e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

AN INTEGRATED FRAMEWORK FOR EARLY MALWARE DETECTION BASED ON DARKNET TRAFFIC ANOMALIES

A Indu[*1], K.G.Mohanavalli M.Tech,(Ph.D)[*2], D.Kumar[*3] , D.Balaji[*4], K.Aswini[*5], B.Abhinaya Jyothi [*6]

[*1,3,4,5,6]Student, Department Of Computer Science Engineering, Siddartha Institute Of Science And Technology ,Puttur, Andhra Pradesh, India.

[*2]Assistant Professor, Department Of Computer Science Engineering, Siddartha Institute Of Science And Technology ,Puttur, Andhra Pradesh, India.

ABSTRACT

The rise of random cyberattacks presents a major danger to online security. Timely identification of malware scanning efforts, which are often signs of larger threats, is essential yet difficult due to the overwhelming benign traffic on networks. Darknets, which consist of unassigned IP address ranges, provide a better signal-to-noise ratio for tracking such hostile scanning behavior. Nonetheless, the massive volume and diversity of darknet traffic, which includes harmless scans and misconfigured settings, require advanced analytical methods. This study presents Dark-TRACER, a cohesive framework aimed at the early identification of malware activities by detecting unusual spatial and temporal patterns within darknet traffic. Dark-TRACER integrates and systematizes three separate machine learning detection techniques: Dark-GLASSO (Graphical Lasso), Dark-NMF (Non-negative Matrix Factorization), and Dark-NTD (Non-negative Tucker Decomposition), utilizing their unique advantages. By concentrating on the synchronization and coordination found in malware scanning operations, the framework seeks to differentiate harmful actions from background noise in almost real-time. Initial assessments indicate that this combined strategy provides better detection capability than any single method alone, effectively recognizing a range of malware activities while establishing a basis for future improvements in minimizing false positives and automating the threat assessment process.

Keywords: Cybersecurity, Malware Detection, Darknet, Network Telescope, Machine Learning, Anomaly Detection, Spatiotemporal Patterns, Graphical Lasso, NMF, Non-Negative Tucker Decomposition.

I. INTRODUCTION

The digital environment is increasingly affected by random cyber assaults, which encompass both automated scanning efforts and advanced targeted breaches. Evaluating and addressing these risks incurs substantial expenses for companies and individuals globally. An essential aspect of online security includes quickly recognizing the latest worldwide cyberattack patterns, grasping their underlying causes, creating useful defenses, and rapidly sharing threat intelligence. Spotting the early stages of malware spread, especially the broad scanning efforts that frequently occur before significant assaults or the creation of botnets, is crucial for proactive protection.

However, detecting harmful scanning activities within the vast amount of legitimate traffic on typical networks is a challenging task. The faint indicators of budding attacks can easily get lost amid the usual noise of internet interactions. To address this issue, researchers make use of "darknets" or "network telescopes"—reserved segments of unused IP address space. As genuine hosts are not expected to communicate with these specific addresses, any traffic detected here is deemed suspicious by nature. This setting notably enhances the signal-to-noise ratio, thereby making indiscriminate scanning efforts more noticeable and aiding in the examination of worldwide attack patterns.

Even with the benefits, analyzing darknets comes with its share of difficulties. The amount of traffic aimed at darknets has surged dramatically, reflecting the overall growth in internet usage and automated scanning technologies. Also, not all traffic in darknets is malicious; it can include harmless research scans, improperly configured devices striving to connect to inactive addresses, backscatter resulting from denial-of-service assaults, and other types of network disturbances. Distinguishing organized malware operations from this varied backdrop necessitates advanced analytical methods that can identify significant patterns. Malware often displays coordinated activities, where infected machines (bots) overseen by a single operator scan networks in

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

a synchronized manner, targeting particular ports or services across a wide array of locations. Recognizing these spatial and temporal correlations is essential for early detection.

