Entropy/IP: Uncovering Structure in IPv6 Addresses

ACM IMC 2016, Santa Monica, USA

Paweł Foremski, David Plonka, Arthur Berger

What's Entropy/IP?

A **system** that automatically learns structures in Internet addresses known to be active

Combines Entropy, Machine Learning, and Probabilistic Graphical Models

Goal: **insight** into addressing plans of IPv6 networks

Application: IPv6 **scanning vulnerability**

Background: IPv6 addressing

- Is IPv6 *addressing* just “more addresses”?

- Quantitative change: **2^32 --> 2^128**
- But… qualitative implications

- **IPv6 made the addressing space *sparse***

More freedom in address assignment

Background: IPv6 examples

- How to assign an IPv6 address? (in general)

[network ID (64 bits)] + [interface ID (64 bits)]

2001:db8:0010:0001::103	fixed
2001:db8:0167:1109::10:901	structured
2001:db8:0000:1cdf:21e:c2ff:fec0:11db	EUI-64
2001:db8:4137:9e76:3031:f3fd:bbdd:2c2a	ephemeral

No Single Algorithm

Background: No Single Algorithm

`[network ID (64 bits)] + [interface ID (64 bits)]`

- Interface Identifier (IID):
  - Stateless Address Autoconfiguration (SLAAC) e.g. RFC 4862
  - **Static / Other**
- Network Identifier:
  - Routing prefixes (e.g. BGP)
  - **Static / Other**

IPv6 networks adopt their own addressing schemes

Background: motivations for Entropy/IP

- Remotely glean IPv6 addressing scheme:
  - Which bits are used / unused ?
  - What are the most common values ?
  - What is the syntax ?
- Provide supportive information for:
  - Classifying addresses (e.g. host reputation)
  - Scanning / defending IPv6 scanning
  - Measuring the growth of IPv6 networks

IPv6 users:

World	>12%
USA	>29%
Belgium	>48%

Entropy/IP: operation overview

1. Entropy Analysis
2. Address Segmentation
3. Segment Mining
4. Bayesian Modeling

1. Entropy Analysis: input

```
2001:0db8:0010:0013:0000:0000:0000:07fe
2001:0db8:0010:0000:0000:0000:0000:0ed3
2001:0db8:0010:0003:0000:0000:0000:0fb5
2001:0db8:0020:d05f:882f:6082:f768:710d
2001:0db8:0010:0004:0000:0000:0000:04dc
2001:0db8:0010:0003:0000:0000:0000:03ce
2001:0db8:0010:0008:0000:0000:0000:0794
2001:0db8:0010:000a:0000:0000:0000:0923
2001:0db8:0010:0006:0000:0000:0000:003c
2001:0db8:0022:1014:aef6:60af:d029:63cd
2001:0db8:0010:0012:0000:0000:0000:0c7b
2001:0db8:0022:10c0:5100:ac7d:96f5:5851
2001:0db8:0010:0002:0000:0000:0000:0de8
2001:0db8:0010:0008:0000:0000:0000:0506
2001:0db8:0022:2053:4e6a:a11a:d57f:e26d
(...)
```

1. Entropy Analysis: operation

```
2001:0db8:0010:0013:0000:0000:0000:07fe
2001:0db8:0010:0000:0000:0000:0000:0ed3
2001:0db8:0010:0003:0000:0000:0000:0fb5
2001:0db8:0020:d05f:882f:6082:f768:710d
2001:0db8:0010:0004:0000:0000:0000:04dc
2001:0db8:0010:0003:0000:0000:0000:03ce
2001:0db8:0010:0008:0000:0000:0000:0794
2001:0db8:0010:000a:0000:0000:0000:0923
2001:0db8:0010:0006:0000:0000:0000:003c
2001:0db8:0022:1014:aef6:60af:d029:63cd
2001:0db8:0010:0012:0000:0000:0000:0c7b
2001:0db8:0022:10c0:5100:ac7d:96f5:5851
2001:0db8:0010:0002:0000:0000:0000:0de8
2001:0db8:0010:0008:0000:0000:0000:0506
2001:0db8:0022:2053:4e6a:a11a:d57f:e26d
(...)
```

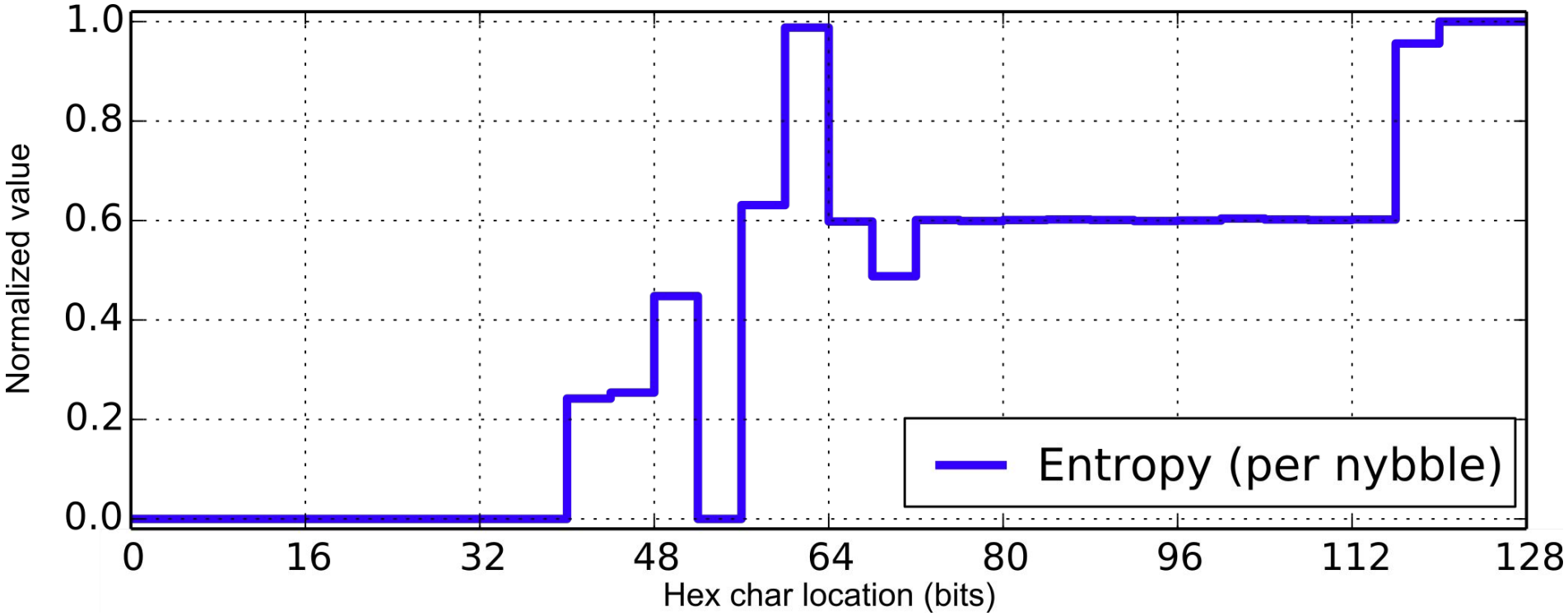
For a discrete random variable X:

$$H(X) = -\sum_{i=1}^{k} P(x_i) \log P(x_i)$$

$H(X_{16}) = 3.8$ /4

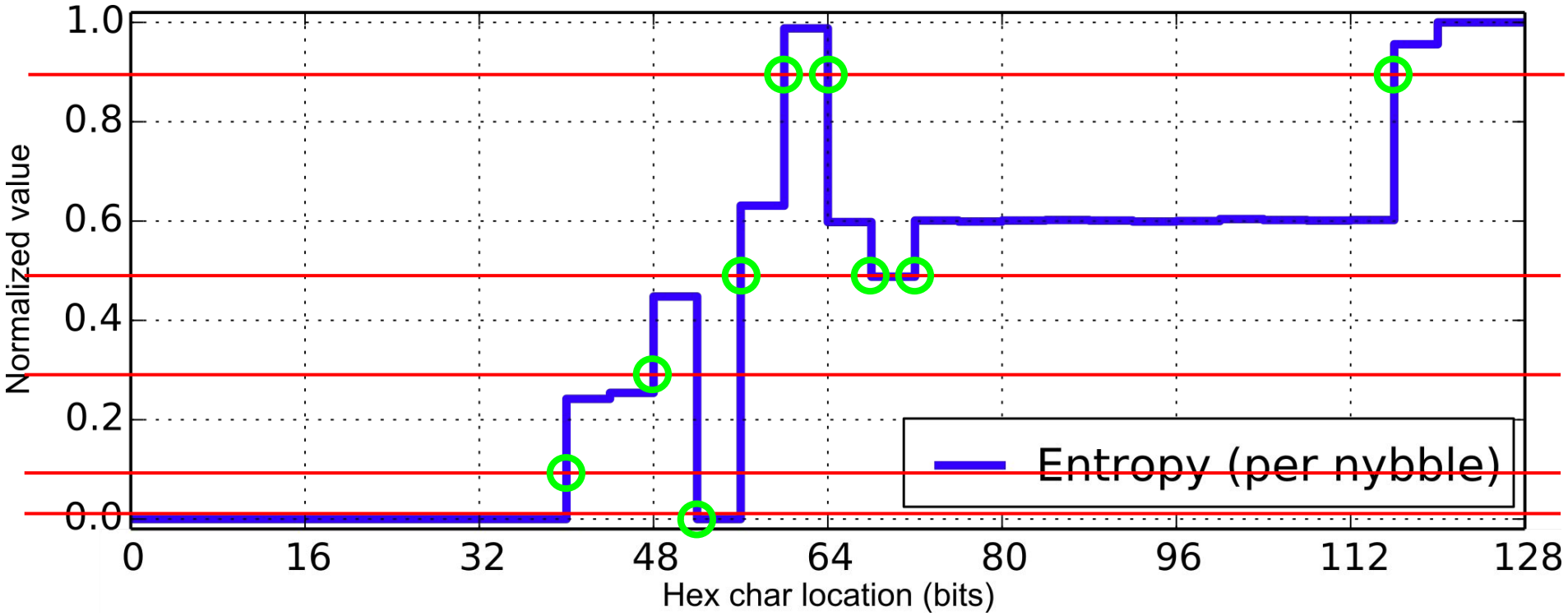
$H(X_{18}) = 2.2$ /4

1. Entropy Analysis: hex character variability



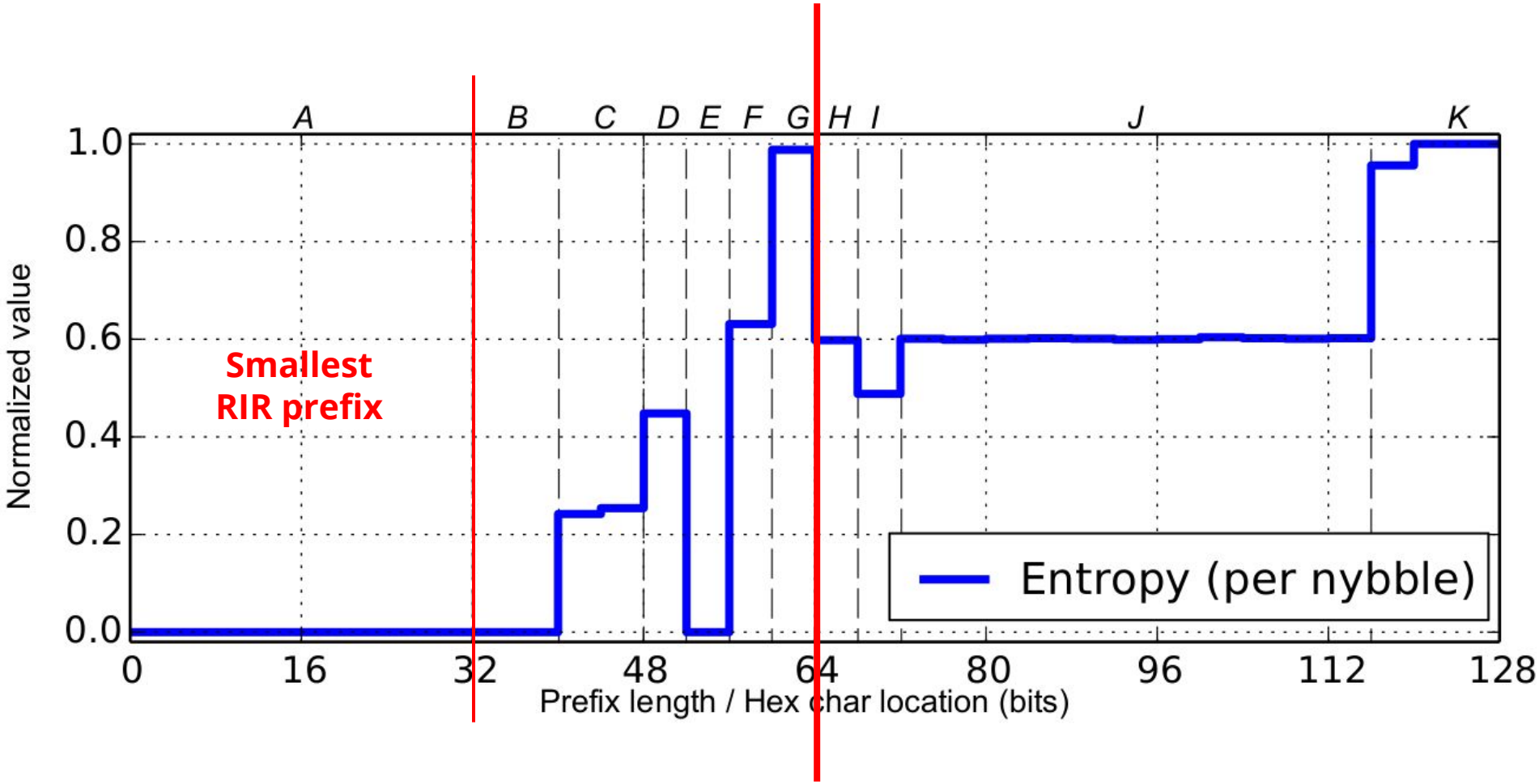
2. Address Segmentation: group by similar entropy

$$T = \{0.025, 0.1, 0.3, 0.5, 0.9\}$$



$$|\hat{H}(X_i) - \hat{H}(X_{i-1})| > T_h \qquad (T_h = 0.05)$$

2. Address Segmentation: list of bit ranges



3. Segment Mining: what's inside?

Extract ***all values*** D_k from given ***segment*** k, and find:

a) Most popular values $> Q_3 + 1.5 \times IQR$
   - ➢ e.g. find constants, enumerations, etc.

