



# Understanding network traffic through Intraflow data

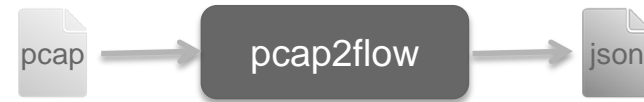
David McGrew and Blake Anderson  
[mcgrew@cisco.com](mailto:mcgrew@cisco.com), [blaander@cisco.com](mailto:blaander@cisco.com)

FloCon 2016

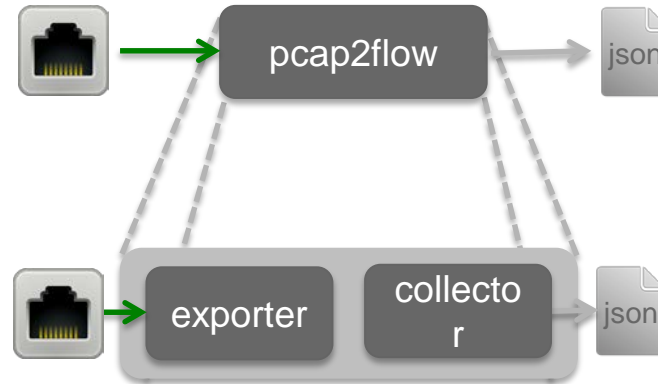
```
"flow": {
  "sa": "66.114.170.46",
  "da": "10.117.10.228",
  "pr": "6",
  "sp": "443",
  "dp": "51693",
  "ob": "29689",
  "op": "82",
  "ib": "0",
  "ip": "81",
  "ts": "1448916579.567474",
  "te": "1448916609.021792",
  "ottl": "237",
  "ittl": "64",
  "non_norm_stats": [
    { "b": "825", "dir": ">", "ipt": "0" },
    { "b": "206", "dir": ">", "ipt": "421" },
    { "b": "96", "dir": ">", "ipt": "65" },
    { "b": "114", "dir": ">", "ipt": "137" },
    { "b": "206", "dir": ">", "ipt": "221" },
    { "b": "825", "dir": ">", "ipt": "618" },
    { "b": "206", "dir": ">", "ipt": "546" },
    { "b": "825", "dir": ">", "ipt": "809" },
    { "b": "96", "dir": ">", "ipt": "421" },
    { "b": "114", "dir": ">", "ipt": "285" },
    { "b": "206", "dir": ">", "ipt": "110" },
    { "b": "825", "dir": ">", "ipt": "177" },
    { "b": "206", "dir": ">", "ipt": "789" },
    { "b": "825", "dir": ">", "ipt": "619" },
    { "b": "54", "dir": ">", "ipt": "431" },
    { "b": "206", "dir": ">", "ipt": "360" },
    { "b": "96", "dir": ">", "ipt": "206" },
    { "b": "114", "dir": ">", "ipt": "474" },
    { "b": "825", "dir": ">", "ipt": "0" },
    { "b": "206", "dir": ">", "ipt": "122" },
    { "b": "825", "dir": ">", "ipt": "703" },
    { "b": "206", "dir": ">", "ipt": "499" },
    { "b": "825", "dir": ">", "ipt": "751" },
    { "b": "206", "dir": ">", "ipt": "442" }
  ]
}
```

# Exploring threat data features at scale

joy



Offline

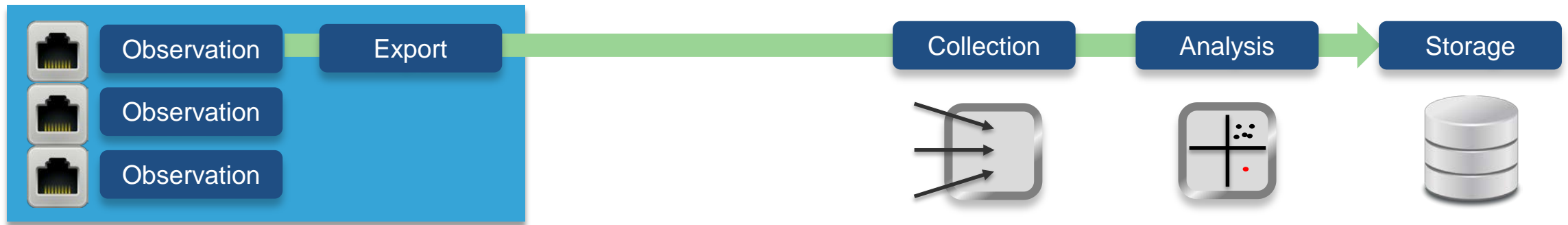


Online

<https://github.com/davidmcgrew/joy>

# Flow Monitoring

srcIP, dstIP, srcPort, dstPort, prot, startTime, stopTime, numBytes, numPackets



*“I need to understand traffic even when it is encrypted”*

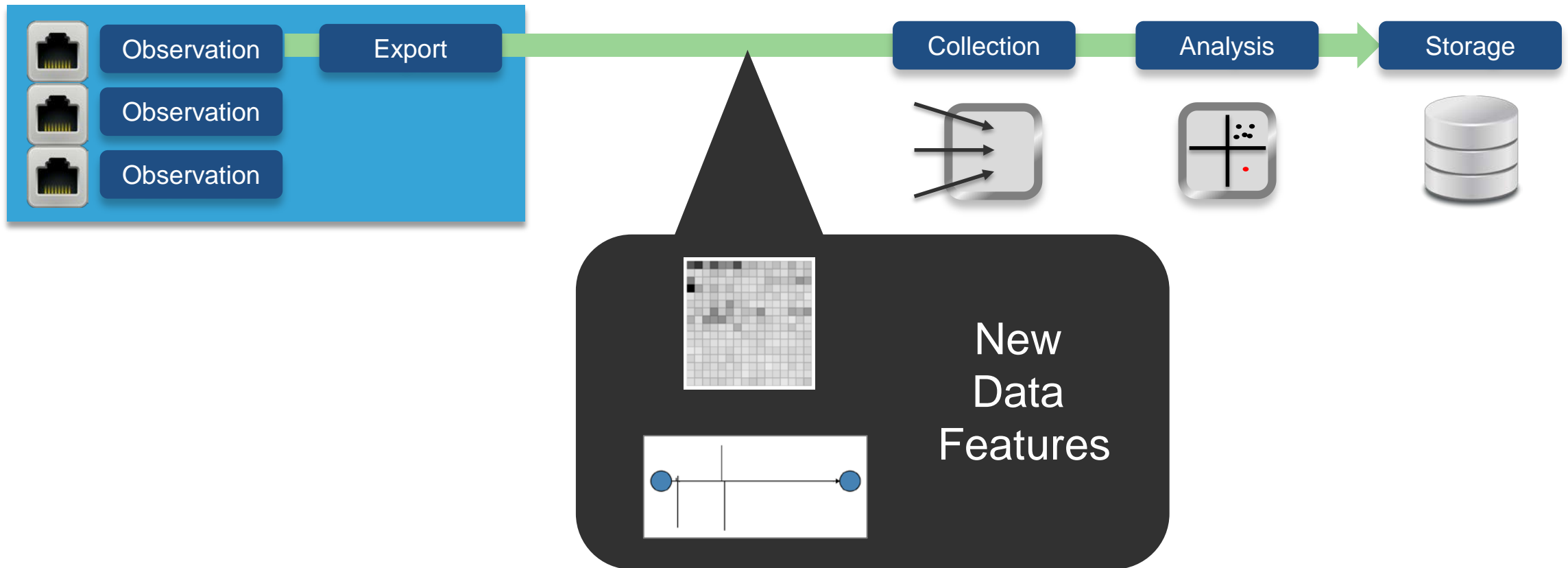
Known threats and malware  
Evasive applications and tunnels  
TLS, SSH, and other encrypted traffic (on **any** port)