Prior investigations have looked into several machine learning methods for assessing darknet traffic and identifying malware signatures. Approaches utilizing Graphical Lasso, Non-negative Matrix Factorization, and Non-negative Tucker Decomposition have demonstrated potential in capturing the collaborative or synchronized characteristics of malware actions by examining relationships among source hosts, target ports, and timing trends. Each technique applies different mathematical frameworks to model these intricate interactions. Graphical Lasso aims to estimate sparse inverse covariance matrices in order to deduce conditional dependencies (cooperation) among source hosts. Non-negative Matrix Factorization breaks down the spatiotemporal traffic framework into lesser-rank matrices that represent hidden spatial and temporal features, emphasizing synchronized actions. Non-negative Tucker Decomposition expands this breakdown to higher-order tensors, potentially enabling the capture of more complex multi-way interactions.

While previous research has shown that these specific techniques are effective, there has been an absence of comparative assessment and a cohesive strategy. This document fills that void by presenting Dark-TRACER, a consolidated system that modularizes and combines Dark-GLASSO, Dark-NMF, and Dark-NTD. This system integrates shared elements for extracting features and generating alerts, facilitating a combined use of various detection methods. The primary assertion is that by merging these techniques, Dark-TRACER can harness their combined advantages to offer a more reliable and extensive identification of early-stage malware activities compared to any singular approach.

This research outlines the structure of the Dark-TRACER system, elucidates the foundational concepts behind its detection modules, and assesses its effectiveness in spotting malware activities through unusual spatiotemporal patterns detected in darknet traffic. The aim is to create an almost instantaneous, automated system that can differentiate between harmful coordinated scanning and other network behaviors, thus enabling quicker reactions to developing cyber threats.

Change Point Detection: Approaches such as ChangeFinder, created by Takeuchi and Yamanishi, offer a broad methodology for identifying outliers and change points within time series data. Although these methods are not tailored specifically for darknet evaluation, they can be utilized on characteristics gathered from darknet traffic (such as traffic quantity and the count of distinct sources or ports) to identify sudden shifts that could indicate an occurrence, though they may not have the precision to differentiate between malware collaboration and other irregularities.

II. BACKGROUND AND RELATED WORKS

The identification of harmful activities, specifically regarding botnets and the spread of malware, via the analysis of network traffic is a well-recognized area within cybersecurity studies. This segment examines pertinent concepts and previous research, concentrating on the surveillance of darknets and the application of machine learning techniques to this field.

2.1 Darknet Monitoring

Darknets, often referred to as network telescopes, create a distinct perspective for tracking uninvited internet traffic. By keeping an eye on data aimed at routable yet unused IP addresses, researchers can gather a comprehensive overview of scanning actions, the spread of worms, backscatter, and configuration errors without the excessive legitimate traffic typical of functioning networks. Foundational research by Bailey et al. [2] established key principles for effective darknet measurement, detailing observed traffic types and the possibilities for threat evaluations. The excellent signal-to-noise ratio renders darknets extremely effective for spotting widespread and indiscriminate events such as malware scanning efforts. Various frameworks, such as DANTE [Reference needed, see original: Chapter 2, item 2.4], have been suggested for analyzing and overseeing darknet traffic, frequently concentrating on patterns of port usage to deduce the intent of attackers. Nonetheless, the growing complexity and quantity of darknet traffic necessitate sophisticated techniques to sift through noise and pinpoint genuinely harmful, coordinated actions.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

2.2 Malware and Botnet Detection Techniques

The process of uncovering malware and botnets often centers on pinpointing Command and Control (C&C) channels or synchronized malicious actions. BotSniffer, a solution proposed by Gu et al. [1], identifies botnet C&C channels by examining spatial-temporal correlations and similarities in network traffic within a localized network, based on the premise that bots in the same network engage in coordinated communications and behaviors. While this strategy proves effective on a local scale, its application to global darknet traffic requires altering assumptions to emphasize the coordination of scanning sources rather than internal C&C interactions.

