Dark –Tracer: Early Detection Framework For Malware Activity Based On Anomalous Spatiotemporal Patterns

Mrs. K. Ramya Laxmi[1*], Vasamsetti Soumya[2], Veerla Sri Nikitha[3], Gottala Sreekar[4], Dharavath Saikiran Nayak[5]

[1*]Assistant professor, Dept of CSE, Sreyas Institute of Engineering and Technology, ramya.kunta@sreyas.ac.in
[2]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[3]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[4]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.
[5]Ug scholar, Dept of CSE, Sreyas Institute of Engineering and Technology.

***Corresponding Author:** Mrs. K. Ramya Laxmi
*Assistant professor**,** Dept of CSE, Sreyas Institute of Engineering and Technology, ramya.kunta@sreyas.ac.in

Abstract

As cyber attacks become increasingly prevalent globally, there is a need to identify trends in these cyber attacks and take suitable countermeasures quickly. The darknet, an unused IP address space, is relatively conducive to observing and analyzing indiscriminate cyber attacks 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. We used darknet traffic data from October 2018 to October 2020 observed in our large-scale darknet sensors (up to /17 subnet scales). The results demonstrate that the weaknesses of the methods complement each other, and the proposed framework achieves an overall 100% recall rate. In addition, Dark-TRACER detects the average of malware activities 153.6 days earlier than when those malware activities are revealed to the public by reputable third-party security research organizations. Finally, we evaluated the cost of human analysis to implement the proposed system and demonstrated that two analysts can perform the daily operations necessary to operate the framework in approximately 7.3 h.

Keywords: Anomalous synchronization estimation, darknet, malware activity, spatiotemporal pattern.

INTRODUCTION

In recent years, an increasingly large number of indiscriminate cyberattacks 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 cyberattack 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 is regular networks. Therefore, we adopted unused IP address spaces (dark nets). The term “darknet” 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 mis configured 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. Such a tendency is also observed on the dark net. Here, the distributions of source hosts and destination ports for packets observed in a temporal variation of these 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-case 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.

LITERATURE SURVEY

In our pervious 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 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 super positions 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 Dark-NMF, and Dark-NTD.

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 frame work. 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 detection performance of malware activity, we used the ground truth reliable malware activity in october2018, 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 and the proposal modules Dark-GLASSO and Dark-NMF, we evaluate Dark-NTD for the rest time using the same criteria. In the second experiment, we manually generated a new ground truth of events 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 modules weakness. In addition the results of the early detection feasibility evaluation show that Dark-TRACER can detect threats 153.6days 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 average of 7.3 h, assuming that one analyst requires 15 min of analysis time per port.

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. 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. To detect the indiscriminate scanning attack activities caused by malware at an early stage before a particular attack becomes a pandemic. 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.

