Towards Behavior Grammar-Driven IoT Network Traffic Generation using MUD Specifications


Shuai Zhang*


Shayan Azizi*


Aditya Joshi+


Gustavo Batista+


Hassan Habibi Gharakheili*

*: *Electrical Engineering and Telecommunications, UNSW Sydney, Australia*
+: *Computer Science and Engineering, UNSW Sydney, Australia*

UNSW SYDNEY

Expanded Network Attack Surfaces

- Rapid growth of IoT devices has increased size and complexity of networks
  - Led to *expanded attack surfaces*

- To maintain secure network operation, *robust methodologies* are needed for
  - analyze, characterize and verify IoT device behaviors

- ML models are employed for traffic analysis and monitoring
  - but their performance rely on training data → *traffic data* is needed

Gap

- Obtaining IoT traffic datasets with high *diversity* and *fidelity* remains a challenge
  - Limited public data
    - expensive and time-consuming to build testbed accurately reflects real-world scenarios
    - real traffic datasets could contain private information

- Our objectives
  - *generate synthetic yet realistic* traffic datasets using formal *descriptions* and *models* of network *behaviors*.

A Possible Solu

Internet Engineering Task Force (IETF)
Request for Comments: 8520
Category: Standards Track
ISSN: 2070-1721

E. Lear
Cisco Systems
R. Droms
Google
D. Romascanu
March 2019

Manufacturer Usage Description Specification

- IETF MUD provides fou of IoT behaviors:
  - provide
  - but not

& endpoint (e.g., UDP/123 →

Sample MUD:

```
"ietf-access-control-list:access-lists" : {
  "acl" : [ {
    "name" : "from-ipv4-samsungsmartcam",
    "type" : "ipv4-acl-type",
    "aces" : {
      "ace" : [ {
        "name" : "from-ipv4-samsungsmartcam-0",
        "matches" : {
          "ipv4" : {
            "protocol" : 17,
            "ietf-acldns:dst-dnsname" : "europe.pool.ntp.org"
          },
          "udp" : {
            "destination-port" : {
              "operator" : "eq",
              "port" : 123
            }
          }
        },
        "actions" : {
          "forwarding" : "accept"
        }
      }, {
```

UDP

server identity

service port (NTP)

action permitted

Our Contributions

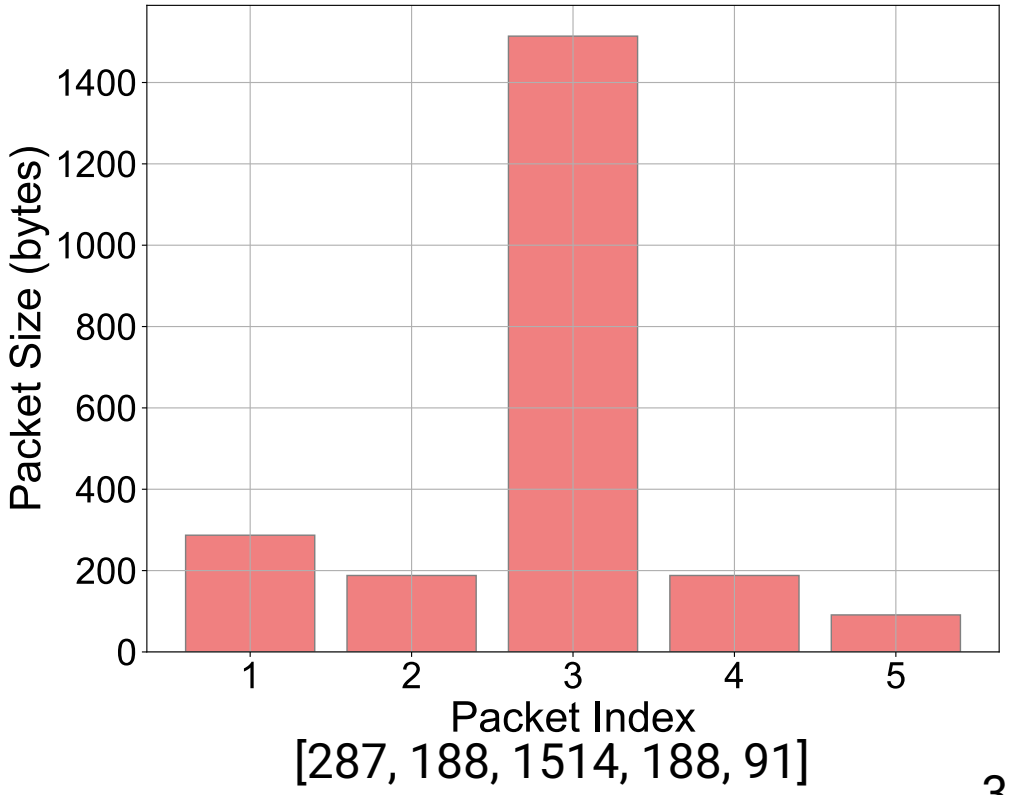
- We develop a common grammar model to describe behaviors of MUD service flows;

