Traffic Measurement and Inference for IP Networks

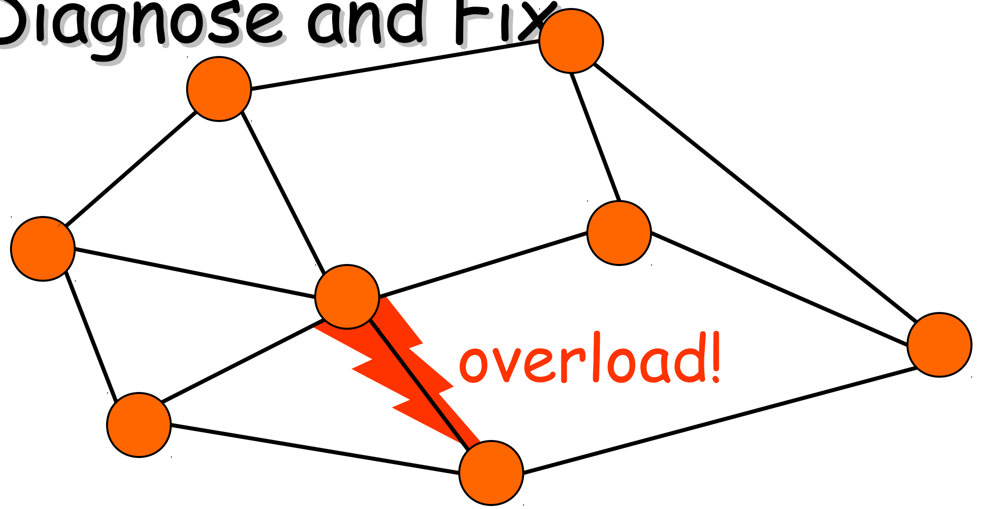
Prof. Francesco Lo Presti

Dipartimento di Informatica
Università di Roma "Tor Vergata"

Thanks to: M. Grossglauser & J. Rexford (AT&T Labs)

