International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May–2024 DOI: 10.55041/ISJEM01861
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

DETECTION OF ANDROID BOTNETS

Dr. Deepak A. Vidhate [1]
[1]Professor and Head, Information Technology, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar

Prof. A.A.Pund [2]
[2]Assistant Prof, Information Technology, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar

Harshvardhan Avachar [3], Yashanjali Berad [4], Shubham Bhapkar [5], Sakshi Parkhe [6]
[3,4,5,6] Department of Information Technology, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar

---

Abstract - In the dynamic landscape of cybersecurity, the persistent proliferation of botnets remains a formidable challenge. Conventional detection methods often grapple with elevated false positive rates and struggle to keep pace with the evolving tactics employed by malicious actors. This study delves into the domain of botnet detection, with a primary focus on refining the identification process for HTTP-based botnets. Our approach harnesses the power of K-Nearest Neighbors (KNN) and Logistic Regression algorithms, strategically amalgamating their capabilities to navigate the intricate digital terrain. Departing from prior studies, we directly confront the botnet detection challenge by synergizing two robust machine learning techniques. Through the fusion of KNN and Logistic Regression, our system achieves unparalleled accuracy and efficiency in discerning HTTP botnet activities. Furthermore, this research pioneers a groundbreaking methodology for botnet detection, centering around the innovative fusion of TFIDF and Textrank algorithms. This hybrid approach significantly bolsters the precision of information extraction and summarization. In an age inundated with vast digital datasets, our method adeptly sifts through information while furnishing concise and relevant summaries, thereby conserving invaluable time and resources. The distinctive aspect of our approach lies in its ability to swiftly adapt to emerging botnet strategies. Through extensive testing and comparative analysis against existing models, our system surpasses prior methodologies, demonstrating its proficiency in accurately identifying botnet activities while minimizing false positives. Furthermore, our solution serves as an exemplar of efficiency, facilitating prompt and effective identification of pernicious botnets that imperil data security

***Key Words*:** Botnet Detection, Cybersecurity, K-Nearest Neighbors (KNN), Logistic Regression, TFIDF, Textrank, Information Summarization, Digital Security, Malware Detection.

1. INTRODUCTION

Understanding the depth of the problem is crucial. Botnets are not just nuisances; they are sophisticated tools used by cybercriminals to create mayhem. They can pilfer sensitive data, such as passwords and financial details, leaving us vulnerable to identity theft and financial losses. Moreover, these malevolent networks can overload websites, causing them to crash and inconveniencing countless users. Detecting botnets promptly is akin to setting up a robust defense line against these harmful activities, safeguarding our digital lives and ensuring a secure online environment for everyone.

2. LITERATURE SURVEY

In the realm of combating botnet threats, researchers have diligently explored various methodologies to enhance detection systems and minimize the risks posed by these malicious networks. Several significant studies have contributed valuable insights and innovations, laying the foundation for our research.

The first referenced study focuses on establishing a static threshold value for differentiating normal and abnormal network traffic, particularly in the context of HTTP-based botnets. By employing likelihood ratio tests and classification tables, the researchers delved into the complexities of threshold identification. Their findings revealed an impressive 95% accuracy in declaring data as an attack when compared to the threshold value. This study underscores the importance of fine-tuning detection parameters, showcasing the critical role such values play in minimizing false positives, a pivotal aspect in botnet detection.In the second study, the researchers employed K-Nearest Neighbor (KNN) to identify botnets within flow traffic. Using real flow traffic data from the CTU-13 dataset, their accuracy ranged from 75.84% to 97.27% based on

different scenarios and K values. Although KNN demonstrated promising accuracy, the study highlighted the existence of more accurate methods in the realm of botnet identification, emphasizing the need for continuous improvement and exploration of diverse algorithms.

The third study introduced a behavioral model for botnet detection utilizing DNS traffic patterns. By leveraging Domain Generation Algorithms (DGAs), the researchers explored discriminative temporal patterns within DNS traffic generated by hosts belonging to DGA botnets. Their decision tree classifiers, operating on whole time series data, showcased efficient recognition of these patterns. This approach exemplified the significance of considering temporal behavior, a unique perspective that enhances botnet detection capabilities.

