Routing High-bandwidth Traffic in Max-min Fair Share Networks

Qingming Ma, Peter Steenkiste, Hui Zhang

Computer Science Department

Carnegie Mellon University

Outline

- Background
  - high-bandwidth traffic
  - max-min fair share networks
- Single-path routing
  - link state & cost metrics
- Multi-path routing
  - multi-level max-min fairness
- Conclusions

Different types of traffic

- Small interactive messages
  - low end-to-end per packet delay
- Continuous-media stream
  - intrinsic rate and delay
- ***Bulk data transfers---high-bandwidth traffic***
  - medical images and large files and data sets
  - low elapsed time:

    set-up + end-to-end-delay + (bytes / rate)
  - high rate is the key to reducing elapsed time

Approaches to improve performance

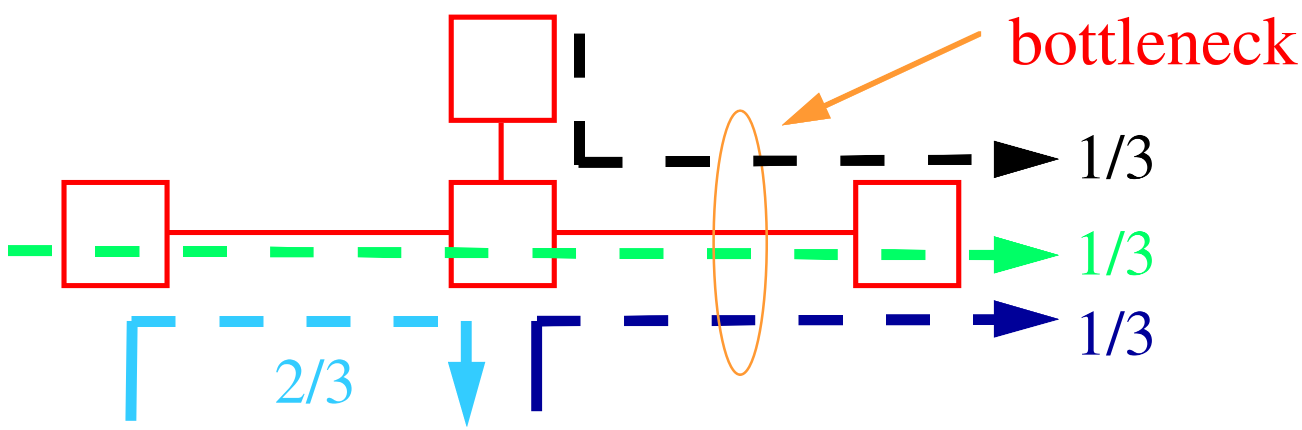
- Traffic control
  - prioritizing, reservation, …
  - preferential treatment: needs pricing and/or policies

OR

- ***Use resources more efficiently***
  - optimized routing algorithm
    - select path with highest rate
  - routing decision based on
    - congestion information
    - application information

Max-min fair share network

- A fair rate allocation criterion
  - [Jaffe 81, Mosley 84, Hahne 86, Charny 94, …]
  - ATM ABR traffic management
- Flows share bandwidth equally unless being bottlenecked by other links
  - max-min fair share rate changes dynamically
  - possible to predict max-min share rate of new flows



Routing high-bandwidth traffic

- Need to access network load information
  - link-state routing
  - e.g., OSPF, ATM PNNI
- Routing decision on a per-flow basis
  - we used source routing
- Select path that delivers high throughput
  - link cost functions based on max-min rate, not residual bandwidth

Link-state information

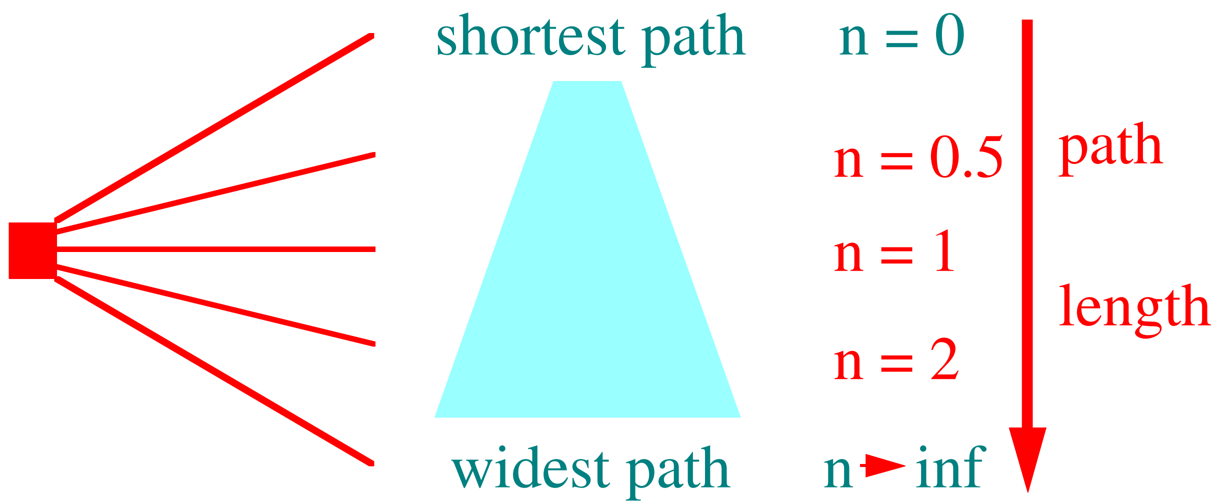
- Per-flow information can be too expensive
  - routing information distribution
  - max-min fair rate computation
- Several alternatives to reduce state info
  - example: rate buckets

10	15	2	1
2 Mb/s	3 Mb/s	20 Mb/s	45 Mb/s

Link metrics: balance length and width

- shortest path
  - consume less resources, but rate can be low
- widest path
  - high rate, but use more resources
  - more likely to lose share to new arriving flows
- use index *n* to balance length and width

