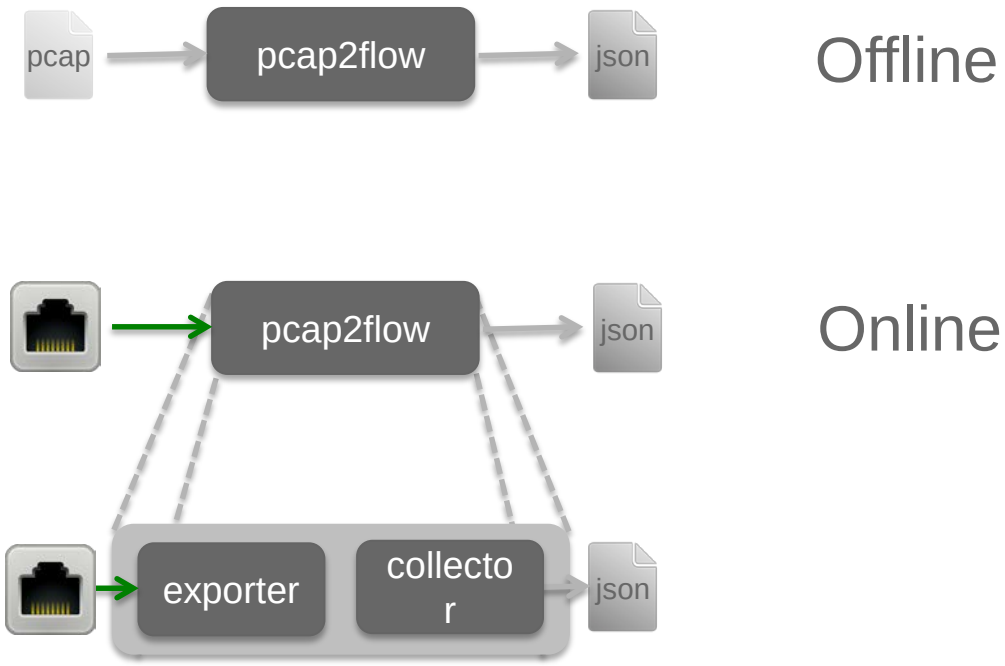
Understanding network traffic through Intraflow data

