IJREAT International Journal of Research in Engineering& Advanced Technology, Volume 8, Issue 4, Aug - Sep, 2020
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

IDENTIFYING MALICIOUS ACTIVITIES BY MONITORING DARKNET ACCESS

[1]Preeti S. Joshi, Department of IT, MMCOE, Pune, Maharashtra, preetijoshi@mmcoe.edu.in

[2]Dinesha H.A., CSE Department, SGBIT, Belagavi, India, dineshameet@gmail.com

***Abstract*—**The evolutionary types of attacks generated by attacker makes it difficult to identify them. Novel tools and techniques are required to detect, analyze and generate reports for threat intelligence. Sensors placed in unused or unassigned IP addresses and used to monitor traffic passively are called Darknets. Attempt to access darknet addresses can be classified as malicious activity. Monitoring traffic reaching darknet addresses can generate cyber intelligence indicating activities going onto the Internet. They are looked as a means for early detection of cyber-attacks. In this paper we present an overview of recent work done in the field of darknet monitoring. We try to bring forth the limitations and challenges related to identification of malicious traffic by passive monitoring of traffic. We compile the suggestions for improvement.

Keywords: Darknet Monitoring Malicious activity, Traffic analysis

I. INTRODUCTION

Cyber criminals try to exploit security vulnerabilities onto networked devices in an effort to replicate, steal, modify, or destroy a specified target for an attack. To identify vulnerable computers, their malicious code may randomly pick any random IP address as target address, or may search across one or more IP address ranges. The first step of any cyber-attack is reconnaissance. In this step the attacker identifies the vulnerabilities in the target system and exploits these vulnerabilities to gain access to the system. The reconnaissance step involves scanning the system or network. If this scanning activity is identified before actual attack takes place, it can prevent further attack to the system or lessen the impact.

Darknet or Network Telescopes are the portion of internet with unassigned or unused IP addresses. Onto these addresses no service or applications are hosted. So traffic targeting darknet addresses are either through misconfiguration or are malicious with intension of cyber-attacks. Darknets have been used for passive monitoring of network traffic and to identify cyber threats. Usually darknets receive traffic only for one of three reasons: traffic sent accidentally/by mistake, backscatter, malicious activity of scanning and worms. Accidental Darknet requests can occur if an individual mistyped an IP address or the URL used had an incorrect Domain Name Service (DNS) entry leading to a darknet. To hide identity of sender, DoS attacks are launched using Spoofed IP address. An acknowledgement generated for completing the three way handshake by victim to these initial TCP packet of DoS is called Backscatter. In comparison to viruses computer worms can propagate automatically without human initiation. Software worms are capable of self-replicating as well asspreading unassisted. This means that worms will make several copies of themselves to send malicious activities to other computers, as per the attacker's script.

Various research studies exist with different deployment schemes, different ways to analyse the captured traffic, find or classify the malicious activities as scanning or probing, DoS, Worms. Some research implementations have used traffic to darknets for visualization, since humans cannot interpret raw network traffic with ease. Visualization techniques also aid in presenting the analysis of captured traffic.

The structure of paper is as follows. Section 2 presents the study report of existing Systems, methods and apparatus for a distributed security that monitors communications to identify access attempts todarknet addresses. Section 3 discusses the challenges and limitations of the systems. Section 4 concludes the paper.

II. REVIEW OF EXISTING SYSTEM

A. Evaluation of Darknet based on Size

Darknets are subnets in IPV4 network that are monitoring traffic for malicious activities on the Internet. Use of IPV6 has increased in few years. Google reported IPV6 adoption by its users to be approximately 1% in 2013, while it is 31.73% in May 2020. Darknet monitoring on IPV6 may not be able to capture such activity, because an attacker will not be able to scan entire IPV6 network due to its size. Also traffic monitoring in IPV6 will be difficult due to encryption. A critical evaluation of capability of darknet based on its size, in addressing threat intelligence is studied in literature. In a research carried out in 2013, darknet on IPV4 and IPV6 address space, it showed that IPV4 reveals more information of scanning activity, while IPV6 showed only packets coming due to misconfiguration [9].

Visibility of darknets varies according to the IP range that is size of the darknets[14], [20]. Using Jaccard similarity index it was found that source of traffic varies significantly according to IP ranges and size of darknet impacts its visibility [20]. In [14] authors advocate the accuracy of small sized darknets for use in organizations for producing threat intelligence.