$$\sum \frac{1}{r_i^{\,n}}$$

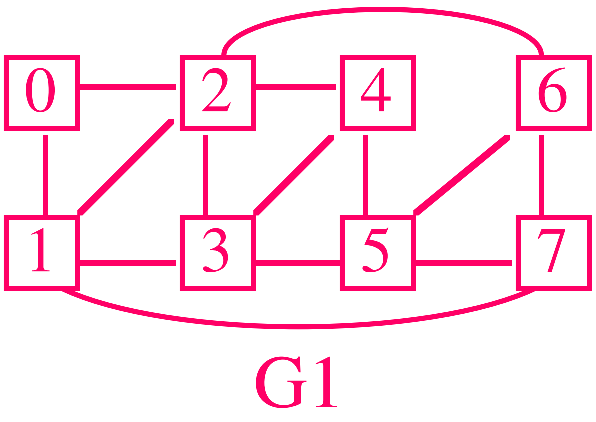


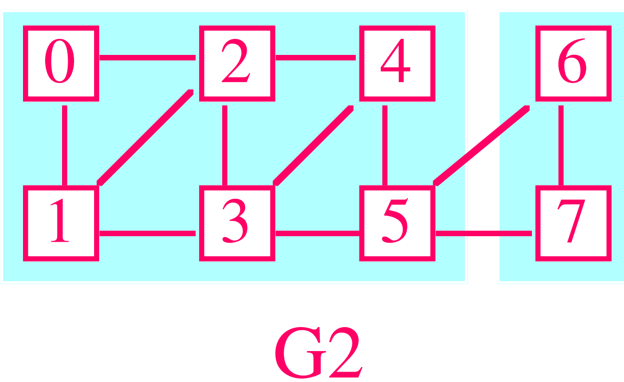
Evaluation methodology

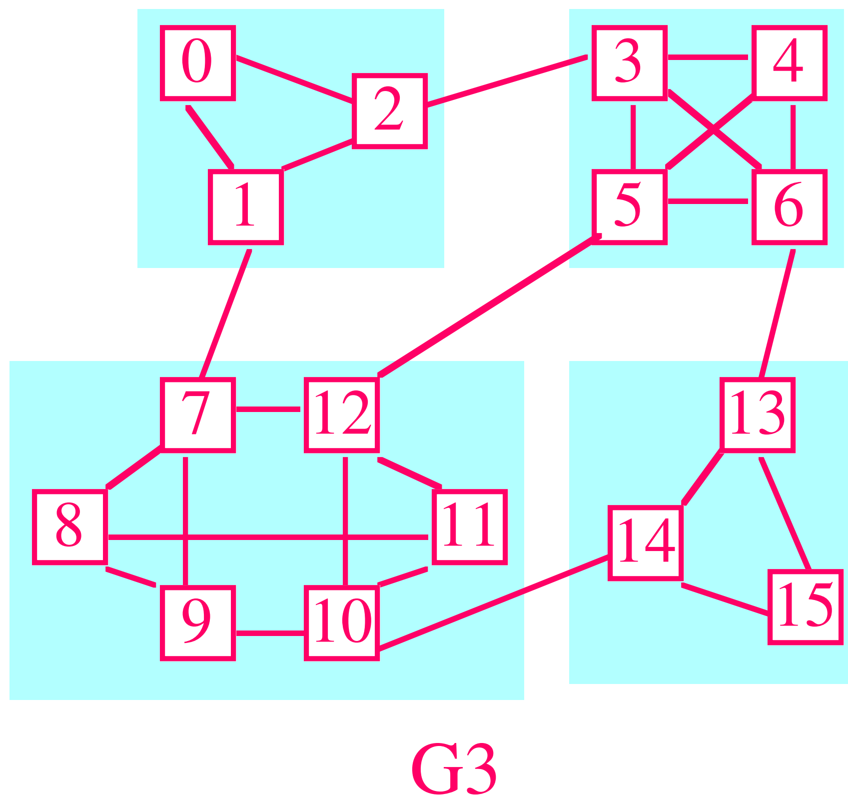
- Session-level event-driven simulator
- Traffic load
  - exponential arrival from hosts
  - ratio of HB traffic vs. non-HB traffic
  - uniform distribution vs. traffic concentration
- Simulation parameters

traffic type	message size	routing overhead	peak rate	setup cost
HB	1MB-1GB	10 ms	155 Mb/s	3 ms
non-HB	1KB-1MB	0 ms	3-5 Mb/s	3 ms

