ECSE-6660
Traffic Engineering

http://www.pde.rpi.edu/

Or

http://www.ecse.rpi.edu/Homepages/shivkuma/

Shivkumar Kalyanaraman

Rensselaer Polytechnic Institute

shivkuma@ecse.rpi.edu

- Introductions:course description & calendar
- Answers to frequently asked questions
- Prerequisites
- Informal Quiz

Without Traffic Engineering

Cars:

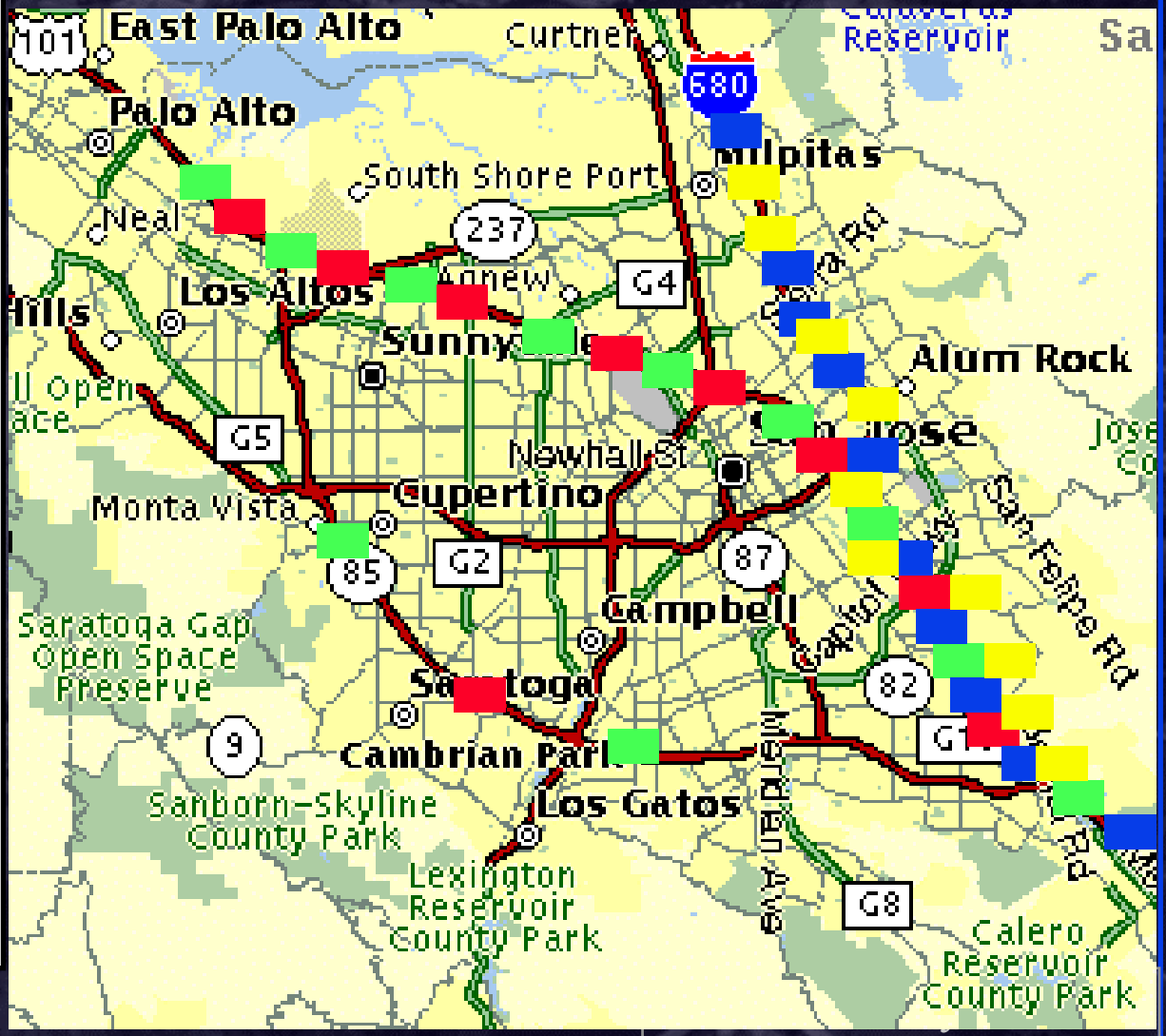
SFO-LAX

SAN-SMF

LAX-SFO

SMF-SAN

No Traffic Engineering analogy to Human Drivers

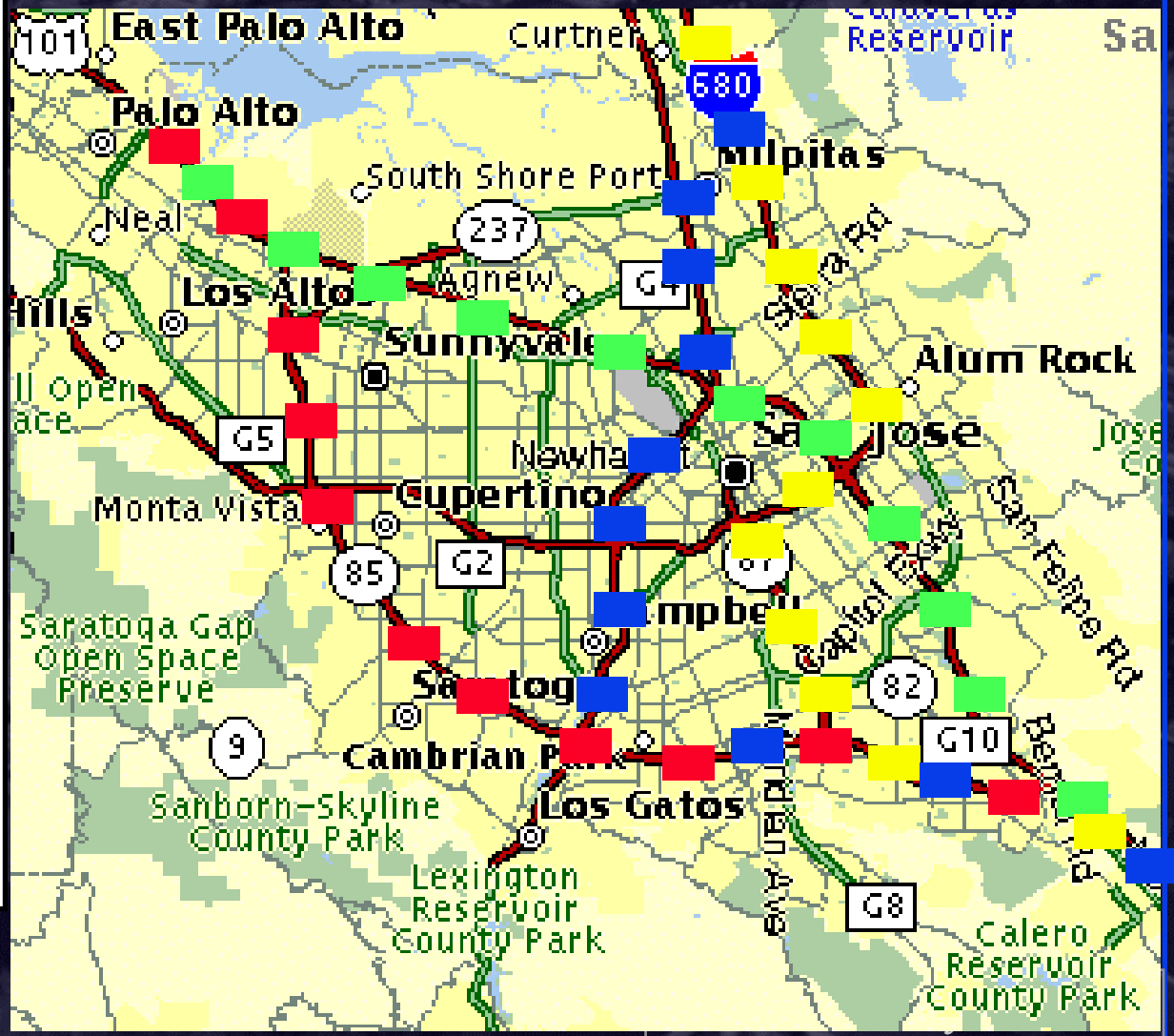


Traffic Engineering: Analogy

Cars:

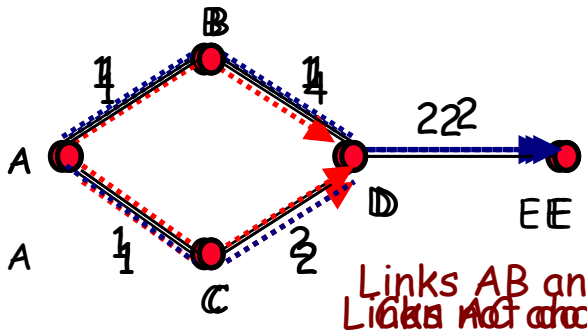
- SFO-LAX
- LAX-SFO
- SAN-SMF
- SMF-SAN

Traffic Engineering analogy



Motivation

- *TE: "…that aspect of Internet network engineering dealing with the issue of performance evaluation and performance optimization of operational IP networks …"*
- 90's approach to TE was by changing link weights in IGP (OSPF, IS-IS) or EGP (BGP-4)
  - Performance limited by the shortest/policy path nature
  - Assumptions: Quasi-static traffic, knowledge of demand matrix



Fundamental Requirements

- Need the ability to:
  - Map traffic to an LSP
  - Monitor and measure traffic
  - Specify explicit path of an LSP
    - Partial explicit route
    - Full explicit route
  - Characterize an LSP
    - Bandwidth
    - Priority/ Preemption
    - Affinity (Link Colors)
  - Reroute or select an alternate LSP

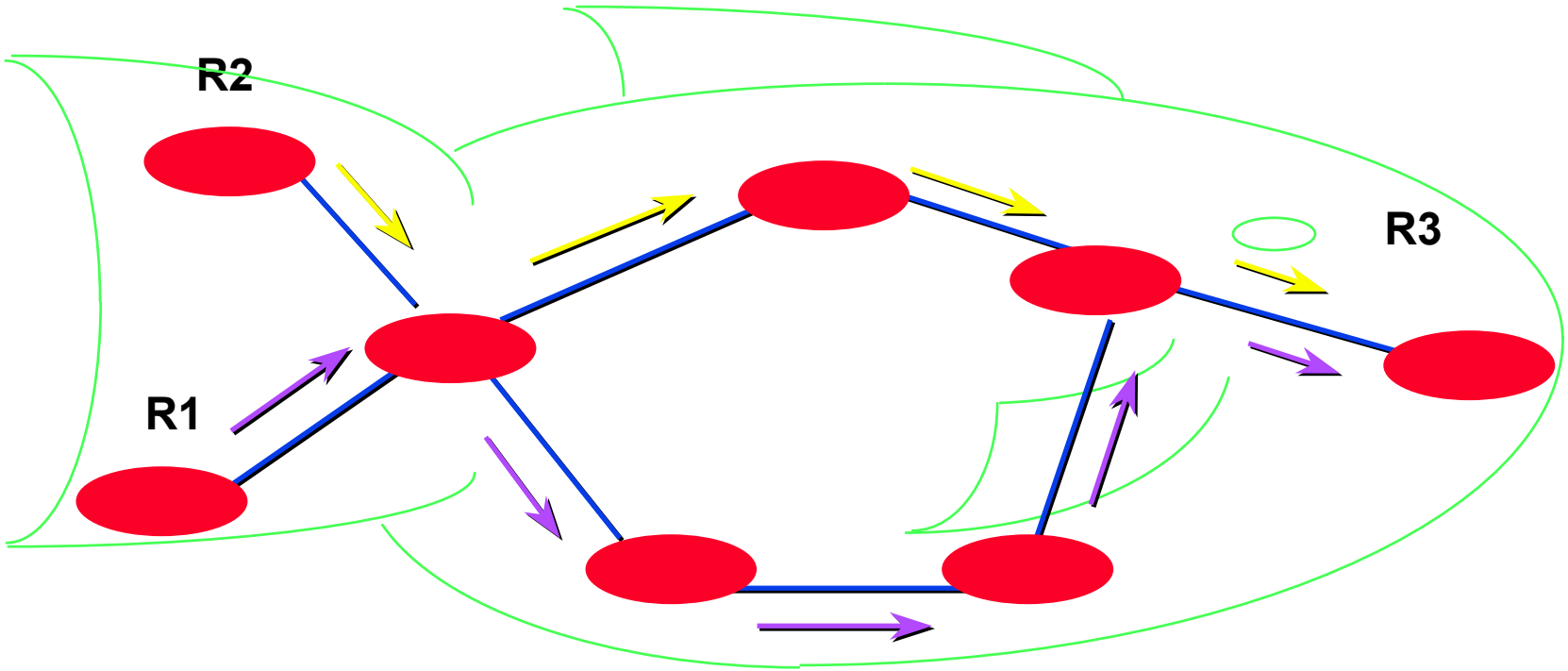
Traffic Engineering Steps

- First, determine how to lay out traffic on the physical topology
  - Measure traffic (e.g., city-pair-wise)
  - Crunch numbers
- Second, do something to convince the packets to follow your plan

Traffic Engineering Options

- BGP – play with communities, filtering
- IGP – play with metrics
  - Linear programming can help
- Source routing
  - ATM
  - MPLS

Routing Solution to Traffic Engineering

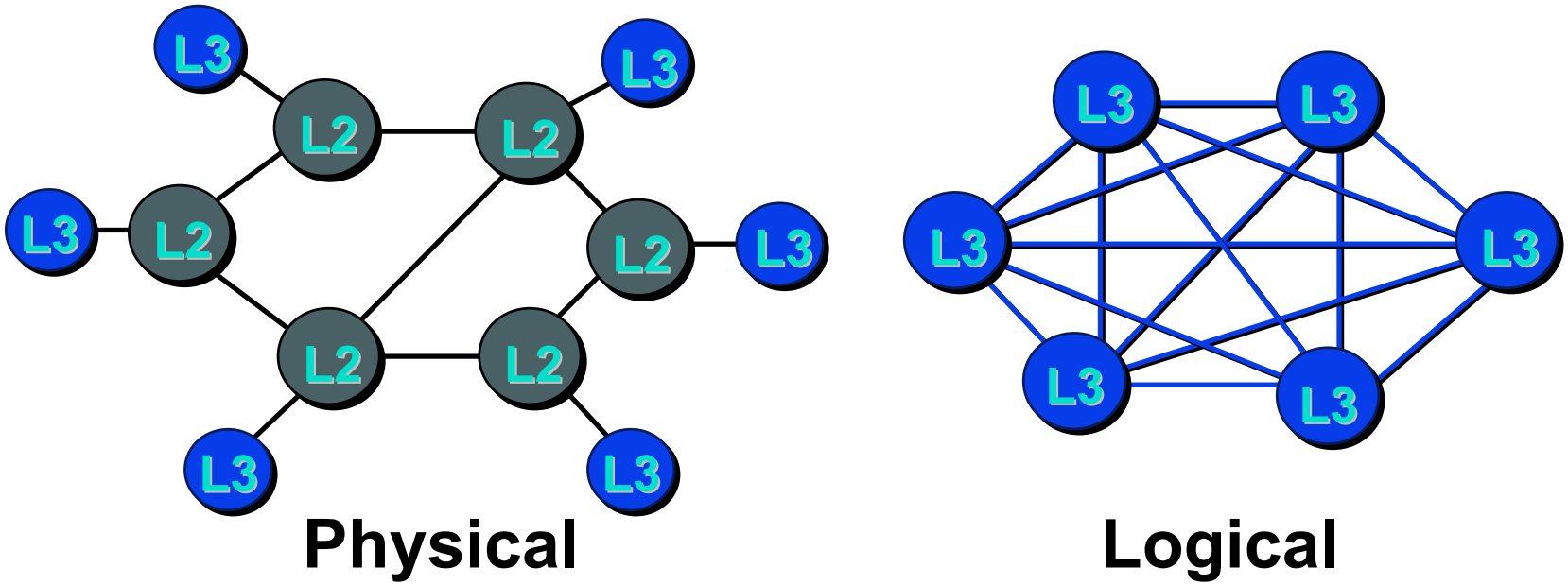


- Construct routes for traffic streams within a service provider in such a way, as to avoid causing some parts of the provider's network to be over-utilized, while others parts remain under-utilized (I.e. load-balance)