David McGrew and Blake Anderson
mcgrew@cisco.com, blaander@cisco.com

FloCon 2016

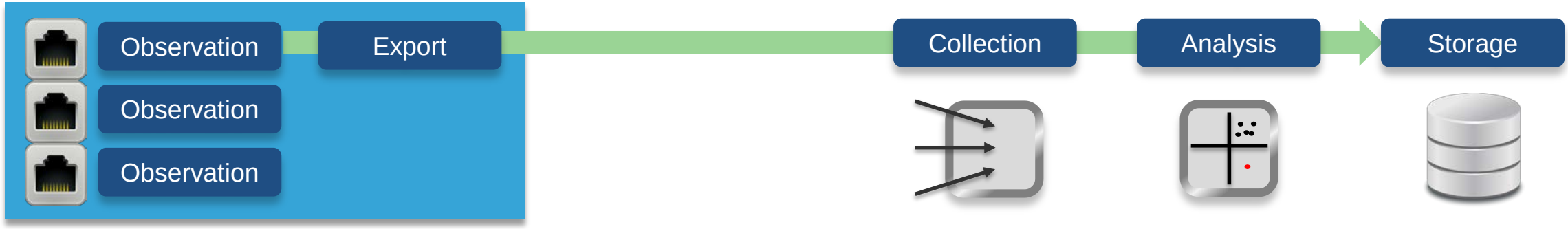
Exploring threat data features at scale



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"*

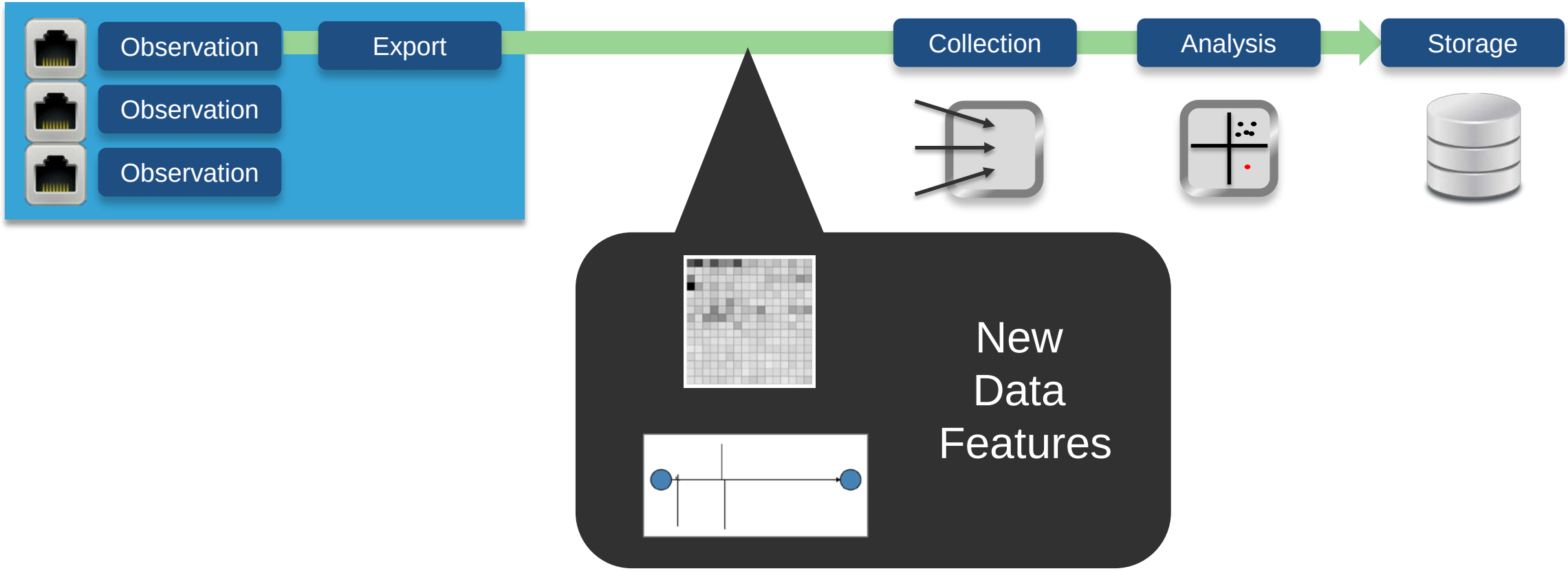
Monitoring internal traffic

Forensics

Crypto usage audit