Three topologies

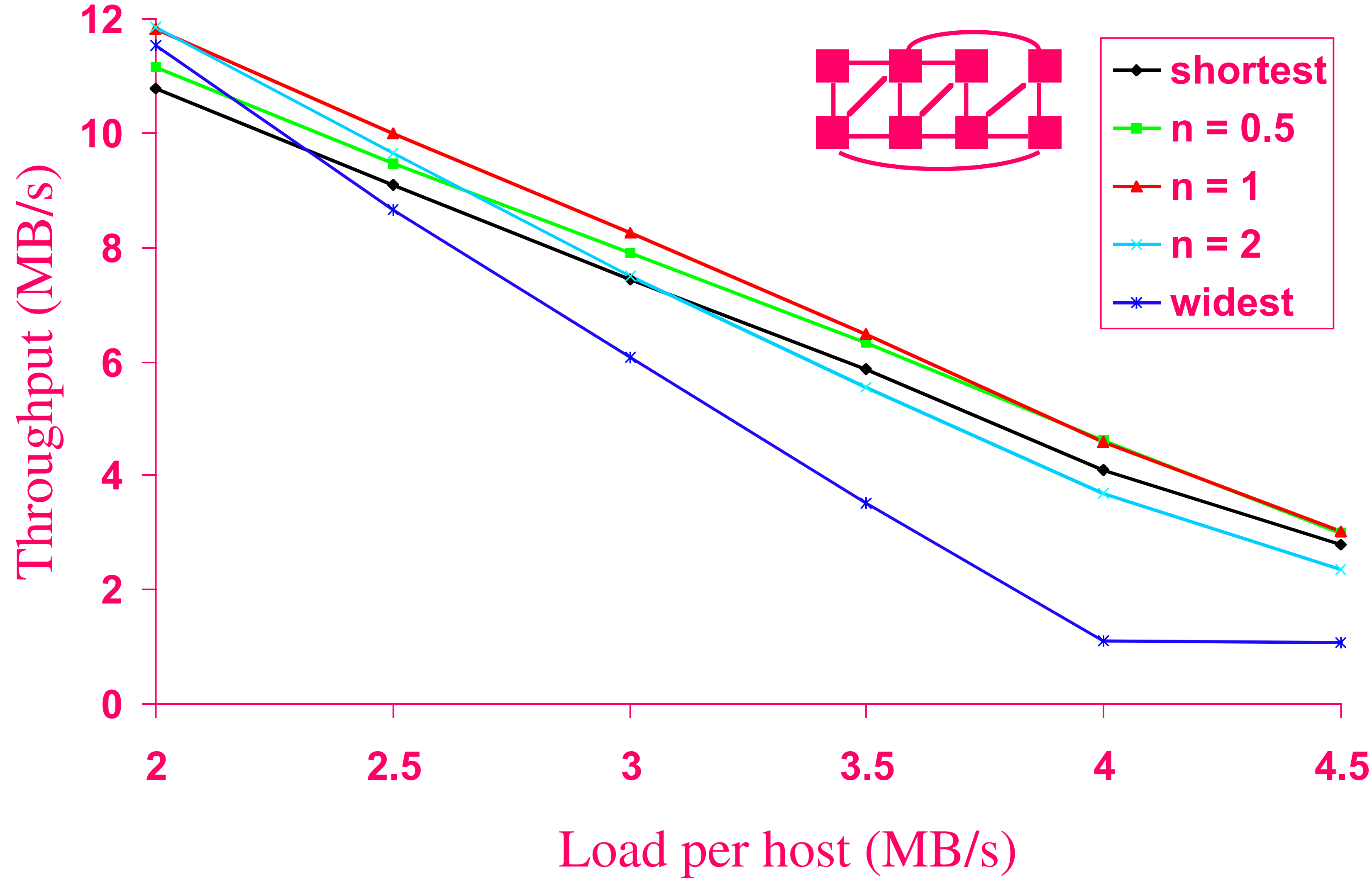




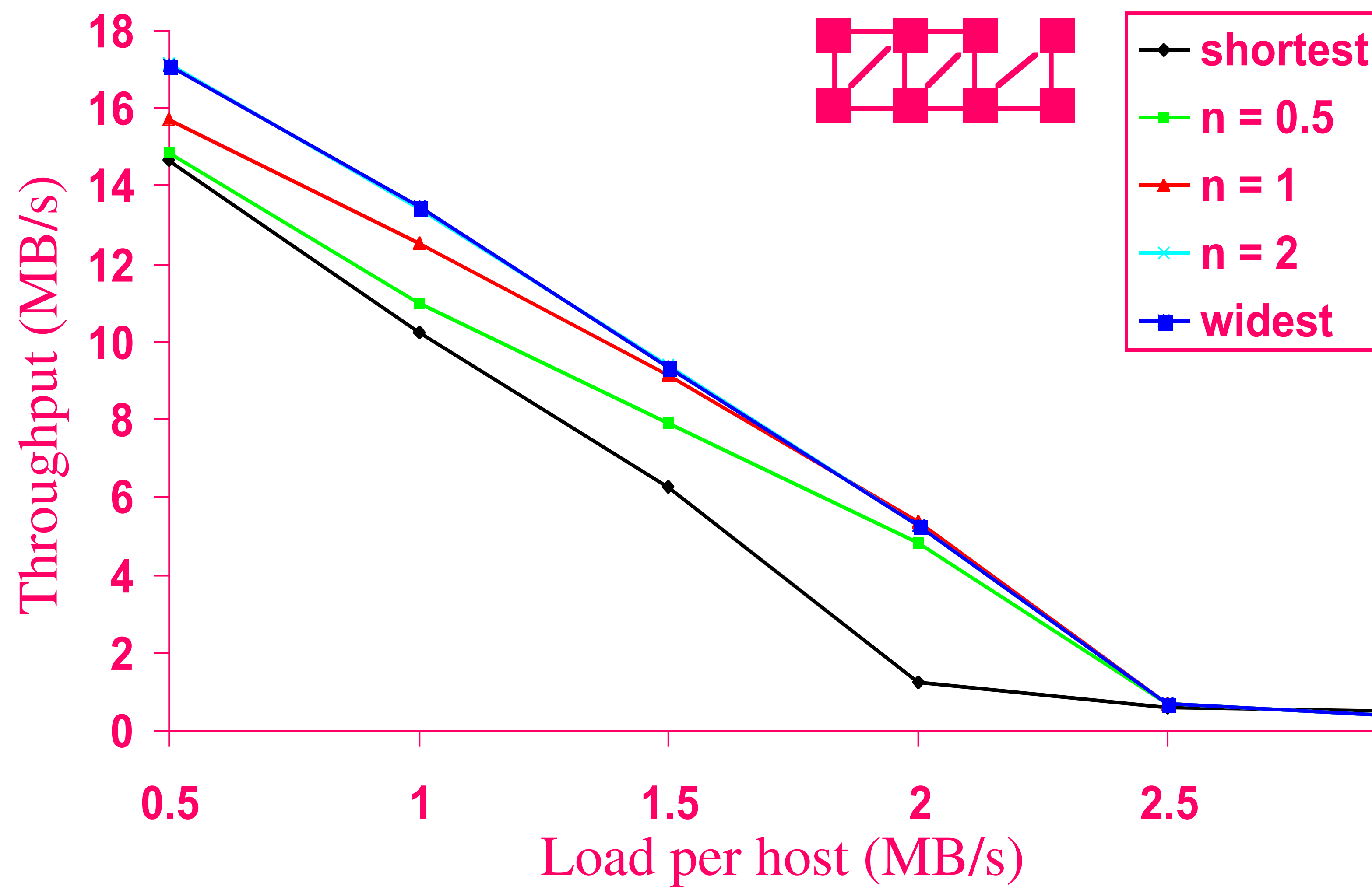


Size, connectivity, symmetry, 8 hosts/switch