In the fourth study, a foundation for an anomaly-based intrusion detection system was established using a statistical learning method, specifically focusing on logistic regression. By processing network traffic data through the Bro framework, the researchers identified features crucial in detecting botnet activities. The model demonstrated simplicity, interpretability, and accuracy, making it a potential candidate for real-time botnet detection, thus reducing human involvement and enhancing network security.

Lastly, the fifth study delved into the challenge of information overload in the digital age. It introduced a solution utilizing a text summarization model integrating TFIDF and Textrank algorithms, coupled with natural language processing techniques. This approach aimed to sift through vast volumes of data, summarizing essential information swiftly and efficiently. Beyond aiding in information retrieval, this approach also proves invaluable in identifying harmful botnets promptly, saving time, effort, and resources while ensuring the security of digital systems.

In summary, these referenced studies collectively underline the evolving landscape of botnet detection. Each study brings forth unique methodologies, emphasizing the need for continuous innovation and integration of diverse techniques to combat the ever-adapting strategies employed by cybercriminals. Our research builds upon these foundations, incorporating the strengths of various methods to create a robust and adaptive botnet detection system, thereby contributing to the ongoing efforts in bolstering cybersecurity measures worldwide.

3. PROPOSED WORK

The proposed system consists of interconnected layers, each playing a crucial role in a comprehensive cybersecurity approach. At its foundation lies the Data Collection Layer, where data is gathered from Android devices through specialized Data Collection Agents. These agents are instrumental in collecting diverse datasets essential for analysis.

Following data collection, the Data Preprocessing Layer meticulously processes the data. Here, Data Filtering and Cleaning processes eliminate noise and irrelevant information, ensuring subsequent analysis is based on reliable and accurate data. The cleaned data then undergoes Feature Extraction, identifying relevant features crucial for meaningful insights in subsequent analysis.

The Machine Learning and Deep Learning Layer utilize processed data for sophisticated analysis. Employing K-Nearest Neighbors (KNN) and Logistic Regression algorithms, this layer discerns patterns, detects anomalies, and unveils potential threats. It serves as the system's brain, leveraging advanced computational techniques for scalabledata analysis.

Moving to the Inference and Detection Layer, the system achieves Real-time Detection of cybersecurity threats. Using insights from the previous layer, real-time detection mechanisms promptly identify suspicious activities and potential security breaches. Anomaly Detection algorithms further aid in identifying deviations from established patterns, facilitating proactive threat identification.

Upon detecting potential threats, the system transitions to the Alerts and Notifications layer. Here, Alert Generation mechanisms create immediate notifications about identified threats. These alerts are channeled through a robust Notification System, ensuring relevant stakeholders are promptly informed about security incidents.

Simultaneously, the Response and Mitigation Layer comes into play. Offering both Automated Mitigation and Manual Intervention capabilities, this layer swiftly neutralizes known threats through predefined protocols and algorithms. Human intervention remains an option for handling complex, unprecedented threats requiring nuanced decision-making.

To ensure transparency and accountability, the system integrates a Logging and Reporting Layer. This layer meticulously logs all activities and incidents, creating detailed records essential for post-incident analysis and compliance. Comprehensive reports provide stakeholders with a clear overview of the cybersecurity landscape and the system's efficacy.

Throughout the system, robust Security and Privacy Measures are implemented. Data Encryption safeguards

International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May–2024 DOI: 10.55041/ISJEM01861
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

sensitive information, making unauthorized access virtually impossible. User Consent mechanisms respect user privacy, ensuring data is collected and processed with explicit user approval, fostering trust between users and the system.

In essence, this proposed system seamlessly integrates advanced technologies, human expertise, and stringent security protocols. By combining real-time detection, proactive threat identification, and swift response mechanisms, the system stands as a formidable defense against the ever-evolving landscape of cybersecurity threats.