Linear Programming

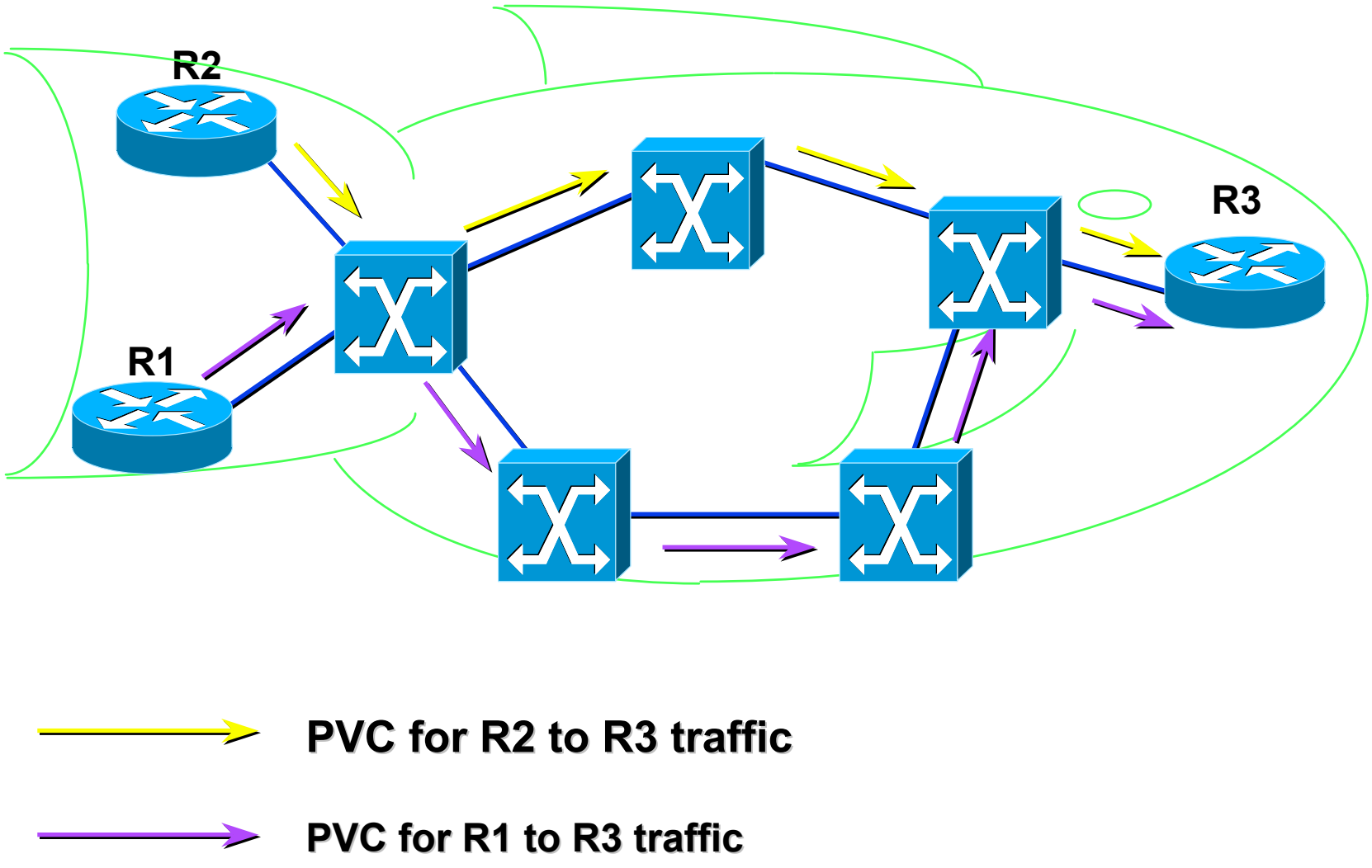
- TE among N cities: N^2 city pairs
- Set up N^2 by N^2 matrix for LP
- Matrix multiplication/inversion is $O(M^3)$ for M x M matrix; simplex is $O(M^3)$ matrix operations
- So, LP problem is $O(N^{12})$
- Also can't deal with "looped routes"

The "Overlay" Solution



- Routing at layer 2 (ATM or FR) is used for traffic engineering
- Analogy to direct highways between SFO-LAX & SAN-SMF. Nobody enters the highway in between.

Traffic engineering with overlay



Connectionless Routing Today

- Internet connectionless routing protocols originally designed to find one route
  - Eg: shortest route or policy route)

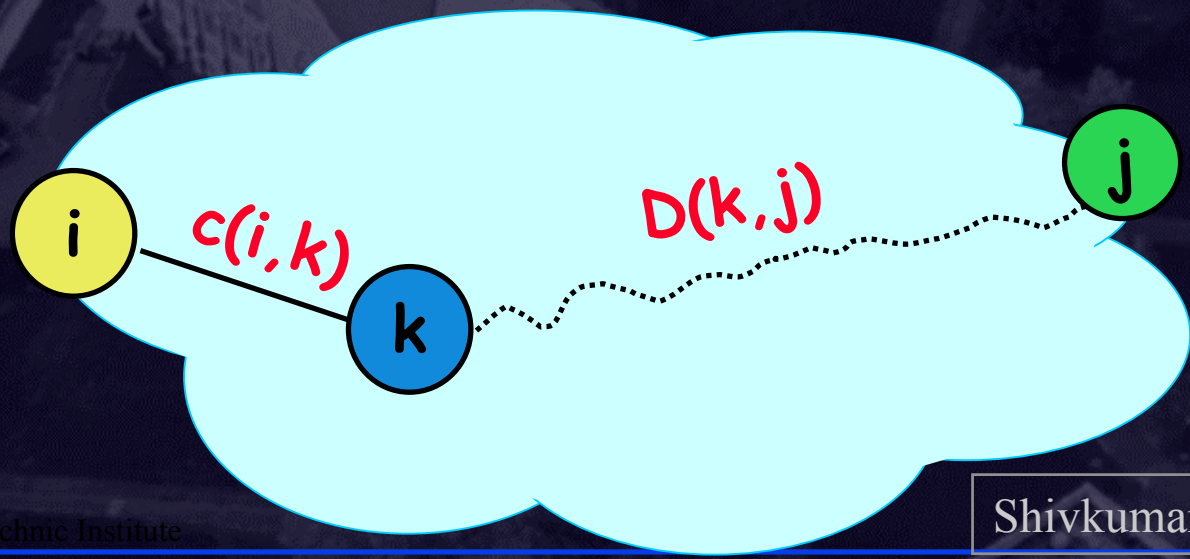
- Connectionless routing relies upon a global consistency criterion (GCC)
  - The GCC is constructed using globally known identifiers (Eg: ASNs, link weights)

