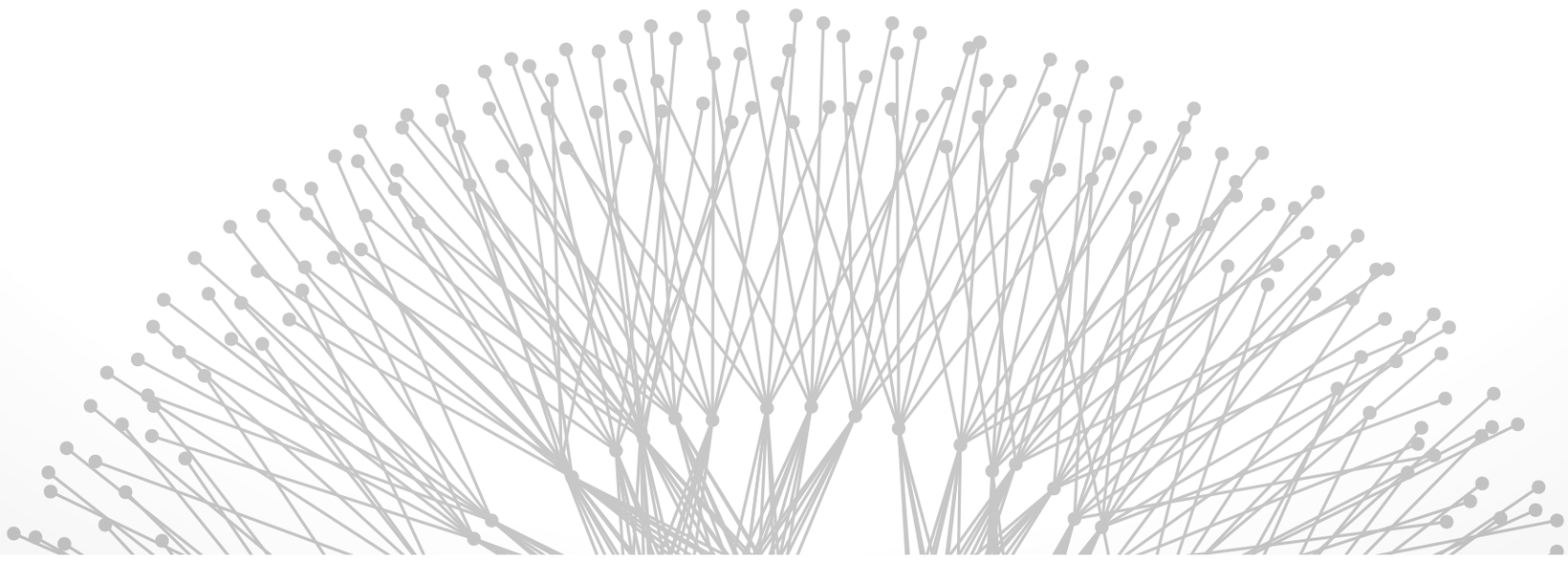
Intradomain Routing

Brighten Godfrey

CS 538 September 20 2012



Dealing with difficult readings

Readings can be difficult to understand

- It gets easier
- Ask questions!

Readings can be difficult to criticize in the reviews

- Goal is to think critically about the paper, not to write the definitive judgement of the work
- This is part of the process of understanding!

Routing

Choosing paths along which messages will travel from source to destination.

Often defined as the job of Layer 3 (IP). But...

- Ethernet spanning tree protocol (Layer 2)
- Distributed hash tables, content delivery overlays, ... (Layer 4+)

Problems for intradomain routing

Distributed path finding

Optimize link utilization (traffic engineering)

React to dynamics

High reliability even with failures

Scale

The two classic approaches

Distance Vector & Link State

Far from the only two approaches!

- We'll see more later...

Distance vector routing

Original ARPANET: distance vector routing

Remember vector of distances to each destination and exchange this vector with neighbors

- Initially: distance 0 from myself
- Upon receipt of vector: my distance to each destination = min of all my neighbors' distances + 1

Send packet to neighbor with lowest dist.