b) Densely packed *ranges* of values $DBSCAN(values)$
   - ➢ e.g. find adjacent subnets

c) Uniform distributions $DBSCAN(histogram)$
   - ➢ e.g. find counters, randoms

d) Summarize what's left $[\ min(D_k),\ max(D_k)\ ]$

3. Segment Mining: output & encoding

	Code	Value	Frequency
C (40-48)	C1	00	67.02%
	C2	01	11.13%
	C3	c2	0.67%
	C4	fe	0.41%
	C5	ff	0.41%
	C6	02-5b	11.94%
	C7	5c-fd	8.42%

2001:0db8:0841:2500:0000:d9a0:5345:0012

2001:0db8:0841:2500:0000:d9a0:5345:0012

(A1, B2, C6, D4, E5, F1, G12, H1, I2, J3)

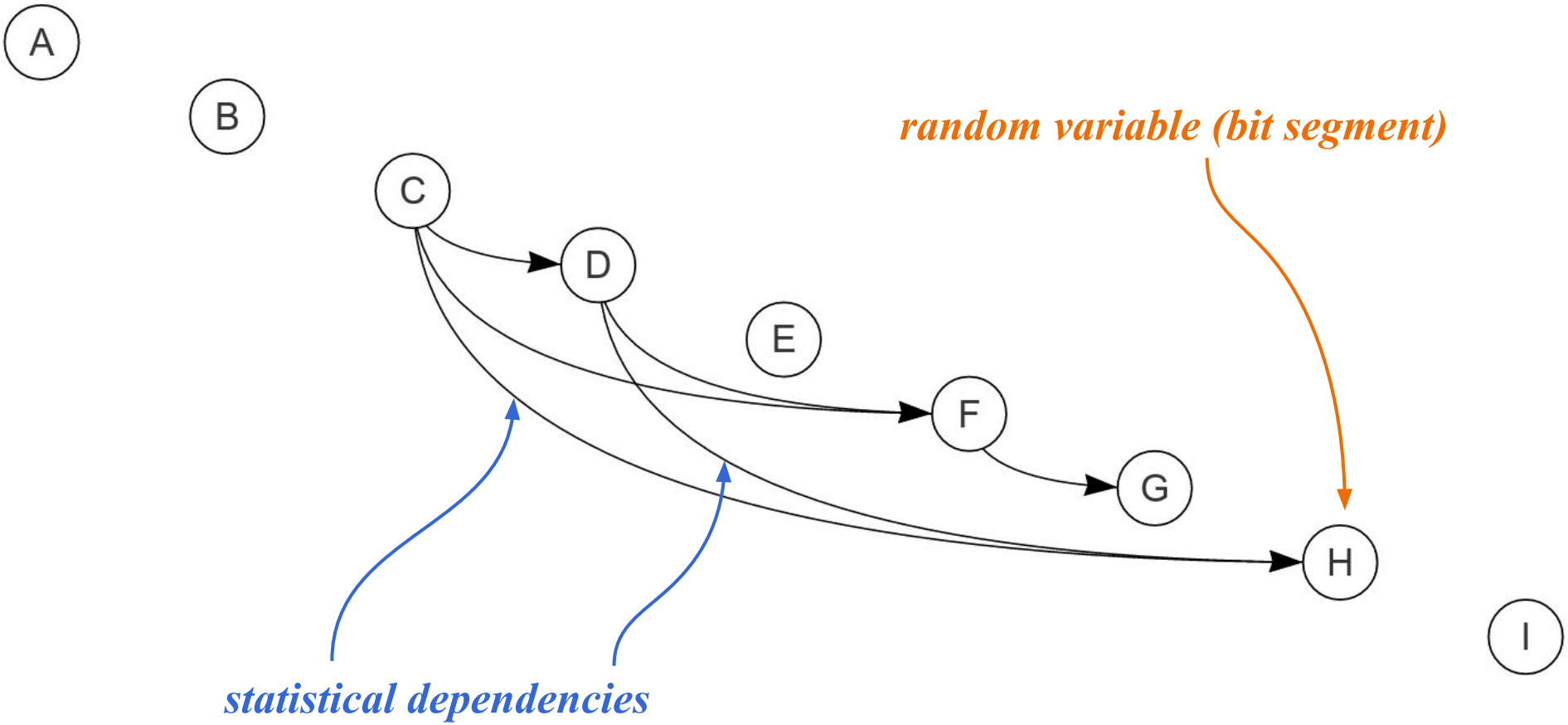
4. Bayesian Network: segment inter-dependencies

```
2001:0db8:0010:0004:0000:0000:0000:03cc
2001:0db8:0010:0003:0000:0000:0000:0f97
2001:0db8:0022:1028:9e83:1334:17c0:897a
2001:0db8:0022:3064:69f5:02d2:f223:8635
2001:0db8:0010:0014:0000:0000:0000:0347
2001:0db8:0010:0014:0000:0000:0000:022a
2001:0db8:0010:0005:0000:0000:0000:03ca
2001:0db8:0010:0015:0000:0000:0000:0ae9
2001:0db8:0021:0056:8032:6eb3:6098:3084
2001:0db8:0010:0003:0000:0000:0000:018b
2001:0db8:0010:0002:0000:0000:0000:0424
2001:0db8:0010:0013:0000:0000:0000:0e2f
2001:0db8:0022:20a4:3eb9:5fca:3ccb:2aae
2001:0db8:0021:0014:3326:6434:74c9:aad6
2001:0db8:0010:000f:0000:0000:0000:07bd
(...)
```

4. Bayesian Network: segment inter-dependencies

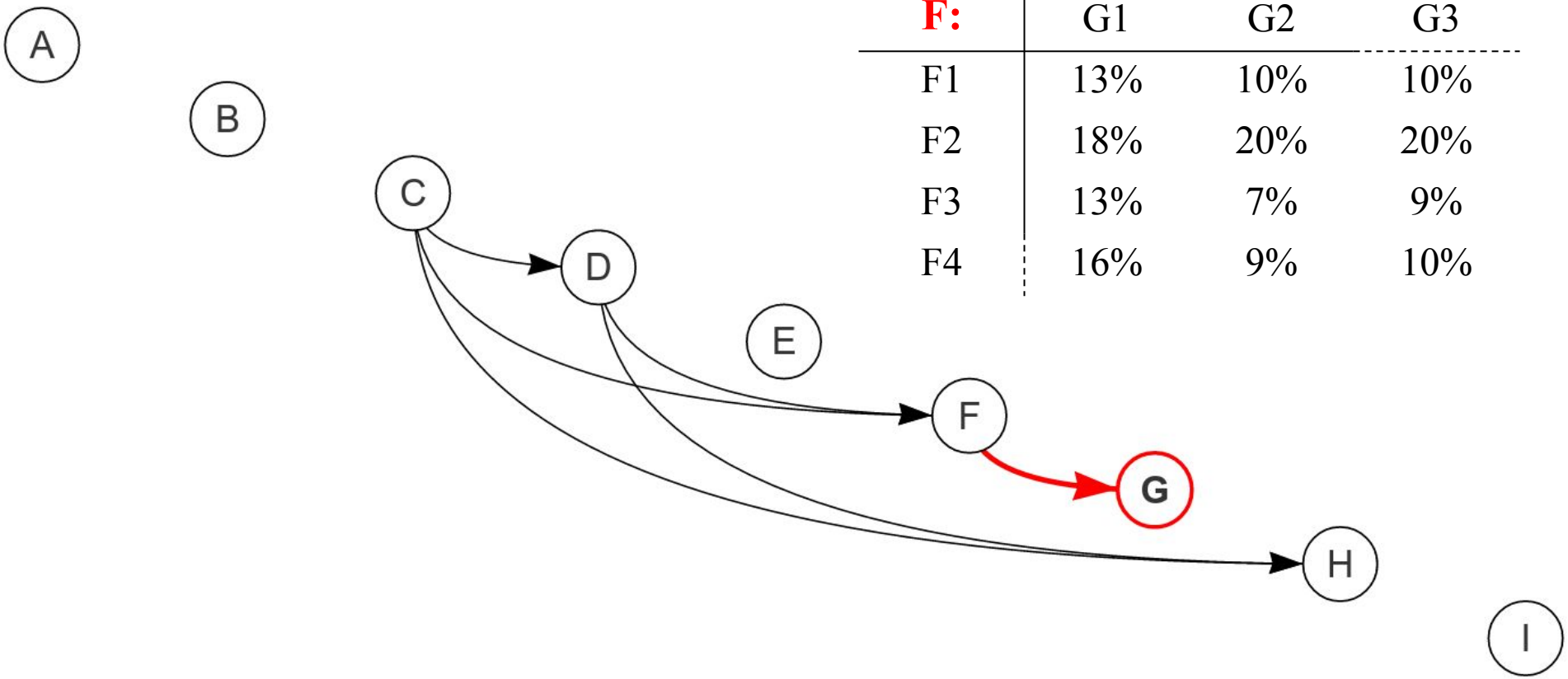
2001:0db8:0010:0004:0000:0000:0000:03cc	(A1, B1, C1, D1, E1, F1, G3, H1, I11)
2001:0db8:0010:0003:0000:0000:0000:0f97	(A1, B1, C1, D1, E1, F1, G1, H1, I11)
2001:0db8:0022:1028:9e83:1334:17c0:897a	(A1, B1, C2, D2, E1, F5, G4, H2, I11)
2001:0db8:0022:3064:69f5:02d2:f223:8635	(A1, B1, C2, D3, E1, F3, G3, H2, I11)
2001:0db8:0010:0014:0000:0000:0000:0347	(A1, B1, C1, D1, E1, F2, G3, H1, I11)
2001:0db8:0010:0014:0000:0000:0000:022a	(A1, B1, C1, D1, E1, F2, G3, H1, I11)
2001:0db8:0010:0005:[illegible]	(A1, B1, C1, D1, E1, F1, G2, H1, I11)
2001:0db8:0010:0015:[illegible]	(A1, B1, C1, D1, E1, F2, G2, H1, I11)
2001:0db8:0021:0056:8032:6eb3:6098:3084	(A1, B1, C3, D1, E1, F4, G8, H2, I11)
2001:0db8:0010:0003:0000:0000:0000:018b	(A1, B1, C1, D1, E1, F1, G1, H1, I11)
2001:0db8:0010:0002:0000:0000:0000:0424	(A1, B1, C1, D1, E1, F1, G8, H1, I11)
2001:0db8:0010:0013:0000:0000:0000:0e2f	(A1, B1, C1, D1, E1, F2, G1, H1, I11)
2001:0db8:0022:20a4:3eb9:5fca:3ccb:2aae	(A1, B1, C2, D4, E1, F6, G3, H2, I11)
2001:0db8:0021:0014:3326:6434:74c9:aad6	(A1, B1, C3, D1, E1, F2, G3, H2, I11)
2001:0db8:0010:000f:0000:0000:0000:07bd	(A1, B1, C1, D1, E1, F1, G8, H1, I11)
(...)	

4. Bayesian Network: dependency graph



4. Bayesian Network: conditional probabilities

F:	G: G1	G2	G3
F1	13%	10%	10%
F2	18%	20%	20%
F3	13%	7%	9%
F4	16%	9%	10%

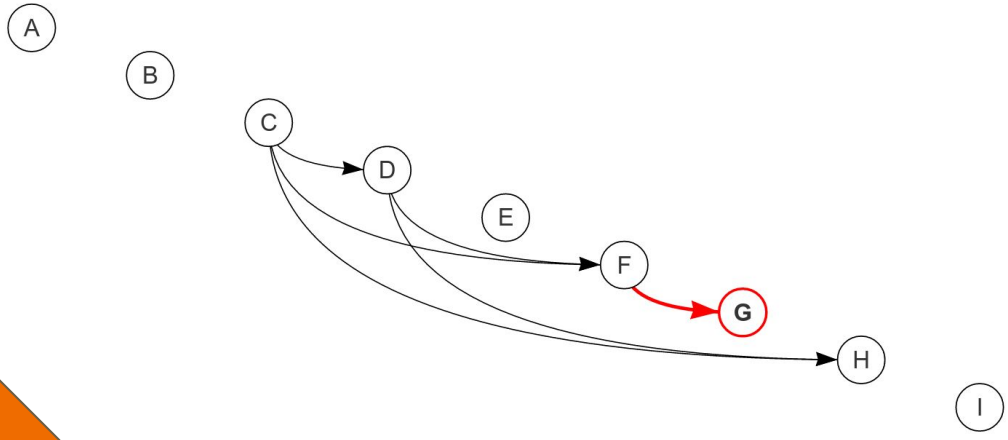


4. Bayesian Network: how to find it?

(A1, B1, C1, D1, E1, F1, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C2, D2, E1, F5, G4, H2, I11)
(A1, B1, C2, D3, E1, F3, G3, H2, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G2, H1, I11)
(A1, B1, C1, D1, E1, F2, G2, H1, I11)
(A1, B1, C3, D1, E1, F4, G8, H2, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)
(A1, B1, C1, D1, E1, F2, G1, H1, I11)
(A1, B1, C2, D4, E1, F6, G3, H2, I11)
(A1, B1, C3, D1, E1, F2, G3, H2, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)

