18-345 – Fall 08

Lecture 17

Network Layer: Packet Switching

Peter Steenkiste

Readings: Chapter 7

Overview Network Layer

- Network layer
- Packet network topology
- Distance vector routing
- Link state routing
- Internet addressing
- Internet protocol
- Routing in the Internet
- Putting things together

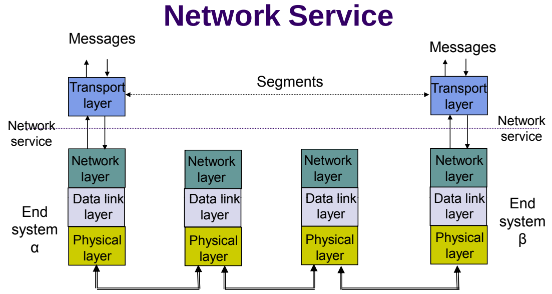
Network Layer Overview

Reading: Sections 7.1 and 7.2

Network Layer: Challenges

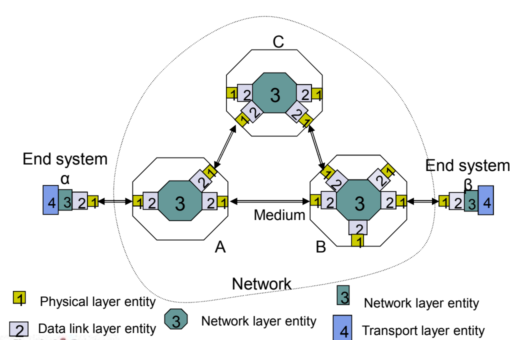
- Requires the coordinated actions of multiple, geographically distributed network elements
  - Networks, routers, …
- Must be able to deal with very large scales
  - People & communicating devices
- Competing providers need to cooperate
  - ISPs are competing for business
- Biggest Challenges
  - Addressing: identify destinations
  - Routing: what path should be used?

Network Service



- Network layer can offer a variety of services to transport layer
  - Connection-oriented service or connectionless service
  - Best-effort versus delay and loss guarantees

Complexity at the Edge or in the Core?



The End-to-End Argument for System Design

- An end-to-end function is best implemented at a higher level than at a lower level
  - End-to-end service requires all intermediate components to work properly
  - Higher-level better positioned to ensure correct operation
- Example: stream transfer service
  - Establishing an explicit connection for each stream across the network requires all network elements (NEs) to be aware of connection;
  - All NEs have to be involved in re-establishment of connections in case of network fault
  - In connectionless network operation, NEs do not deal with each explicit connection and hence are much simpler in design

➔ IP provides simple connectionless service without guarantees

Network Layer Functions

Essential

- **Routing**: mechanisms for determining the set of best paths for routing packets requires the collaboration of network elements
- **Forwarding**: transfer of packets from network element (NE) inputs to outputs
- **Priority & Scheduling**: determining order of packet transmission in each NE

Optional

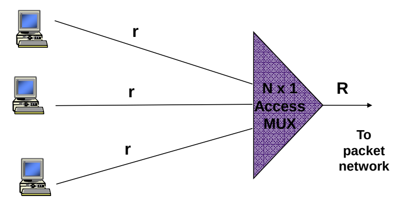
Congestion control, segmentation & reassembly, security

End-to-End Packet Network

- Individual packet streams are highly bursty
  - Statistical multiplexing is used to concentrate streams
- User demand can undergo dramatic change
  - Peer-to-peer applications stimulated huge growth in traffic volumes
- Internet structure highly decentralized
  - Paths traversed by packets can go through many networks controlled by different organizations
  - No single entity responsible for end-to-end service

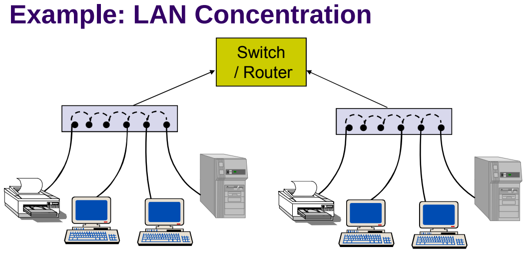
➔ Statistical multiplexing at many levels