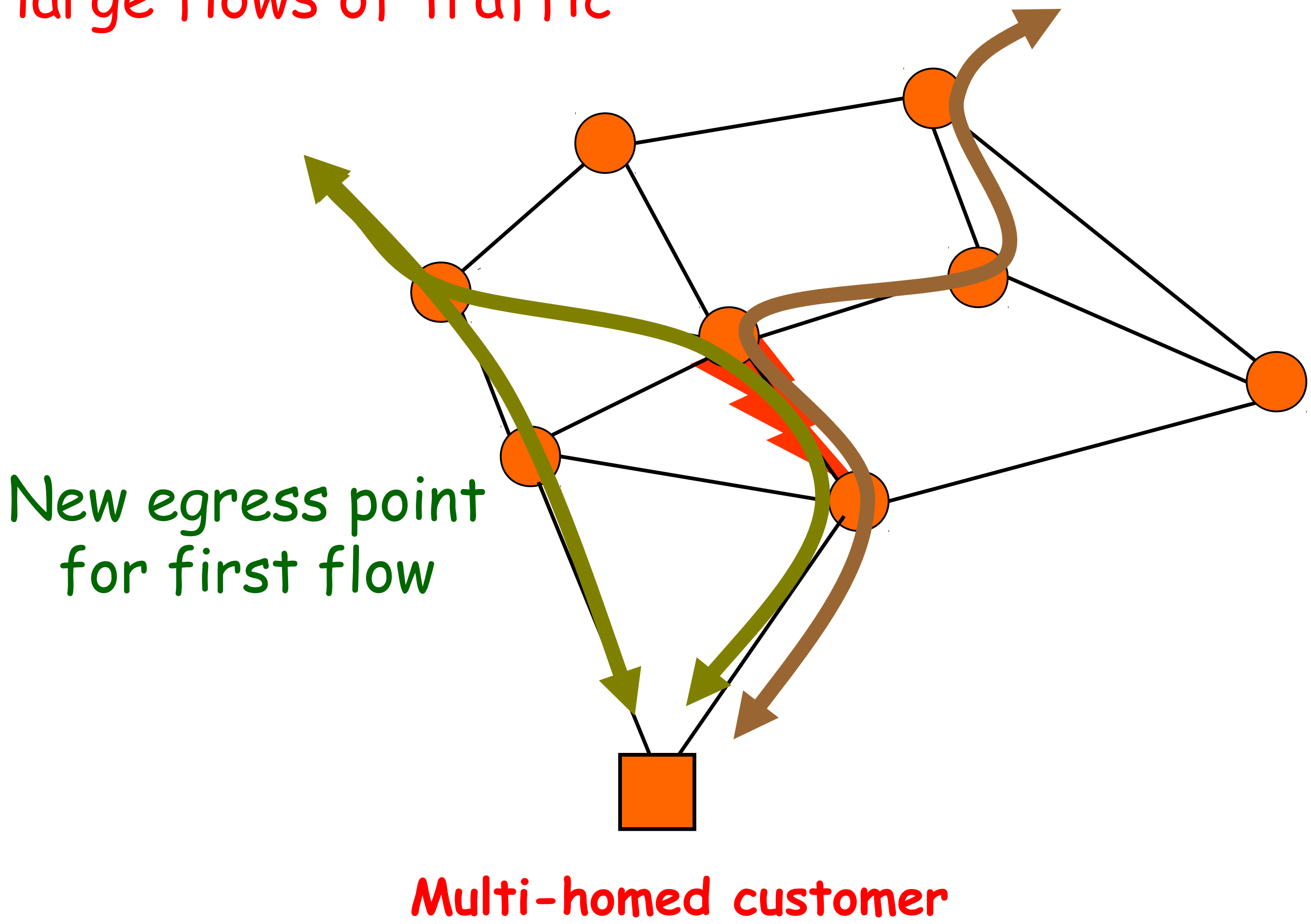
Motivation: Network Operations

- Detecting, Diagnose and Fix

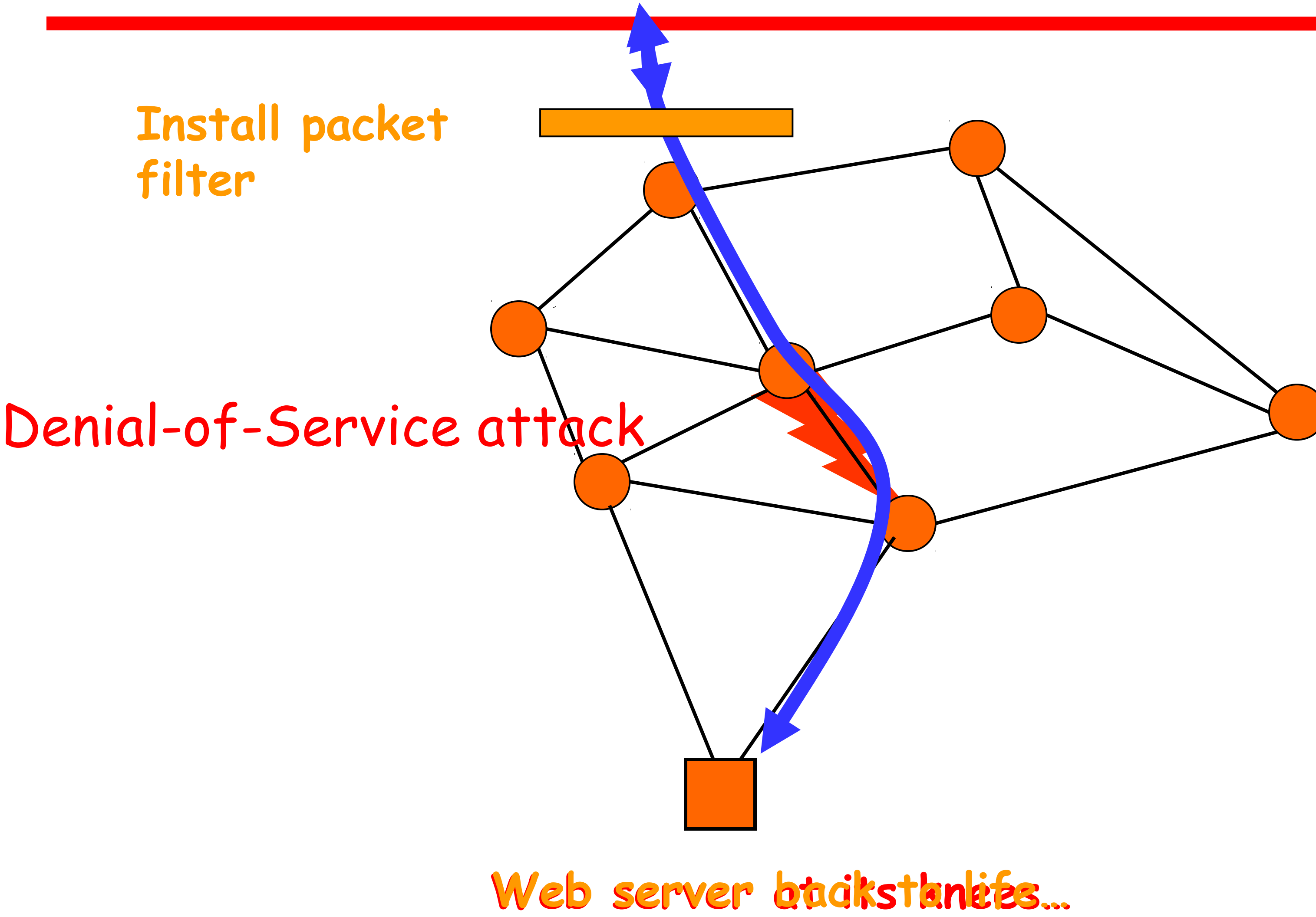


Excess Traffic

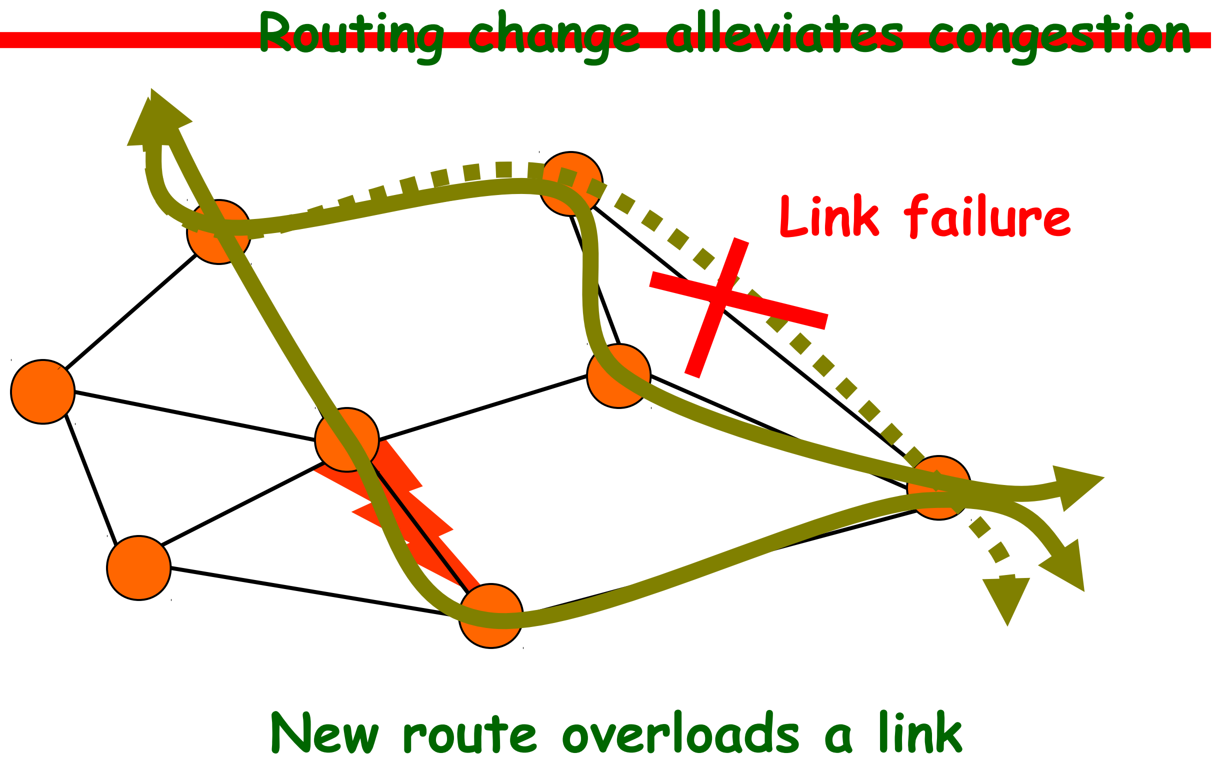
Two large flows of traffic



DoS Attack


Install packet filter
Denial-of-Service attack
Web server back to life…
Web server at its knees…

Link Failure



Summary of the Examples

- How to *detect* that a link is congested?
  - Periodic polling of link statistics
  - Active probes measuring performance
  - Customer complaints
- How to *diagnose* the reason for the congestion?
  - Change in user behavior
  - Denial of service attack
  - Router/link failure or policy change
- How to *fix* the problem???
  - Interdomain routing change
  - Installation of packet filters
  - Intradomain routing change

Network measurement plays a key role in each step!