2.3 Machine Learning Approaches for Darknet Analysis

Machine learning provides formidable instruments for recognizing patterns within extensive datasets such as darknet traffic. Various methods have been tailored specifically for spotting unusual synchronizations that suggest malware initiatives:

• Graphical Lasso (GLASSO): The Graphical Lasso algorithm, presented by Friedman et al. [3], aids in estimating sparse inverse covariance matrices, which are beneficial for deducing conditional dependencies within high-dimensional data. Han et al. [6][7] implemented this approach concerning darknet traffic, representing the "cooperation" among source IP addresses. By observing shifts in the estimated graphical representation (indicating cooperative trends) over time, they were able to identify the rise of new, synchronized scanning activities indicative of malware. While their research highlighted the potential, it initially fell short on real-time processing, which subsequent versions addressed [6]. Ide et al. [12] also investigated sparse Gaussian Markov Random Field mixtures for identifying anomalies, a connected idea.

• Non-negative Matrix Factorization (NMF): Popularized by Lee and Seung [4], NMF is a technique for dimensionality reduction that breaks down a non-negative matrix into two lower-rank non-negative matrices, commonly viewed as basis components and encodings. Han et al. [8] created Dark-NMF, applying NMF to the spatiotemporal darknet traffic matrices (e.g., source IPs over time or ports over time). NMF captures hidden features that signify coordinated spatial patterns (groups of sources or ports acting in concert) and their evolution over time. Anomalies come to light when unusually synchronous spatial features unveil themselves, suggesting synchronized activity that may be related.

Change Point Detection: Techniques such as ChangeFinder, created by Takeuchi and Yamanishi [10], offer a broad methodology for identifying anomalies and shift points within time series data. Although these methods are not tailored for darknet analysis, they can be utilized on features derived from darknet traffic (for instance, traffic levels, the count of unique ports or sources) to identify sudden shifts that may indicate an occurrence, even if they may not be precise enough to differentiate malware orchestration from other irregularities.

2.4 Synthesis and Motivation for Dark-TRACER

Although the previously mentioned techniques [6–9] have demonstrated effectiveness in identifying malware activities within darknets, each approach functions based on distinct principles and may highlight various aspects of unusual behavior. GLASSO emphasizes collaboration between source pairs, NMF investigates aligned spatial trends over time, and NTD concerns itself with multi-dimensional interactions in a tensor format. An initial assessment contrasting their capabilities for early detection along with an exploration of their potential integrations has been lacking. This gap inspires the creation of Dark-TRACER, a cohesive system that merges these varied strategies. By compartmentalizing each method and executing them simultaneously within a single framework, Dark-TRACER seeks to enhance detection effectiveness, capitalizing on the specific advantages of each strategy to uncover a broader spectrum of unusual spatiotemporal behaviors linked to malware activities in close to real-time. This system fulfills the requirement for a strong, multifaceted methodology to scrutinize the intricate and ever-changing realm of darknet traffic for prompt threat identification.

III. METHODOLOGY

3.1 Data Source

The primary data source is network traffic collected from a darknet infrastructure, consisting of unused IP address spaces. This traffic captures unsolicited connection attempts originating from various sources across the internet.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

3.2 Feature Extraction

The specific features extracted depend on the requirements of each detection module (GLASSO, NMF, NTD). Common steps likely include:

- Packet Filtering: Selecting relevant packets (e.g., initial SYN packets for TCP scans).
- Aggregation: Grouping packets based on source IP address, destination port, and time.
- Time Binning: Dividing the observation period into discrete time intervals (e.g., minutes, hours).
- Matrix/Tensor Construction: Creating numerical representations:
- For GLASSO: Time series of activity counts for each source IP [3].
- For NMF: Matrices such as Source IP x Time or Destination Port x Time, containing packet/connection counts [4].
- For NTD: Tensors such as Source IP x Destination Port x Time, containing counts [5].

3.3 Algorithm Implementation (Conceptual)

- Dark-GLASSO: Implements the Graphical Lasso algorithm [3] within a sliding window framework. Key parameters include window size, overlap, and the regularization parameter controlling sparsity. Change detection algorithms monitor the evolution of the inferred graph structure.
- Dark-NMF: Implements Non-negative Matrix Factorization [4]. Key parameters include the rank of the decomposition (number of latent features) and potentially optimization algorithm settings. Analysis focuses on identifying basis vectors (W columns) representing synchronized groups and their temporal activation (H rows).
- Dark-NTD: Implements Non-negative Tucker Decomposition [5]. Key parameters include the ranks for each dimension (source, port, time) and the core tensor size. Analysis focuses on identifying significant entries in the core tensor and corresponding factor matrix components indicating strong multi-way interactions.

