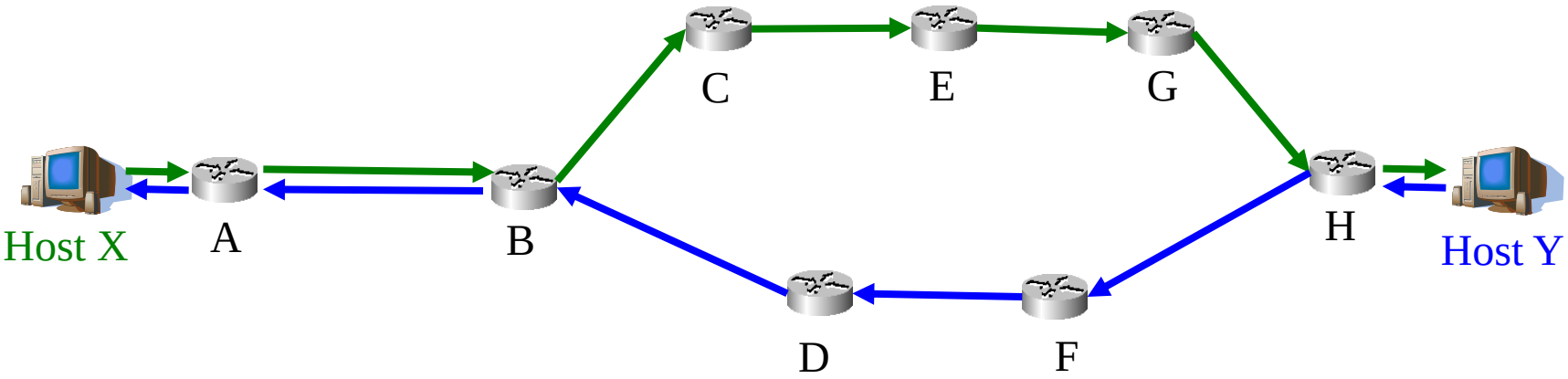
Estimating Routing Symmetry on Single Links by Passive Flow Measurements

Wolfgang John
Chalmers University of Technology, Sweden
Maurizio Dusi
Università degli Studi di Brescia, Italy
kc claffy
CAIDA, UC San Diego, USA

Estimating Routing Symmetry on Single Links by Passive Flow Measurements

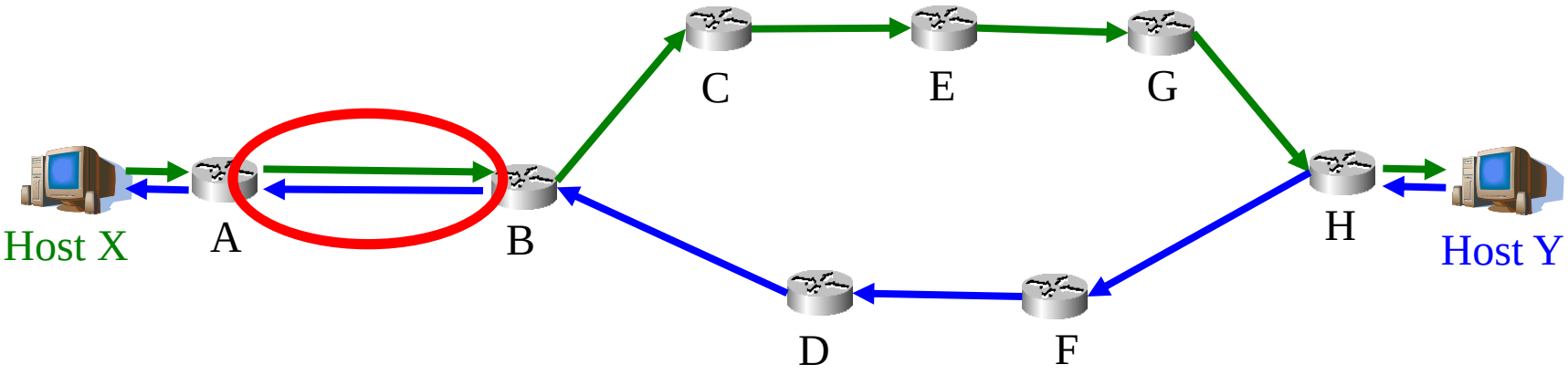


- Routing symmetry: End-to-End perspective
  - Asymmetric route between X and Y

    X->Y via 6 hops: [A,B,C,E,G,H]

