Henry Clausen, David Aspinall

Controlling traffic micro-structures for model probing

SecureComm 2021

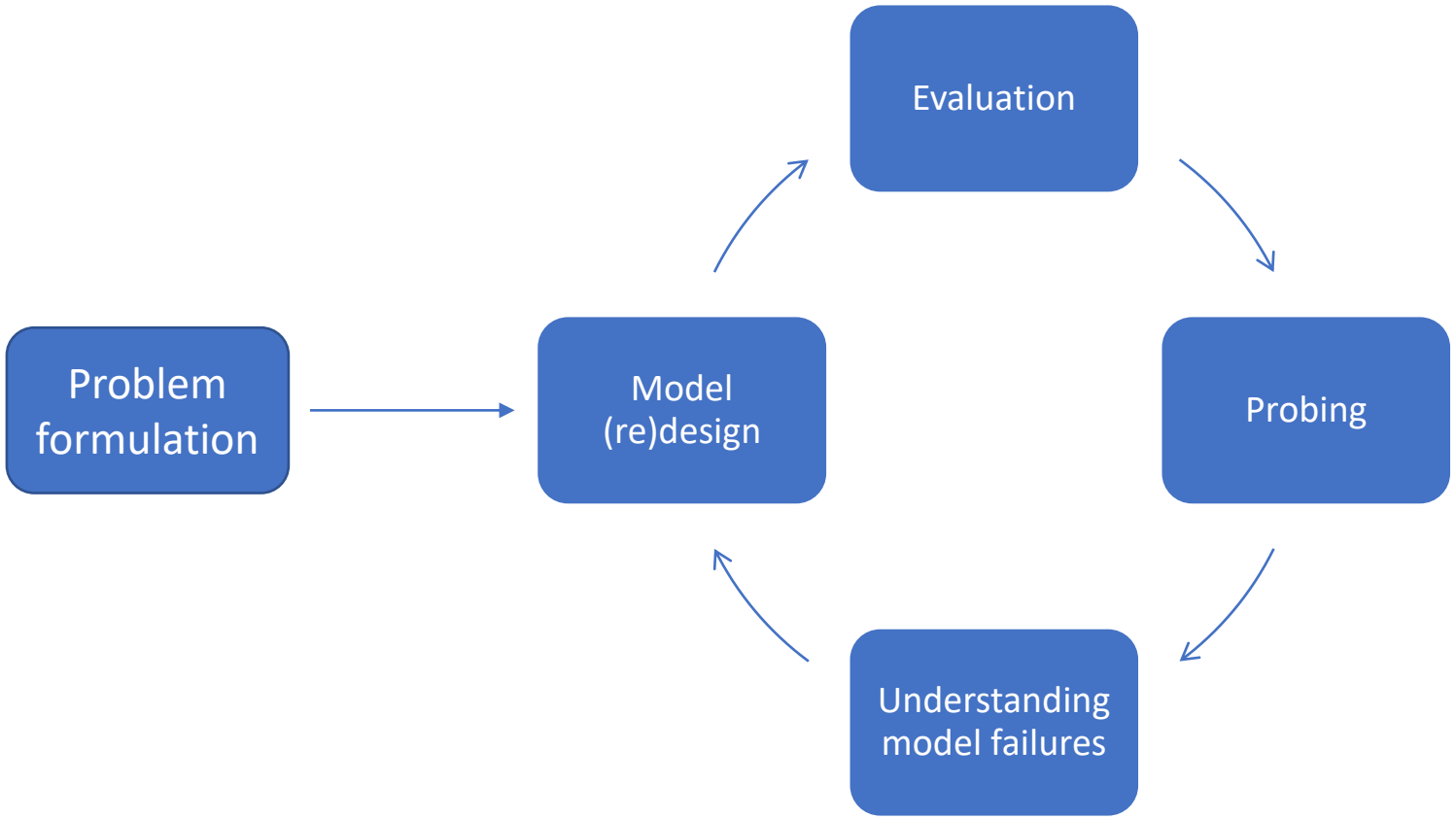
THE UNIVERSITY *of* EDINBURGH
informatics

EPSRC
Pioneering research and skills

The Alan Turing Institute

BT

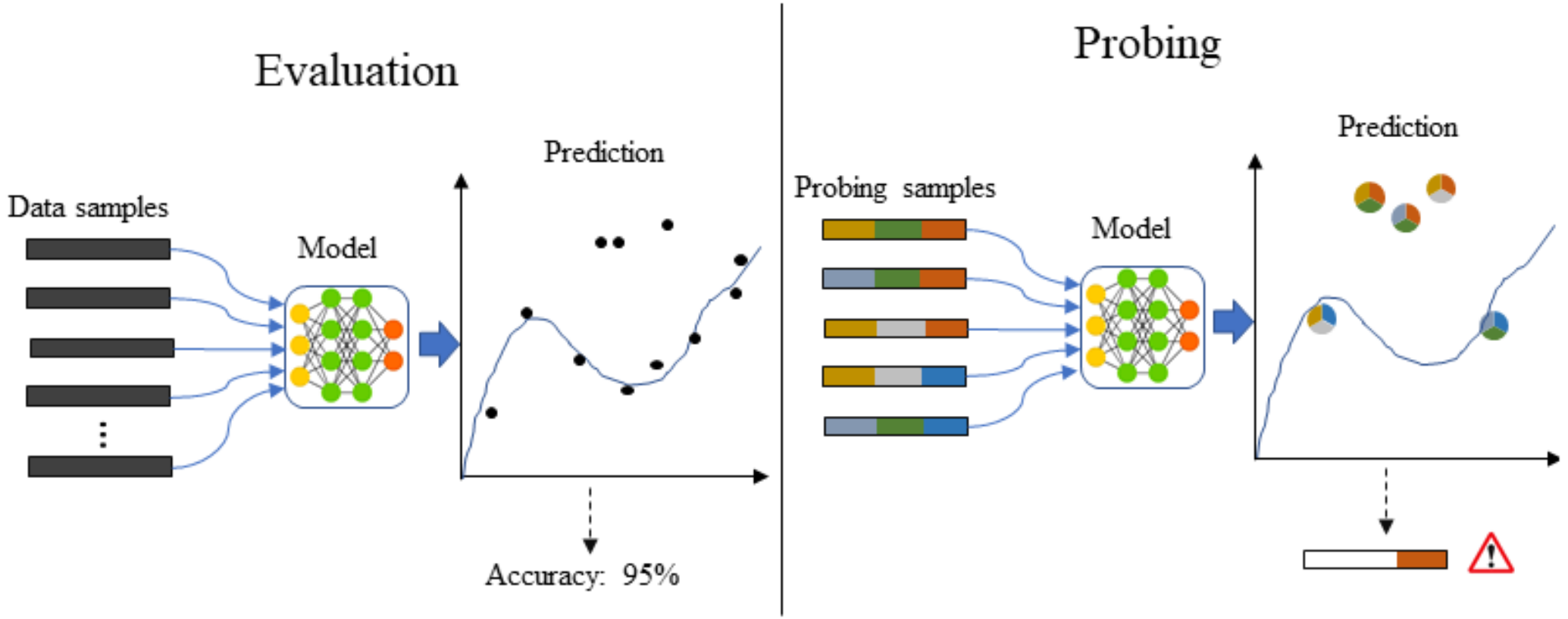
Machine learning progress



- Ambiguous words in Translation
  - Attention layer
- CNN biased to texture
  - Image stylization
- Object sizes in video enhancement
  - Multi-scale encoders