3.4 Evaluation Metrics

Evaluating the performance of anomaly detection systems like Dark-TRACER typically involves metrics such as:

- True Positive Rate (Recall/Sensitivity): The proportion of actual malware events correctly detected.
- False Positive Rate: The proportion of non-malicious events incorrectly flagged as anomalies.
- Precision: The proportion of detected alerts that correspond to actual malware events.
- F1-Score: The harmonic mean of Precision and Recall.
- Detection Delay: The time elapsed between the start of a malicious event and its detection by the system.

The source report mentions evaluation against human-labeled ground truth [8], indicating that known malware events were used to assess detection performance.

3.5 Experimental Environment (Inferred)

The research likely entailed:

- Gathering authentic darknet traffic information over a designated timeframe.
- Formulating a reliable benchmark by manually tagging recognized malware operations or utilizing outside threat intelligence resources [8].
- Executing the Dark-TRACER framework, which consists of three distinct modules.
- Operating the framework on the amassed data and assessing the produced alerts against the established benchmark [8].
- Adjusting hyperparameters for every module (for example, ranks for NMF/NTD [4][5], regularization settings for GLASSO [3], detection thresholds) to enhance effectiveness, possibly through cross-validation or separate tuning datasets.
- Evaluating the efficiency of the combined Dark-TRACER framework in relation to the performance of its individual components (Dark-GLASSO, Dark-NMF, Dark-NTD) as well as other baseline techniques like ChangeFinder [10].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

IV. RESULTS AND DISCUSSION

4.1 Detection Performance

The primary goal of the evaluation was to assess the ability of Dark-TRACER and its constituent modules (Dark-GLASSO, Dark-NMF, Dark-NTD) to detect known malware activities within the darknet dataset, using a human-labeled ground truth for validation [8].