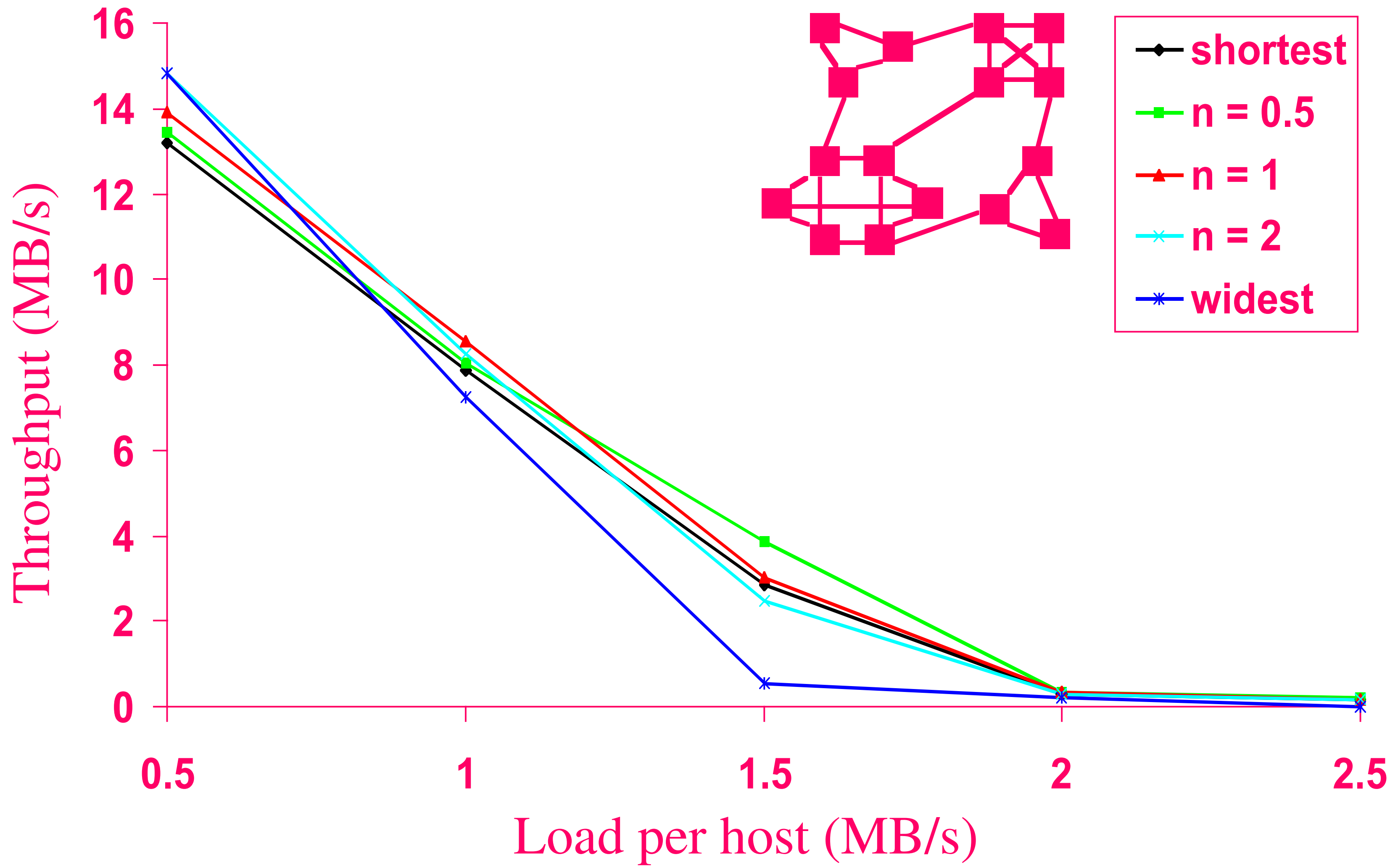
Average throughput for different load: G1



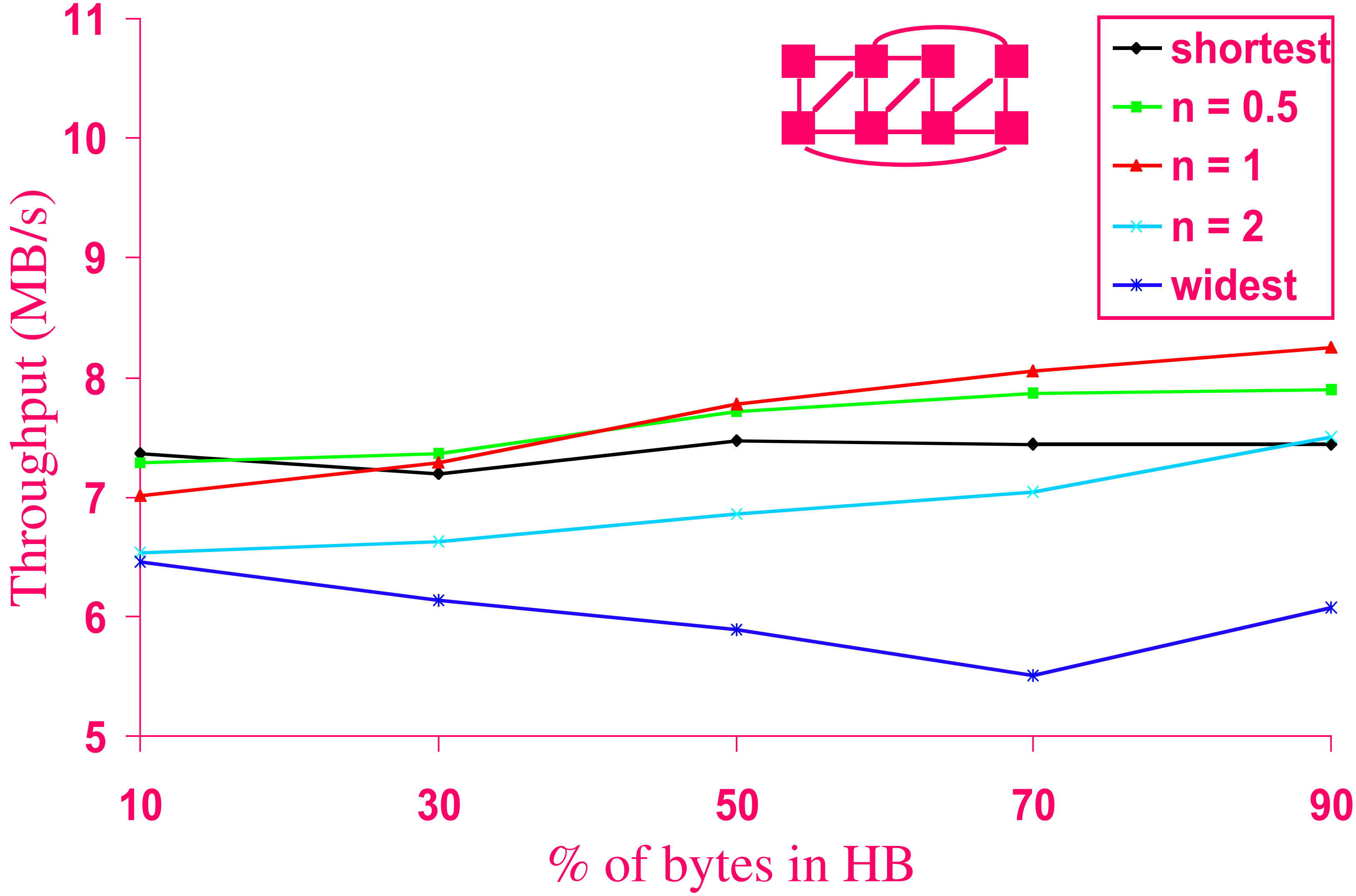
Average throughput for different load: G2