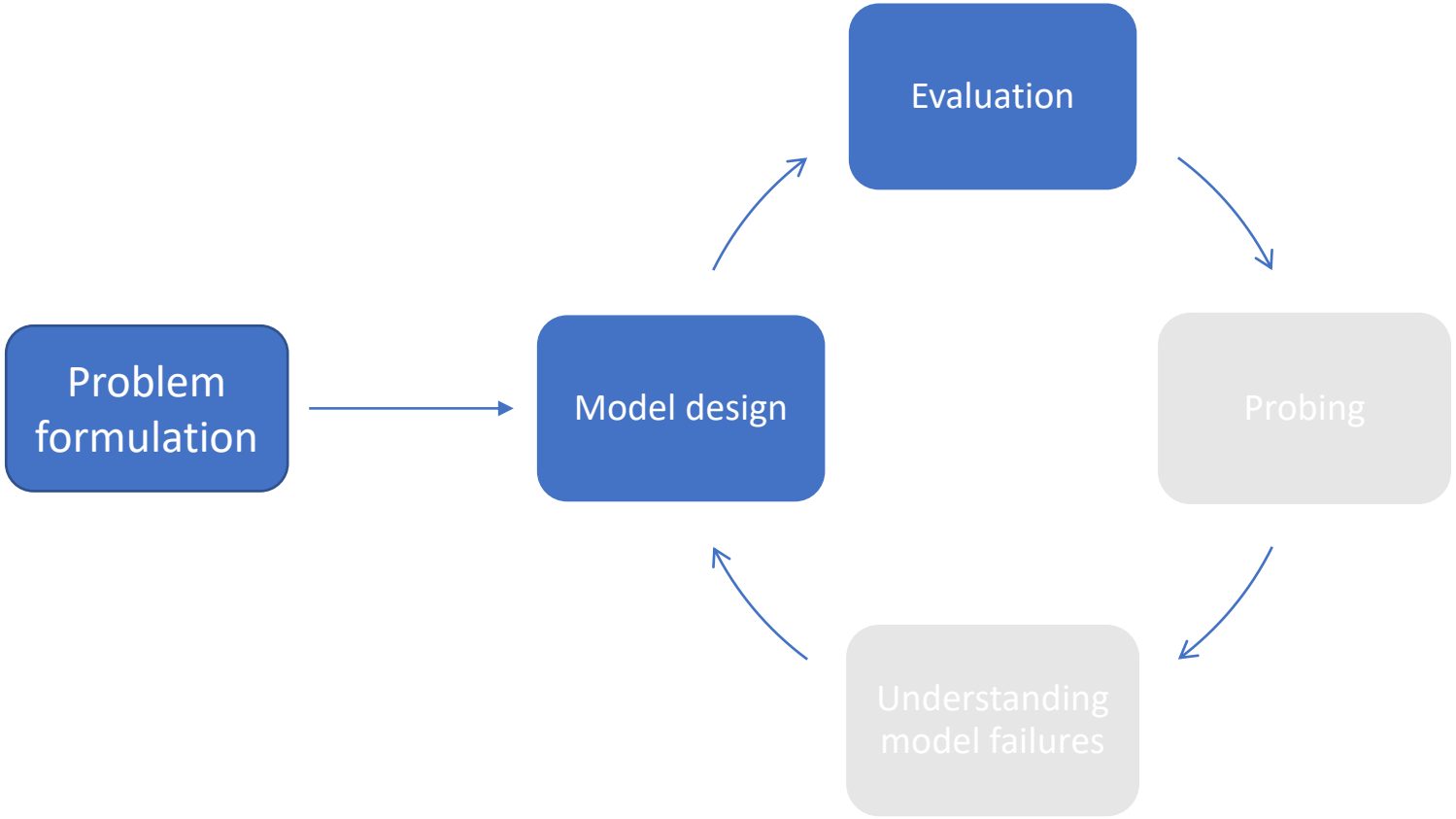
THE UNIVERSITY of EDINBURGH informatics

Model evaluation vs probing



THE UNIVERSITY of EDINBURGH
informatics

Machine learning progress in NID



NID-datasets

- Sparse labelling
- Difficult to read
- Hard to alter specific structures

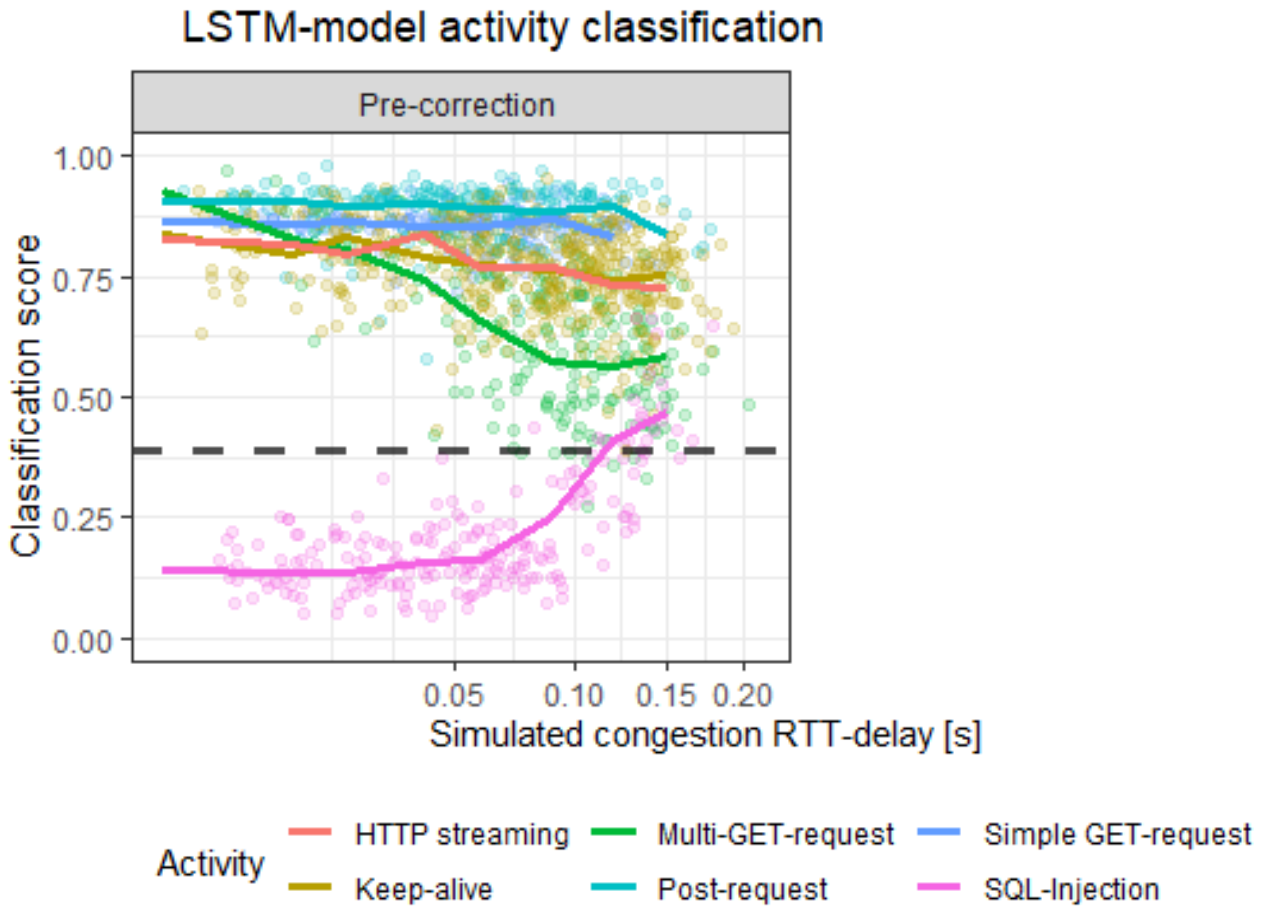
THE UNIVERSITY of EDINBURGH informatics

Structure

- NID probing example
- Influences on traffic microstructures
- DetGen: Controlling traffic microstructures
- Determinism of DetGen

NID probing example

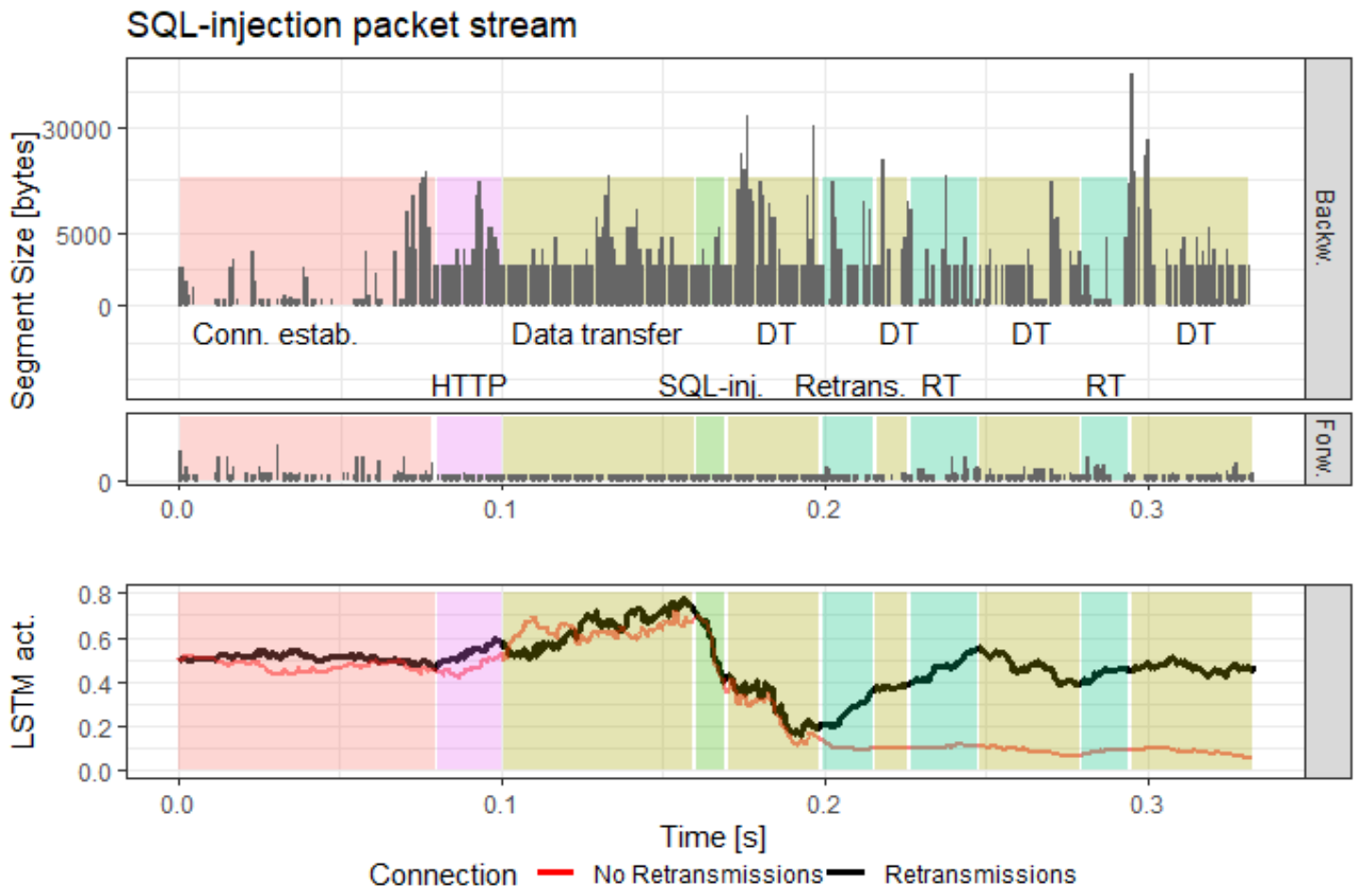
- Packet-stream LSTM-classifier by Hwang et al. 2019
- CICIDS-17 data (85%) + DetGen traffic (15%)
  - 96% DR, 2.7% FPR
- Probe with randomized labelled traffic
- Correlation between errors and latency