*“I need to understand **all** the traffic in my network, not just traffic that passes through a security appliance”*

Monitoring internal traffic  
Forensics  
Crypto usage audit

# Flow Monitoring

srcIP, dstIP, srcPort, dstPort, prot, startTime, stopTime, numBytes, numPackets



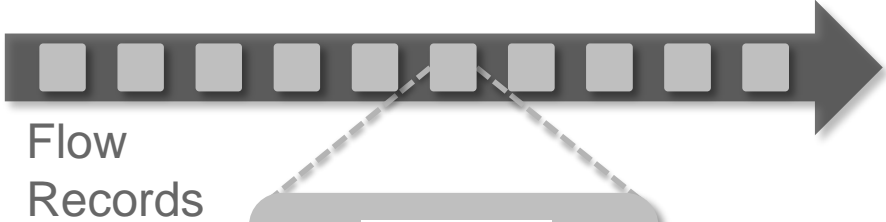
# Intraflow data

*Information about events or data inside of flows that can be conveniently collected, stored, and analyzed within a flow monitoring framework*

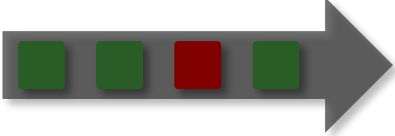
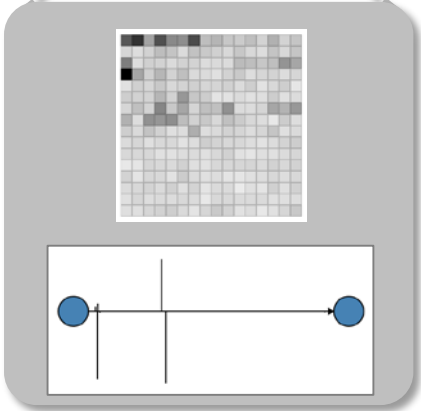
# Intraflow data

- Economical observation
  - Unidirectional
  - Minimal computation
  - Small snaplen
- Application/protocol independence
- Compactness
  - Observation
  - Transmission and storage
- Composability

# Architecture



Flow  
Records

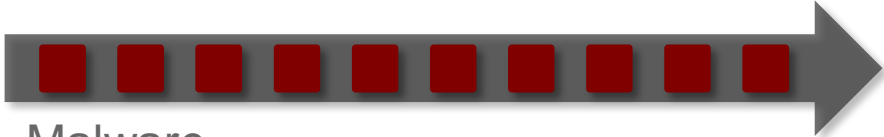




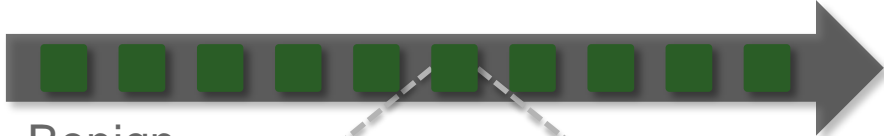
# Training architecture



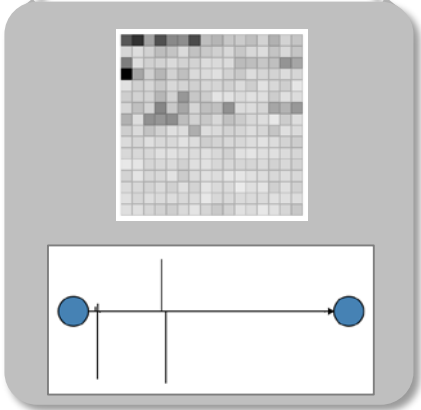
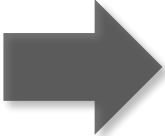
Malware  
Detonation



Malware  
Records



Benign  
Records



# New Telemetry Data Features

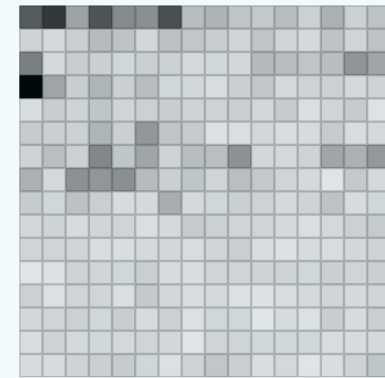
```
"flow": {
  "flow": {
    "sa": "66.114.170.46",
    "da": "10.117.10.228",
    "pr": 6,
    "sp": 443,
    "dp": 51693,
    "ob": 29689,
    "op": 82,
    "ib": 0,
    "ip": 81,
    "ts": 1448916579.567474,
    "te": 1448916609.021792,
    "ottl": 237,
    "ittl": 64,
    "non_norm_stats": [
      {
        "b": "825", "dir": ">", "ipt": 0,
        "b": "206", "dir": ">", "ipt": 421,
        "b": "96", "dir": ">", "ipt": 65,
        "b": "114", "dir": ">", "ipt": 137,
        "b": "825", "dir": ">", "ipt": 221,
        "b": "206", "dir": ">", "ipt": 618,
        "b": "825", "dir": ">", "ipt": 546,
        "b": "206", "dir": ">", "ipt": 809,
        "b": "825", "dir": ">", "ipt": 421,
        "b": "96", "dir": ">", "ipt": 285,
        "b": "114", "dir": ">", "ipt": 110,
        "b": "206", "dir": ">", "ipt": 177,
        "b": "825", "dir": ">", "ipt": 789,
        "b": "206", "dir": ">", "ipt": 619,
        "b": "825", "dir": ">", "ipt": 421,
        "b": "96", "dir": ">", "ipt": 474,
        "b": "114", "dir": ">", "ipt": 0,
        "b": "825", "dir": ">", "ipt": 122,
        "b": "206", "dir": ">", "ipt": 703,
        "b": "825", "dir": ">", "ipt": 499,
        "b": "206", "dir": ">", "ipt": 751,
        "b": "825", "dir": ">", "ipt": 442
      ]
    ]
  }
}
```