Access Multiplexing



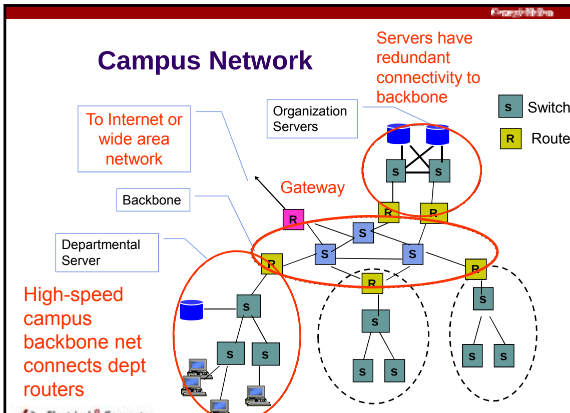
- Packet traffic from users multiplexed at access to network into aggregated streams
- Statistical multiplexing allows oversubscription: R << X x r
- Many examples: LAN switches, Cable Modem Termination System, DSL Access Mux, …

Example: LAN Concentration

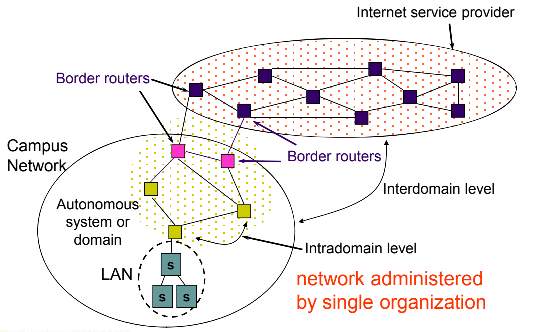


- LAN Hubs and switches in the access network also aggregate packet streams that flows into switches and routers

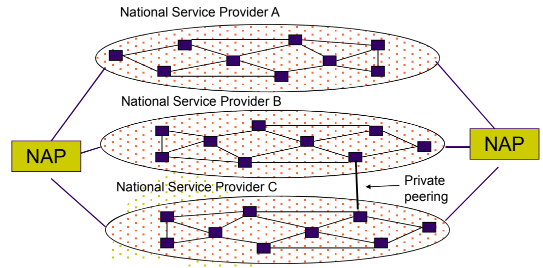
Campus Network



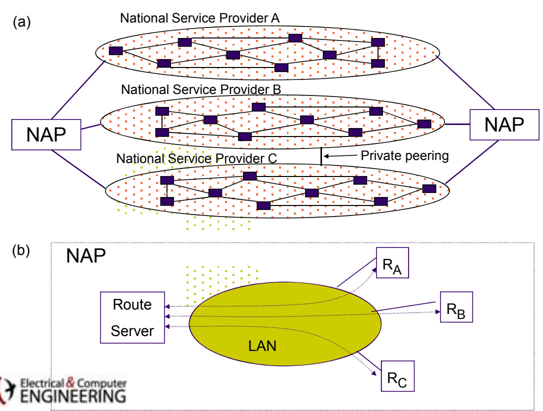
Connecting to Internet Service Provider



Internet Backbone



- Network Access Points: set up during original commercialization of Internet to facilitate exchange of traffic
- Private Peering Points: two-party inter-ISP agreements to exchange traffic



Key Role of Routing

How to get packet from here to there?

- Decentralized nature of Internet makes routing a major challenge
  - Interior gateway protocols (IGPs) are used to determine routes within a domain
  - Exterior gateway protocols (EGPs) are used to determine routes across domains
  - Routes must be consistent & produce stable flows
- Scalability required to accommodate growth
  - Hierarchical structure of IP addresses essential to keeping size of routing tables manageable

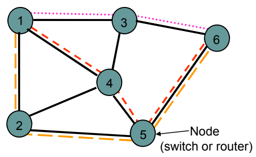
Two Levels of Routing

- Routing inside a autonomous system
  - Focus is typically on optimizing efficiency
- Routing across autonomous systems
  - Focus is on policy and business relationships
  - Am I being paid to do this?

Routing in Packet Networks

Readings: Sections 7.4 and 7.5

Routing in Packet Networks



- Three possible (loop-free) routes from 1 to 6:
  - 1-3-6, 1-4-5-6, 1-2-5-6
- What does "best" mean?
  - Min delay? Min hop? Max bandwidth? Min cost? Max reliability?

