Part I

Path Inferrence and Bandwidth Estimation

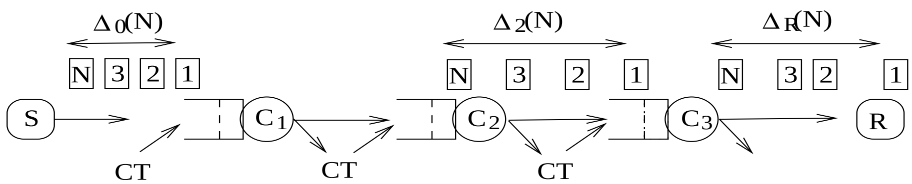
Inferring Path Characteristics Based on End-to-End Measurements

Constantinos Dovrolis

College of Computing
Georgia Tech

Path inferrence

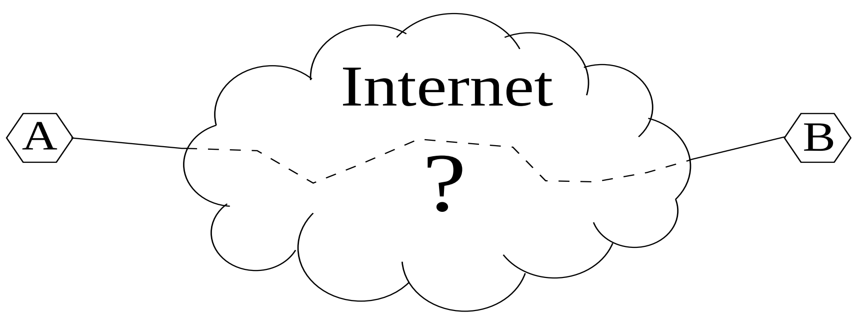
- Network path: sequence of links & routers from sender S to receiver R
- Path characteristics: round-trip time, delay jitter, loss rate, bandwidth, etc
- Inference techniques use special probing packets or application packets to measure path characteristics



- Objectives: accuracy, non-intrusiveness, timeliness

Looking inside a cloud..

- From the users' perspective, the Internet is a big **black** cloud
- Routers do not send explicit feedback to end-systems
- An important requirement for network simplicity and scalability



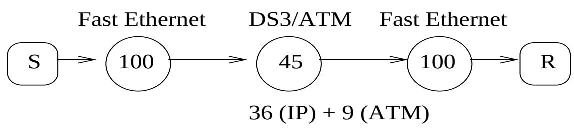
- End-systems need to *infer* network characteristics through end-to-end measurements
- Example: TCP Round-Trip Time (RTT) estimation

Applications of bandwidth estimation

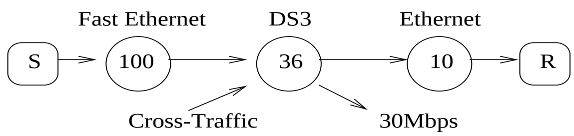
- Congestion control and TCP: automatic socket buffer sizing
- Overlay networks: configure overlay routes
- Content distribution networks: select best server
- Streaming applications: adjust encoding rate
- SLA and QoS verification: monitor path load
- End-to-end admission control: check for sufficient bandwidth
- Peer-to-peer networks: construct application-layer topology
- Interdomain traffic engineering: select egress ISP
- And many more..

Bandwidth estimation

- **Capacity:** maximum possible IP-layer throughput in path



- **Available bandwidth:** non-utilized part of path's capacity



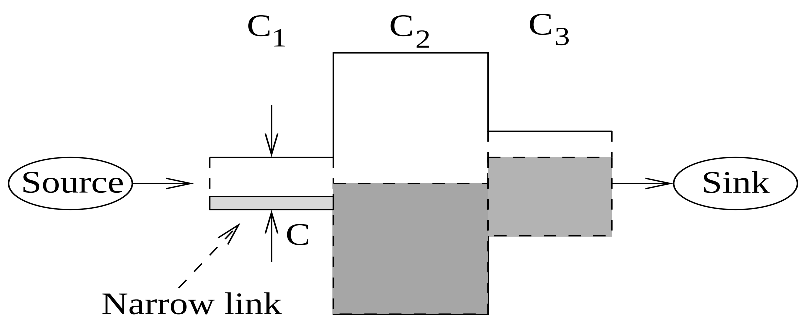
- **Bulk-Transfer Capacity (BTC):** average network-limited TCP throughput