Average throughput for different load: G3



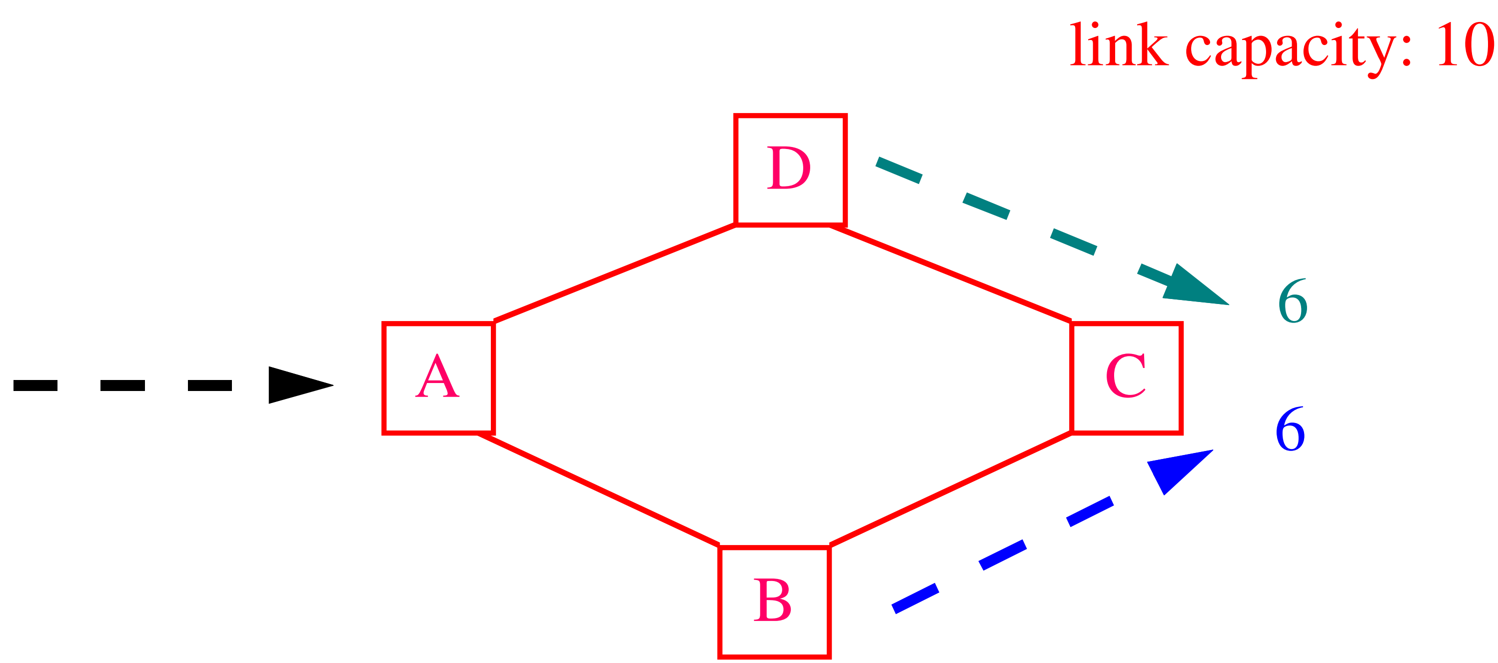
Average throughput for different HB volume



Summary: single-path routing

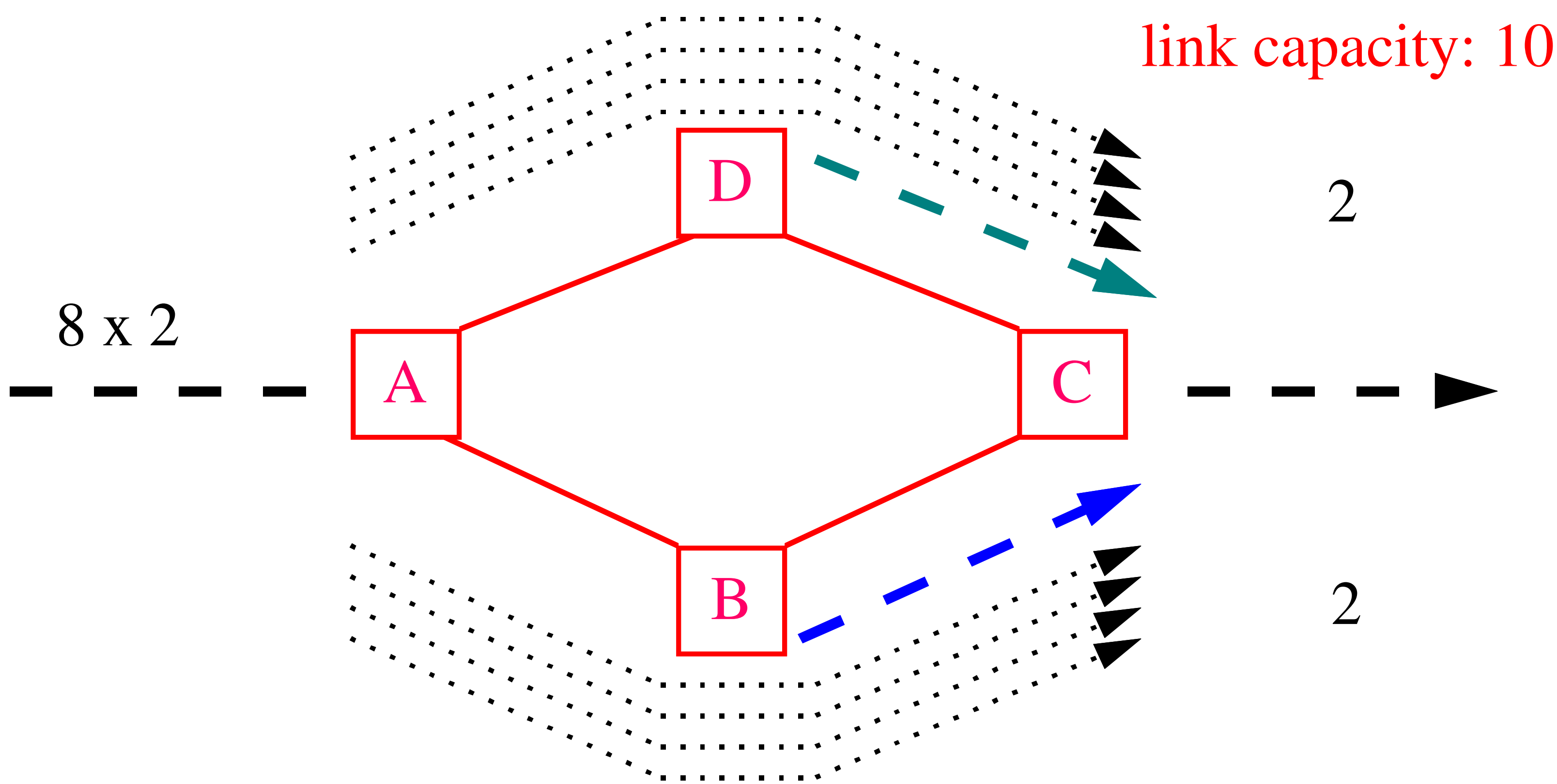
- Shortest or widest path can perform poorly, depending on traffic load and topology
- Link metrics 1/r performs well under
  - both heavy and light load
  - uniformly distributed & concentrated traffic
  - different topology and size of the network
  - different volume of HB traffic
- Other traffic classes also benefit from improved performance for HB traffic