    Y->X via 5 hops: [A,B,D,F,H]

Estimating Routing Symmetry on Single Links by Passive Flow Measurements



- Routing symmetry: Single link perspective
  - Asymmetry on most links e.g. B->C, H->F…
  - BUT symmetry on link A<->B

Estimating Routing Symmetry on Single Links by Passive Flow Measurements

- Passively collected flows:
  sequences of packets between endpoints
- We use CoralFlow on packet-level traces:
  - unidirectional flows
  - defined by 5-tuple {sIP,sPort,dIP,dPort,proto}
  - separated by timeout intervals

  - Symmetric flow: forward AND backward packets seen
  - Asymmetric flow: forward OR backward packets seen

Estimating Routing Symmetry on Single Links by Passive Flow Measurements

- Traffic analysis and classification deals with flows collected at specific links
  - Symmetric traffic on a link = bidirectional flows
    - allows easier inference of sessions
    - offers additional feature-space for traffic classification (better classification results[1])

[1] Crotti et al. in GLOBECOM 2009

Motivation

? Fraction of flows on a link observed symmetric?

! Well, this sounds easy…

… but depends on traffic composition of the link:

- Inherently asymmetric flows (UDP, ICMP)
  - UDP is dominated by single-packet flows[2]
- TCP background radiation
  - Scanning traffic can dominate in terms of flow numbers[3]

[2] John et al. in PAM 2008
[3] Allman et al. in IMC 2007

Contribution

? So, how many flows are symmetric?

! We present …

- a method to assess flow-level symmetry, the Flow-level Symmetry Estimator (FSE)
- a FSE tool based on CoralFlow
- results of FSE on a diverse dataset
  - Impact of
    - inherently asymmetric traffic
    - observation interval
    - traffic granularity (only in paper)

Datasets

	Date	Length	#Flows	Location
U-Brescia (Validation)	2009-12	2x10min	60 K	Access Link (100Mbps) Univ. Brescia, Italy
GigaSUNET	2006-04	6x10min	9 M	Tier 2 (10Gbps) Sweden
	2006-11		16 M	
OptoSUNET	2009-01		57 M	Tier 2-Tier 1 (10Gbps) Sweden
	2009-02		62 M	
Eq-Chicago	2008-04	1x60min	119 M	Tier 1 (10Gbps) Illinois-Washington
	2008-05		134 M	
Eq-SanJose	2008-07		145 M	Tier 1 (10Gbps) California
	2008-08		139 M	

Method: Flow-based Symmetry Estimation (FSE)

```
1: given a time interval of traffic trace
2:       consider TCP data traffic only
3:       T_f (T_b) = set of tuples going forward (backward)
4:       T_f ∧ T_b = set of symmetric tuples T_s
5:       packets (bytes) in T_s = set of symmetric packets (bytes)
```

TCP data traffic =

TCP traffic without signaling flags, but with ACK

- Disregarding non-TCP traffic
  - TCP is guaranteed bidirectional (=symmetric)
  - TCP is responsible for majority of traffic volume
- Filtering out TCP signaling traffic (SYN/FIN/RST)
  - TCP Background radiation is mainly SYN probing
  - Established, valid connections have data or ACK packets in both directions

