Topological Analysis and Visualisation of Network Monitoring Data: Darknet case study

Marc Coudriau[1,2], Abdelkader Lahmadi[3], Jérôme François[2]

[1]ENS Ulm, Paris, France

[2]Inria Nancy Grand Est, Villèrs-les-Nancy, France

[3]LORIA, Université de Lorraine, France

NMRG meeting, IETF 99, Prague

Overview

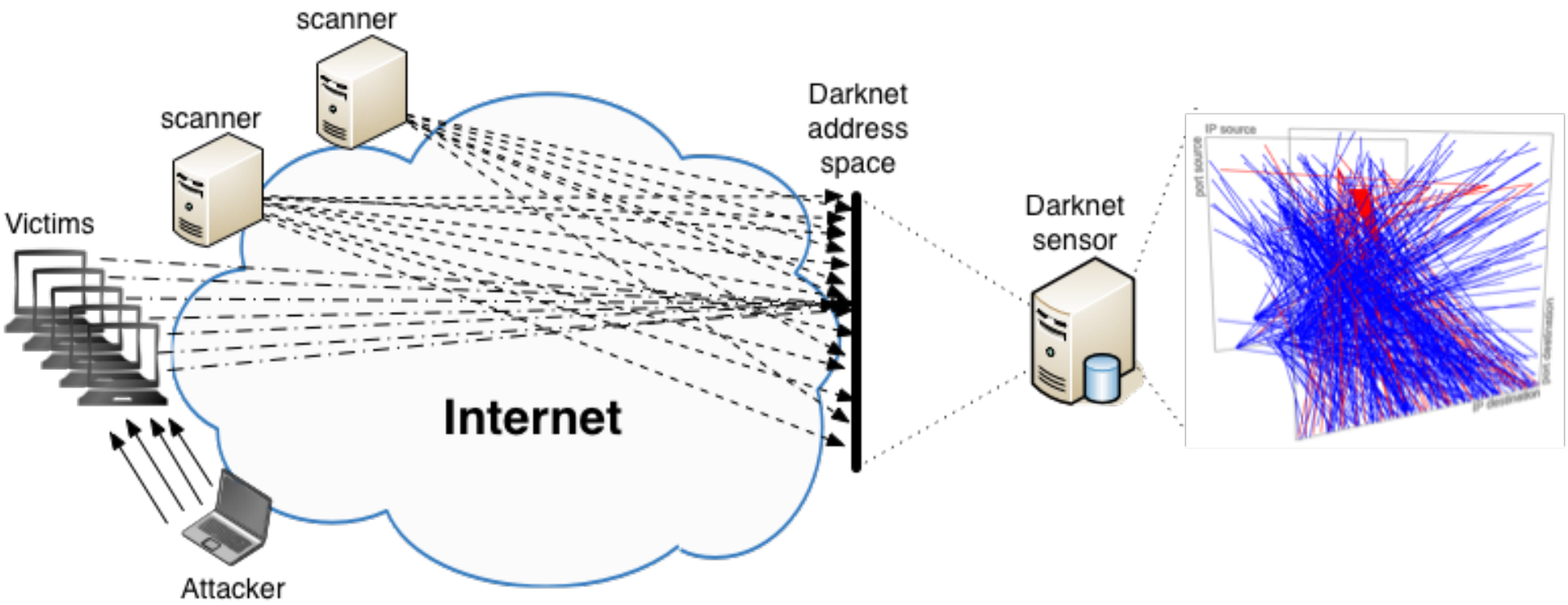
- Motivation
- Background and related work
- Methodology
- Experimental results
  - Topologies of scanning activities
  - Topologies of DDoS activities
- Conclusion and future work

Network Monitoring Data

- Widely used for security, forensics and anomaly detection
- Identify malicious activities: traffic patterns and alerts triggering
- Internet Background Radiation: IBR
  - network telescopes, darknets
  - noisy traffic, but important source of forensic data
  - considerable volume and wide range of services and sources
  - extraction of structures and components
  - prediction and modeling of Internet malicious activities

Darknets

- Traffic sent to unused IP addresses
- Nonproductive traffic: no legitimate traffic
- Silently collecting all incoming packets, i.e. without replying to any of them



Problem statement

- What are the components of a darknet traffic ?
- How can we filter this traffic to extract types of malicious activities ?

Characterization of IBR

- First characterisation of IBR traffic : composition of observed protocols and ports [Pang el al, 2004]
- Probability to observe DoS attacks with a telescope [Moore et al, 2006]
- Characterization of IBR traffic over multiple darknets to extract invariant features and level of pollution of destination IP addresses [Wustrow el al, 2010]

[Pang et al, 2004] R. Pang, et al,"Characteristics of internet background radiation," in Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 27–40.