# Enhanced Telemetry Data Types

- SPLT – Sequence of Packet Lengths and Arrival Times

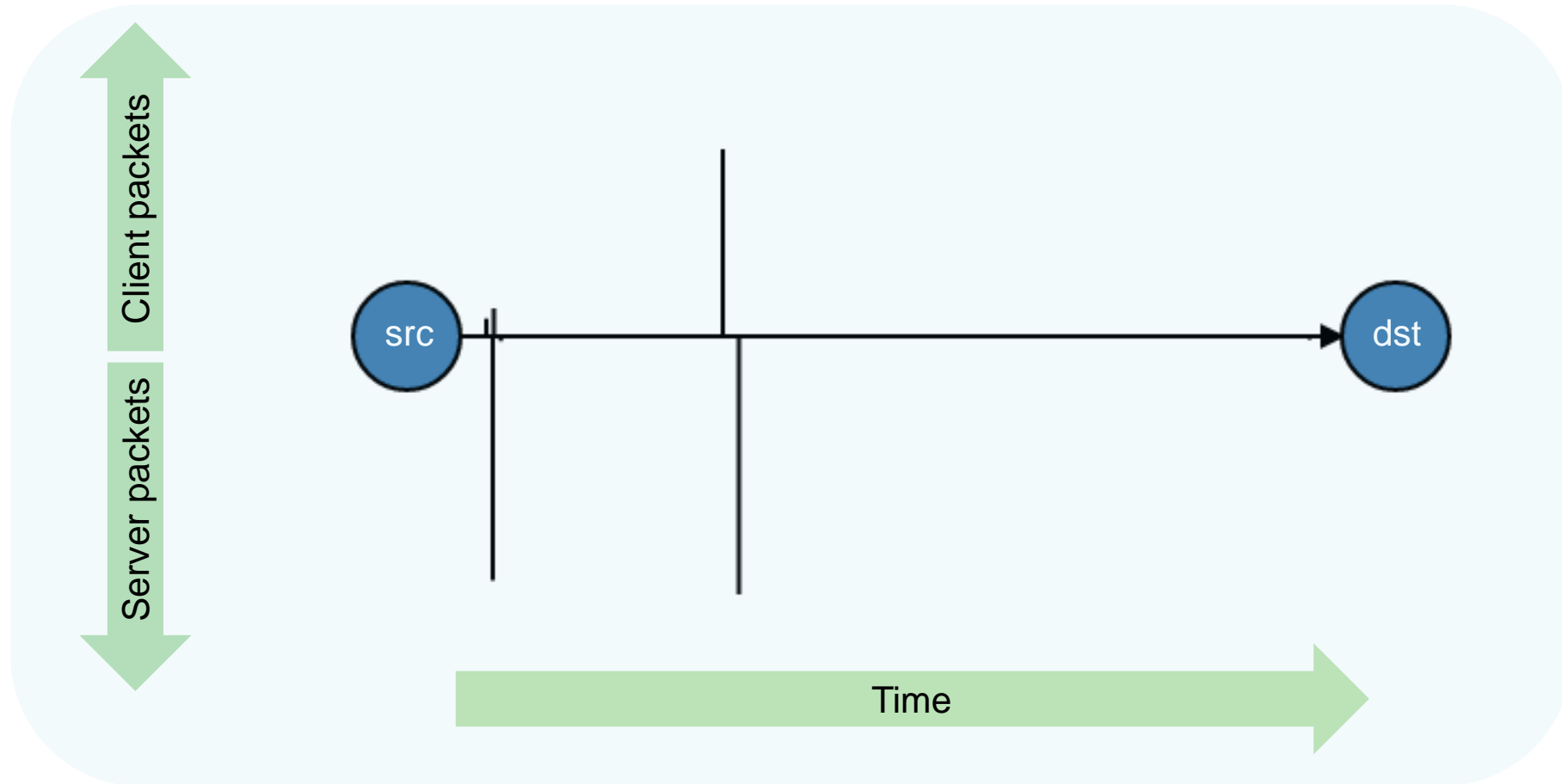


- Byte Distribution
  - Relative frequency for each byte in a flow
- Byte Entropy