- We build SynGen, an automated tool that takes grammar files and generates realistic traffic in a containerized, virtualized network;

- We evaluate the efficacy of our grammar files for 19 MUD service flows.

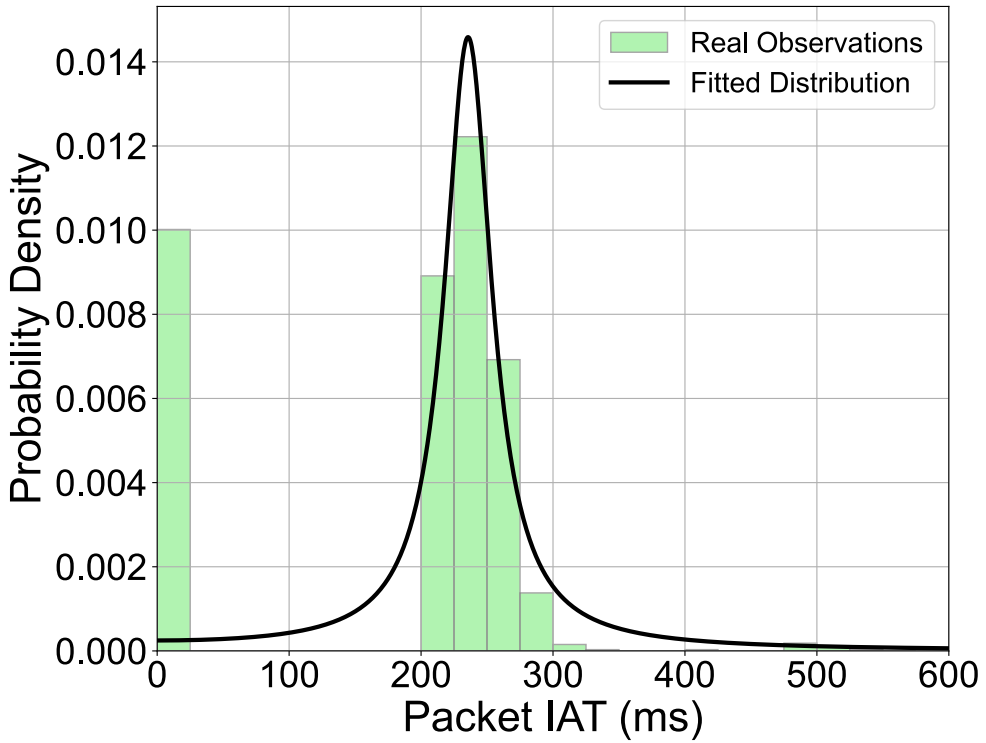
AmazonEcho:
TCP/443 ➔ dcape-na.amazon.com

packet sizes per flow



[287, 188, 1514, 188, 91]