Slow convergence and looping problems

- E.g., consider case of disconnection from destination
- Fix for loops in BGP: store path instead of distance

Link state routing

Protocol variants

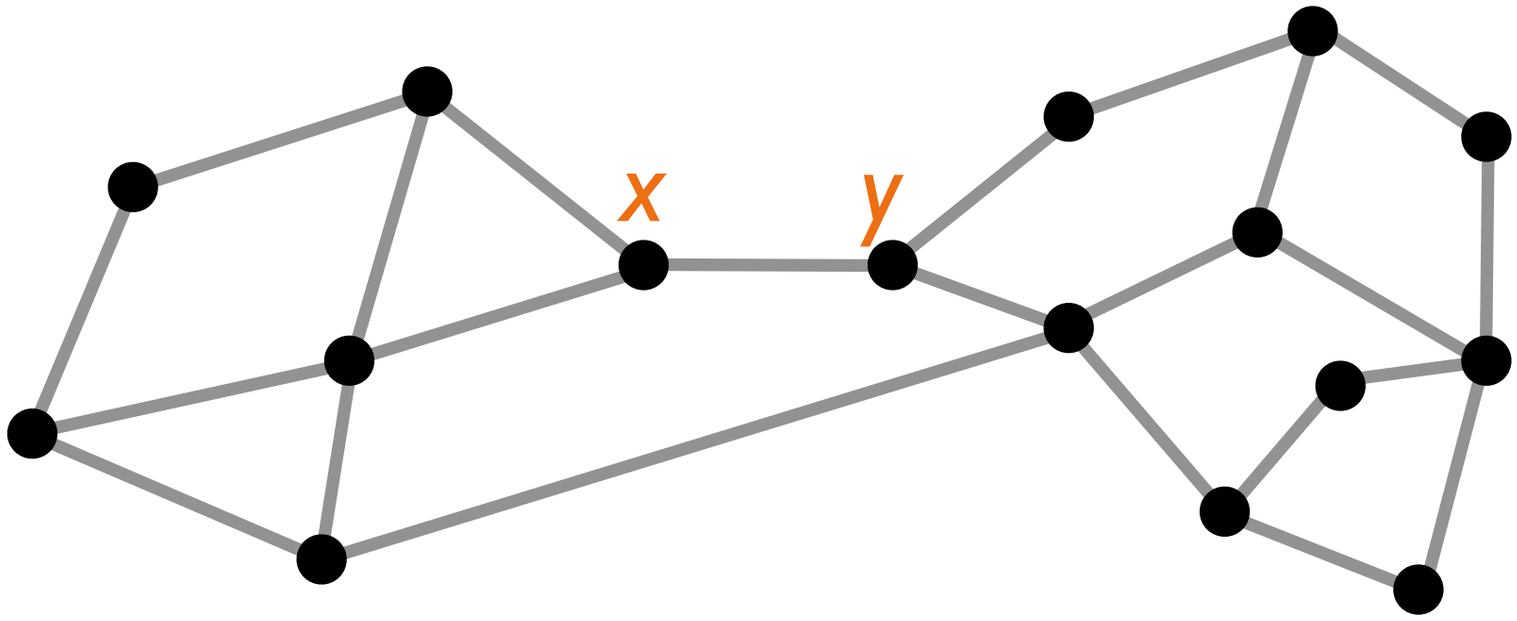
- ARPANET: McQuillan, Richer, Rosen 1980; Perlman 1983
- Intermediate System-to-Intermediate System (IS-IS)
- Open Shortest Path First (OSPF)

Algorithm

- Broadcast the entire topology to everyone
- Forwarding at each hop:
  - Compute shortest path (Dijkstra's algorithm)
  - Send packet to neighbor along computed path