THE UNIVERSITY of EDINBURGH informatics

NID probing example

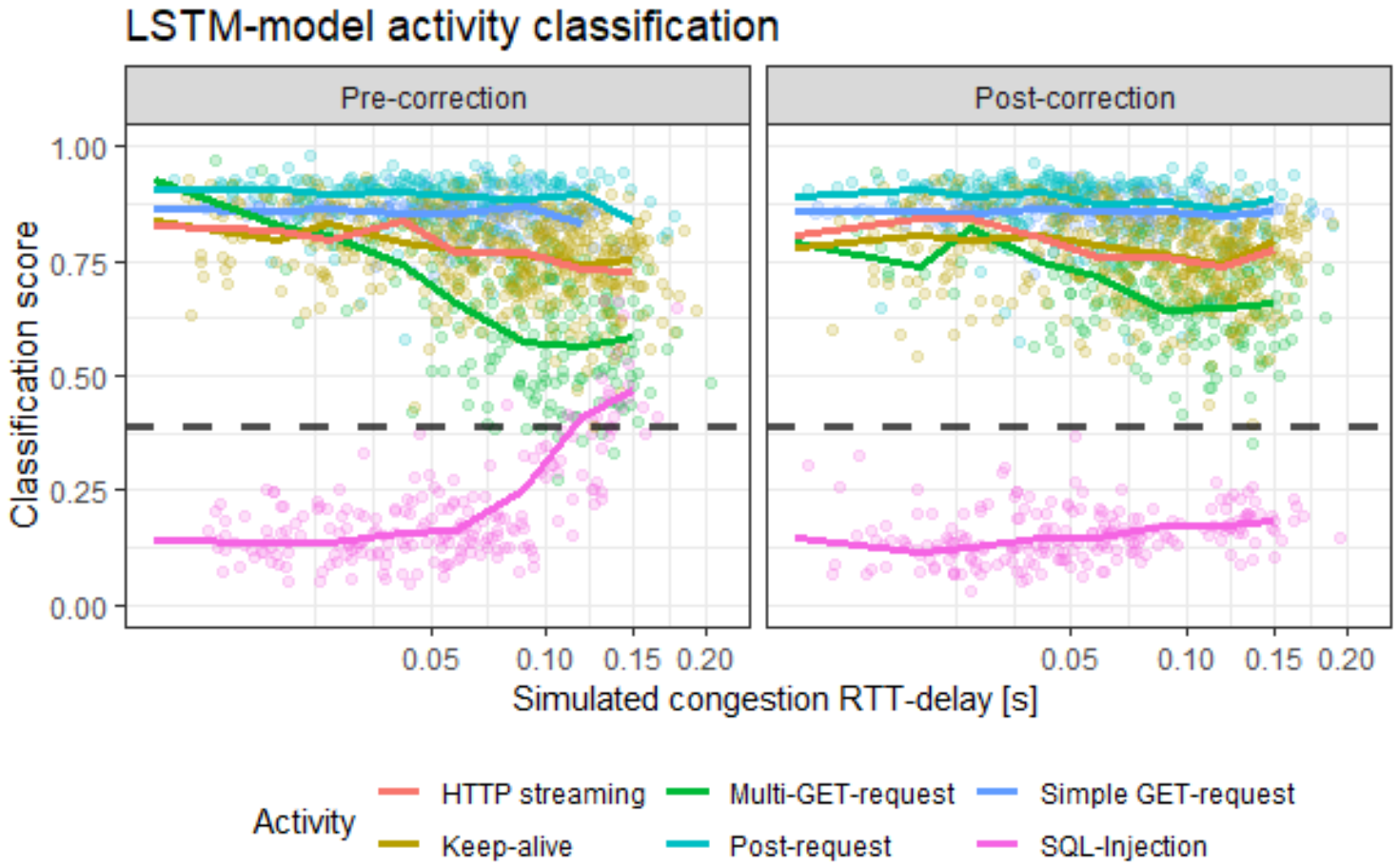
- Generate two SQL-injection connections
  - Constant microstructures
  - One with high latency
- Retransmission sequences deplete activation
- Filter RT-sequences
  → 98% DR and 0.4% FPR



THE UNIVERSITY of EDINBURGH
informatics

NID probing example

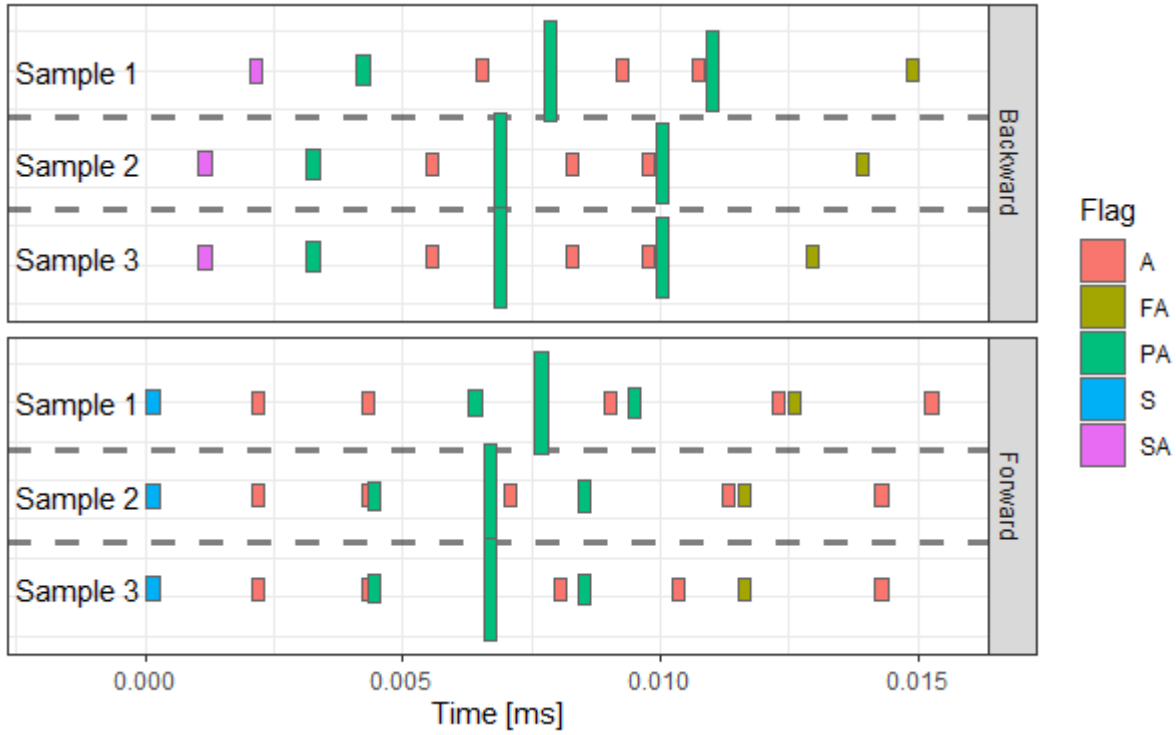
- Generate two SQL-injection connections
  - Constant microstructures
  - One with high latency
- Retransmission sequences deplete activation
- Filter RT-sequences
  → 98% DR and 0.4% FPR