TABLE I REVIEW OF STUDY OF DARKNET SIZE

Publication	Addressing	Size	Conclusion
[14]	IPV4	Subnets of size /24 in 146/8 and 155/8 from South Africa's University	Feasibility of small sized network for collecting IBR collection
[20]	IPV4	Subnet size /19 in Brazil, /15Network in Netherlands, /24 Network in Italy	Sources of traffic significantly varies according to the IP range, and the size of the darknet impacts its visibility

TABLE II. REVIEW OF SCAN DETECTION

Publication	Method/Tools	Objective	Dataset	Contribution
[21]	Topological data Analysis, DBSCAN	Visualization	/20 Network	Pattern of malicious event and Visualization
[13]	IP Scan, Port Scan, Hybrid Scan	Traffic flow analysis method for grouping malicious events	/20 Network	Malicious event grouping
[2]	ML algorithms, Random Forest, Light GBM	Framework for threat Detection	SURFnet	Patterns of advanced threat
[1]	Transition graphs, AR and VAR model	Analysis of targeted service, modelling probing activity ,geo-location of source, Prediction of probing rate	/20 Network	Analysis of probing and rate detection
[6]	Correlate the flows observed at a low-interaction honeypot with packets observed in the darknet	Vulnerability scanning using honeypot and Darknet	Twelve /24 networks	Attack scale identification

A. Scanning, DoS and Backscatter

Darknet were used to observe patterns in traffic received that are missed by IDS suricata in [21] and to find malicious events by analysing traffic flow in[13]. In [2] a framework for threat detection is presented. Traffic from darknet is use to train a machine learning classifier and adjust performance based on drift in performance from the threshold due to any used parameter.

Multiple malwares randomly scan the network locate a target for attack. Current research carried out on detection of scanning activity captured on darknet presents some meaningful statistics. Paper [1] identifies the most targeted port, service targeted by top network probers, model of probing pattern using transition graph and prediction of probing rates by AR and VAR models. The method used to find top network probers does not consider distribution of probing activity. This way it might have missed the probes sent by bots. The paper also presents the probing activity carried out country wise, but this method of finding origin of probing may not be correct, since attacker may have used botnets using spoofed address. DoS attacks onto Internet are quantified by the backscatter traffic (TCP,ICMP,and UDP) collected at the darknet in[11].The 'source origin test' using Gaussian distribution notifies if the spoofed address is picked randomly. DoS attack by DNS amplification, causes root level or Top level Domain to send a reply to a victim to overwhelm it. The query type used is ANY leading to all possible information sent by the DNS server. To infer large-scale DNS-based DRDoS activities, flow generation, rate classification and clustering approach is used in [12] instead of backscatter data.

The study of work done in darknet monitoring helps to understand the trend of the ongoing scan, DoS, backscatter attack features along with the tools and techniques used till now and provides the capability to diagnose correctly and respond suitably.

B. Coordination and Cooperation of Bots

The coordinated and cooperated activity for scanning or sending malware by means of botnet with command and control server (C2) is studied by researchers in [3], [4]. [3] Presents a real time algorithm to generate alert for coordinated and cooperated scanning. The traffic observed is categorised as cyber-attack, survey scans and Sporadically-focused traffic. These cyber-attacks are further classified as IoT malwares as Mirai, Hajime, and HNS. In [4] authors use NTD method of tensorflow factorization method for extracting co-occurrence pattern. The components considered for finding coordination are time series, source IP and destination port. A multilevel analysis, host level and group level and using SVM machine learning algorithms used to identify DDoS and botnet coordinated activity by association rule mining [7]. Authors suggest that as the number of IP addresses in darknet increases it can gather essential information for malware identification. A

measurement and analysis of scan activity by sality botnet is presented in [16]using Hilberts curve. Geolocation of such botnets are identified by MaxMind

TABLE III. REVIEW OF DoS ATTACK DETECTION

Publication	Method	Objective	Dataset	Contribution
[12]	Low generation, K-means clustering	Inferring significant DNS amplification DRDoS activities	/13 network	Identify DRDoS attack from DNS amplification.
[11]	Statistical approach	Quantify TCP,ICMP and UDP backscatter	/16 Network	Detection of DoS through Backscatter in Darknet
[18]	Neural Network-RAN-LSH Classifier	Detection of Backscatter generated from DDoS	NICT Japan	Detection of Backscatter generated from DDoS