[Moore et al] D. Moore, et al, "Inferring internet denial-of-service activity," ACM Trans. Comput. Syst., vol. 24, no. 2, May 2006.

[Wustrow et al, 2010] E. Wustrow, et al, "Internet background radiation revisited," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 62–74

Characterization of darknet data

- Analysis of main activities of a Darknet (scanning, worms propagation) using clustering and visualisation techniques [Fachka et al, 2016]
- Analysis of DNS queries to identify DRDoS (Distributed Reflection Denial of Service) [Fachka et al, 2015]

[Fachka et al, 2016] C. Fachkha et al, "Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization," IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 1197–1227, Second quarter 2016.

[Fachka et al, 2015] C. Fachkha et al,"Inferring distributed reflection denial of service attacks from darknet," Computer Communications, vol. 62, pp. 59-71,2015.

Visualisation of Darknet data

- InetVis plots darknet data on a 3D scatter plot and highlights visual patterns using IDS alerts like Bro or Snort [Van Riel et al, 2006]
- 3D visualisation tool to monitor darknet traffic in real time [Inoue et al, 2012]

[Van Riel et al, 2006] J-P. van Riel et al, "Inetvis, a visual tool for network telescope traffic analysis," in Proceedings of the 4th International Conference on Computer Graphics. ACM, 2006.

[Inoue et al, 2012] D. Inoue et al, "Daedalus- viz: Novel real-time 3d visualization for darknet monitoring-based alert system," in

Proceedings of the Ninth International Symposium on Visualization for Cyber Security, ser. VizSec '12, 2012, pp. 72–79.

Topological Data Analysis (TDA)

Definition

Branch of mathematics to analyze high dimensional and complex data by extracting invariant geometrics features that might help us discover relationships and patterns in data.

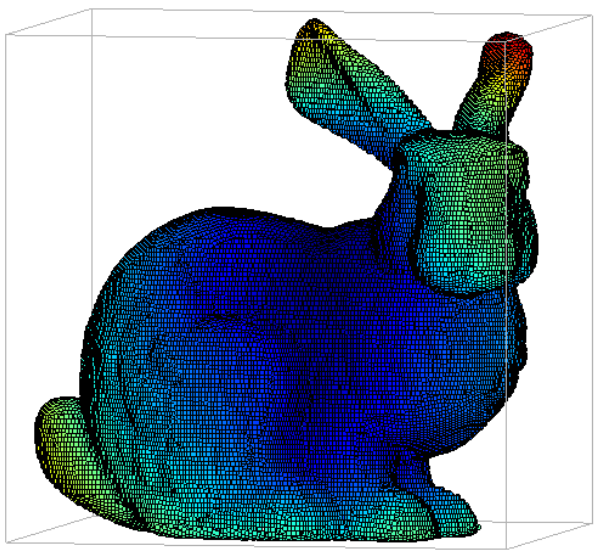
Fundamental properties

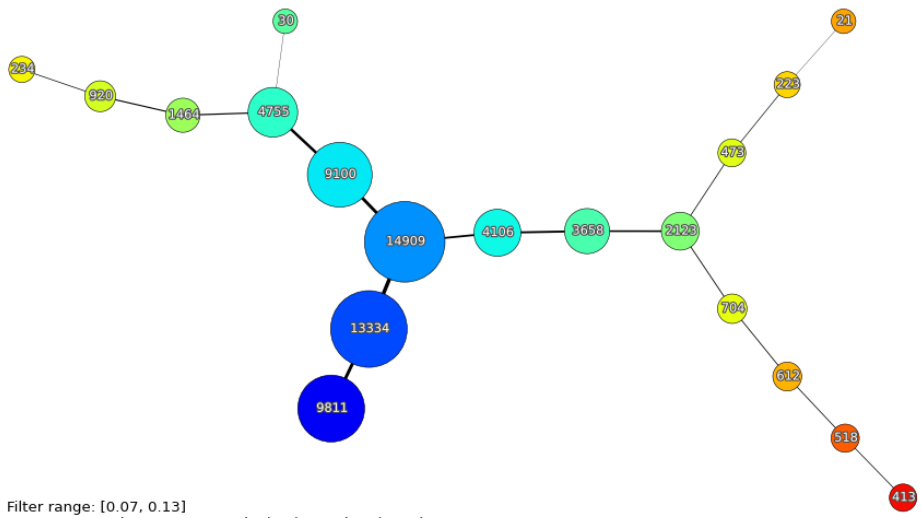
- Coordinate invariance
  - does not depend on coordinate system
  - analyze data collected from different platforms
- Deformation invariance
  - less sensitive to noise
  - handle approximate data
- compressed representation

