What do Packet Dispersion Techniques Measure?

C. Dovrolis, P. Ramanathan. D. Morre

Abstract

- Provide some understanding to *estimate* the capacity of a path.
- packet dispersion of a loaded path has a *multimodal* distribution.
- path capacity is often not a global mode.
- Unlike conventional wisdom, maximum packet size is not an optimal way to perform packet pair or packet train.
- Increase the length of packet train reduces the measurement variance but the estimates converge to an *asymptotic dispersion rate* (ADR), which is less than the capacity.
- Provide a tool called *pathrate*.

Motivation

- We need bandwidth monitoring tool.
- User can use the tool to check whether they gain access to the Internet with the bandwidth that they paid for.
- Basic tool to check for bandwidth provisioning.
- Bandwidth capacity planning
- To detect congestion or underutilized link
- To find *good path* for application-level multicast.
- To find *good path* for peer-2-peer applications.

Introduction

- A network path is a sequence of links that forward packets from source to sink.
- Two bandwidth metrics:
  - path capacity C.
  - path available bandwidth A.
- The path capacity is the maximum IP-layer throughput that the path can provide to a flow, when there is no competing traffic load (cross traffic).
- The available bandwidth is the maximum IP-layer throughput that the path can provide to a flow, given that path's current cross traffic load.

- C is determined by the minimum transmission rate of a link along the path.
- A is determined by the minimum unused capacity of a link along the path.
- If H is the number of hops in a path, C_i is the transmission rate of link i, C_0 is the transmission rate of the source:

$$C = \min_{i=0,\ldots,H} C_i \quad (1)$$

$$A = \min_{i=0,\ldots,H} C_i(1-\mu_i) \quad (2)$$

where μ_i is the utilization of link i

Packet Pair

- A packet pair consists of two packets sent out back-to-back at the source, each packet is of size L.
- If link i has capacity C_i, then the probe packet will have a transmission delay of

$$\tau_i = \frac{L}{C_i}$$

- Without any cross traffic in the path, the packet pair will reach the receiver *dispersed* (spaced) by the transmission delay of the **narrow link** $\tau_n = L/C$.
- The receiver can compute the capacity of a path C from the *measured dispersion* Δ, as $C = L/\Delta$.

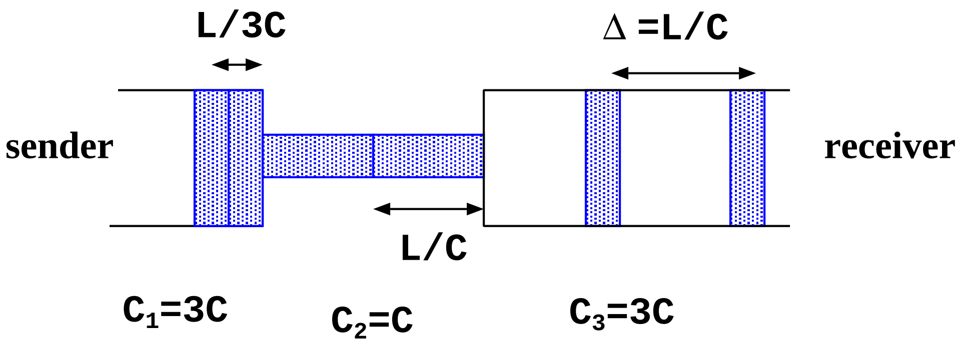


Figure 1: An illustration of the packet pair technique

- But is it really that simple? What does it really measure? What does it mean when there is cross traffic?
- If we use packet train (e.g., more than 2 packets), can we get more accurate result?

