



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

Dark-TRACER Early Detection Framework for Malware Activity Based on Anomalous Spatiotemporal Patterns

¹M.Sathwika, ²M.Uday Kiran, ³M.Sathwik, ⁴P.Shiva teja, ⁵K. ASHOK BABU,

^{1,2,3,4} U.G.Scholar, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

⁵Professor, Department of ECE, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

Abstract

As cyberattacks become increasingly prevalent globally, there is a need to identify trends in these cyberattacks and take suitable countermeasures quickly. The darknet, an unused IP address space, is relatively conducive to observing and analyzing indiscriminate cyberattacks because of the absence of legitimate communication. Indiscriminate scanning activities by malware to spread their infections often show similar spatiotemporal patterns, and such trends are also observed on the darknet. To address the problem of early detection of malware activities, we focus on anomalous synchronization of spatiotemporal patterns observed in darknet traffic data. Our previous studies proposed algorithms that automatically estimate and detect anomalous spatiotemporal patterns of darknet traffic in real time by employing three Independent machine learning methods. In this study, we integrated the previously proposed methods into a single framework, which we refer to as Dark-TRACER, and conducted quantitative experiments to evaluate its ability to detect these malware activities.

Keywords—Cyberattacks, Darknet, Legitimate communication, Spatiotemporal Patterns, Detection.

I INTRODUCTION

In recent years, an increasingly large number of indiscriminate cyber-attacks have been observed on the Internet, and it is therefore becoming increasingly costly to analyze these attacks. To maintain security of the Internet, it is necessary to quickly recognize global cyber-attack trends, specify their causes, devise countermeasures, and alert the world of the details of the threat. For this purpose, it is important to detect the indiscriminate scanning attack activities caused by malware at an early stage before a particular attack becomes a pandemic.

However, it is challenging to identify malware scanning attacks among the massive amount of benign traffic in regular networks. Therefore, we adopted unused IP address spaces (dark nets). The term "dark net" refers to observation networks, also known as "network telescopes," and should not be confused with

anonymous communication networks such as Tor. In the dark net, legitimate communication (noise) does not occur; therefore, indiscriminate scanning communication (signal) is observed more noticeably. Thus, the signal-to-noise ratio is high. This makes it an effective way to identify trends and tendencies in global cyber attacks.

However, the volume of traffic observed in the dark net is increasing each year exponentially. Moreover, there are many communications whose intentions are unknown, as only the initial communications are observed. For example, in a darknet, we observe numerous independent cyber-attacks occurring simultaneously, as well as many communications that are unrelated to attacks, such as scanning activities that are conducted for benign investigation purposes, communications with unknown causes, and misconfigured communications. As a research target, we should distinguish such noisy communications from malicious attack communications in detail.

Devices infected with similar malware, that is, ones which share scanning modules, tend to scan in a similar spatiotemporal pattern to compromise new infection targets [1]. Such a tendency is also observed on the dark net [2]. Here, the distributions of source hosts and destination ports for packets observed in a certain period are referred to as spatial features. The features observed in the temporal variation of these spatial features are thus referred to as spatiotemporal patterns. The hosts and destination ports that send packets with similar spatiotemporal patterns are then referred to as being synchronized. Even in case of small-scale infection activity of malware, a high degree of synchronicity is expected to occur in the associated spatiotemporal patterns, and early detection of malware activity can be realized by estimating the synchronicity and detecting anomalies. In our previous studies, we focused on such synchronization and attempted to detect potential malware activities by estimating the group of source hosts with high synchronization in their spatiotemporal patterns on a large-scale dark net. We adopted the following three different machine learning methods in this study: *Graphical Lasso* [3], nonnegative matrix factorization (NMF) [4], and nonnegative Tucker decomposition (NTD) [5] to estimate the synchronization of spatiotemporal patterns from packet counts by spatial feature per unit time in dark net traffic data. The *Graphical Lasso* algorithm can sparsely estimate conditionally independent variable pairs that are not synchronous from a covariance matrix. The NMF and NTD algorithms can decompose synchronous latent frequent patterns from data matrices or tensors into superpositions of multiple groups. We previously proposed the following different methods to estimate the synchronization in real time to automatically use the aforementioned algorithms and detect the source host space groups that show abnormal synchronization: *Dark-GLASSO* [6], [7], *Dark-NMF* [8], and *Dark-NTD* [9].

In our previous studies, we confirmed that each method is capable of detecting malware activities well. However, we did not comparatively evaluate the methods and examine their early malware activity detection performance.