DV: Global Consistency Criterion

- ***The subset of a shortest path is also the shortest path between the two intermediate nodes.***
- If the shortest path from node i to node j, with distance D(i,j) passes through neighbor k, with link cost c(i,k), then:

$$D(i,j) = c(i,k) + D(k,j)$$

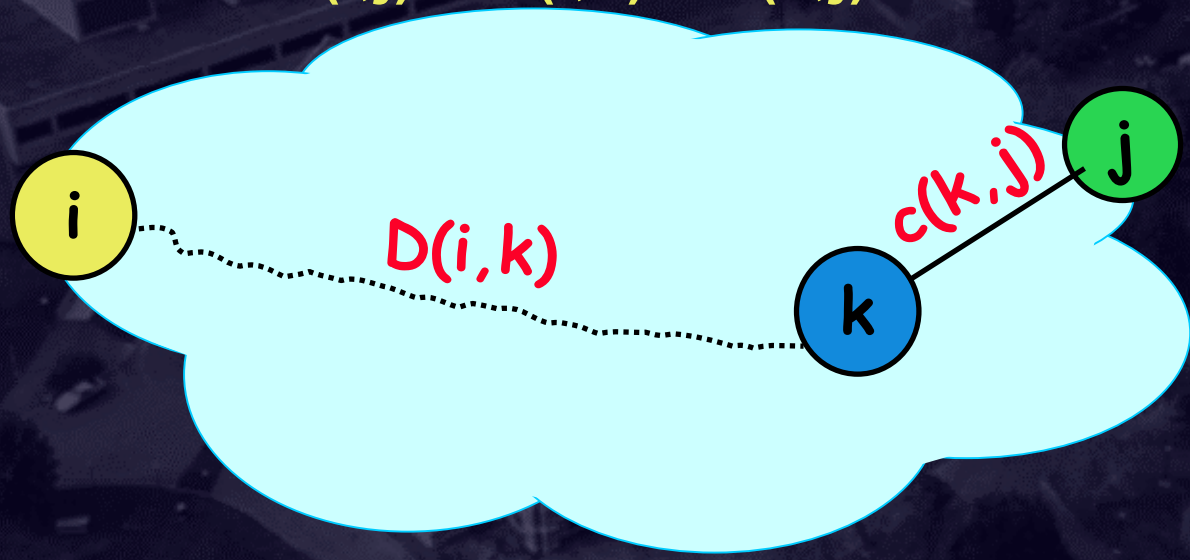
- **D(i,*)** is a distance vector at node i.



Link State (LS): Global Consistency Criterion

- The *link state (Dijkstra) approach is iterative, but it* pivots around destinations j, and their predecessors k = p(j)
  - Alternative version of the consistency condition:

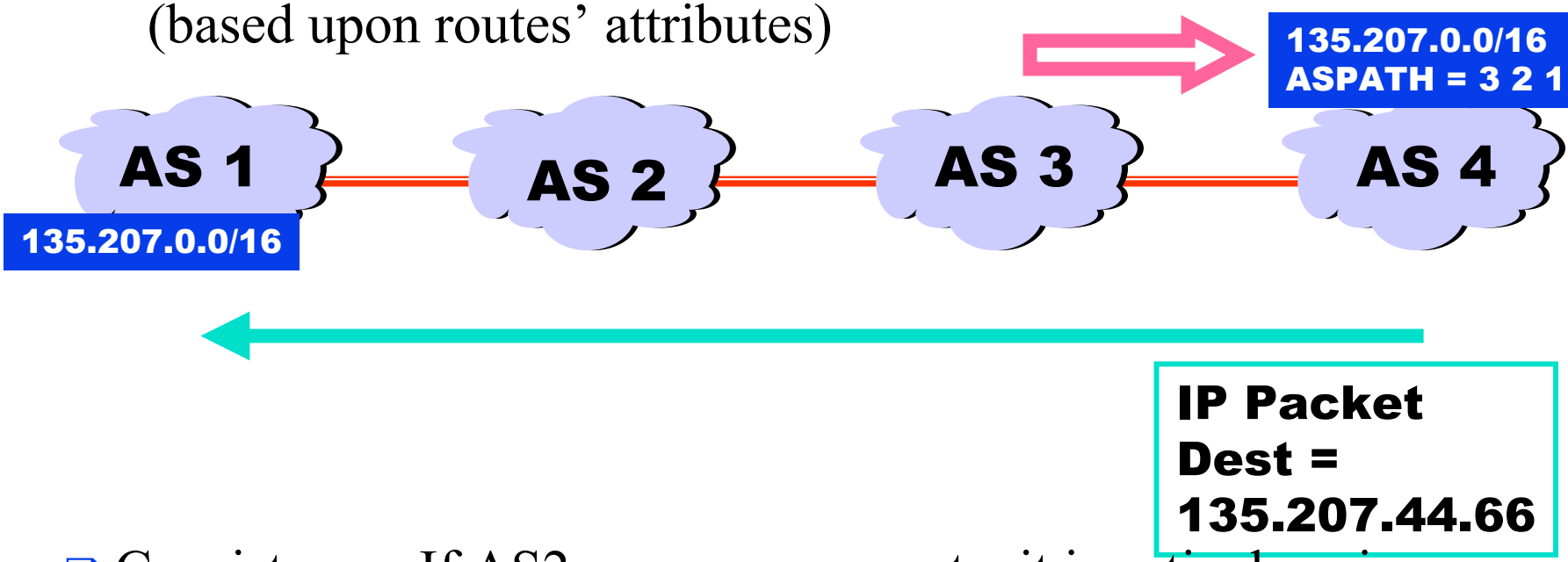
D(i,j) = D(i,k) + c(k,j)



- Each node i collects all link states **c(*,*)** first and runs the complete Dijkstra algorithm locally.

Path-Vector: BGP's Consistency Criterion

- Policy-based routing:
  - Arbitrary preference among a menu of available routes (based upon routes' attributes)



- Consistency: If AS2 announces a route, it is actively using the route, and will honor forwarding requests on that route

Acknowledgement: Based upon Dr. Tim Griffin's SIGCOMM Tutorial Slides

Limitations of Today's Connectionless TE

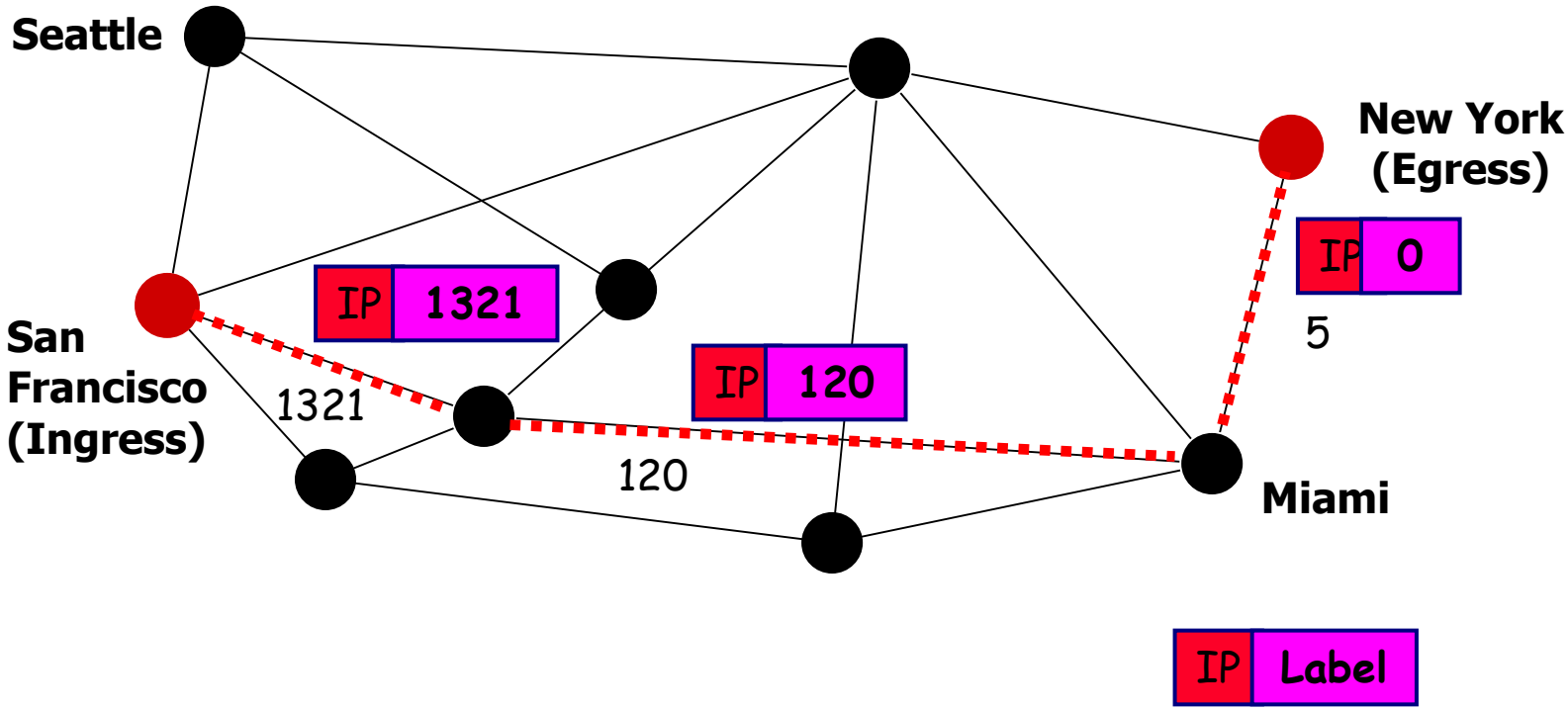
- Traffic mapping coupled with route availability
  - Changing parameters changes routes AND changes the traffic mapped to the routes
- Priority rules only:
  - LOCAL-PREF, MED, longest-prefix match
  - Cannot split traffic to same destination among two paths