Packet Pair Dispersion

- Consider an H-hop path $P = \{C_0, C_1, \ldots, C_H\}$ and two packets of size L are sent back-to-back.
- The *dispersion* of the packet pair is the interval from the instant the last bit of the first packet is received at certain path point to the instant the last bit of the second packet is received at that point.
- Let $\Delta_0 = L/C_0$ after the source and Δ_i after the link i.
- When the packet pair reaches the sink, the dispersion Δ_H is measured and the receiver computes the bandwidth estimation as $b = L/\Delta_H$.
- Since Δ_H varies in general, we repeat the experiment many times and b becomes a random variable with distribution $\mathcal{B}$.

Packet Pair: No Cross Traffic

If there is no cross traffic:

- Δ_i cannot be lower than the dispersion at the previous hop Δ_{i-1} and the transmission delay $\tau_i = L/C_i$.

$$\Delta_i = max\{\Delta_{i-1}, \tau_i\}$$

(Illustrate)

- Applying this, we have

$$\Delta_H = \max_{i=0,\ldots,H} \tau_i = \frac{L}{\min_{i=0,\ldots,H}\{C_i\}} = \frac{L}{C_n} = \tau_n \qquad (3)$$

where τ_n and C_n are the transmission and capacity of the narrow link.

Packet Pair: With Cross Traffic (CT)

With CT, packet can experience queueing delays due to the cross traffic.

- Let d_i^1 be the queueing delay of the first probe packet at link i.
- Let d_i^2 be the queueing delay of the second probe packet at link i *after the first packet has been transmitted at that link.*

$$\Delta_i = \begin{cases} \tau_i + d_i^2 & \text{if } \tau_i + d_i^1 \geq \Delta_{i-1} \\ \Delta_{i-1} + (d_i^2 - d_i^1) & \text{otherwise} \end{cases} \tag{4}$$

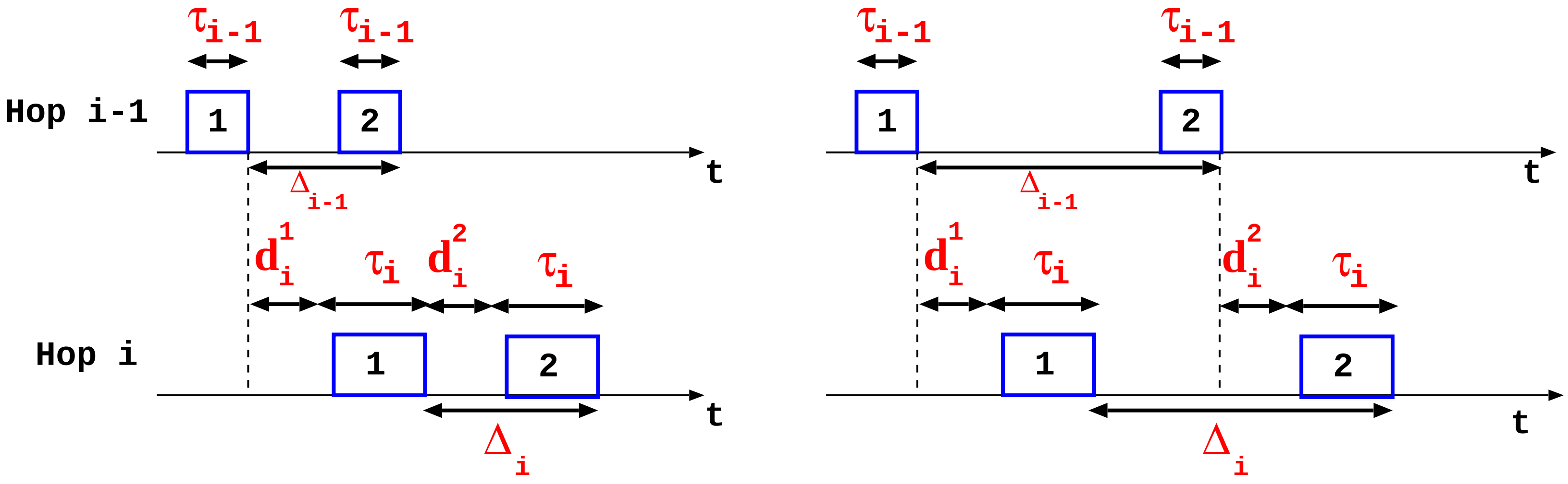
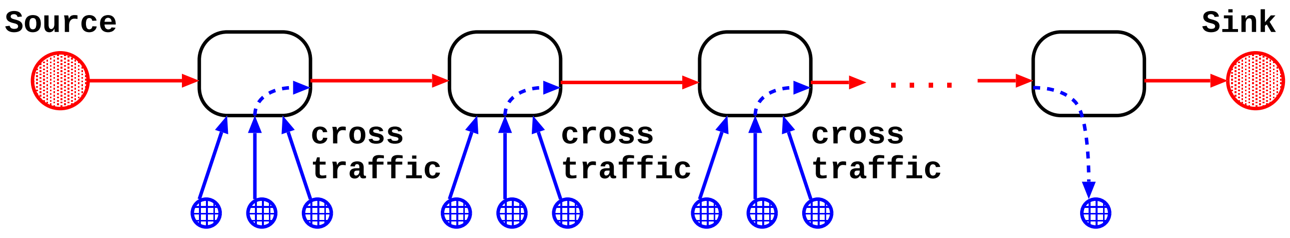