Flow Monitoring

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



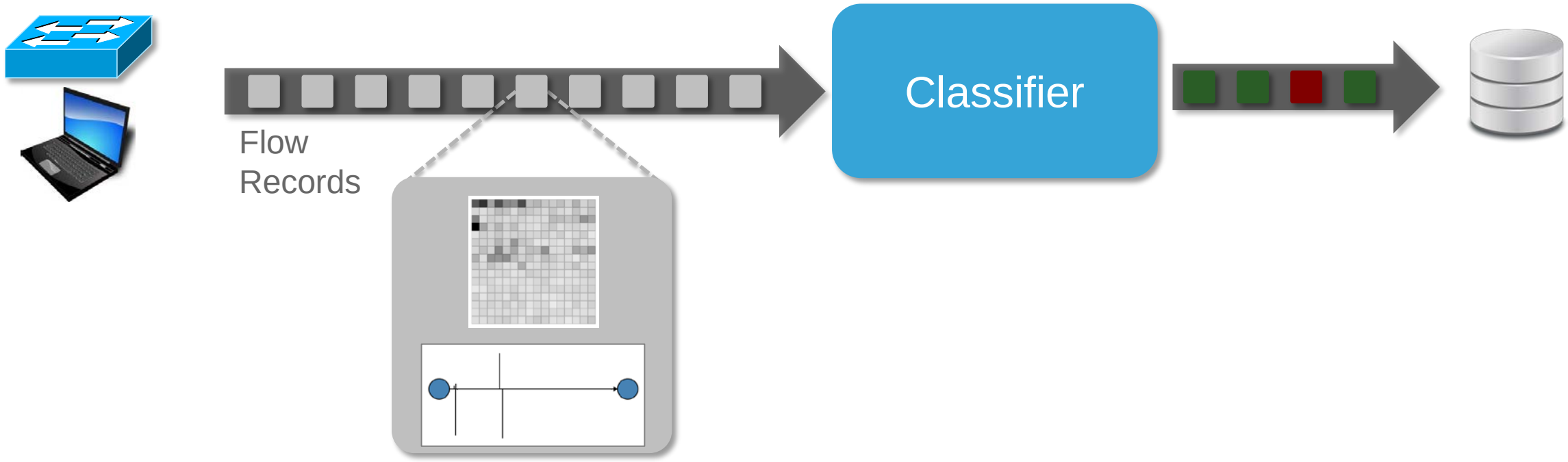
Intraflow data

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

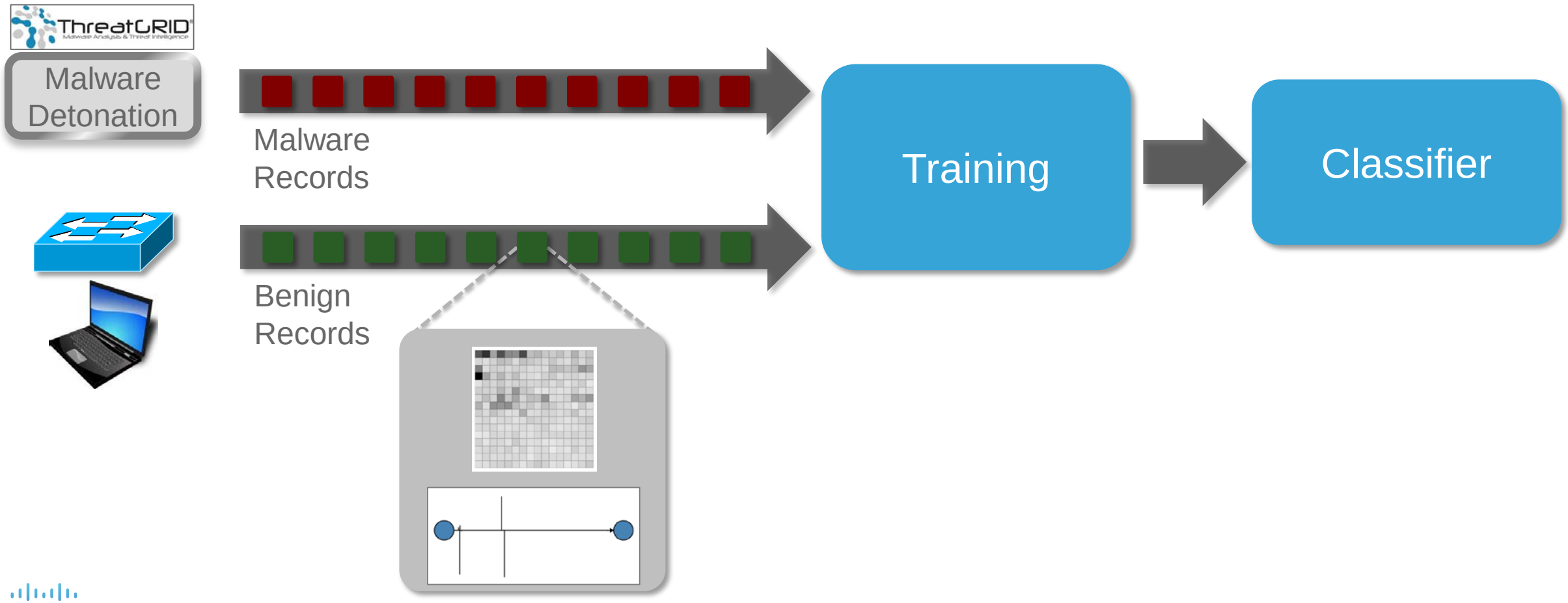
Intraflow data

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

Architecture