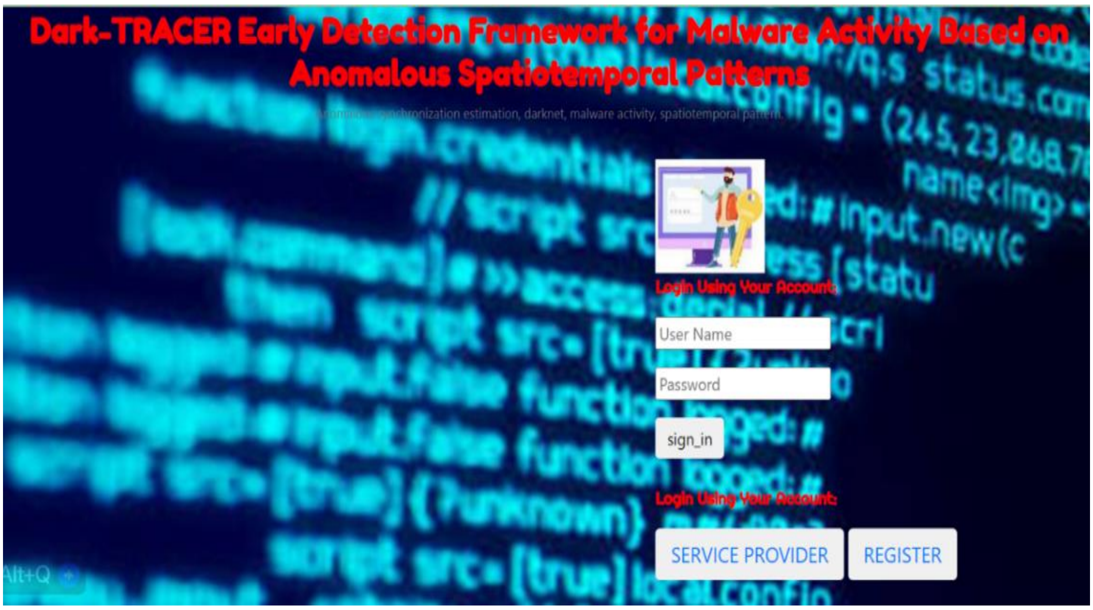
Dainotti *et al.* contributed to a census-like analysis of how the IP address space is used by developing malware and evaluating methods to remove spoofed traffic from darknets and live networks. Durumeric *et al.* analyzed a large-scale darknet to investigate Internet-wide scanning activities and identify 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. The system is not implemented a large-scale darknet observation system, the NICTER project and which aims to understand global trends in indiscriminate cyber attacks. An existing system attempted to detect potential malware activities by estimating the group of source hosts with high synchronization in their spatiotemporal patterns on a large-scale darknet.

PROPOSED SYSTEM CONFIGURATION

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. advantages of proposed system are The proposed system can reduce the effect of benign noise communication in the darknet traffic and highlight the malicious communication. In addition, malware activities that are difficult to trace by conventional manual operations, such as threats that are small-scale, orchestrated, or have no visible explicit spikes, can be captured before the malware infection becomes widespread by detecting anomalously synchronized spatial features. Finally, if a malware activity is found to be synchronized with other malware activities at a time when the scale of infection is small (i.e., before it spreads in earnest), it can be detected at that early stage.

User needs to register or sign-in first by filling out the required details:



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

Anomalous synchronization estimation, darknet, malware activity, spatiotemporal pattern.

Login Using Your Account:

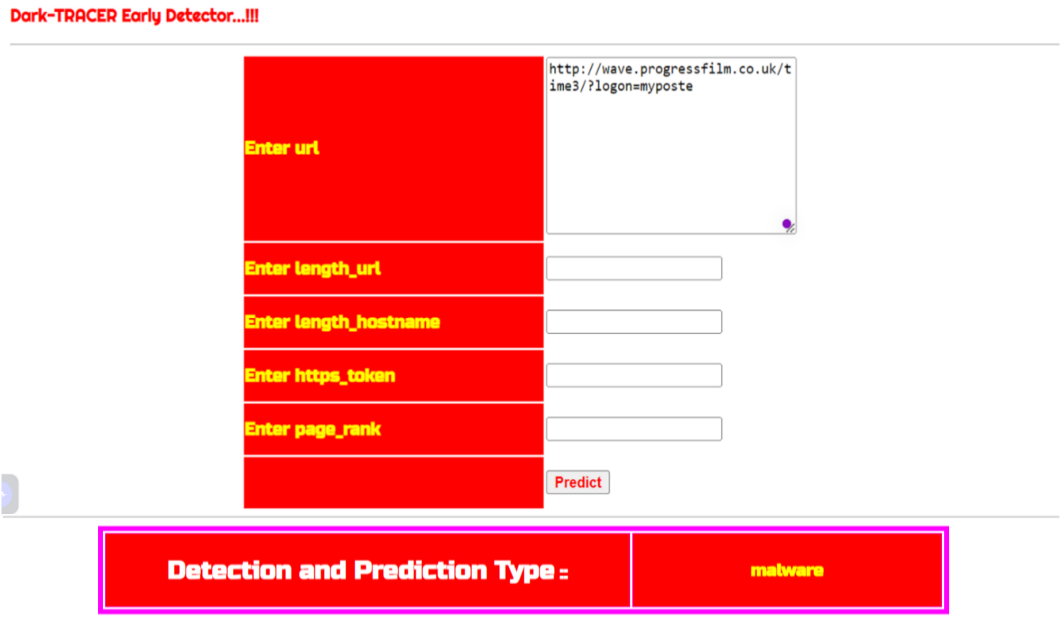
User Name

Password

Login Using Your Account:

Fig.1 User needs to register or sign-in first by filling out the required details

After signing in enter the URL of the website user needs to check as shown in below figure:



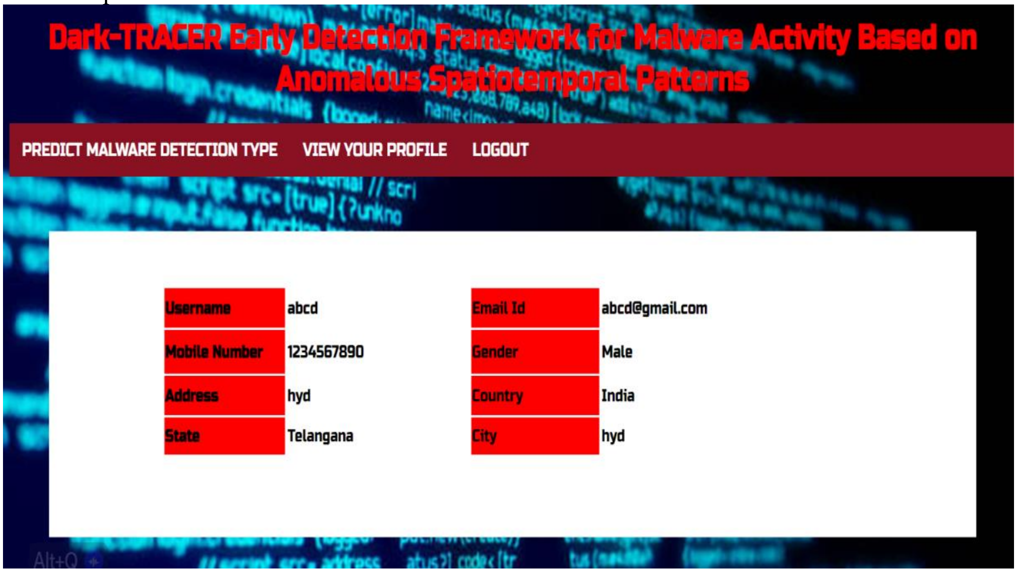
Dark-TRACER Early Detector...!!!

Enter url	http://wave.progressfilm.co.uk/time3/?logon=myposte
Enter length_url	
Enter length_hostname	
Enter https_token	
Enter page_rank	
	Predict

Detection and Prediction Type :	malware

Fig.2 After signing in enter the URL of the website user needs to check as shown

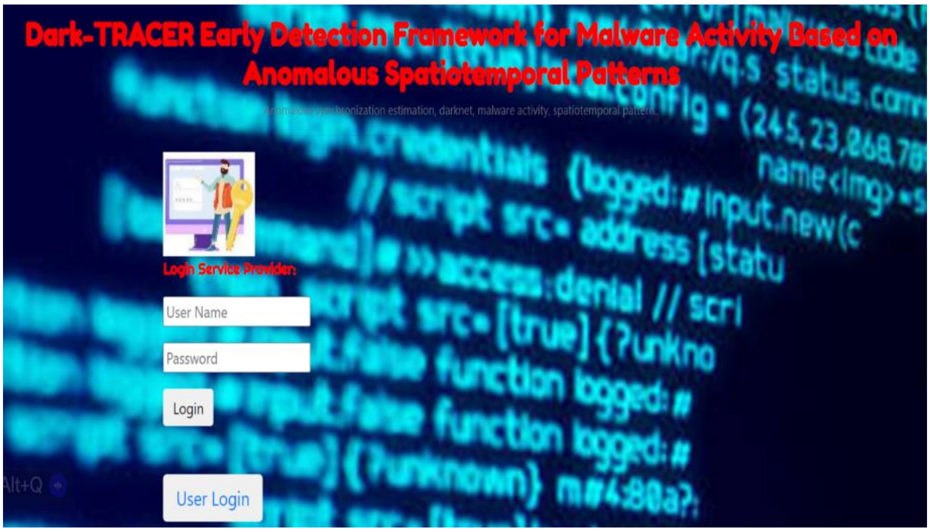
User can view his/her profile:



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

Username	abcd	Email Id	abcd@gmail.com
Mobile Number	1234567890	Gender	Male
Address	hyd	Country	India
State	Telangana	City	hyd

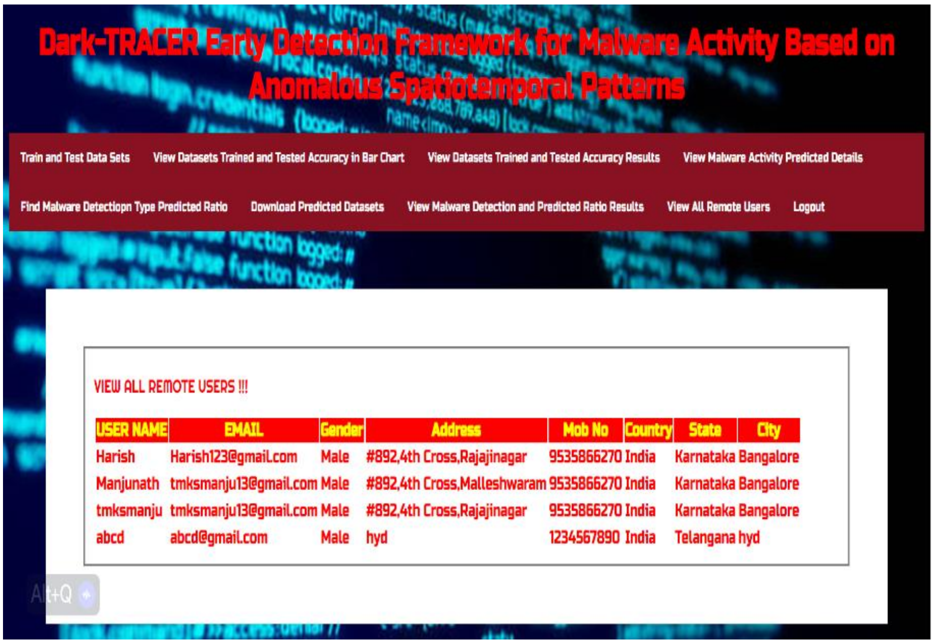
Fig.3 User can view his/her profile



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

Anomalous synchronization estimation, darknet, malware activity, spatiotemporal pattern.

Fig 4. Service Provider can login through the project homepage as shown



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

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 Detectiopn Type Predicted Ratio
Download Predicted Datasets
View Malware Detection and Predicted Ratio Results
View All Remote Users
Logout
Alt+Q

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
Harish	Harish123@gmail.com	Male	#892,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	Male	#892,4th Cross,Malleshwaram	9535866270	India	Karnataka	Bangalore
tmksmanju	tmksmanju13@gmail.com	Male	#892,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
abcd	abcd@gmail.com	Male	hyd	1234567890	India	Telangana	hyd