Defi nition of capacity

- Link capacity: C_i for link i
- Path capacity:

$$C = \min_{i=0\ldots H} \{C_i\} = C_n$$

- Capacity is limited by *narrow link* n:

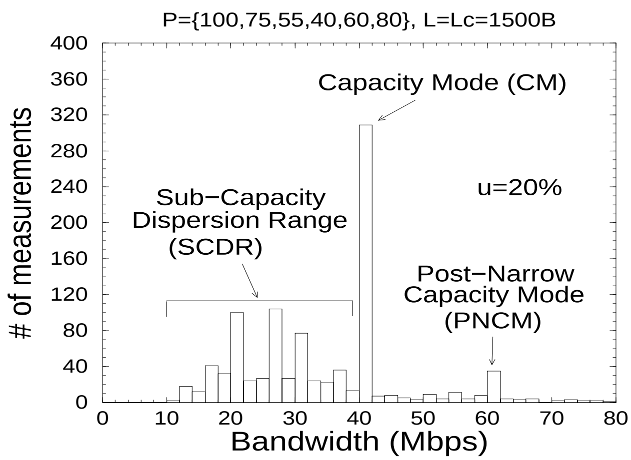


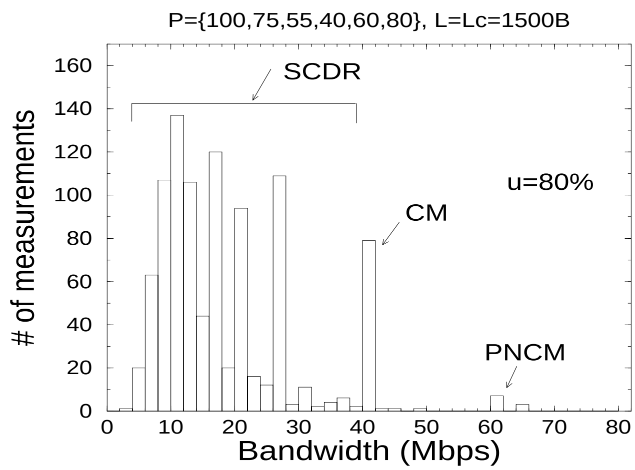
Part II

Capacity Estimation

But is it that simple?

- Bandwidth distribution from 1000 packet pair experiments

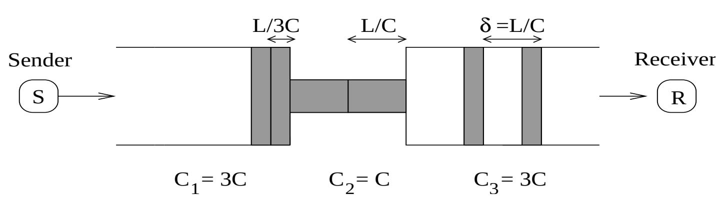




- Cross traffic creates local modes below and above capacity
- Objective: identify local mode that corresponds to capacity

The packet pair idea

- Transmission time of L-byte packet at link with capacity C: $\tau = \frac{L}{C}$
- Send two packets back-to-back from source to receiver
- Measure *dispersion* Δ of packet pair at receiver