Multi-path routing

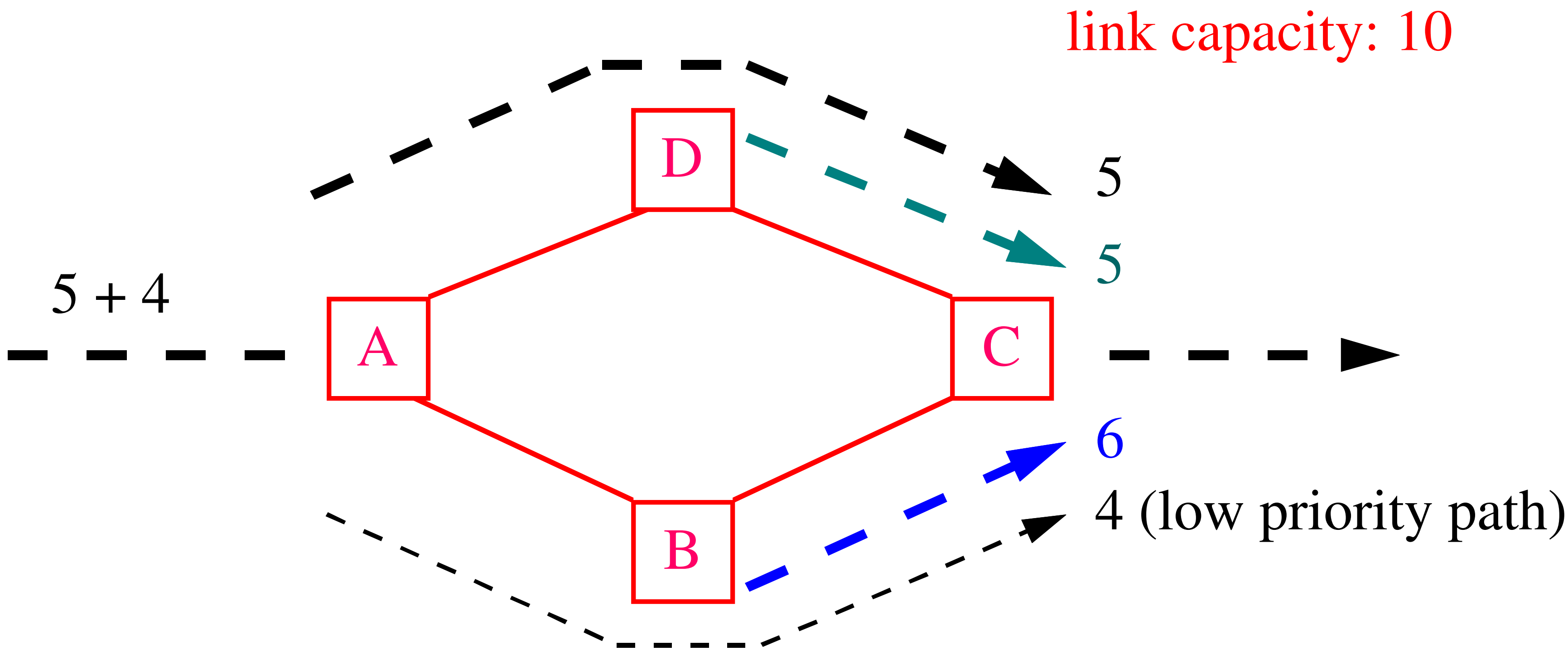


- Use unused bandwidth
- Protect users of single path

A scenario of bandwidth squeezing

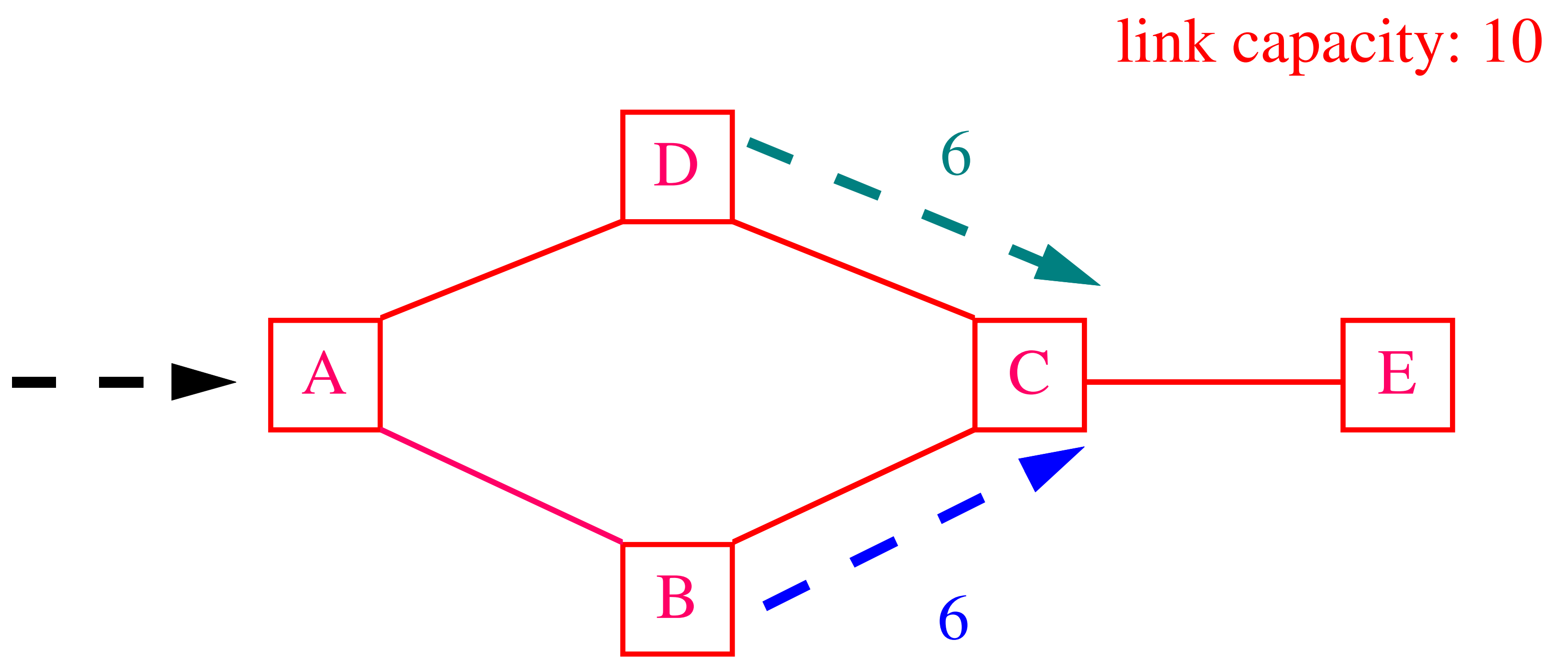


Prioritizing multi-path



- Secondary path uses low priority

Why not disjoint paths?