packet IAT per flow

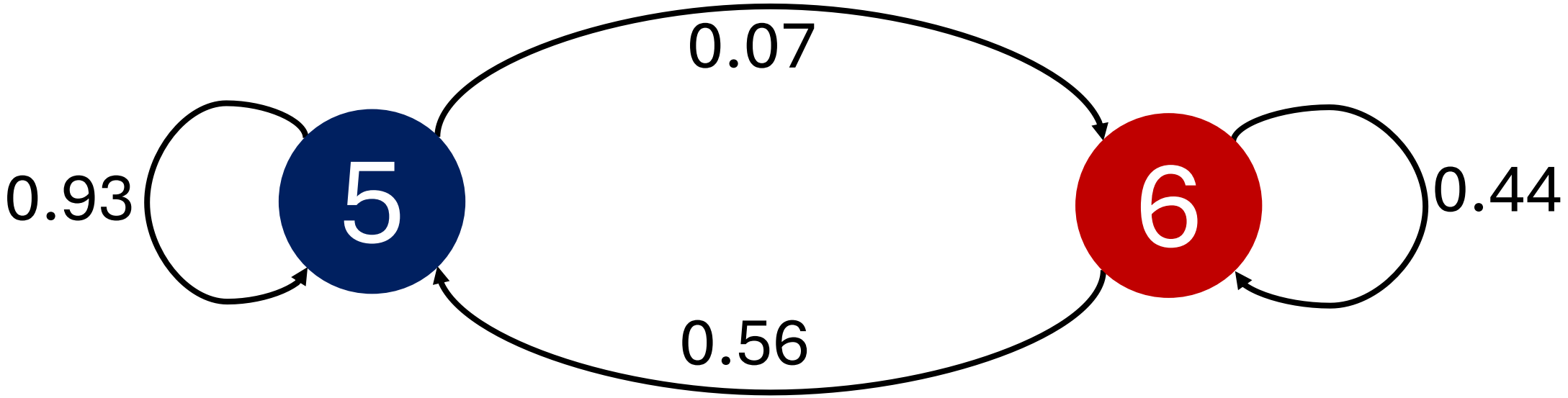


327 samples

UNSW SYDNEY