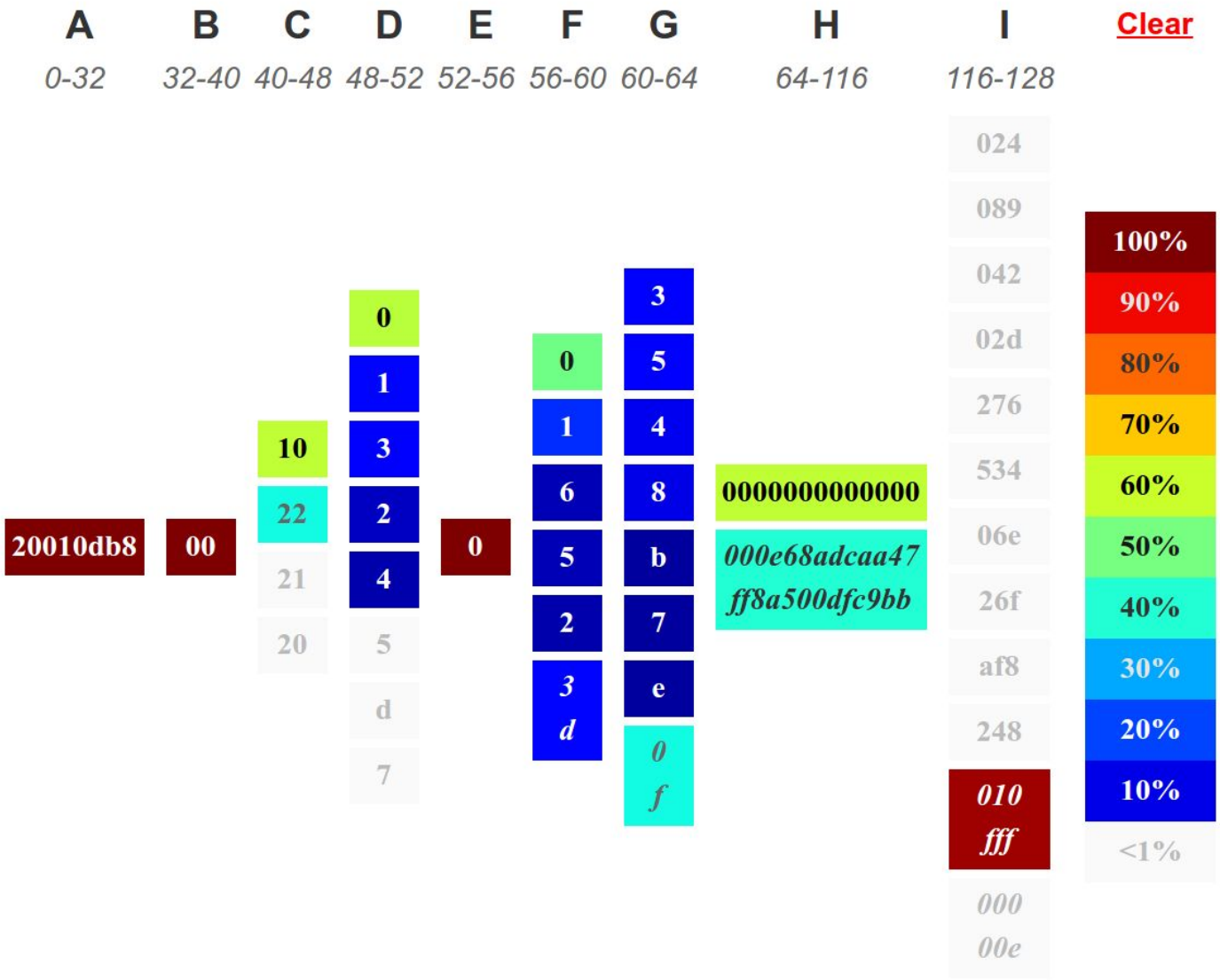
4. Bayesian Network: BNfinder

(A1, B1, C1, D1, E1, F1, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C2, D2, E1, F5, G4, H2, I11)
(A1, B1, C2, D3, E1, F3, G3, H2, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G2, H1, I11)
(A1, B1, C1, D1, E1, F2, G2, H1, I11)
(A1, B1, C3, D1, E1, F4, G8, H2, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)
(A1, B1, C1, D1, E1, F2, G1, H1, I11)
(A1, B1, C2, D4, E1, F6, G3, H2, I11)
(A1, B1, C3, D1, E1, F2, G3, H2, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)

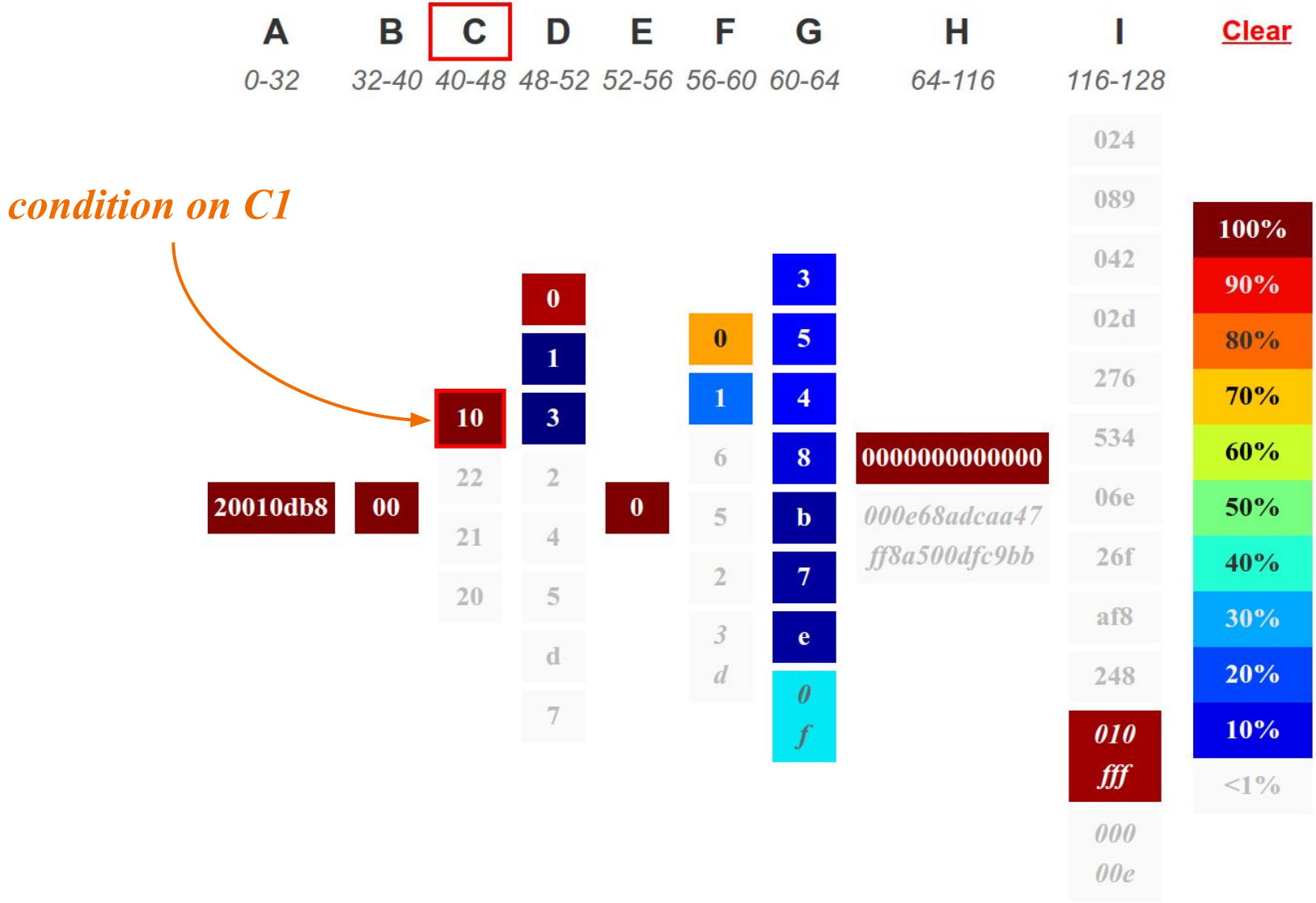


	G:		
F:	G1	G2	G3
F1	13%	10%	10%
F2	18%	20%	20%
F3	13%	7%	9%
F4	16%	9%	10%

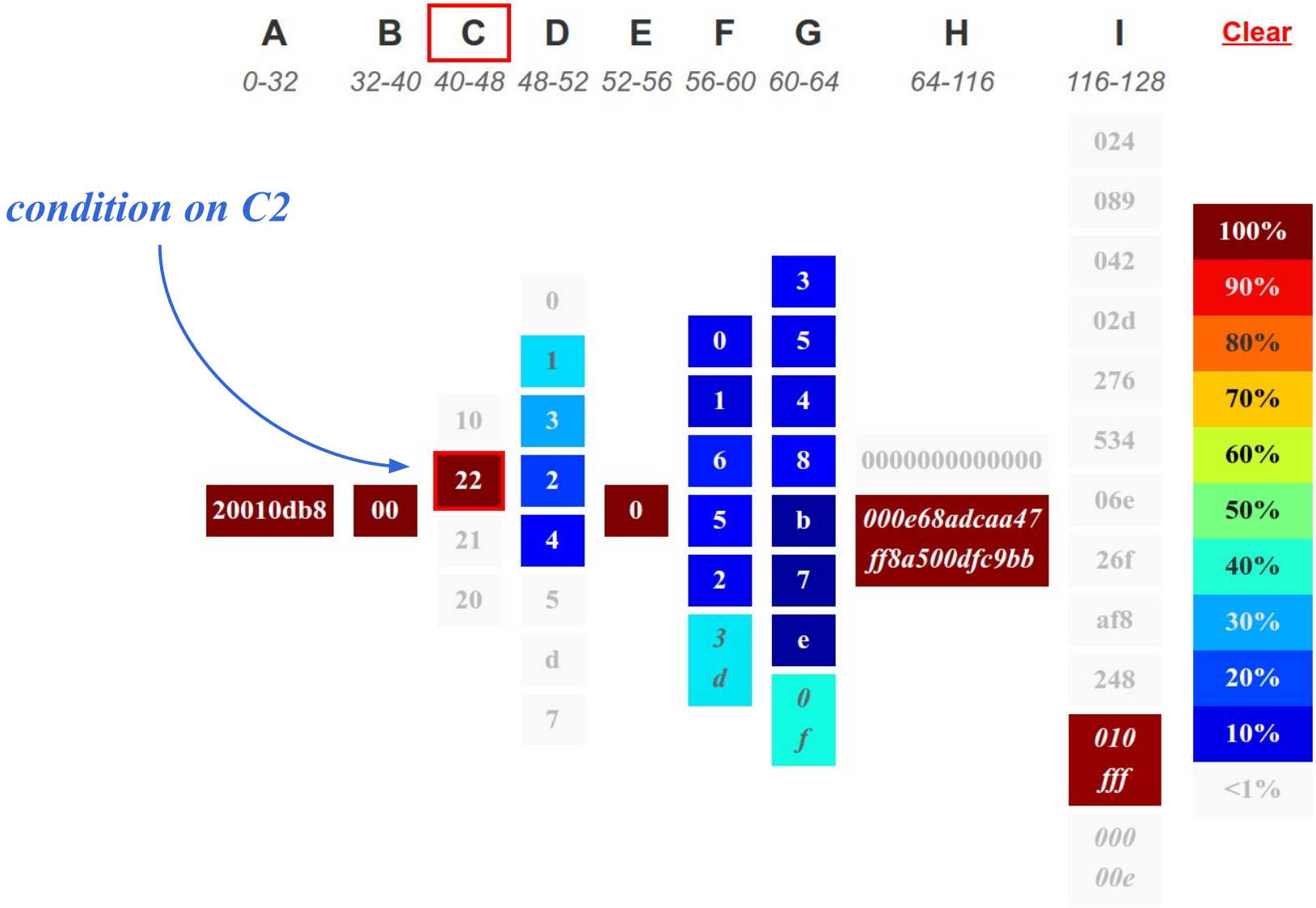
4. Bayesian Network: visualization



4. Bayesian Network: visualization (2)



4. Bayesian Network: visualization (3)



Evaluation: data

- Q1 2016
- 3.5 billion IPs
- DNS
- Traceroutes
- CDN logs

Type	ID	Data Sources				
		DNSDB	FDNS	rDNS	TR	CDN
Servers	S1	110 K	180 K	-		
	S2	290 K	4.7 K	-		
	S3	7.5 K	65 K	-		
	S4	12 K	5.7 K	-		
	S5	33 K	1.7 K	30 K		
	AS	790 K				
Routers	R1	-	28 K	1.8 K	6.7 M	
	R2	-	55 K	-	180 K	
	R3	460	10 K	11 K	7.5 K	
	R4	50	-	2.5 K	900	
	R5	10	-	1.3 K	380	
	AR				12 M	
Clients	C1					83 M
	C2					8.2 M
	C3					530 M
	C4					39 M
	C5					43 M
	AC					3.5 G