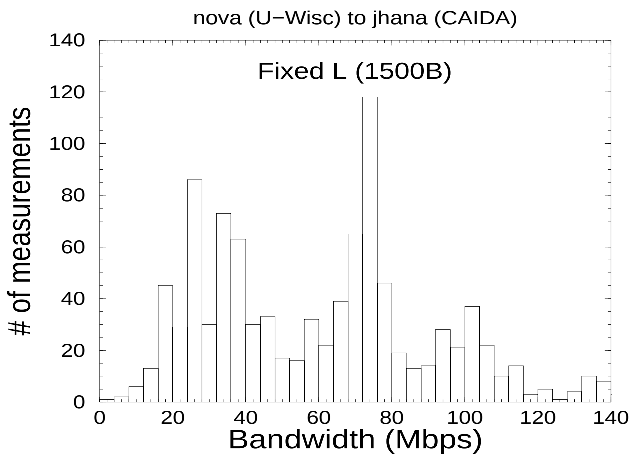


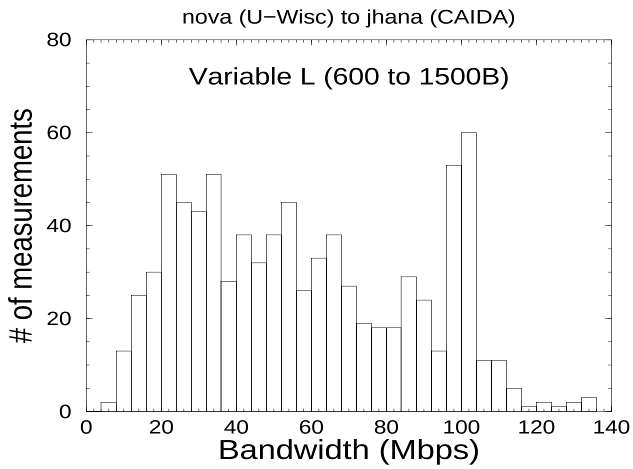
$$\Delta = \max_{i=0 \ldots H} \tau_i = \frac{L}{\min_{i=0 \ldots H}\{C_i\}} = \frac{L}{C}$$

- Basic capacity estimate: $\hat{C} = \frac{L}{\Delta}$

The physicist's approach

- Cross traffic packets tend to have a few common sizes (e.g., 40B, 1500B)

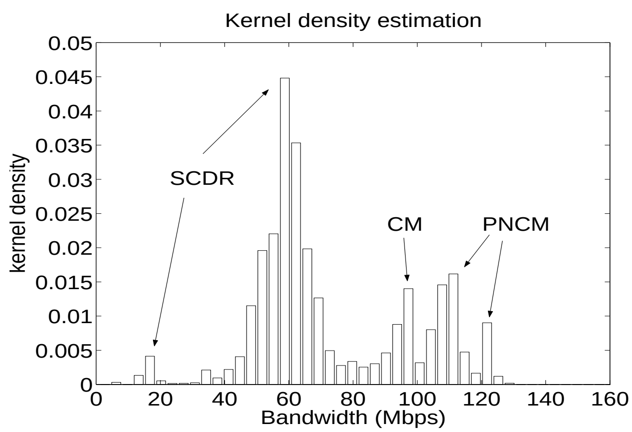


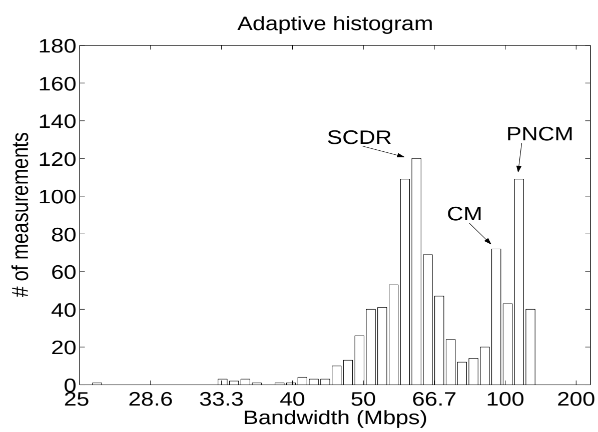


- Variable probing packet sizes reduce the intensity of local SCDR modes
- Also, dispersion rate of long packet trains is lower bound on capacity

