

# Automated Malware and Phishing Website Detection Using Cluster Ensemble Techniques for Cybercrime Prevention

Nega. B, Rithika. K, Rithika. V, D. Suganthi, J. Mythili, Dr. N. Prabhu

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

**Abstract-** Cybercrime is a specialised field that use internet communication networks to enhance the identification of cyber offenders via cyber laws. Extensive study is being undertaken to provide appropriate legal methodologies for preventing and regulating cybercriminal activity. Malware and phishing detection have become as prominent subjects in the last decade because to the harm they inflict on internet users. The identification of phishing websites is a novel area in the discipline. Phishing websites are regarded as a significant threat for the exploitation of personal information for the benefit of cybercriminals. This research presents an automated classification system designed to identify malware and phishing websites by integrating several clustering techniques using a cluster ensemble approach.

**Index Terms-**Forensics, Cybercrime, Social Networks, Phishing website, Clustering

## I. INTRODUCTION

The study of forensic cybercrime investigations in social networks use digital forensic tools to examine cybercrime incidents inside these platforms. This research aims to gather digital evidence for legal processes to identify, prosecute, and punish cybercriminals. Network forensics is the collection and analysis of data on network traffic to detect and investigate security events. This entails the acquisition and examination of network packets, together with the reconstruction of network activity to ascertain the origin and characteristics of any possible assaults [2].

Network forensics encompasses several approaches and technologies, such as packet sniffers, intrusion detection systems, and network analysis instruments. These technologies enable forensic investigators to record and analyse network data, detect anomalous behaviour, and reconstruct network activities to ascertain the origin and extent of security breaches. The objective of network forensics is to collect comprehensive information on a security incident, including the chronological sequence of events, the systems and people implicated, and the magnitude of any damage or breach. This information may then be used to discover and address vulnerabilities inside the network, as well as to formulate measures to avert future security issues. Emerging network forensics is the investigation of network-related crimes or security events by the analysis of network traffic, logs, and other data sources.

## II. FORENSIC CYBERCRIME INVESTIGATIONS IN TYPES OF SOCIAL NETWORKS

Forensic cybercrime investigations on social networks include the collection and analysis of digital evidence to identify, locate, and prosecute offenders who use social media platforms for crimes such as fraud, harassment, identity theft, and online grooming. The following categories of social networks are used for forensic cybercrime investigations:

- Facebook
- Twitter
- LinkedIn
- Instagram
- Snapchat

Forensic cybercrime investigations include the collection and analysis of digital evidence, such as text messages, photographs, videos, and social media content, to reconstruct the crime and ascertain the perpetrator's identity. This procedure requires specialised knowledge and experience in digital forensics, cybercrime, and social media analysis [10]. Network forensics involves the capture, analysis, and interpretation of network data to discover and analyse security risks and occurrences. It entails using a variety of tools and methodologies to gather data on network traffic, reconstruct network occurrences, and analyse the information to detect indications of a security breach.

### III. ONLINE SOCIAL NETWORK FORENSICS (OSNF)

Online social networks (OSNs) have become essential to our lives and, thus, a venue for illegal activity. Online Social Network Forensics (OSNF) involves the collection, analysis, and preservation of electronic evidence from online social networks to investigate illegal actions. Nonetheless, several obstacles are linked to OSNF, including:

- Privacy
- Data Volume
- Data Validity
- Jurisdiction
- Encryption

Table 1. Comparison of Malware sample Detection rate per 10

Clustering Methods	Malware in the rate of 10
K-Means	8.5
K-Medoids	8.2
Hierarchical Clustering	8
DB Scan	8.9

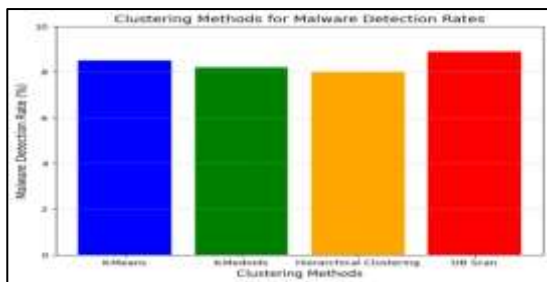


Table 2. Comparison of Phishing Websites Detection rate per

Clustering Methods	Phishing pages 10 per search
K-Means	8.8
K-Medoids	8.9
Hierarchical Clustering	8.2
DB Scan	8.7