Evaluation: data

- Q1 2016
- 3.5 billion IPs
- DNS
- Traceroutes
- CDN logs

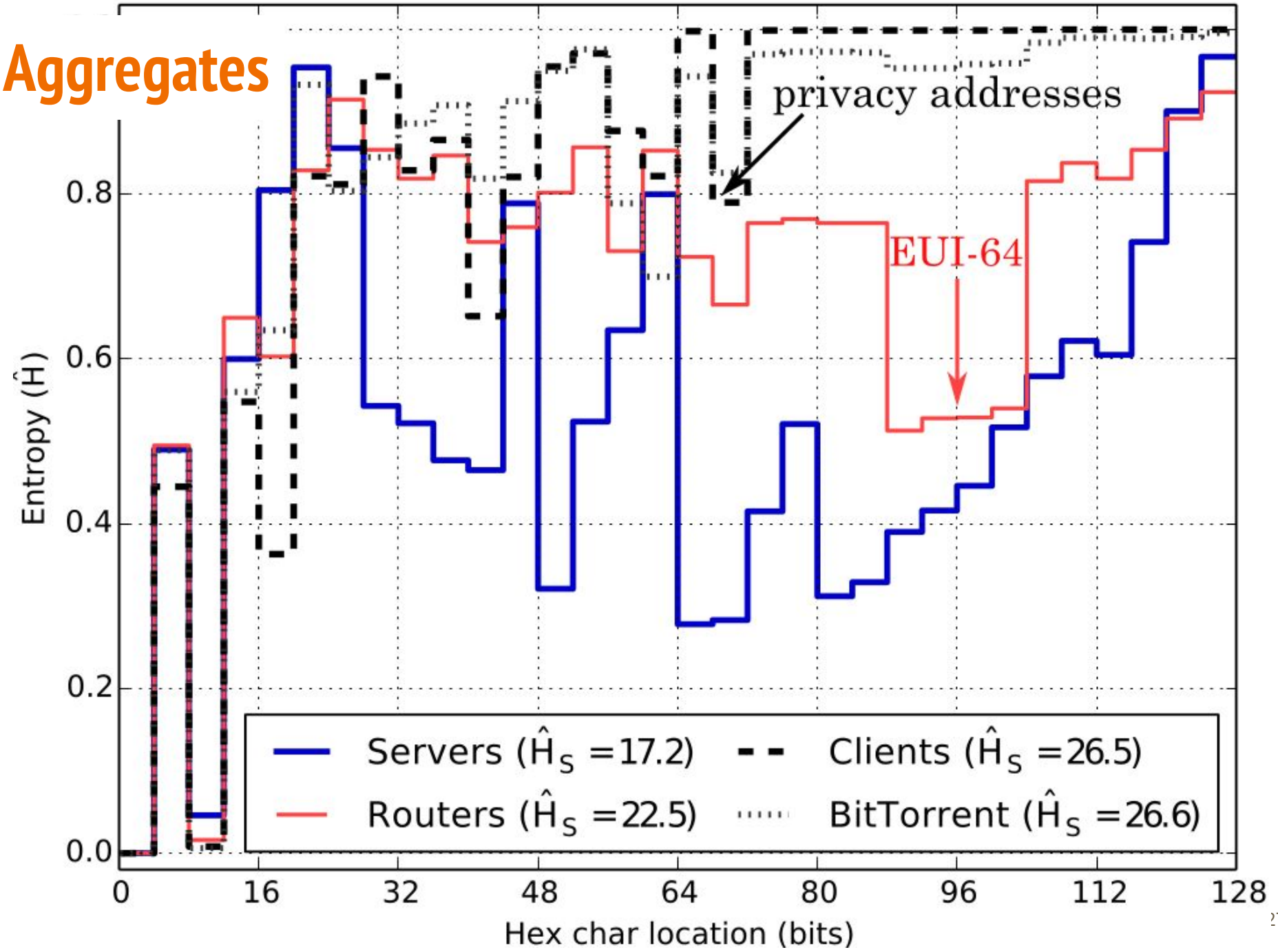
Type	ID	Data Sources				
		DNSDB	FDNS	rDNS	TR	CDN
Servers	S1	110 K	180 K	-		
	S2	290 K	4.7 K	-		
	S3	7.5 K	65 K	-		
	S4	12 K	5.7 K	-		
	S5	33 K	1.7 K	30 K		
	AS	790 K				
Routers	R1	-	28 K	1.8 K	6.7 M	
	R2	-	55 K	-	180 K	
	R3	460	10 K	11 K	7.5 K	
	R4	50	-	2.5 K	900	
	R5	10	-	1.3 K	380	
	AR				12 M	
Clients	C1					83 M
	C2					8.2 M
	C3					530 M
	C4					39 M
	C5					43 M
	AC					3.5 G

Evaluation: data

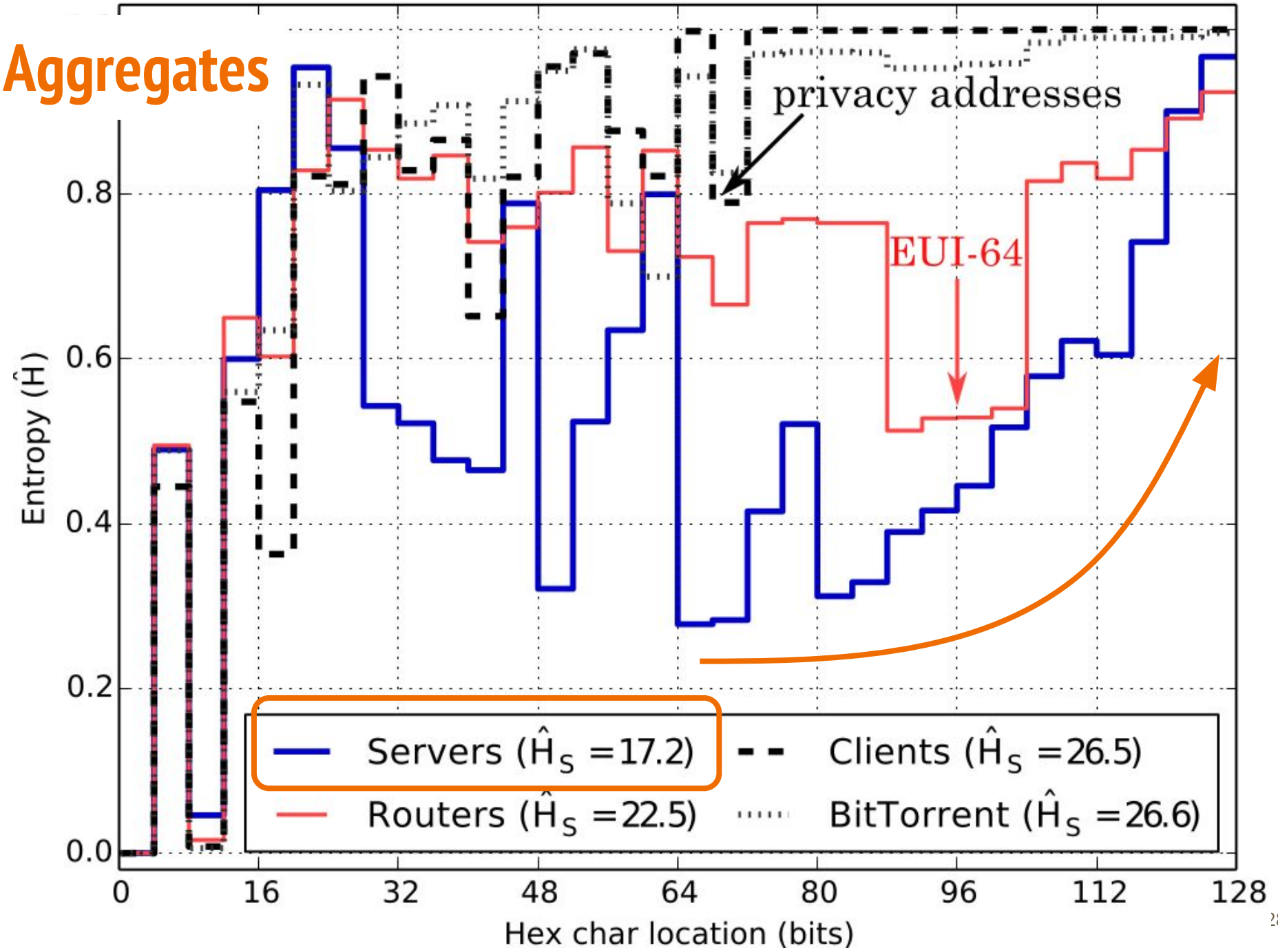
- Q1 2016
- 3.5 billion IPs
- DNS
- Traceroutes
- CDN logs

Type	ID	Data Sources				
		DNSDB	FDNS	rDNS	TR	CDN
Servers	S1	110 K	180 K	-		
	S2	290 K	4.7 K	-		
	S3	7.5 K	65 K	-		
	S4	12 K	5.7 K	-		
	S5	33 K	1.7 K	30 K		
	AS	790 K				
Routers	R1	-	28 K	1.8 K	6.7 M	
	R2	-	55 K	-	180 K	
	R3	460	10 K	11 K	7.5 K	
	R4	50	-	2.5 K	900	
	R5	10	-	1.3 K	380	
	AR				12 M	
Clients	C1					83 M
	C2					8.2 M
	C3					530 M
	C4					39 M
	C5					43 M
	AC					3.5 G

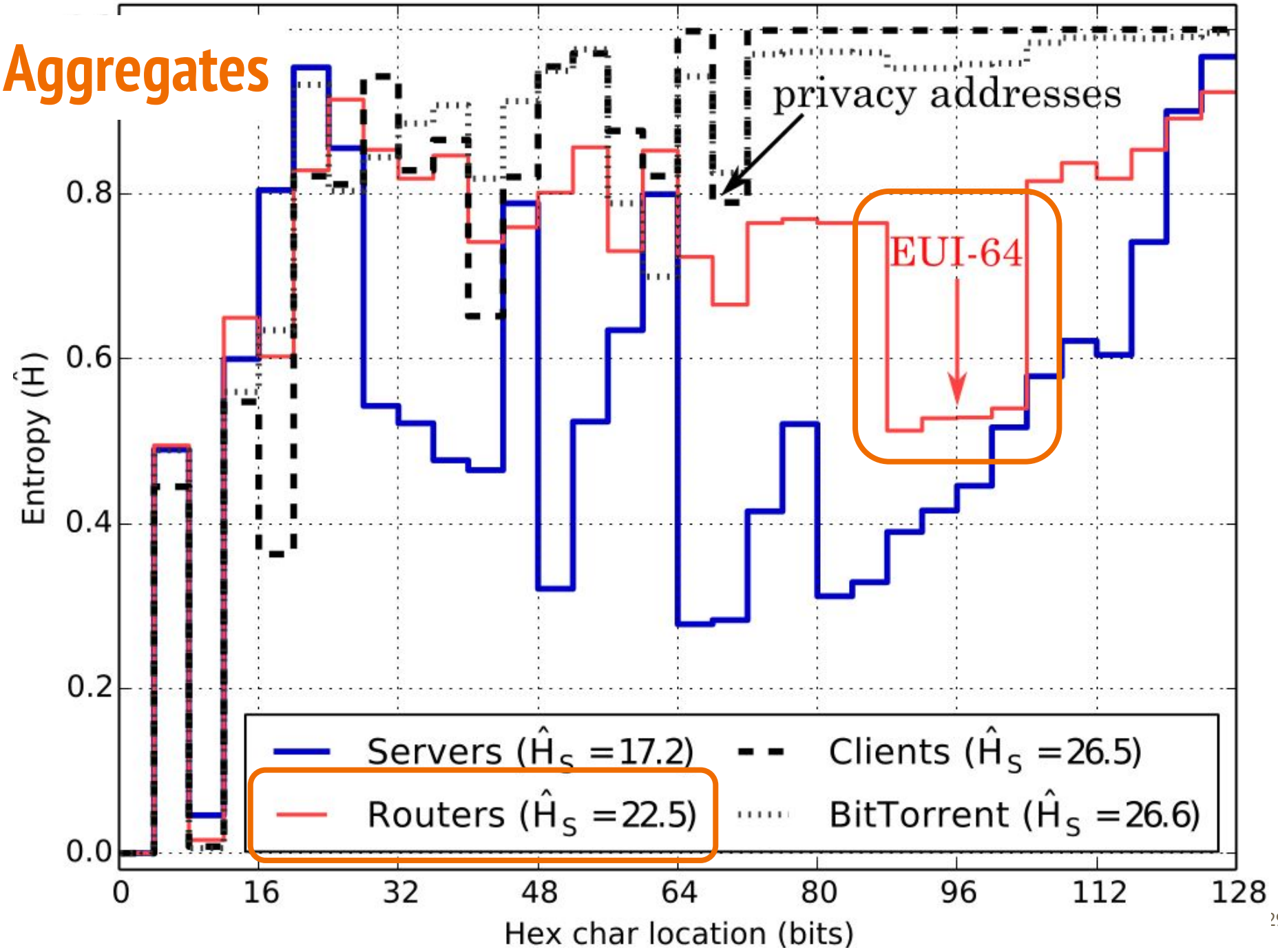
Aggregates


privacy addresses
EUI-64
Entropy ($\hat{H}$)
0.0
0.2
0.4
0.6
0.8
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Servers ($\hat{H}_S$ =17.2)
Clients ($\hat{H}_S$ =26.5)
Routers ($\hat{H}_S$ =22.5)
BitTorrent ($\hat{H}_S$ =26.6)

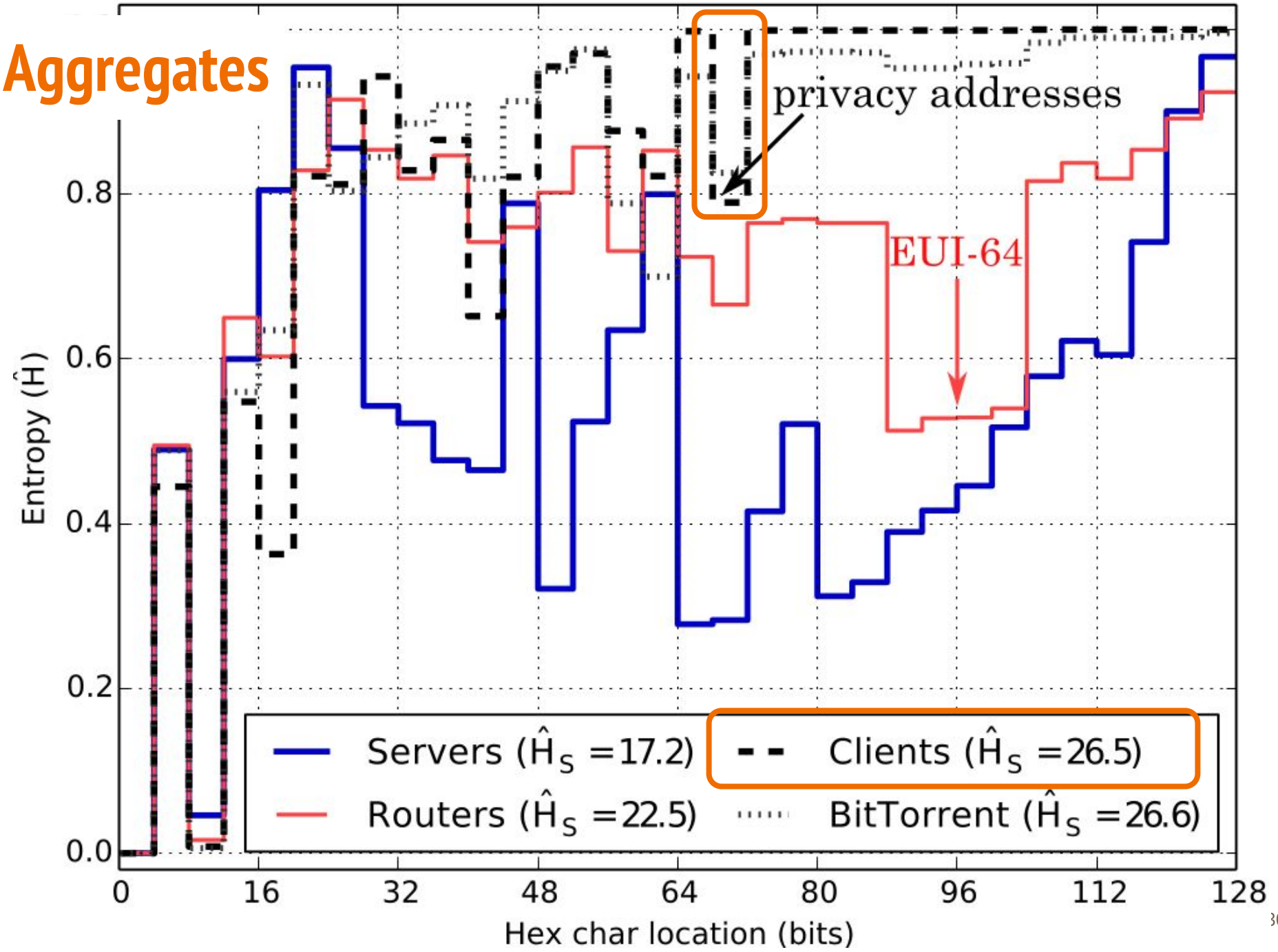
Aggregates


privacy addresses
EUI-64
Entropy ($\hat{H}$)
0.0
0.2
0.4
0.6
0.8
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Servers ($\hat{H}_S$ =17.2)
Clients ($\hat{H}_S$ =26.5)
Routers ($\hat{H}_S$ =22.5)
BitTorrent ($\hat{H}_S$ =26.6)