[Carlson, 2009] G. Carlsson, "Topology and data," Bulletin of the American Mathematical Society, vol. 46, no. 2, pp. 255–308,

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges

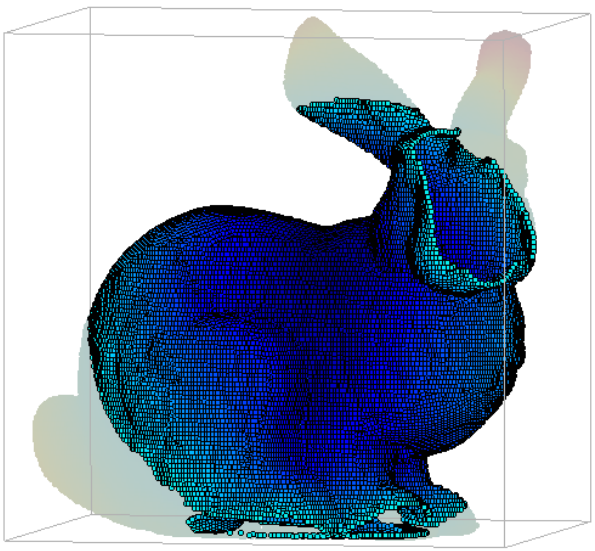


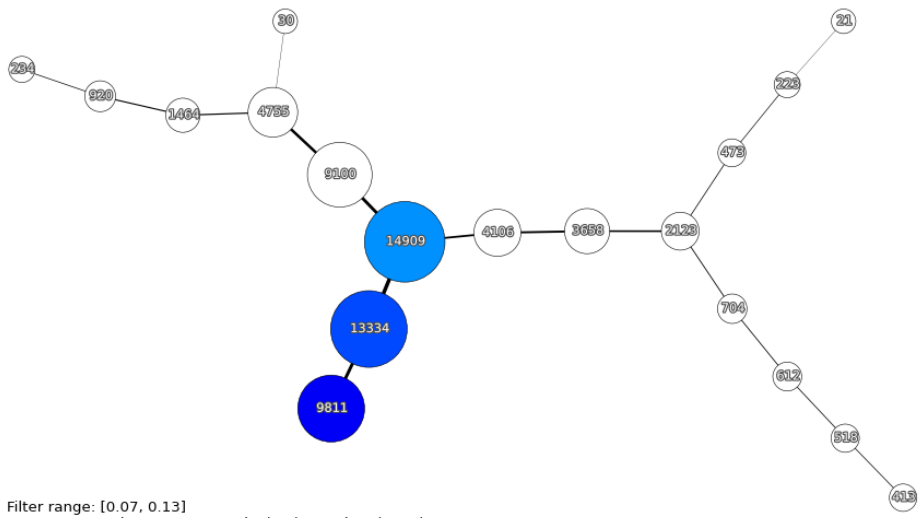


[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges



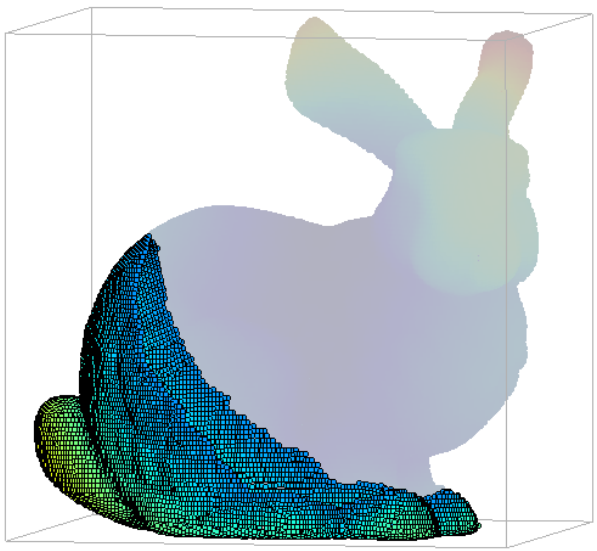


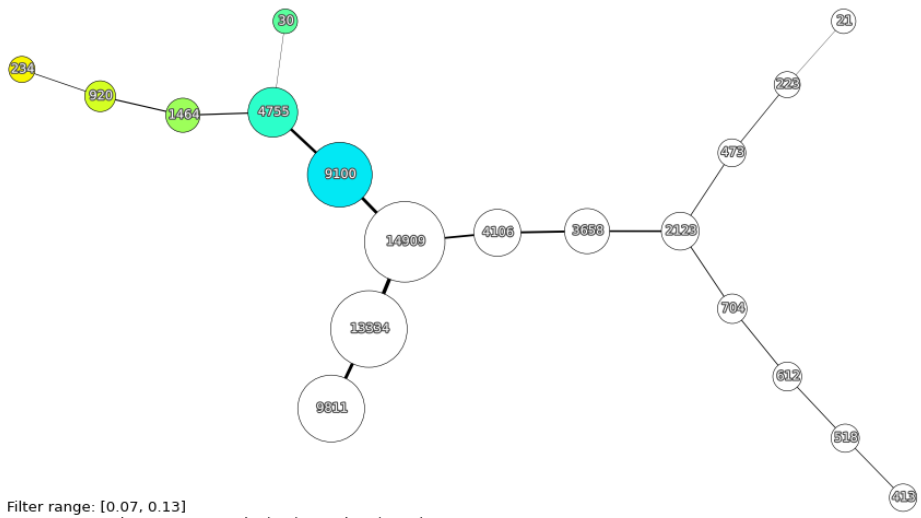
Filter range: [0.07, 0.13]
Cover: Hypercube cover. Intervals: (10,). Overlap: (50.0,)
Clustering method: Single linkage clustering
Cutoff: First gap of relative width 0.1
Size range: [21,14909]