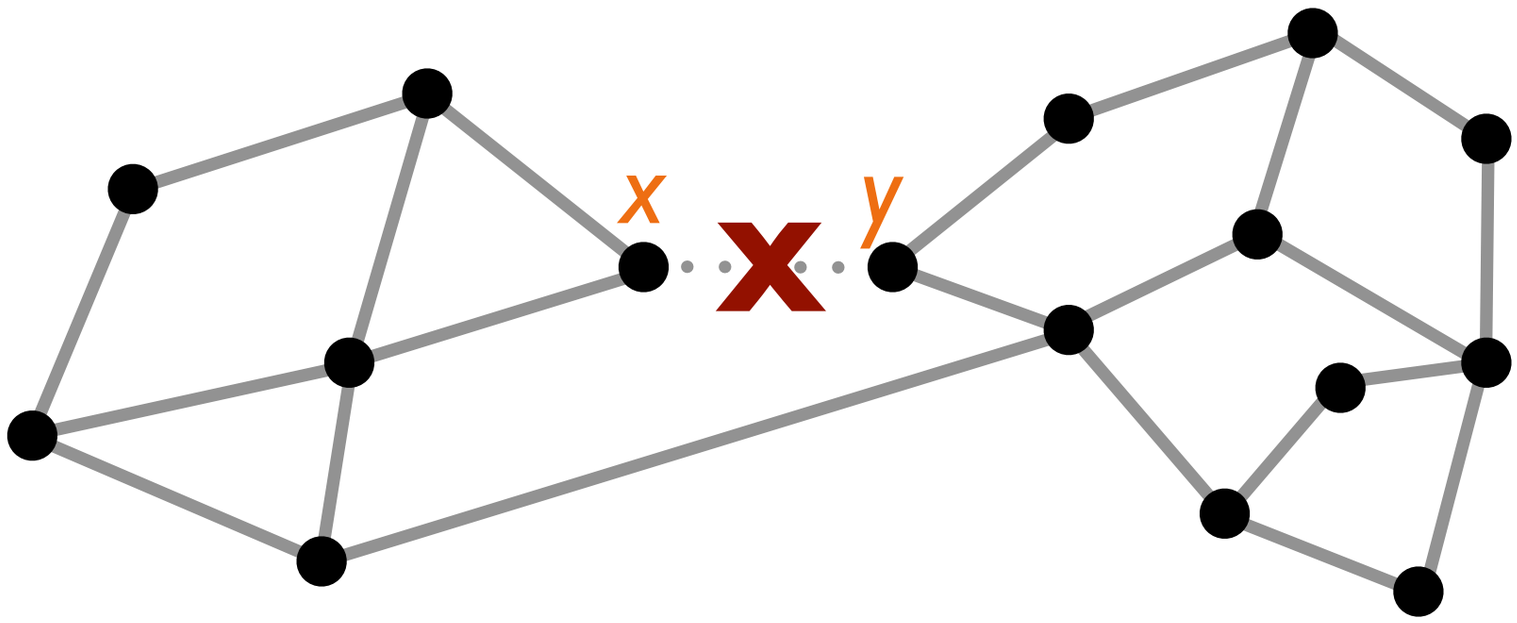
Question

We have a network...



Question

A link fails. How many messages does *x* send in immediate response?