In this study, we first modularized the previously proposed methods and integrated common

components such as feature extraction and alert issuing into a single framework. We refer to this integrated framework *Dark-TRACER*. As the main challenge, we conducted two experiments on *Dark-TRACER* one is to evaluate the quantitative detection performance, and the other is to evaluate the feasibility of early detection. In the first experiment, to quantitatively evaluate the detection performance of malware activity, we used the ground truth of reliable malware activity in October 2018, which was manually created, and performed parameter tuning to minimize false negatives and false positives in each module. Although we have previously presented the evaluation results of a conventional method *Change Finder* [10] and the proposed modules *Dark-GLASSO* and *Dark-NMF*, we evaluate *Dark-NTD* for the first time using the same criteria. In the second experiment, we manually generated a new ground truth of events (from June 2019 to October 2020) that clearly shows the time of infection spread of malware activities and used it to evaluate the feasibility of the proposed framework for early detection.

As a result, *Dark-GLASSO*, *Dark-NMF*, and *Dark-NTD* achieved 97.1%, 100%, and 97.1% recall, respectively. We also identified the pros and cons of each module and found that the integration of all the proposed modules into a single framework, *Dark-TRACER*, complemented each individual

module's weaknesses. In addition, the results of the early detection feasibility evaluation show that *Dark-TRACER* can detect threats 153.6 days earlier than when the threats were revealed to the public by reputable third-party security research organizations. We also assessed the human analysis cost and found that daily operation with two analysts could be performed in an average of 7.3 h, assuming that one analyst requires 15 min of analysis time per port.

In summary, this study afforded the following contributions:

- We integrated our three prior methods (modules) into a single framework, *Dark-TRACER*. To the best of our knowledge, our approach is the first method that focuses on the synchronization of spatiotemporal patterns of darknet traffic. *Dark-TRACER* can detect malware activities that show anomalous synchronization.
- This work is also the most advanced practical study that quantitatively evaluated the detection performance of malware activities and the feasibility of early detection.
- We found that *Dark-TRACER* complements the weaknesses of each module, and achieves a 100% recall rate. In addition, the results demonstrate that *Dark-TRACER* detects threats on average 153.6 days earlier than when the threats are revealed to the public. We also demonstrated that two analysts can conduct the necessary daily operations of the framework in approximately 7.3 h.

Currently, *Dark-TRACER* is being implemented in real world contexts for actual operation. It is expected to provide information on detected global malware activities to organizations such as the

Computer Security Incident Response Team (CSIRT) and the Security Operation Center (SOC), and to assist in their ability to implement prompt countermeasures such as investigating the causes and conducting detailed analysis.

II LITERATURE SURVEY

"Malware Detection Techniques: A Comprehensive Review"

This literature survey provides an extensive review of existing malware detection techniques. It covers traditional signature-based methods, anomaly detection, behavior-based approaches, and machine learning algorithms. The survey serves as a foundational resource for understanding the landscape of malware detection techniques and their limitations.

"Anomaly Detection in Cybersecurity: Approaches and Challenges"

In this survey, we focus on anomaly detection techniques in cybersecurity. It explores the methodologies, algorithms, and challenges associated with detecting anomalous behavior in networks and systems. The survey provides insights into the applicability of anomaly detection in identifying potential malware activities.

"Spatiotemporal Analysis in Cyber Threat Detection"

This survey delves into spatiotemporal analysis techniques in cyber threat detection. It discusses how spatial and temporal data are utilized to detect patterns and anomalies in network traffic and system behavior. The survey reviews the effectiveness of spatiotemporal analysis in identifying potential malware activity.

"Early Detection Frameworks for Emerging Cyber Threats."

In this survey, we explore early detection frameworks designed for emerging cyber threats. It investigates proactive measures and predictive models aimed at identifying novel malware activities and potential threats in their early stages. The survey discusses the role of early detection in mitigating cybersecurity risks.

III EXISTING SYSTEM

Dainotti *et al.* contributed to a census-like analysis of how the IP address space is used by developing malware and evaluating methods to remove spoofed traffic from darknets and live networks [49]. Durumeric *et al.* analyzed a large-scale darknet to investigate Internet-wide scanning activities and identify patterns of extensive horizontal scanning operations

Fachkha *et al.* devised an inference and characterization module to identify and analyze the probing activities of cyberphysical systems (CPS) by extracting various features from large amounts of darknet data and performing correlational analyses. Jonker *et al.* introduced a

framework to protect against DoS attacks based on various data sources, including darknet traffic data . They found that one-third of all /24 networks on the Internet had suffered at least one DoS attack in the past two years. Shaikh *et al.* identified unsolicited IoT devices by collecting IP header information from darknet traffic data and classifying them using several machine learning algorithms

Akiyoshi *et al.* proposed a method to detect emerging scanning activities and their scale by analyzing the correlation between traffic in honeypots and darknets . Most of the measurement analysis studies using darknets have been applied to understand the general trend of malicious communications observed in darknets. Thus, for detailed analysis, many studies use not only darknet data but also trap- based monitoring systems such as honeypots