Signaled Approach (eg: MPLS)

- Nice features:
  - In MPLS, choice of a route (and its setup) is orthogonal to the problem of traffic mapping onto the route
  - Signaling maps global IDs (addresses, path-specification) to local IDs (labels)
  - Nice label stacking, tunneling features

Label-Switched Forwarding

- San Francisco prepends MPLS header to the IP packet
- MPLS label is swapped at each hop along the LSP
- Forwarding is done based on a label table



What Does MPLS Offer?

- Tunnels
  - Drop a packet in, and out it comes at the other end *without being IP routed*
- Explicit (source) routing (circuits)
- Label stack
  - 2-label stack: “outer” label defines the tunnel; “inner” label de-multiplexes
- Layer 2 independence

Why Tunnels?

- *Can't* IP route
  - Non-IP packets
  - IP packets with private addresses
- *Don't want to* IP route
  - "BGP-free" core
  - Don't like IP multicast model

Tunnel Comparison

MPLS (LDP) tunnels

- Small header
- Label stacking
- Signaling for demux
- Automagic tunnels
- Tracks IP routing
- Harder to spoof
- No data security

IP tunnels

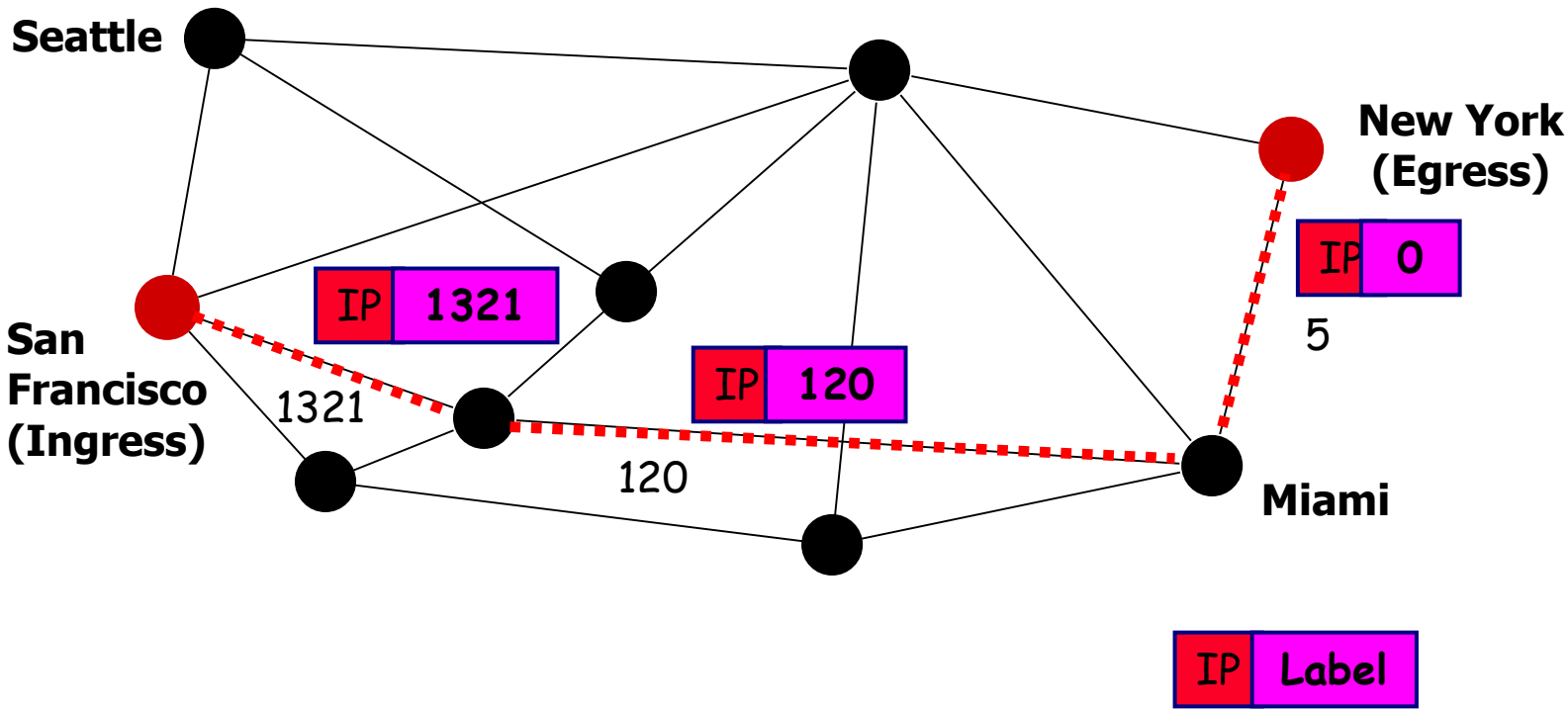
- Big header
- No stacking (*)
- No signaling (yet)
- Configured tunnels
- Duh!
- Spoofable
- IPSec

Bottom Line on Tunnels

- Don't *need* MPLS for tunnels
- But MPLS tunnels have some nice properties
- Decision (should be) based on cost of deploying new protocol vs. benefits

MPLS Signaling and Forwarding Model

- MPLS label is swapped at each hop along the LSP
- Labels = LOCAL IDENTIFIERS …
  - Signaling *maps global identifiers* (addresses, path spec) to *local identifiers*



Limitations of Signaled TE Approach

- Requires extensive upgrades in the network
- Hard to inter-network beyond area boundaries
- Very hard to go beyond *AS boundaries*
  - Even within the same organization/ISP !
  - Note: large ISPs (eg: ATT) have several AS'es
- Impossible for inter-domain routing across *multiple organizations*
  - Inter-domain TE has to be connectionless