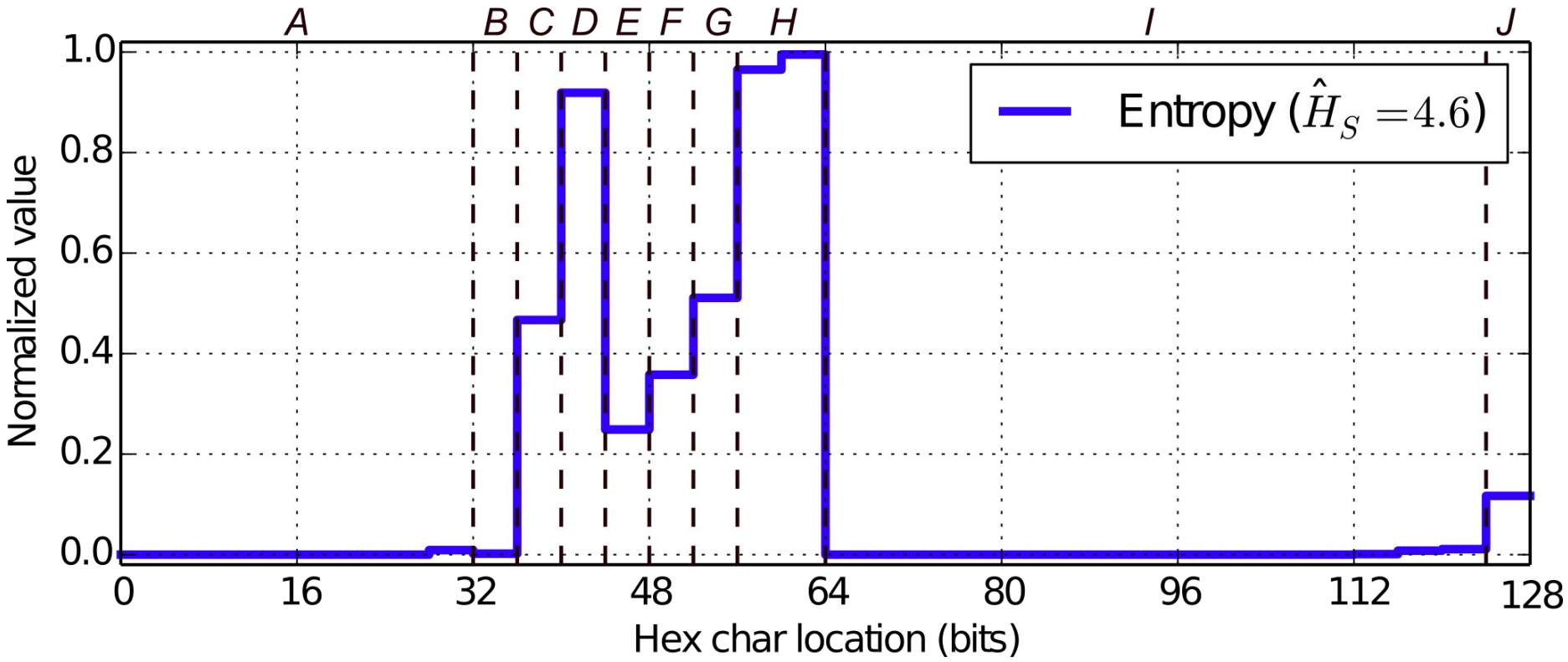
Aggregates

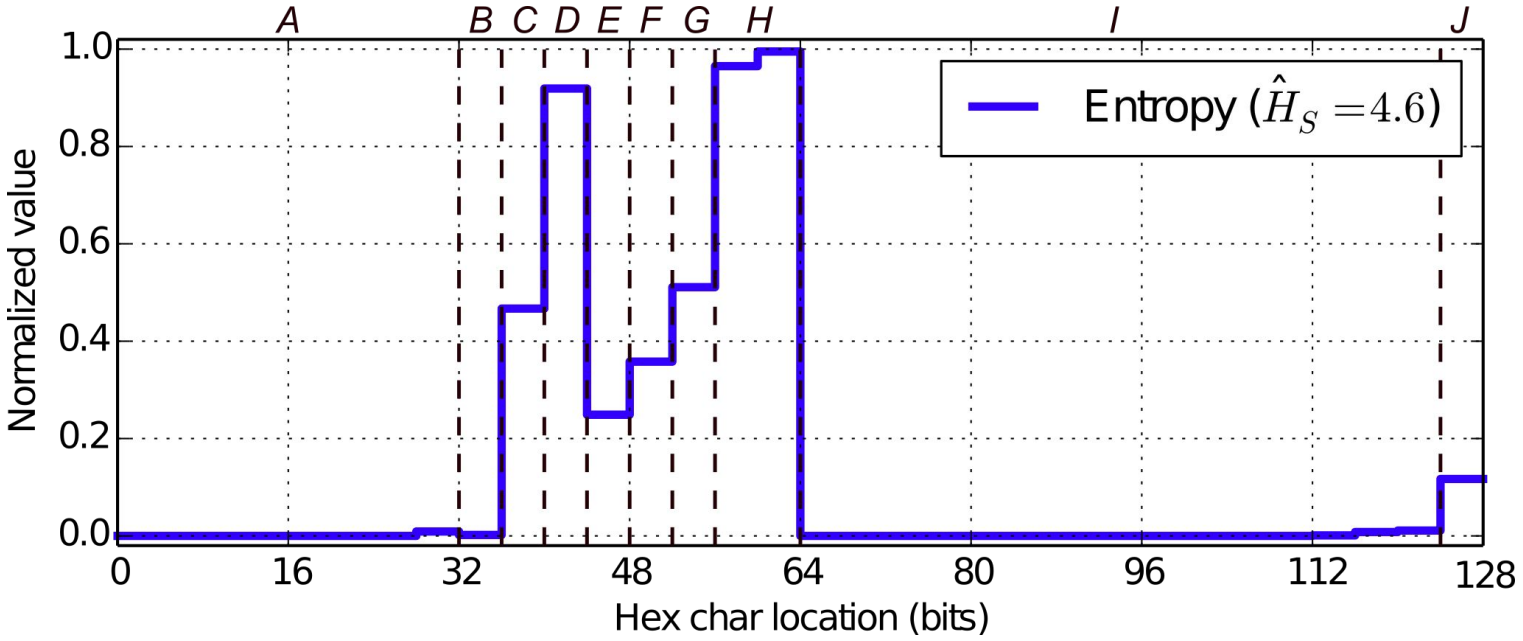

privacy addresses
EUI-64
Entropy ($\hat{H}$)
0.0
0.2
0.4
0.6
0.8
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Servers ($\hat{H}_S$ =17.2)
Clients ($\hat{H}_S$ =26.5)
Routers ($\hat{H}_S$ =22.5)
BitTorrent ($\hat{H}_S$ =26.6)

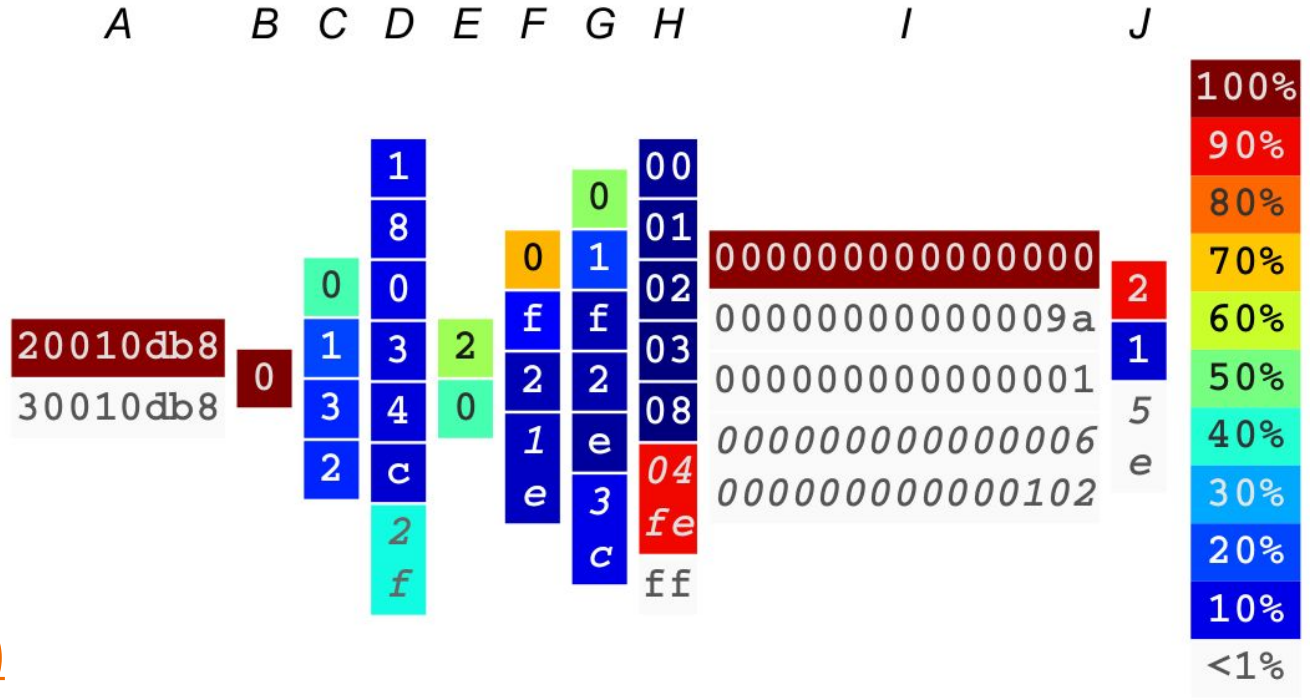
Aggregates


privacy addresses
EUI-64
Entropy ($\hat{H}$)
0.0
0.2
0.4
0.6
0.8
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Servers ($\hat{H}_S$ =17.2)
Routers ($\hat{H}_S$ =22.5)
Clients ($\hat{H}_S$ =26.5)
BitTorrent ($\hat{H}_S$ =26.6)

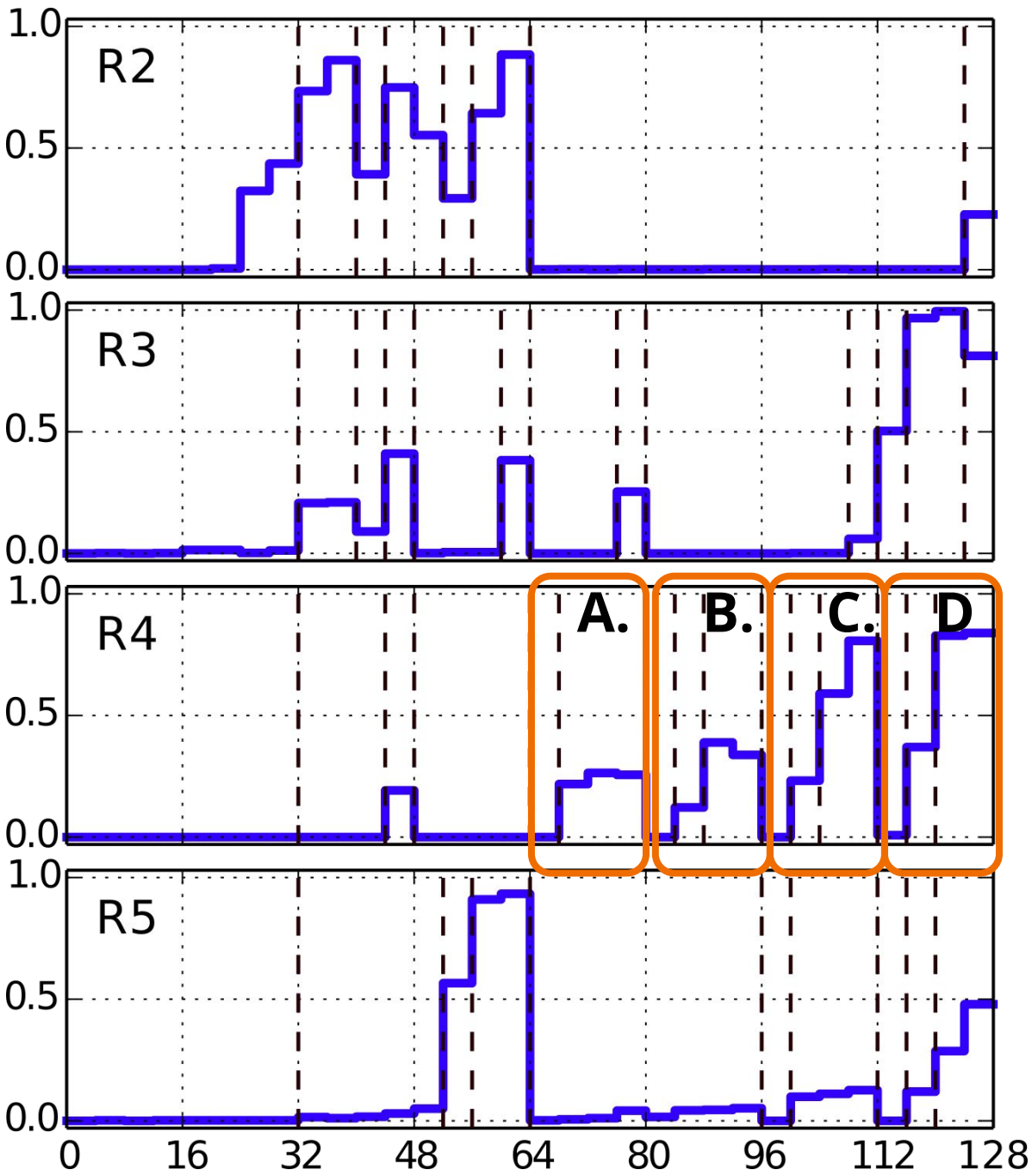
Evaluation: R1 (routers, global Internet carrier)




