Wide-Area Traffic Management

COS 597E: Software Defined Networking

Jennifer Rexford
Princeton University

MW 11:00am-12:20pm

Traffic Management

- Assigning resources to traffic
  - Optimize some objective
    - Min congestion, max utility, min delay, …
  - Given network resource constraints
- Three main "knobs"
  - Routing: what path(s) the traffic takes
  - Link scheduling: how to share each link
  - Rate control: how much a source can send
- Host/network split
  - Host: rate control
  - Network: routing and link scheduling

2

Simple Traffic Management

- Protocols adapt automatically
  - 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)

3

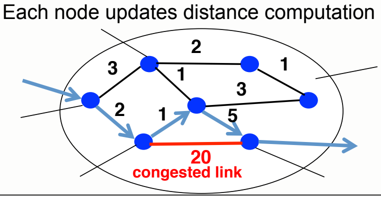
Automatically Adapting the Link Weights

ARPAnet Routing

4

Original ARPAnet Routing (1969)

- Routing
  - Shortest-path routing based on link metrics
  - *Distance-vector* algorithm (i.e., Bellman-Ford)
- Link metrics
  - *Instantaneous* queue length plus a constant
  - Each node updates distance computation



5

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

6

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 computation leads to loops
  - New: Link-state protocol where each router computes paths based on the complete topology
- Reduce frequency of updates
  - Old: Too many update messages
  - New: Send updates if change passes a threshold

7

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
  - Routers may send packets on *longer* paths

8

Revised ARPAnet Metric (1987)

- Limit path length
  - Bound the value of the link metric
  - "This link is busy enough to go two extra hops"
- Prevent over-reacting
  - Shed traffic from a congested link gradually
  - Starting with alternate paths that are *slightly* longer
  - Through weighted average in computing the metric, and limits on the change from one period to the next
- New algorithm
  - New way of computing the link weights
  - No change to routing protocol or path computation

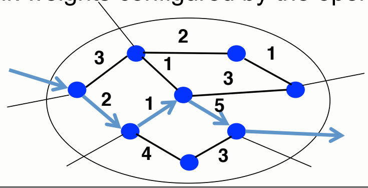
9

Optimizing the "Static" Link Weights

10

Routing With "Static" Link Weights

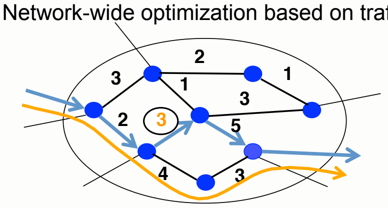
- Routers flood information to learn topology
  - Determine "next hop" to reach other routers…
  - Compute shortest paths based on link weights
- Link weights configured by the operator



11

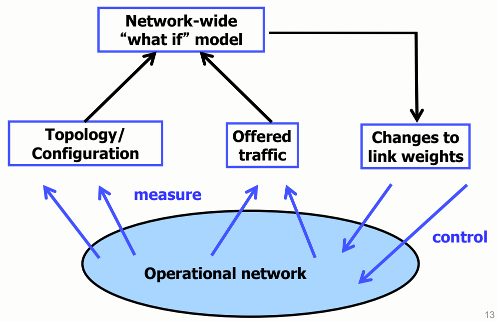
Setting the Link Weights

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



12

Measure, Model, Control



13

Pros and Cons

- Advantages
  - Network-wide optimization
  - Avoids oscillation
  - No changes to the routing protocols
- Disadvantages
  - Overhead of collecting the measurements
  - Limited splitting of traffic over multiple paths
  - Computational complexity of the optimization
  - Transient disruptions during weight changes
- So, performed at a slow time scale (hours)

14

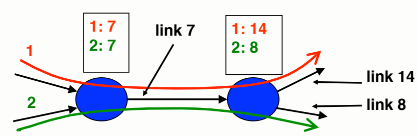
MPLS-TE

MultiProtocol Label Switching

15

Explicit End-to-End Paths

- Establish end-to-end path in advance
  - Learn the topology (as in link-state routing)
  - End host or router computes and signals a path
- Routers supports virtual circuits
  - Signaling: install entry for each circuit at each hop
  - Forwarding: look up the circuit id in the table



16

MPLS-TE

- Learn about congestion
  - Dynamically changing link weights
- Reserve resources on paths
  - Pick a path, and signal to reserve resources
- Change paths during congestion
  - Pick a new path, and reserve resources
- More flexible, but still some limitations
  - Uncoordinated decisions at different nodes
  - Suboptimal decisions, and non-deterministic
  - Complex interaction of several protocols

17