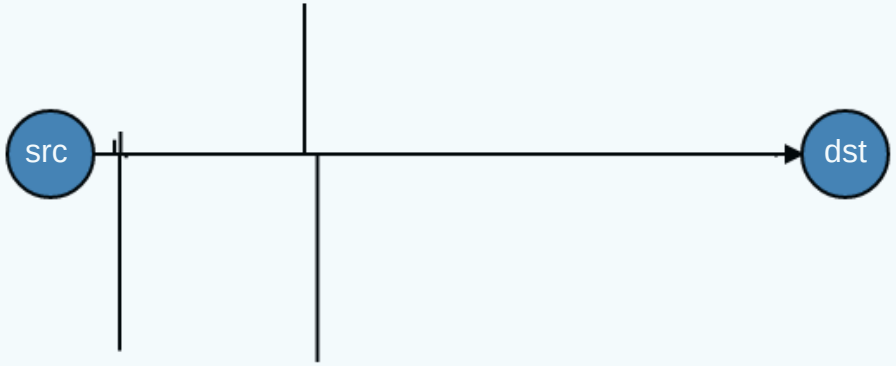
Training architecture



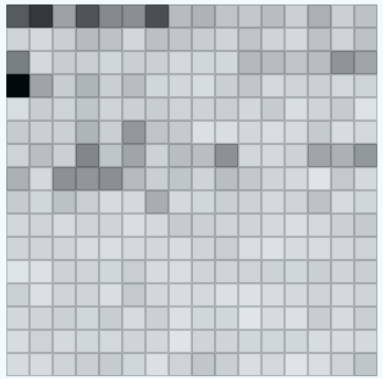
New Telemetry Data Features

Enhanced Telemetry Data Types

- SPLT – Sequence of Packet Lengths and Arrival Times

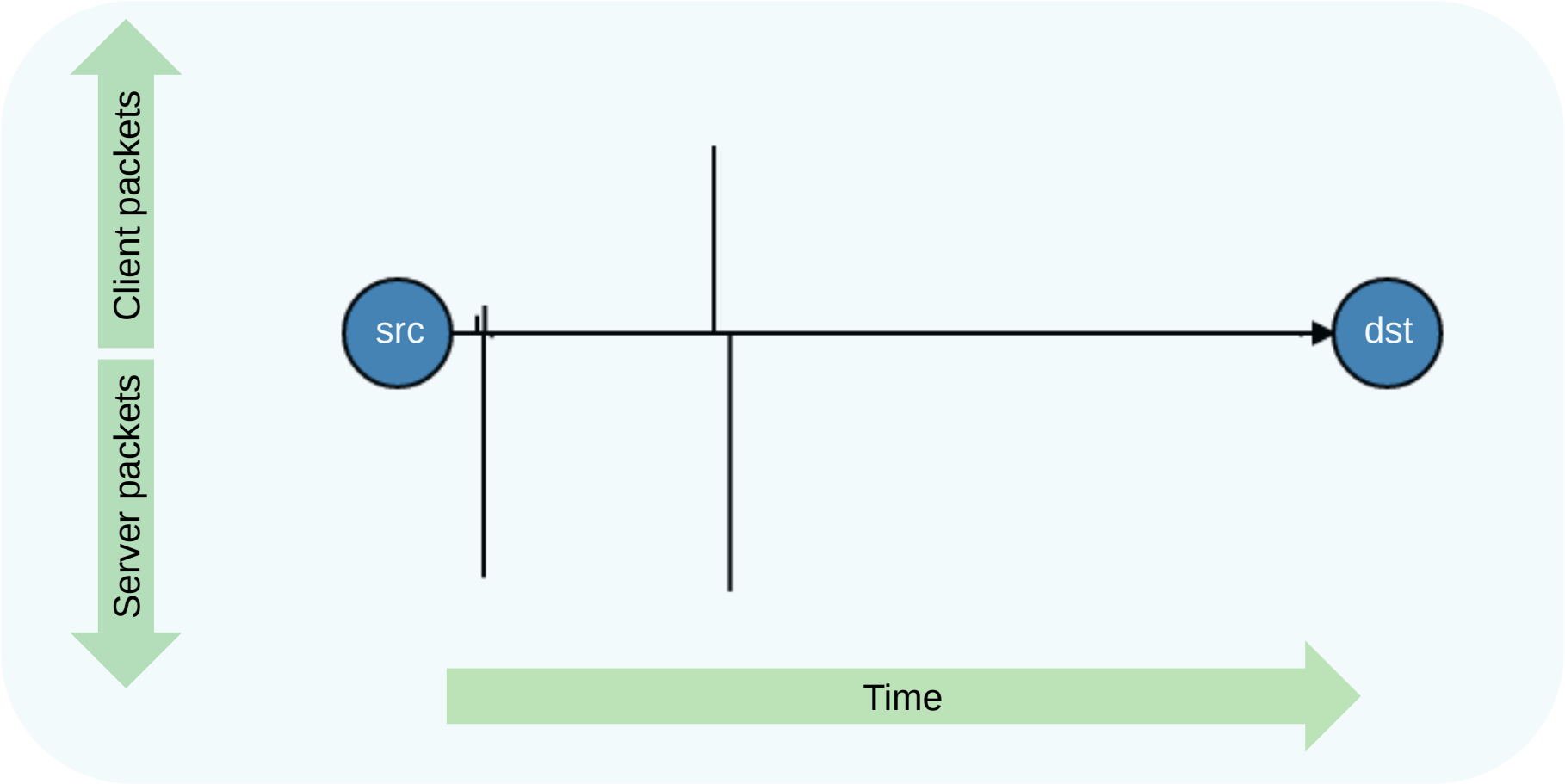


- Byte Distribution
  - Relative frequency for each byte in a flow
- Byte Entropy