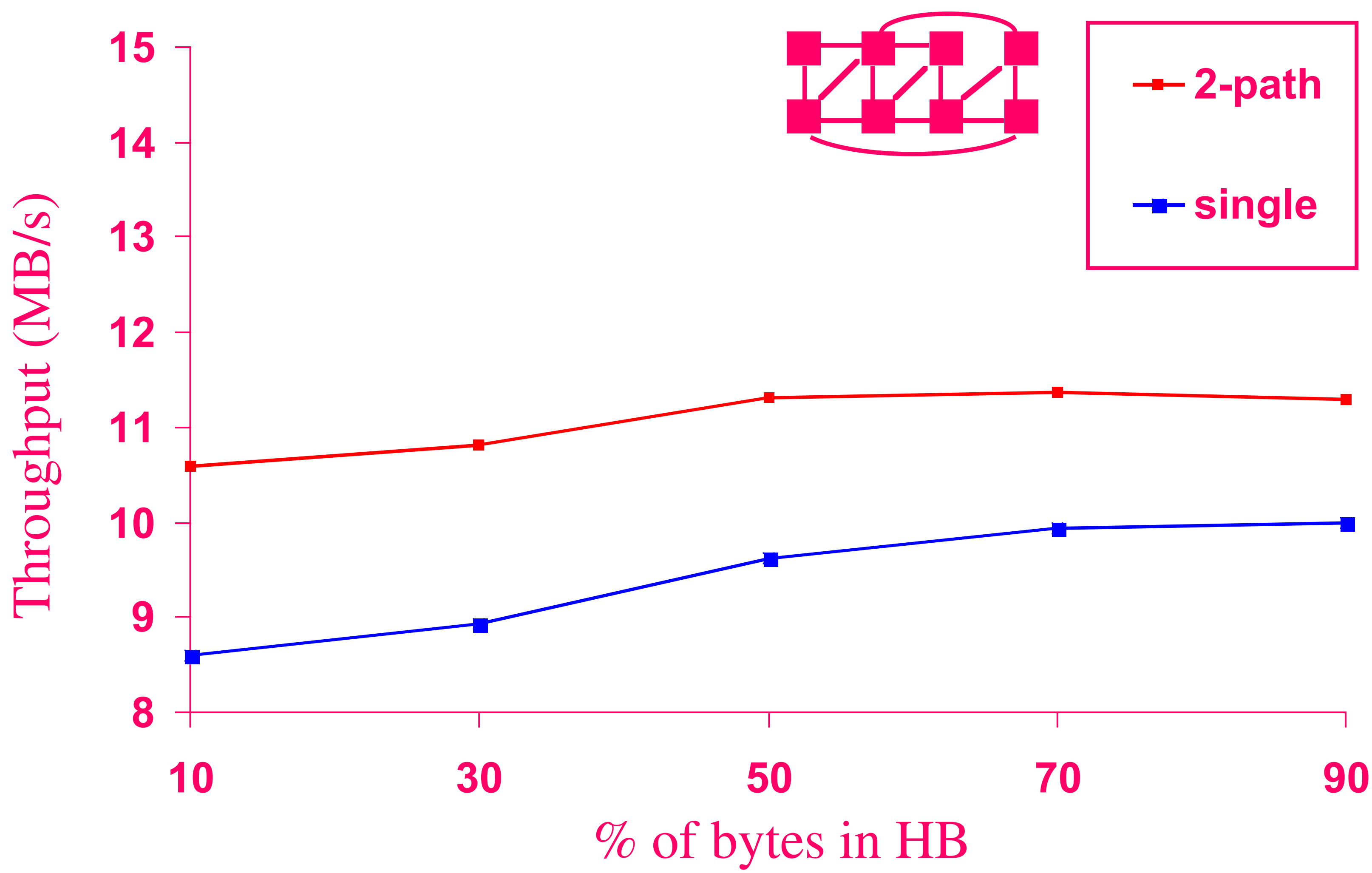


- No two disjoint paths from switch A to E

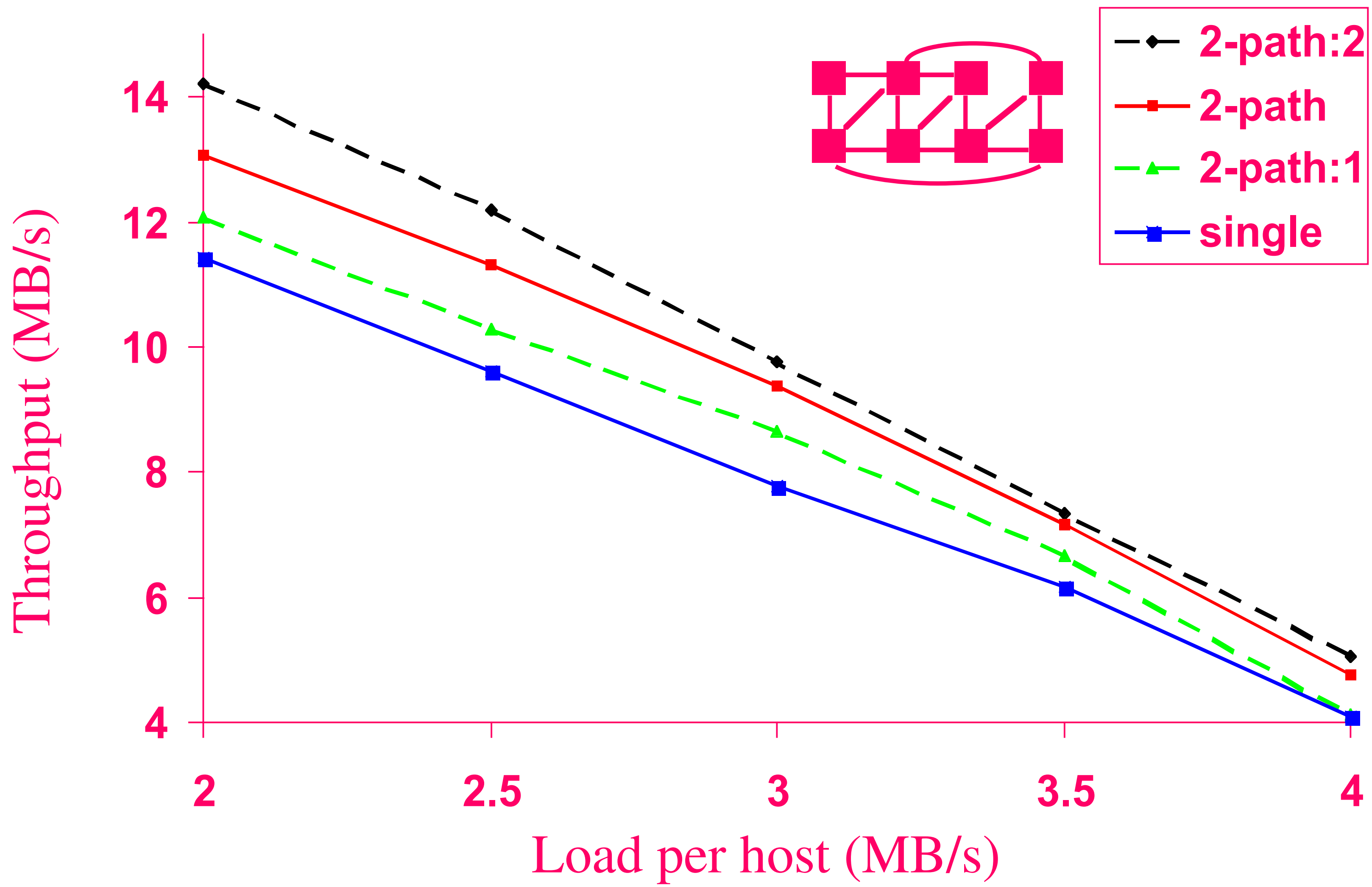
Multi-level prioritized max-min fairness