- Initial Data Packet

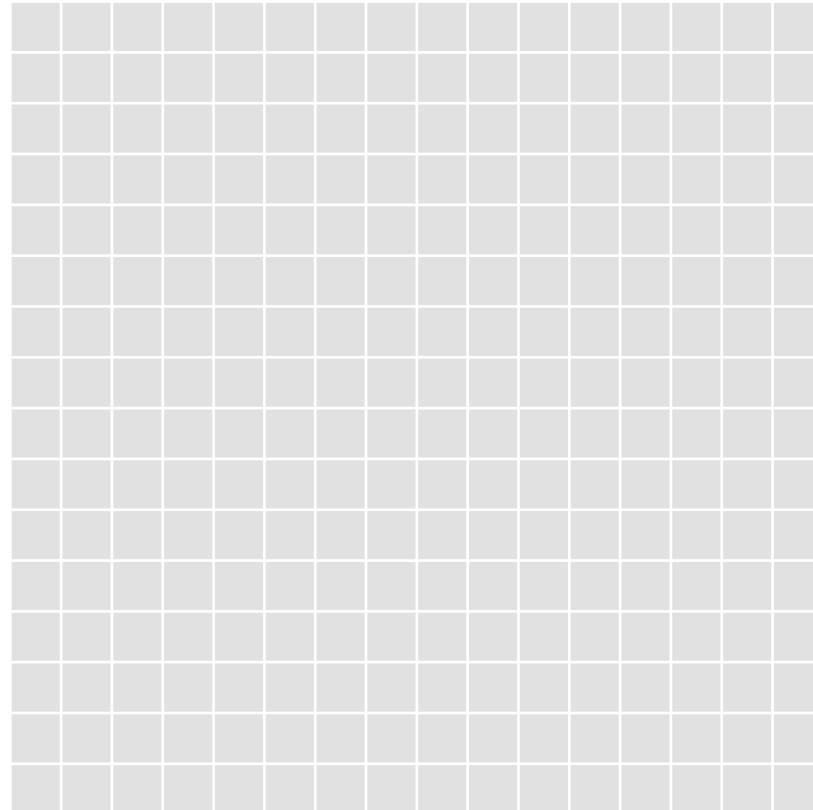
# Sequence of Packet Lengths and Times



# Byte Distribution

H T T P / 1 . 1 2 0 0 O K

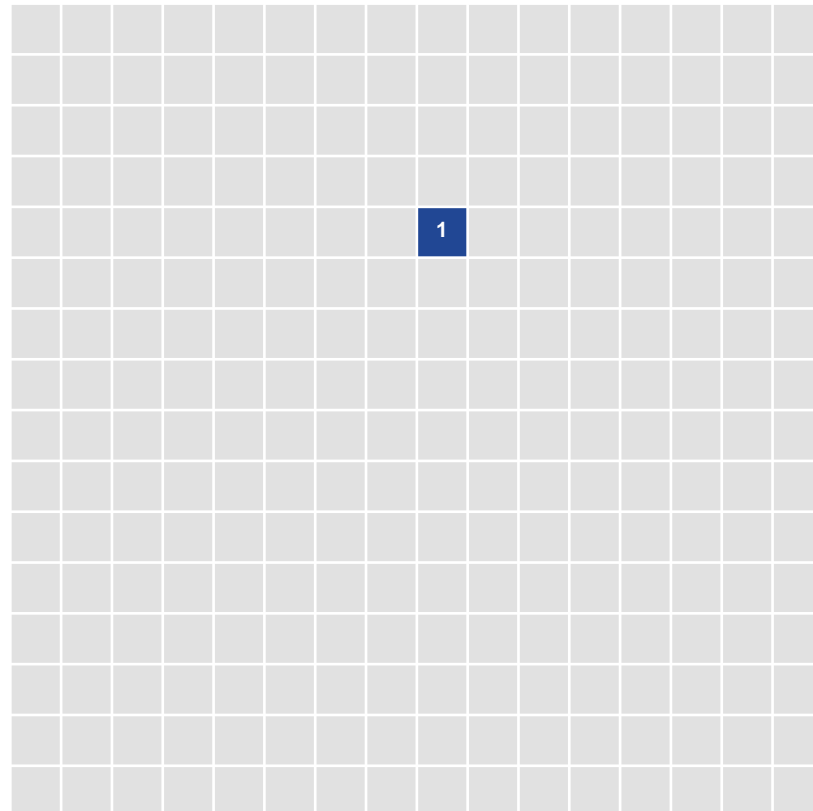
□ 48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b



# Byte Distribution

H T T P / 1 . 1 2 0 0 O K

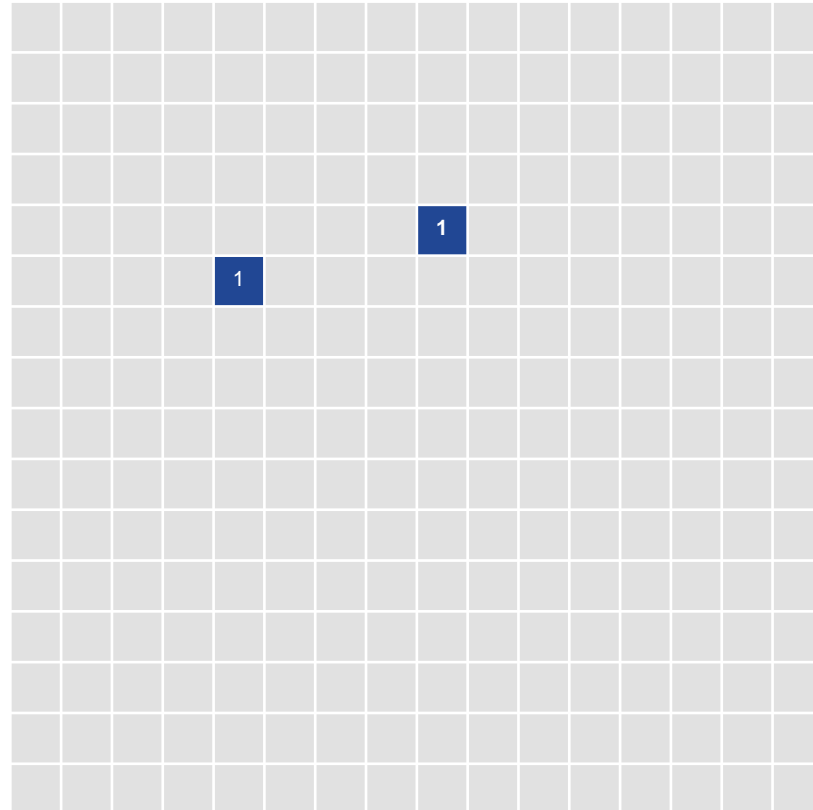
48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b



# Byte Distribution

H T T P / 1 . 1 2 0 0 O K

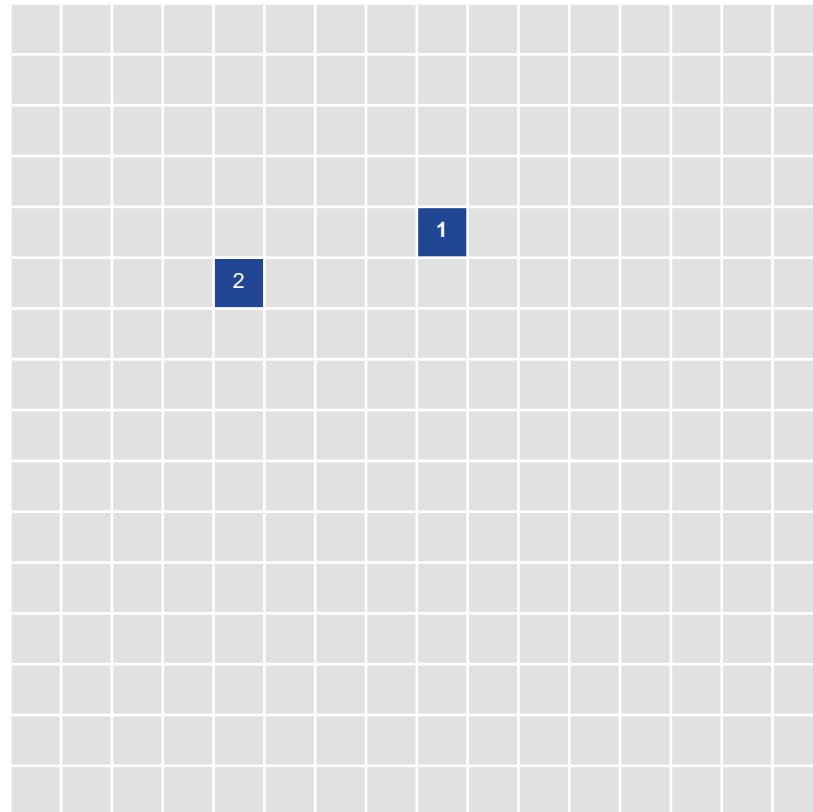
□ 48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b