A
B C D E F G H
I
J
1.0
0.8
0.6
0.4
0.2
0.0
Normalized value
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Entropy ($\hat{H}_S$ =4.6)

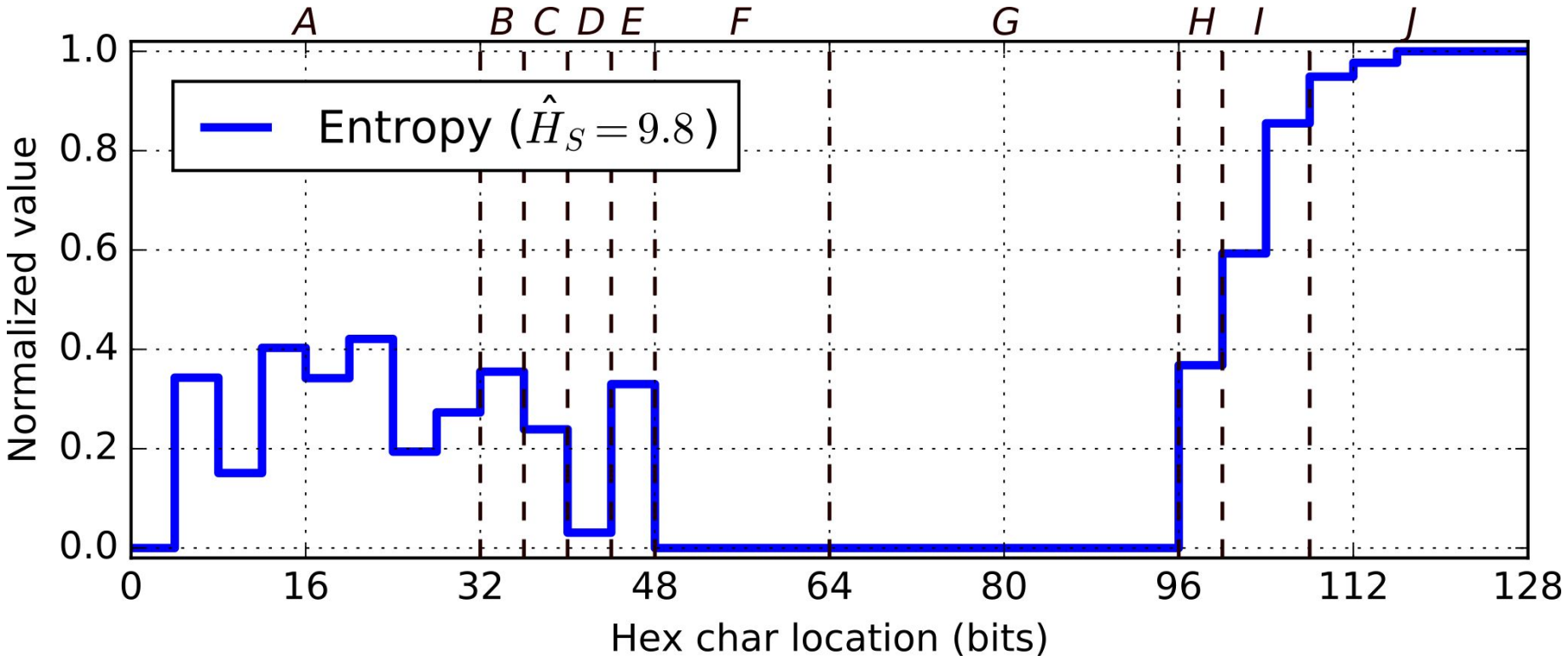

A
B
C
D
E
F
G
H
I
J
20010db8
30010db8
0
0
1
3
2
1
8
0
3
4
c
2
f
2
0
0
f
2
1
e
0
1
f
2
e
3
c
00
01
02
03
08
04
fe
ff
0000000000000000
000000000000009a
0000000000000001
0000000000000006
0000000000000102
2
1
5
e
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
<1%

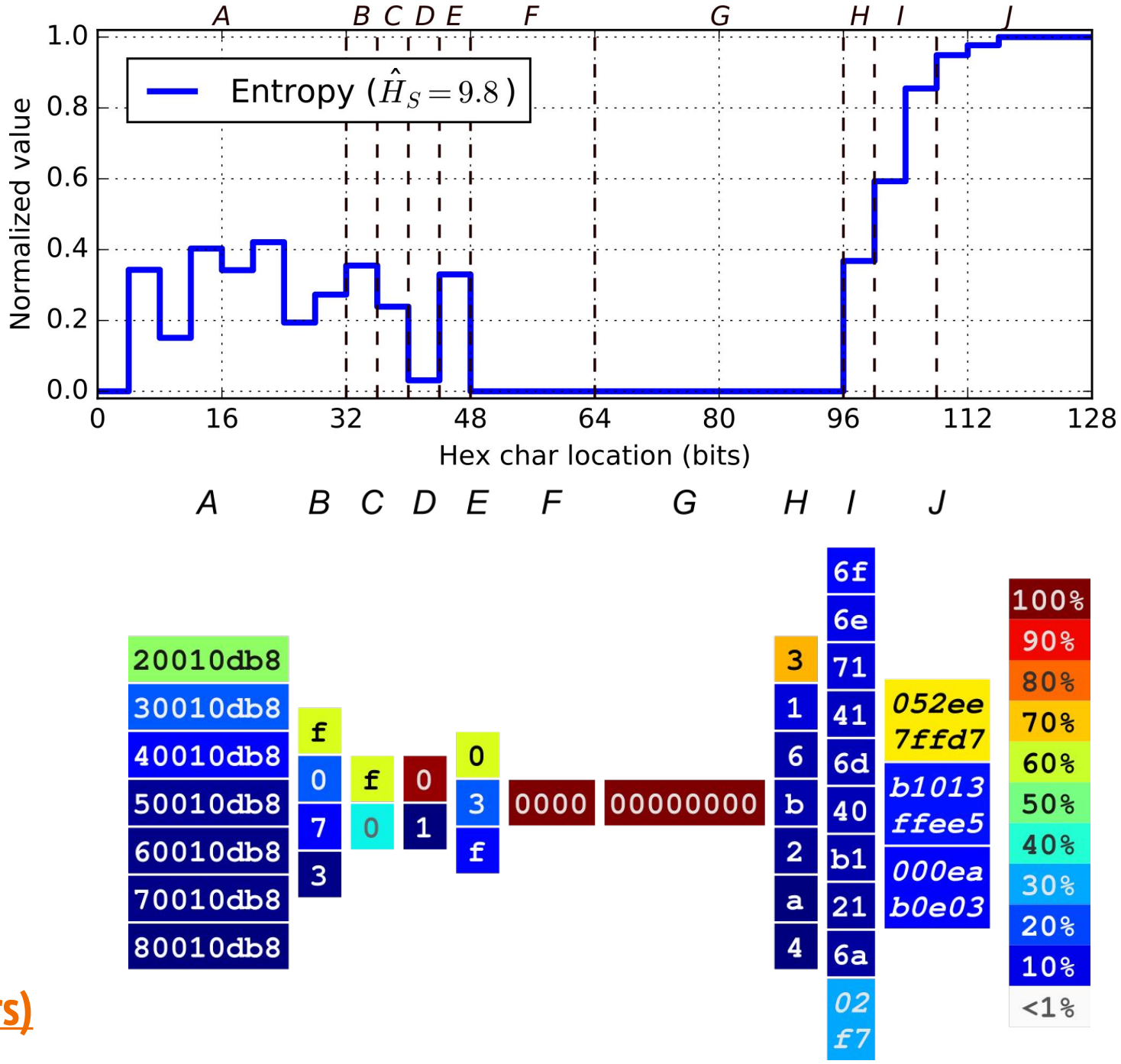
R1 (routers)



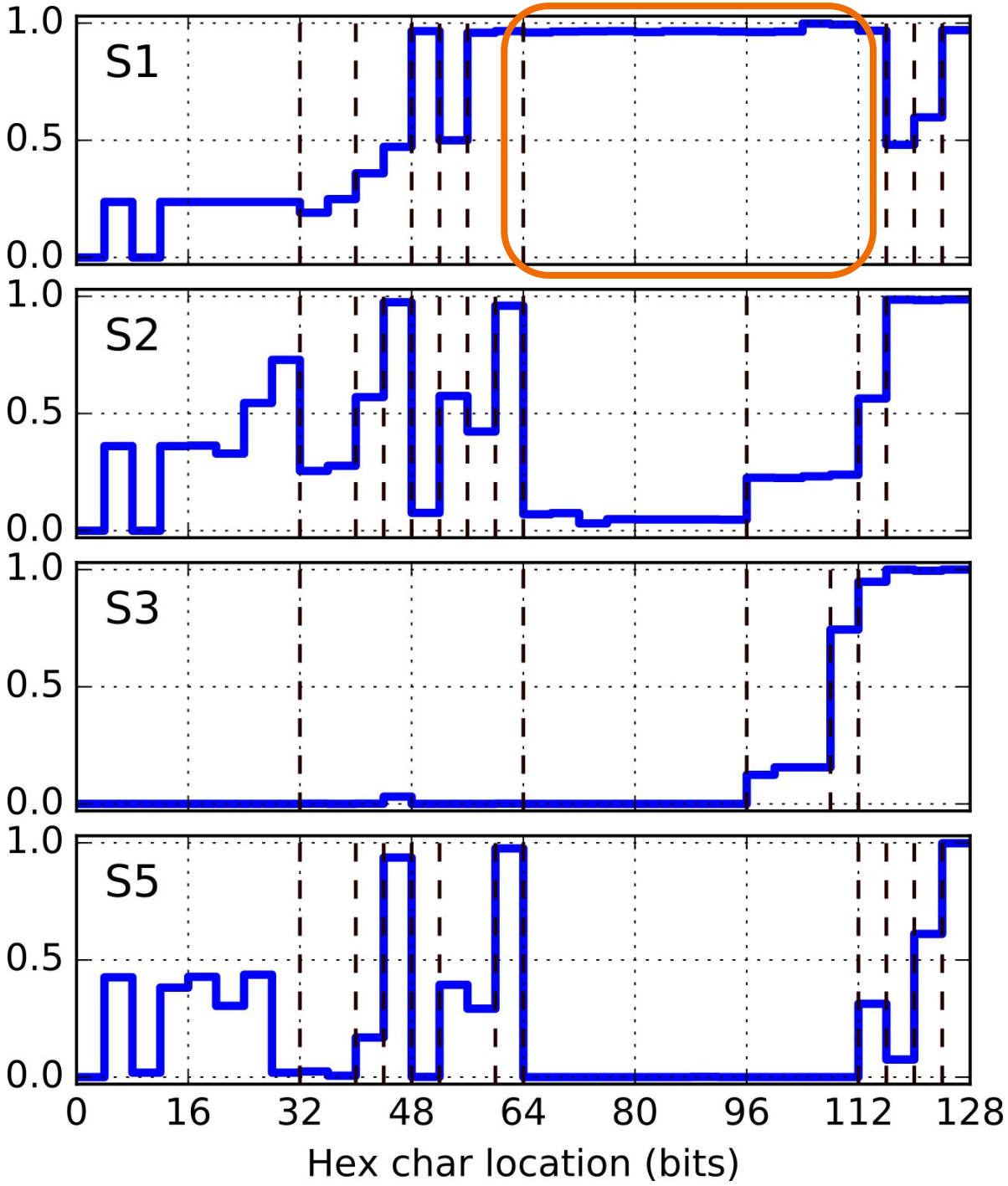
Routers (brief)

Evaluation: S4 (servers, leading cloud operator)