THE UNIVERSITY of EDINBURGH
informatics

Microstructures: Influence factors

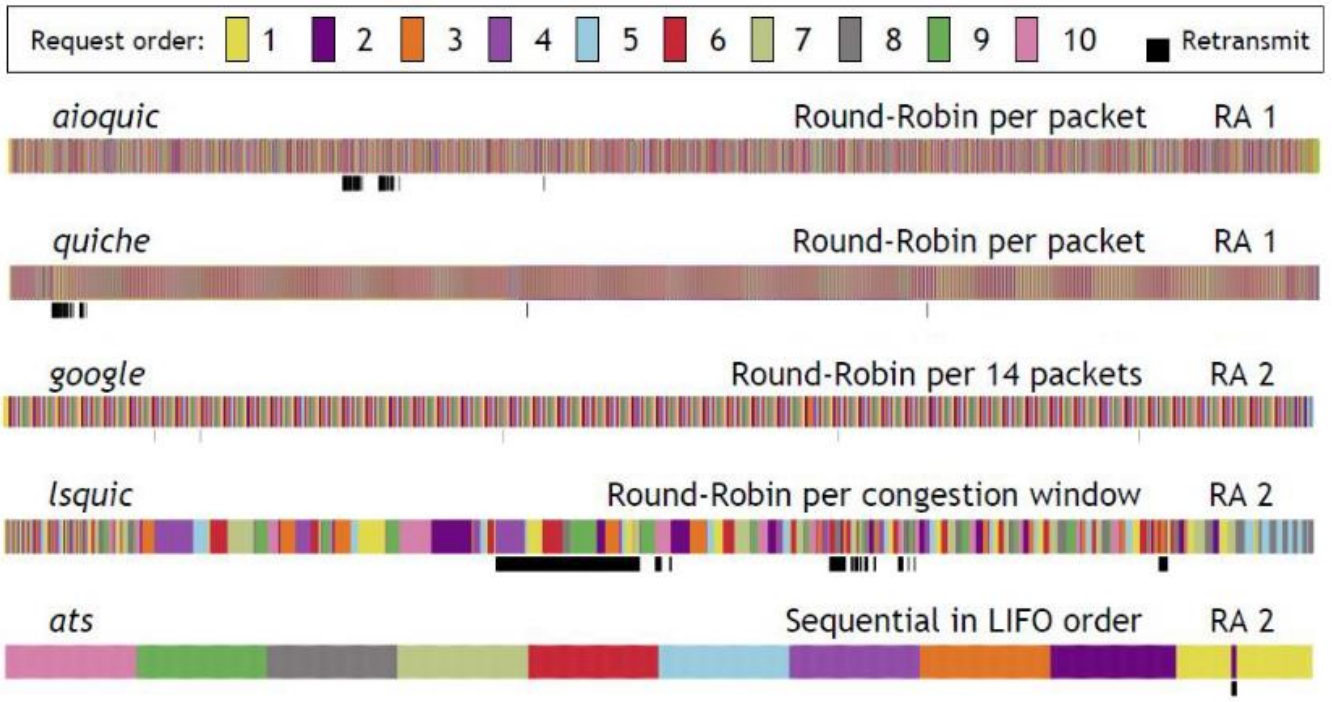
- Short-term structures at packet or connection level
- Manifest themselves in
  - IATs,
  - Packet sizes
  - Flags
  - ….
- Used by state-of-the-art models such as DeepCorr, Kitsune etc.



THE UNIVERSITY of EDINBURGH
informatics

Microstructures: Influence factors

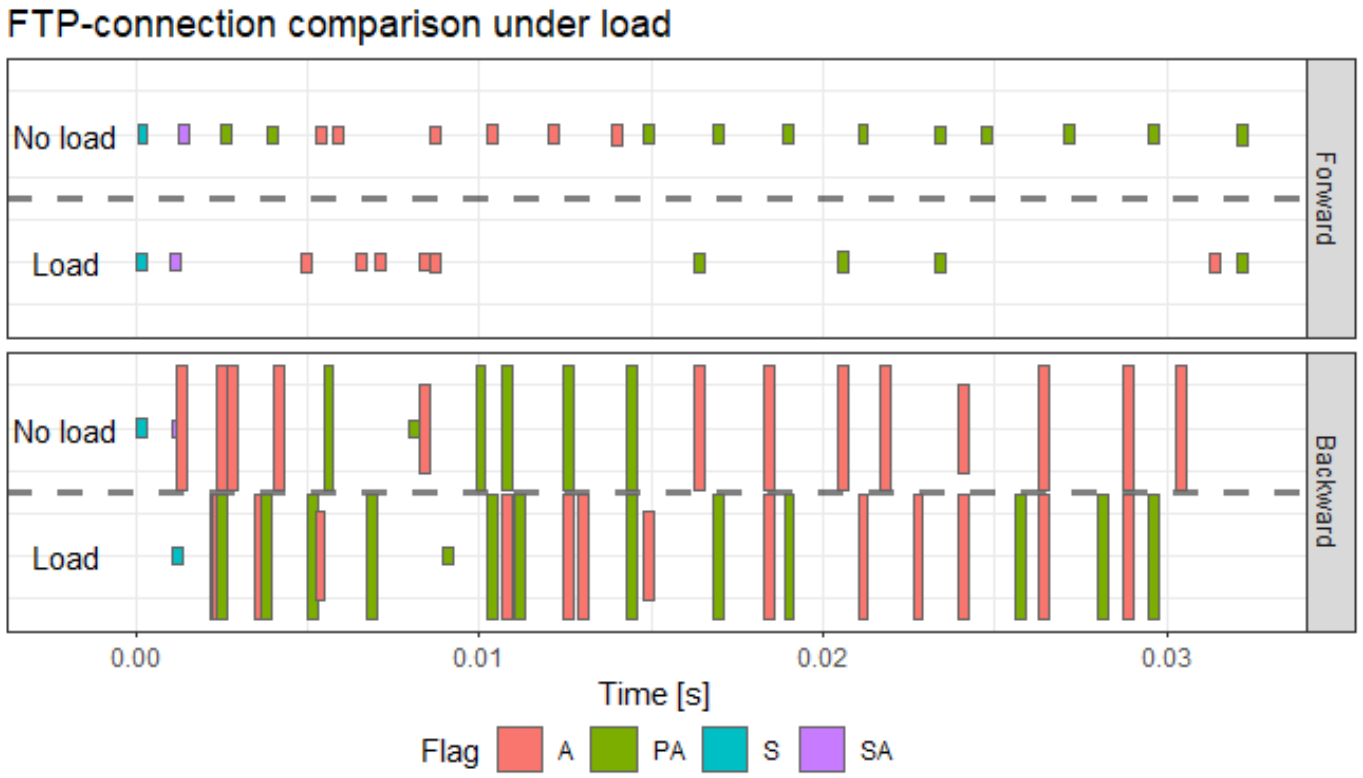
- Application/task
- Implementation version
- Network congestion
- Host load
- Caching/repetition
- Background traffic



Marx et al. 2020

Microstructures: Influence factors

- Application/task
- Implementation version
- Network congestion
- Host load
- Caching/repetition
- Background traffic



THE UNIVERSITY of EDINBURGH
informatics

Microstructures: Influence factors

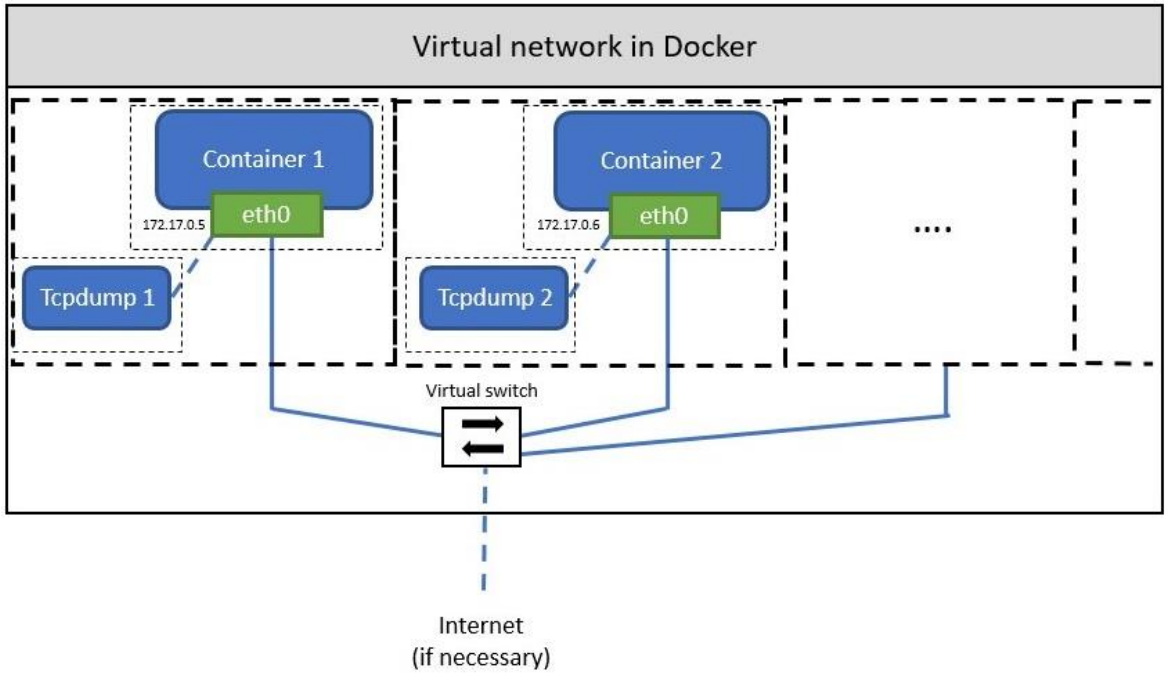
- Application/task
- Implementation version
- Network congestion
- Host load
- Caching/repetition
- Background traffic

