International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Forensic Analysis Model for Investigating Cybercrime Over the Network

Midhunya.P.S, Adhulya. D, Merlin Jenifer. L, D. Suganthi, J. Mythili, Dr. N. Prabhu
Department of Computer Science with Cognitive Systems, PSGR Krishnammal College for Women, Coimbatore, India

Abstract- Despite significant investments in security protocols, the frequency of cybersecurity incidents continues to rise, with traditional methods proving ineffective against complex cyber-attacks. This research aims to address this issue by using a publicly accessible dataset on Advanced Persistent Threats (APTs) to develop a data-driven approach for identifying APT phases within the Cyber Kill Chain framework. APTs are sophisticated and targeted attack strategies that can bypass conventional intrusion detection systems, posing a major challenge for security professionals. The study incorporates several machine learning classifiers, including Naïve Bayes, Bayes Net, KNN, Random Forest, and Support Vector Machine (SVM), to analyze the dataset and identify APT phases, offering a promising method for improving cybersecurity detection and response.

Index Terms-Cyber Kill Chain, cyber-attack, attack prediction, forensics, and machine learning

I. INTRODUCTION

The rapid increase in cybercrime incidents over the past decade has made network forensics a crucial field for investigating and mitigating security breaches. Traditional security measures often fall short in detecting sophisticated attacks, such as Advanced Persistent Threats (APTs) and other complex cybercrimes, which can evade conventional intrusion detection systems.

As cybercriminals continue to refine their techniques, the need for robust forensic analysis models has become more critical to understanding and addressing the tactics used by attackers. Forensic analysis of network traffic provides an effective method for uncovering traces of malicious activities, identifying attack phases, and determining the scope and impact of a security breach. By leveraging advanced methodologies, such as machine learning algorithms and data-driven approaches, investigators can better identify patterns and anomalies within network traffic, ultimately improving their ability to detect and respond to cyber threats.

This research focuses on developing a comprehensive forensic analysis model tailored to investigating cybercrime over the network, utilizing cutting-edge tools and algorithms for detection and analysis. The model incorporates a variety of techniques, including feature selection, classification methods, and the application of the Cyber Kill Chain framework, which breaks down cyber-attacks into distinct stages for more targeted analysis. The study employs publicly available datasets to train machine learning classifiers, such as Naïve Bayes, Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN), to identify phases of cyber-attacks and gain deeper insights into attacker behaviors. By combining established forensic principles with modern data analysis techniques, this model aims to provide a reliable and efficient approach to network forensics, offering valuable support for security professionals and investigators in the ongoing battle against cybercrime.

II. LITERATURE

APTs are advanced, long-term cyberattacks frequently targeted at governments and big organizations. Unlike traditional threats, APTs use low-and-slow attack patterns, hence hard to identify using signature-based methods [1]. Han et al. (2020) presented UNICORN, a provenance-based system for detecting APTs using graph-based anomaly detection, outperforming existing detection mechanisms in detecting stealthy threats (Han et al., 2020).

Current studies highlight the inclusion of machine learning (ML) methodologies to discover and interfere with APTs at every phase of the CKC. Machine learning application in cybersecurity has made notable progress, with the ability to identify sophisticated attacks automatically.

Dasgupta et al. (2020) offered an extensive review of ML methods for cybersecurity, including their success in intrusion and malware discovery and highlighting weaknesses like adversarial attacks (Dasgupta et al., 2020). Abrahams et al. (2024) investigated ML-based threat detection approaches with major focus on feature engineering and anomaly detection to enhance accuracy (Abrahams et al., 2024).

© 2024 IJSRET

Table 1. Feature Selection Techniques

Feature Selection	Accuracy	DR	FAR	F1
OneR	92.3	92.3	1.1	92.4
Gain Ratio	91	91	1	91.2
InfoGain	88.5	88.5	1.7	88.6

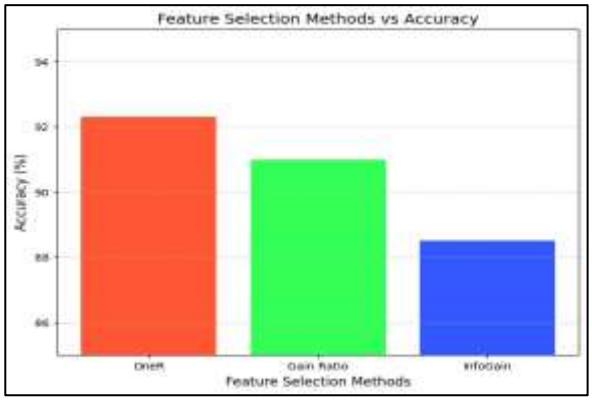

Figure 1. Accuracy for Feature Selection Methods

Table 2. Accuracy for Classification

No.of features	Naïve Bayes	Bayes net	k-NN	Random forest	SVM
5	88.7	80.5	84	83.6	88.12
7	87.43	91.35	89.6	87.87	91.41
7	88.24	90.85	89.86	87.68	90.79
10	91.1	90.73	89.05	87.43	90.79
14	87.31	82.15	83.58	84.32	87.43