# Byte Distribution

H T **T** P / 1 . 1 2 0 0 O K

□ 48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b

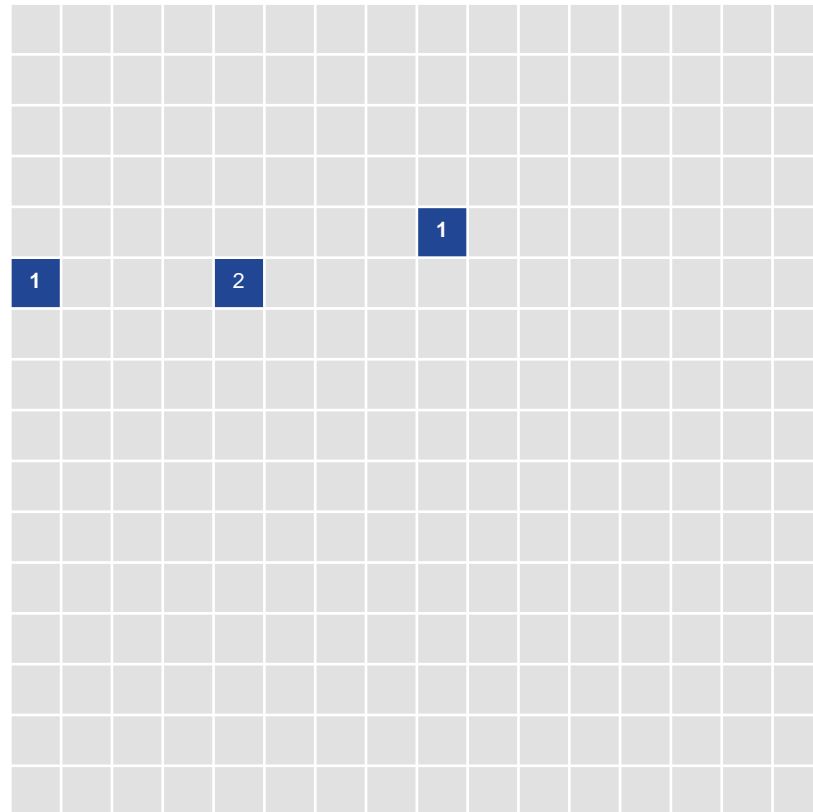




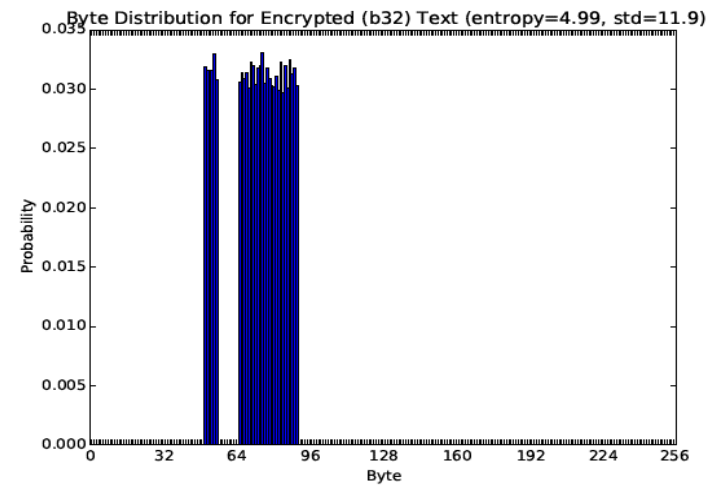
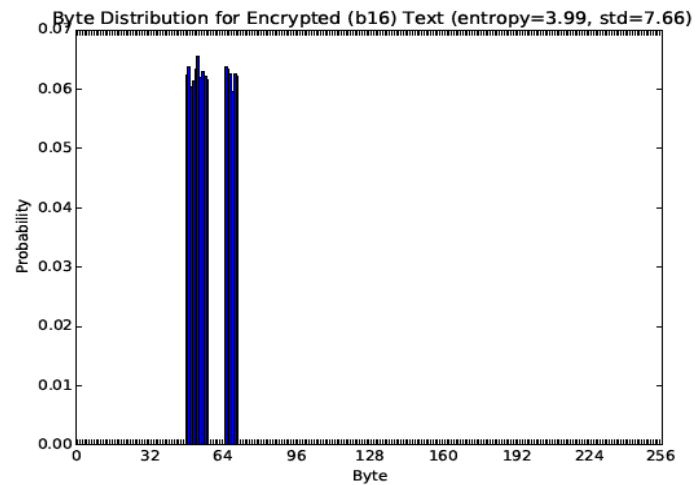
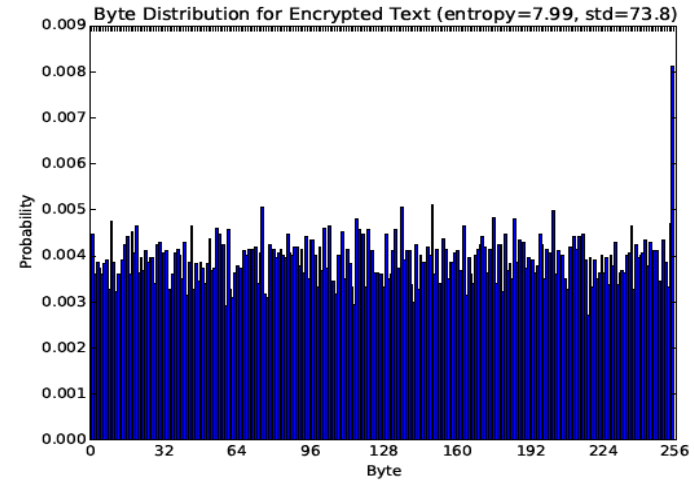
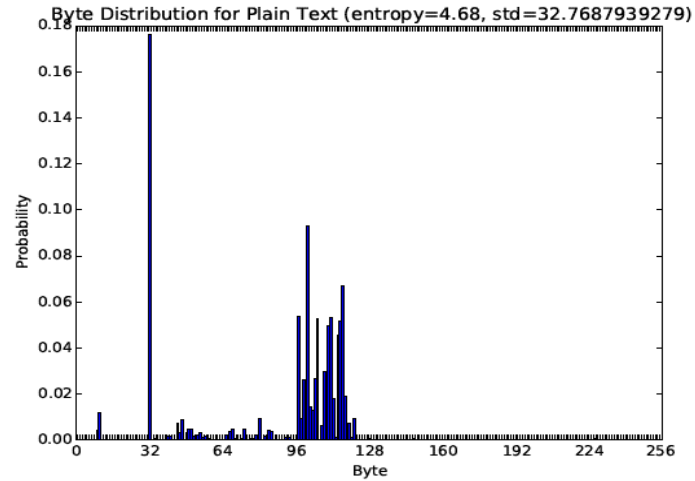
# Byte Distribution

