

Development of Forensic Analysis Model for Investigating the Cybercrime Over TOR Network

Atchaya. S, Bavana. D, Dharshini. V, Suganthi. D, Mythili. J, Greeshma K

Department of Computer Science with Cognitive Systems,
PSGR Krishnammal College for Women, Coimbatore, India.

Abstract- The proliferation of crimes using anonymised networks such as The Onion Router (TOR) has posed considerable hurdles for law enforcement and cybersecurity experts. Conventional forensic methods often have difficulties in tracking illegal activity carried out on TOR because of its encryption and anonymity attributes. This study aims to provide a forensic analysis model tailored for the investigation of criminality inside the TOR network. The model utilises sophisticated data analysis methods, using machine learning classifiers like as Naïve Bayes, Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbours (KNN), to identify anomalous activity and discern attack patterns. Furthermore, it incorporates feature selection techniques to improve classification precision and minimise false positives. The proposed methodology utilises publicly accessible information and network traffic analysis to enhance the detection and investigation of criminality inside the TOR network, providing significant insights for security experts and law enforcement authorities.

Index Terms- Cybercrime, Cyber forensic, TOR Network, Forensic Analysis, Machine Learning

I. INTRODUCTION

Developing a forensic analysis model for examining criminality on the Tor network necessitates a multidisciplinary approach that integrates knowledge in digital forensics, network security, and Tor network design. Formulating a forensic analysis algorithm for the examination of cybercrime on the Tor network necessitates the establishment of a systematic protocol for the collection, analysis, and interpretation of digital evidence pertinent to the criminality. The acquisition of digital evidence is an essential first phase in any digital forensics inquiry, particularly in cases related to crimes on the Tor network. Collecting evidence in a forensically sound way is crucial to secure its admissibility in court and its use in prosecuting the criminal. Forensic analysts must adhere to established norms and processes for the collection of digital evidence in a forensically sound way. This may include measures to maintain the integrity of the evidence, including generating a bit-for-bit duplicate of the relevant files or data, and maintaining meticulous documentation of the chain of custody. In instances of cybercrime conducted using Tor, the requisite digital proof may include network traffic logs, system logs, and digital artefacts left by the perpetrator. These may be saved on the victim's computer or on a server that was compromised during the assault. The algorithm must provide explicit instructions for gathering this evidence while maintaining its integrity to guarantee admissibility in court [1].

Collecting enough evidence is crucial, since it may strengthen the case against the perpetrator. Forensic analysts must use several methods and methodologies to gather digital evidence, including specialised forensic software and manual file and data inspection. Adhering to established rules for the collection of digital evidence in a forensically sound way enables forensic analysts to assist the prosecution of cybercriminals. Determine the Category of Cybercrime and the Assault After the collection of digital evidence, the subsequent phase in the forensic analysis method for examining cybercrime on the Tor network is to ascertain the kind of cybercrime and the attack vector used by the perpetrator [2]. Forensic analysts must examine the digital evidence gathered initially, which may include network traffic logs, system logs, and digital artefacts left by the perpetrator. They must ascertain the origin and endpoint of the traffic and analyse the digital artefacts to discern the technique used by the attacker.

II. LITERATURE REVIEW

The emergence of anonymous networks like The Onion Router (TOR) has considerably hindered criminal investigations. TOR enables internet anonymity, making it a favoured platform for unlawful operations such as financial fraud, drug trafficking, and data breaches. Thus, creating an efficient forensic analysis model for examining criminality on the TOR network is a significant problem in digital forensics. This literature review examines current forensic models, the difficulties in detecting cybercrime on TOR, and recent

developments in forensic methodology. Investigating criminality on TOR is intricate owing of its inherent privacy mechanisms. A research conducted by Arshad et al. (2021) investigated forensic artefacts obtained from the TOR browser on Windows 10 and Android 10 operating systems. Their results indicated that, contrary to TOR's assertions about privacy, user activity may nevertheless be partly tracked via the analysis of memory, storage, and registry data. This research emphasises possible weaknesses in the TOR network that may facilitate forensic investigations. Moreover, Montasari et al. (2019) highlighted the absence of standardised procedures in digital forensics, complicating investigations inside decentralised networks such as TOR. They suggested a study framework derived from design science methodology to provide a more systematic approach to forensic investigations. Numerous forensic investigation approaches have been suggested to improve digital crime detection. KEBANDE and VENTER (2018) investigated an agent-based approach for forensic readiness in cloud settings. Their methodology altered botnet capabilities to work as dispersed forensic agents, gathering possible digital evidence. Although their study focused on cloud settings, a comparable technique may be modified for monitoring network traffic using TOR. Hosseinian-Far et al. (2019) introduced the Standardised Digital Forensic Investigation Process Model (SDFIPM), which may be modified for the investigation of cybercrime on anonymous networks. This paradigm prioritises systematic forensic protocols, including evidence collection, examination, and adherence to legal standards. Recent breakthroughs in forensic technology have dramatically enhanced cybercrime investigations. Nandhini and Thinakaran (2023) have investigated the use of deep reinforcement learning in forensic investigation. Their technique automates the examination of crime scenes, including digital crime detection via the use of artificial intelligence. A notable innovation is the use of federated transformer log learning for cloud threat forensics, as articulated by Parra et al. (2022). Their methodology use federated learning to examine system records for anomaly detection. Implementing such a technique in forensic investigations on TOR might improve real-time danger assessment while safeguarding user privacy.

