Clustering Malware's Network Behavior using Simple Sequential Features

Azqa Nadeem[1Φ] Carlos H. Ganan[2] Sicco Verwer[1]

[1] Cyber Security Group, Department of Intelligent Systems*

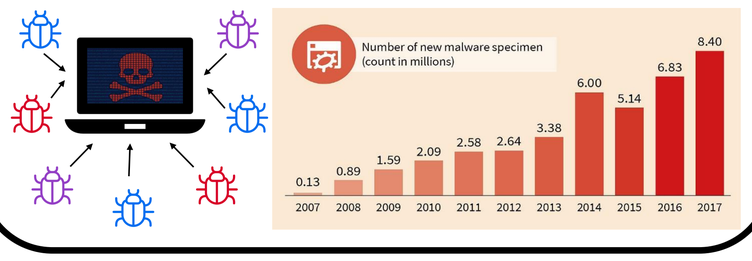
[2] Organization and Governance Group, Department of Multi Actor Systems*

*Delft University of Technology

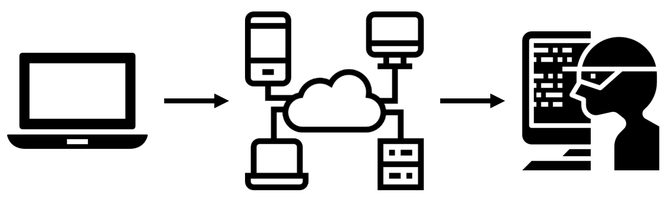
Φazqa.nadeem@tudelft.nl

Problem: Malware variant characterization is difficult

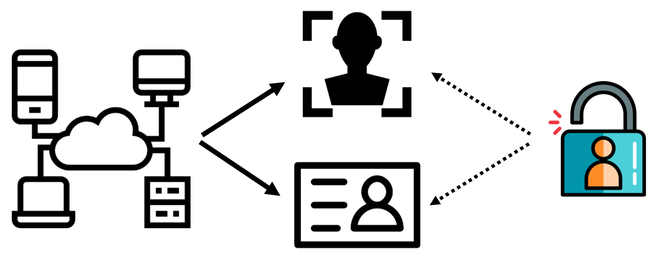
Too many malware variants to analyze



Network traffic shows core malware behavior

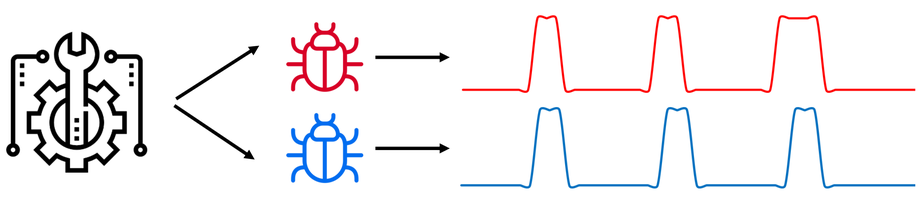


Network traffic is privacy-sensitive

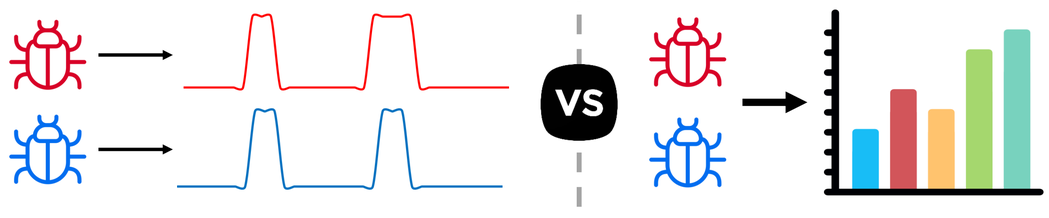


Intuitions

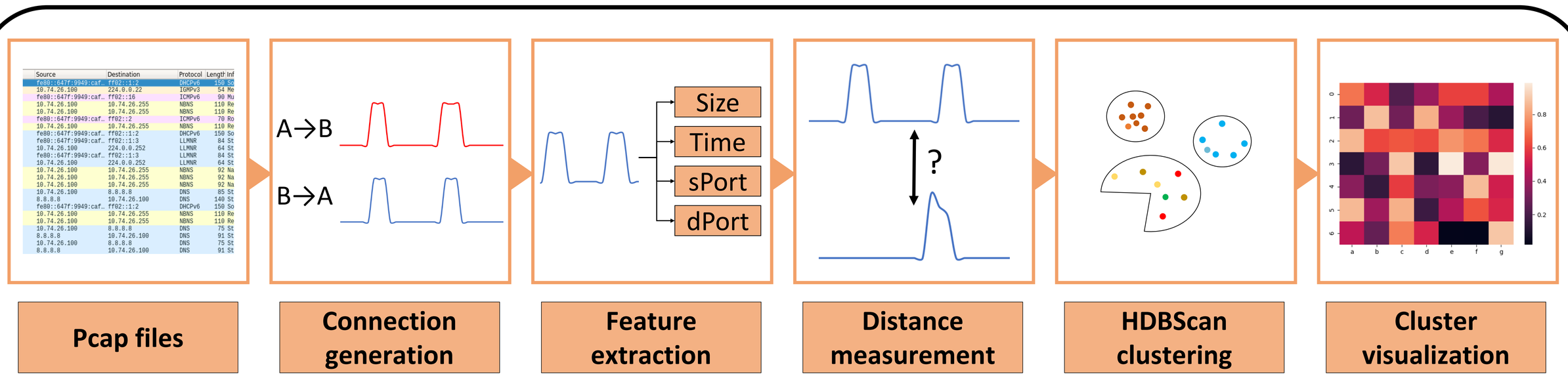
For malware samples, if same underlying infrastructure → malware behaves similarly