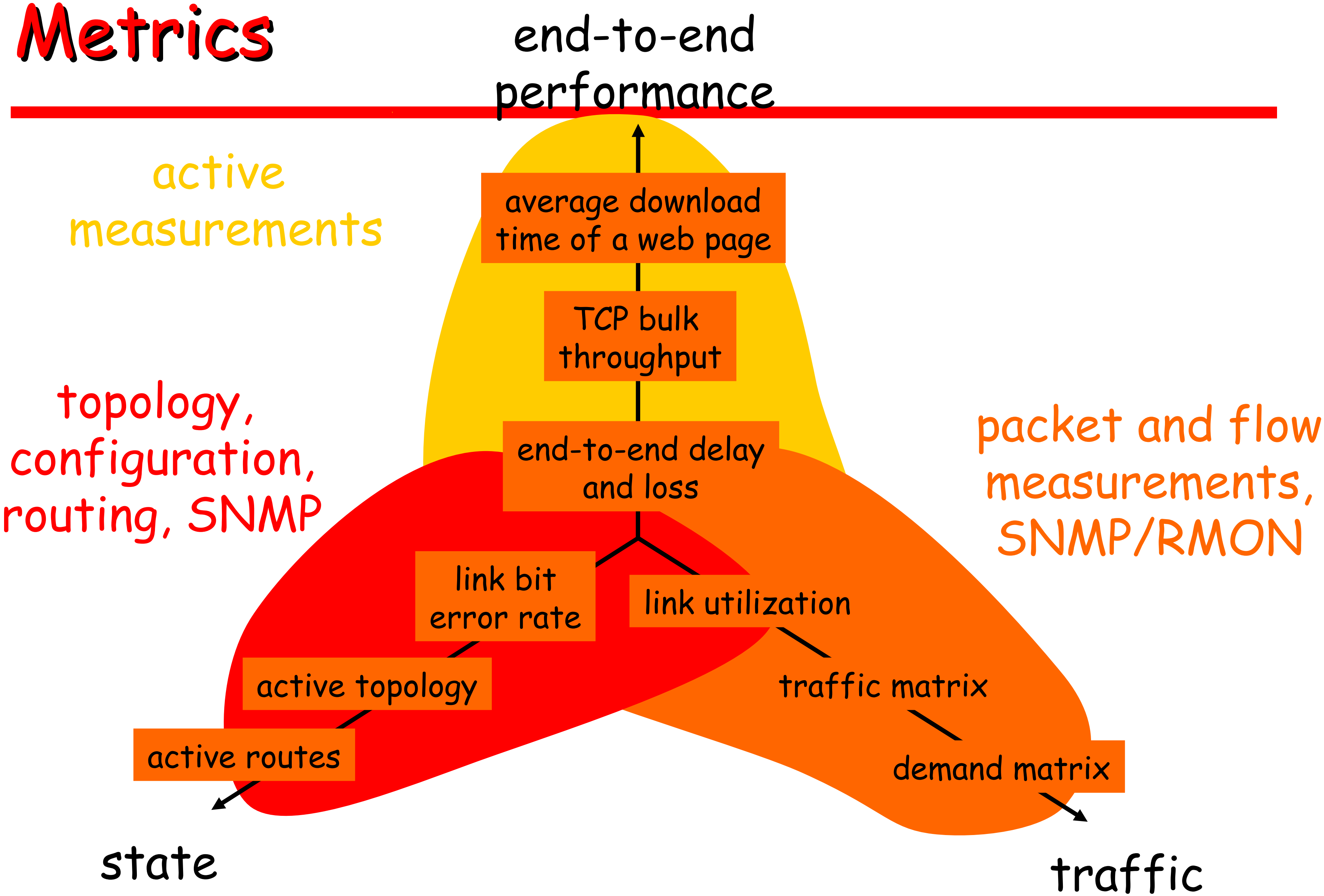
The Role of Traffic Measurement

- Operations (control)
  - Generating reports for customers and internal groups
  - Diagnosing performance and reliability problems
  - Tuning the configuration of the network to the traffic
  - Planning outlay of new equipment (routers, proxies, links)
- Application (performance and reliability)
  - Choose among several servers/replicas (CDN)
- Science (discovery)
  - End-to-end characteristics of delay, throughput, and loss
  - Verification of models of TCP congestion control
  - Workload models capturing the behavior of Web users
  - Understanding self-similarity/multi-fractal traffic