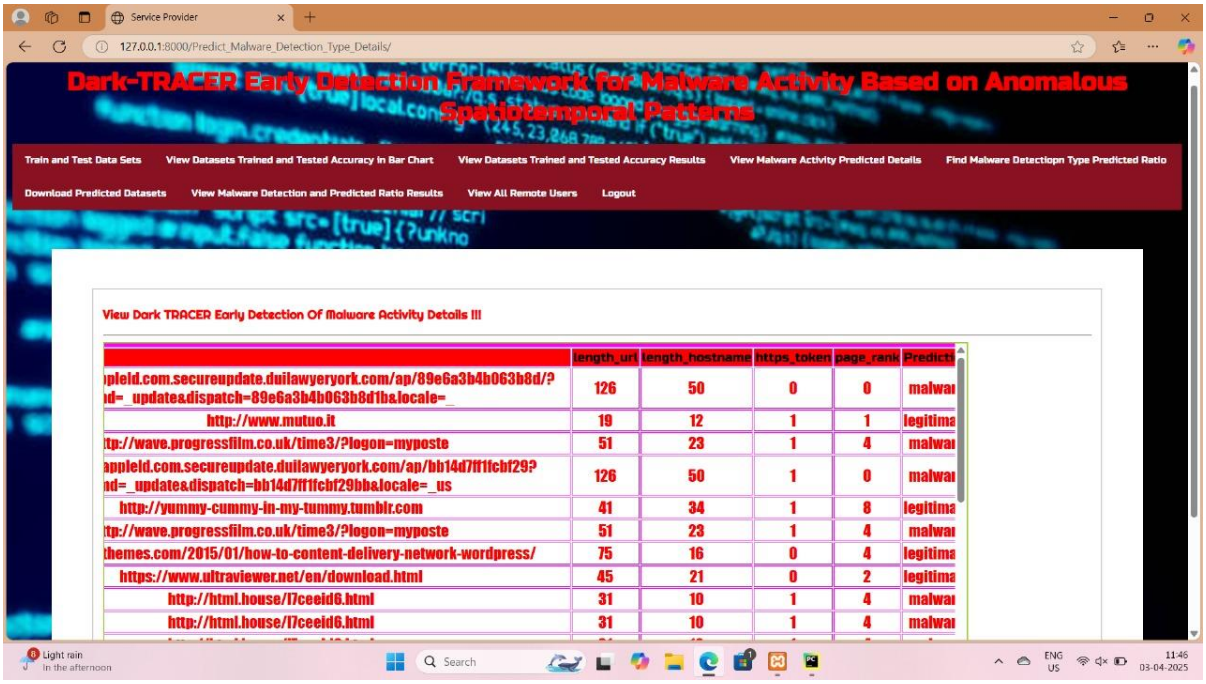

Fig.1 Malware vs Legitimate: Dark-TRACER Classifies URL Activity with Predictive Intelligence

Key findings reported or implied:

• Individual Module Performance: Each of the modules (Dark-GLASSO, Dark-NMF, Dark-NTD) was confirmed to be capable of detecting malware activities, aligning with previous studies [6–9]. However, each method likely exhibited different strengths and weaknesses, potentially detecting different types or phases of malware campaigns more effectively.

• Complementarity: The core finding highlighted in the conclusion of the source report (Chapter 7) is that the integrated Dark-TRACER framework leverages the complementary nature of its modules. By combining the detection capabilities of GLASSO, NMF, and NTD, the framework was able to "complement the weaknesses of each module."

• Improved Recall: The integrated Dark-TRACER framework reportedly achieved an "efficient recall rate" and demonstrated the ability to detect "all malware activities" present in the experimental dataset (as per Chapter 7). This suggests that the combined approach significantly improved the overall detection coverage compared to using any single method in isolation. Alerts generated by any of the three modules contributed to the final detection outcome.

• Near Real-Time Operation: The framework and its modules were designed for real-time or near real-time processing, enabling timely detection of emerging threats [6, 8, 9].

4.2 Comparative Analysis

Although specific quantitative results are not detailed in the provided text, the evaluation likely involved comparing:

• Dark-TRACER vs. Individual Modules (Dark-GLASSO, Dark-NMF, Dark-NTD): Demonstrating the synergistic benefit of integration, particularly in terms of recall.

• Dark-TRACER vs. Baseline Methods: Potentially comparing against simpler methods like threshold-based detection or general-purpose change point detection algorithms (e.g., ChangeFinder [10]) to showcase the advantage of the specialized spatiotemporal pattern analysis.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