[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges

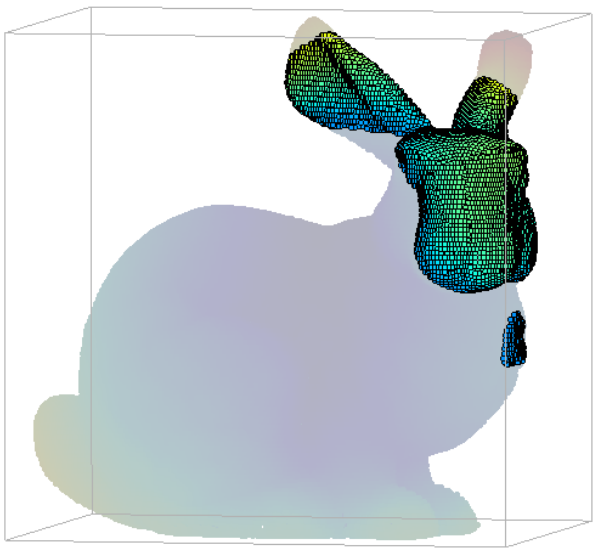


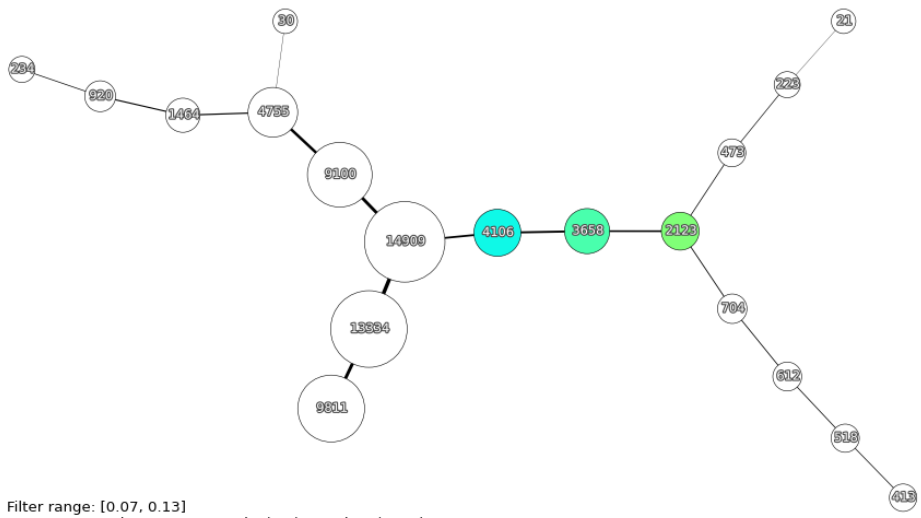


[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges

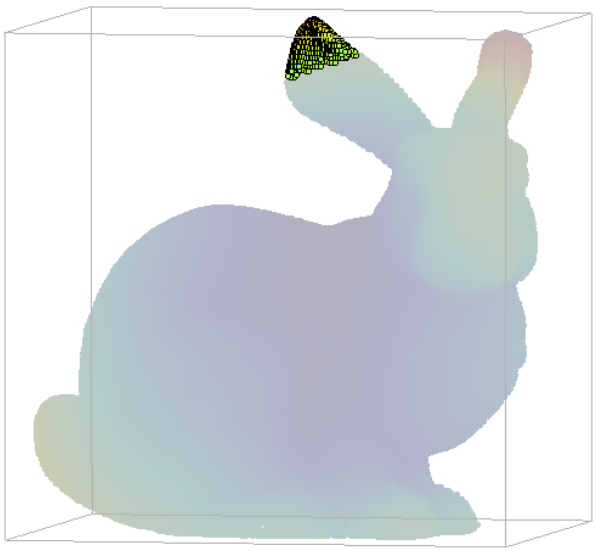


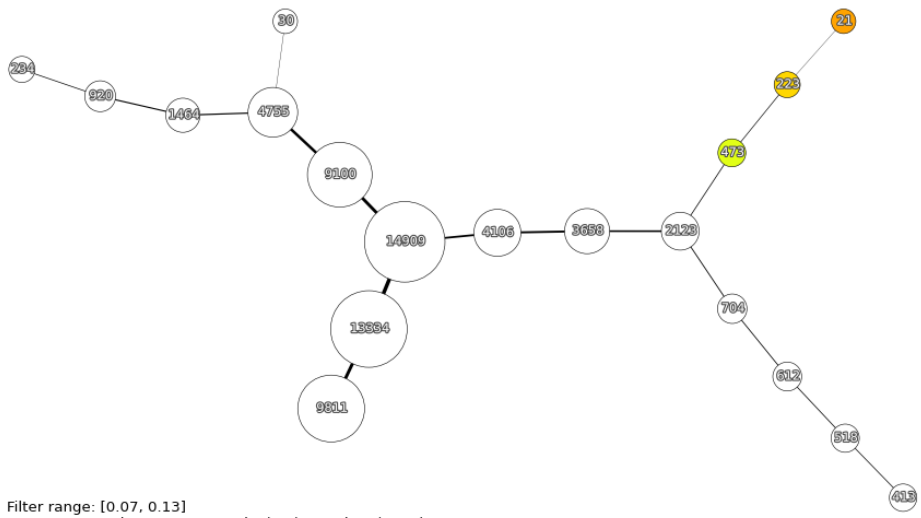