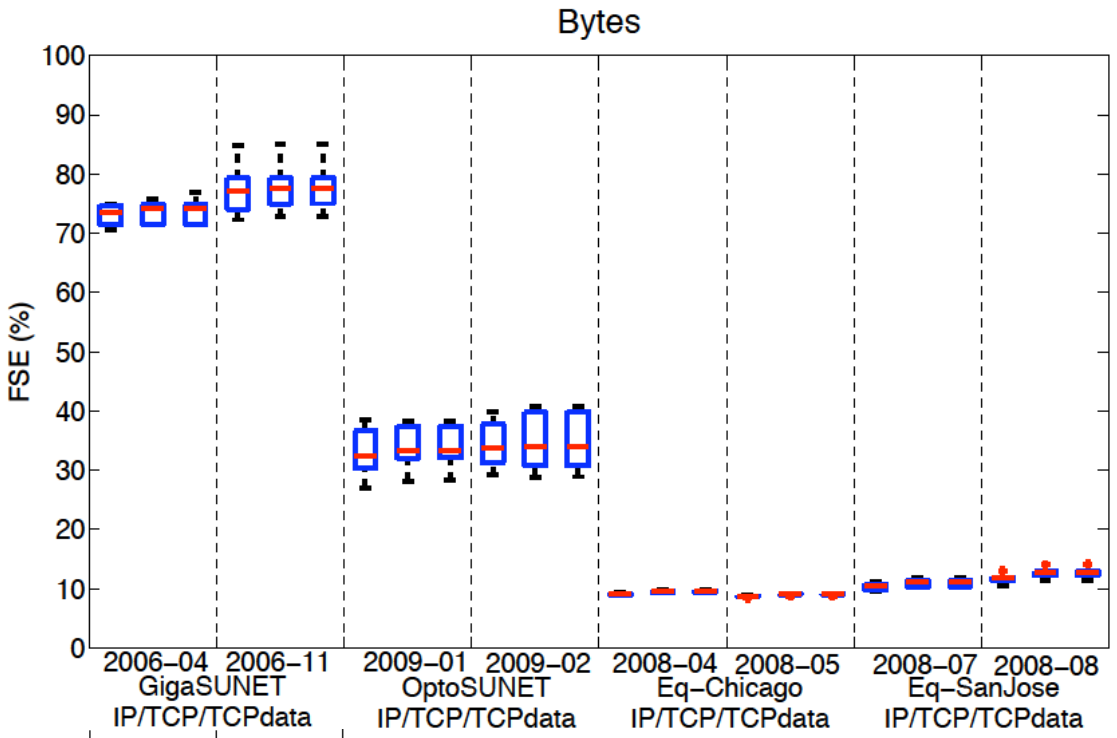
FSE: TCP Background Radiation

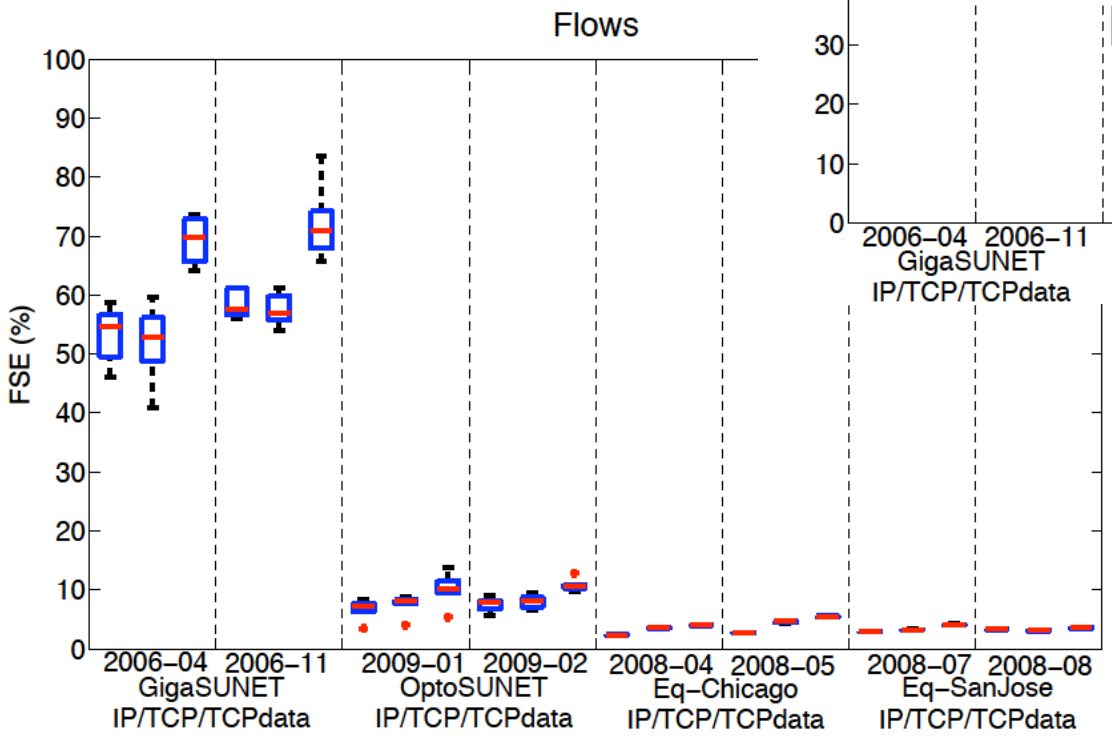
- TCP Traffic removed by FSE method
  = estimate of background radiation