The statistician's approach

- Approach-1: Kernel-density estimation
- Approach-2: Adaptive histogram (variable bin width)





- In general, unsuccessful to estimate capacity when a simple histogram would not work either

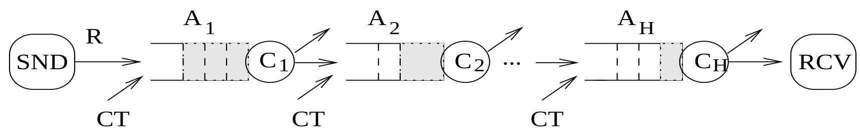
Part III

Available Bandwidth Estimation

Capacity estimation literature

- Jacobson, Keshav, Bolot: preliminary work on packet-pairs
- Carter & Crovella: Performance Evaluation '96, Infocom'97
  - Used variable-sized packets and union/intersection filtering
- Paxson: Sigcomm'97
  - Identified multiple bandwidth modes; used both pairs and trains
- Lai & Baker: Infocom'99
  - Used kernel-density estimation and max-sized probing packets
- Dovrolis, Ramanathan & Moore: Infocom'01
  - Explained bandwidth distribution of packet pairs & trains through queueing effects
- Pasztor & Veitch: IWQoS'02
  - Used dispersion histogram and peak detection; showed effect of L2 headers
- Harfoush, Bestavros & Byers: Infocom'03
  - Estimated capacity of targetted path segments

Self-Loading Periodic Streams (SLoPS)



- SND sends a periodic UDP packet stream of rate R
- Stream characteristics: K packets, size L, period T, rate $R = L/T$
- Measured One-Way Delay (OWD): $D^k = T^{RCV}_{arrive} - T^{SND}_{send}$
- OWD variation: $\Delta D^k = D^{k+1} - D^k$ (independent of clock offset)
- With a stationary & fluid model for the cross traffic, and FIFO queues:

  If $R > A = \min A_i$, *then* $\Delta D^k > 0$ *for* $k = 1, \ldots, K-1$

  Else, $\Delta D^k = 0$ *for* $k = 1, \ldots, K-1$

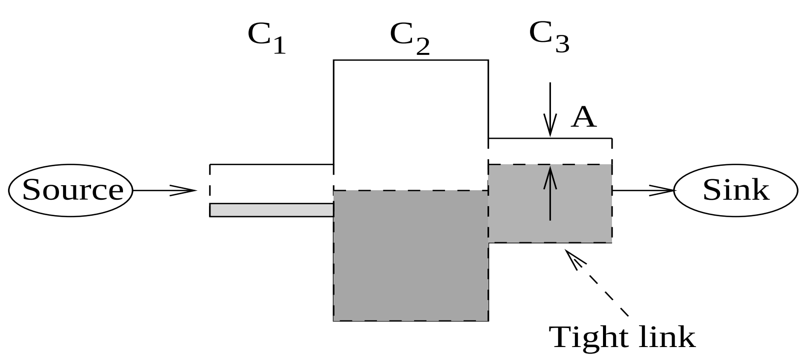
NeXtworking'03 June 23-25, 2003, Chania, Crete, Greece

Defi nition of available bandwidth (avail-bw)

- u_i: Average utilization of link i in a time interval of length τ ($0 \leq u_i \leq 1$)
- Avail-bw of link i: $A_i = C_i\,(1 - u_i)$

  End-to-end avail-bw: $A = \min_{i=0 \ldots H} A_i = \min_{i=0 \ldots H} C_i\,(1 - u_i)$