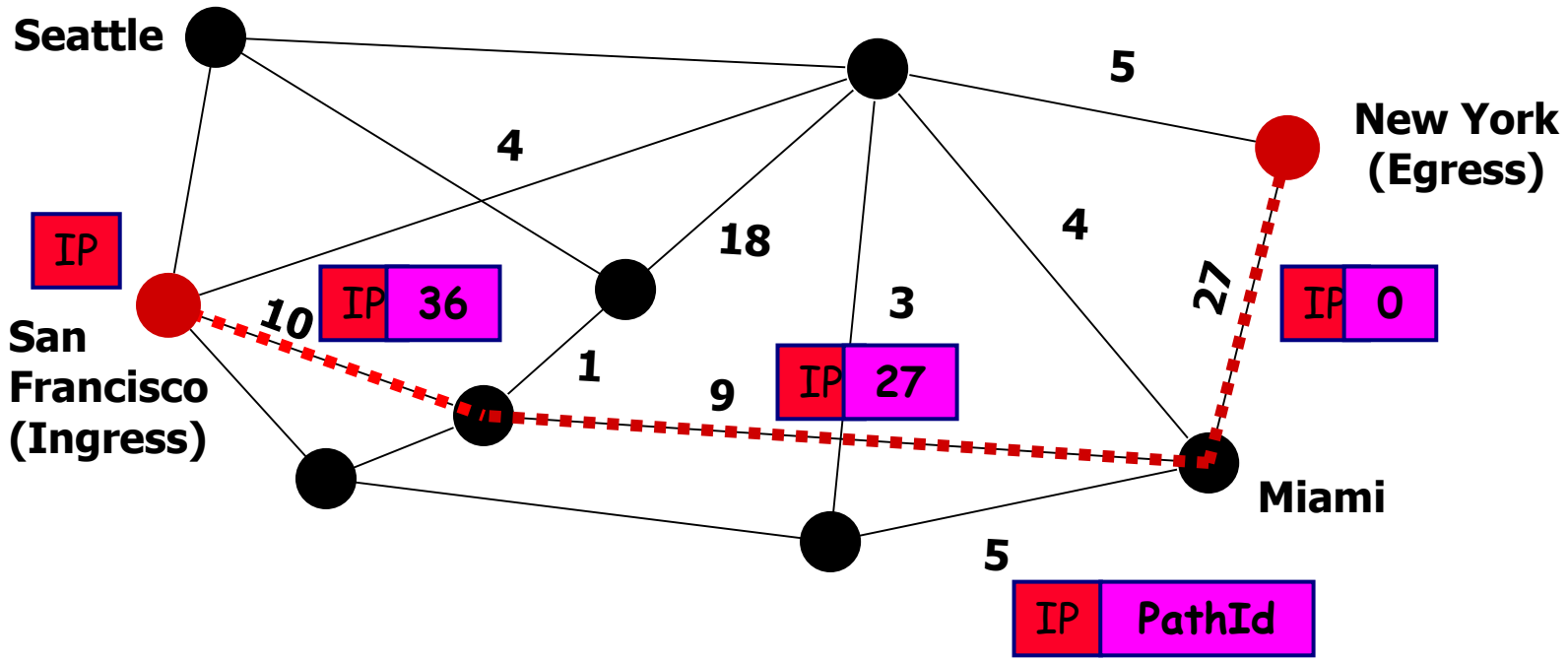
Traffic Engineering w/o Signaling?

- Fine-grained Traffic Engineering needs some form of source routing
- Specific incremental changes much easier with source routing
  - Change a single city-pair flow
  - Reacting to a link failure

- *Can we do source-routing efficiently in connectionless protocols?*

Idea!

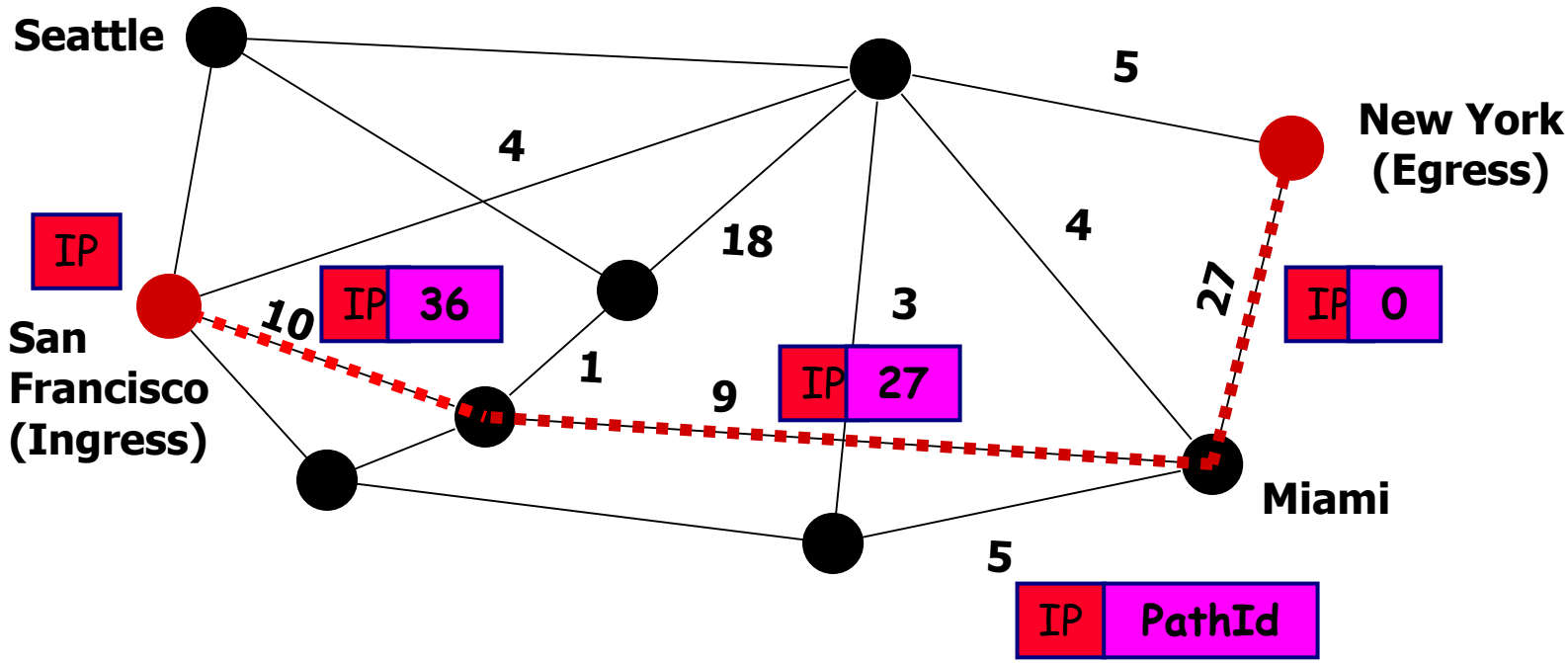
- Instead of using *local path identifiers* (Labels in MPLS), use *global path identifiers*



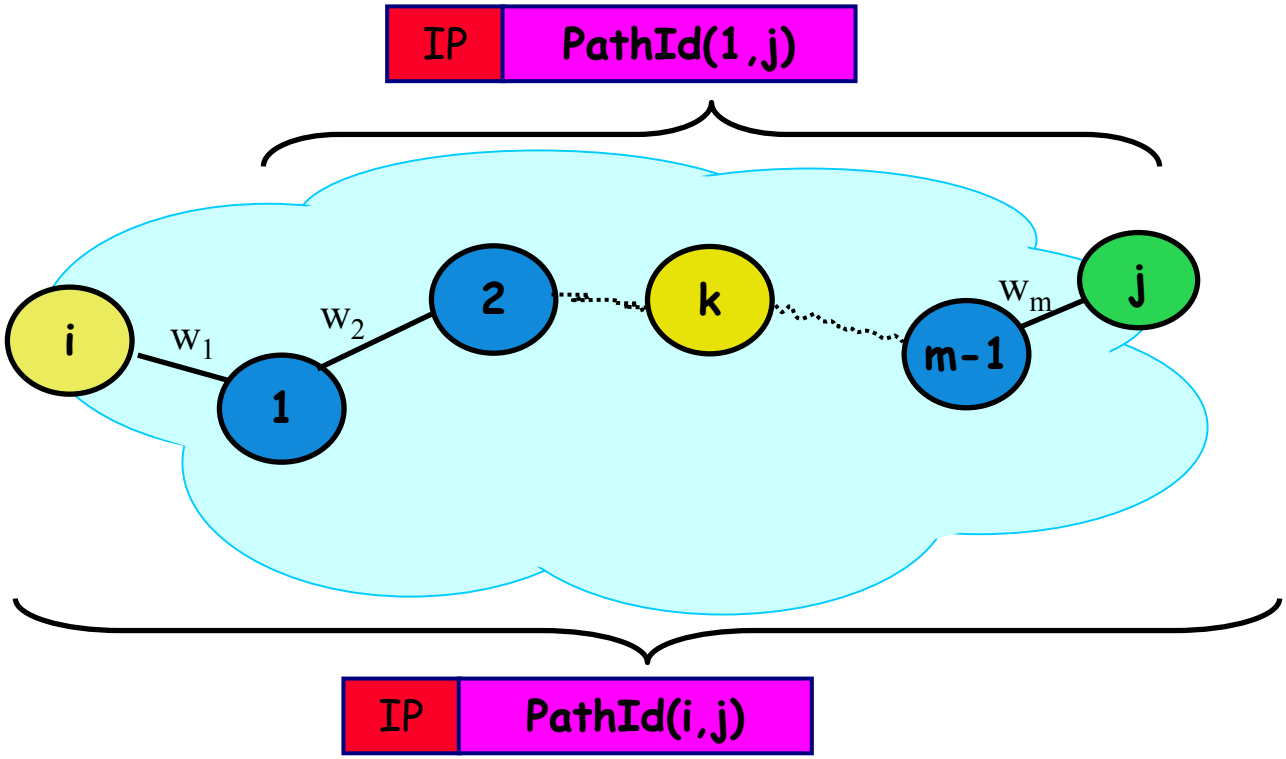
Routers have capability to compute multiple paths using map from IGP (OSPF/IS-IS)