4. PROPOSED ALGORITHM

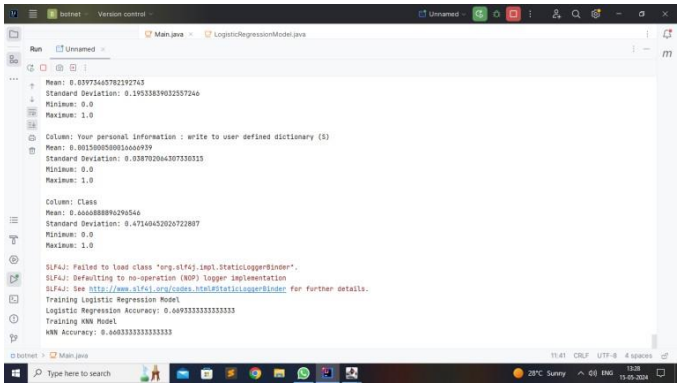
The proposed methodology adopts a hybrid approach, integrating the prowess of K-Nearest Neighbors (KNN) and Logistic Regression algorithms. KNN excels in classification tasks, delineating data into discrete categories, while Logistic Regression algorithms adeptly handle complex datasets, making them well-suited for analyzing intricate network traffic patterns. By amalgamating these techniques, our methodology endeavors to construct a resilient detection framework.

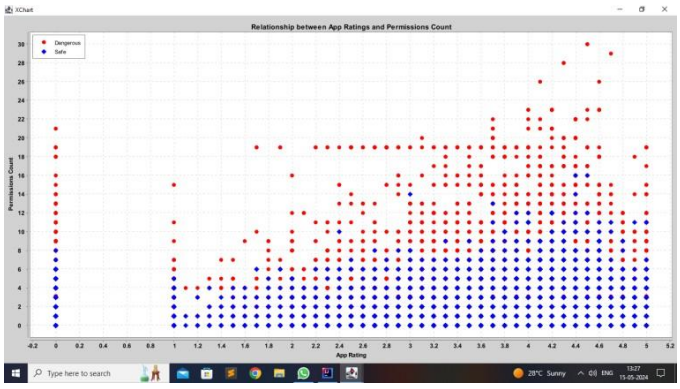
Additionally, the project advocates for the incorporation of Domain Generation Algorithms (DGAs) in DNS traffic analysis. DGAs, frequently utilized by botnets to locate and communicate with command and control servers, imprint unique temporal patterns within DNS traffic. Leveraging decision tree classifiers to recognize these patterns ensures efficient botnet detection within expansive Internet Service Provider (ISP) networks.

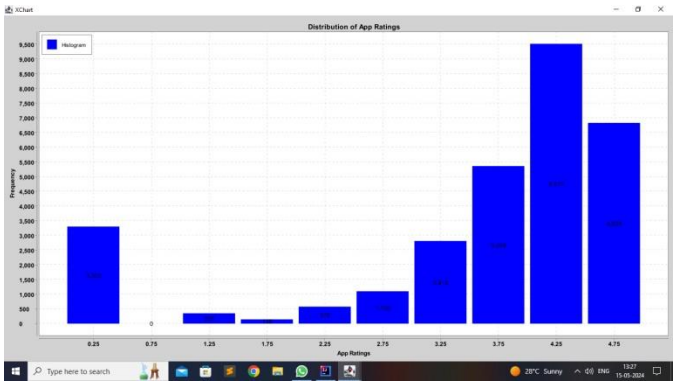
Moreover, the project integrates advanced text summarization techniques, employing a fusion of TFIDF and Textrank algorithms alongside natural language processing methods. This text summarization mechanism sifts through voluminous digital data, furnishing concise and pertinent summaries. This not only facilitates efficient data analysis but also expedites the identification of detrimental botnets, thereby conserving valuable time and resources.

By synergizing these innovative methodologies, the project endeavors to cultivate an intelligent, adaptive, and efficient botnet detection system. Such a system will constitute a substantial contribution to the cybersecurity domain, bolstering the precision and agility of botnet identification and fortifying digital defenses against the ever-evolving panorama of cyber threat.

