Delay Tolerant Bulk Data Transfers on the Internet

by N. Laoutaris et al., SIGMETRICS'09

Ilias Giechaskiel

Cambridge University, R212
ig305@cam.ac.uk

March 4, 2014

Conclusions

Takeaway Messages

- Need to transfer multiple terabytes daily
  - Postal system for infrequent transfers
  - Direct transfer for small timezone differences
  - Store and forward otherwise
- Take advantage of off-peak bandwidth through "water-filling"
- Mathematical analysis with cost estimates and deadlines
- No concrete implementation!

Context

Motivation

- 1PB of data every day at CERN
  - 10GB of data transfers every second at peak!
  - `http://home.web.cern.ch/about/computing`
- Data tolerates delays from several hours to a few days
- Postal system and dedicated networks too expensive

95-percentile Pricing

- Allow 5% of traffic to be burst traffic beyond committed rate
- Charges based on peak rate!
- Lots of bandwidth wasted

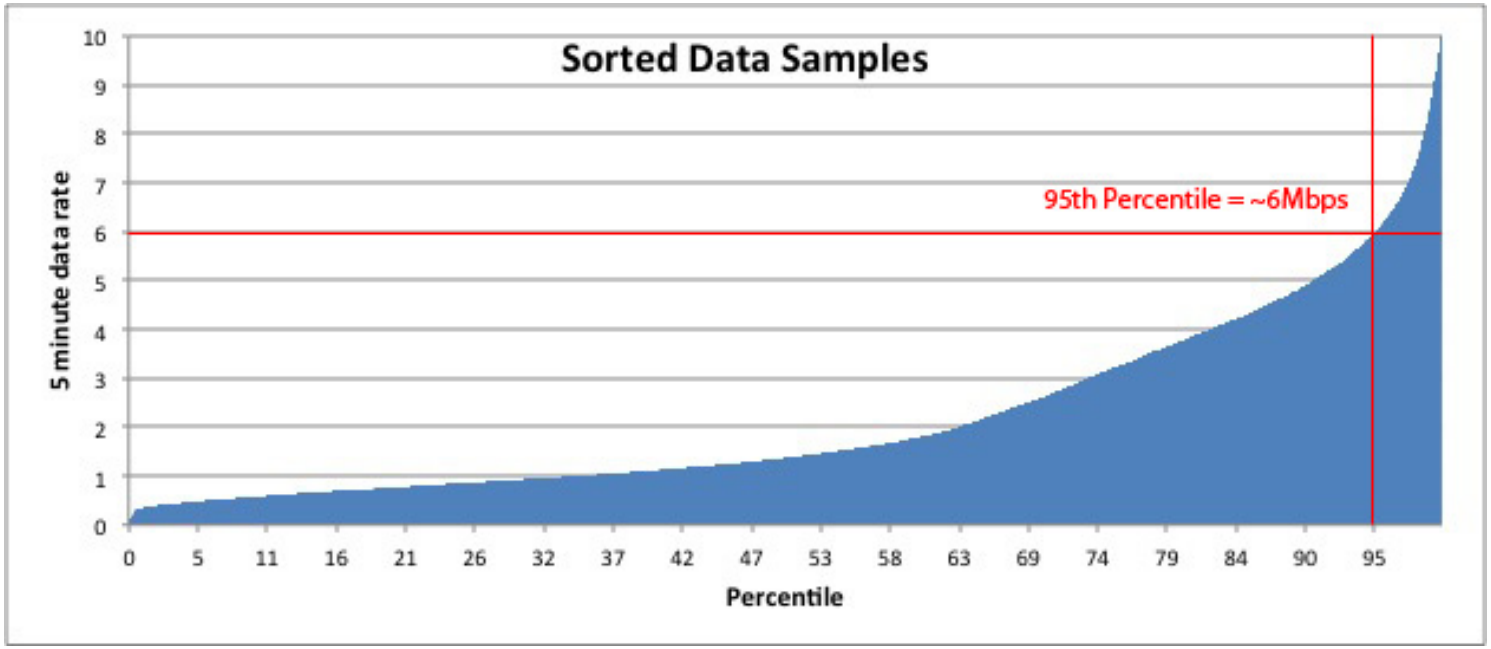


Figure: http://www.semaphore.com/blog/2011/04/04/95th-percentile-bandwidth-metering-explained-and-analyzed