TABLE IV. REVIEW OF BOT DETECTION

Publication	Real Time	Coordination of Bots	Malware Detected	Alert	Dataset
[3]	Yes	Sparse structure learning algorithm - graphical lasso	Mirai, Hajime, HNS	Timestamp, targeted destination TCP port, Source IP addresses & number of addresses	https://csdataset.nict.go.jp/darknet
[4]	Yes	Tensor factorization and memory-efficient NTD by utilizing two methods, LRA-NTD and FSTD.	-	Packet timestamps, Source IP addresses & Destination port number	5 sensors over 35000 IP addresses
[7]	-	Association rule mining for botnet coordination	Carna	DDoS attack	NICTER Project with 30 sensors
[16]	-	Hilbert-curve map Visualization	Sality	Geo-location of bots, Fingerprinting of source OS	/8 UCSD Network
[22]	Yes	Time series Analysis and Clustering	-	Fingerprinting of source OS	/13 Networks sensors

GeoLite database and p0f is used as fingerprinting tool for identification of operating system on source address.
Attacks held by botnets campaign are ever changing the pattern of scan & attack. Early detection of such botnet activities through darknet monitoring helps to reduce the damage. Due to varying signature of malwares, identification at IDS or firewall becomes challenging.

C. Honeypots and Darknets

Finding malicious traffic from dark is difficult task as the data generated may be too huge to handle. Another difficulty is that even though packets do reach the darknet there is no interaction from darknet to outside internet. Packets coming to darknet are only initial packets like TCP –SYN, UDP or ICMP or backscatter. Information that can be deduced from the packets carrying payloads is not available. Honeypots are security systems that are deliberately placed onto network attract attacker to investigate unauthorized accesses in order to discover potential vulnerabilities in operational systems, and reduce the risks. Using honeypots along with darknets can help gain more insight into malicious transmissions reaching darknet. A wide range of honeypots are discussed in [5] with the taxonomy and performance. Traffic collected by Lurker, a low interaction honeypot, Sis correlated with traffic on darknet to estimate the scale of attack [6]. The low interaction honeypot responds only to TCP SYN and ICMP packets and acquires the first payload after a three way handshake.

D. New trends- IoT and Darknets

Technology like Internet of Things connects various other devices apart from computing machines. Security is less thought of topic in IoT devices leading to these devices being used in malicious activities of working as bots along with launching DDoS attack, sending backscatter. Newer characteristics of scanning the network are seen and they differ earlier methods.With growing penetration of IoT

devices on Internet, a scalable approach for detecting maliciousness of IoT devices is required. Darknets proves to be useful tool in identifying the malicious traffic and infected IoT devices. Based on 5 TB data collected at darknet, compromised IoT devices and those targeted by DoS are characterized in [10]. In this work, identification of IOT devices is done by a search engine of IoT devices 'shodan'.

TABLE V. DETECTION OF IoT MALWARE FROM DARKNET DATA

Publication	Dataset	Identification of compromised IoT devices	Malware identified	Campaign Detection	Contribution
[15]	/8 Darknet data from CAIDA	CNN Algorithm	Mirai, Fbot, xmrMiner	Hierarchical agglomerative clustering	Fingerprinting of malicious IoT botnets, demonstrate evolving IoT botnets with cryptojacking capabilities,
[10]	/8 Darknet data from CAIDA	Shodan	Ramnit, Starman Kryptik, Nivdort Razy,Zusy,BayrodArtemis,MSIL Vupa, Allaple	correlation algorithm	Backscatter generated by DDoS attack, Malware activity & scanning by IoT devices, IoT device information
[17]	CAIDA	Shodan	Mirai, Satori, Fbot,ADB.Miner, Lightaidra, Tsunami, Gafgyt-A	DBSCAN	Evolution of IoT-tailored malware/botnet
[19]	CAIDA	Censys, Shodan	-	-	Classification –Scanning, Backscatter, Misconfigration

The results are also compared by Çymon API that provides threat intelligence data. In another work in [15] malicious IOT devices working as botnets are identified from data received at darknet. The identification/fingerprinting of compromised IoT devices is carried out by binary classifier based upon CNN.IoT devices infected by malwarestend to work in a coordinated manner with common campaign objective of scan or DoS attack. Surveillance of these campaigns and fingerprinting of the IoT devices helps to generate a cyber-threat intelligence.