Time	Source-IP	Destination-IP	**Dest. Port**
13:45:56.8	192.168.10.9	192.168.10.50	**21**
13:45:56.9	192.168.10.9	192.168.10.50	**10602**
13:45:57.5	192.168.10.9	69.168.97.166	**443**
13:45:59.1	192.168.10.9	192.168.10.3	**53**
13:46:00.1	192.168.10.9	205.174.165.73	**8080**

DetGen: Controlling microstructures

Scope:

- Precise control over traffic influence factors
- Ground-truth labels on traffic origins
- Scalability and modularity

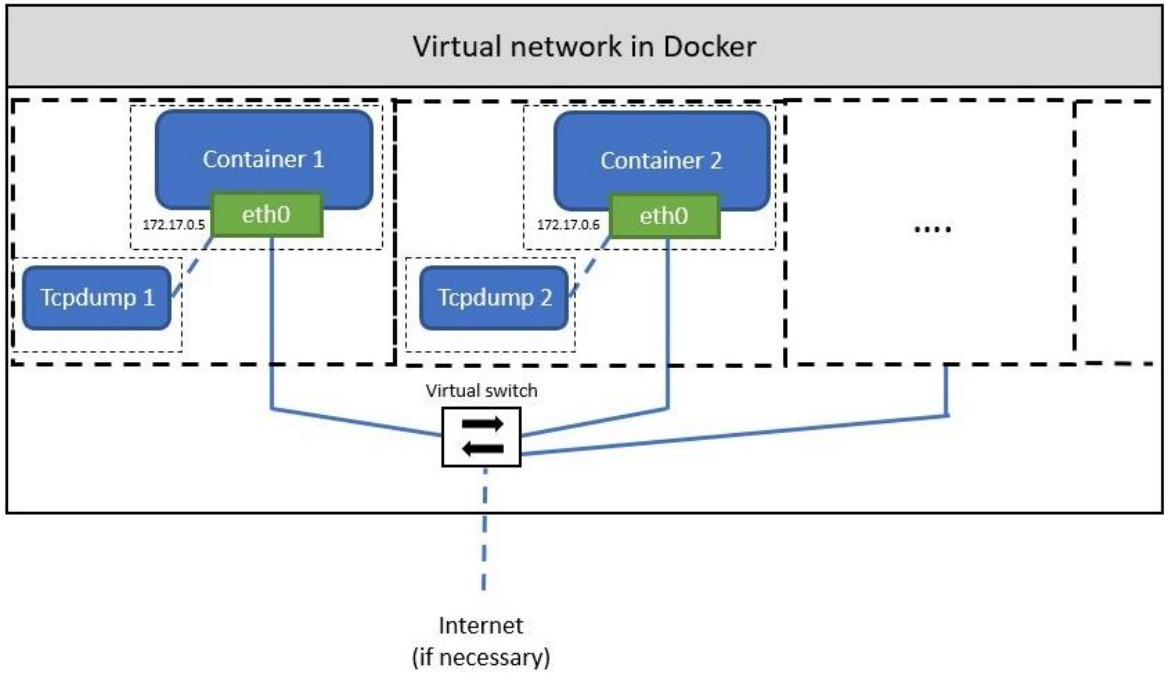


THE UNIVERSITY of EDINBURGH
informatics

DetGen: Controlling microstructures

Design:

- Scripting of diverse scenarios + subscenarios
- Isolation through containerization
- Simulation of external effects
- Randomisation at every stage



More details:

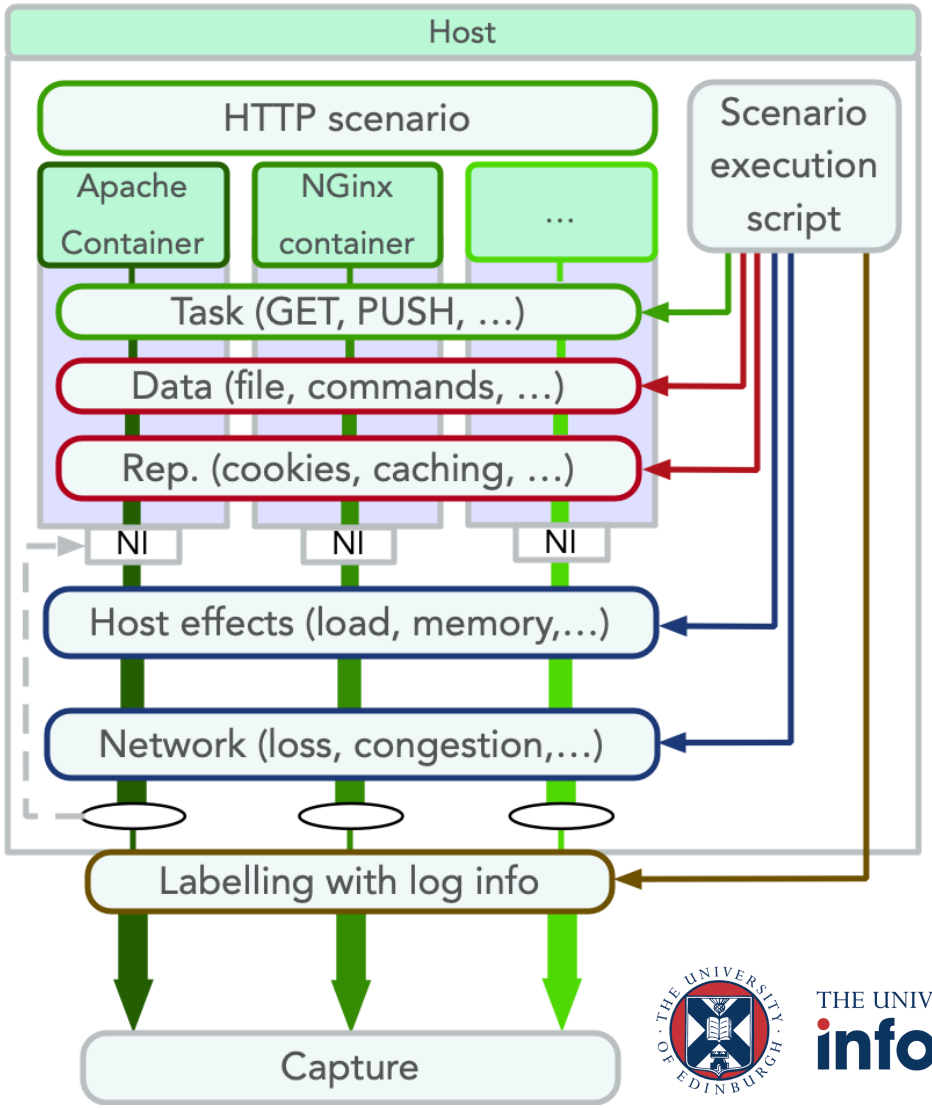
- “Traffic generation using containerization”, 2019
- github.com/detlearsom/DetGen

THE UNIVERSITY of EDINBURGH
informatics

DetGen: Controlling microstructures

Design:

- Scripting of diverse scenarios + subscenarios
- Isolation through containerization
- Simulation of external effects
- Randomisation at every stage



More details:

- "Traffic generation using containerization", 2019
- github.com/detlearsom/DetGen

THE UNIVERSITY of EDINBURGH informatics

Determinism of DetGen

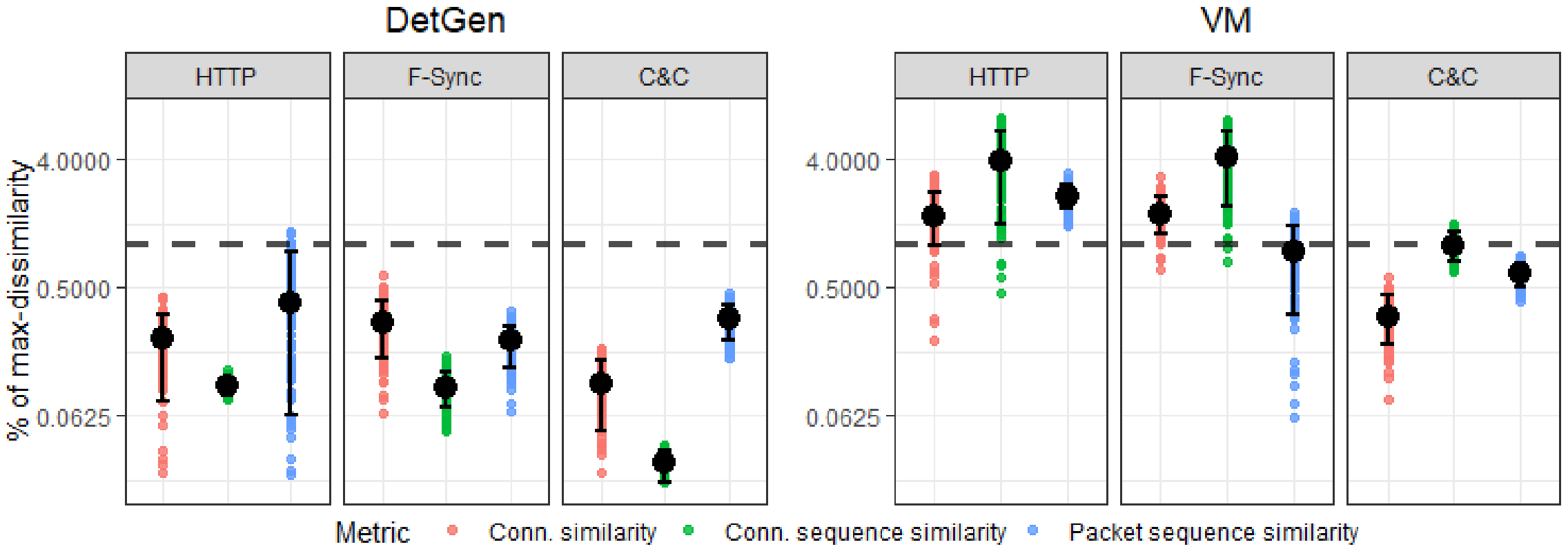
- How well can DetGen control traffic influence factors?
- Is containerisation improving control?