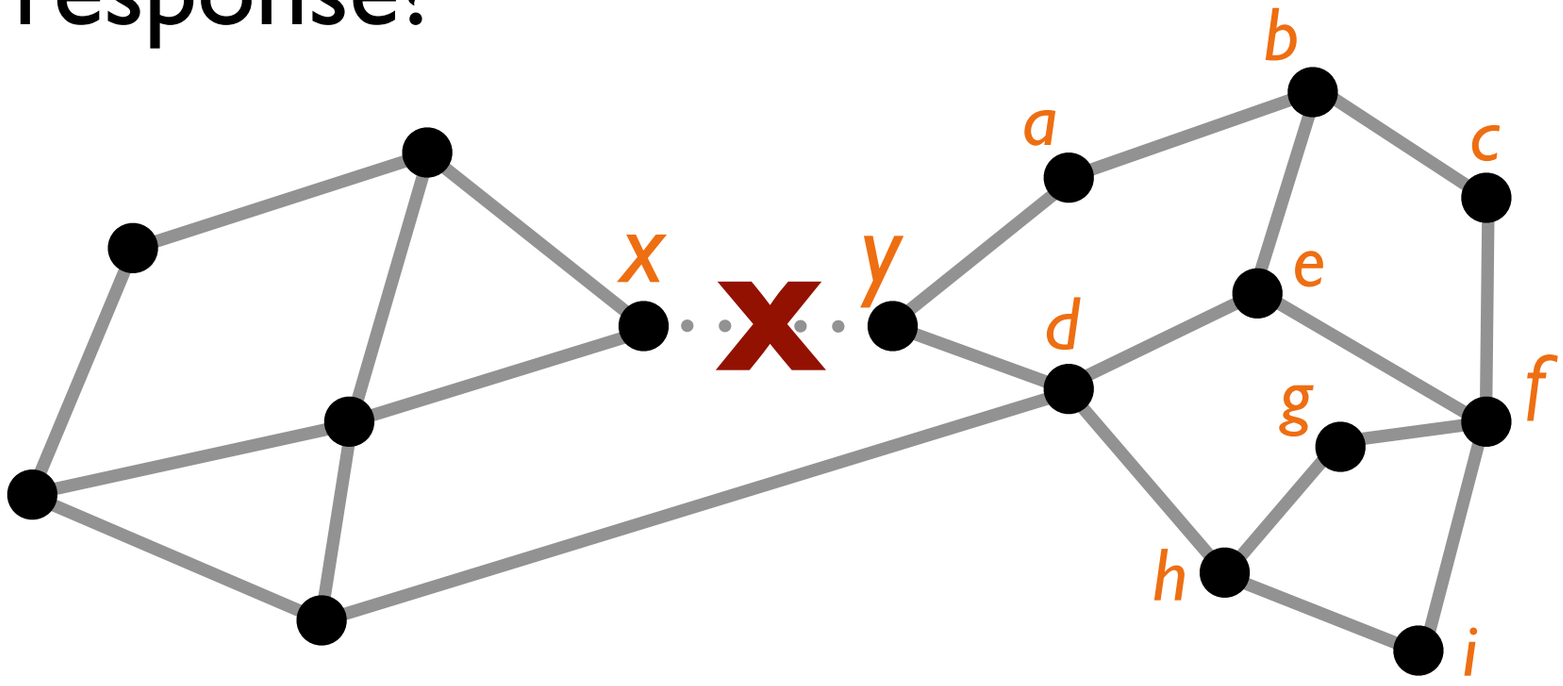


...using distance vector?

...using link state?

Question

A link fails. How many messages does *x* send in immediate response?



...using distance vector?

20

"My distance to *y* changed!
My distance to *a* changed!
My distance to *b* changed!
...
My distance to *i* changed!"
...to each of 2 neighbors

...using link state?

2

"Oh hey, link *x*-*y* failed"
...to each of 2 neighbors

Link state vs. distance vector

Disadvantages of LS

- Need consistent computation of shortest paths
  - Same view of topology
  - Same metric in computing routes
- Slightly more complicated protocol

Advantages of LS

- Faster convergence
- Gives unified global view
  - Useful for other purposes, e.g., building MPLS tables

Q: Can link state have forwarding loops?

LS variant: Source routing

Algorithm:

- Broadcast the entire topology to everyone
- Forwarding at source:
  - Compute shortest path (Dijkstra's algorithm)
  - Put path in packet header
- Forwarding at source and remaining hops:
  - Follow path specified by source

Q: Can this result in forwarding loops?