- Initial Data Packet

Sequence of Packet Lengths and Times



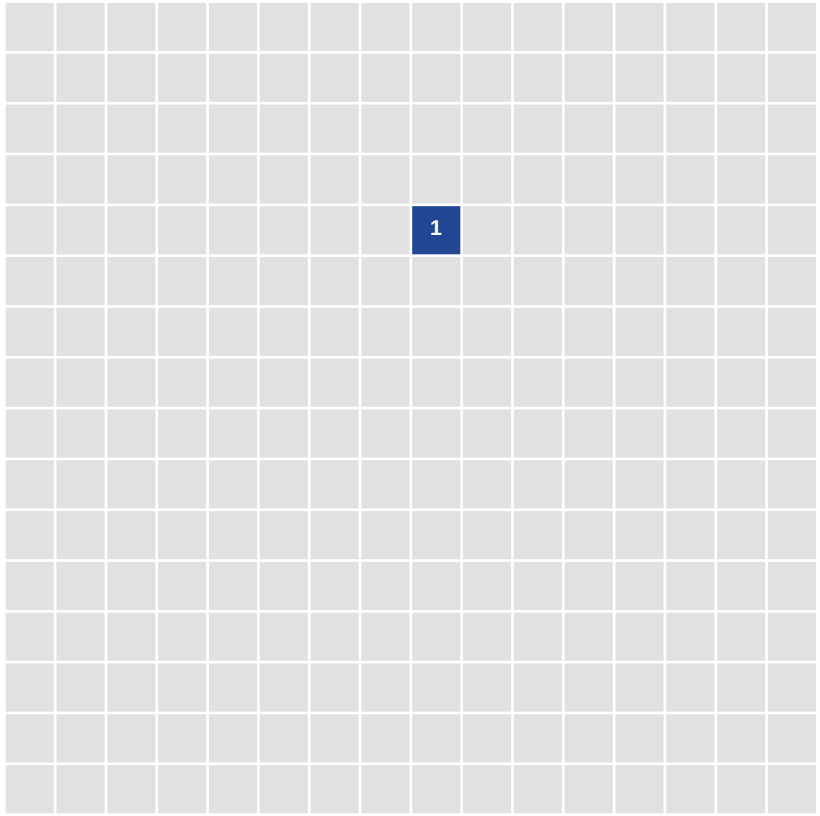
Byte Distribution

```
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

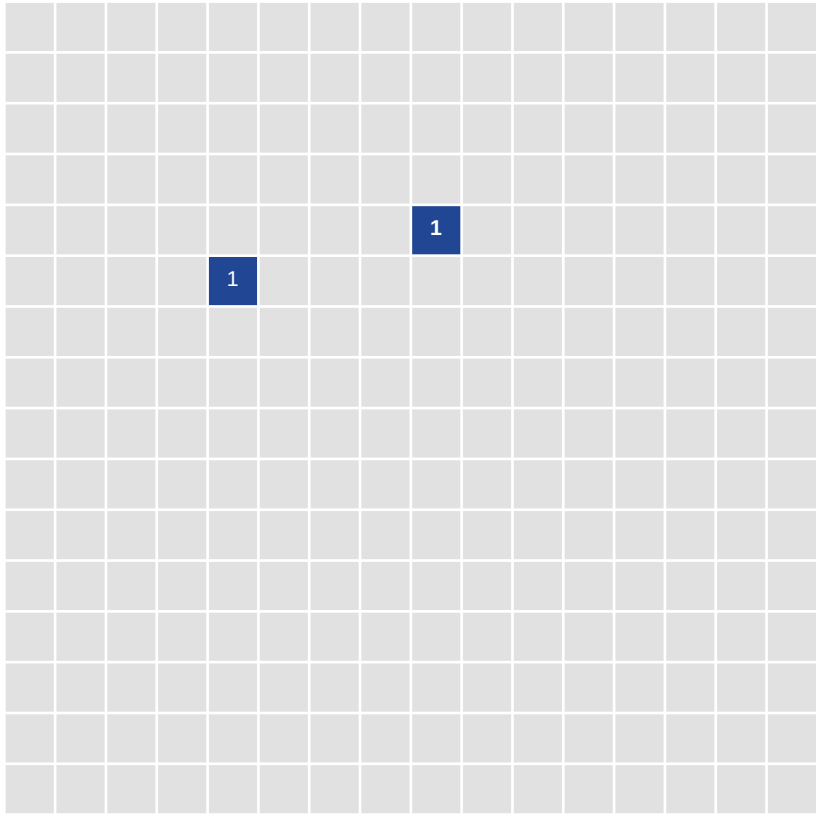
H	T	T	P	/	1	.	1		2	0	0		O	K

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



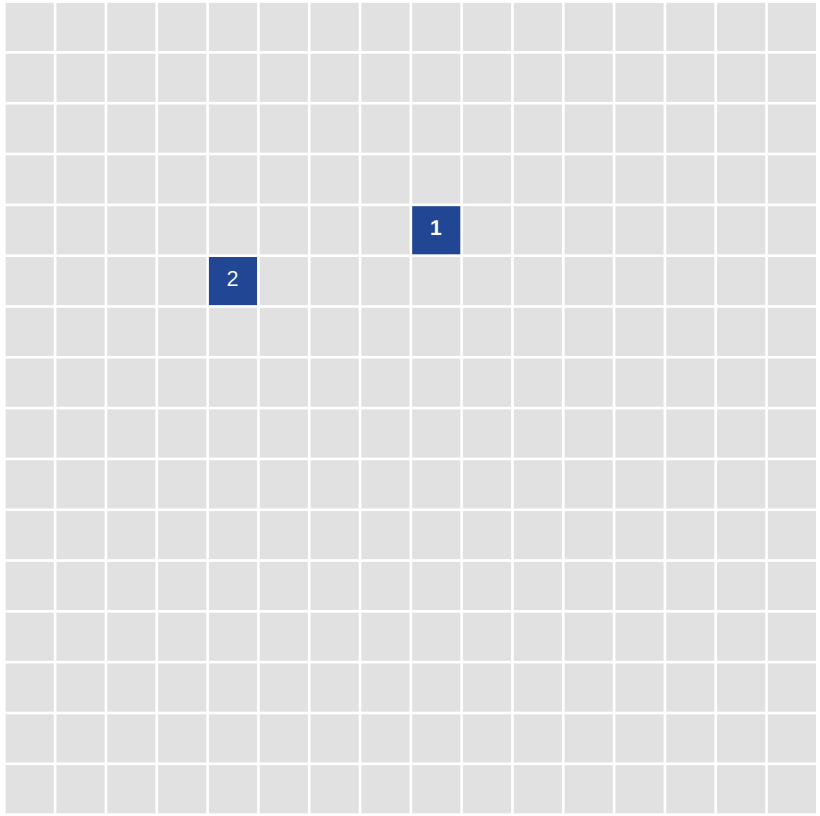
Byte Distribution

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



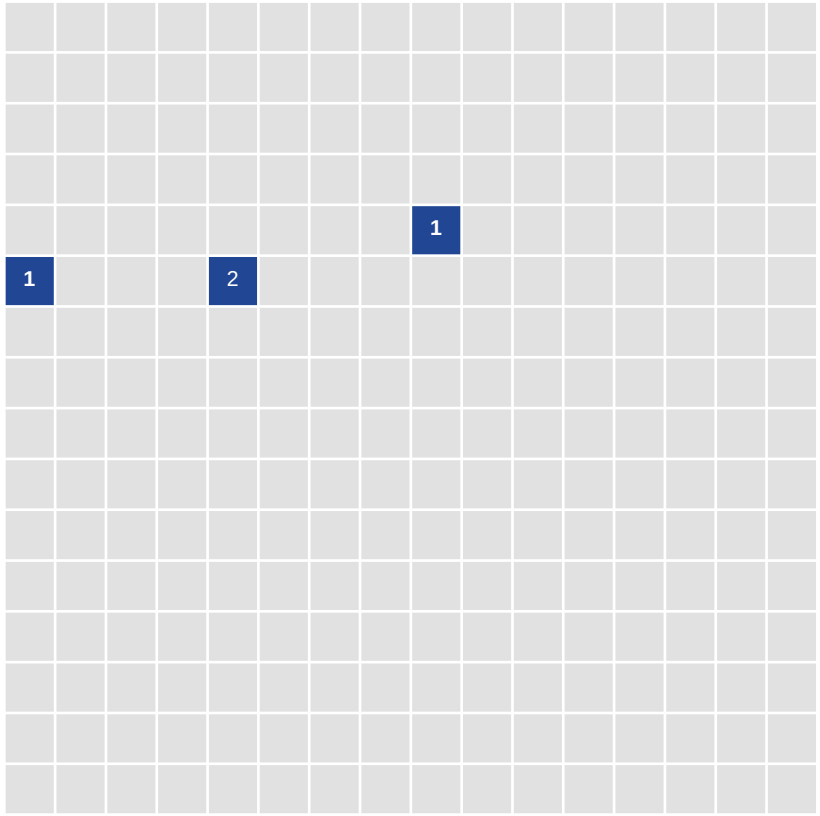
Byte Distribution