TCP signaling traffic		% flows	% packets	% bytes
GigaSUNET	2006-04	32.36	4.85	0.15
	2006-11	27.86	1.95	0.15
OptoSUNET	2009-01	34.81	2.05	0.08
	2009-02	34.74	2.05	0.09
Eq-Chicago	2008-04	19.19	5.60	0.51
	2008-05	23.62	4.31	0.34
Eq-SanJose	2008-07	25.27	8.04	0.83
	2008-08	19.41	7.75	0.78

Results: Impact of Inherently Asymmetric Traffic

Symmetry of bytes increases only slightly

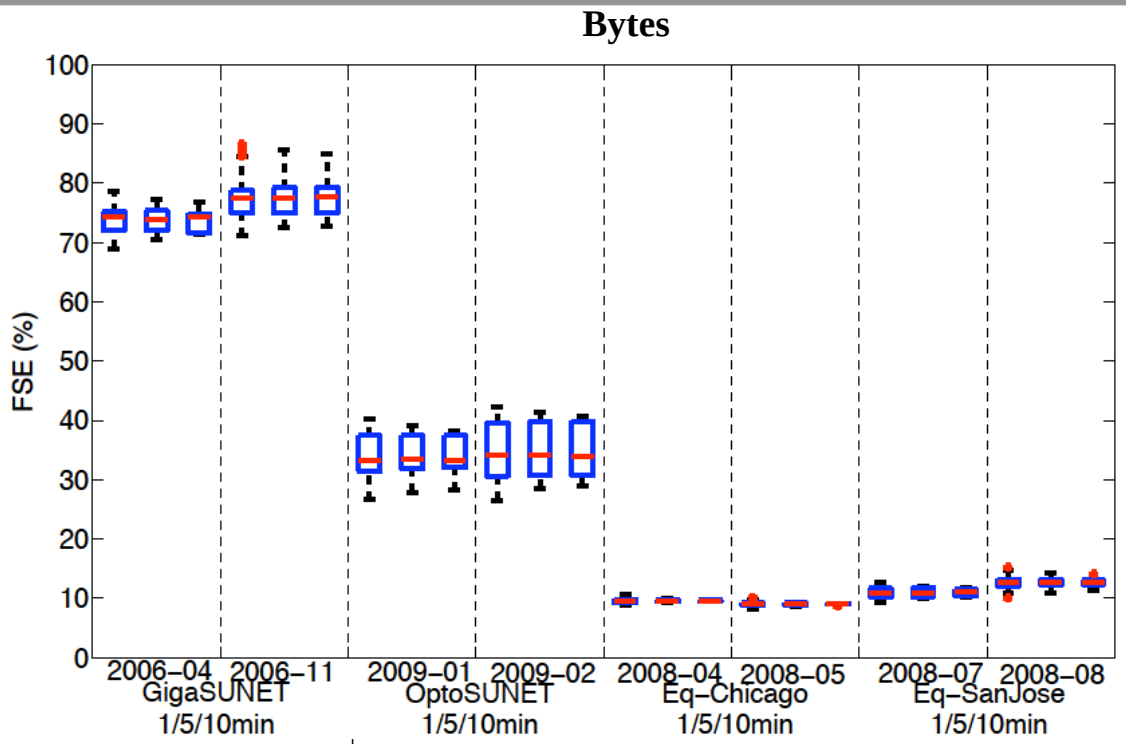


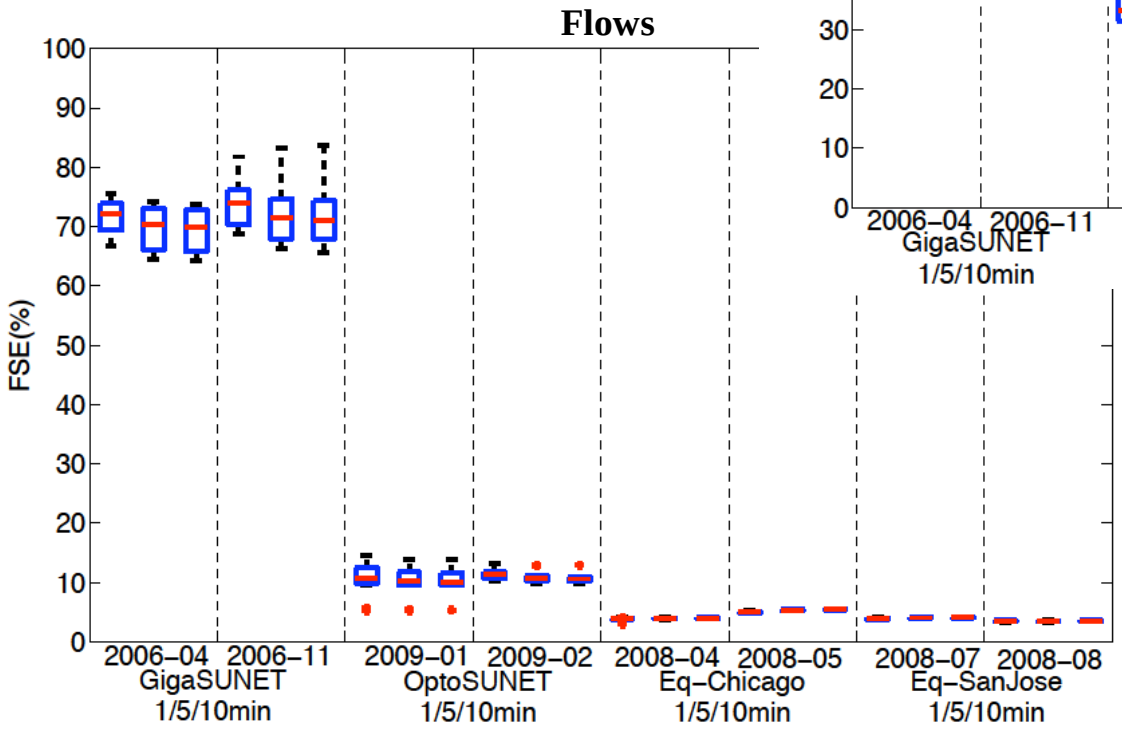


FSE filter increases symmetry of flow-tuples significantly

Results: Impact of Observation Interval

Little effect on symmetry estimates



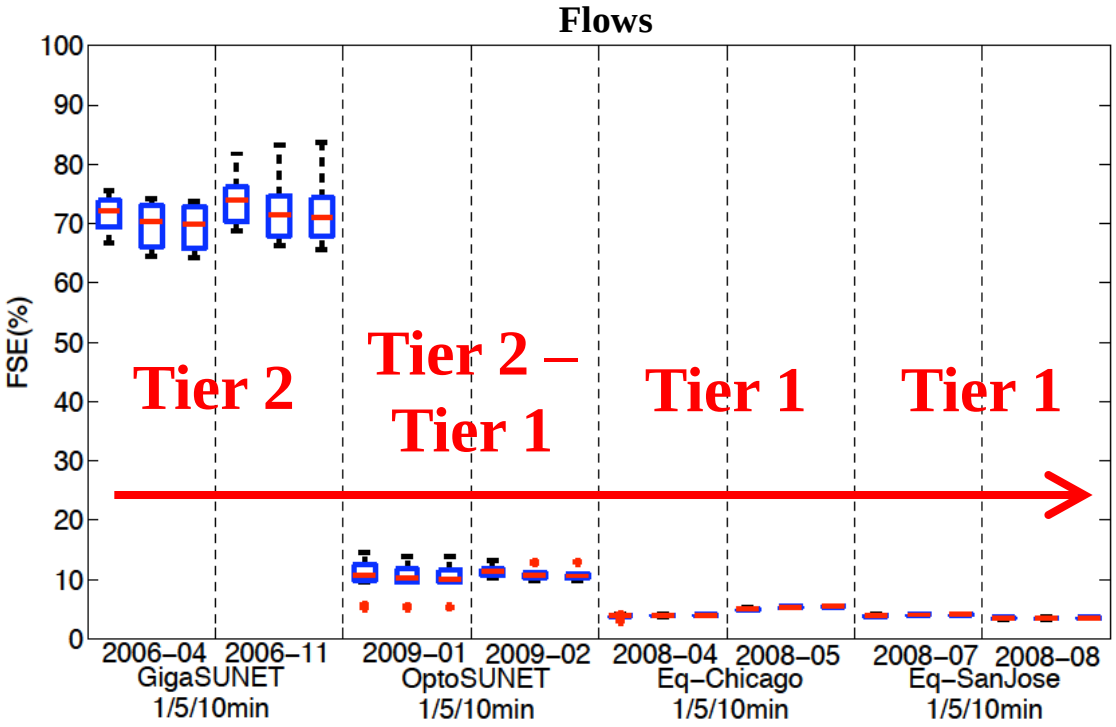


Results: Impact of Observation Interval

Bytes

2 further observations:

- Symmetry estimates are stable (within minutes and months)