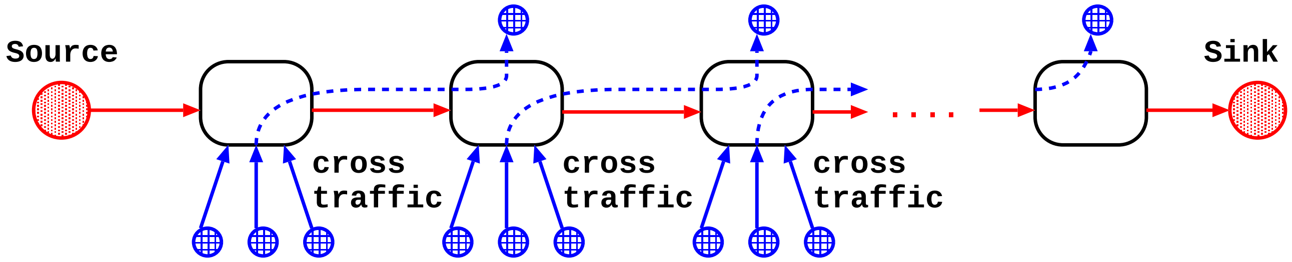
Figure 2: The Two Cases

Note that when $\tau_i + d_i^1 < \Delta_{i-1}$ and $d_i^2 < d_i^1$, **the dispersion decreases from hop $i-1$ to hop i.** This causes a dispersion at the receiver that is *lower* than the dispersion at the narrow link, i.e., $\Delta_H < \tau_n = L/C$. So the capacity of a path is NOT simply the minimum measured dispersion.