The problem

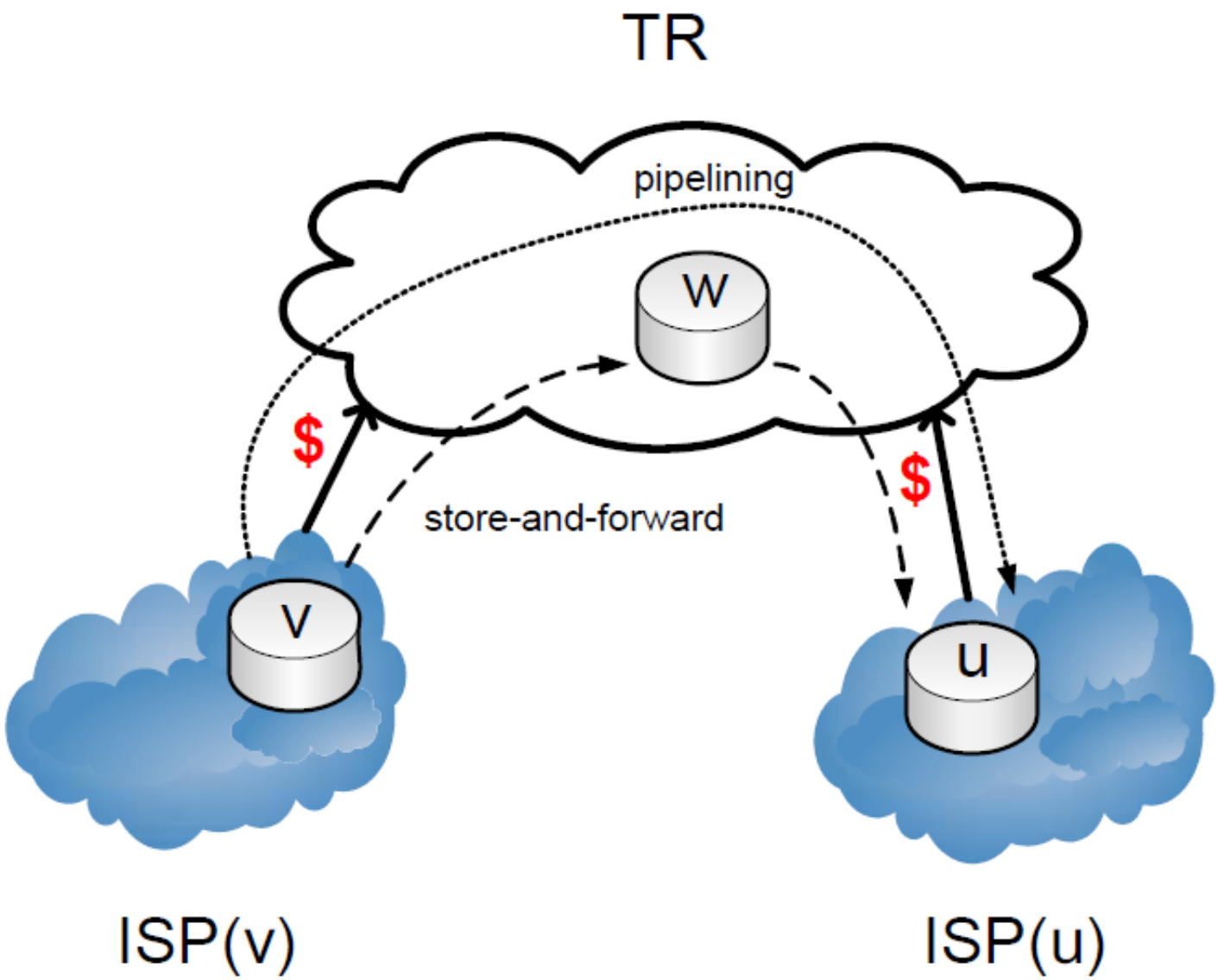
Goals

- Transfer data between data centers without dedicated network
- Avoid increasing 95-percentile cost for sender and receiver
- Avoid impact on QoS of interactive traffic