IV PROBLEMSTATEMENT

In the realm of cybersecurity, the identification and mitigation of malware threats pose a significant challenge due to the evolving nature of malicious activities. Current detection systems often struggle to identify novel malware strains and sophisticated attack patterns in their early stages. The lack of proactive measures to detect and counteract malware activities results in potential data breaches, system vulnerabilities, and compromised network integrity. Thus, there is a critical need for an advanced detection framework that can effectively identify and predict anomalous spatiotemporal patterns associated with emerging malware threats, enabling early detection and proactive defense mechanisms to safeguard digital systems and networks

V PROPOSEDSYSTEM

We integrated our three prior methods (modules) into a single framework, Dark-TRACER. To the best of our knowledge, our approach is the first method that focuses on the synchronization of spatiotemporal patterns of the darknet traffic. Dark-TRACER can detect malware activities that show anomalous synchronization.

This work is also the most advanced practical study that quantitatively evaluated the detection performance of malware activities and the feasibility of early detection.

We found that Dark-TRACER complements the weaknesses of each module, and achieves a 100% recall rate. In addition, the results demonstrate that Dark- TRACER detects threats on average 153.6 days earlier than when the threats are revealed to the public. We also demonstrated that two analysts can conduct the necessary daily operations of the framework in approximately 7.3 h

VI IMPLEMENTATION

ServiceProvider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train and Test Data Sets, View Datasets Trained and Tested Accuracy in Bar Chart, View Datasets Trained and Tested Accuracy Results, View Malware Activity Predicted Details, Find Malware Detection Type Predicted Ratio, Download Predicted Datasets, View Malware Detection and Predicted Ratio Results, View All Remote Users.

ViewAndAuthorizeUsers

In this module, the admin can view the list of users who are all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

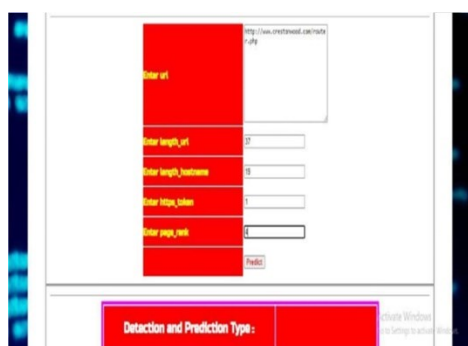
RemoteUsers

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Register And Login, Predict Malware Detection Type, View Your Profile.



VII RESULTS





In this study, we introduced three independent machine learning methods to automatically estimate the synchronization of the spatiotemporal patterns of dark net traffic in real time and to detect anomalies. Those three methods are: Dark-GLASSO, Dark-NMF, and Dark-NTD. We also proposed Dark-TRACER, which integrates all three methods into a single framework. We found that *Dark-TRACER* was able to complement the weaknesses of each module, achieving a 100% recall rate and detecting all malware activities in the experiment. It detected the malware on average 153.6 days earlier than the time when the threats were revealed to the public by reputable third-party security research organizations. In addition, we found that two analysts could perform the daily operations necessary to detect these threats in approximately 7.3 h. Currently, our most serious challenge is the large number of false positives. In this study, we confirmed that even a simple rule-based approach can effectively reduce the number of false-positive alerts. As described in Sections VD and VI-C, our future work is to reduce the number of false

positives by identifying the fingerprints of investigative scanners and building a model to track them. By reducing the number of false positives, the analysis cost can be lowered. In addition, we intend to automate the secondary collision analysis mentioned in Section V-E to elucidate the causes and details of the alerts detected by Dark-TRACER. Finally, we plan to deploy Dark-TRACER in the real world and detect threats and malware activities in real-time to aid a rapid response.

REFERENCES

- [1] 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.
- [2] 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.
- [3] C. Han, J. Shimamura, T. Takahashi, D. Inoue, M. Kawakita, J. Takeuchi, and K. Nakao, "Real-time detection of malware activities by analyzing darknet traffic using graphical lasso," in *Proc. 18th IEEE Int. Conf. Trust, Security. Privacy Comput. Commun. (TrustCom)*, Aug. 2019, pp. 144_151.
- [4] 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.
- [5] 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.
- [6] J. Gibberd and J. D. B. Nelson, "High dimensional changepoint detection with a dynamic graphical lasso," in *Proc. IEEE Int. Conf.*
- [7] D. Cohen, Y. Mirsky, M. Kamp, T. Martin, Y. Elovici, R. Puzis, and Shabtai, "DANTE: A framework for mining and monitoring darknet traffic," in *Proc. 25th Eur. Symp. Res. Comput. Secur. (ESORICS)*. Springer, 2020, pp. 88_109.