Experiments



(a) Path persistent cross traffic

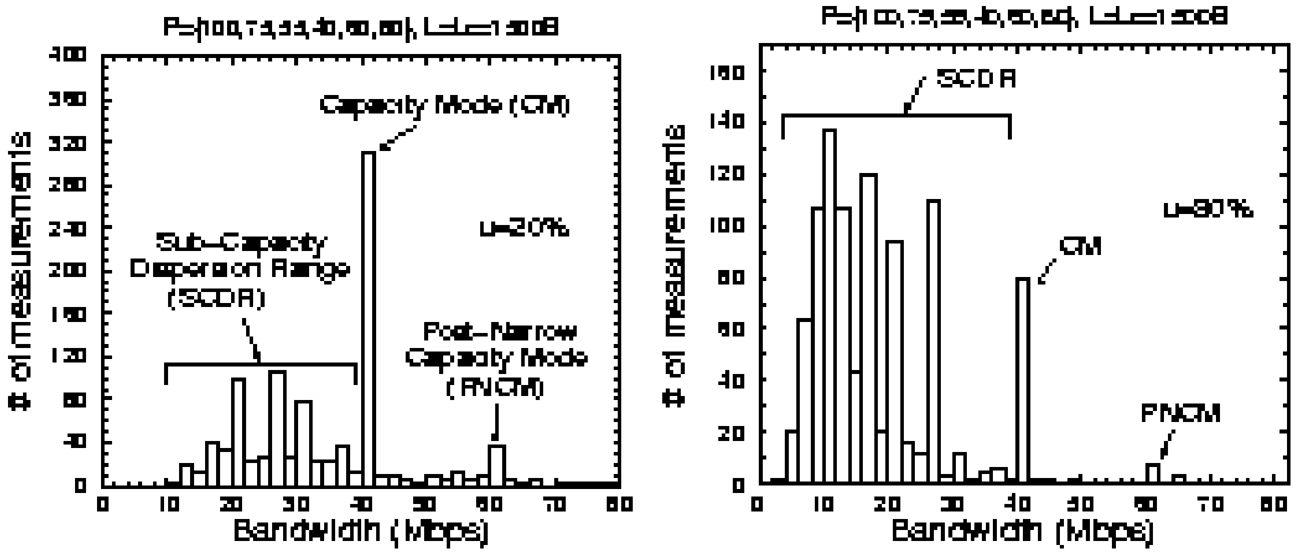


(a) One-hop persistent cross traffic

NS-2 simulation: cross traffic is generated from sixteen Pareto sources at each hop. Link is FCFS. Estimate bandwidth $\mathcal{B}$.

Experiments

$\mathcal{P} = \{100, 75, 55, 40, 60, 80\}.$



(a) Light load conditions (b) Heavy load conditions

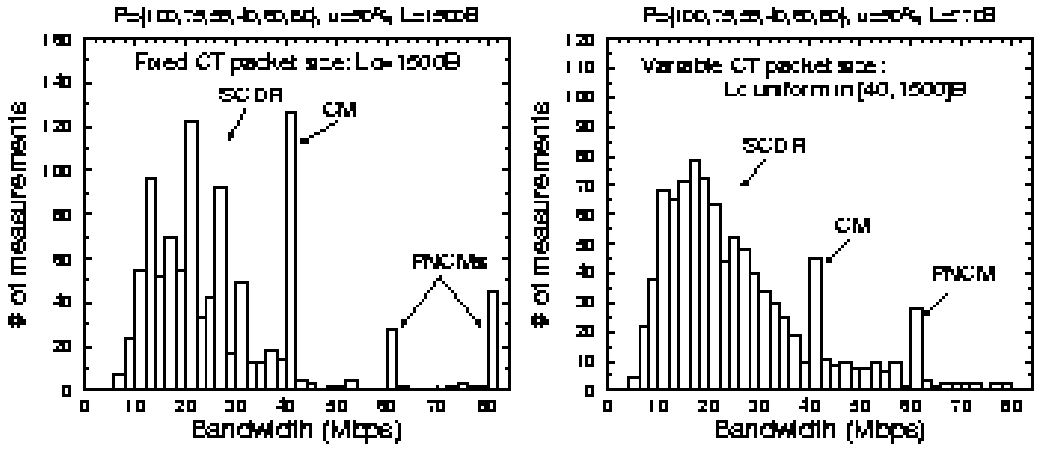
Fig. 4. The B distribution in two different path loads.

- When the path is lightly loaded (20%), we can detect the path capacity (40 Mbps) is prevalent in $\mathcal{B}$, forming the *Capacity Mode* (CM).
- Bandwidth estimates that are lower than CM are caused by CT, they are called *Sub-capacity Dispersion Range* (SCDR).
- Bandwidth estimates that are higher than CM are caused in the post-narrow links when the first probing packet is delayed more than the second, we called this the *Post-Narrow Capacity Modes* (PNCM).