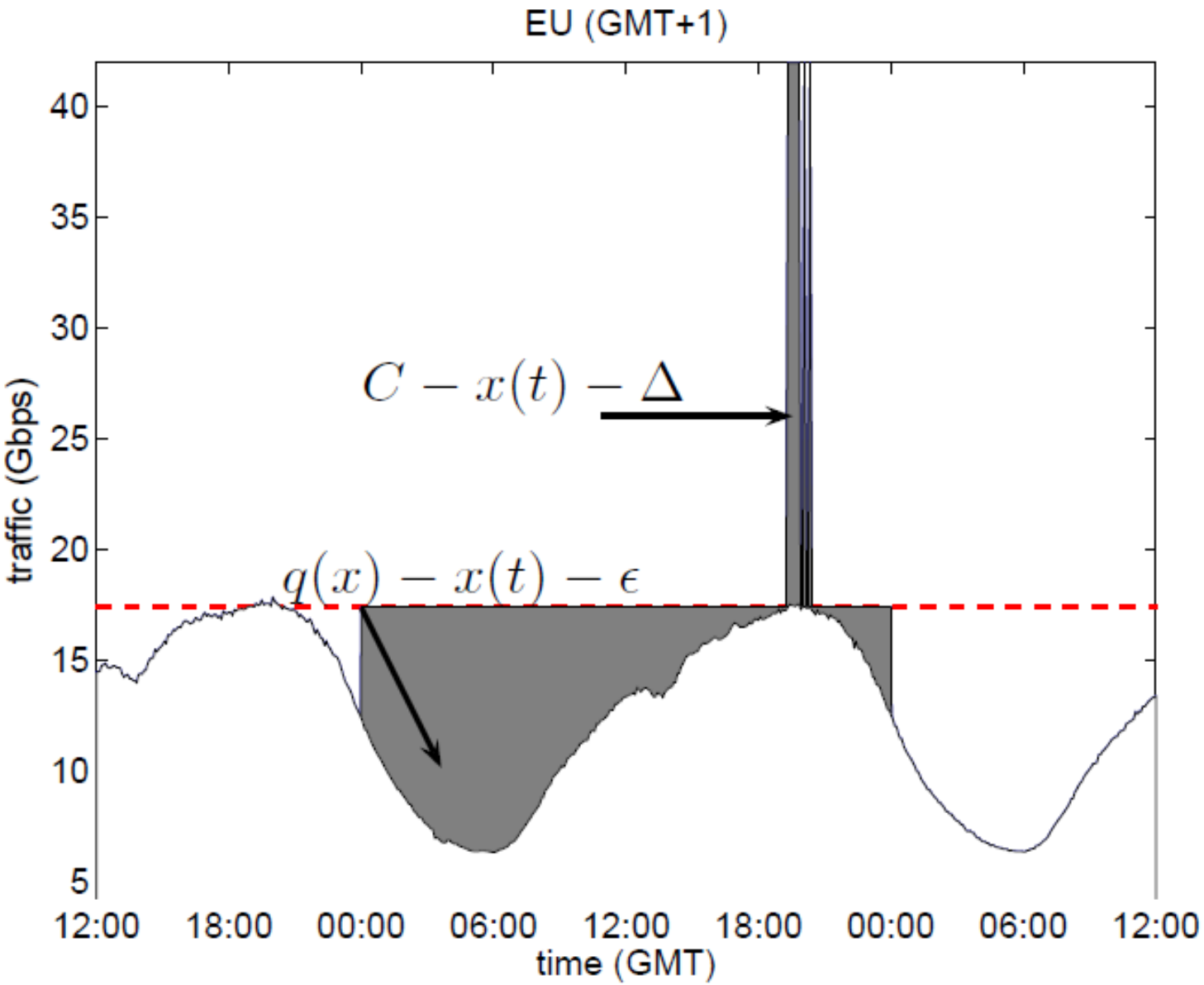
Approach

- Transmit during off-peak hours *of both sender and receiver*
  - Directly when centers close-by (*E2E*)
  - With intermediate storage otherwise (*SnF*)
- Evaluate using bandwidth costs and estimates

Network Model



Water-Filling



Bulk Transfer Policies

End-to-End with Source Scheduling

- Water-filling that respects sender and receiver charge volumes
- If enough to send for free, just use it!

Store-and-Forward

- Two independent water-fillings
- Send minimum of two and store to or transfer from transit

Implementation

Required Predictions

- Next slot load
  - Successive loads highly correlated
- Total charged volume
  - Use current so far or previous month