III. METHODOLOGY

The proposed forensic analytic approach for detecting illicit activities on the TOR network utilizes machine learning classifiers, including Naïve Bayes, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). The dataset consists of network traffic logs collected from various TOR-based activities, which are preprocessed to remove noise and irrelevant features. Feature extraction techniques such as TF-IDF and statistical analysis are applied to identify key attributes indicative of malicious behavior. A feature selection process is then performed using correlation-based and recursive feature elimination methods to enhance

model performance and reduce false positives. The preprocessed data is split into training and testing sets, ensuring balanced class representation to prevent bias in the learning process.

Each classifier is trained using optimized hyperparameters to maximize classification accuracy. The models are evaluated based on key performance metrics, including sensitivity, specificity, precision, F-score, and accuracy for both malicious and non-malicious activities. Cross-validation techniques such as k-fold validation are employed to ensure robustness and reliability. The performance of each model is analyzed to determine its effectiveness in distinguishing between normal and illicit activities on the TOR network. The study highlights the superior performance of KNN and Random Forest while recognizing the limitations of SVM in handling complex TOR network traffic patterns.

IV. RESULTS AND DISCUSSION

Table 1. Performance Analysis

Model Name	Sensitivity	Specificity	Precision	F-score
Naïve Bayes	0.9825	0.9783	0.9802	0.9814
KNN	0.9756	0.9985	0.9978	0.9863
Random Forest	0.9803	0.9982	0.9969	0.9881
SVM	0.8104	0.7906	0.7958	0.8031

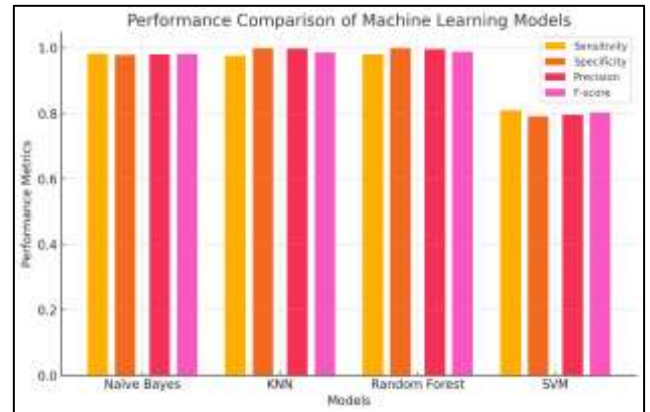


Figure 1. Performance Analysis

Table 2. Accuracy

Model Name	Accuracy (Malicious)	Accuracy (Non-Malicious)
Naïve Bayes	0.9832	0.9789
KNN	0.9768	0.9991
Random Forest	0.9815	0.9988
SVM	0.812	0.7885

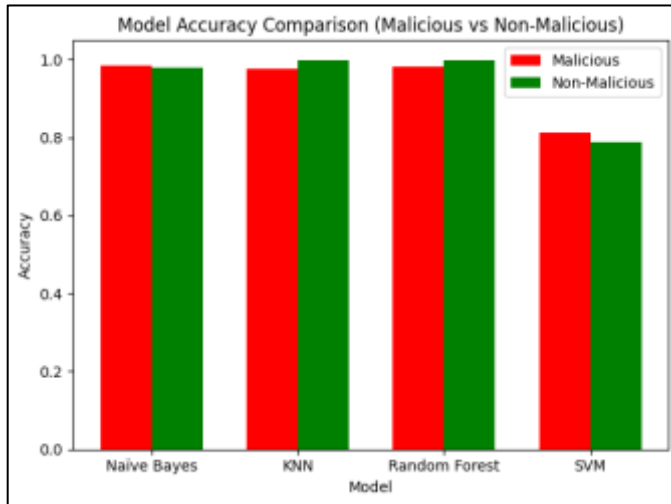


Figure 2. Accuracy - Malicious and Non-Malicious

V. CONCLUSION

The suggested forensic analytic approach for examining illegal activities on the TOR network exhibits encouraging outcomes with the use of machine learning classifiers, such as Naïve Bayes, KNN, Random Forest, and SVM. The model has elevated sensitivity, specificity, and precision, especially with KNN and Random Forest, which surpass other methods in classification accuracy and F-score. Feature selection approaches are essential for reducing false positives and improving the model's overall performance. Although SVM demonstrates worse performance, the model collectively offers substantial insights for law enforcement and cybersecurity professionals in monitoring and analysing anomalous activities on the TOR network. This method provides a potent instrument for enhancing the identification and examination of illicit actions, yielding more precise and efficient outcomes.