Source routing vs. link state

Advantages

- Essentially eliminates loops
- Compute route only once rather than every hop
- Forwarding table (FIB) size = #neighbors (not #nodes)
- Flexible computation of paths at source

Disadvantages

- Flexible computation of paths at source
- Header size (fixable if paths not too long)
  - Use local rather than global next-hop identifiers
  - $\log_2$(#neighbors) per hop rather than $\log_2$(#nodes)
- Source needs to know topology
- Harder to redirect packets in flight (to avoid a failure)

Traffic engineering

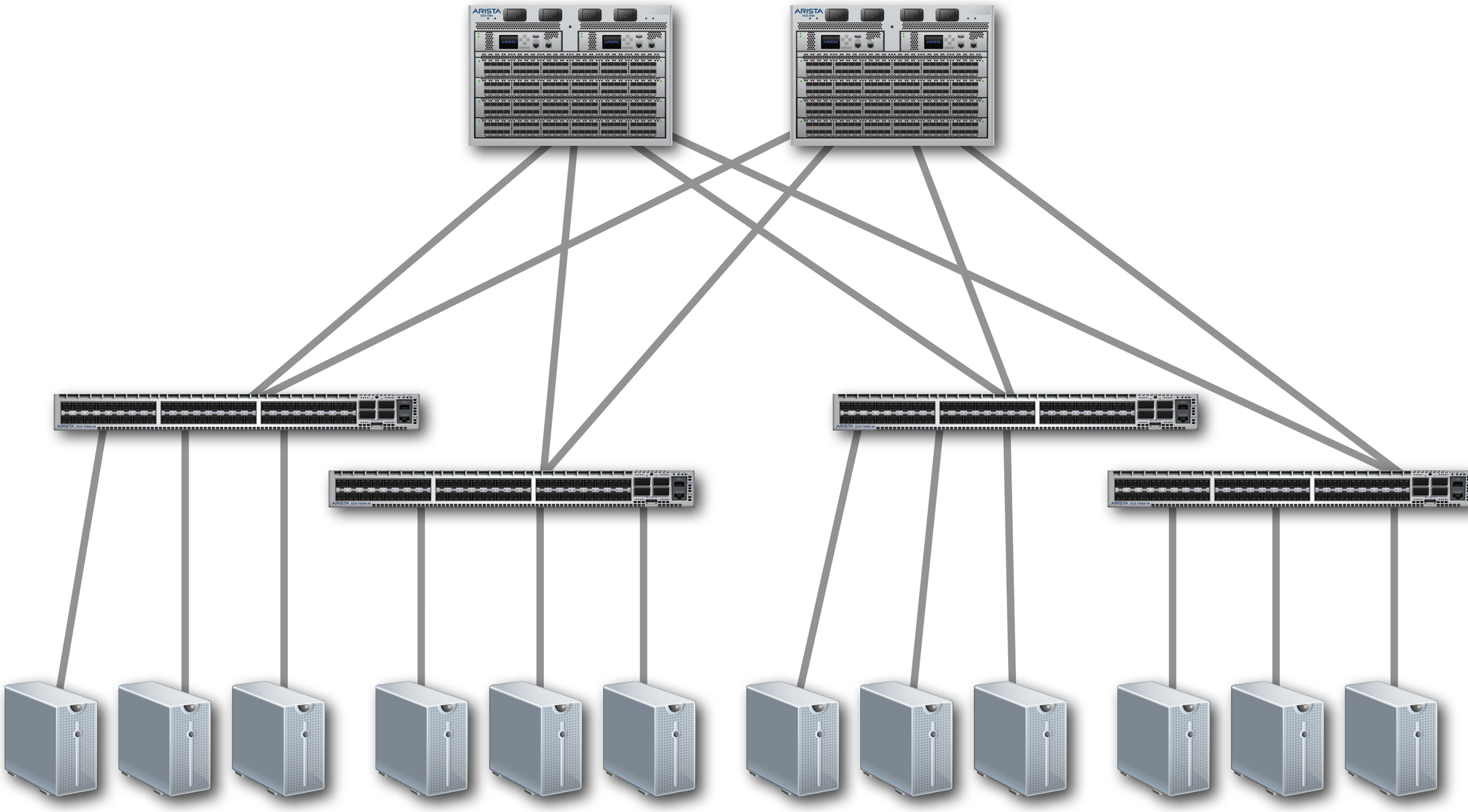
Key task of intradomain routing: optimize utilization

No TE: Shortest path routing

- How well does this work?