Electrical & Computer ENGINEERING

Routing Protocol: Functions

- Are responsible for creating the forwarding table in each router
  - Control function: longer time scale
- Need to distributed information on state of links
  - What information: Link up/down; congested; delay; etc.
  - How often is information exchanged?
  - Exchange with neighbors; Broadcast or flood
- Use a routing algorithm to computer routes
  - Results are used to create or update the forwarding table
  - Single metric; multiple metrics
  - Single route; alternate routes

Electrical & Computer ENGINEERING

Routing Algorithm Requirements

- Responsiveness to changes
  - Topology or bandwidth changes, congestion
  - Rapid convergence of routers to consistent set of routes
  - Freedom from persistent loops
- Optimality
  - Resource utilization, path length
- Robustness
  - Continues working under high load, congestion, faults, equipment failures, incorrect implementations
- Simplicity
  - Efficient software implementation, reasonable processing load

Electrical & Computer ENGINEERING

Centralized vs Distributed Routing

- Centralized Routing
  - All routes determined by a central node
  - All state information sent to central node
  - Problems adapting to frequent topology changes
  - Does not scale
- Distributed Routing
  - Routes determined by routers using distributed algorithm
  - State information exchanged by routers
  - Adapts to topology and other changes
  - Better scalability

Electrical & Computer ENGINEERING

Static vs Dynamic Routing

- Static Routing
  - Set up manually, do not change; requires administration
  - Works when traffic predictable & network is simple
  - Used to override some routes set by dynamic algorithm
  - Used to provide default router
- Dynamic Routing
  - Adapt to changes in network conditions
  - Automated
  - Calculates routes based on received updated network state information

Electrical & Computer ENGINEERING

Routing Metrics – Shortest Path

Means for measuring desirability of a path

- Path Length = sum of link costs
- Possible metrics
  - Hop count: rough measure of resources used
  - Reliability: link availability; BER
  - Delay: sum of delays along path; complex & dynamic
  - Bandwidth: "available capacity" in a path
  - Load: Link & router utilization along path
  - Cost: $$$

Electrical & Computer ENGINEERING

Shortest Path Routing Approaches

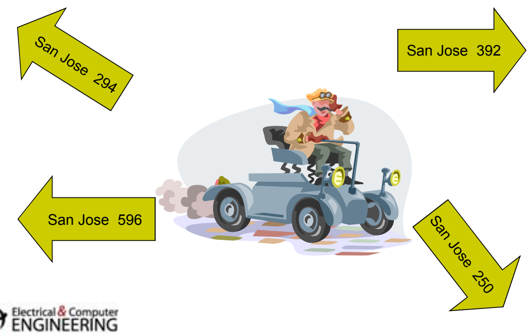
Distance Vector Protocols

- Neighbors exchange list of distances to destinations
- Best next-hop determined for each destination
- Ford-Fulkerson (distributed) shortest path algorithm

Link State Protocols (next lecture)

- Link state information (link up/down?) flooded to all routers
- Routers have complete topology information
- Shortest path (& hence next hop) calculated
- Dijkstra (centralized) shortest path algorithm

Distance Vector
Do you know the way to San Jose?



Distance Vector

Local Signpost

- Direction
- Distance

Routing Table

For each destination list:

- Next Node
- Distance

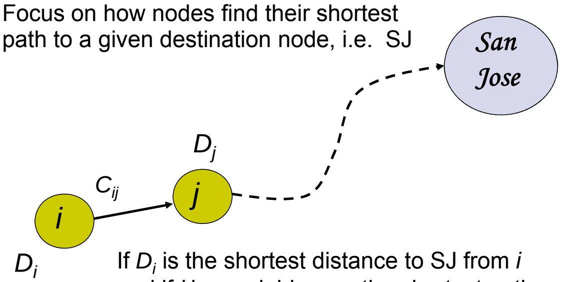
dest	next	dist

Table Synthesis

- Neighbors exchange table entries
- Determine current best next hop
- Inform neighbors
  - Periodically
  - After changes

Shortest Path to SJ

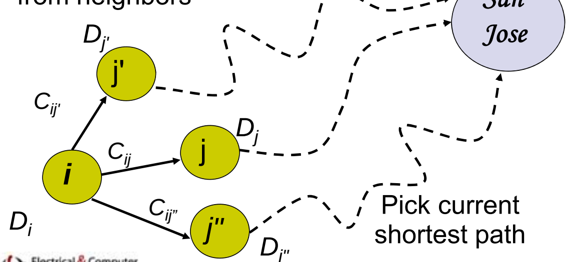
Focus on how nodes find their shortest path to a given destination node, i.e. SJ