[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges



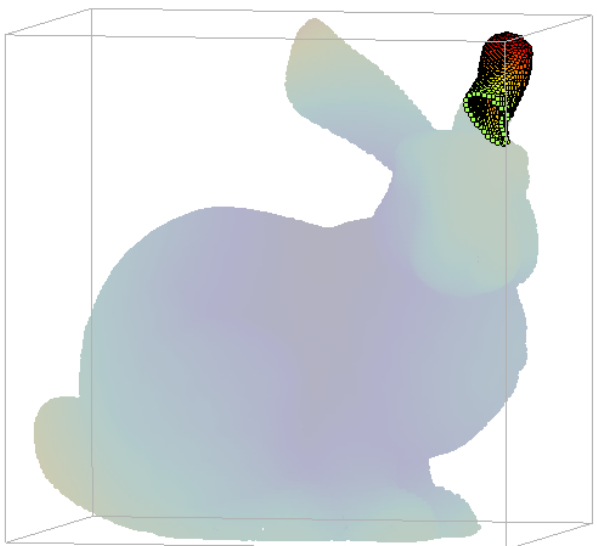


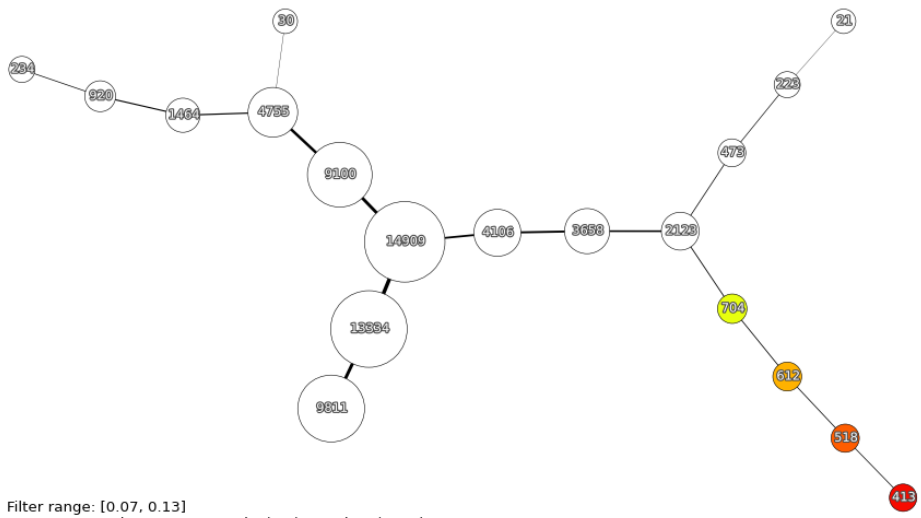
Filter range: [0.07, 0.13]
Cover: Hypercube cover. Intervals: (10,). Overlap: (50.0,)
Clustering method: Single linkage clustering
Cutoff: First gap of relative width 0.1
Size range: [21,14909]