Fig 5. Service Provider can View all remote users

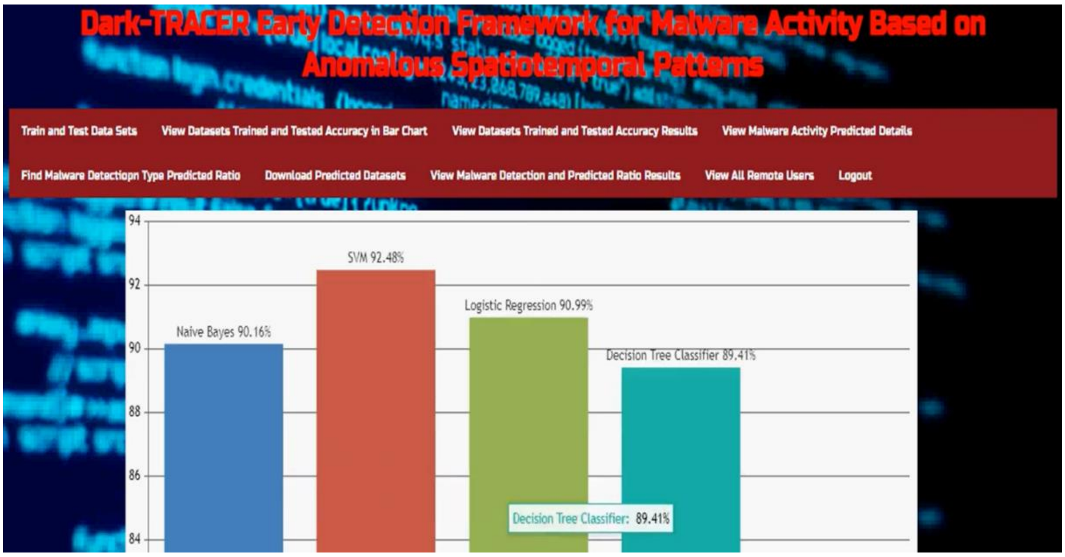


Fig 6. Service Provider can View accuracy of datasets trained and tested in bar chart

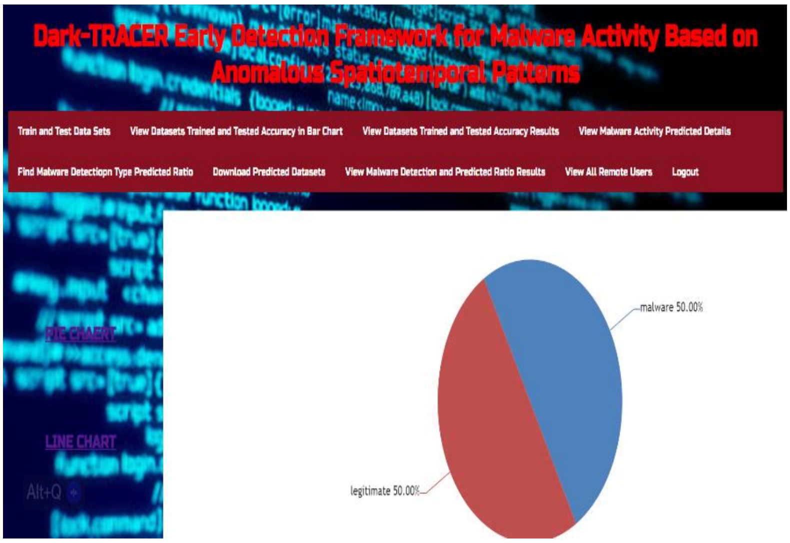
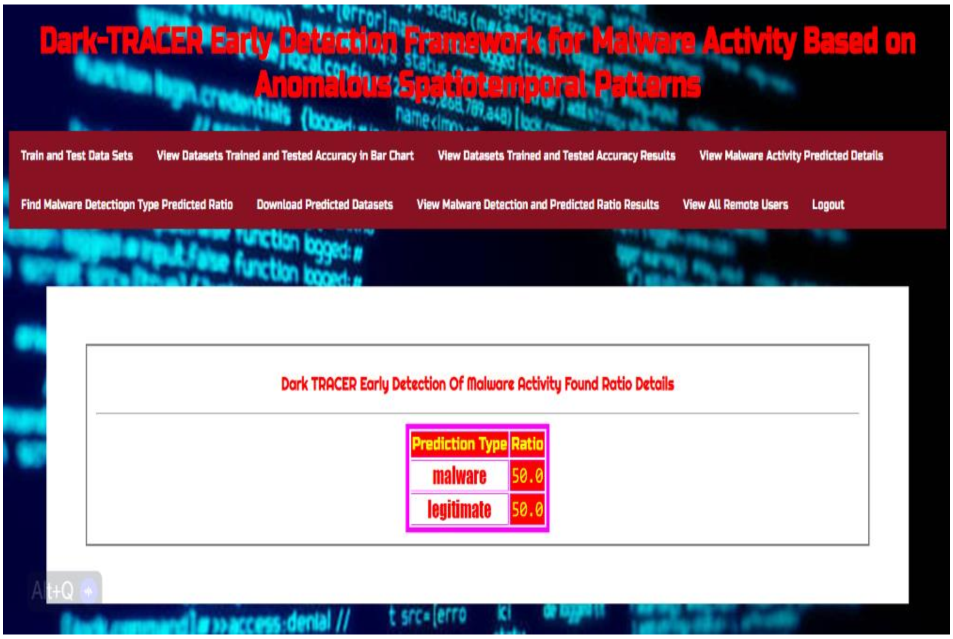