- Symmetry estimates decrease radically closer to the core



Results: Summary of FSE

	Date	Non-TCP Flows	TCP Radiation	Location	Symm. All IP	Symm. TCP data
U-Brescia (Validation)	09-12	57 %	16 %	Access Link	79.0 %	**98.6 %**
GigaSUNET	06-04	58 %	32 %	Tier 2	54.6 %	**70.0 %**
	06-11		27 %		57.7 %	**71.0 %**
OptoSUNET	09-01	80 %	35 %	Tier 2-Tier 1	7.1 %	**10.2 %**
	09-02		35 %		7.8 %	**11.7 %**
Eq-Chicago	08-04	50 %	19 %	Tier 1	2.4 %	**4.0 %**
	08-05		24 %		2.7 %	**5.5 %**
Eq-SanJose	08-07	50 %	25 %	Tier 1	2.9 %	**4.1 %**
	08-08		19 %		3.3 %	**3.6 %**

Results: Summary of FSE

	Date	Non-TCP Flows	TCP Radiation	Location	Symm. All IP	Symm. TCP data
U-Brescia (Validation)	09-12	57 %	16 %	Access Link	79.0 %	**98.6 %**
GigaSUNET	06-04	58 %	32 %	Tier 2	54.6 %	**70.0 %**
	06-11		27 %		57.7 %	**71.0 %**
OptoSUNET	09-01	80 %	35 %	Tier 2-Tier 1	7.1 %	**10.2 %**