Analyze *sequence* of actions rather than *statistical* aggregates



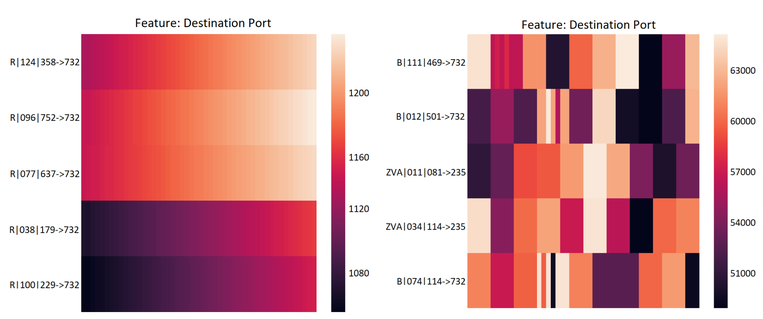
Proposed Framework



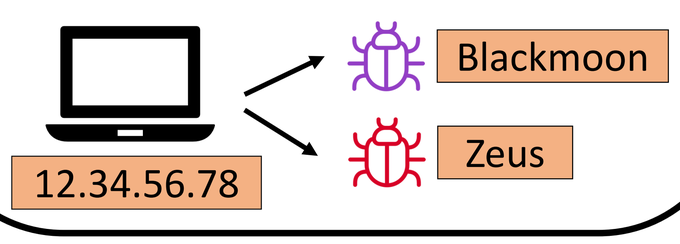
Key Results

Attacking Capabilities Detected

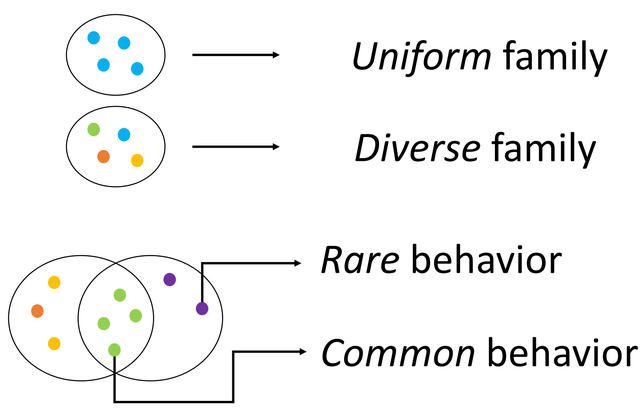
- Port scans



- Reuse of C&C server



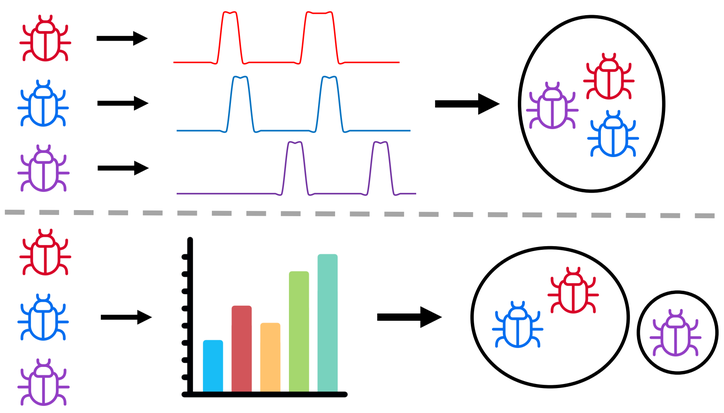
Family Behavioral Analysis



E.g., malware Gozi is diverse:

- 16 out of 18 *diverse* behaviors
- Among those, 3 *rare* behaviors

Sequential vs. Statistical Features



Statistical features perform worse:

- 78% broken clusters
- 18% more noise

* Icons made by Eucalyp, Bryn Taylor, Freepik, Vectors Market, dDara from www.flaticon.com | Statistics taken from https://www.gdatasoftware.com/blog/2018/03/30610-malware-number-2017