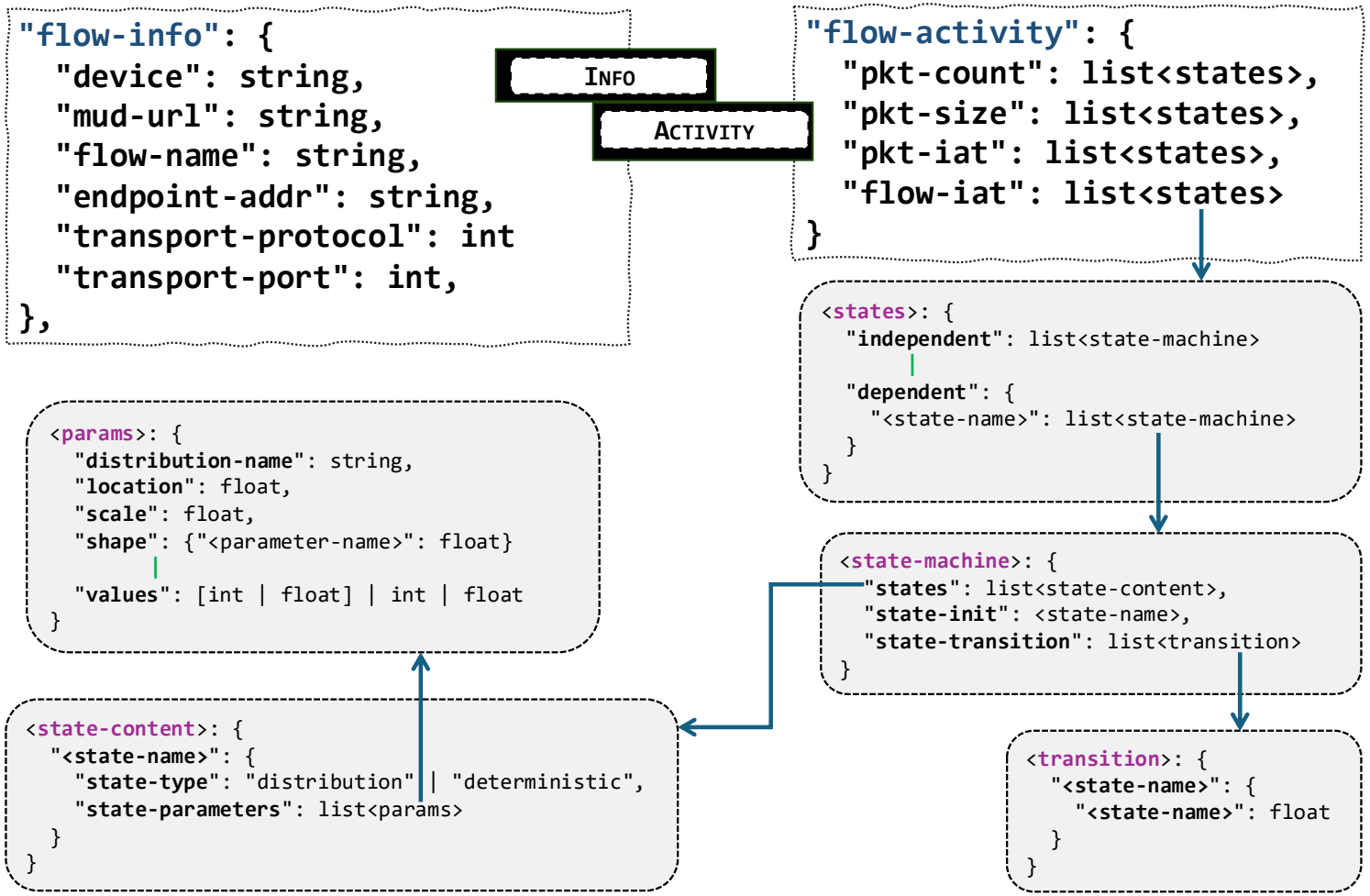
Awair Air Quality:
`TCP/443` ➔ `ota.awair.is`