Experiment:

- Generate traffic with constant settings
- Measure sample similarity

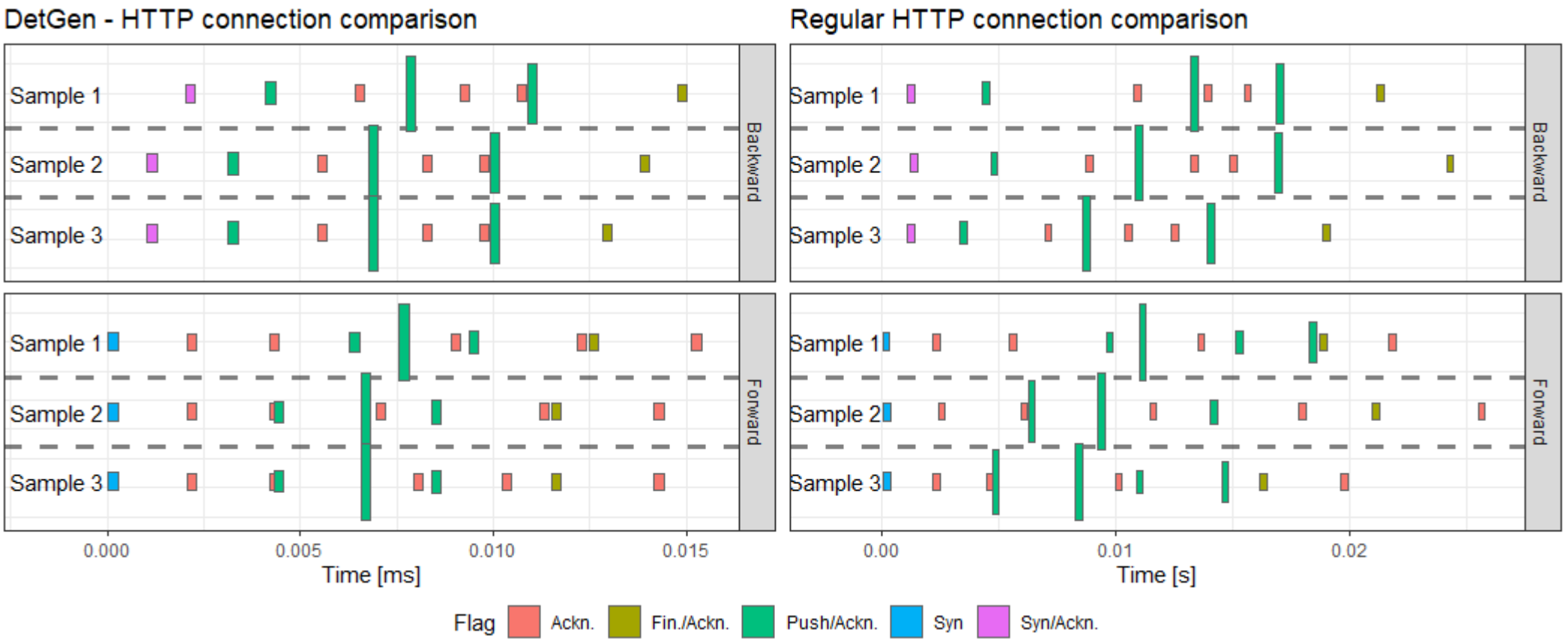
- Metrics:
  - Connection (size, IATs, etc.)
  - Packet seq.
  - Connection seq.

