enables faster identification and response to cyber-attacks. The system is highly scalable and can handle data from multiple sources such as networks, cloud, and IoT devices. It enhances visibility by integrating data across different platforms. Automated detection and response mechanisms reduce human effort and errors. The system also supports better risk assessment and decision-making. Predictive capabilities allow organizations to prevent attacks before they occur. It improves overall system reliability and performance. The use of advanced frameworks like Hadoop and Spark increases processing efficiency. Despite its advantages, challenges such as data privacy and computational cost still exist. Continuous research is required to improve model accuracy and efficiency. Future enhancements may include blockchain and advanced deep learning techniques. The system can be applied across various sectors like banking, healthcare, and government. It strengthens the overall cybersecurity infrastructure of organizations. In conclusion, the proposed system

provides a powerful and intelligent approach to secure digital environments against evolving cyber threats.

IV REFERENCES

- [1] S. Mittal, J. Joshi, and S. Finin, "Cyber Threat Intelligence," *IEEE Computer*, vol. 52, no. 8, pp. 16–19, 2019.
- [2] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.
- [3] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A Deep Learning Approach for Network Intrusion Detection System," in *Proc. IEEE ICST*, 2016.
- [4] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in *IEEE Symposium on Security and Privacy*, 2010.
- [5] J. Manyika et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," *McKinsey Global Institute*, 2011.
- [6] T. Chen, Y. Zhang, and Q. Chen, "Big Data Based Cyber Security Intelligence: A Survey," *IEEE Access*, vol. 7, pp. 451–470, 2019.
- [7] G. Creech and J. Hu, "A Semantic Approach to Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns," *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 807–819, 2014.
- [8] W. Lee and S. Stolfo, "Data Mining Approaches for Intrusion Detection," in *Proc. USENIX Security Symposium*, 1998.