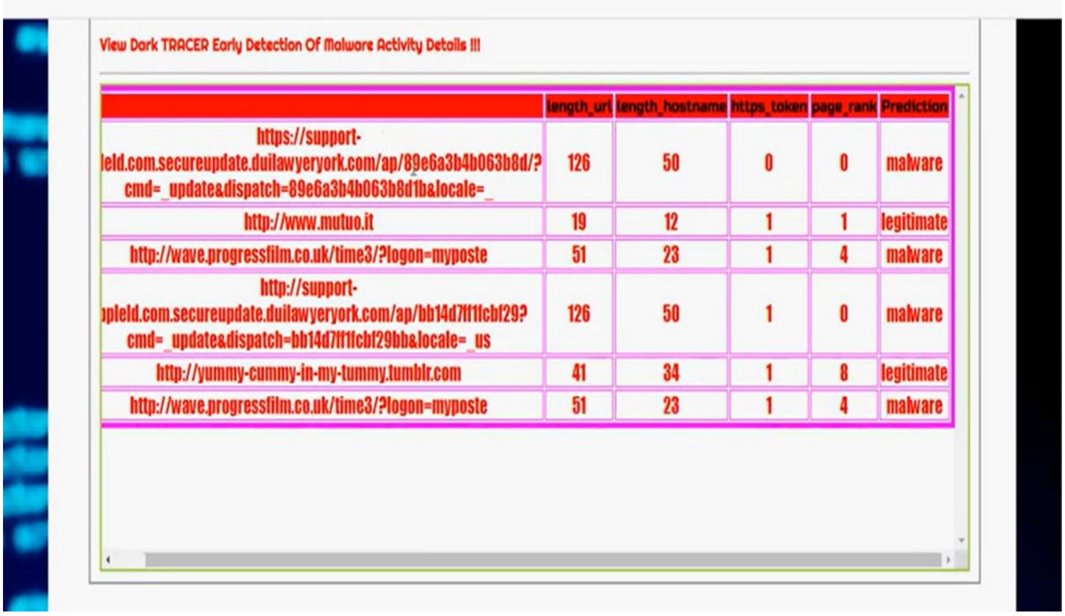
Fig 7. Viewing malware predicted and detected ration as shown


Dark-TRACER Early Detection Framework for Malware Activity Based on Anomalous Spatiotemporal Patterns
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 Detectiopn Type Predicted Ratio
Download Predicted Datasets
View Malware Detection and Predicted Ratio Results
View All Remote Users
Logout
Alt+Q