Meeting Deadlines

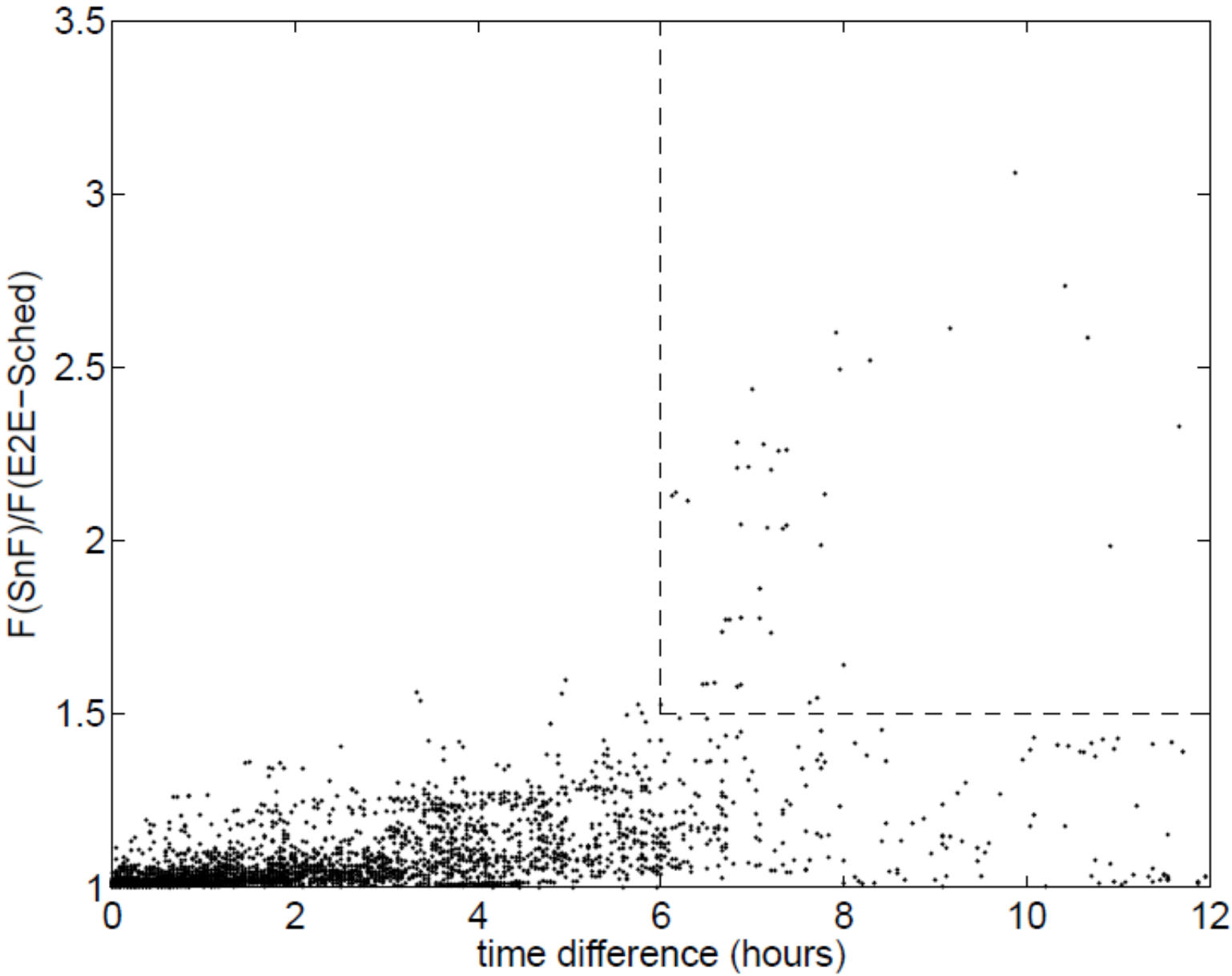
- Not all volumes can be sent for free!
- Use existing approach, but modify cost volumes allowed
  - Polynomial exact search or greedy approximation for min cost
- Need prediction for entire month
  - Use same day of previous week
  - 1-2% worse than actually knowing future values