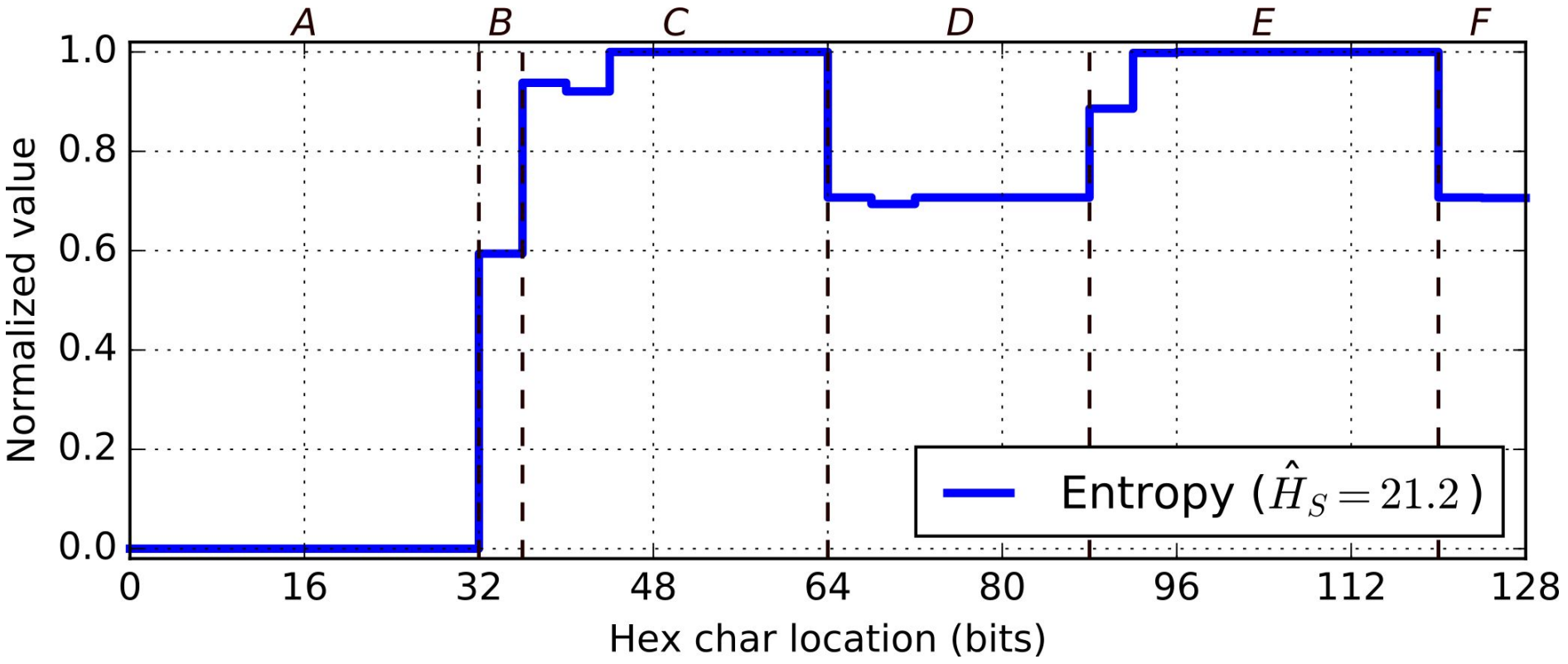

A
B C D E
F
G
H I
J
Entropy ($\hat{H}_S = 9.8$)
Normalized value
1.0
0.8
0.6
0.4
0.2
0.0
0
16
32
48
64
80
96
112
128
Hex char location (bits)
A
B C D E
F
G
H I
J
20010db8
30010db8
40010db8
50010db8
60010db8
70010db8
80010db8
f
0
7
3
f
0
0
1
0
3
f
0000
00000000
3
1
6
b
2
a
4
6f
6e
71
41
6d
40
b1
21
6a
02
f7
052ee
7ffd7
b1013
ffee5
000ea
b0e03
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
<1%

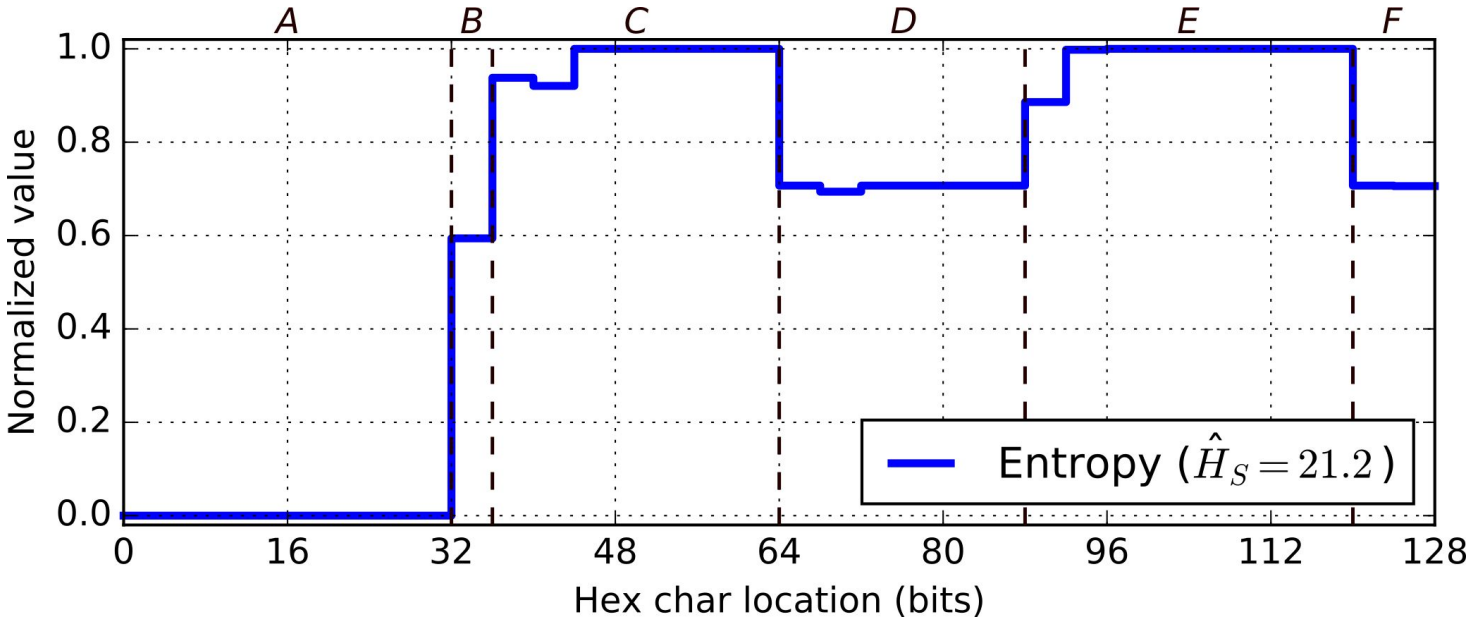
S4 (servers)

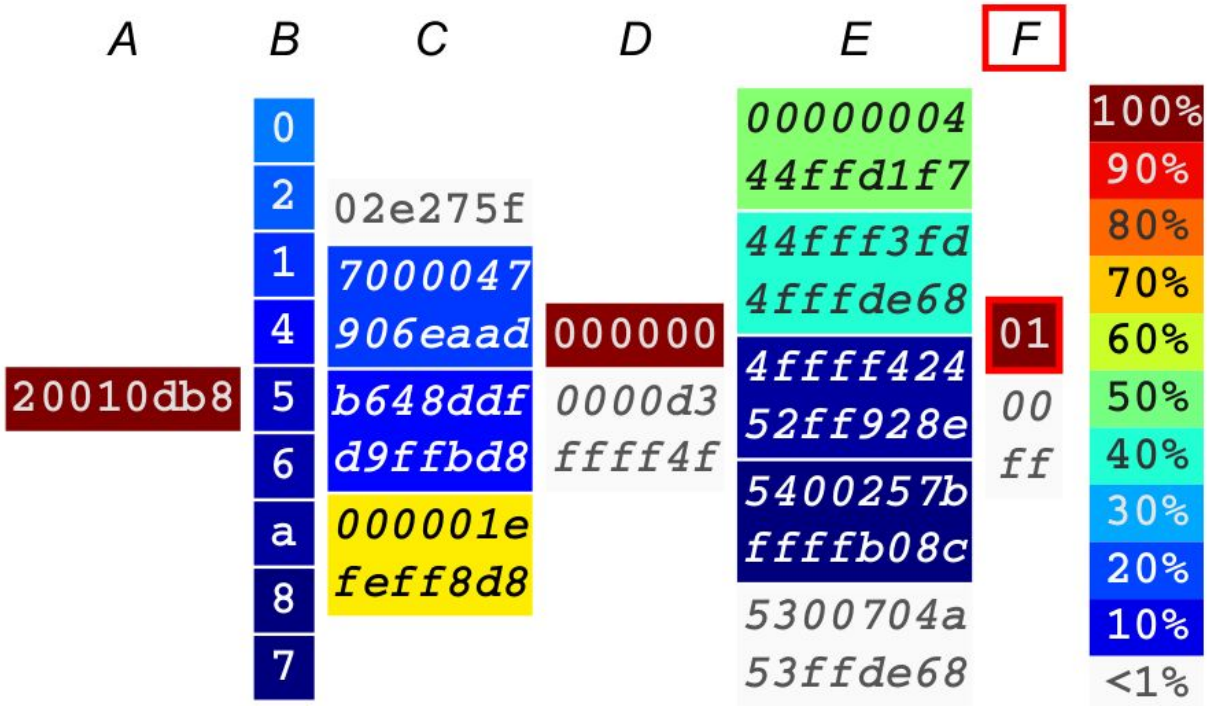


Servers (brief)

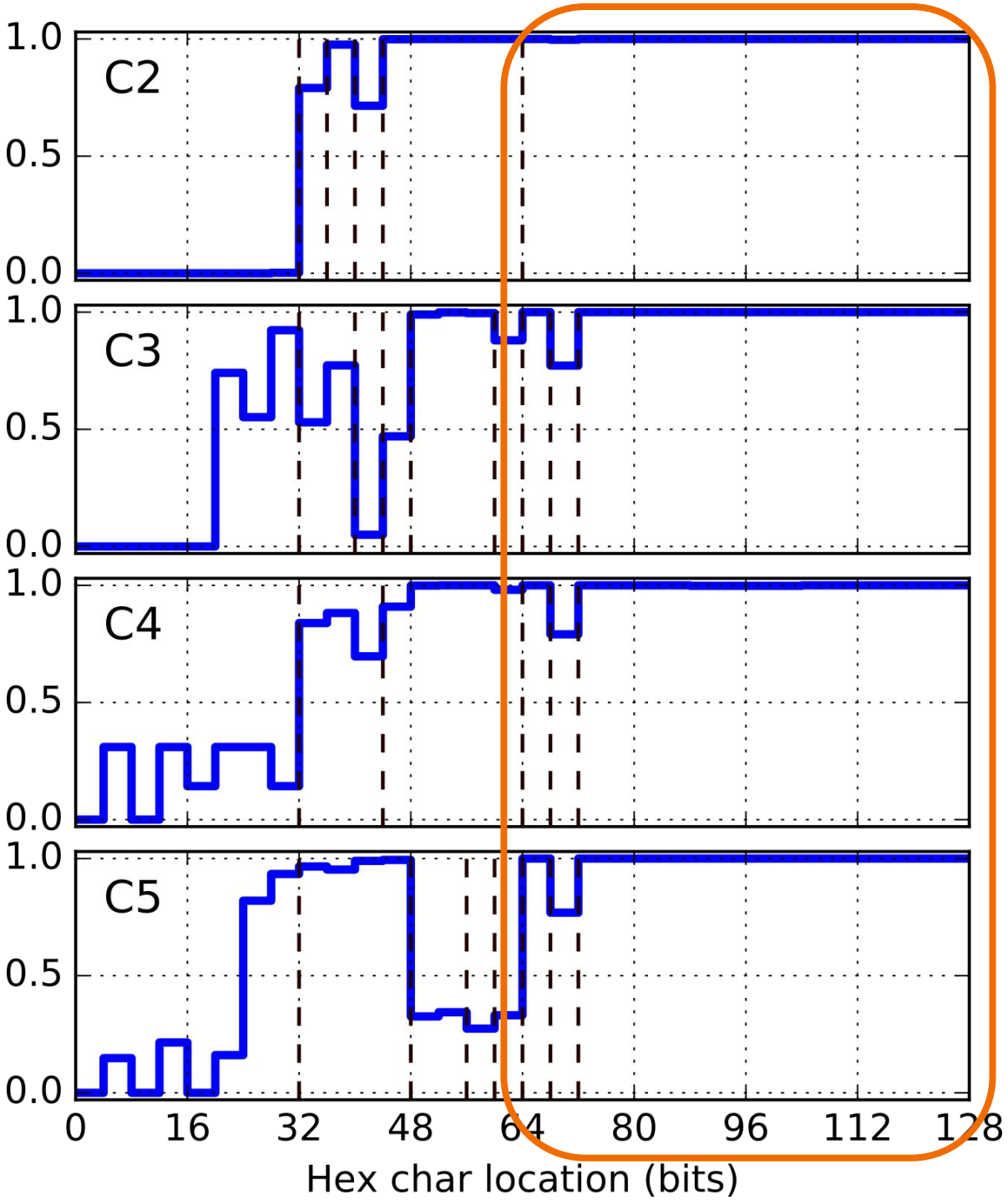
Evaluation: C1 (clients, large mobile operator)




A
B
C
D
E
F
1.0
0.8
0.6
0.4
0.2
0.0
Normalized value
0
16
32
48
64
80
96
112
128
Hex char location (bits)
Entropy ($\hat{H}_S = 21.2$)


A
B
C
D
E
F
20010db8
0
2
1
4
5
6
a
8
7
02e275f
7000047
906eaad
b648ddf
d9ffbd8
000001e
feff8d8
000000
0000d3
ffff4f
00000004
44ffd1f7
44fff3fd
4fffde68
4ffff424
52ff928e
5400257b
ffffb08c
5300704a
53ffde68
01
00
ff
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
<1%

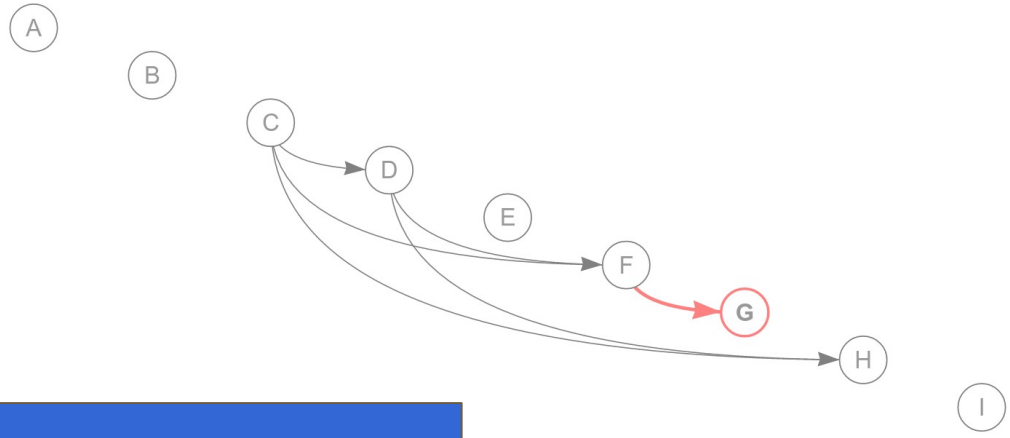
C1 (clients)



Clients (brief)