Measurement Challenges

- Network-wide view
  - Crucial for evaluating control actions
  - Multiple kinds of data from multiple locations
- Large scale
  - Large number of high-speed links and routers
  - Large volume of measurement data
- Poor state-of-the-art
  - Working within existing protocols and products
  - Technology not designed with measurement in mind
- The "do no harm" principle
  - Don't degrade router performance
  - Don't require disabling key router features
  - Don't overload the network with measurement data

Terminology: Measurements vs Metrics



Active Measurement

- Definition:
  - Injecting measurement traffic into the network
  - Computing metrics on the received traffic
- Scope
  - Closest to end-user experience
  - Least tightly coupled with infrastructure
  - Comes first in the detection/diagnosis/correction loop
- Outline
  - Tools for active measurement: probing, traceroute
  - Operational uses: intradomain and interdomain
  - Inference methods: peeking into the network
  - Standardization efforts

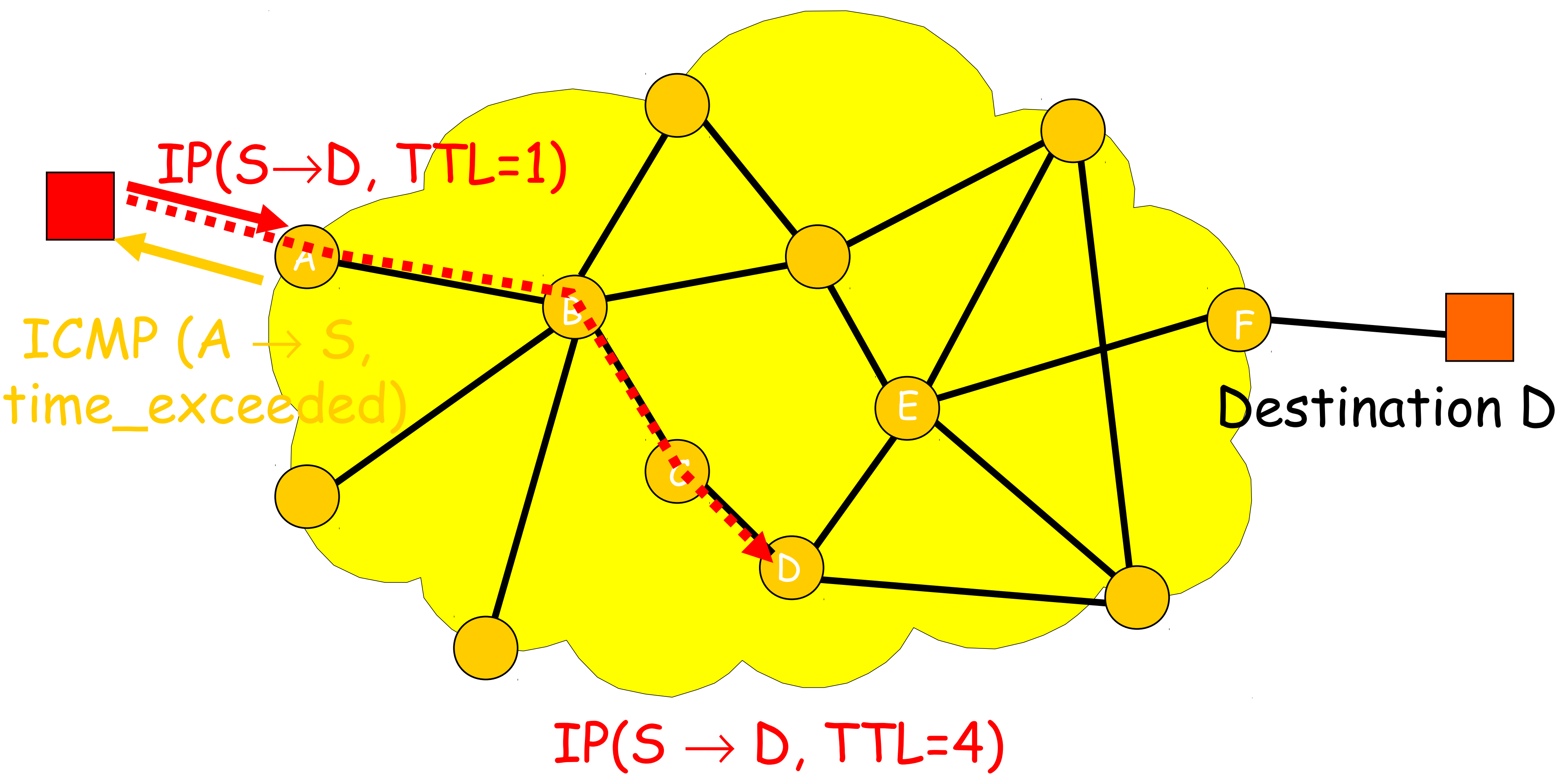
Tools: Ping

- ICMP-echo request-reply
- Advantage: wide availability (in principle, any IP address)
- Drawbacks:
  - pinging routers is bad! (except for troubleshooting)
    - delay measurements very unreliable/conservative

Tools: Traceroute

- Exploit TTL (Time to Live) feature of IP
  - When a router receives a packet with TTL=1, packet is discarded and ICMP_time_exceeded returned to sender
- Operational uses:
  - Can use traceroute towards own domain to check reachability
    - list of traceroute servers: http://www.traceroute.org
  - Debug internal topology databases
  - Detect routing loops, partitions, and other anomalies

Tools: Traceroute



Operational Uses: Intradomain

- Types of measurements:
  - » loss rate
  - » average delay
  - » delay jitter
- Various homegrown and off-the-shelf tools
  - » Ping, host-to-host probing, traceroute,...