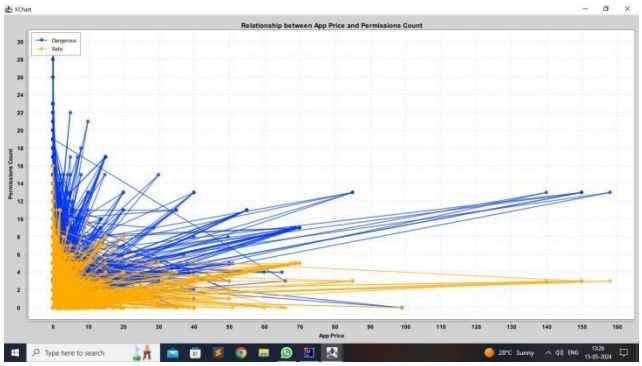
5. RESULT

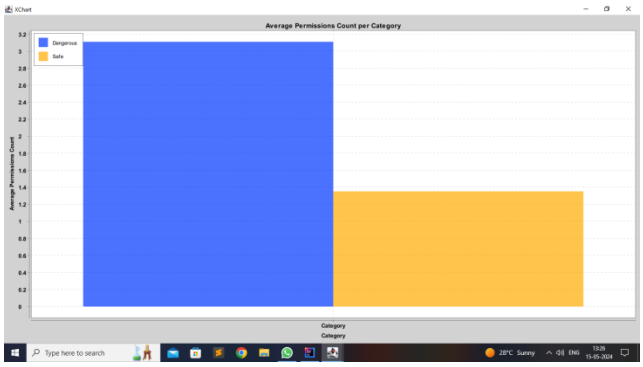


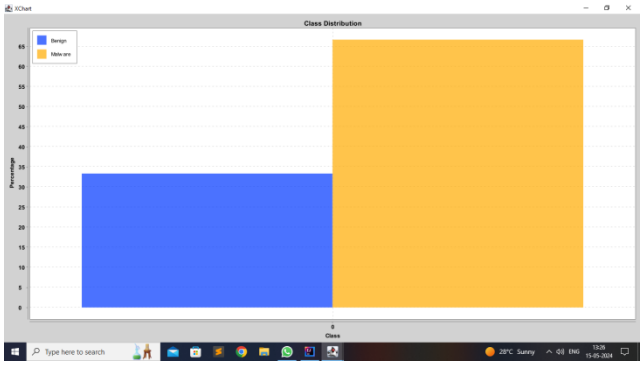


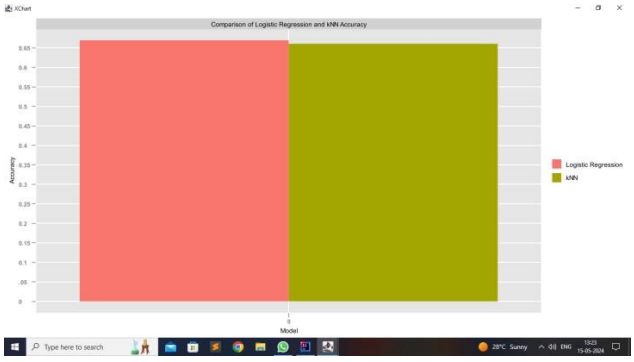


International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May–2024 DOI: 10.55041/ISJEM01861
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata









6. CONCLUSIONS

In conclusion, the Data Flow Diagrams (DFDs) presented in this discussion offer a comprehensive visualization of the system's data flow, processes, and interactions. DFDs are invaluable tools for understanding the system's architecture and can serve as a foundation for system design, analysis, and improvement. The Level 0 DFD (Context Diagram) provides a holistic view of the entire system, outlining its boundaries and interactions with external entities. The Level 1 DFD elaborates on specific processes, data flows, and data stores, providing detailed insights into the system's internal workings.

REFERENCES

[1] - THRESHOLD IDENTIFICATION FOR HTTP BOTNET DETECTION by 1 NUR HIDAYAH M. S, 1 FAIZAL M. A, 2 WAN AHMAD RAMZI W. Y, 1 RUDY FADHLEE M. D at www.jatit.org.

[2] - Botnet Identification Based on Flow Traffic by Using K-Nearest Neighbor by Dani Gunawan, Tika Hairani, Ainul Hizriadi at IEEE Xplore.

[3]-DGA Bot Detection with Time Series Decision Trees by Anaël Bonneton ,Daniel Migault ,Stephane Senecal and Nizar Kheir at IEEE

[4] –Identifying Malicious Botnet Traffic using Logistic Regression by Rohan Bapat, Abhijith Mandya, Xinyang Liu, Brendan Abraham, Donald E. Brown, Hyojung Kang, and Malathi Veeraraghavan at IEEE

[5] –MOBILE BOTNET DETECTION: A MACHINE LEARNING APPROACH USING LOGISTICREGRESSION by Akbar Shaikh1, Ganesh Landage2, Sanket Thorat3, Prasad Shinde4, Mamta Sharma at IEEE