[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

TDA in practice

- Input data: 3D point cloud representing the Stanford Bunny (35947 points)
- Filter function: $f(x_i) \rightarrow \text{eccentricity}(x_i)$
- Output : network with 19 vertices and 18 edges





Filter range: [0.07, 0.13]
Cover: Hypercube cover. Intervals: (10,). Overlap: (50.0,)
Clustering method: Single linkage clustering
Cutoff: First gap of relative width 0.1
Size range: [21,14909]

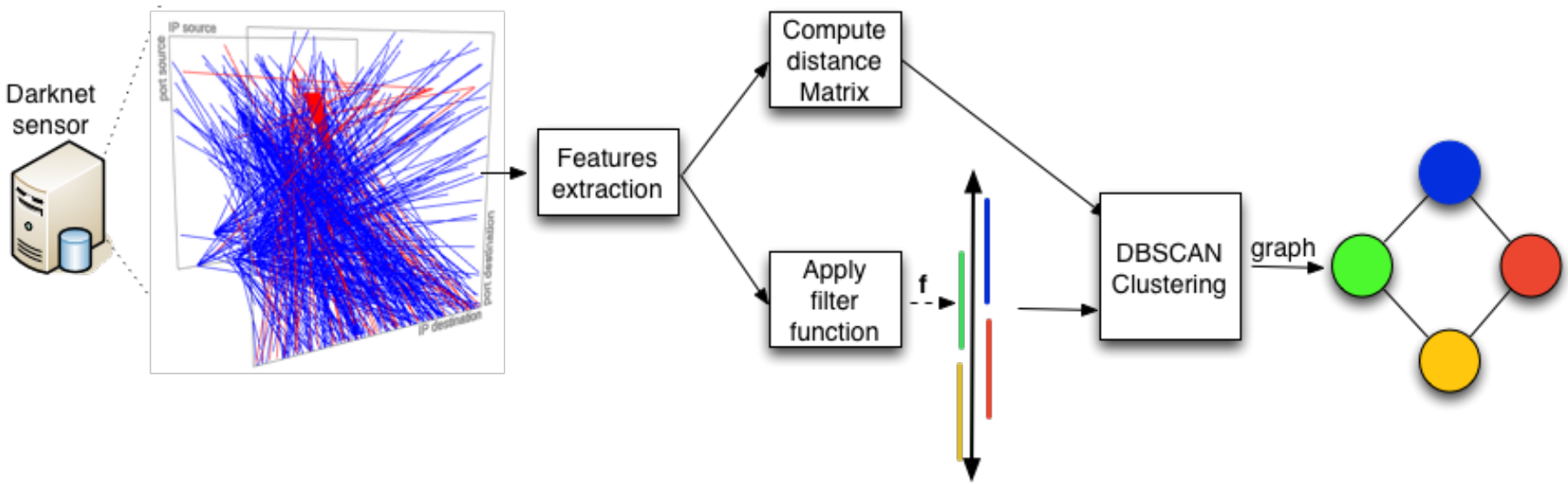
[Lum et al., 2013] Lum et al. "Extracting insights from the shape of complex data using topology". Scientific Reports, 3:1236.

Method overview

Objective

- extracting activities from noisy monitoring data collected by LHS darknet (/20 subnetwork)
- data set: a month of collected data with a rate of 3 millions packets per day

Apply Mapper method from TDA on darknet traffic to extract attack patterns (scanning, DDoS)



Mapper method details

- Input : feature vectors of darknet packets (the timestamp, the source and destination IP addresses and ports, and the protocol)
- Parameters: number of intervals (resolution), overlapping percentage (zoom)
- output :

1. Filter function f: $\mathbb{R}^6 \rightarrow \mathbb{R}^6$
2. Put data into overlapping bins : $f^{-1}(a_i, b_i)$
3. Cluster each bin using DBSCAN and a distance function
4. Create a graph
   - Vertex: a cluster of a bin
   - Edge: nonempty intersection between clusters

Partial clustering details

- Apply DBSCAN clustering within each hypercube
- Two parameters
  - ϵ: the maximum distance between two points to be considered in the same cluster
  - *minpts*: the number of neighbors that a point should have to be considered as a cluster
- Used distance function
  - Difference for timestamp attribute, IP destination and source addresses
  - Equality metric for protocol and ports names : 0 or 1