```
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
```

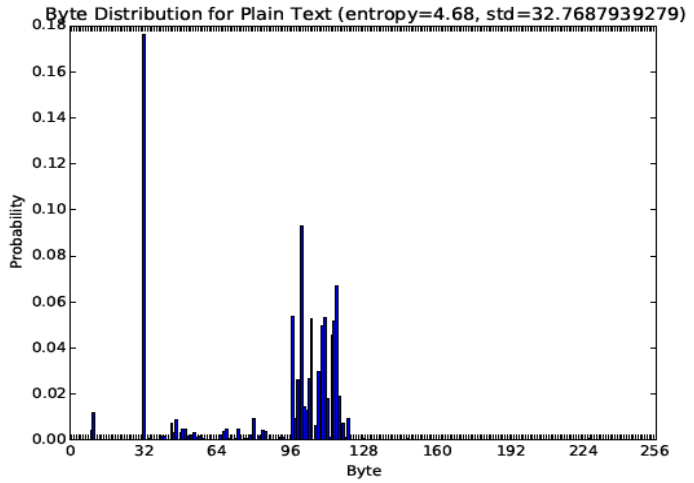


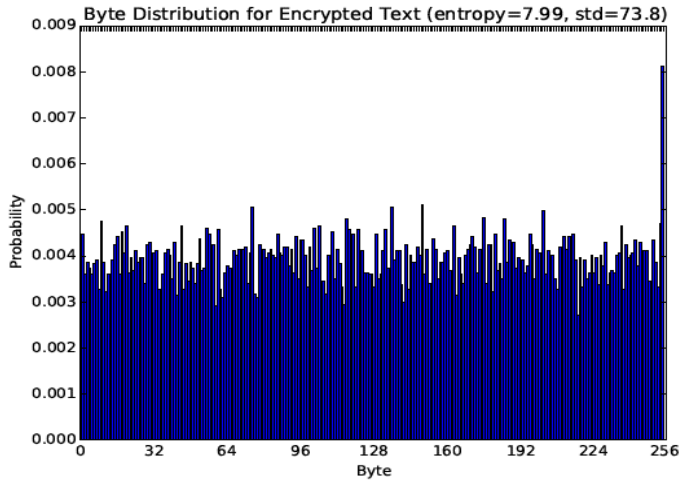
Byte Distribution

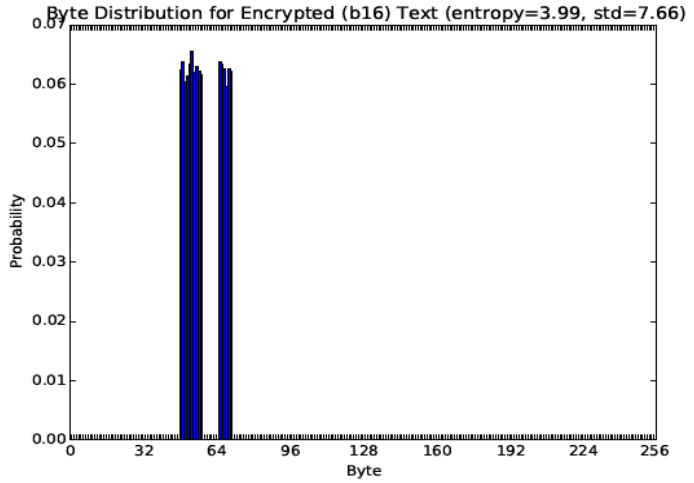
```
H  T  T  P  /  1  .  1     2  0  0     O  K
48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b
```

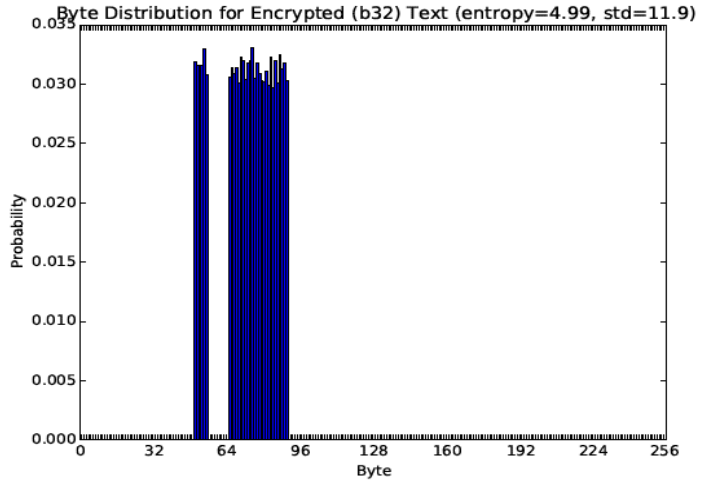


Byte Distribution for different encodings









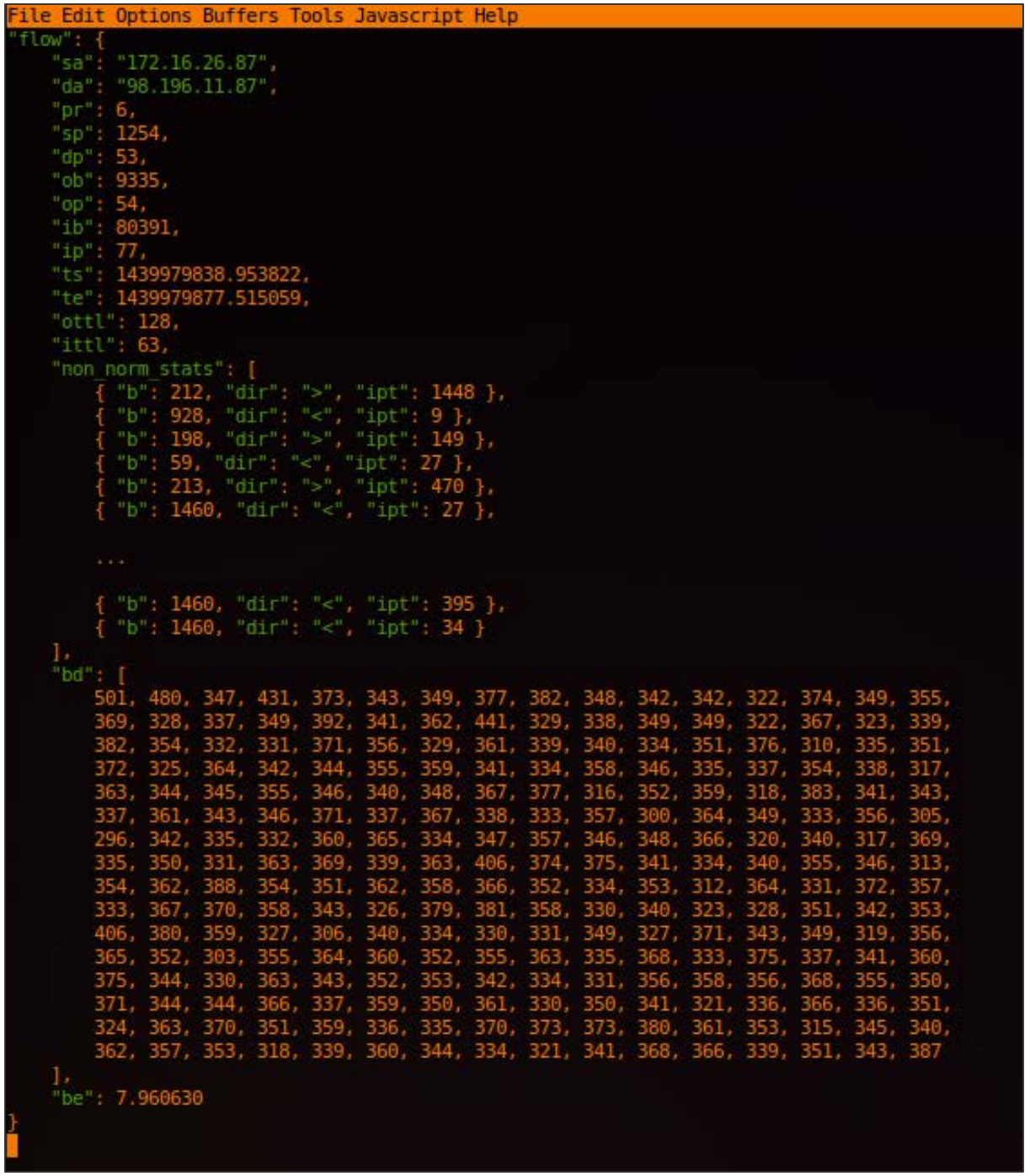
JSON flow data

Conventional flow data

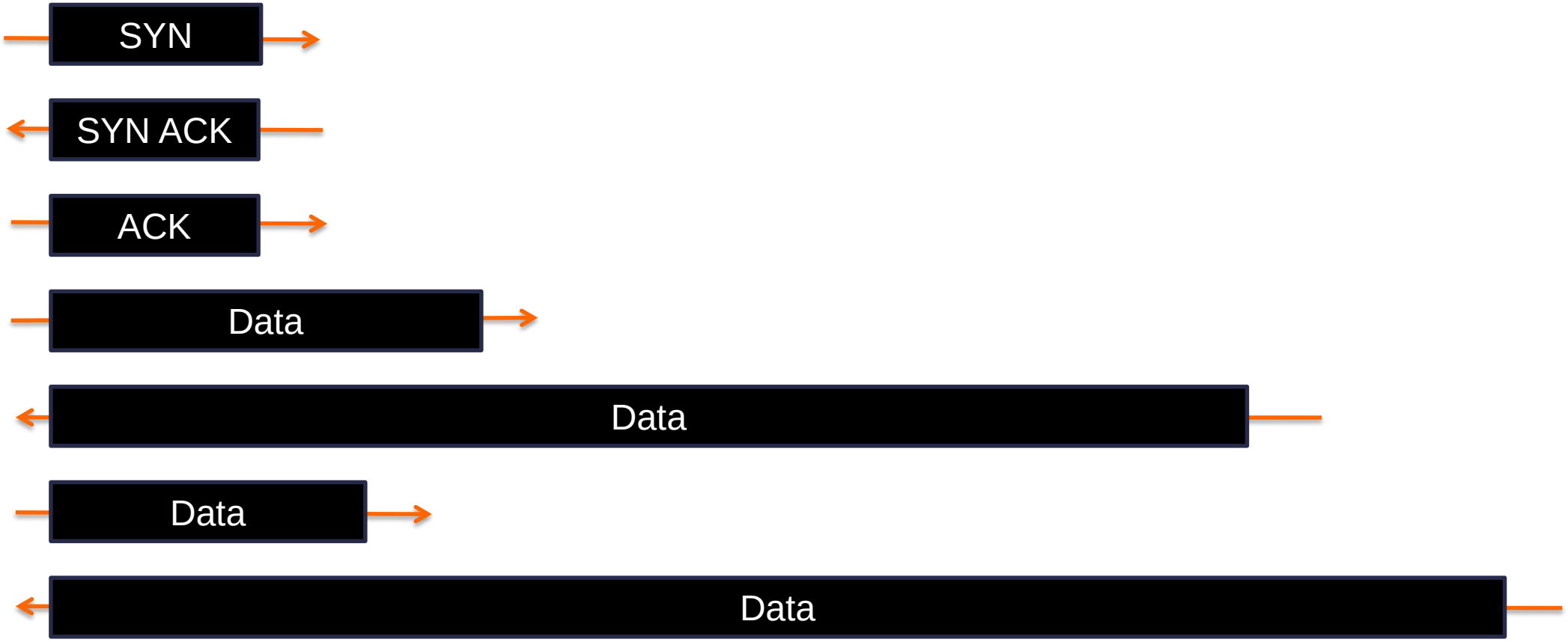
Intraflow data

Extracted parameters

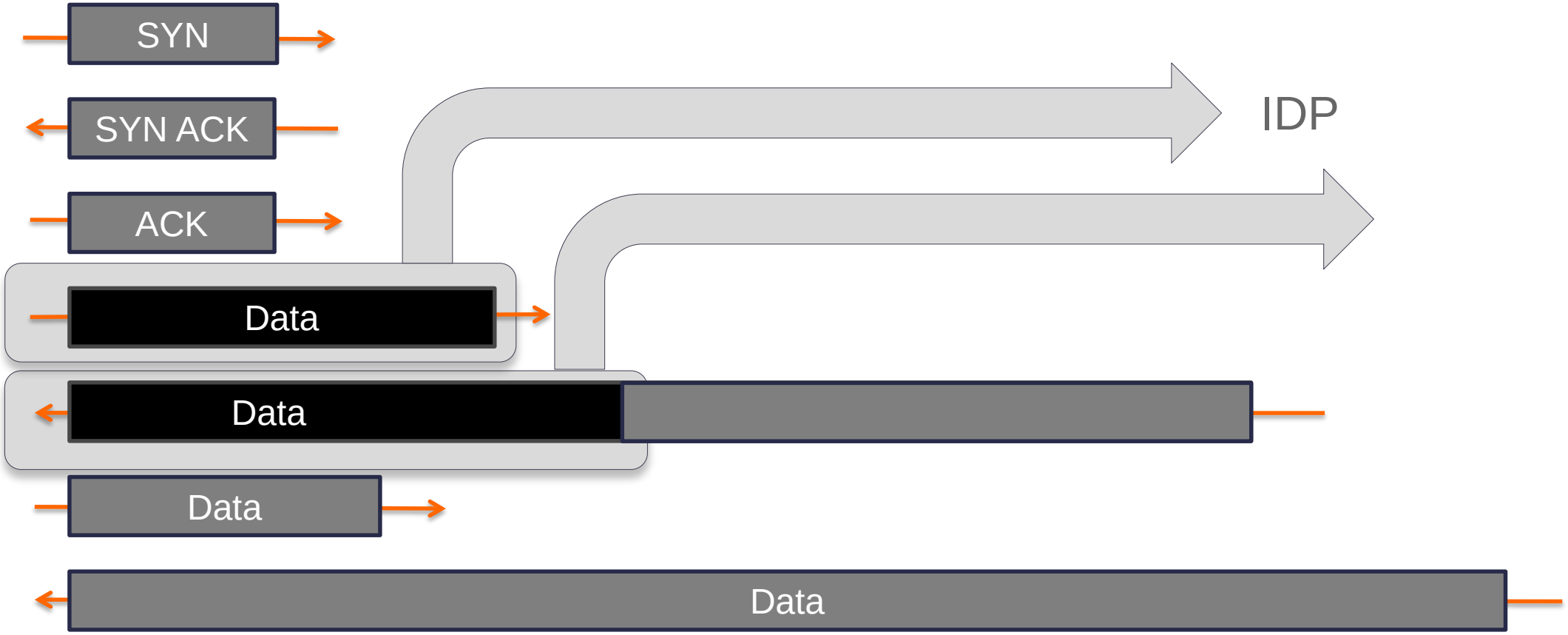
Easy to use with data analytics and machine learning tools