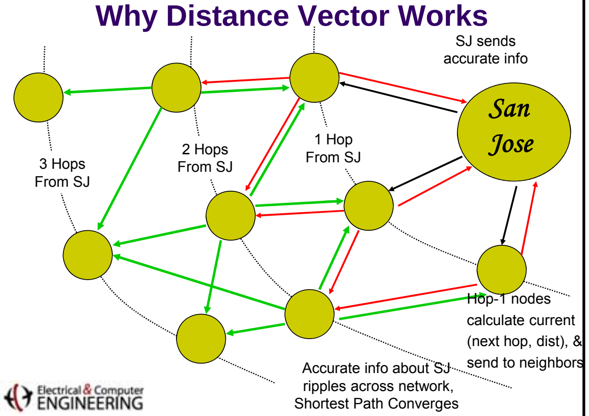
If D_i is the shortest distance to SJ from *i* and if *j* is a neighbor on the shortest path, then $D_i = C_{ij} + D_j$

But we don't know the shortest paths

i only has local info from neighbors



Why Distance Vector Works



Bellman-Ford Algorithm

- *Consider computations for one destination d*
- *Initialization*
  - Each node table has 1 row for destination *d*
  - Distance of node *d* to itself is zero: $D_d=0$
  - Distance of other node *j* to *d* is infinite: $D_j=\infty$, for $j \neq d$
  - Next hop node n_j = -1 to indicate not yet defined for $j \neq d$
- *Send Step*
  - Send new distance vector to immediate neighbors across local link
- *Receive Step*
  - At node *j*, find the next hop that gives the minimum distance to *d*,
    - $Min_j \{ C_{ij} + D_j \}$
    - Replace old $(n_j, D_j(d))$ by new $(n_j^*, D_j^*(d))$ if new next node or distance
  - Go to send step

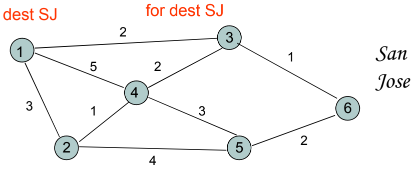
Bellman-Ford Algorithm

- *Now consider parallel computations for all destinations d*
- *Initialization*
  - Each node has 1 row for each destination *d*
  - Distance of node *d* to itself is zero: $D_d(d)=0$
  - Distance of other node *j* to *d* is infinite: $D_j(d)=\infty$, for $j \neq d$
  - Next node n_j = -1 since not yet defined
- *Send Step*
  - Send new distance vector to immediate neighbors across local link
- *Receive Step*
  - For each destination *d*, find the next hop that gives the minimum distance to *d*,
    - $Min_j \{ C_{ij} + D_j(d) \}$
    - Replace old $(n_j, D_i(d))$ by new $(n_j^*, D_j^*(d))$ if new next node or distance found
  - Go to send step

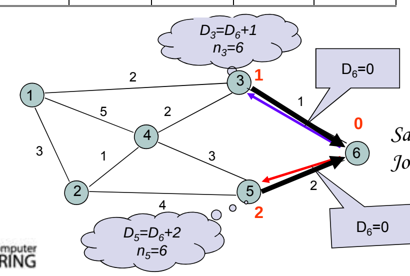
Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)
1					
2					
3					

Table entry @ node 1 for dest SJ

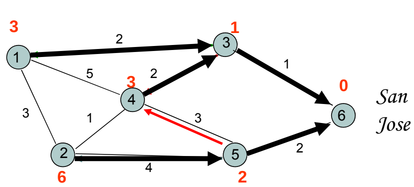
Table entry @ node 3 for dest SJ



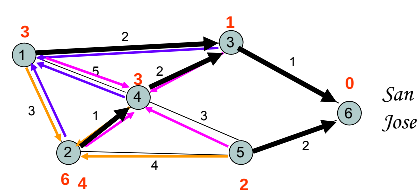
Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)
1	(-1, ∞)	(-1, ∞)	(6,1)	(-1, ∞)	(6,2)
2					
3					



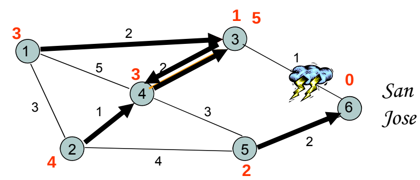
Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)
1	(-1, ∞)	(-1, ∞)	(6, 1)	(-1, ∞)	(6,2)
2	(3,3)	(5,6)	(6, 1)	(3,3)	(6,2)
3					



Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)	(-1, ∞)
1	(-1, ∞)	(-1, ∞)	(6, 1)	(-1, ∞)	(6,2)
2	(3,3)	(5,6)	(6, 1)	(3,3)	(6,2)
3	(3,3)	(4,4)	(6, 1)	(3,3)	(6,2)

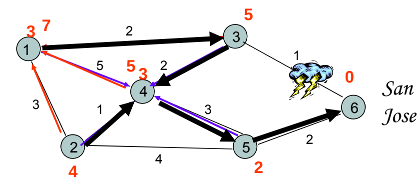


Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(3,3)	(4,4)	(6, 1)	(3,3)	(6,2)
1	(3,3)	(4,4)	(4, 5)	(3,3)	(6,2)
2					
3					



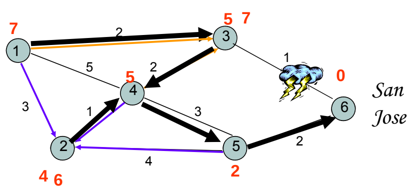
Network disconnected; Loop created between nodes 3 and 4

Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(3,3)	(4,4)	(6, 1)	(3,3)	(6,2)
1	(3,3)	(4,4)	(4, 5)	(3,3)	(6,2)
2	(3,7)	(4,4)	(4, 5)	(5,5)	(6,2)
3					



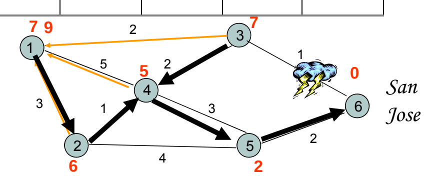
Node 4 could have chosen 2 as next node because of tie

Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
Initial	(3,3)	(4,4)	(6, 1)	(3,3)	(6,2)
1	(3,3)	(4,4)	(4, 5)	(3,3)	(6,2)
2	(3,7)	(4,4)	(4, 5)	(5,5)	(6,2)
3	(3,7)	(4,6)	(4, 7)	(5,5)	(6,2)



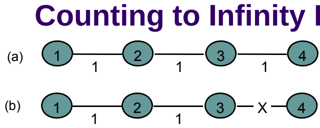
Node 2 could have chosen 5 as next node because of tie

Iteration	Node 1	Node 2	Node 3	Node 4	Node 5
1	(3,3)	(4,4)	(4, 5)	(3,3)	(6,2)
2	(3,7)	(4,4)	(4, 5)	(2,5)	(6,2)
3	(3,7)	(4,6)	(4, 7)	(5,5)	(6,2)
4	(2,9)	(4,6)	(4, 7)	(5,5)	(6,2)



Node 1 could have chose 3 as next node because of tie

Counting to Infinity Problem



Nodes believe best path is through each other

(Destination is node 4)

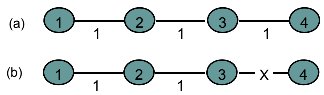
Update	Node 1	Node 2	Node 3
Before break	(2,3)	(3,2)	(4, 1)
After break	(2,3)	(3,2)	(2,3)
1	(2,3)	(3,4)	(2,3)
2	(2,5)	(3,4)	(2,5)
3	(2,5)	(3,6)	(2,5)
4	(2,7)	(3,6)	(2,7)
5	(2,7)	(3,8)	(2,7)
…	…	…	…

Problem: Bad News Travels Slowly

Remedies

- Split Horizon
  - Do not report route to a destination to the neighbor from which route was learned
- Poisoned Reverse
  - Report route to a destination to the neighbor from which route was learned, but with infinite distance
  - Breaks erroneous direct loops immediately
  - Does not work on some indirect loops

Split Horizon with Poison Reverse



Nodes believe best path is through each other

Update	Node 1	Node 2	Node 3	
Before break	(2, 3)	(3, 2)	(4, 1)	
After break	(2, 3)	(3, 2)	(-1, ∞)	Node 2 advertizes its route to 4 to node 3 as having distance infinity; node 3 finds there is no route to 4
1	(2, 3)	(-1, ∞)	(-1, ∞)	Node 1 advertizes its route to 4 to node 2 as having distance infinity; node 2 finds there is no route to 4
2	(-1, ∞)	(-1, ∞)	(-1, ∞)	Node 1 finds there is no route to 4

Electrical & Computer ENGINEERING