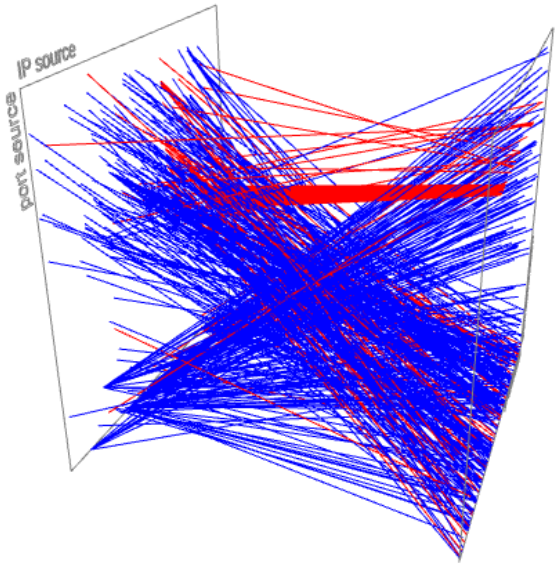
Separating patterns

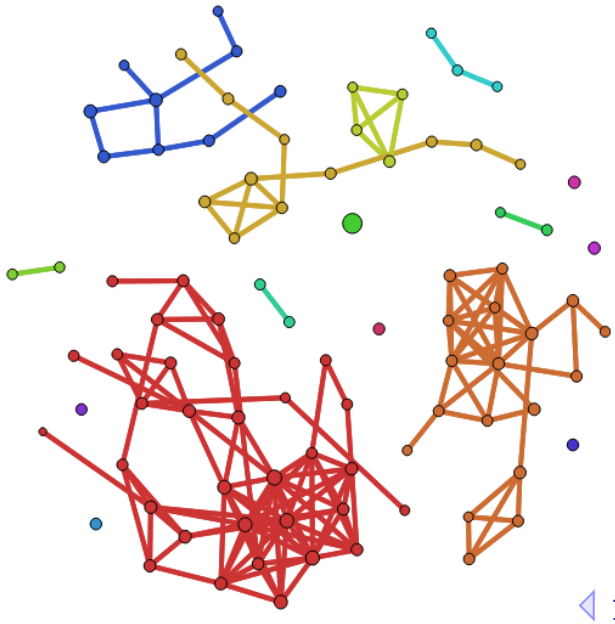
Mapper parameters

- 1000 packets with $\epsilon = 0.5$ and minpts=3 and overlap $= 10\%$

Extracted patterns

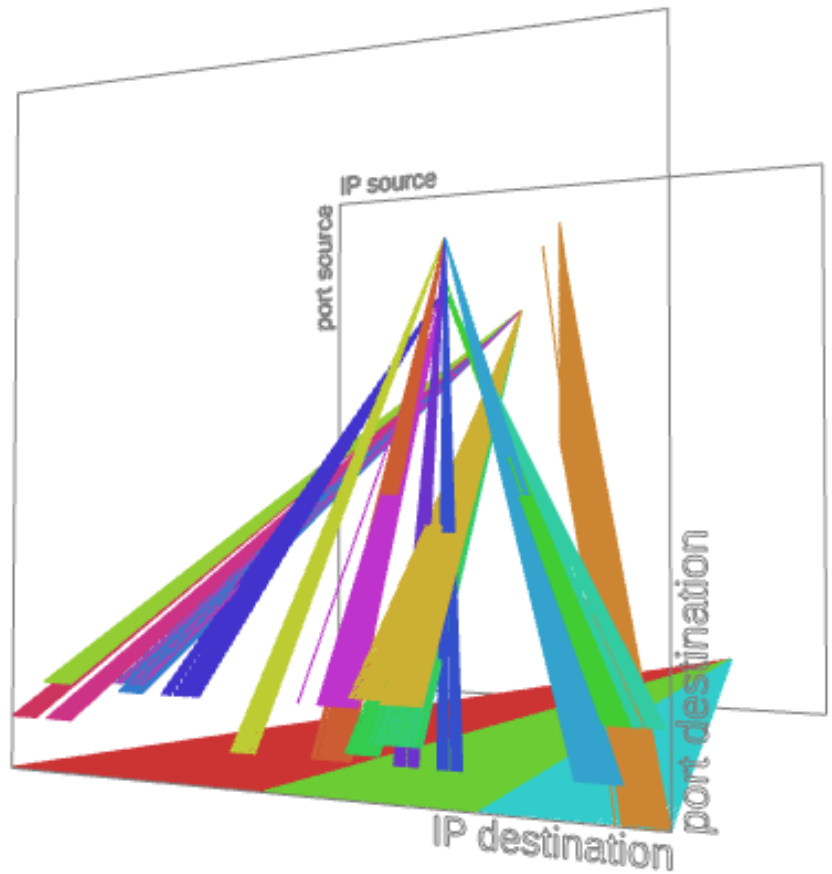
- large green dot: scanning activity on port 53413 (known exploit)
- red component: probing Telnet and SSH accesses
- orange component: sparse scans
- yellow component: two randomized scans and some noise