Determinism of DetGen



THE UNIVERSITY of EDINBURGH
informatics

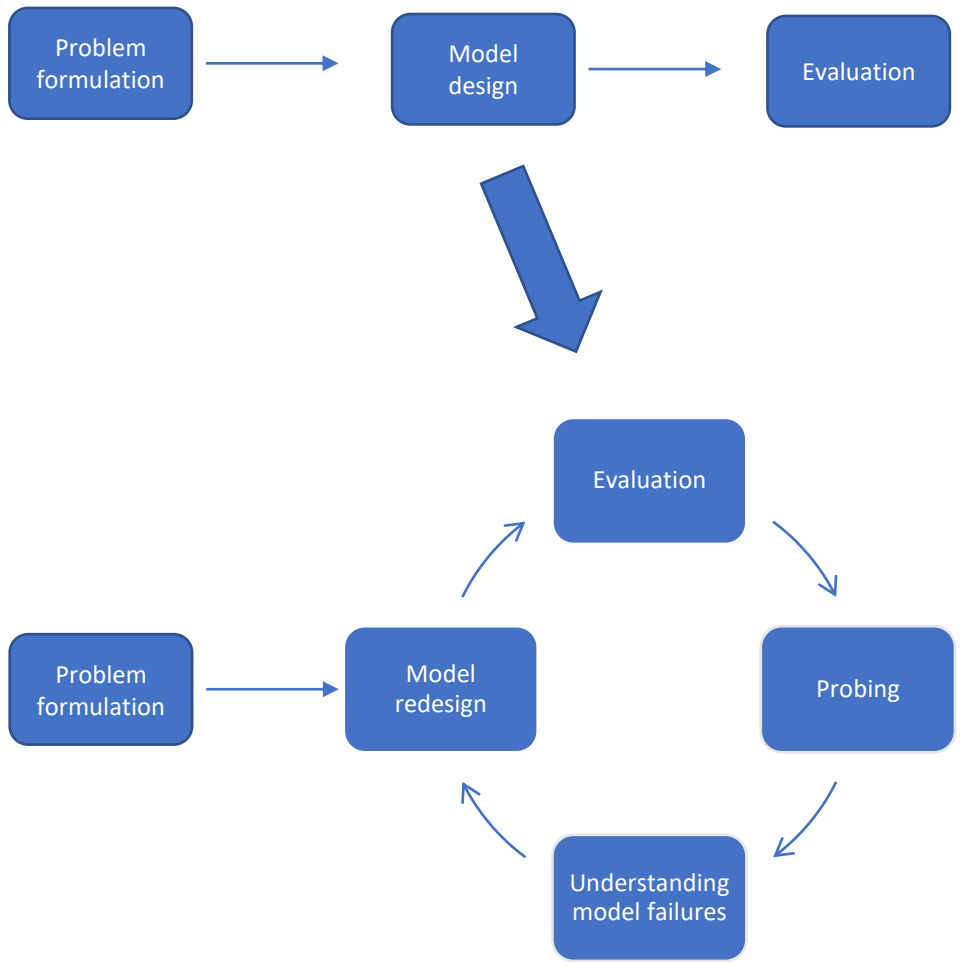
Determinism of DetGen



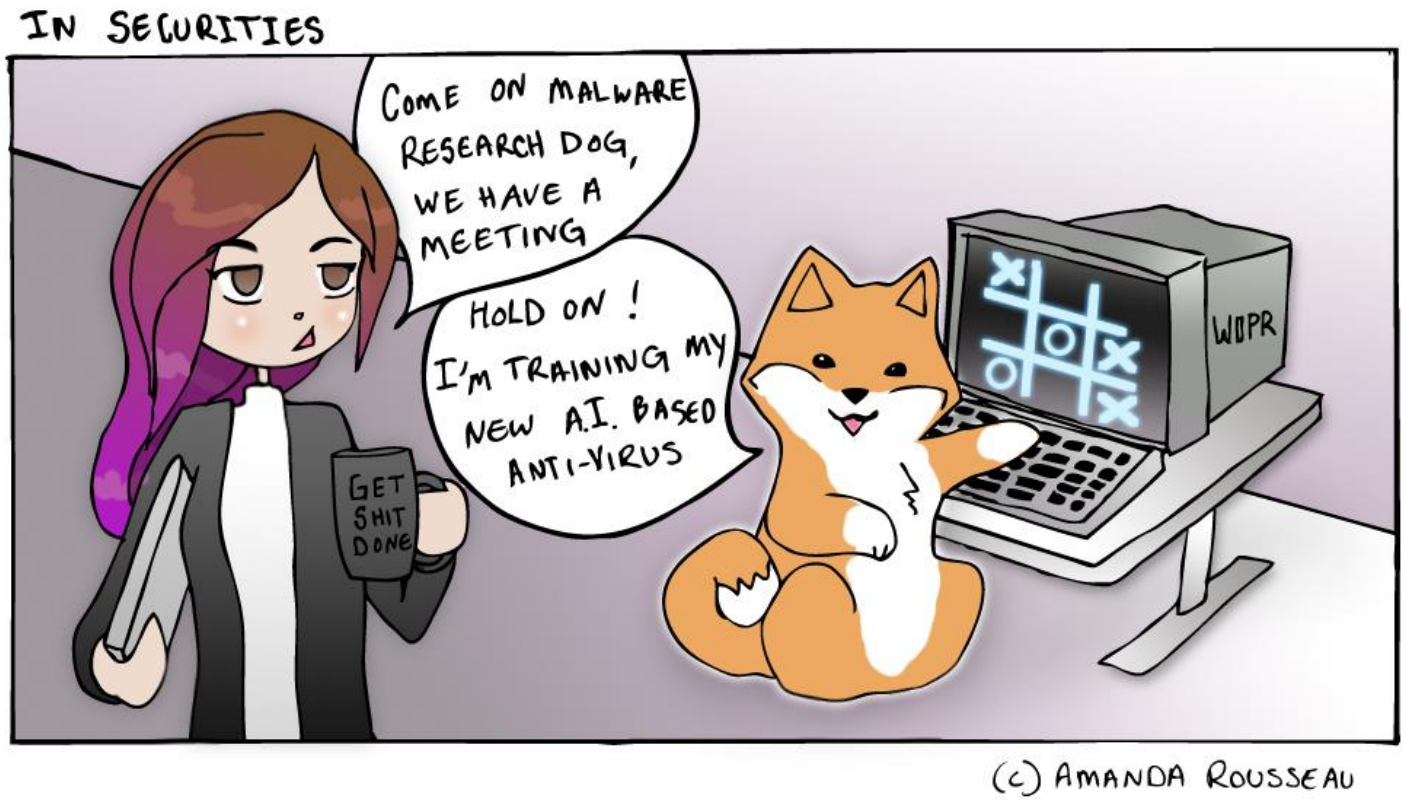
THE UNIVERSITY of EDINBURGH
informatics

Conclusion

- Targeted probing can identify model failures
- Labelling for misclassification correlation
- Control traffic microstructures
  - Randomise for broad probing
  - Reduce variations for close examination
- github.com/detlearsom/DetGen



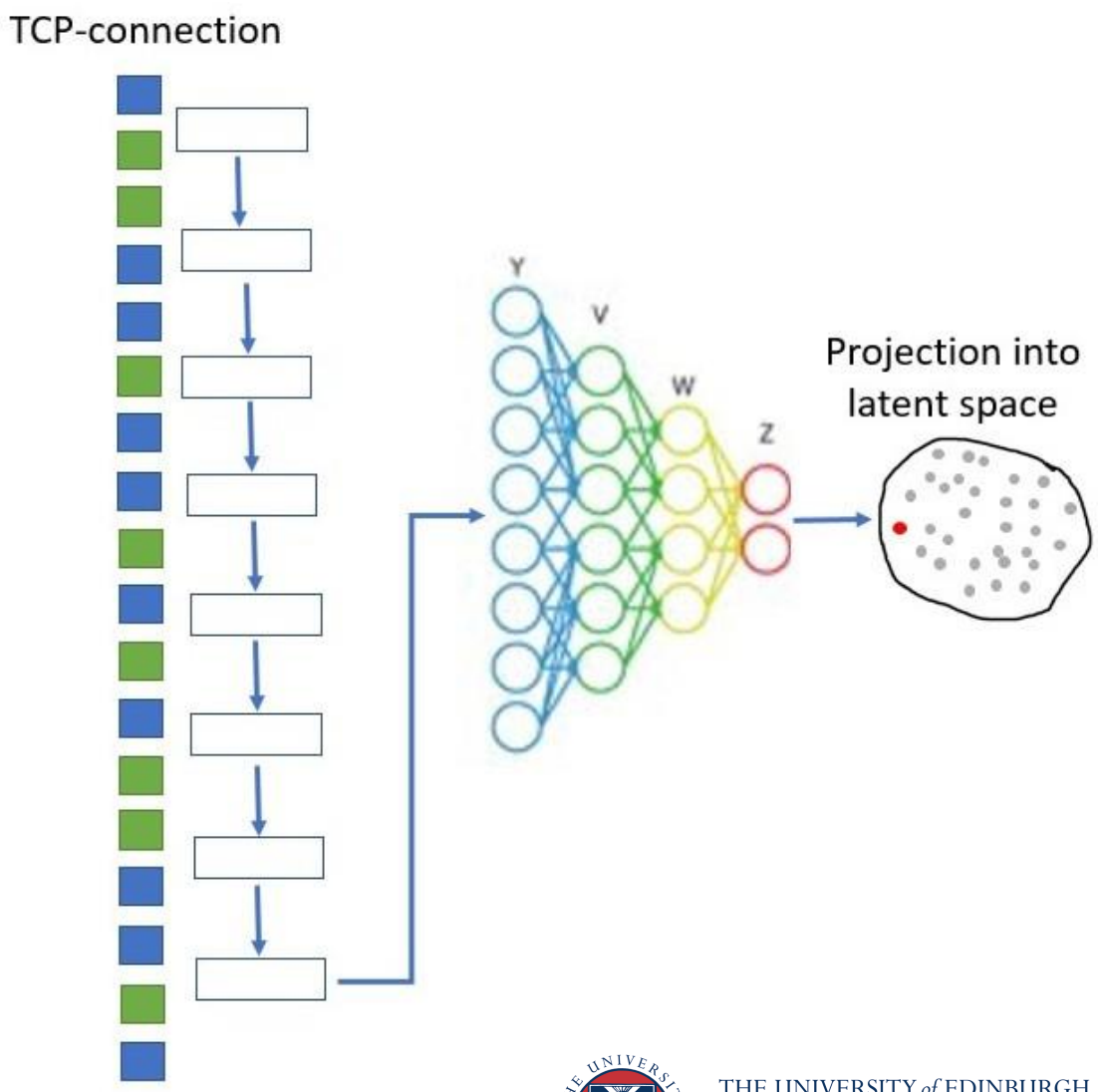
Thank you for your attention!



THE UNIVERSITY of EDINBURGH
informatics

Projection sensitivity

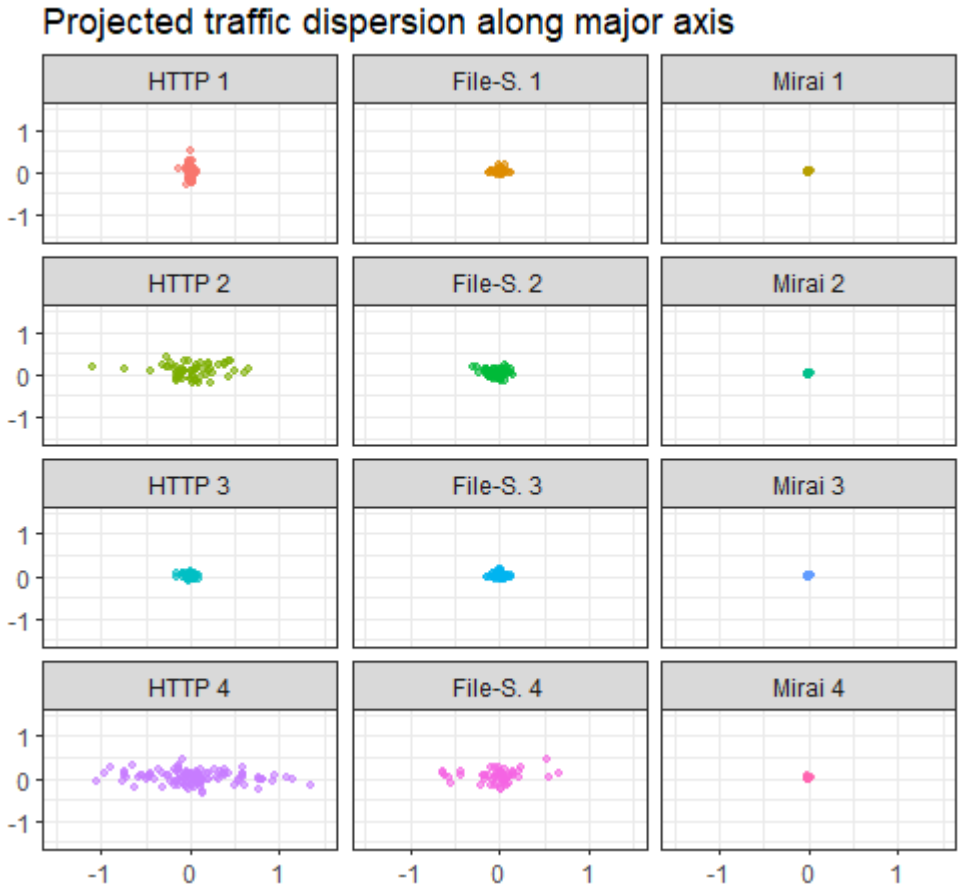
- Kitsune 2018
  - Seq-encoding for anomaly detection
  - Botnet, man-in-middle, Brute-force,…
- Traffic groups with constant settings
- Projections should be consistent
- Sensitive to
  - connection IATs
  - half-open connections