packet counts per flow



82 samples

(1) Data Grammar for MUD Service Flows Behaviors



Realizing Behavior Data Grammar files

- We created the behavior grammar files for 19 MUD service flows extracted from 10 MUD files and corresponding real PCAPs


```
"pkt-count": {
  "independent": {
    "states": {
      "state-1-pkt-count": {
        "state-type": "deterministic",
        "state-parameters": {
          "values": 5
        }
      },
      "state-2-pkt-count": {
        "state-type": "deterministic",
        "state-parameters": {
          "values": 6
        }
      }
    },
    "state-init": "state-1-pkt-count",
    "state-transition": {
      "state-1-pkt-count": {
        "state-1-pkt-count": 0.9315,
        "state-2-pkt-count": 0.0685
      },
      "state-2-pkt-count": {
        "state-1-pkt-count": 0.5556,
        "state-2-pkt-count": 0.4444
      }
    }
  }
}
```



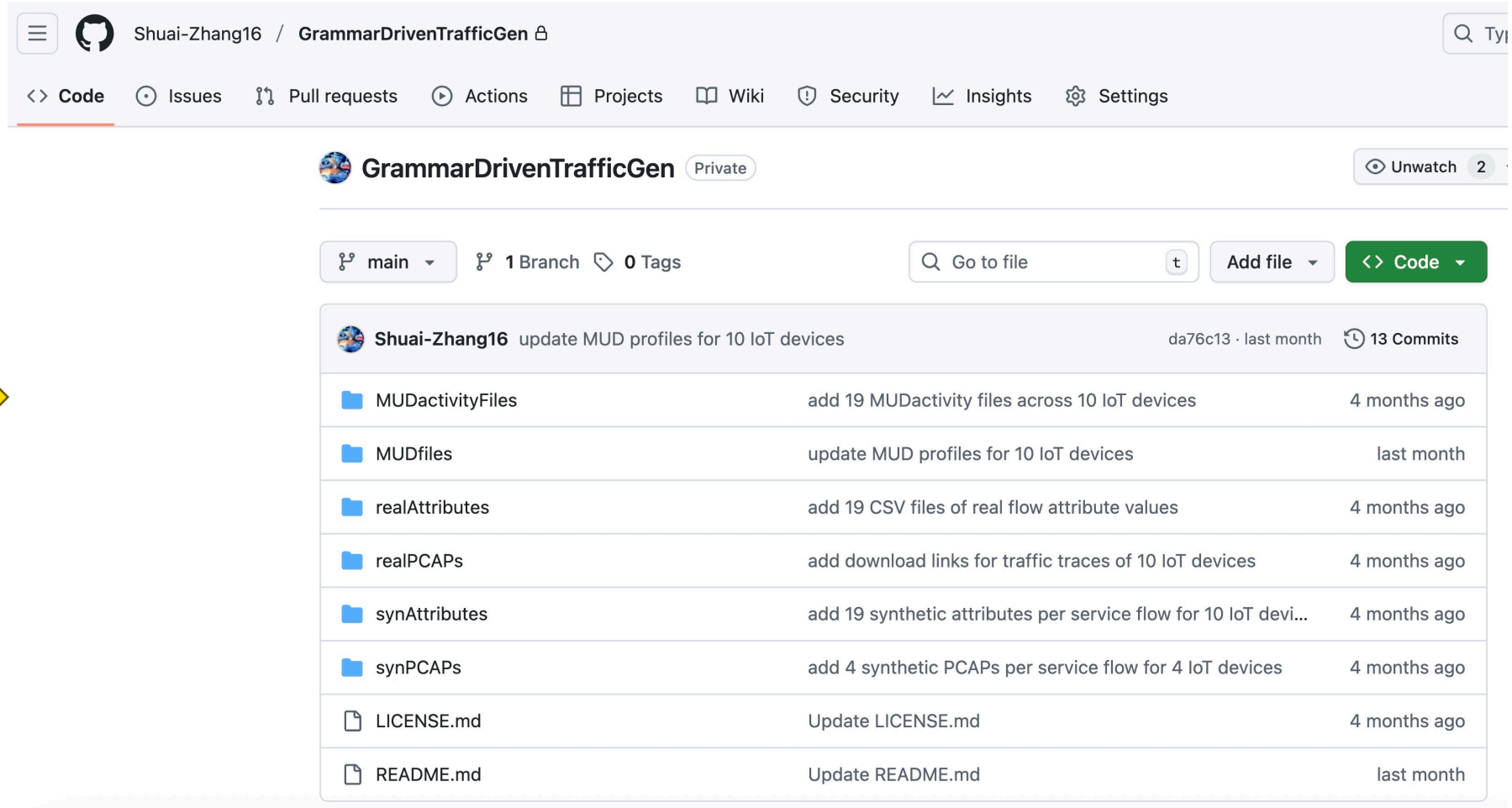
```
"flow-iat": {
  "independent": {
    "states": {
      "state-1-flow-iat": {
        "state-type": "distribution",
        "state-parameters": {
          "distribution-name": "Folded Normal",
          "location": 3004.1552508967943,
          "scale": 252.16620469306287,
          "shape": {
            "c": 0.9579907644266459
          }
        }
      }
    },
    "state-init": "state-1-flow-iat",
    "state-transition": {
      "state-1-flow-iat": {
        "state-1-flow-iat": 1.0
      }
    }
  }
}
```