	09-02		35 %		7.8 %	**11.7 %**
Eq-Chicago	08-04	50 %	19 %	Tier 1	2.4 %	**4.0 %**
	08-05		24 %		2.7 %	**5.5 %**
Eq-SanJose	08-07	50 %	25 %	Tier 1	2.9 %	**4.1 %**
	08-08		19 %		3.3 %	**3.6 %**

Results: Summary of FSE

	Date	Non-TCP Flows	TCP Radiation	Location	Symm. All IP	Symm. TCP data
U-Brescia (Validation)	09-12	57 %	16 %	Access Link	79.0 %	**98.6 %**
GigaSUNET	06-04	58 %	32 %	Tier 2	54.6 %	**70.0 %**
	06-11		27 %		57.7 %	**71.0 %**
OptoSUNET	09-01	80 %	35 %	Tier 2-Tier 1	7.1 %	**10.2 %**
	09-02		35 %		7.8 %	**11.7 %**
Eq-Chicago	08-04	50 %	19 %	Tier 1	2.4 %	**4.0 %**
	08-05		24 %		2.7 %	**5.5 %**
Eq-SanJose	08-07	50 %	25 %	Tier 1	2.9 %	**4.1 %**
	08-08		19 %		3.3 %	**3.6 %**

Summary

Analysis often done on flows from one location

- Can we see symmetric (bi-directional) flows?