Dark TRACER Early Detection Of Malware Activity Found Ratio Details

Prediction Type	Ratio
malware	50.0
legitimate	50.0

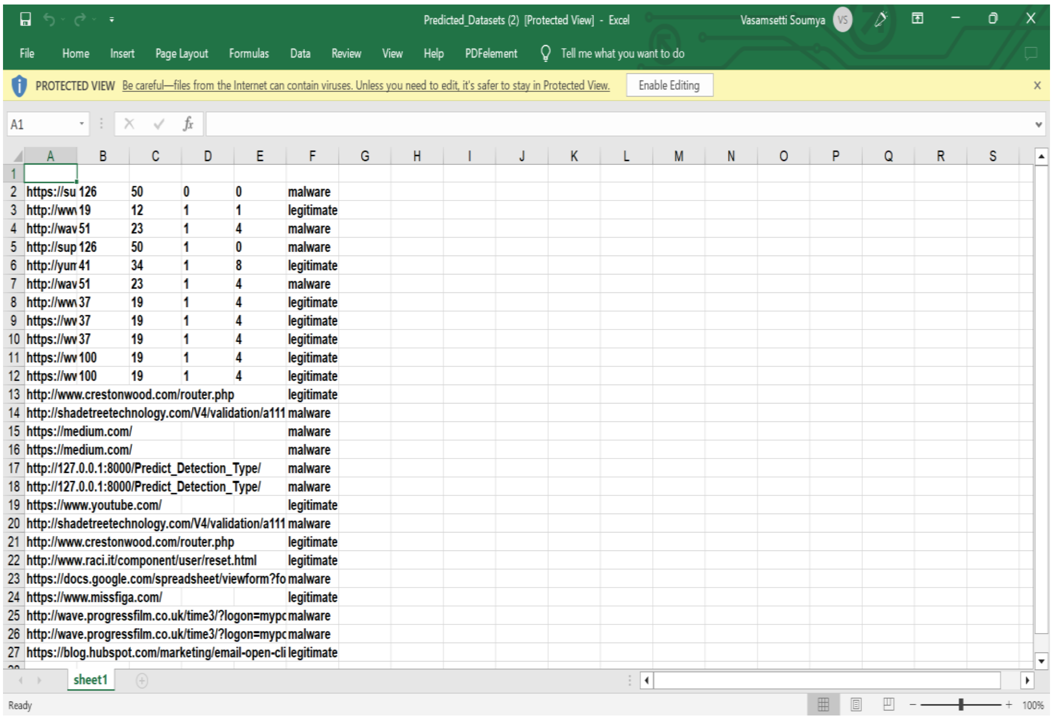
Fig 8. Screenshot Of Results



View Dark TRACER Early Detection Of Malware Activity Details !!!

	length_url	length_hostname	https_token	page_rank	Prediction
https://support- leld.com.secureupdate.duilawyeryork.com/ap/89e6a3b4b063b8d/? cmd=_update&dispatch=89e6a3b4b063b8d1b&locale=_	126	50	0	0	malware
http://www.mutuo.it	19	12	1	1	legitimate
http://wave.progressfilm.co.uk/time3/?logon=myposte	51	23	1	4	malware
http://support- ppleld.com.secureupdate.duilawyeryork.com/ap/bb14d7ff1fcbf29? cmd=_update&dispatch=bb14d7ff1fcbf29bb&locale=_us	126	50	1	0	malware
http://yummy-cummy-in-my-tummy.tumblr.com	41	34	1	8	legitimate
http://wave.progressfilm.co.uk/time3/?logon=myposte	51	23	1	4	malware

