Labeling Network Telescope Data: Challenges and New Directions (Abstract)

Michalis Kallitsis
mgkallit@merit.edu

Network telescopes consist of networking infrastructure that record and receive unsolicited Internet-wide activities. A network telescope or “darknet” is configured to listen to traffic destined to an *unused* but *routed* IP space. Since this “dark IP space” serves no legitimate network services such as Web servers, DNS resolvers, etc., any traffic arriving to the darknet is inherently suspicious/malicious. Thus, network and security researchers have been utilizing network telescopes to study a plethora of macroscopic Internet activities such as shedding light into botnets [1], [2], obtaining insights about network outages [3], [4], understanding certain types of denial of service attacks [5], [6], [7], examining the behavior of IoT devices [8], and detecting Internet misconfigurations [9], [10], etc.

Despite the indisputable utility of network telescopes, annotations and labeling of darknet activities usually happen on an opportunistic basis (e.g., to study, in a post mortem manner, a known event, such as the onset of the Mirai botnet [1]). The problem of *automated* and *ongoing* labeling has been a challenging one, exacerbated by the sheer volume of darknet data that the analyst needs to consider. In this presentation, we would like to describe recent efforts on renewing Merit Network’s network telescope and adding meaningful annotations on the collected data. Specifically, we would like to present the ORION (Observatory for cyber-Risk Insights and Outages of Networks) infrastructure [11] and its data pipeline that extracts Darknet events of interest (such as scanning activities and spoofing-based denial of service attacks) while also uploading—in near-real-time—such annotated data into Google’s BigQuery for ease of processing and analysis. The ORION data pipeline enriches the identified Darknet events with several other useful meta-data (such as routing, DNS and geolocation information) along with useful fingerprints/labels that can be extracted from packet headers (i.e., the Mirai, Masscan and ZMap fingerprints).

The ORION network telescope has also enabled new opportunities for data labeling. We would like to discuss our work on clustering darknet data using AI/ML techniques such as unsupervised clustering [12]. In this work, our focus is on assigning network-based features to scanning IPs appearing in the ORION network telescope, and using unsupervised learning techniques, such as K-means, to assign the darknet scanners into groups of similar activities. To deal with the high-dimensionality of the input feature space, we employ “deep representation learning” techniques to “compress” the input data into a lower-dimensional space of embeddings that we would then use for clustering. We will discuss how the clustering output can be leveraged for the detection of temporal changes in the structure of the darknet (i.e., finding when “new” activities may be appearing in the darknet), thus enabling novel ways to offer meaningful threat intelligence to the cybersecurity community.

References

[1] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis et al., “Understanding the mirai botnet,” in 26th USENIX Security Symposium (USENIX Security 17), 2017.

[2] A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescap´e, “Analysis of a /0; stealth scan from a botnet,” IEEE/ACM Transactions on Networking, vol. 23, no. 2, pp. 341–354, April 2015

[3] K. Benson, A. Dainotti, k. Claffy, and E. Aben, “Gaining insight into as-level outages through analysis of internet background radiation,” ser. CoNEXT Student 2012, 2012.

[4] A. Dainotti, R. Amman, E. Aben, and K. C. Claffy, “Extracting benefit from harm: Using malware pollution to analyze the impact of political and geophysical events on the internet,” SIGCOMM CCR 2012

[5] J. Czyz, M. Kallitsis, M. Gharaibeh, C. Papadopoulos, M. Bailey, and M. Karir, “Taming the 800 pound gorilla: The rise and decline of NTP DDoS attacks,” in Proceedings of the 2014 Conference on Internet Measurement Conference, 2014

[6] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, “Inferring internet denial-of-service activity,” ACM Trans. Comput. Syst.

[7] M. Jonker, A. King, J. Krupp, C. Rossow, A. Sperotto, and A. Dainotti, “Millions of targets under attack: A macroscopic characterization of the dos ecosystem,” in Proceedings of the 2017 IMC, 2017

[8] F. Shaikh, E. Bou-Harb, J. Crichigno, and N. Ghani, “A machine learning model for classifying unsolicited iot devices by observing network telescopes,” in 2018 14th International Wireless Communications and Mobile Computing Conference (IWCMC), 2018, pp. 938–943.

[9] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, “Internet background radiation revisited,” 10th ACM SIGCOMM Conference on Internet Measurement, 2010

[10] J. Czyz, K. Lady, S. G. Miller, M. Bailey, M. Kallitsis, and M. Karir, “Understanding IPv6 Internet background radiation,” in IMC 2013, 2013.

[11] Merit Network, Inc., ORION Network Telescope: Observatory for cyber-Risk Insights and Outages of Networks, https://www.merit.edu/initiatives/orion-network-telescope/. Note: An NSF-funded project with PIs Michalis Kallitsis, Zakir Durumeric and Stilian Stoev.

[12] M. Kallitsis, R. Prajapati, V. Honavar, D. Wu and J. Yen, "Detecting and Interpreting Changes in Scanning Behavior in Large Network Telescopes," in *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3611-3625, 2022, doi: 10.1109/TIFS.2022.3211644.