A start: Equal Cost Multipath Protocol (ECMP)

- Each router splits traffic across equally short next-hops
- Hash header to pin flow to a pseudorandom path (why?)
- When do you think this works well?


ARISTA
DCS-7050S-64

Traffic engineering

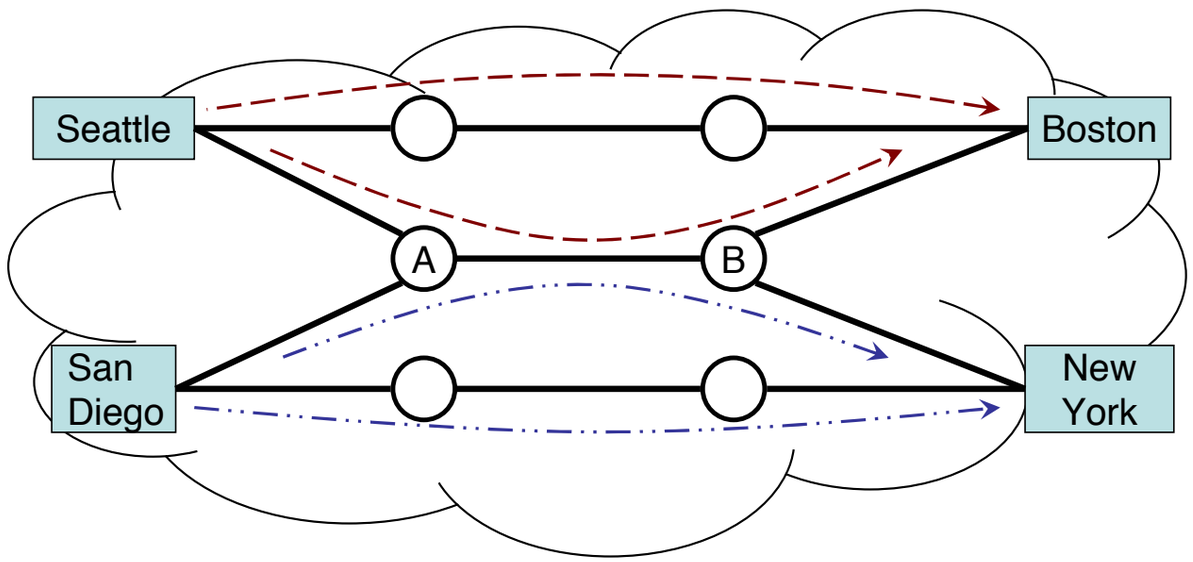
Key task of intradomain routing: optimize utilization

Classic TE: optimize OSPF weights

- Need to propagate everywhere: can't change often
- Single path to each destination