- Operational tool to verify network health, check service level agreements (SLAs)
  - » Promotional tool for ISPs:
  - » advertise network performance

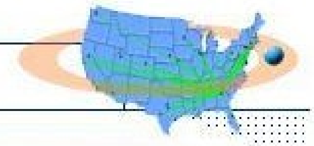
Example: AT&T

AT&T

AT&T DATA & IP SERVICES
Networking the New Economy

AT&T DATA & IP SERVICES
HOME
CURRENT PERFORMANCE
BACKBONE DELAY AND LOSS
MAY AVERAGES
NETWORK STATUS INFORMATION
METHODOLOGY
GLOSSARY

Delay and Loss



12 Backbone Nodes at a Glance

BACKBONE DELAY

Thresholds are distance sensitive

Current Average 36 ms

	Att	Cam	Chi	Dal	Den	LA	NY	Orl	SF	StL	Sea
Cam	25										
Chi	37	23									
Dal	14	41	20								
Den	35	62	40	21							
LA	50	70	48	36	32						
NY	20	5	17	33	55	65					
Orl	9	35	46	23	44	59	30				
SF	58	62	40	44	24	9	57	67			
StL	30	29	6	14	34	40	23	39	40		
Sea	73	68	45	59	39	24	62	82	16	55	
Was	14	10	23	29	47	56	6	24	56	16	67

Legend | Increasing Delay

BACKBONE LOSS

Expressed as a %

Current Average 0.0%

	Att	Cam	Chi	Dal	Den	LA	NY	Orl	SF	StL	Sea
Cam	0.0										
Chi	0.0	0.0									
Dal	0.0	0.0	0.0								
Den	0.0	0.0	0.0	0.0							
LA	0.0	0.0	0.0	0.0	0.0						
NY	0.0	0.0	0.0	0.0	0.0	0.0					
Orl	0.0	0.0	0.0	0.0	0.0	0.0	0.0				
SF	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0			
StL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
Sea	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Was	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Legend | Less than 5% | 5% to 10% | More than 10%

Operational Uses: Interdomain

- Infrastructure efforts:
  - NIMI (National Internet Measurement Infrastructure)
    - measurement infrastructure for research
    - shared: access control, data collection, management of software upgrades, etc.
  - RIPE NCC (Réseaux IP Européens Network Coordination Center)
    - infrastructure for interprovider measurements as service to ISPs
    - interdomain focus
- Main challenge: Internet is large, heterogeneous, changing
  - How to be representative over space and time?