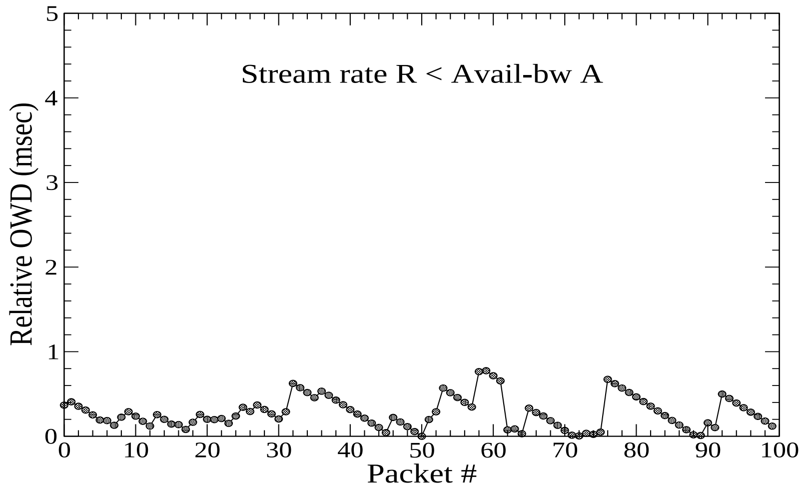
- Time interval length τ: averaging timescale



- Avail-bw is limited by *tight link*

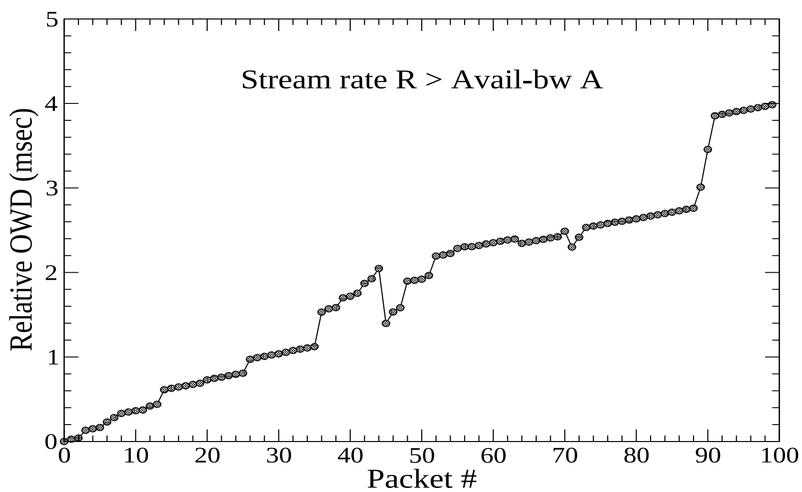
Non-increasing delay trend: $R < A$

- Path: Univ-Oregon to Univ-Delaware (12-hops)
- A=74Mbps (MRTG), R=37Mbps (K=100 packets, T=100μs, L=462B)



Increasing delay trend: $R > A$

- Path: Univ-Oregon to Univ-Delaware (12-hops)
- A=74Mbps (MRTG), R=96Mbps (K=100 packets, T=100μs, L=1200B)



Part IV

BTC Estimation

Available bandwidth literature

- Carter & Crovella: Performance Evaluation '96, Infocom'97
  - Estimated avail-bw from dispersion of long packet trains
- Melander, Bjorkman, & Gunningberg: Global Internet Symposium '00
  - Trains Of Packet Pairs (TOPP); like SLoPS, but with linear rate probing
- Jain & Dovrolis: Sigcomm '02
  - Self-Loading Periodic Streams (SLoPS); binary search rate probing
- Hu & Steenkiste: JSAC '03 (to appear)
  - Based on packet trains; aims to be faster than SLoPS
- Ribeiro et al.: PAM '03
  - Used chirp-like packet streams for probing at multiple rates