Application: generating candidate targets

(A1, B1, C1, D1, E1, F1, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C2, D2, E1, F5, G4, H2, I11)
(A1, B1, C2, D3, E1, F3, G3, H2, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F2, G3, H1, I11)
(A1, B1, C1, D1, E1, F1, G2, H1, I11)
(A1, B1, C1, D1, E1, F2, G2, H1, I11)
(A1, B1, C3, D1, E1, F4, G8, H2, I11)
(A1, B1, C1, D1, E1, F1, G1, H1, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)
(A1, B1, C1, D1, E1, F2, G1, H1, I11)
(A1, B1, C2, D4, E1, F6, G3, H2, I11)
(A1, B1, C3, D1, E1, F2, G3, H2, I11)
(A1, B1, C1, D1, E1, F1, G8, H1, I11)



F:	G: G1	G2	G3
F1	13%	10%	10%
F2	18%	20%	20%
F3	13%	7%	9%
F4	16%	9%	10%

Application: generating candidate targets

2001:0db8:0010:0004:0000:0000:0000:03cc	(A1, B1, C1, D1, E1, F1, G3, H1, I11)
2001:0db8:0010:0003:0000:0000:0000:0f97	(A1, B1, C1, D1, E1, F1, G1, H1, I11)
2001:0db8:0022:1028:9e83:1334:17c0:897a	(A1, B1, C2, D2, E1, F5, G4, H2, I11)
2001:0db8:0022:3064:69f5:02d2:f223:8635	(A1, B1, C2, D3, E1, F3, G3, H2, I11)
2001:0db8:0010:0014:0000:0000:0000:0347	(A1, B1, C1, D1, E1, F2, G3, H1, I11)
2001:0db8:0010:0014:0000:0000:0000:022a	(A1, B1, C1, D1, E1, F2, G3, H1, I11)
2001:0db8:0010:0005:0000:0000:0000:03ca	(A1, B1, C1, D1, E1, F1, G2, H1, I11)
2001:0db8:0010:0015:0000:0000:[illegible]	[illegible] D1, E1, F2, G2, H1, I11)
2001:0db8:0021:0056:8032:6eb3:[illegible]	[illegible] D1, E1, F4, G8, H2, I11)
2001:0db8:0010:0003:0000:0000:0000:018b	(A1, B1, C1, D1, E1, F1, G1, H1, I11)
2001:0db8:0010:0002:0000:0000:0000:0424	(A1, B1, C1, D1, E1, F1, G8, H1, I11)
2001:0db8:0010:0013:0000:0000:0000:0e2f	(A1, B1, C1, D1, E1, F2, G1, H1, I11)
2001:0db8:0022:20a4:3eb9:5fca:3ccb:2aae	(A1, B1, C2, D4, E1, F6, G3, H2, I11)
2001:0db8:0021:0014:3326:6434:74c9:aad6	(A1, B1, C3, D1, E1, F2, G3, H2, I11)
2001:0db8:0010:000f:0000:0000:0000:07bd	(A1, B1, C1, D1, E1, F1, G8, H1, I11)
(...)	

Scanning: experiment

1. Train on **1K samples**
2. Evaluate **on 1M generated candidates**
3. Check number of **valid** addresses in:
    - Testing set
    - Ping requests
    - Reverse DNS

Scanning: Servers and Routers (1K sample)

Dataset	Found IPv6 addresses				Success rate	New /64s
	Test set	Ping	rDNS	Overall		
S1	0	0	0	0	0.0%	0
S2	6.4 K	160 K	29	160 K	16%	12 K
S3	62 K	430 K	0	430 K	43%	0
S4	480	23 K	0	23 K	2.3%	0
S5	44 K	53 K	18 K	66 K	6.6%	1.2 K
R1	20 K	33 K	37 K	44 K	4.4%	24 K
R2	14 K	9.7 K	22	19 K	1.9%	8.2 K
R3	11 K	10 K	16 K	19 K	1.9%	70
R4	1.6 K	400	1.7 K	1.7 K	1.7%	0
R5	4.3 K	3.3 K	2.3 K	5.5 K	0.55%	380
	160 K	720 K	75 K	770 K		46 K

Discovering structure even in client networks

Dataset	Predicted /64s		Success rate (7-day)
	Mar 17	Mar 17-23	
C1	12 K	54 K	5.4%
C2	2.0 K	11 K	1.1%
C3	7.5 K	8.3 K	0.83%
C4	37 K	120 K	12%
C5	150 K	200 K	20%
	210 K	390 K	

Conclusions

- **IPv6 networks *are* scannable:**
  - For most Server & Router networks we tried
  - For Clients, network IDs are predictable
  - But… only to some degree (% success rate)
- **IPv6 addresses *are* structured**
  - Can build probabilistic models for them (BNs)
  - Entropy uncovers semantically separate segments
- **Entropy/IP automatically learns these structures**
  - Interactive browser
  - Can generate targets for scanning
  - Can help in securing against scanning

Home Demo Contact

www.entropy-ip.com

Entropy/IP: Uncovering Structure in IPv6 Addresses

P. Foremski, D. Plonka, A. Berger

This website is a Live Demo of Entropy/IP, our analysis system for sets of IPv6 addresses. The system employs Shannon's Entropy, Machine Learning, and Bayesian Networks to automatically discover, describe, and visualize structures in IPv6 addresses of Servers, Routers, and Web Clients. In addition, the system uses its models to generate candidate target addresses for scanning portions of the IPv6 Internet.

See the Sample Report › | Read the Paper (PDF) › | Run it! ›

Our paper has been accepted to appear in the ACM Internet Measurement Conference 2016 (IMC 2016)

Reports for Servers

- Dataset S1 (Web Hosting)
- Dataset S2 (Content Delivery Network)
- Dataset S3 (Content Delivery Network)
- Dataset S4 (Cloud Network)
- Dataset S5 (many services)

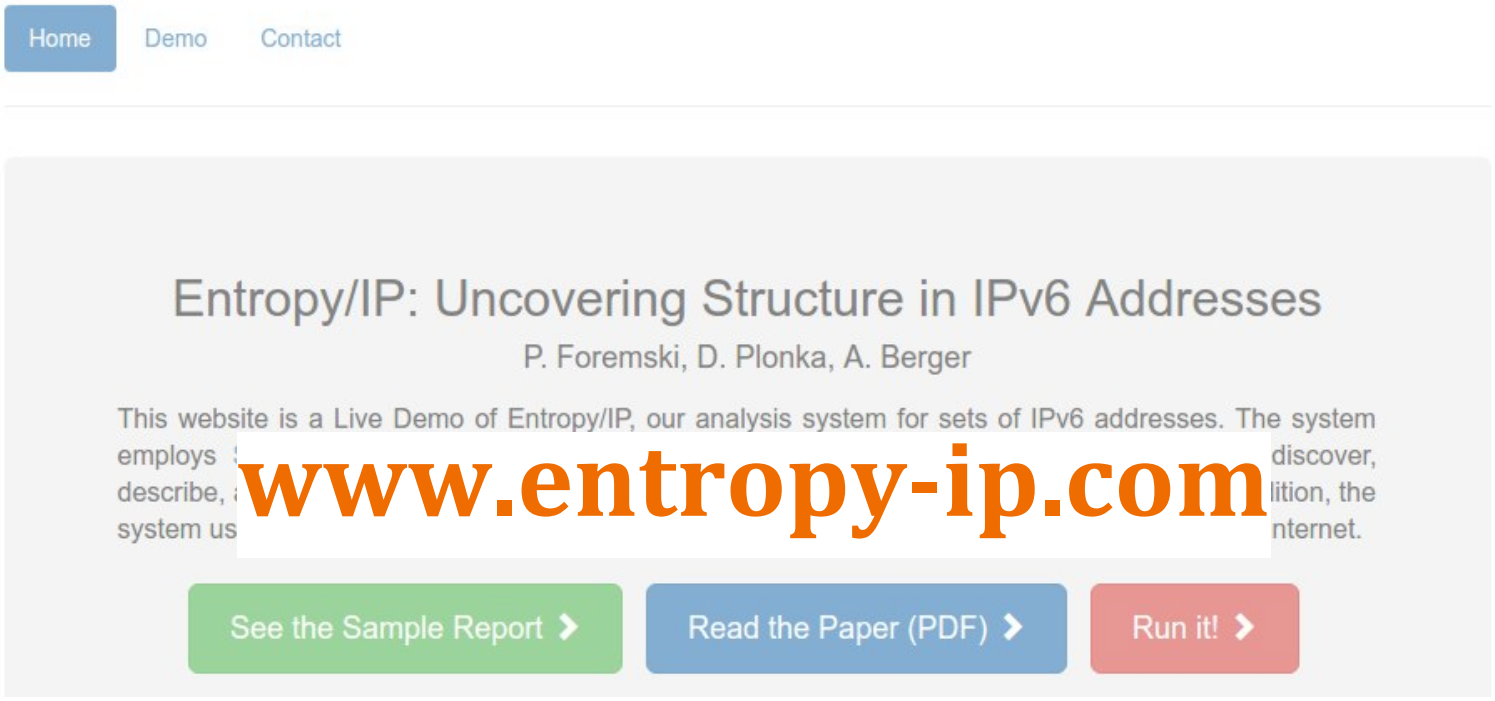
Reports for Routers

- Dataset R1
- Dataset R2
- Dataset R3
- Dataset R4
- Dataset R5

Reports for Clients

- Dataset C1 (a Mobile Operator)
- Dataset C2 (a Mobile Operator)
- Dataset C3
- Dataset C4
- Dataset C5

The above are reports for *anonymized* IPv6 addresses, collected from DNS, `traceroute`, and logs of a CDN.



Thank You!

Paweł Foremski

Institute of Theoretical and Applied Informatics

Polish Academy of Sciences

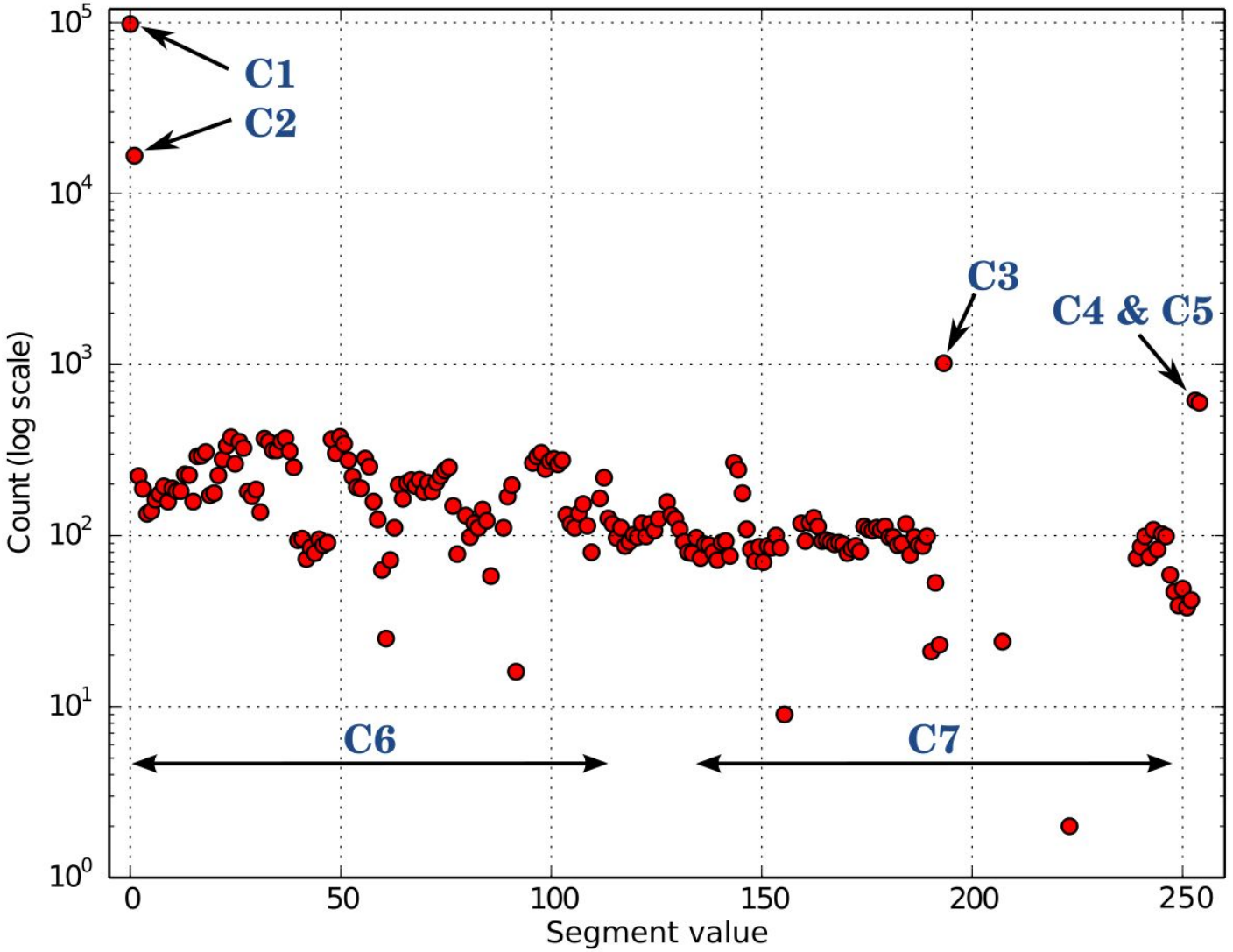
Email: pjf@iitis.pl

Twitter: @pforemski

Scanning: Sample size vs. success rate

Dataset	Training sample size				
	100	1 K	10 K	100 K	1 M
S5	5.2%	6.7%	7.0%	2.9%	-
R1	4.1%	4.4%	4.4%	-	-
C5	16%	20%	21%	20%	14%

3. Segment Mining: example



1. Entropy Analysis: variability

```
2001:0db8:0010:0013:0000:0000:0
2001:0db8:0010:0000:0000:0000:0
2001:0db8:0010:0003:0000:0000:0
2001:0db8:0020:d05f:882f:6082:f
2001:0db8:0010:0004:0000:0000:0
2001:0db8:0010:0003:0000:0000:0
2001:0db8:0010:0008:0000:0000:0
2001:0db8:0010:000a:0000:0000:0
2001:0db8:0010:0006:0000:0000:0
2001:0db8:0022:1014:aef6:60af:d
2001:0db8:0010:0012:0000:0000:0
2001:0db8:0022:10c0:5100:ac7d:9
2001:0db8:0010:0002:0000:0000:0
2001:0db8:0010:0008:0000:0000:0
2001:0db8:0022:2053:4e6a:a11a:d
(...)
```