Shedding Light Into the Darknet: Scanning Characterization and Detection of Temporal Changes

Rupesh Prajapati, Vasant Honavar, Dinghao Wu, John Yen and Michalis Kallitsis
Penn State University, Merit Network, Inc.
rxp338@psu.edu,vuh14@psu.edu,dwu12@psu.edu,juy1@psu.edu and mgkallit@merit.edu

Introduction

Network telescopes provide a unique window into Internet-wide malicious activities associated with malware propagation, denial of service attacks, network reconnaissance, and others. Analyses of this telescope data can highlight ongoing malicious events in the Internet which can be used to prevent or mitigate cyberthreats in real-time. However, large telescopes observe millions of events on a daily basis which renders the task of transforming this knowledge to meaningful insights challenging. In order to address this, we present a novel framework for characterizing Internet's background radiation and for tracking its temporal evolution. The proposed framework:

1. Extracts a high dimensional representation of telescope scanners composed of features distilled from telescope data and learns an information-preserving low-dimensional representation of these events that is amenable to clustering
2. Performs clustering of resulting representation space to characterize the scanners
3. Utilizes the clustering outcomes as "signatures" to detect temporal changes in the network telescope

Network Telescope

Network telescopes or "Darknets" receive and record unsolicited traffic destined to an unused but routed address space and hence, provide a unique opportunity for characterizing Internet-wide malicious activities in a timely manner. This "dark IP space" hosts no services or devices, and therefore, any traffic arriving to it is inherently malicious.

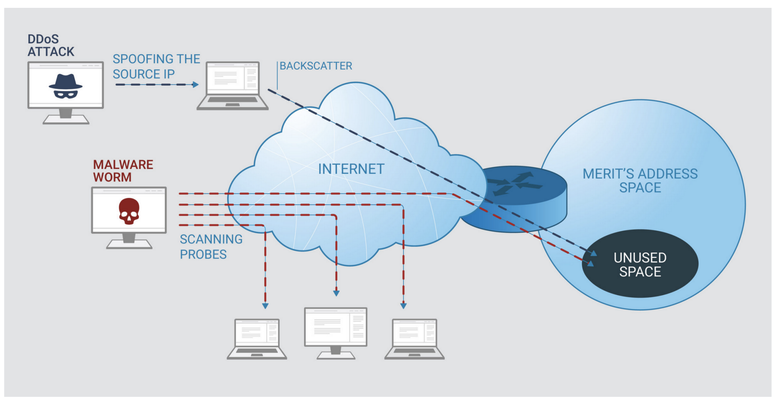


Figure: Scanning and backscatter traffic captured in the Darknet.

Problem Formulation

Our primary objective is to identify the patterns of scanning behaviors and track dynamic changes in these patterns.

1. High dimensional Data
   - Features like the scanned ports need to be one-hot encoded
2. Pattern Recognition
3. Temporal Change Detection

Acknowledgements

This work is partially supported by the U.S. DHS under Grant Award Number 17STQAC00001-05-00 and by the NSF CNS-1823192 award. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies of the sponsor.

References

[1] Félix Iglesias and Tanja Zseby. Pattern discovery in internet background radiation. *IEEE Transactions on Big Data*, 2017.

[2] Paul Barford, Yan Chen, Anup Goyal, Zhichun Li, Vern Paxson, and Vinod Yegneswaran. *Employing Honeynets For Network Situational Awareness*, pages 71–102. Springer US, Boston, MA, 2010.

Methodology

Our starting point is raw traffic traces from a large network telescope (our team has access to a /13 Darknet spanning approximately 500,000 unique IPs; on a typical day, more than 100 GB of compressed Darknet data is collected consisting of some 3 billion packets on average).