THE UNIVERSITY of EDINBURGH informatics

Projection sensitivity

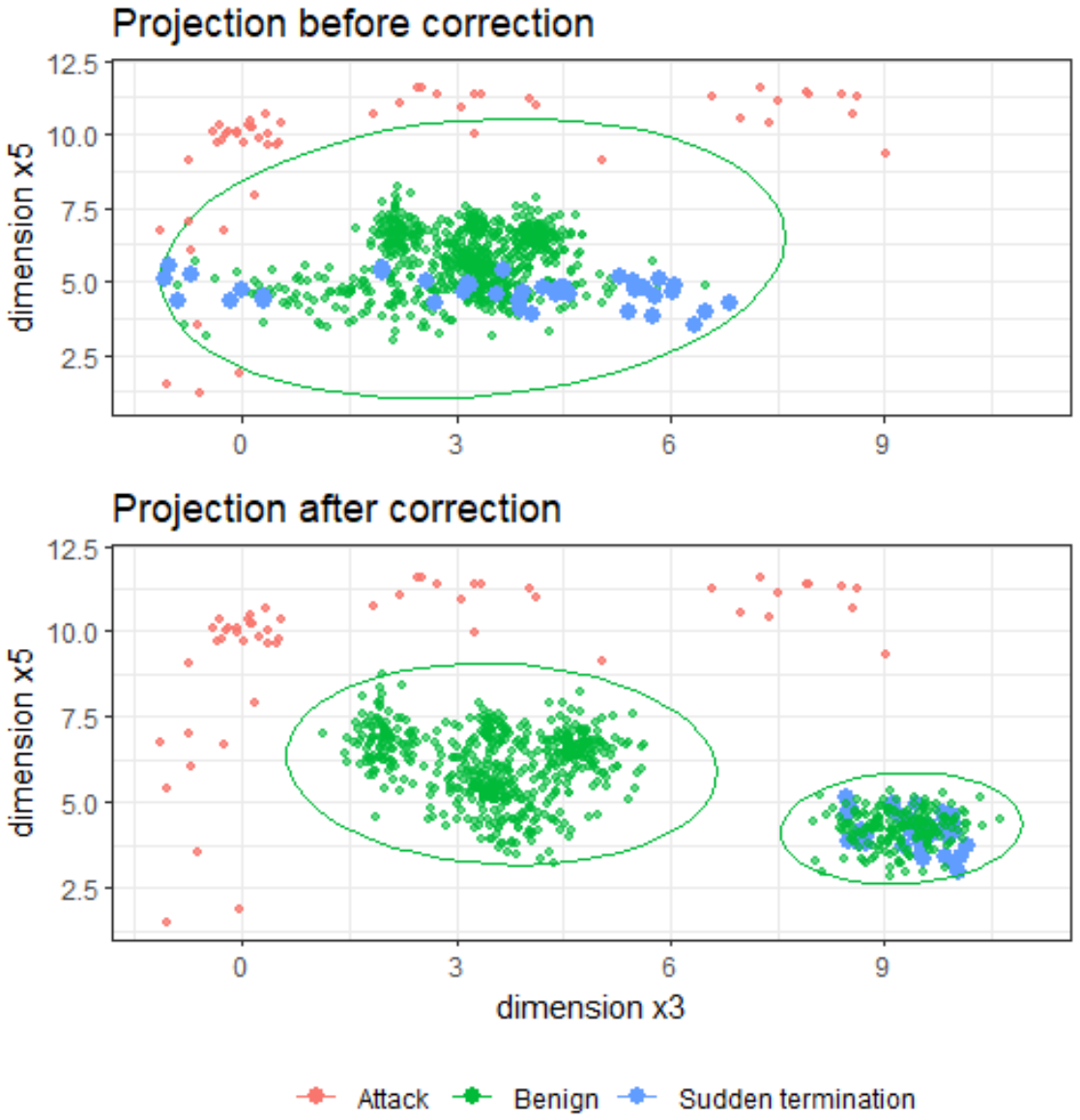
- Kitsune 2018
  - Seq-encoding for anomaly detection
  - Botnet, man-in-middle, Brute-force,…
- Traffic groups with constant settings
- Projections should be consistent
- Sensitive to
  - connection IATs
  - half-open connections



THE UNIVERSITY of EDINBURGH
informatics

Projection sensitivity

- Kitsune 2018
  - Seq-encoding for anomaly detection
  - Botnet, man-in-middle, Brute-force,…
- Traffic groups with constant settings
- Projections should be consistent
- Sensitive to
  - connection IATs
  - half-open connections



THE UNIVERSITY of EDINBURGH
informatics

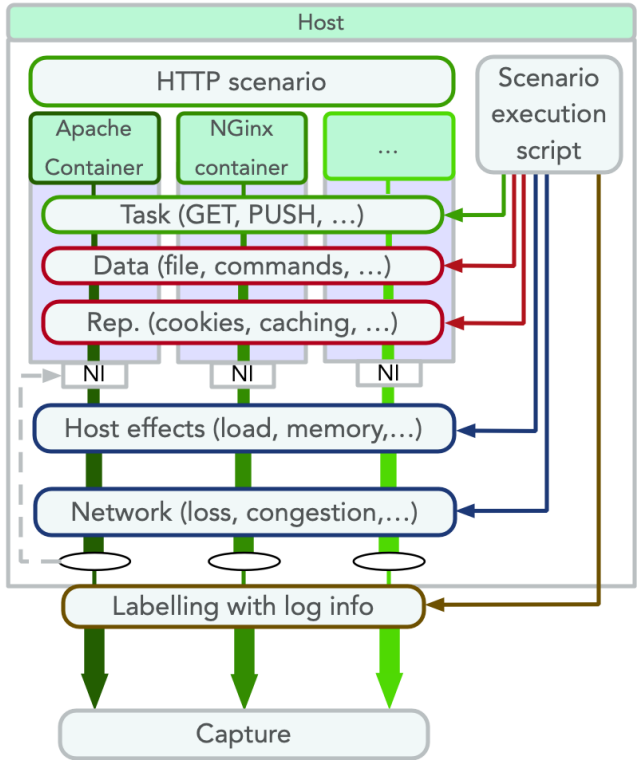
Controlling traffic microstructures

DetGen Clausen et al., SecureComm 2021

- Traffic generation tool

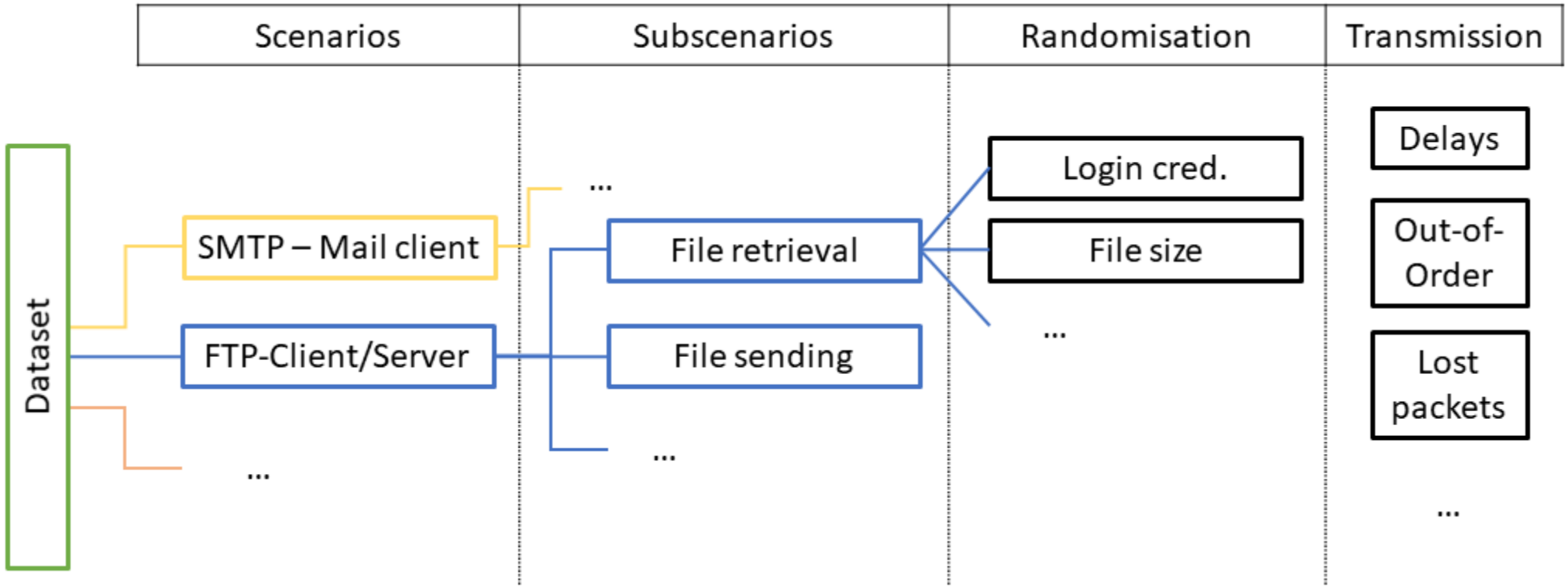
- Controlling and labelling microstructures:
  - Performed task/application
  - Implementations
  - Congestion
  - Failures
  - …

- github.com/detlearsom/DetGen



THE UNIVERSITY of EDINBURGH informatics

DetGen: Controlling microstructures



THE UNIVERSITY of EDINBURGH
informatics