III. LIMITATIONS AND CHALLENGES

Darknet are usually used to accumulate large traffic for research on potentially malicious activities by deploying one or more machines onto unused address range. However, these approaches may be impractical, based on the expense of accessing and handling large blocks of darknets. Larger the size of darknet more information it gathers for threat detection. Furthermore, because IP addresses are a finite resource, attempts have been made to prevent large blocks of IP addresses from not being used.

Huge amount of traffic is seen on darknets. Storing, processing and analysing is a difficult task. IP addresses in darknet are not assigned and no services are hosted onto these range of IP addresses. The only traffic coming to this range are probing packets either UDP, TCP SYN, or ICMP. No packet containing payload are received by these addresses. So little information is obtained by analysing the traffic. The attackers are using malwares with different forms, patterns or signatures. Identification of new malware becomes a challenging task.

Using low interaction honeypot to collect traffic for analysis along with darknet helps to gain more insight into identification of attack scale and attacks to some extent. A large scale deployment of honeypots with high interaction may alert the attacker.

Tools like zmap and masscan can scan entire IPV4 network containing 2^{32} addresses in less than an hour. Scanning a small part of IPV6 network by any tool will be impossible due to its huge size of 2^{128}. Use of darknets onto IPV6 address space to find malicious and scanning activity is difficult. Scanning activity on IPV6 is identified by DNS backscatter in [8] reason given that, darknets will be ineffective

IV. CONCLUSION

Data on communication links of internet is ever-growing. Cyber threats are also changing their forms using new techniques and tools, making it difficult to analyse for security systems. Darknet is used for passively monitoring and identification of malicious activities. They are sensors placed in unused IP addresses. Darknet or network Telescope are also called as Internet backgroundradiation are used as a method or technique to identify the cyber threat. Data collected from darknet can be used to profile an attack strategy, draw statistical conclusion about the size & location of attack, develop a threat model for the network, and attribute the device used. Darknets are researched for use in IPV4 address space. However expansion of Internet

with Internet of things will lead to more use of IPV6 address space. The working area of research can be expanded with use of IPV6 and IOT based systems.

ACKNOWLEDGMENT

Our sincere thanks to Dr. K.G. Vishwanath, Principal and Director, CSE Research centre head, Jain College if Engineering, Belagavi, and Dr. S.M. Deshpande, Principal, MarathwadaMitraMandal's College ofEngineering, Pune for the encouragement.

REFERENCES

[1]Zakroum, M., et al.: Exploratory data analysis of a network telescope traffic and prediction of port probing rates. In: 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 175–180. IEEE (2018)

[2]S. Kumar, H. Vranken, J. v. Dijk and T. Hamalainen, "Deep in the Dark: A Novel Threat Detection System using Darknet Traffic," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 4273-4279, doi: 10.1109/BigData47090.2019.9006374.

[3] C. Han et al., "Real-Time Detection of Malware Activities by Analyzing Darknet Traffic Using Graphical Lasso," 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 2019, pp. 144-151, doi: 10.1109/TrustCom/BigDataSE.2019.00028.

[4] Hideaki Kanehara, Yuma Murakami, JumpeiShimamura, Takeshi Takahashi, Daisuke Inoue, and Noboru Murata. 2019. Real-time botnet detection using nonnegative tucker decomposition. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing(SAC '19). Association for Computing Machinery, New York, NY, USA, 1337–1344. DOI:https://doi.org/10.1145/3297280.3297415

[5]W. Fan, Z. Du, D. Fernández and V. A. Villagrá, "Enabling an Anatomic View to Investigate Honeypot Systems: A Survey," in IEEE Systems Journal, vol. 12, no. 4, pp. 3906-3919, Dec. 2018, doi: 10.1109/JSYST.2017.2762161.

[6] R. Akiyoshi, D. Kotani and Y. Okabe, "Detecting Emerging Large-Scale Vulnerability Scanning Activities by Correlating Low-Interaction Honeypots with Darknet," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, 2018, pp. 658-663, doi: 10.1109/COMPSAC.2018.10314.

[7] T. Ban and D. Inoue, "Practical darknet traffic analysis: Methods and case studies," 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA, 2017, pp. 1-8, doi: 10.1109/UIC-ATC.2017.8397445.