We proposed:

- A method: Flow-based Symmetry Estimation (FSE)
  - Utilizing on common tool (CoralReef)
  - Disregarding inherently asymmetric traffic
  - Normalized metric allows fair comparison
- Results: Symmetry assessments from a large, diverse dataset

Conclusions

- Traffic normalization (filtering) is important for flow-symmetry assessment
  - Exclusive access link may be considered asymmetric
- FSE is robust against different flow definitions
  - 1, 5 and 10min timeout intervals
  - Timeout-based and signaling-based flows (see paper)
- Traffic granularity has an effect (see paper)
  - IP Pair flows result in higher FSE's than 5-tuple flows
- Symmetry is quite stable (minutes to months)
- Symmetry radically decreases with link 'coreness'

Analysis methods should NOT assume symmetric flows
(except on exclusive stub access links)

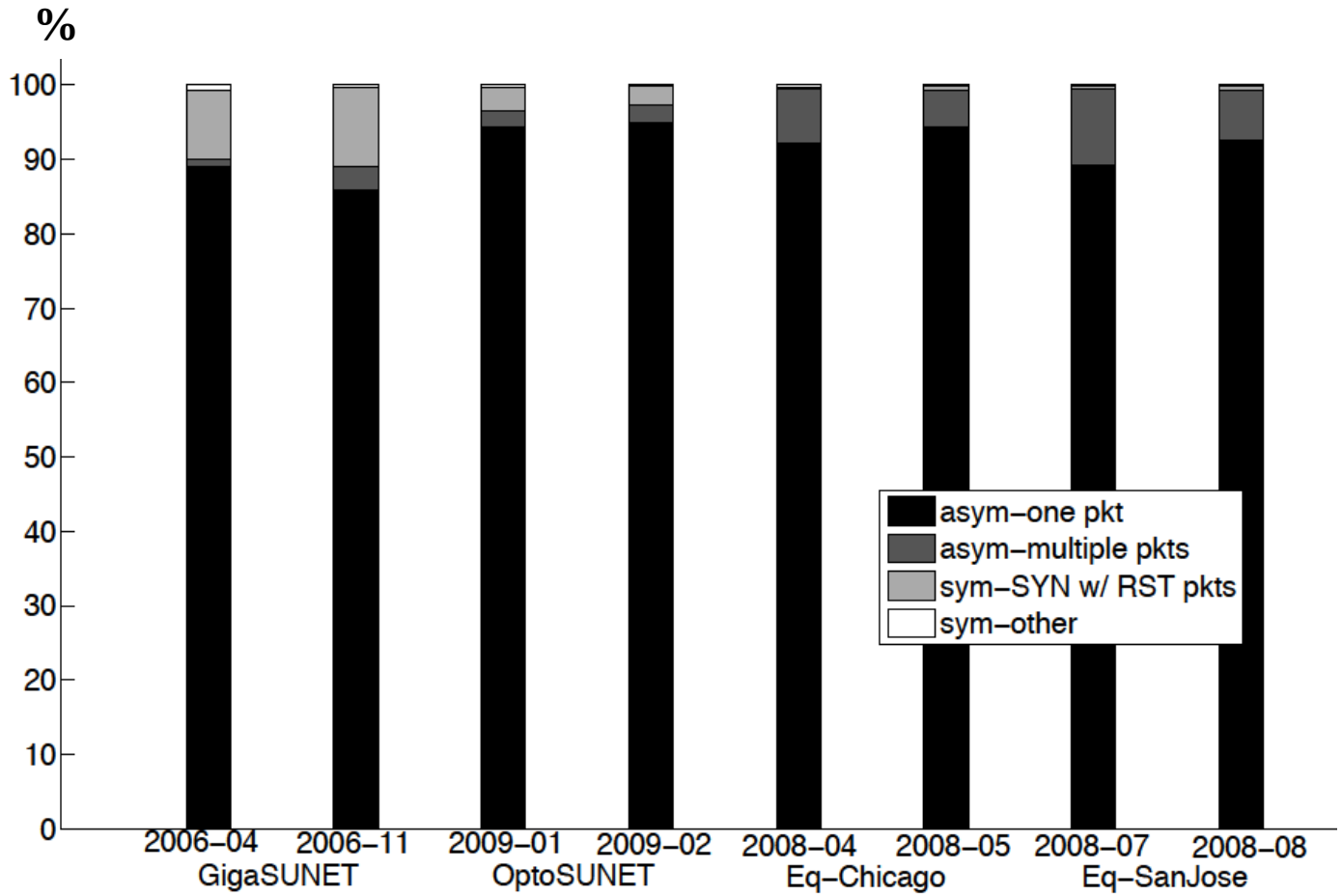
FSE (Flow-based Symmetry Estimation) Tool:
http://www.cse.chalmers.se/~johnwolf/FSE