Part V

Research Challenges in Path Inferrence

BTC: definition and estimation

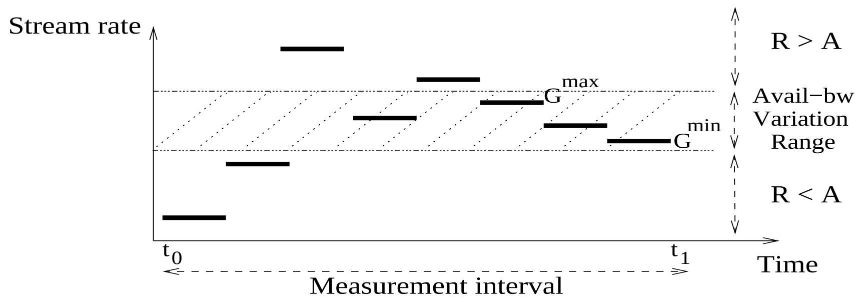
- Bulk-Transfer Capacity (BTC): long-term average TCP throughput
  - Congestion-limited transfer, i.e., sufficiently large receiver window
  - Depends on exact TCP implementation at sender & receiver
- BTC measurements are generally intrusive
  - BTC probing saturates measured path (depends on tight link buffering)
  - BTC probing affects cross traffic through increased RTTs and losses
- BTC non-intrusive estimation. Is it possible?
  - Use TCP throughput analytical models, such as:

$$\text{Throughput} = \frac{c\,\text{MSS}}{\text{RTT}\sqrt{\text{lossrate}}}$$

  - Ignores that TCP connection would increase RTT and lossrate!

Callenge-2: Variability

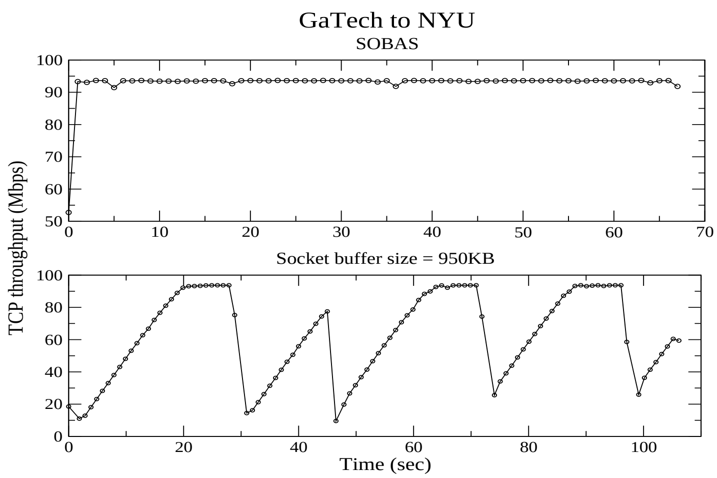
- With SLOPS, TOPP, etc, we can only estimate range of variation of avail-bw



- Avail-bw variability increases with load
- Also, variability depends on aggregation level
- How should applications deal with possibly large such variations?

Challenge-1: Integration with applications

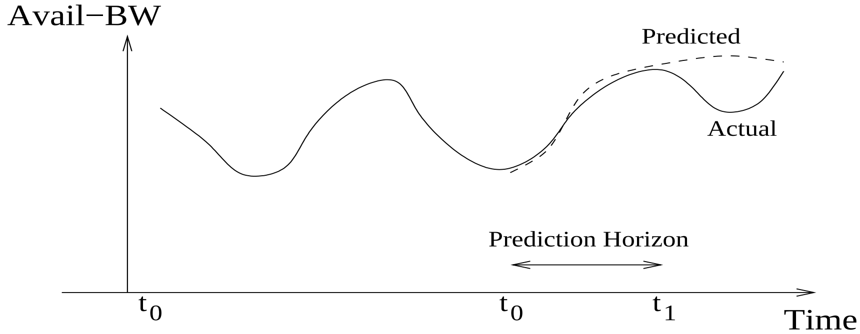
- Example: use bandwidth estimation in TCP socket buffer sizing (SOBAS)



- Optimal socket buffer size at maximum lossless receive-throughput
- Other approaches: TCP Westwood, TCP swift start etc

Callenge-3: Predictability

- How far in the future can we predict avail-bw?
- With what accuracy can we predict avail-bw?



- Note: Long-Range Dependency of network traffic improves predictability
- Preliminary work by Paxson and others shows that BTC is quite predictable