[8] Kensuke Fukuda and John Heidemann. 2018. Who Knocks at the IPv6 Door? Detecting IPv6 Scanning. In Proceedings of the Internet Measurement Conference 2018 (IMC '18). Association for Computing Machinery, New York, NY, USA, 231–237. DOI:https://doi.org/10.1145/3278532.3278553

[9] Jakub Czyz, Kyle Lady, Sam G. Miller, Michael Bailey, Michael Kallitsis, and Manish Karir. 2013. Understanding IPv6 internet background radiation. In Proceedings of the 2013 conference on Internet measurement conference (IMC '13). Association for Computing Machinery, New York, NY, USA, 105–118. DOI:https://doi.org/10.1145/2504730.2504732

[10] S. Torabi, E. Bou-Harb, C. Assi, M. Galluscio, A. Boukhtouta and M. Debbabi, "Inferring, Characterizing, and Investigating Internet-Scale Malicious IoT Device Activities: A Network Telescope Perspective," 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Luxembourg City, 2018, pp. 562-573, doi: 10.1109/DSN.2018.00064.

[11] Norbert Blenn, Vincent Ghiëtte, and Christian Doerr. 2017. Quantifying the Spectrum of Denial-of-Service Attacks through Internet Backscatter. In Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17). Association for Computing Machinery, New York, NY, USA, Article 21, 1–10. DOI:https://doi.org/10.1145/3098954.3098985

[12] C. Fachkha, E. Bou-Harb and M. Debbabi, "Inferring Distributed Reflection Denial of Service Attacks from Darknet", Comput. Commun, vol. 62, no. C, pp. 59-71, 2015.

[13] Pang, S., Komosny, D., Zhu, L. et al. Malicious Events Grouping via Behavior Based Darknet Traffic Flow Analysis. Wireless PersCommun 96, 5335–5353 (2017). https://doi.org/10.1007/s11277-016-3744-4

[14] Stones DalitsoChindipha, Barry Irwin, and Alan Herbert. 2019. Quantifying the Accuracy of Small Subnet-Equivalent Sampling of IPv4 Internet Background Radiation Datasets. In Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019 (SAICSIT '19). Association for Computing Machinery, New York, NY, USA, Article 20, 1–8. DOI:https://doi.org/10.1145/3351108.3351129

[15]MortezaSafaei Pour, Antonio Mangino, Kurt Friday, Matthias Rathbun, Elias Bou-Harb, Farkhund Iqbal, Khaled Shaban, and AbdelkarimErradi. 2019. Data-driven Curation, Learning and Analysis for Inferring Evolving IoT Botnets in the Wild. In Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19). Association for Computing Machinery, New York, NY, USA, Article 6, 1–10. DOI:https://doi.org/10.1145/3339252.3339272

[16] A. Dainotti, A. King, K. Claffy, F. Papale and A. Pescapé, "Analysis of a "/0" Stealth Scan From a Botnet," in IEEE/ACM Transactions on Networking, vol. 23, no. 2, pp. 341-354, April 2015, doi: 10.1109/TNET.2013.2297678.

[17] S. Torabi, E. Bou-Harb, C. Assi, E. B. Karbab, A. Boukhtouta and M. Debbabi, "Inferring and Investigating IoT-Generated Scanning Campaigns Targeting A Large Network Telescope," in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2020.2979183.

[18] S. H. A. Ali, S. Ozawa, T. Ban, J. Nakazato and J. Shimamura, "A neural network model for detecting DDoS attacks using darknet traffic features," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 2016, pp. 2979-2985, doi: 10.1109/IJCNN.2016.772757

[19] F. Shaikh, E. Bou-Harb, J. Crichigno and N. Ghani, "A Machine Learning Model for Classifying Unsolicited IoT Devices by Observing Network Telescopes," 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, 2018, pp. 938-943, doi: 10.1109/IWCMC.2018.8450404.

[20] F. Soro, I. Drago, M. Trevisan, M. Mellia, J. Ceron and J. J. Santanna, "Are Darknets All The Same? On Darknet Visibility for Security Monitoring," 2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), Paris, France, 2019, pp. 1-6, doi: 10.1109/LANMAN.2019.8847113.

[21] M. Coudriau, A. Lahmadi and J. François, "Topological analysis and visualisation of network monitoring data: Darknet case study," 2016 IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, 2016, pp. 1-6, doi: 10.1109/WIFS.2016.7823920.