Contact:
http://www.chalmers.se/cse/EN/people/john-wolfgang
johnwolf@chalmers.se

Backup Slides

FSE: Characteristics of TCP Background Radiation

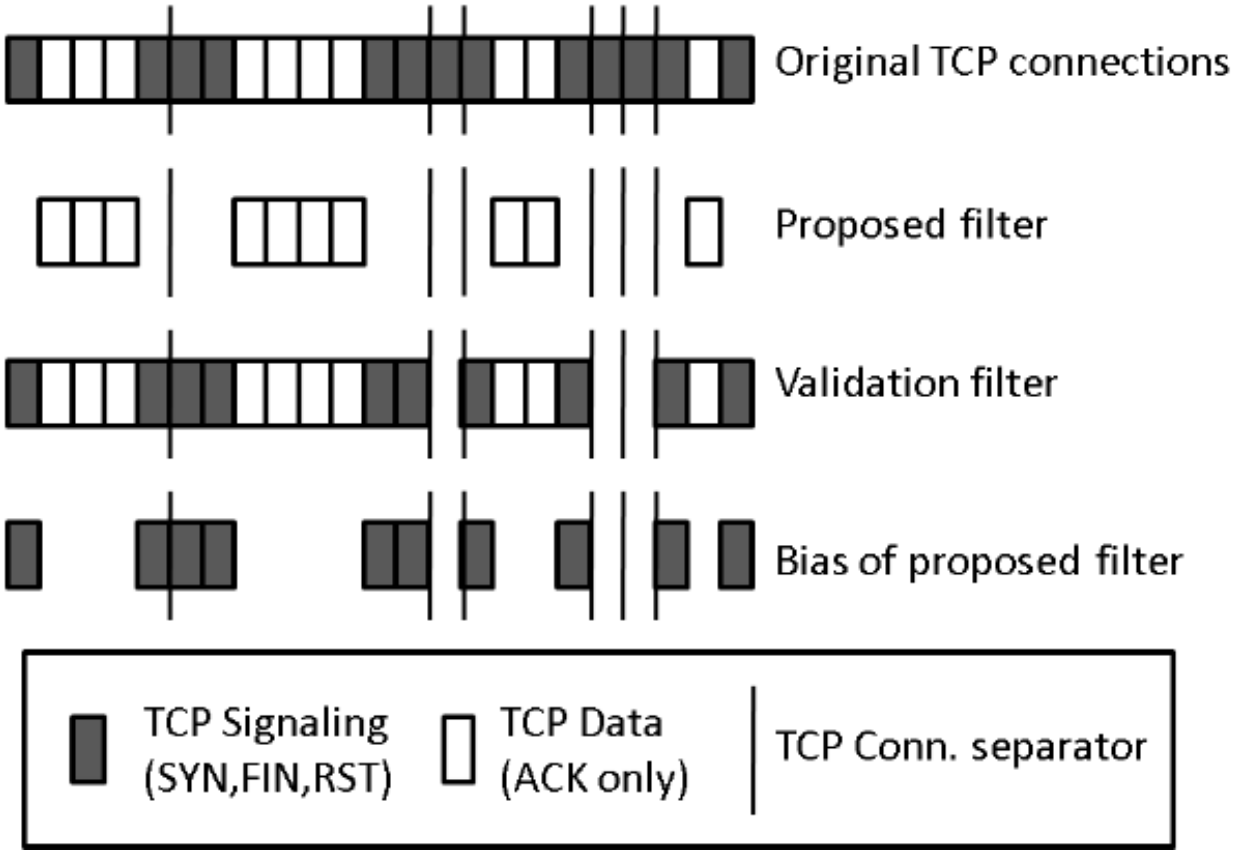
- Background radiation indeed mainly asymmetric



(Results from Validation method, details in paper)

FSE: Filter Bias

- Bias in terms of packet/bytes numbers



FSE: Bias on Estimates

- Quantified in numbers

Eq-Chicago 2008-05

Packets

	Sym.	Tot.	Sym.%
V	47.2M	469.8M	10.05%
F	45.7M	455.5M	10.04%
Diff.	1.5M	14.3M	

Bytes

	Sym.	Tot.	Sym.%
V	39.4G	433.6G	9.09%
F	39.3G	432.5	9.09%
Diff.	0.1GB	1.1G	

Results: Impact of Traffic Granularity

- IP pairs vs flows