Fig 9. It will Detected Malware activity details are displayed as shown



PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

	A	B	C	D	E	F
1						
2	https://su	126	50	0	0	malware
3	http://ww	19	12	1	1	legitimate
4	http://wav	51	23	1	4	malware
5	http://sup	126	50	1	0	malware
6	http://yun	41	34	1	8	legitimate
7	http://wav	51	23	1	4	malware
8	http://ww	37	19	1	4	legitimate
9	https://wv	37	19	1	4	legitimate
10	https://wv	37	19	1	4	legitimate
11	https://wv	100	19	1	4	legitimate
12	https://wv	100	19	1	4	legitimate
13	http://www.crestonwood.com/router.php					legitimate
14	http://shadetreetechnology.com/V4/validation/a111					malware
15	https://medium.com/					malware
16	https://medium.com/					malware
17	http://127.0.0.1:8000/Predict_Detection_Type/					malware
18	http://127.0.0.1:8000/Predict_Detection_Type/					malware
19	https://www.youtube.com/					legitimate
20	http://shadetreetechnology.com/V4/validation/a111					malware
21	http://www.crestonwood.com/router.php					legitimate
22	http://www.raci.it/component/user/reset.html					legitimate
23	https://docs.google.com/spreadsheet/viewform?fo					malware
24	https://www.missfiga.com/					legitimate
25	http://wave.progressfilm.co.uk/time3/?logon=mypc					malware
26	http://wave.progressfilm.co.uk/time3/?logon=mypc					malware
27	https://blog.hubspot.com/marketing/email-open-cli					legitimate

Fig 10. Predicted datasets can also be downloaded and viewed as shown

CONCLUSION

In this study, we introduced three independent machine learning methods to automatically estimate the synchronization of the spatiotemporal patterns of darknet 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. malware activity based on anomalous spatiotemporal patterns shows promise in improving the security posture of organizations by providing proactive and accurate detection of malware threats. Continued research and development in this field can contribute to the advancement of cybersecurity practices, ensuring better protection against evolving malware attacks.

REFERENCES

1. G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting botnet command and control channels in network traffic," 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 traffic 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.
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.
13. J. Gibberd and J. D. B. Nelson, "High dimensional changepoint detection with a dynamic graphical lasso," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2014, pp. 2684–2688.
14. T. Idé, A. C. Lozano, N. Abe, and Y. Liu, "Proximity-based anomaly detection using sparse structure learning," in Proc. SIAM Int. Conf. Data Mining, Apr. 2009, pp. 97–108.
15. S. Liu, T. Suzuki, and M. Sugiyama, "Support consistency of direct sparse-change learning in Markov networks," in Proc. 29th AAAI Conf. Artif. Intell., 2015, pp. 2785–2791.
16. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," IEEE Comput., vol. 42, no. 8, pp. 30–37, Aug. 2009.
17. Q. Zhao, C. F. Caiafa, D. P. Mandic, L. Zhang, T. Ball, A. Schulze-Bonhage, and A. and Cichocki, "Multilinear subspace regression: An orthogonal tensor decomposition approach," in Proc. 25th Annu. Conf. Neural Inf. Process. Syst., 2011, pp. 1269–1277.
18. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," IEICE Nonlinear Theory Appl., vol. 1, no. 1, pp. 37–68, 2010.
19. Anandkumar, P. Jain, Y. Shi, and U. N. Niranjan, "Tensor vs. matrix methods: Robust tensor decomposition under block sparse perturbations," in Proc. 19th Int. Conf. Artif. Intell. Statist., (AISTATS), vol. 51, 2016, pp. 268–276.
20. F. Caiafa and A. Cichocki, "Generalizing the column–row matrix decomposition to multi-way arrays," Linear Algebra its Appl., vol. 433, no. 3, pp. 557–573, Sep. 2010.
21. G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Efficient nonnegative tucker decompositions: Algorithms and uniqueness," IEEE Trans. Image Process., vol. 24, no. 12, pp. 4990–5003, Dec. 2015.
22. M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. and Zhou, "Understanding the Mirai botnet," in Proc. 26th USENIX Secur. Symp., 2017, pp. 1093–1110.
23. H. Griffioen and C. Doerr, "Examining Mirai's battle over the Internet of Things," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., Oct. 2020, pp. 743–756.