- Under heavy load (e.g., 80%), the CT interfering is large.
- The global mode of $\mathcal{B}$ is not the CM, but rather, the SCDR.
- The path capacity cannot be always correctly estimated by statistical techniques that extract the most common bandwidth value or range.

Experiments

$\mathcal{P} = \{100, 75, 55, 40, 60, 80\}$, $\rho = 0.5$, Packet size of CT L_c is 1500 and probe packet L is 1500 in (a). L_c is uniformly distributed between $[40, 1500]$ and L is 770 in(b).



(a) Fixed CT packet size (b) Variable CT packet size

Fig. 5. Fixed versus variable CT packet size L_c.

- When $L_c = L = 1500$, we can explain the SCDR. For example, when $\mathcal{P} = \{100, 60, 40\}$, a local mode of 30 Mbps can by caused by a packet interfering with the packet pair at 60 Mbps and it causes a dispersion after the narrow link of

$$\Delta_n = \frac{L}{40} + (2\frac{L}{60} - \frac{L}{40}) = \frac{L}{30}$$

- A 20 Mbps is caused by a packet interfering with packet pair at 40 Mbps or by two packets interfering at 60 Mbps.
- When $L_c \neq L$, result is more difficult to explain since it can produce a range of SCDR.
- Nevertheless, CM and PNCMs are still distinct in the distribution.

Packet Train

- The source sends $N > 2$ back-to-back packets of size L to the sink.
- The sink measures the total dispersion $\Delta(N)$ from the first to the last probing packets, computes the bandwidth estimate as

$$b(N) = \frac{(N-1)L}{\Delta(N)}$$

  Many experiments form the bandwidth distribution $\mathcal{B}(N)$.
- The original argument is to use packet train so that measurement is more robust to noise caused by CT.
- Although the $\Delta(N)$ becomes larger as N increases, so does the "noise" in the measured values of $\Delta(N)$. Why? Because it

becomes more likely that CT packets will interfere in the packet train. This will lead to *Bandwidth underestimation*.