Evaluation

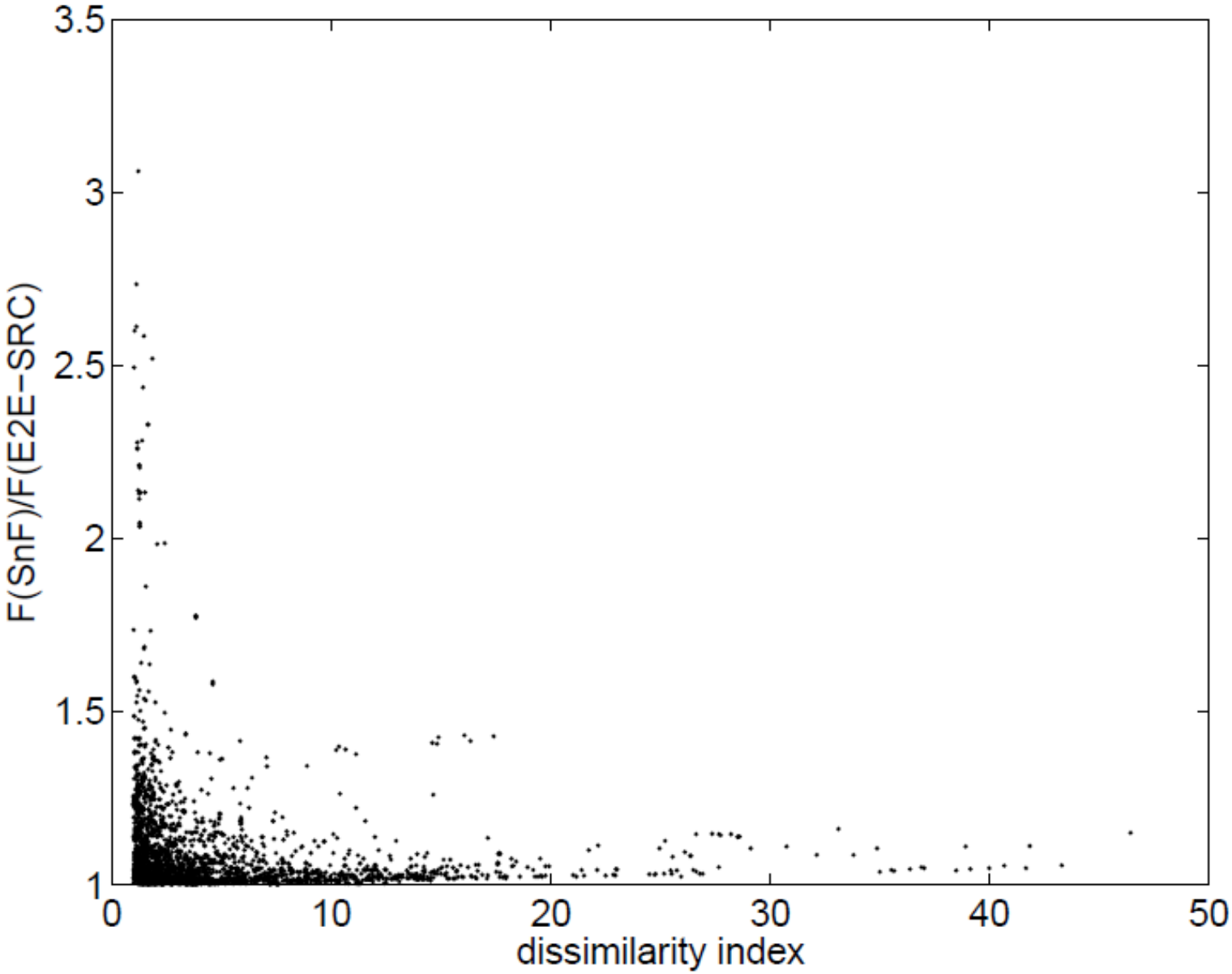
Methodology

- Data given by large Transit Provider for 2008 Q1
- 448 links with 140 ISPs
- Keep 280 that have $> 1 Gbps$ capacity
- Several are unpaid peerings
- Deadline 1 day
- Repeat for all working days of week