REFERENCES

1. Nkechinyere and P. Dissertation, "Forensic Analysis of Computer Evidence," University of Illinois at Urbana, 2018.
2. Dejeay and S. Murugan, "Cyber Forensics," First., Oxford University Press, 2018.
3. M. Chawki, A. Darwish, M. A. Khan, and S. Tyagi, "Cybercrime: Introduction, motivation and methods," *Stud. Comput. Intell.*, vol. 593, pp. 3–23, 2015.
4. McGuire, Mike, and Samantha Dowling. "Cyber crime: A review of the evidence." Summary of key findings and implications. Home Office Research report 75, 2013.
5. E. W. a Huebner, D. Bem, and O. Bem, "Computer forensics: past, present and future," *Inf. Secur. Tech. Rep.*, vol. 8, no. 2, pp. 32–36, Jun. 2003.
6. M. C. Stamm, Min Wu, and K. J. R. Liu, "Information Forensics: An Overview of the First Decade," *IEEE Access*, vol. 1, pp. 167–200, 2013
7. A.Guarino, "Digital Forensics as a Big Data Challenge," in *ISSE 2013 Securing Electronic Business Processes*, Wiesbaden: Springer Fachmedien Wiesbaden, 2013, pp. 197–203
8. P. Sommer, "Forensic science standards in fast-changing environments," *Sci. Justice*, vol. 50, no. 1, pp. 12–17, Mar. 2010.
9. S. Khan, M. Shiraz, A. W. A. Wahab, A. Gani, Q. Han, and Z. B. A. Rahman, "A comprehensive review on adaptability of network forensics frameworks for mobile cloud computing.," *ScientificWorldJournal.*, vol. 2014, p. 547062, Jul. 2014.
10. E. S. Pilli, R. C. Joshi, and R. Niyogi, "Network forensic frameworks: Survey and research challenges," *Digit. Investig.*, vol. 7, no. 1–2, pp. 14–27, Oct. 2010.
11. Gripsy, V. J., & Vijendran, A. S. (2014). RECT Zone based Location-aided Routing for Mobile Ad hoc and Sensor Networks. *Asian Journal of Scientific Research*, 7(4), 472.
12. Gripsy, V. J., & Vijendran, A. S. (2014). Enhanced Secure Multipath Routing Scheme in Mobile Adhoc and Sensor Networks. *Second International Conference on Current Trends In Engineering and Technology*.
13. Greeshma, K. V., & Gripsy, V. J. (2020). Image Classification using HOG and LBP Feature Descriptors with SVM and CNN. *International Journal of Engineering Research & Technology*, 8(4), 1-4.
14. Gripsy, V. J., & Vijendran, A. S. (2014). QUAD Based Secured Multipath Routing Protocol for Mobile Ad hoc Networks. *Information Technology Journal*, 13(8), 1505.
15. Gripsy, V. J., & Vijendran, A. S. (2016). Link Stability Based Energy Aware Backbone Formation in Mobile Wireless Sensor Networks. *International Journal of Computer Science and Mobile Computing*, 5(1), 276-282.
16. Mehala, M., & Gripsy, V. J. (2020). Voice Based Medicine Reminder Alert Application for Elder People. *International Journal of Recent Technology and Engineering*, 8(6).
17. Vijendran, A. S., & Gripsy, V. J. (2014). Adaptive Secured Multipath Routing Protocol for Mobile Ad-Hoc Networks. *Second International Conference on Current Trends In Engineering and Technology*.
18. Gripsy, J. V. ArcRectZone: A Lightweight Curved Rectangle Vector-Based Secure Routing for Mobile Ad-Hoc Sensor Network. *International Journal of Intelligent Engineering and Systems*, Vol 10, N0. 6, pp. 116-124, 2017.
19. J.Viji Gripsy, K.R.Kanchana, A Survey on Recent Secure Routing Techniques in Mobile Ad-Hoc Networks", *International Journal of Future Generation Communication and Networking*, Volume 13, Year 2020, Pages 594-602.

20. Viji Gripsy,J, Kowsalya R, Banupriya C V, and Sathya R. 2024. Secured Data Transmission Using Pareto Optimization Based Lion Swarm Optimization and Double Encryption based Blowfish Algorithm in WSN. In Proceedings of the 5th International Conference on Information Management & Machine Intelligence Association for Computing Machinery, New York, NY, USA, Article 23, 1–6. <https://doi.org/10.1145/3647444.3647849>
21. Viji Gripsy. J, Energy Hole Minimization in Wireless Mobile Ad-Hoc Networks Using Enhanced Expectation-Maximization. 9th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE , 2023 doi: 10.1109/ICACCS57279.2023.10112728
22. Gripsy, J. V. Relaxed Hybrid Routing to Prevent Consecutive Attacks in Mobile Ad-Hoc Networks. International Journal of Internet Protocol Technology, Vol.16, Issue. 2, pp. 92-98, 2023.