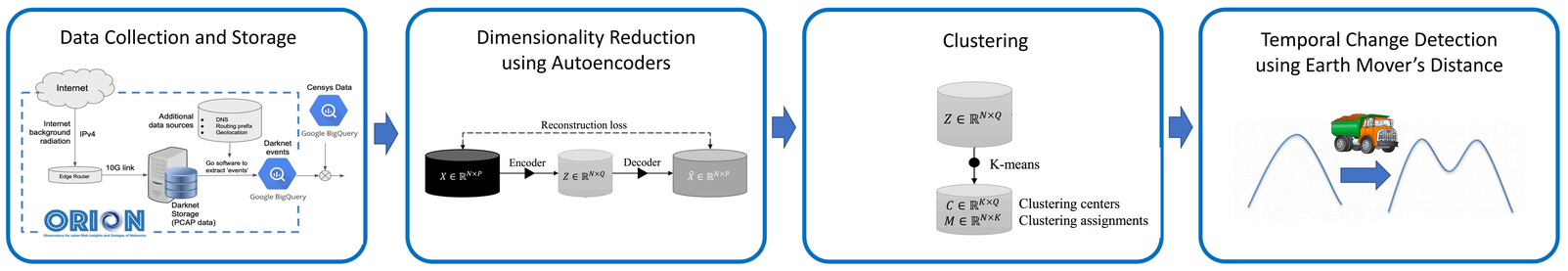
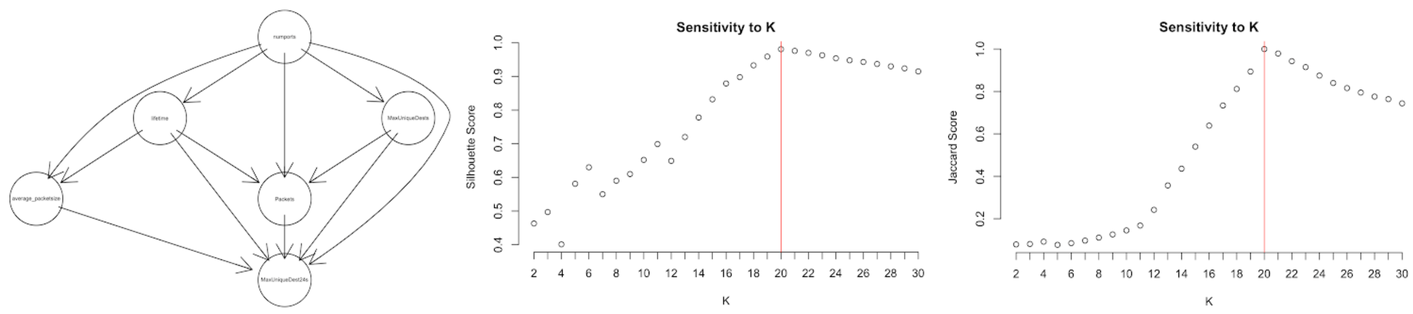


Figure: Complete pipeline from data collection to temporal change detection

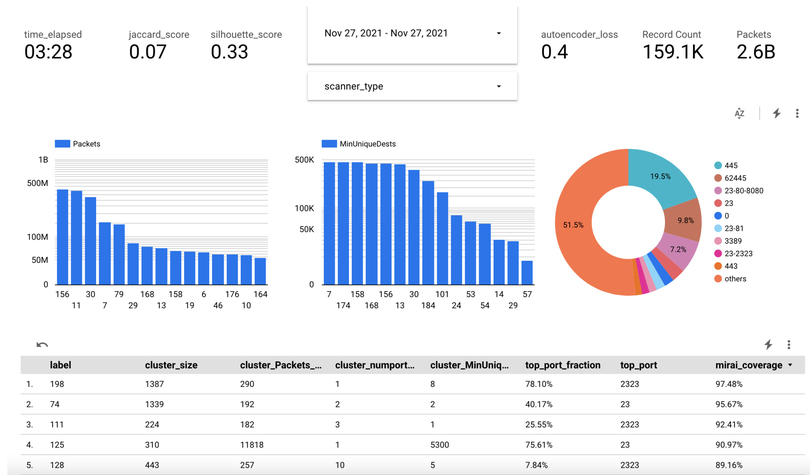
Evaluation

Before applying this methodology on real data, we applied this pipeline on various sets of synthetic data. A Bayesian network trained on the real data under certain constraints generates synthetic data for which we know the real cluster assignments.

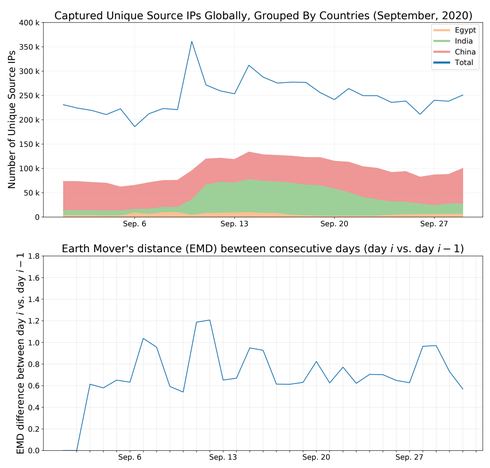


(a) Bayesian Network (b) Silhouette Score Evaluation (c) Jaccard Score Evaluation

Preliminary Results and Discussion



(a) Dashboard for analysts to study clusters.



(b) EMD metric captures the temporal changes.

The figure b) shows a significant increase in the EMD distance between the clustering outcome of Sep. 5^{th} and Sep. 6^{th} which indicates a structural change in our Darknet. This increase in EMD distance is corroborated by the sudden increase in scanning traffic originating from countries like India and Egypt. In our clustering results for Sep. 6^{th}, we observe novel clusters with scanners scanning a particular set of ports: 23, 80, 2323, 7574, 8080, 37215, 49152 and 52869.

Summary and Next Steps

We presented a novel framework towards network situational awareness. In addition to Darknet characterization[1], our approach utilizes the clustering outcomes to detect structural changes in the Darknet. As part of ongoing work, we plan to: a) expand the set of features we select [2]; b) integrate additional data sources into our system (e.g., VirusTotal, ExploitDB, honeypot data, etc.); c) add new tools/analyses for cluster interpretation (e.g., decision trees).