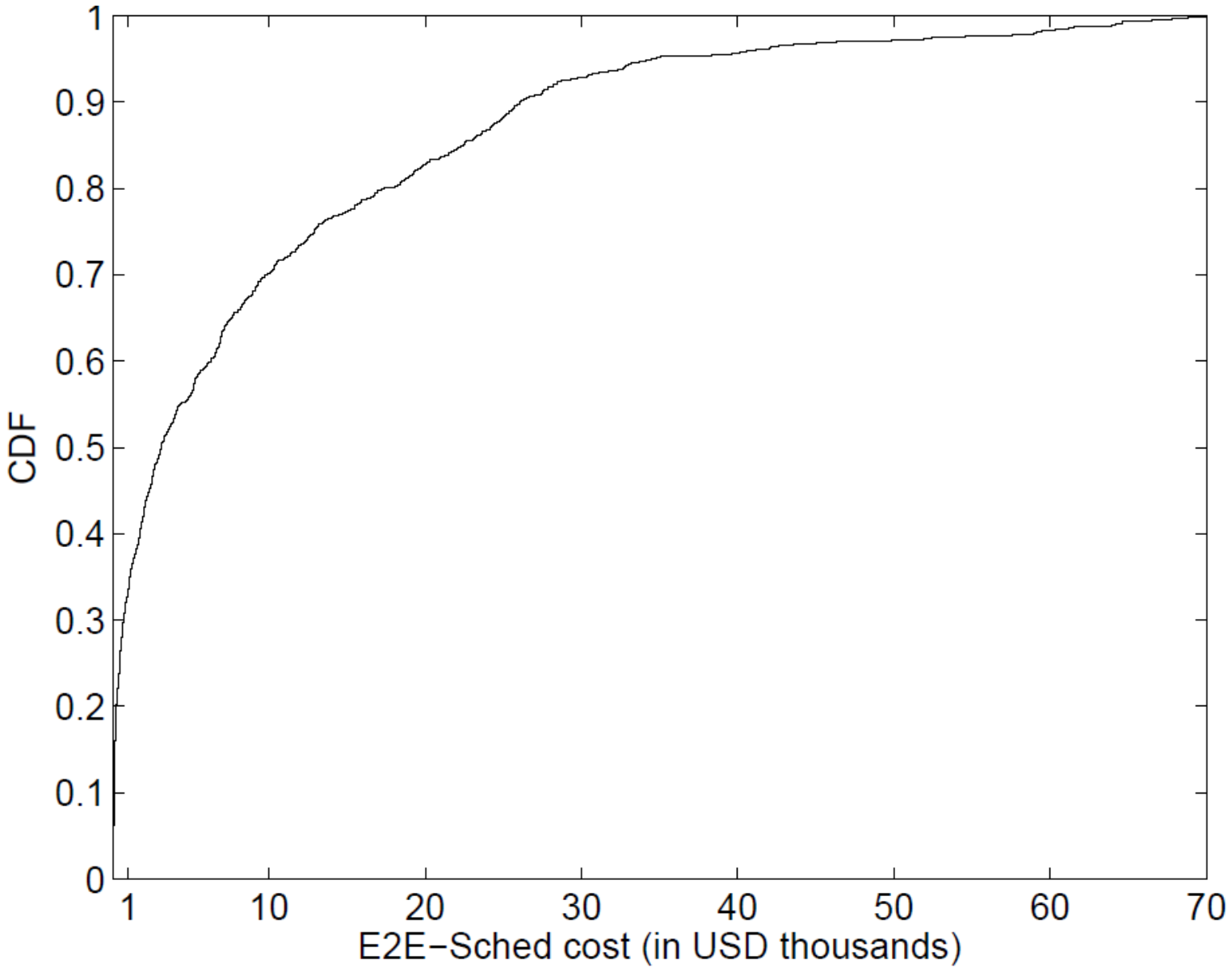
Free Volumes



Dissimilarity



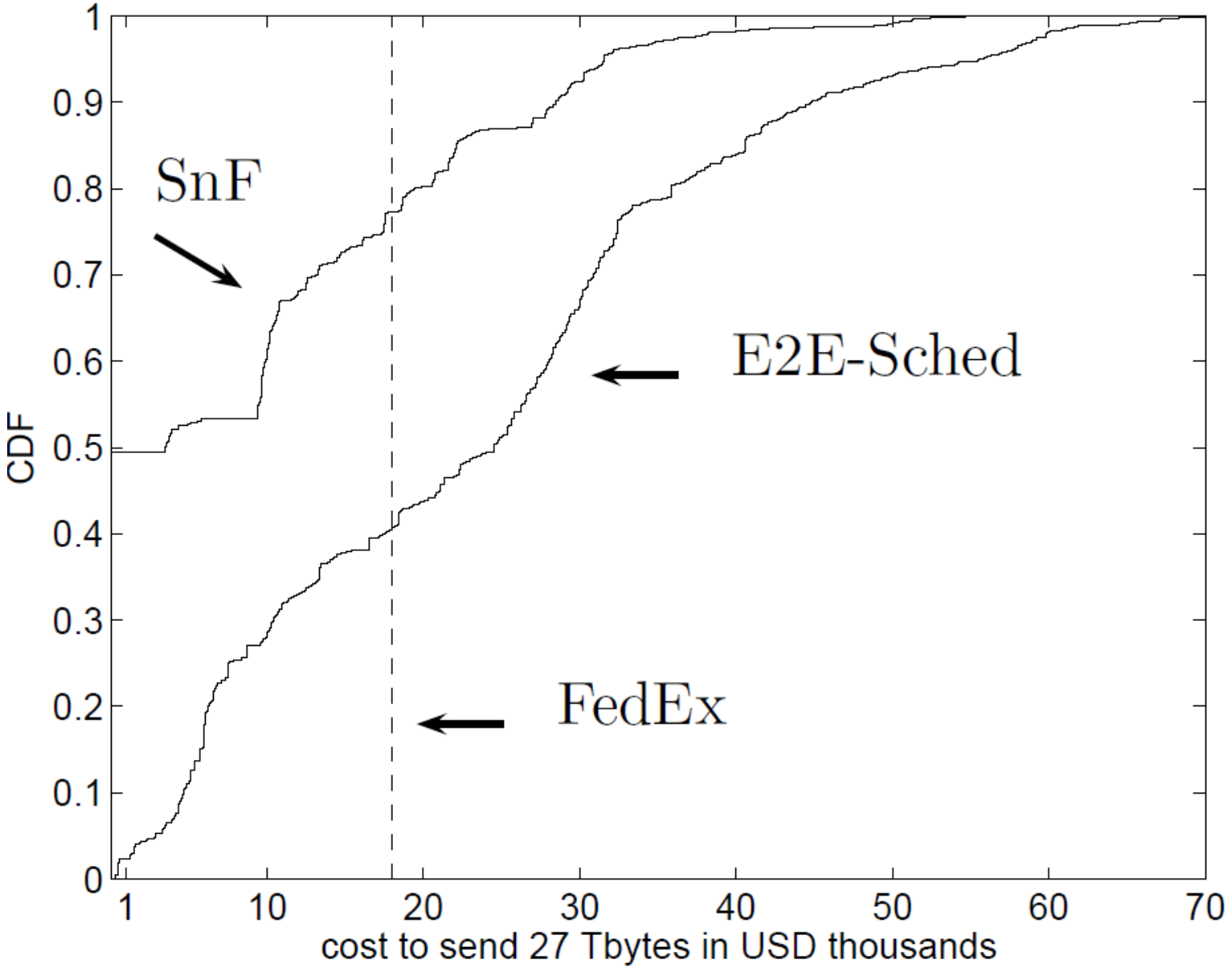
E2E vs. SnF


CDF
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
10
20
30
40
50
60
70
E2E−Sched cost (in USD thousands)

