Lecture 12:
WAN Routing Alternatives

CSE 222A: Computer Communication Networks
Alex C. Snoeren

Thanks: Brighten Godfrey
Lecture 12 Overview

- WAN Internet routing challenges
- Pathlet routing
- Intro to Traffic Engineering
Internet routing challenges

Multipath

reliability

path quality

Scalability

Policy
Internet routing challenges

- Multipath
  - reliability
  - path quality
- Scalability
- Policy

[ F. Wang, Z. M. Mao, J. Wang, L. Gao, R. Bush ’06]

Failure injected
Internet routing challenges

- Multipath
  - reliability
  - path quality
- Scalability
- Policy
- Lowest latency path
- Highest bandwidth path
- Path the network picked for you
Internet routing challenges

- Multipath
  - reliability
  - path quality
- Scalability
- Policy

Internet forwarding table size [Huston ’09]
Internet routing challenges

Multipath
  reliability
  path quality

Scalability

Policy
Pathlet routing

**vnode** virtual node

**pathlet** fragment of a path: a sequence of vnodes

Source routing over pathlets.

**virtual graph:** flexible way to define policy constraints

**provides many path choices for senders**
Flexibility

- can emulate BGP, source routing, MIRO, LISP, NIRA

- local transit policies provide multipath and small forwarding tables

- coexistence of different styles of routing policy
“Design for variation in outcome, so that the outcome can be different in different places, and the tussle takes place within the design, not by distorting or violating it.”

Clark, Wroclawski, Sollins & Braden, 2002

“Tussle in Cyberspace”
Pathlet routing

vnode  virtual node

pathlet fragment of a path:
a sequence of vnodes

Source routing over pathlets.
vnodes

vnodedefinition: virtual node within an AS

Walla Walla
New York
Crumstown
San Diego
Roosterville
Pathlets

Packet route field

Forwarding table

CSE 222A – Lecture 12: WAN Routing Alternatives
Pathlets

Packet route field

Forwarding table

| A | 3 |
| B | 7 |
| C | 2 |
| D | |

| 3 |
| 7,2 |
| 2 |

| ... | ... |
| ... | ... |
| 3 | push 7,2; fwd to B |
| ... | ... |
| 7 | fwd to C |
| ... | ... |
| 2 | fwd to D |

delivered!
Dissemination

- Global gossip fine, except for scalability
- So, let routers choose not to disseminate some pathlets
- Leads to (ironic) use of path vector — only for pathlet dissemination, not route selection
Local transit policies

Each ingress→egress pair is either allowed or disallowed.

Subject to this, any path allowed!

Represented with few pathlets: small FIB
"All valley-free" is local

“customers can route to anyone; anyone can route to customers”

Ingress from a provider

Ingress from a customer

Egress to a provider

Egress to a customer

Forwarding table size: 3 + #neighbors
Local transit policies provide some policy control for networks, while enabling a large number of paths for senders.
Emulating BGP

128.2.0.0/16

CSE 222A – Lecture 12: WAN Routing Alternatives
Mixed policies
Improved connectivity

BGP-style
Mixed
LT policies
Tiny forwarding tables

Forwarding table size CDF

Internet-Like

**BGP**

current Internet (CAIDA/APNIC):

132,158+ entries: one per IP prefix

pathlet routing, valley-free LT policies

2,264 entries, max
8.48 entries, mean
Control overhead

2.23x more messages,
1.61x more memory
in LT than PV

This can likely be improved.
Comparing protocols

- Feedback-based routing
- Loose source routing
- Strict source routing
- MIRO
- LISP
- BGP
- NIRA

Pathlet routing

Routing deflections, path splicing
Conclusion

- Pathlet routing: source routing over a virtual topology formed by pathlets and vnodes
- Highly flexible; supports both “local” policies with small forwarding tables and many paths, and complex BGP policies
- Challenges for source routing: Incentives to provide multiple paths; selecting paths; security; ...

Do IP Networks Manage Themselves?

In some sense, yes:
- TCP senders send less traffic during congestion
- Routing protocols adapt to topology changes

But, does the network run efficiently?
- Congested link when idle paths exist?
- High-delay path when a low-delay path exists?

How should routing adapt to the traffic?
- Avoiding congested links in the network
- Satisfying application requirements (e.g., delay)

... essential questions of traffic engineering
ARPAnet Routing (1969)

- **Routing**
  - Shortest-path routing based on link metrics; distance vector

- **Metrics**
  - *Instantaneous* queue length plus a constant
  - Each node updates distance computation periodically
Problems With the Algorithm

- Instantaneous queue length
  - Poor indicator of expected delay
  - Fluctuates widely, even at low traffic levels
  - Leading to routing oscillations

- Distance-vector routing
  - Transient loops during (slow) convergence
  - Triggered by link weight changes, not just failures

- Protocol overhead
  - Frequent dissemination of link metric changes
  - Leading to high overhead in larger topologies
New ARPAnet Routing (1979)

- Averaging of the link metric over time
  - Old: Instantaneous delay fluctuates a lot
  - New: Averaging reduces the fluctuations

- Link-state protocol
  - Old: Distance-vector path computation leads to loops
  - New: Link-state protocol where each router computes shortest paths based on the complete topology

- Reduce frequency of updates
  - Old: Sending updates on each change is too much
  - New: Send updates if change passes a threshold
Performance of New Algorithm

- Light load
  - Delay dominated by the constant part (transmission delay and propagation delay)

- Medium load
  - Queuing delay is no longer negligible on all links
  - Moderate traffic shifts to avoid congestion

- Heavy load
  - Very high metrics on congested links
  - Busy links look bad to all of the routers
  - All routers avoid the busy links
  - Routers may send packets on longer paths
Over-Reacting to Congestion

- Routers make decisions based on old information
  - Propagation delay in flooding link metrics
  - Thresholds applied to limit number of updates
- Old information leads to bad decisions
  - All routers avoid the congested links
  - … leading to congestion on other links
  - … and the whole things repeats

"SigAlert on the 5" on radio triggers back-up in Del Mar
Newer ARPAnet Metric (1987)

- Prevent over-reacting
  - Shed traffic from a congested link gradually
  - Starting with alternate paths that are just slightly longer
  - Through weighted average in computing the metric, and limits on the change from one period to the next

- Limit path length
  - Bound the value of the link metric
  - “This link is busy enough to go two extra hops”

- New algorithm
  - New way of computing the link weights
  - No change to link-state routing or shortest-path algorithm
Today: “Static” Link Weights

- Routers flood information to learn topology
  - Determine “next hop” to reach other routers…
  - Compute shortest paths based on link weights
- Link weights configured by network operator
Setting the Link Weights

How to set the weights
- Inversely proportional to link capacity?
- Proportional to propagation delay?
- Network-wide optimization based on traffic?
Measure, Model, and Control

Network-wide “what if” model

Topology/Configuration

Offered traffic

Changes to the network

Operational network

measure

control

CSE 222A – Lecture 12: WAN Routing Alternatives
For Next Class…

- Read the Entact paper
- Keep moving on term projects!