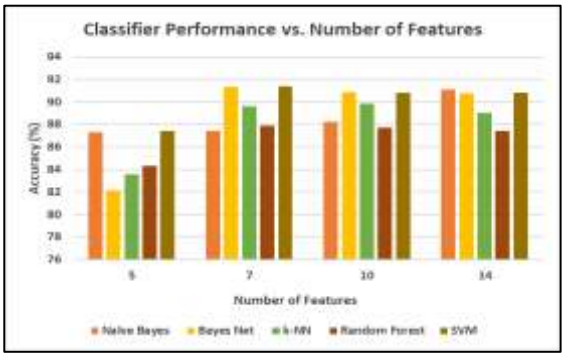

Figure 4. Classification Accuracy.

III. CONCLUSION

In conclusion, the feature selection procedures shown in Table 1 demonstrate the efficacy of several ways in enhancing model correctness. Among the methodologies, OneR attains the best accuracy at 92.3%, closely followed by Gain Ratio at 91%. InfoGain has a little worse performance at 88.5%. The findings indicate that feature selection techniques, especially OneR and Gain Ratio, may substantially improve model performance by enhancing classification accuracy, as shown by metrics such as Detection Rate (DR), False Alarm Rate (FAR), and F1 score. The classification accuracy, as seen in Table 3, indicates that the efficacy of various classifiers fluctuates according on the amount of characteristics used. Naïve Bayes consistently exhibits strong performance across all feature sets, especially with 10 features (91.1%), while SVM achieves the maximum accuracy at 91.41% with 7 features. Conversely, classifiers such as Random Forest and k-NN demonstrate comparatively steady although worse performance relative to Naïve Bayes and SVM. The findings demonstrate that the selection of classifier and feature selection technique may profoundly influence the accuracy of network classification models, underscoring the need of both feature selection and classifier optimisation for peak performance.

REFERENCES

1. GReAT, “APT trends report Q1 2020 APT trends report Q1 2020,” SECURELIST, Kaspersky, 2020. [Online]. Available: https://securelist.com/apt-trends-report-q1-2020/96826/ [Accessed: Septem-ber 03, 2020].
2. T. Yadav and A. M. Rao, “Technical aspects of cyber kill chain,” in Proc. Int. Symp. on Security in Computing and Communications, Kochi, India, pp. 438–452, 2015.
3. D. Kiwia, A. Dehghantanha, K.-K. R. Choo and J. Slaughter, “A cyber kill chain based taxonomy of banking Trojans for evolutionary computational intelligence,”Journal of Computational Science,vol.27, pp. 394–409, 2018.
4. P. N. Bahrami, A. Dehghantanha, T. Dargahi, R. M. Parizi, K.-K. R. Choo et al.,“Cyber kill chain-based taxonomy of advanced persistent threat actors: Analogy of tactics, techniques, and procedures,” Journal of Information Processing Systems, vol. 15, no. 4, pp. 865–889, 2019.
5. F. A. Garba, S. B. Junaidu, I. Ahmad and M. Tekanyi, “Proposed framework for effective detection and prediction of advanced persistent threats based on the cyber kill chain,”Scientific &Practical Cyber Security Journal, vol. 3, pp. 1–11, 2019.
6. J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino et al.,“Feature selection: A data perspective,” ACM Computing Surveys, vol. 50, no. 6, pp. 1–45, 2017.
7. Vijayarani, S., Suganya, E., & Navya, C. (2021). Crime analysis and prediction using enhanced Arima model. Journal homepage: www. ijrpr. com ISSN, 2582, 7421.
8. D. Stabili, M. Marchetti and M. Colajanni, “Detecting attacks to internal vehicle networks through Hamming

© 2024 IJSRET

distance," inProc. AEIT Int. Annual Conf., Cagliari, Italy, pp. 1–6, 2017.
9. S. J. Lee, P. D. Yoo, A. T. Asyhari, Y. Jhi, L. Chermak et al.,"IMPACT: Impersonation attack detection via edge computing using deep autoencoder and feature abstraction," IEEE Access, vol. 8, pp. 65520– 65529, 2020.
10. F. Vanhoenshoven, G. Napoles, R. Falcon, K. Vanhoof and M. Koppen, "Detecting malicious URLs using machine learning techniques," inProc. IEEE Symp. Series on Computational Intelligence,Athens, Greece, pp.1–8, 2016.
11. B. Anderson and D. McGrew, "OS fingerprinting: New techniques and a study of information gain and obfuscation," inProc. IEEE Conf. on Communications and Network Security, Las Vegas, 2017.
12. 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
13. 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 (ICIMMI '23). Association for Computing Machinery, New York, NY, USA, Article 23, 1–6. https://doi.org/10.1145/3647444.3647849

© 2024 IJSRET