Initial Data Packet



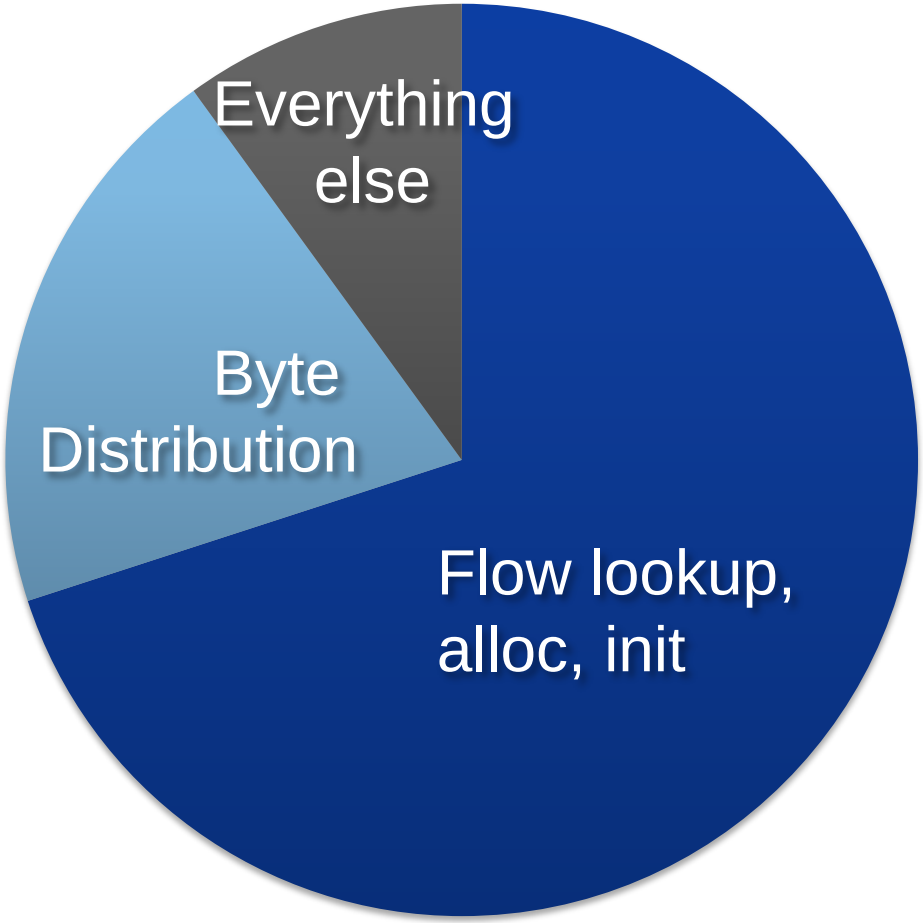
Initial Data Packet



Experimental results

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

Analytics User Interface

LAUI 0.2a Home Contact Local Analytics

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

LAUI 0.2a Home Contact Local Analytics

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

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	Byte Distribution
Economical observation	Yes	Yes
Unidirectional	Yes	Yes
Minimal computation	Yes	No
Small snaplen	Yes	Yes
Application/protocol independence		
Compactness		
Observation	10 bytes	256 bytes
Transmission and storage	10 bytes	256 bytes
Composability	Yes	Yes

Intraflow data

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

	SPLT 10 packets	Byte Distribution
Economical observation	Yes	Yes
Unidirectional	Yes	Yes
Minimal computation	Yes	No
Small snaplen	Yes	Yes
Application/protocol independence		
Compactness	10 bytes	256 bytes
Observation	10 bytes	256 bytes
Transmission and storage	Yes	Yes
Composability		

16 bytes

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": {
"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": 431 },
{ "b": 54, "dir": ">", "ipt": 360 },
{ "b": 206, "dir": ">", "ipt": 206 },
{ "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 },

Thank You

Joy applications