- Prioritizing multi-path---one path per priority
  - low-priority paths use leftover bandwidth from high-priority paths
  - paths of the same priority form a max-min plane
- Low overhead
  - source sets priority for each connection
  - switch does the same max-min calculation, using the bandwidth left over from higher priority plane
  - priority used in rate allocation, not in scheduling, no changes in scheduling and queueing

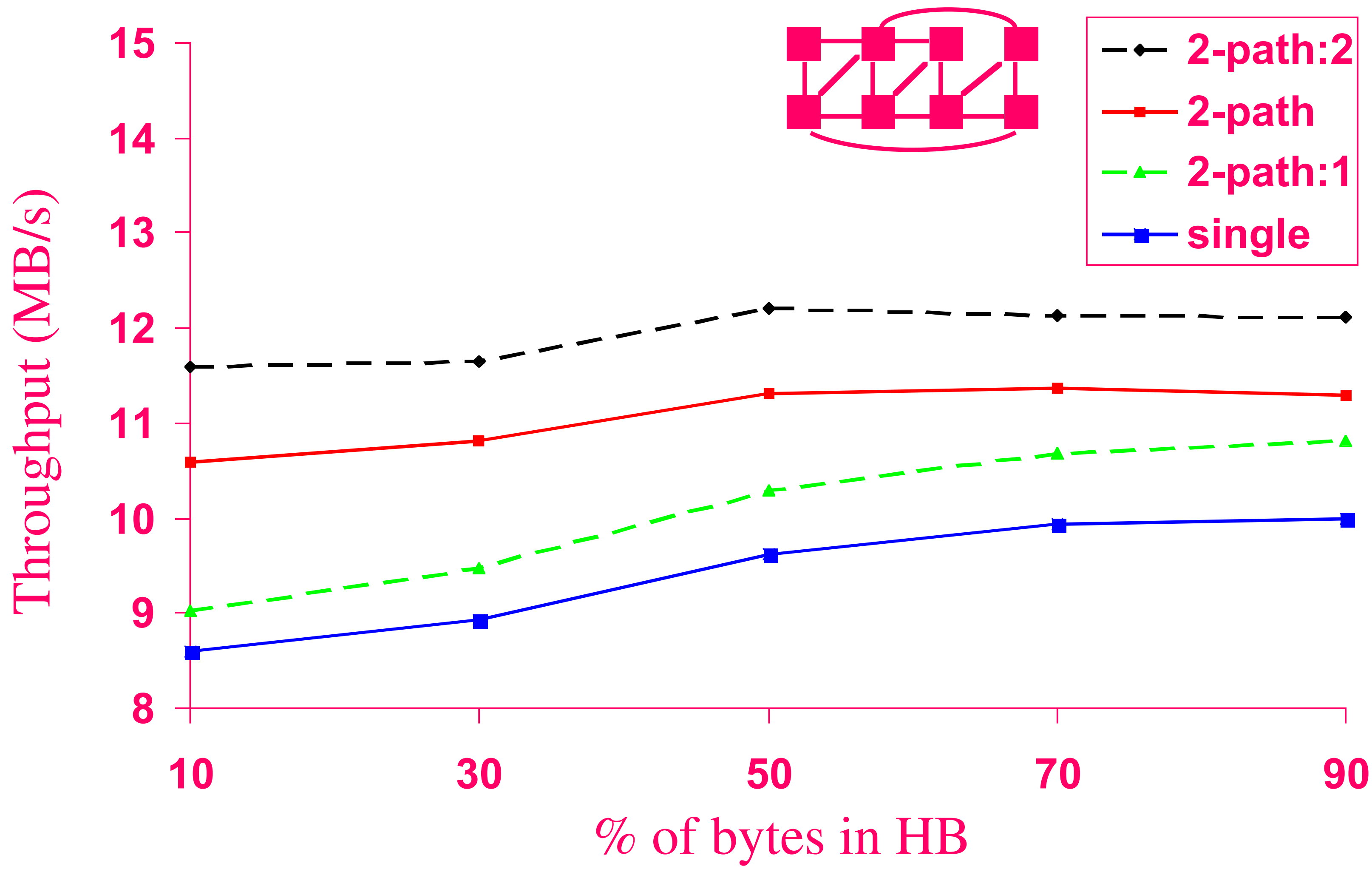
2-path routing---changing HB volume: G1



2-path routing---changing network load:G1



2-path routing---changing HB volume:G1



Sensitivity analysis

- Routing information update interval
  - widest path is more sensitive to out-of-date info.
- Peak rate for flows other than HB traffic
  - multi-path benefits more from low peak rate
- Routing cost
  - little impact
- Cutoff between HB and other traffic classes
  - little impact

Conclusions

- Routing support for HB traffic
  - improves both user satisfaction and resource efficiency
- Single path routing
  - uses rate information as link-state
  - a spectrum of link metrics: 1/r does consistently well
- Multi-path routing
  - significant throughput increase for 2-path users
  - throughput increase for others using single path only