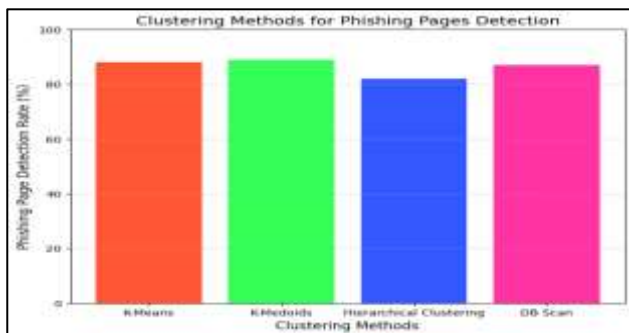


Figure 1. Comparison of Phishing Websites Detection rate per 10Pages.

### IV. CONCLUSION

This methodology is gaining significance as the prevalence of online criminal behaviour increases, with perpetrators using advanced ways to evade detection. Cybercrime denotes illicit crimes perpetrated using computers or the internet, including hacking, phishing, and identity theft. The rise of digital technology has become cybercrime a significant problem for people, corporations, and governments globally. Social network analysis (SNA) is a methodology for examining social networks via the evaluation of interactions among people, groups, and organisations. This methodology is applicable across several disciplines, including sociology, psychology, and marketing. In the realm of cybercrime, Social Network Analysis (SNA) may be used to examine the connections among people and organisations engaged in illicit activity, including botnets and other forms of cyberattacks. Investigators may get insights into the motives, techniques, and strategies of cybercriminals by analysing their social networks, which may aid in preventing future assaults [4].

### REFERENCES

1. Aburrous M., Hossain M.A., Dahal K. and Thabtah K. (2010), 'Predicting phishing websites using classification mining techniques with experimental case studies' in Proc. 7th Int. Conf. Inf. Technol., pp. 176–181
2. Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. Society, 54(2), 138-149. doi:10.1007/s12115-017-0114-0
3. Chen, L., Xu, L., Yuan, X., & Shashidhar, N. (2015). Digital forensics in social networks and the cloud: Process, approaches, methods, tools, and challenges. International Conference on Computing, Networking and Communications (ICNC) (pp. 1132-1136). Garden Grove: IEEE. doi:10.1109/ICNC.2015.7069509
4. Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., & Weippl, E. (2011). Social snapshots: digital forensics for online social networks. Proceedings of the 27th Annual Computer Security Applications Conference (pp. 113-122 ). Orlando: ACM.
5. Vijayarani, S., Suganya, E., & Navya, C. (2021). Crime analysis and prediction using enhanced Arima model. Journal homepage: www.ijrpr.com ISSN, 2582, 7421.
6. Jain, A., & Chhabra, G. S. (2014). Anti-Forensics Techniques: An Analytical Review. Seventh International Conference on Contemporary Computing (IC3). Noida: IEEE. doi:10.1109/IC3.2014.6897209
7. Javed, A., Burnap, P., & Rana, O. (2019, May). Prediction of drive-by download attacks on Twitter. Information Processing & Management, 56(3), 1133-1145. doi:https://doi.org/10.1016/j.ipm.2018.02.003
8. Lillis, D., Becker, B. A., O'Sullivan, T., & Scanlon, M.

- (2016). Current challenges and future research areas for digital forensic investigation. arXiv preprint arXiv:1604.03850, 1-11. Retrieved from <https://arxiv.org/pdf/1604.03850.pdf>
9. Myhre, J. W., Mehl, M. R., & Glisky, E. L. (2017). Cognitive Benefits of Online Social Networking for Healthy Older Adults. *The Journals of Gerontology: Series B*, 72(5), 752-760. doi:10.1093/geronb/gbw025
  10. Dr.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
  11. 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>