H T T **P** / 1 . 1 2 0 0 O K

□ 48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b



# Byte Distribution for different encodings



# JSON flow data

Conventional flow data

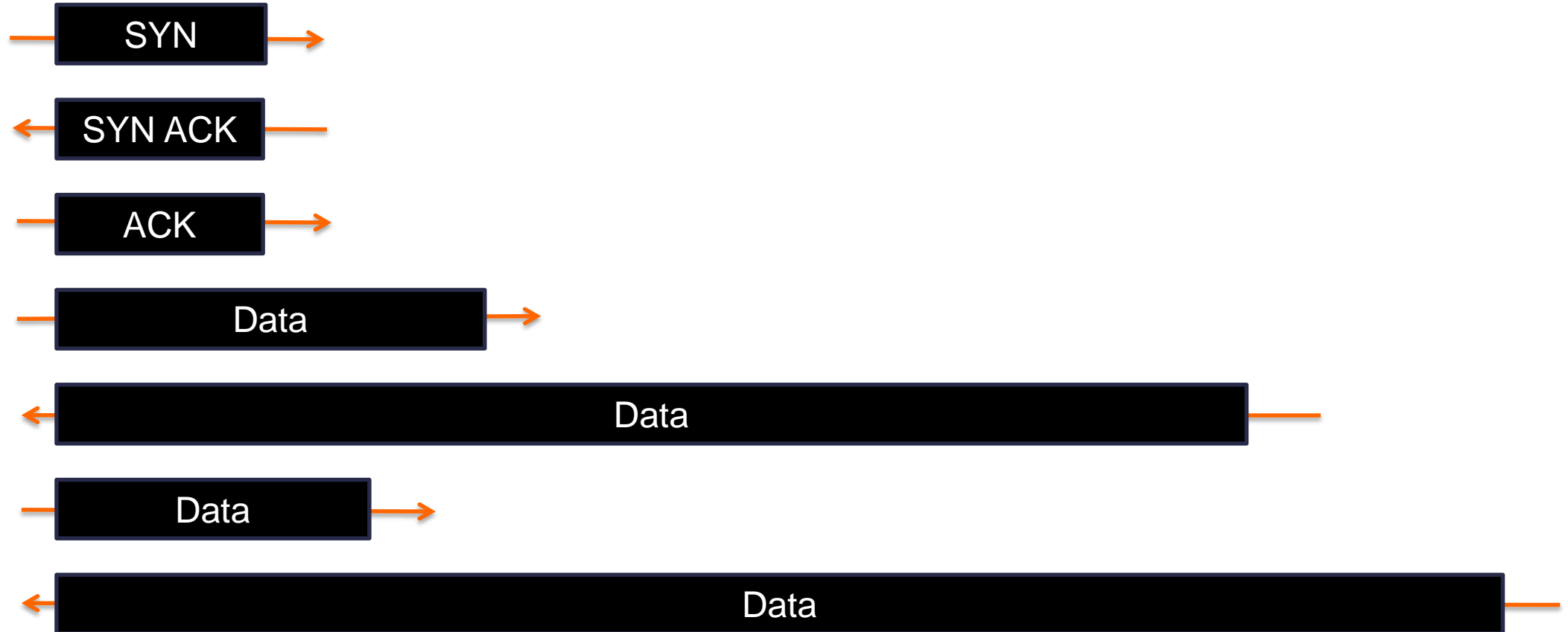
Intraflow data

Extracted parameters

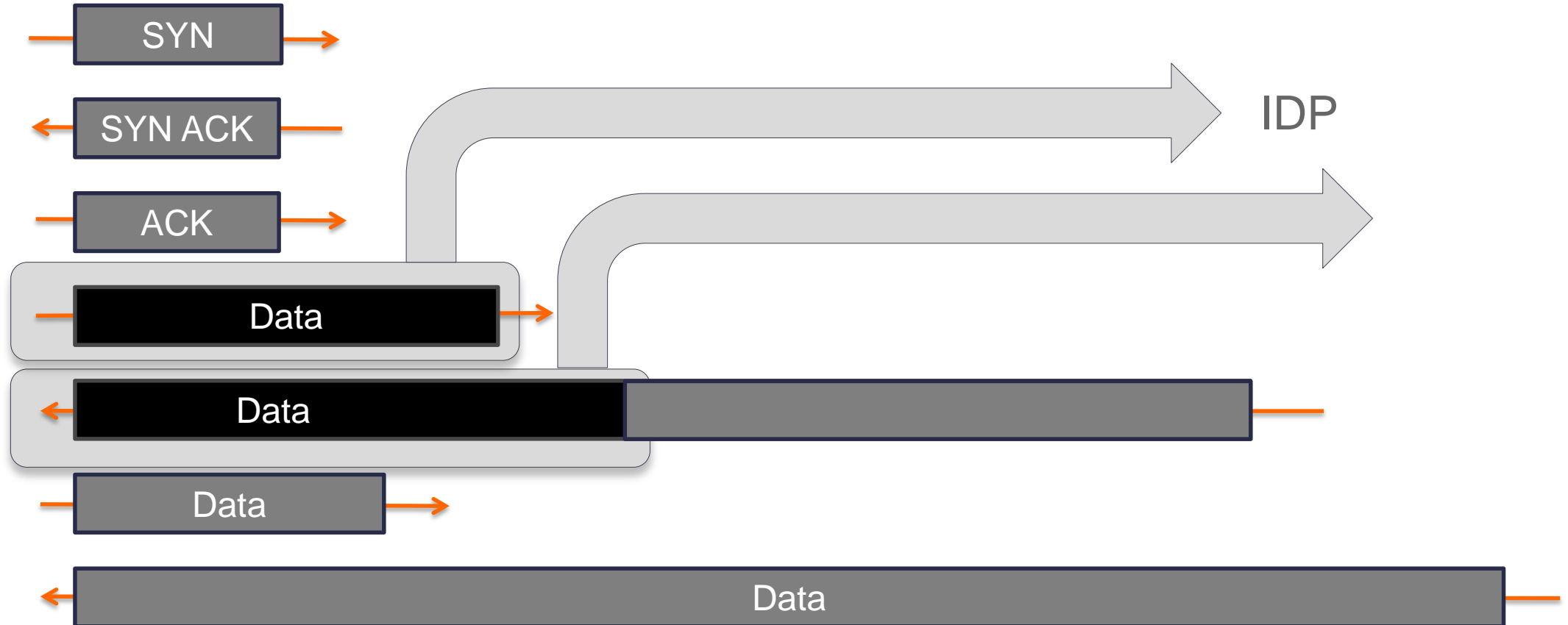
Easy to use with data analytics  
and machine learning tools

```
File Edit Options Buffers Tools Javascript Help
"flow": {
  "sa": "172.16.26.87",
  "da": "98.196.11.87",
  "pr": 6,
  "sp": 1254,
  "dp": 53,
  "ob": 9335,
  "op": 54,
  "ib": 80391,
  "ip": 77,
  "ts": 1439979838.953822,
  "te": 1439979877.515059,
  "ottl": 128,
  "ittl": 63,
  "non_norm_stats": [
    { "b": 212, "dir": ">", "ipt": 1448 },
    { "b": 928, "dir": "<", "ipt": 9 },
    { "b": 198, "dir": ">", "ipt": 149 },
    { "b": 59, "dir": "<", "ipt": 27 },
    { "b": 213, "dir": ">", "ipt": 470 },
    { "b": 1460, "dir": "<", "ipt": 27 },
    ...
    { "b": 1460, "dir": "<", "ipt": 395 },
    { "b": 1460, "dir": "<", "ipt": 34 }
  ],
  "bd": [
    501, 480, 347, 431, 373, 343, 349, 377, 382, 348, 342, 342, 322, 374, 349, 355,
    369, 328, 337, 349, 392, 341, 362, 441, 329, 338, 349, 349, 322, 367, 323, 339,
    382, 354, 332, 331, 371, 356, 329, 361, 339, 340, 334, 351, 376, 310, 335, 351,
    372, 325, 364, 342, 344, 355, 359, 341, 334, 358, 346, 335, 337, 354, 338, 317,
    363, 344, 345, 355, 346, 340, 348, 367, 377, 316, 352, 359, 318, 383, 341, 343,
    337, 361, 343, 346, 371, 337, 367, 338, 333, 357, 300, 364, 349, 333, 356, 305,
    296, 342, 335, 332, 360, 365, 334, 347, 357, 346, 348, 366, 320, 340, 317, 369,
    335, 350, 331, 363, 369, 339, 363, 406, 374, 375, 341, 334, 340, 355, 346, 313,
    354, 362, 388, 354, 351, 362, 358, 366, 352, 334, 353, 312, 364, 331, 372, 357,
    333, 367, 370, 358, 343, 326, 379, 381, 358, 330, 340, 323, 328, 351, 342, 353,
    406, 380, 359, 327, 306, 340, 334, 330, 331, 349, 327, 371, 343, 349, 319, 356,
    365, 352, 303, 355, 364, 360, 352, 355, 363, 335, 368, 333, 375, 337, 341, 360,
    375, 344, 330, 363, 343, 352, 353, 342, 334, 331, 356, 358, 356, 368, 355, 350,
    371, 344, 344, 366, 337, 359, 350, 361, 330, 350, 341, 321, 336, 366, 336, 351,
    324, 363, 370, 351, 359, 336, 335, 370, 373, 373, 380, 361, 353, 315, 345, 340,
    362, 357, 353, 318, 339, 360, 344, 334, 321, 341, 368, 366, 339, 351, 343, 387
  ],
  "be": 7.960630
}
```