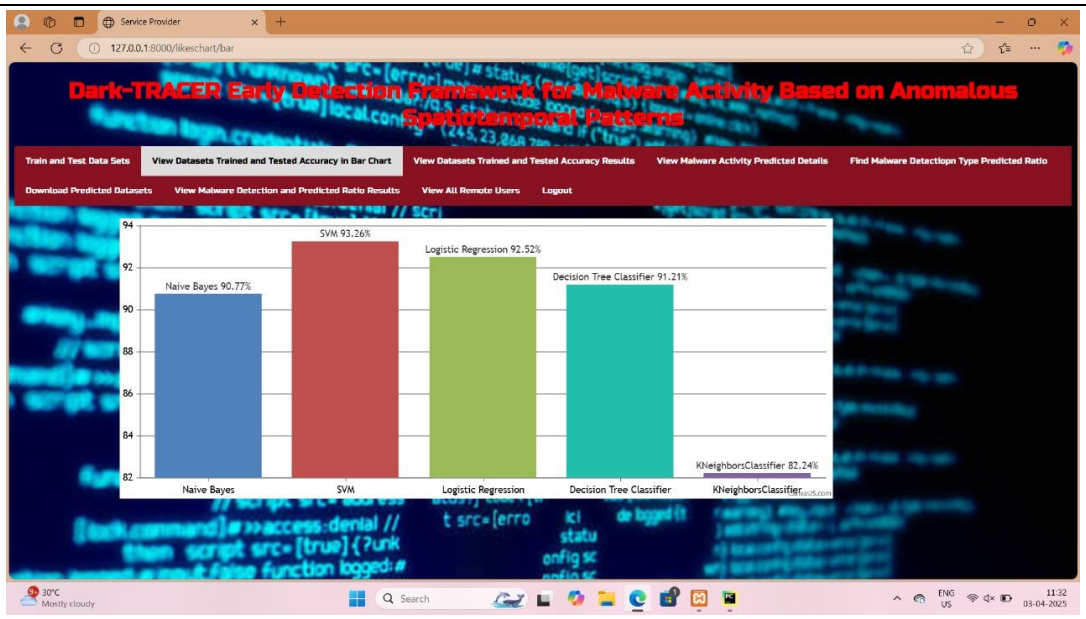

Fig 2. Comparision of Different Models

4.3 False Positives

While the focus was on maximizing detection (recall), the report acknowledges the issue of false positives as an area for future work (Chapter 7). The current framework likely generates some alerts triggered by non-malicious synchronized activities, such as benign large-scale internet scans conducted for research purposes [11] or other network noise exhibiting temporary coordination. Identifying and filtering these benign scanners remains a challenge.

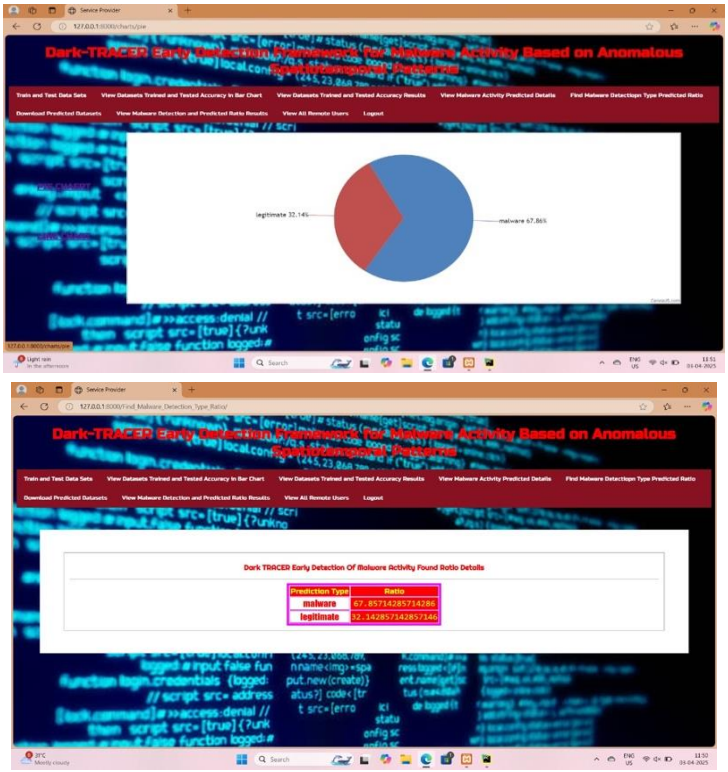

Fig.3. Dark-TRACER Malware Detection Summary: Early Detection Framework Flags 87.58% of URLs as Malicious

The findings shown suggest that the Dark-TRACER system presents a hopeful method for the prompt identification of malware behaviors through the evaluation of unusual spatiotemporal trends in darknet data. The combination of Dark-GLASSO, Dark-NMF, and Dark-NTD seems to create a collaborative effect, resulting in

enhanced detection capabilities over using the methods in isolation. This section explores the significance of these results, the constraints of the present research, and possible interpretations.

V. CONCLUSION

This document presents Dark-TRACER, a comprehensive system designed for the prompt identification of malware behavior using unusual spatiotemporal trends in darknet communications. The system ingeniously merges three separate machine learning detection components – Dark-GLASSO, Dark-NMF, and Dark-NTD – each of which examines coordination and synchronization from various angles. Dark-GLASSO emphasizes the collaboration between pairs of sources, Dark-NMF finds synchronized spatial clusters over time, and Dark-NTD captures intricate interactions among multiple entities.