Global Path Identifiers

- Instead of using *local path identifiers* (Labels in MPLS), we propose the use of *global path identifiers*

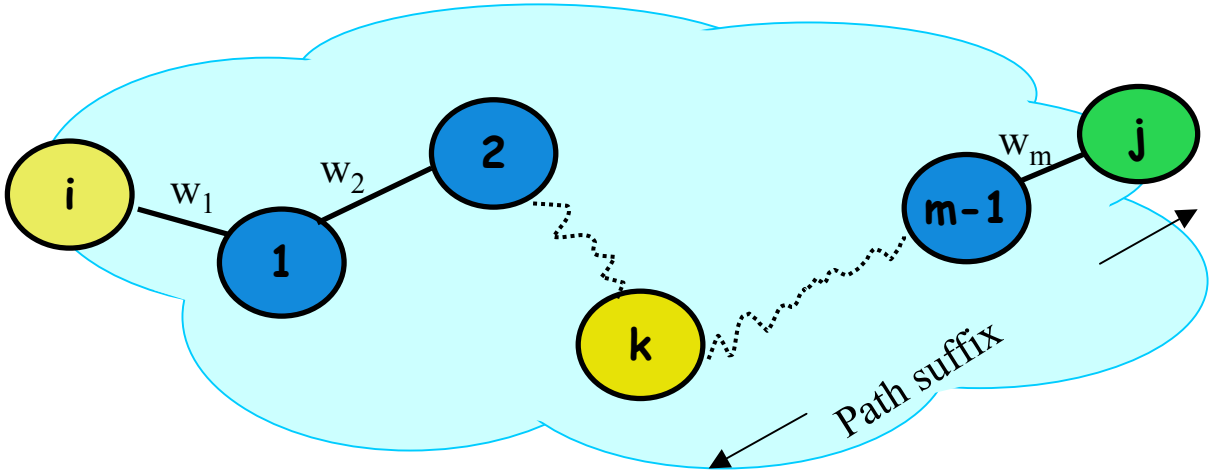


Global Path Identifier



Central idea: Swap ***global*** pathids instead of ***local*** labels!

Global Path Identifier (contd)



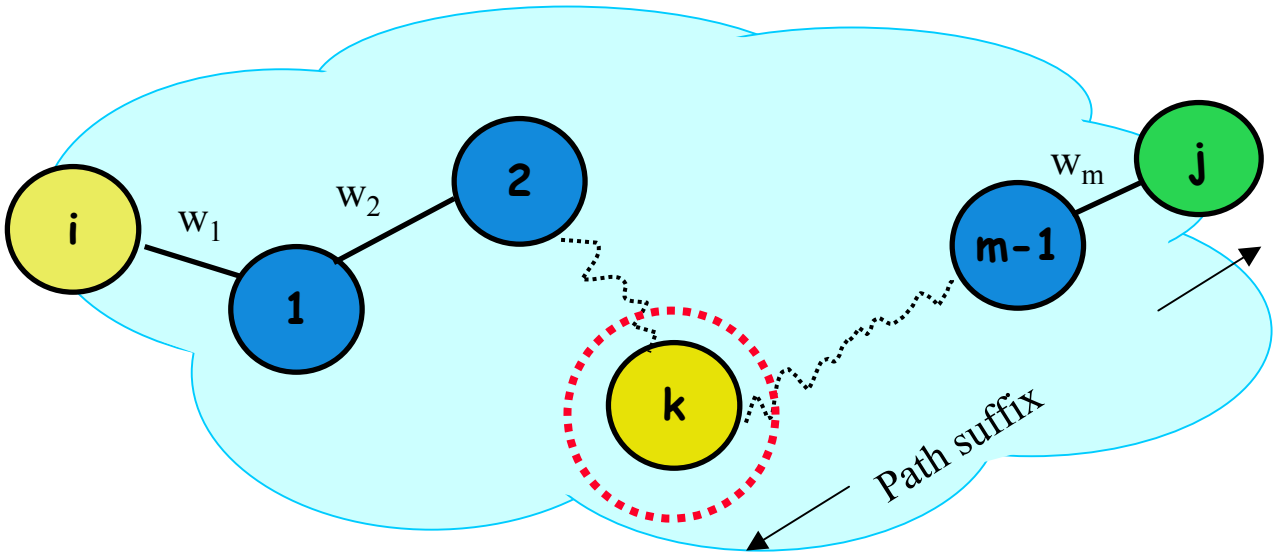
- **Path = {i, w_1, 1, w_2, 2, …, w_k, k, w_{k+1}, … , w_m, j}**
  - Sequence of globally known node IDs & Link weights
  - Global Path ID is a **hash** of this sequence => **locally computable without the need for signaling!**

- Potential hash functions:
  - [j, { h(1) + h(2) + …+h(k)+ … +h(m-1) } mod 2^b]: **node ID sum**
  - **MD5 one-way hash, XOR, 32-bit CRC** etc…
  - We propose the use of MD5 hashing of the subsequence of nodeIDs followed by a CRC-32 to get a 32-bit hash value
    - **Very low collision (I.e. non-uniqueness) probability**

Abstract Forwarding Paradigm

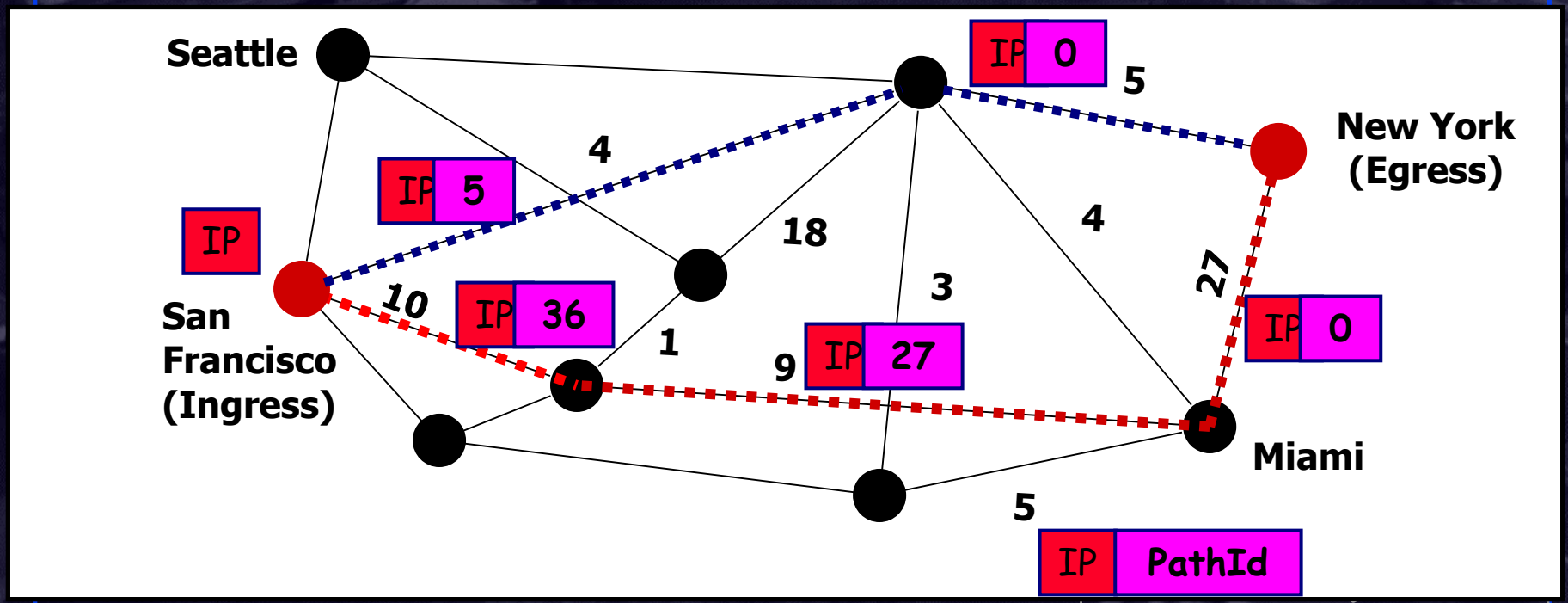
- Forwarding table (Eg; at **Node k**):
  - **[Destination Prefix,PathID] → [Next-Hop, SuffixPathID]**
  - **[j, H{k, k+1, … , m-1}] → [k+1, H{k+1, … , m-1}]**