Modern TE: load balance among multiple MPLS paths

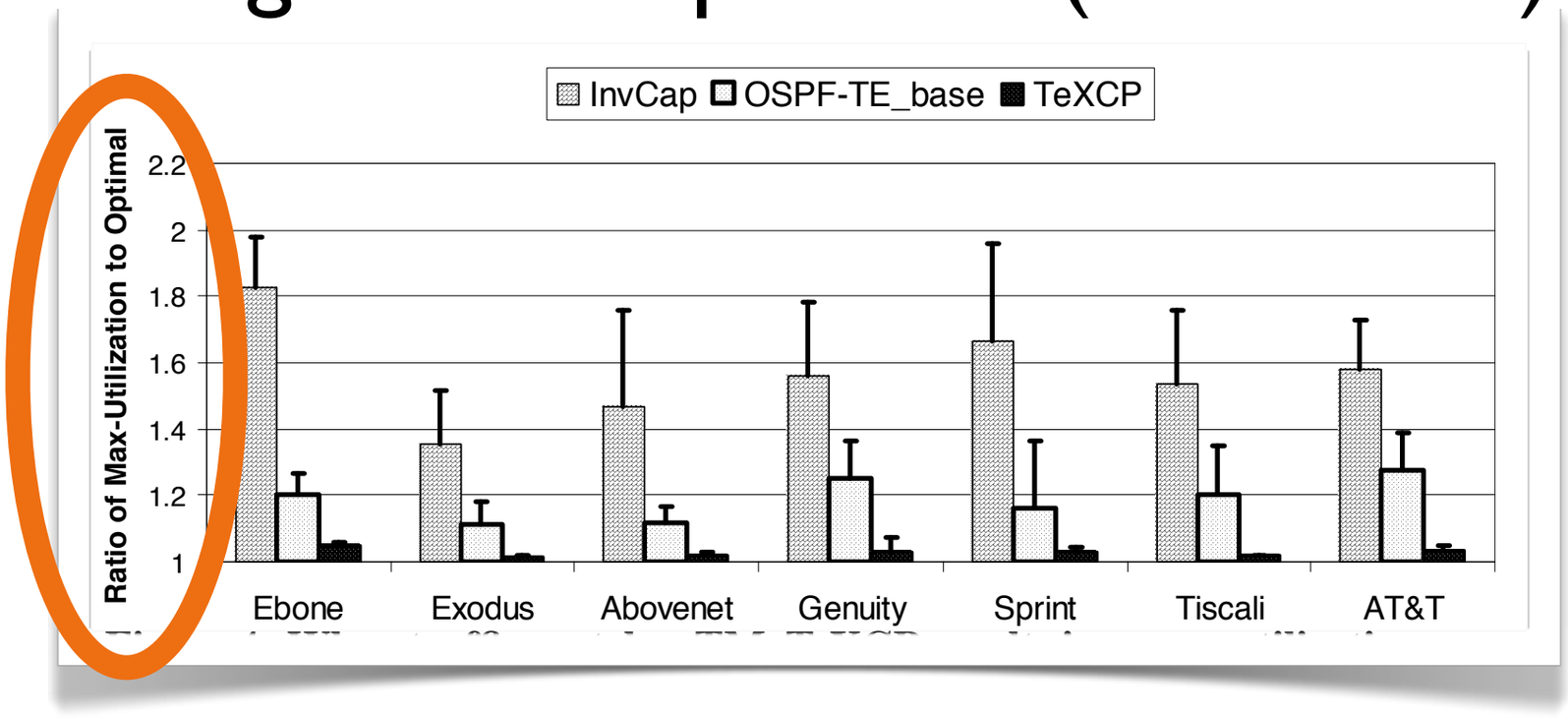
- e.g., TeXCP [Kandula, Katabi, Davie, Charny, 2005]



[Kandula et al, "Walking the Tightrope", SIGCOMM 2005]

TeXCP discussion

In OSPF-TE, "Finding optimal link weights that minimize the max-utilization is NP-hard". Why is this harder than finding the best possible (non-OSPF) solution?



Can TeXCP-style rate control be used end-to-end? [Carmen]

TeXCP discussion

Is minimizing max utilization what we are really looking for? [Pratik]

Improvement: allow incoming traffic to specify what type of metric it wishes to optimize [Uttam]

Announcements

What's to come

By tomorrow: presentation topic matching

Tuesday:

- interdomain routing basics
- project proposals due!

Upcoming meetings: advanced routing challenges

- scalability
- reliability
- selfishness
- security
- complexity

Project proposals

Project proposals due 11:59 p.m. Tuesday

- Submit via email to Brighten
- 1/2 page, plaintext or PDF

Format (see course syllabus):

- the problem you plan to address
- what will be your first steps
- related work
  - > 3 full academic references
  - why it has not addressed your problem
- if there are multiple people on your project team, who they are and how you plan to partition the work

BIG DATA & MACHINE LEARNING

AT NEUSTAR

Join Neustar Chief Technology Officer, **Dr. Mark Bregman,** as he discusses Big Data and Machine Learning at Neustar. He will also discuss how Data Analytics will help make the Internet more efficient, IP Geolocation more accurate, and marketing campaigns more profitable.

Location:
1404 Siebel Center

Time:
September 20th
6:00–7:30pm