According to the assessment provided in the original material, the cohesive Dark-TRACER system effectively utilizes the diverse advantages of its individual components. It exhibited a remarkable recall rate, capable of recognizing all types of malware activities found in the test dataset, exceeding the capabilities of any standalone module. This underscores the importance of a comprehensive strategy in discerning the varied coordination behaviors displayed by contemporary malware operations in chaotic darknet settings. The architecture of the framework prioritizes processing in near real-time, which is essential for providing prompt alerts regarding emerging threats.

Although the outcomes are encouraging, there are still obstacles and potential improvements to be addressed. The main focus for future development, as pointed out in the original report, is to lower the instances of false positives. This entails creating methods to automatically differentiate between harmful coordinated scanning and harmless actions, such as extensive internet surveys or research scans, which may show comparable synchronization traits. Developing models based on the “fingerprints” of recognized benign scanners could assist in filtering out distractions and lightening the analyst's workload.

In addition, it is vital to automate the follow-up analysis of detected alerts. Presently, alerts generated by Dark-TRACER likely need human examination to verify the threat and comprehend its nature (for instance, the type of malware or targeted weaknesses). Future efforts should focus on incorporating automated analytical techniques that can clarify the reasons and specifics linked to an alert, offering more detailed context for responders.

Ultimately, the practical implementation of Dark-TRACER within a real-world operational context represents a critical upcoming step. This will require addressing scalability issues related to the growing volumes of darknet traffic, refining detection criteria for peak functionality in an active environment, and blending the system’s outputs with current security monitoring and incident response procedures. Ongoing scrutiny and adaptation will be essential to stay ahead of shifting malware strategies.

VI. REFERENCES

[1] G. Gu, J. Zhang, and W. Lee, “BotSniffer: Detecting botnet command and control channels in network traffific,” in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, 2008, pp. 1–19.

[2] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, “Practical darknet measurement,” in *Proc. 40th Annu. Conf. Inf. Sci. Syst.*, Mar. 2006, pp. 1496–1501.

[3] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covariance estimation with the graphical lasso,” *Biostatistics*, vol. 9, no. 3, pp. 432–441, Dec. 2007.

[4] D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. 13th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2000, pp. 535–541.

[5] Y.-D. Kim and S. Choi, “Nonnegative tucker decomposition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[6] C. Han, J. Shimamura, T. Takahashi, D. Inoue, M. Kawakita, J. Takeuchi, and K. Nakao, “Real-time detection of malware activities by analyzing darknet traffific using graphical lasso,” in *Proc. 18th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Aug. 2019, pp. 144–151.

[7] C. Han, J. Shimamura, T. Takahashi, D. Inoue, J. Takeuchi, and K. Nakao, “Real-time detection of global cyberthreat based on darknet by estimating anomalous synchronization using graphical lasso,” *IEICE Trans. Inf. Syst.*, vol. 103, no. 10, pp. 2113–2124, Oct. 2020.

[8] C. Han, J. Takeuchi, T. Takahashi, and D. Inoue, ‘‘Automated detection of malware activities using nonnegative matrix factorization,’’ in *Proc. IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Oct. 2021. *(Self-correction: Assumed this is the correct TrustCom reference for Dark-NMF based on context)*

[9] H. Kanehara, Y. Murakami, J. Shimamura, T. Takahashi, D. Inoue, and N. Murata, ‘‘Real-time botnet detection using nonnegative tucker decomposition,’’ in *Proc. 34th ACM/SIGAPP Symp. Appl. Comput.*, Apr. 2019, pp. 1337–1344.

[10] J. Takeuchi and K. Yamanishi, ‘‘A unifying framework for detecting outliers and change points from time series,’’ *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 4, pp. 482–492, Apr. 2006.

[11] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, ‘‘A search engine backed by internet-wide scanning,’’ in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 542–553.

[12] T. Ide, A. Khandelwal, and J. Kalagnanam, ‘‘Sparse Gaussian Markov random field mixtures for anomaly detection,’’ in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 955–960.