Experiments

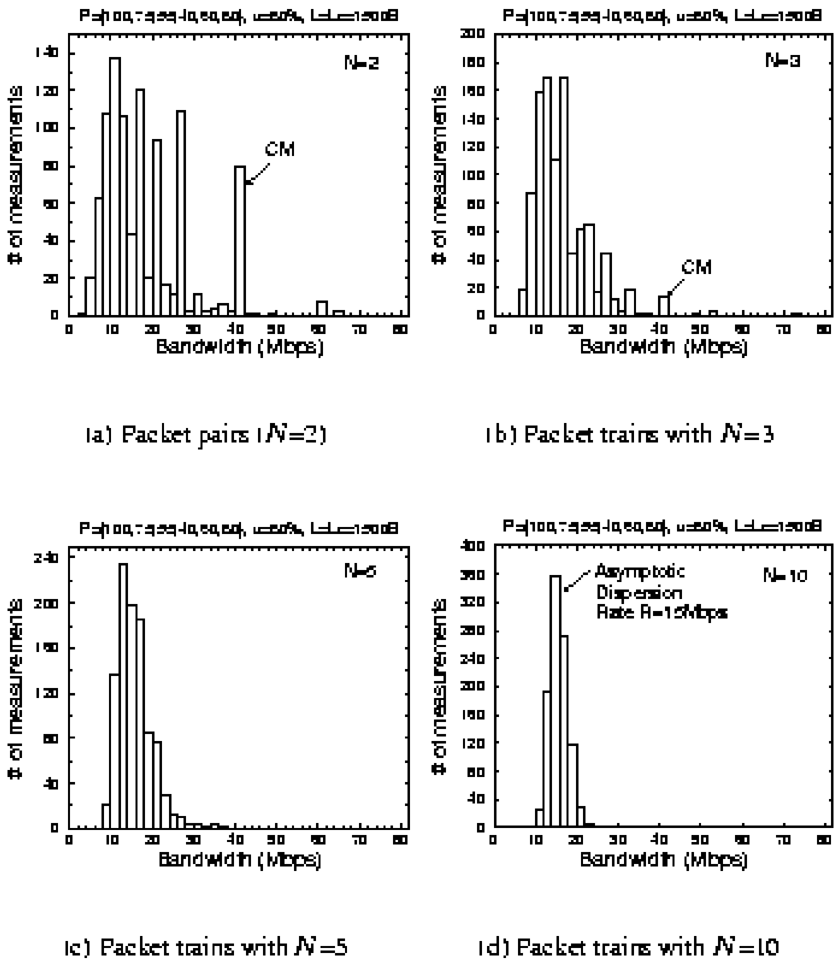


Figure 3: The effect of the packet train length, $\rho = 0.8$, $\mathcal{P} = \{100, 75, 55, 40, 60, 80\}$

- As we increase N, the CM and PCNMs become weaker and then disappear.
- SCDR prevails in $\mathcal{B}(N)$.
- As we increase the length of the packet train, almost all packet trains encounter additional dispersion due to the CT packets.
- As N increases, $\mathcal{B}(N)$ becomes unimodal. We called this the Asymptotic Dispersion Rate (ADR).
- As N increases, the variance of $\mathcal{B}(N)$ decreases, this implies the variance in the amount of cross traffic interfering with the packet train decreases.

To understand ADR

Simple case first: $\mathcal{P} = \{C_0, C_1\}$ where $C_0 \geq C_1$, packet train of length N.

- The initial dispersion $\Delta_0 = L(N-1)/C_0$.
- Let r_1 be the average incoming rate of CT in link 1. The average amount of CT arrives in link 1 during Δ_0 is $X_1 = \Delta_0 r_1$. This becomes the *interference* of the packet train.
- The dispersion in link 1 is:

$$\Delta_1 = \frac{(N-1)L + X_1}{C_1} = \frac{(N-1)L}{C_1}(1 + \mu_1 \frac{C_1}{C_0}) \tag{5}$$

where $\mu_1 = r_1/C_1$ is the utilization of the narrow link to cross traffic.

- The measure bandwidth estimate (or ADR) is:

$$R = \frac{(N-1)L}{\Delta_1} = \frac{C_1}{1+\mu_1 \frac{C_1}{C_0}} < C_1 \qquad (6)$$

- the ADR is lower than the link capacity.
- If C_1 and C_2 are known, we can compute the cross traffic utilization μ_1 and then compute the **available bandwidth** $A = C_1(1-\mu_1)$.

H-hops ADR

Consider H-hops with $C_0 \geq C_1 \geq \cdots \geq C_H$. Let r_i be the average rate of CT on link i. Assume they are path-persistent CT.

- The average dispersion at the exit of link i is: $\Delta_i = \Delta_{i-1}(C_{i-1} + r_i)/C_i$ (derive).
- The ADR is:

$$R = \frac{(N-1)L}{\Delta_H} = \frac{C_H}{\prod_{i=1}^{H}(1 + \frac{r_i}{C_{i-1}})} \quad (7)$$

- For example, for $\mathcal{P} = \{C_0, C_1, C_2\}$ where $C_0 \geq C_1 \geq C_2$: $R = \frac{C_2}{1+\frac{r_1}{C_0}+\frac{r_2}{C_2}+\frac{r_1 r_2}{C_0 C_1}}$
- Authors derived upper and lower bound of ADR for general setting (refer to paper).

Pathrate (tool for capacity estimation)

- In phase 1, use packet pair to uncover all local modes of $\mathcal{B}$. We call these modes $\mathcal{M}$. If $\mathcal{B}$ is unimodal, the mode is the CM (this is the case for lightly loaded path).
- In phase 2, use packet train. We can increase the packet train length to eliminate CM and $PNCMs$ from $\mathcal{M}$. If we continue, we can determine the ADR.
- For detail implementation, refer to paper or the package.