# Initial Data Packet



# Initial Data Packet

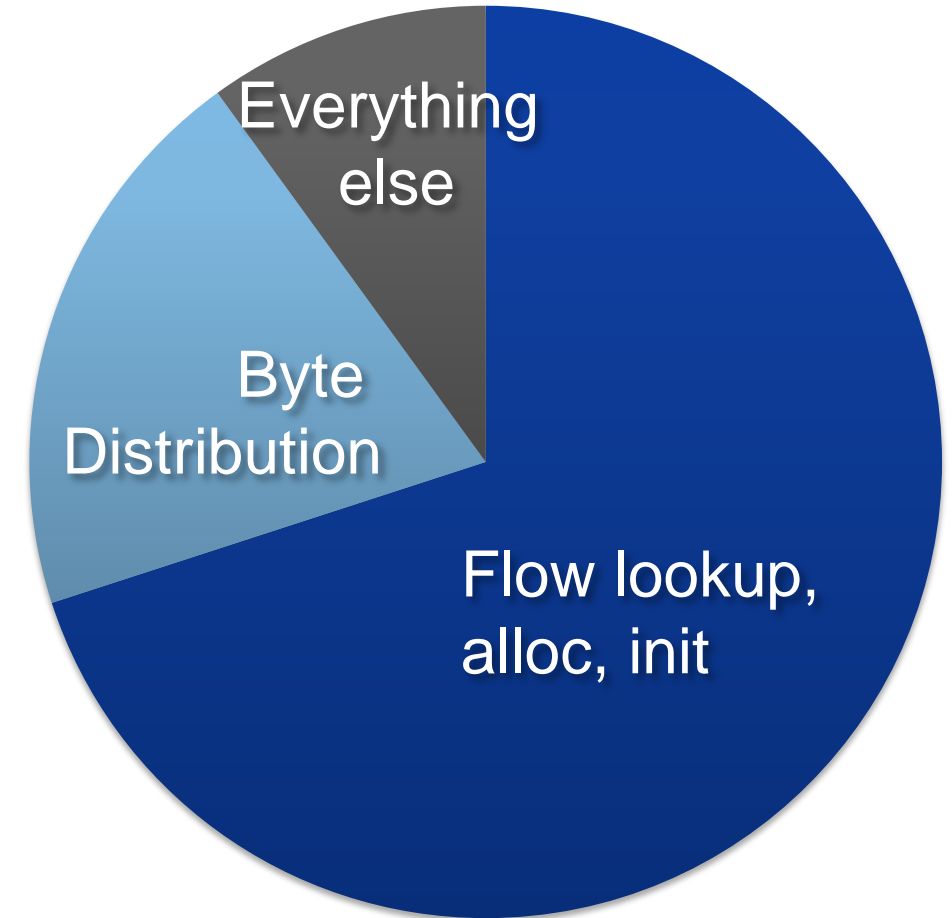


# Experimental results

```
"flow": {
  "flow": {
    "sa": "66.114.170.46",
    "da": "10.117.10.228",
    "pr": 6,
    "sp": 443,
    "dp": 51693,
    "ob": 29689,
    "op": 82,
    "ib": 0,
    "ip": 81,
    "ts": 1448916579.567474,
    "te": 1448916609.021792,
    "ottl": 237,
    "ittl": 64,
    "non_norm_stats": [
      {
        "b": "825", "dir": "dir", "ipt": "ipt", "0",
        "b": "206", "dir": "dir", "ipt": "ipt", "421",
        "b": "96", "dir": "dir", "ipt": "ipt", "65",
        "b": "114", "dir": "dir", "ipt": "ipt", "137",
        "b": "825", "dir": "dir", "ipt": "ipt", "221",
        "b": "206", "dir": "dir", "ipt": "ipt", "618",
        "b": "825", "dir": "dir", "ipt": "ipt", "546",
        "b": "206", "dir": "dir", "ipt": "ipt", "809",
        "b": "825", "dir": "dir", "ipt": "ipt", "421",
        "b": "96", "dir": "dir", "ipt": "ipt", "285",
        "b": "114", "dir": "dir", "ipt": "ipt", "110",
        "b": "206", "dir": "dir", "ipt": "ipt", "177",
        "b": "825", "dir": "dir", "ipt": "ipt", "789",
        "b": "206", "dir": "dir", "ipt": "ipt", "619",
        "b": "825", "dir": "dir", "ipt": "ipt", "431",
        "b": "54", "dir": "dir", "ipt": "ipt", "360",
        "b": "206", "dir": "dir", "ipt": "ipt", "206",
        "b": "96", "dir": "dir", "ipt": "ipt", "474",
        "b": "114", "dir": "dir", "ipt": "ipt", "0",
        "b": "825", "dir": "dir", "ipt": "ipt", "122",
        "b": "206", "dir": "dir", "ipt": "ipt", "703",
        "b": "825", "dir": "dir", "ipt": "ipt", "499",
        "b": "206", "dir": "dir", "ipt": "ipt", "751",
        "b": "825", "dir": "dir", "ipt": "ipt", "442"
      ]
    ]
  }
}
```