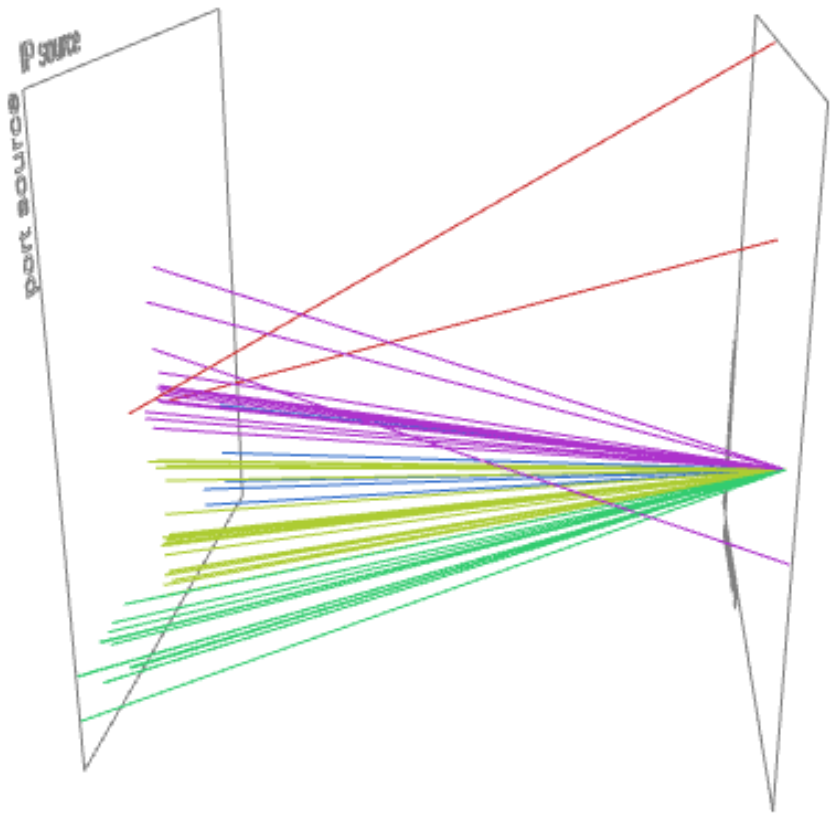
Extracting scanning activities

- 8000 packets, $\epsilon = 0.05$ and minpts=20, overlap=5%
- Parameters estimation: trial-and-error method, but remains stable when found
- Suricata 3.0 detects only 4 scanning activities: grouping packets



Extracting DDoS activities

- 310 000 UDP packets (DNS responses to a spoofed darknet IP address)
- $\epsilon = 0.03$ and minpts=100, overlap=1%



Performance analysis

- Results obtained with a machine having a Quad Core CPU at 2.83GHz, 15 GB RAM and running Linux Mint
- Mapping and clustering of 1024 packets takes a processing time between 0.4s to 0.9s
- Analyzing 3 millions of packets (a darknet day) requires 11 minutes
- Partial clustering in hypercubes: more efficient then global clustering
- What a known attacker sent today ?
  - 32768 packets analyzed in two minutes
- Increasing performance
  - More computing power
  - Parallelization of the tool to make near real-time analysis

Conclusion and future work

- Topological Data Analysis applied to darknet traffic
- Mapper method: filter function (number of intervals and their overlap) and partial clustering using DBSCAN
- Extraction of activities: packets belonging to the same activity (scans and DDoS)
- Experimental results: discovering more patterns than the well-used Suricata IDS

Future work

- Including more packet features
- Extract more activities and analyze their persistance

Acknowledgment

This work was partially funded by HuMa, a project funded by Bpifrance and Region Lorraine under the FUI 19 framework. It is also supported by the High Security Lab hosted at Inria Nancy Grand Est.

Topological Analysis and Visualisation of Network Monitoring Data: Darknet case study

Marc Coudriau[1,2], Abdelkader Lahmadi[3], Jérôme François[2]

[1]ENS Ulm, Paris, France

[2]Inria Nancy Grand Est, Villèrs-les-Nancy, France

[3]LORIA, Université de Lorraine, France

NMRG meeting, IETF 99, Prague