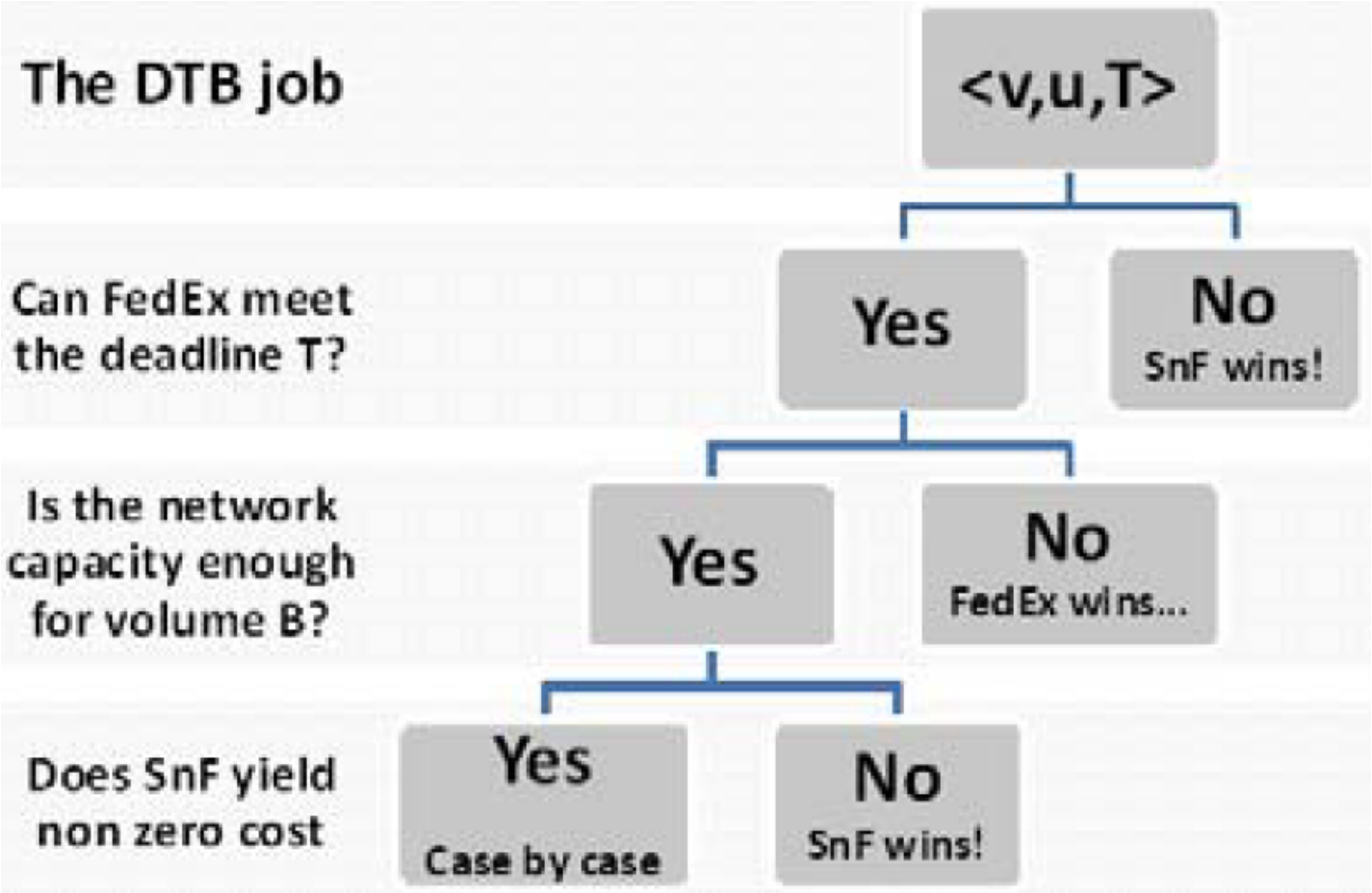
Storage Costs

- Back of the envelope calculation
- \$300/TB storage
- Server cost \$10,000
- Server lifetime 2 years
- Double for maintenance
- $\$ < 1K$ amortized cost
  - \$5K median for E2E
  - \$100,000s for constant-rate bulk without scheduling

27TB/day EU to LAT



Decision Tree



Criticisms

Pricing Model

- Need model based on peak demand
- Network costs defined by peak traffic
- Change for percentile for all traffic
  - Increase, e.g. to 99%, helps SnF
  - Decrease, e.g. to 50%, punishes non-DTB clients
- Transit ISPs claim part of transfer profit?
- Similar idea with electricity [QWB+09]
  - Would undermine relationship agreement

Criticisms

Estimates

- Estimates too rough
- Little data
- Irrelevant data
- No transit bottlenecks modeled

Implementation

- Evaluation too theoretical
- Follow-up work NetStitcher [LSYR11]
  - Introduced more intermediate hops
  - Allowed estimation error correction
- GRESE for specific types of bandwidth elasticity [NP12]
- Jetaway for video traffic [FLL12]

Conclusions

Key Contributions

- Model for free transfers (10-30TB for 10-40Gbps links)
- Simple decision tree choice (SnF usually wins)
- SnF useful in different time-zones with similar capacities
- E2E more expensive by \$5K in 50% of cases
- Courier better for occasional transfers

Key Questions

- Would more intermediate hops help?
- Is there no transit bottleneck?
- How can you combine jobs and optimize traffic?
- Will the price model change?
- Your questions?

Bibliography I


Yuan Feng, Baochun Li, and Bo Li, *Jetway: Minimizing costs on inter-datacenter video traffic*, Proceedings of the 20th ACM International Conference on Multimedia (New York, NY, USA), MM '12, ACM, 2012, pp. 259–268.


Nikolaos Laoutaris, Georgios Smaragdakis, Pablo Rodriguez, and Ravi Sundaram, *Delay tolerant bulk data transfers on the internet*, SIGMETRICS Perform. Eval. Rev. **37** (2009), no. 1, 229–238.


Nikolaos Laoutaris, Michael Sirivianos, Xiaoyuan Yang, and Pablo Rodriguez, *Inter-datacenter bulk transfers with netstitcher*, Proceedings of the ACM SIGCOMM 2011 Conference (New York, NY, USA), SIGCOMM '11, ACM, 2011, pp. 74–85.

Bibliography II

Thyaga Nandagopal and Krishna P. N. Puttaswamy, *Lowering inter-datacenter bandwidth costs via bulk data scheduling*, Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (Ccgrid 2012) (Washington, DC, USA), CCGRID '12, IEEE Computer Society, 2012, pp. 244–251.

Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs, *Cutting the electric bill for internet-scale systems*, SIGCOMM Comput. Commun. Rev. **39** (2009), no. 4, 123–134, `http://doi.acm.org/10.1145/1594977.1592584`.