# Performance

- CPU: Intel Xeon @ 2.70GHz
  - 17% utilization at 1.0e7 bytes/sec, 1.2e4 packets/sec
  - Approx 870 Mbits/sec at full utilization
- RAM: 8Gbyte
  - 2.7% utilization (216 Mbyte)



# Detecting malware with SPLT and Byte Distribution

```
"flow": {
  "flow": {
    "sa": "66.114.170.46",
    "da": "10.117.10.228",
    "pr": 6,
    "sp": 443,
    "dp": 51693,
    "ob": 29689,
    "op": 82,
    "ib": 0,
    "ip": 81,
    "ts": 1448916579.567474,
    "te": 1448916609.021792,
    "ottl": 237,
    "ittl": 64,
    "non_norm_stats": [
      { "b": 825, "dir": ">", "ipt": 0 },
      { "b": 206, "dir": ">", "ipt": 421 },
      { "b": 96, "dir": ">", "ipt": 65 },
      { "b": 114, "dir": ">", "ipt": 137 },
      { "b": 825, "dir": ">", "ipt": 221 },
      { "b": 206, "dir": ">", "ipt": 618 },
      { "b": 825, "dir": ">", "ipt": 546 },
      { "b": 206, "dir": ">", "ipt": 809 },
      { "b": 825, "dir": ">", "ipt": 421 },
      { "b": 114, "dir": ">", "ipt": 285 },
      { "b": 206, "dir": ">", "ipt": 110 },
      { "b": 825, "dir": ">", "ipt": 177 },
      { "b": 206, "dir": ">", "ipt": 789 },
      { "b": 825, "dir": ">", "ipt": 619 },
      { "b": 54, "dir": ">", "ipt": 431 },
      { "b": 96, "dir": ">", "ipt": 360 },
      { "b": 114, "dir": ">", "ipt": 206 },
      { "b": 825, "dir": ">", "ipt": 474 },
      { "b": 206, "dir": ">", "ipt": 0 },
      { "b": 825, "dir": ">", "ipt": 122 },
      { "b": 206, "dir": ">", "ipt": 703 },
      { "b": 825, "dir": ">", "ipt": 499 },
      { "b": 206, "dir": ">", "ipt": 751 },
      { "b": 825, "dir": ">", "ipt": 442 }
    ]
  }
}
```



# Analytics User Interface



Number of flows classified: 244

P(Malware) ▾	P(TLS) ▾	Source Address ▾	Dest.Address ▾	Source Port ▾	Dest. Port ▾	Inbound Packets ▾	Outbound Packets ▾
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1034	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1035	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1036	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1044	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1045	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1064	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1065	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1070	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1071	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1072	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1073	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1074	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1076	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1077	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1080	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1115	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1116	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1117	443	16	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1118	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1119	443	15	11
■ 1.0	■ 0.99	172.16.45.20	104.237.132.39	1120	443	15	11

Number of flows classified: 244

P(Malware)	P(TLS)	Source Address	Dest.Address	Source Port	Dest. Port	Inbound Packets	Outbound Packets
1.0	0.99	172.16.45.20	104.237.132.39	1034	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1035	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1036	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1044	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1045	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1064	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1065	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1070	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1071	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1072	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1073	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1074	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1076	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1077	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1080	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1115	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1116	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1117	443	16	11
1.0	0.99	172.16.45.20	104.237.132.39	1118	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1119	443	15	11
1.0	0.99	172.16.45.20	104.237.132.39	1120	443	15	11

Malware Classification

# Classifying flows as malicious/benign

- L1-logistic regression
  - SPLT + 5-tuple
- L1-logistic regression
  - SPLT + 5-tuple + BD

# Classifying flows as malicious/benign

- L1-logistic regression
- SPLT + 5-tuple
  - 160 non-zero parameters
  - 0.01 FDR: 51.11%
  - Total Accuracy: 98.44%
- L1-logistic regression
- SPLT + 5-tuple + BD
  - 128 non-zero parameters
  - 0.01 FDR: 98.92%
  - Total Accuracy: 99.81%

# Intraflow data

- Economical observation
  - Unidirectional
  - Minimal computation
  - Small snaplen
- Application/protocol independence
- Compactness
  - Observation
  - Transmission and storage

•  Composability

SPLT  
10 packets

Yes

Yes

Yes

Yes

10 bytes

10 bytes

Yes

Byte  
Distribution

Yes

Yes

No

Yes

256 bytes

256 bytes

Yes

# Intraflow data

- Economical observation
  - Unidirectional
  - Minimal computation
  - Small snaplen
- Application/protocol independence
- Compactness
  - Observation
  - Transmission and storage

• Composability

SPLT  
10 packets

Yes

Yes

Yes

Yes

10 bytes

10 bytes

Yes

Byte  
Distribution

Yes

Yes

No

Yes

16  
bytes

256 bytes

256 bytes

Yes

# Conclusions

- Intraflow data is feasible to implement, enables useful inferences
- SPLT is valuable and relatively cheap
- Byte Distribution is valuable but more costly
- Training classifiers is key
  - Data fusion



```
"flow": {  
  "flow": {  
    "sa": "66.114.170.46",  
    "da": "10.117.10.228",  
    "pr": "6",  
    "sp": "443",  
    "dp": "51693",  
    "ob": "29689",  
    "op": "82",  
    "ib": "0",  
    "ip": "81",  
    "ts": "1448916579.567474",  
    "te": "1448916609.021792",  
    "ottl": "237",  
    "ittl": "64",  
    "non_norm_stats": [  
      {  
        "b": "206", "dir": "in", "ipt": "0",  
        "b": "206", "dir": "out", "ipt": "421",  
        "b": "96", "dir": "in", "ipt": "65",  
        "b": "114", "dir": "in", "ipt": "137",  
        "b": "825", "dir": "in", "ipt": "221",  
        "b": "206", "dir": "in", "ipt": "618",  
        "b": "825", "dir": "in", "ipt": "546",  
        "b": "206", "dir": "in", "ipt": "809",  
        "b": "825", "dir": "in", "ipt": "421",  
        "b": "96", "dir": "in", "ipt": "285",  
        "b": "114", "dir": "in", "ipt": "110",  
        "b": "206", "dir": "in", "ipt": "177",  
        "b": "825", "dir": "in", "ipt": "789",  
        "b": "206", "dir": "in", "ipt": "619",  
        "b": "825", "dir": "in", "ipt": "431",  
        "b": "54", "dir": "in", "ipt": "360",  
        "b": "206", "dir": "in", "ipt": "206",  
        "b": "96", "dir": "in", "ipt": "474",  
        "b": "114", "dir": "in", "ipt": "0",  
        "b": "825", "dir": "in", "ipt": "122",  
        "b": "206", "dir": "in", "ipt": "703",  
        "b": "825", "dir": "in", "ipt": "499",  
        "b": "206", "dir": "in", "ipt": "751",  
        "b": "825", "dir": "in", "ipt": "442",  
      }  
    ]  
  }  
}
```

Thank You



# Joy applications