- Incoming Packet Hdr: Destination address (**j**) & PathID = **H{k, k+1, … , m-1}**
- Outgoing Packet Hdr: **[j, PathID = H{k+1, … , m-1}]**
  - **Longest prefix match + exact label match + label swap!**
  - PathID mismatch => map to shortest (default) path, and set PathID = 0
  - *No signaling* because of globally meaningful pathIDs!



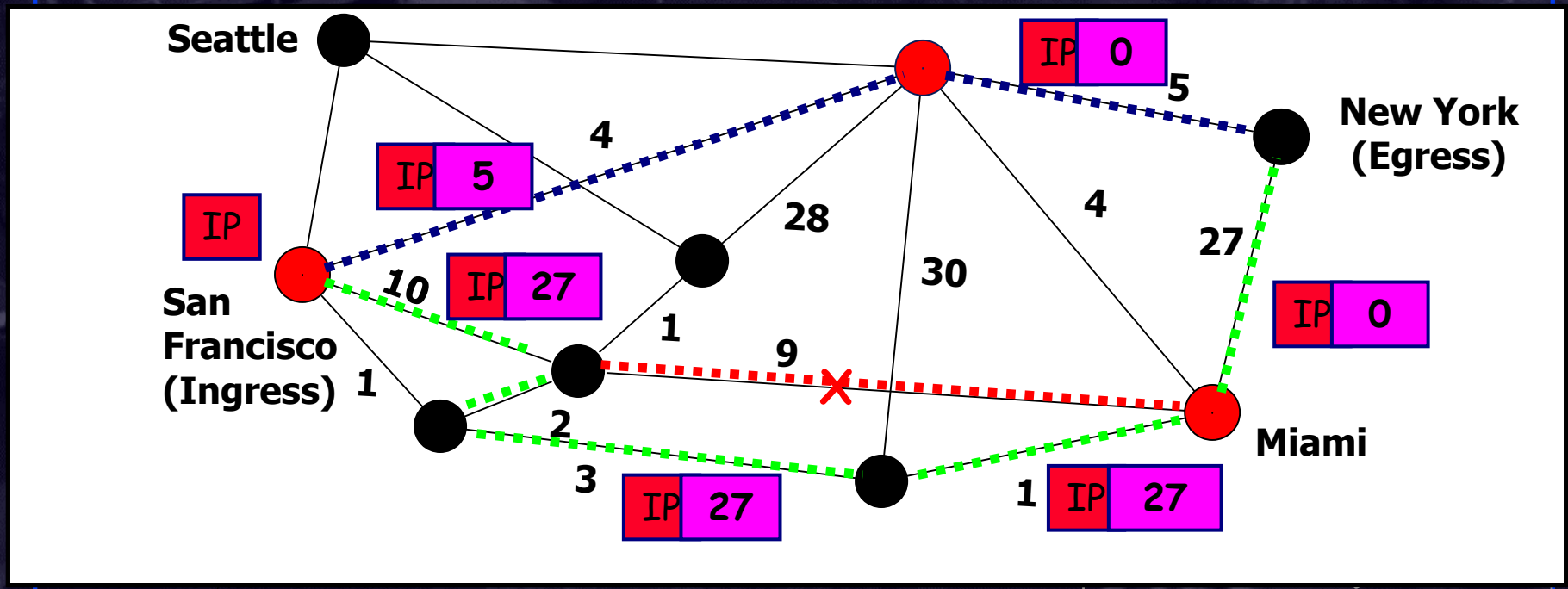
BANANAS TE: Explicit, Multi-Path Forwarding…

- ***Explicit Source-Directed Routing:*** Not limited by the shortest path nature of IGP
  - Different PathIds => different next-hops (***multi-paths***)
  - ***No signaling*** required to set-up the paths
- Traffic splitting is ***decoupled*** from route computation



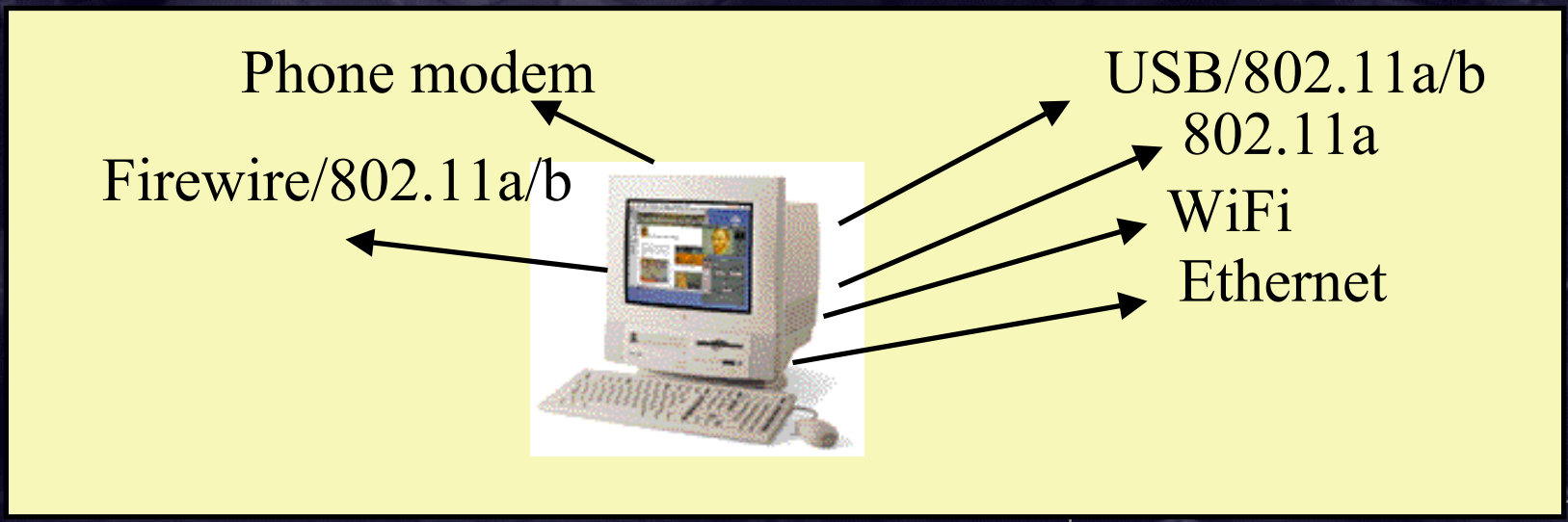
BANANAS TE: Partial Deployment

- Only “red” routers are upgraded
  - Link State Advertisements (LSAs) may indicate (with 1 bit) which routers are upgraded
  - Non-upgraded routers forward everything on the shortest path (default path): forming a “***virtual hop***”



Multiplicity Paradigm

- Unlike telephony, data networking can get statistical multiplexing gains from simultaneously using:
  - Multiple transmission modes (802.11a/b, 3G etc)
  - Multiple exits (USB, Firewire, Ethernet, modem)
  - Multiple paths (routes)
    - Lightweight distributed QoS on each path
- Can then quickly meet the performance thresholds of high-quality multimedia apps!



Eg: Multipath MPEG using Multi-band 802.11a/b Community Wireless Networks