Awair Air Quality flow-2:
TCP/443 → ota.awair.is

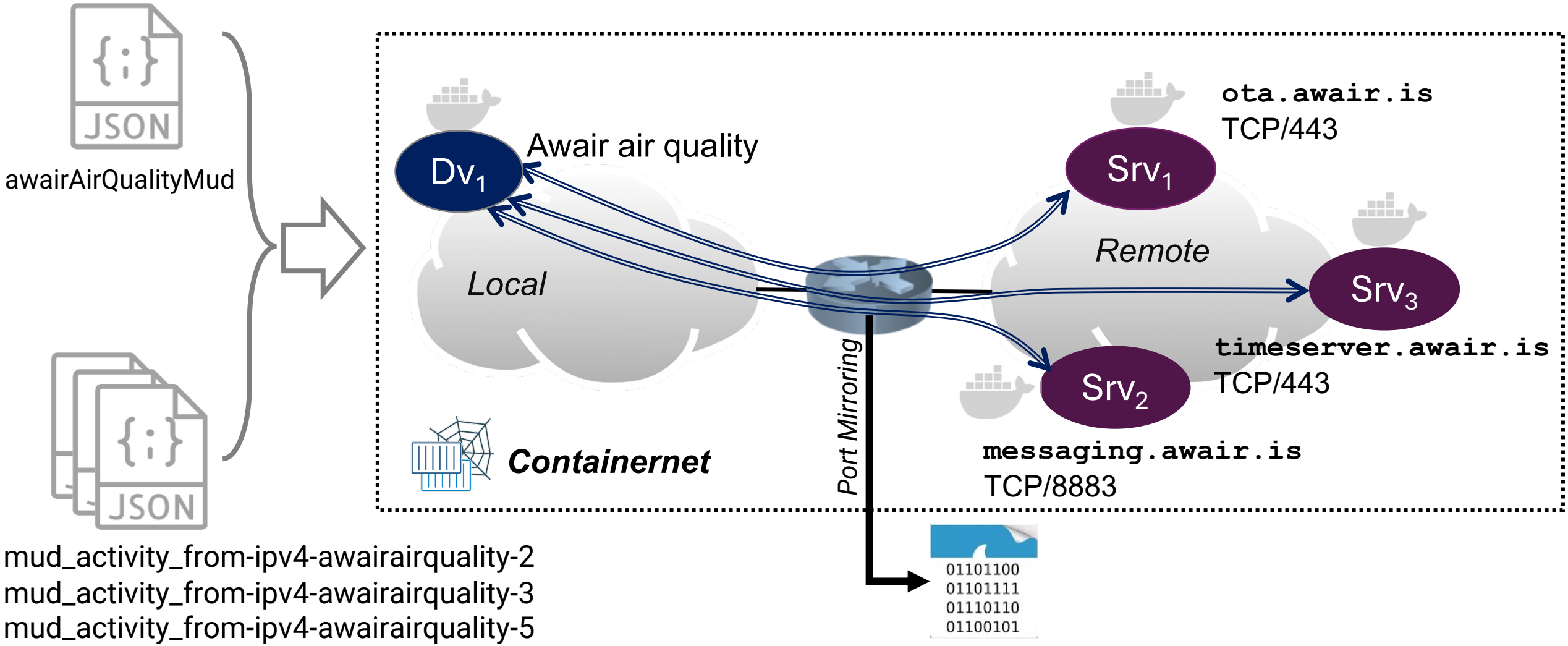
Amazon Echo flow-0:
TCP/443 → dnape-na.amazon.com

UNSW SYDNEY

Behavior descriptions and Synthetic data Repo



(2) SynGen for generating synthetic traffic data



(3) Evaluating quality of synthetic traffic data

- We employed KS test and WD test to quantify the quality of synthetic traffic data

  - `pkt-count` and `pkt-size` achieved high and moderate similarity
  - `pkt-IAT` passed majority WD tests, while `flow-iat` remains the most challenging attribute

Flow Attributes	# Srv-Flows passed KS test	# Srv-Flows with WD≤ 1	# Srv-Flows with WD≤ 10
Pkt-Count	18 (95%)	15 (79%)	16 (84%)
Pkt-Size	11 (58%)	7 (37%)	12 (63%)
Pkt-IAT	4 (21%)	15 (79%)	18 (95%)
Flow-IAT	4 (21%)	0 (0%)	0 (0%)

Conclusion

- ML models are increasingly used to monitor network attack surfaces of IoT devices
  - need high-quality data

- We developed a method to describe network behaviors by grammar files that can be fed to a tool to generate synthetic traffic

- We evaluated the efficacy of our method by applying it to 19 MUD service flows

- We publicly released our synthetic traffic traces and behavior grammar models

Thanks for your attention ☺

Shuai Zhang

shuai.zhang4@student.unsw.edu.au

More data and tools can be found on our research website: **https://iotanalytics.unsw.edu.au/**