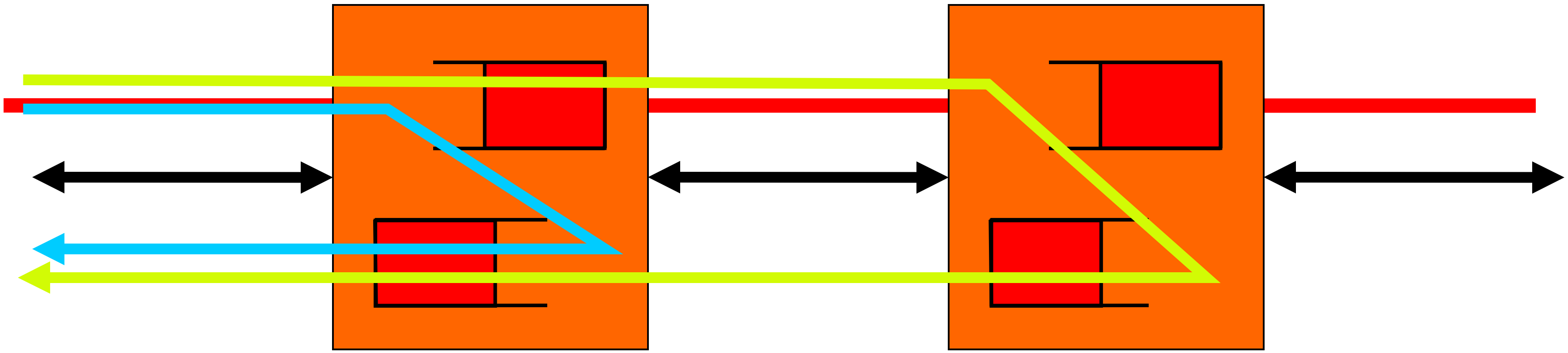
Inference Methods

- ICMP-based
  - Pathchar: variant of traceroute, more sophisticated inference
- Multicast-based inference
  - MINC: infer topology, link loss, delay
  - Also extended to Unicast-based inference

Pathchar

- Similar basic idea as traceroute
  - Sequence of packets per TTL value
- Infer per-link metrics
  - Loss rate
  - Propagation + queueing delay
  - Link capacity
- Operator
  - Detecting & diagnosing performance problem
  - Measure propagation delay (this is actually hard!)
  - Check link capacity

Pathchar (cont.)



$$rtt(i+1) = rtt(i) + d + L/c + \varepsilon$$

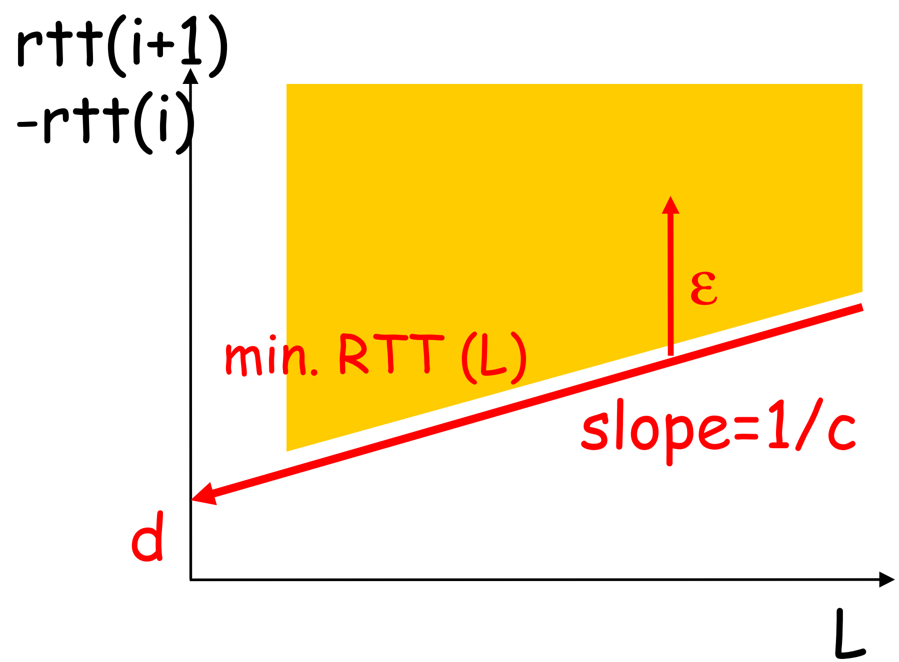
i : initial TTL value

c : link capacity

L : packet size

Three delay components:

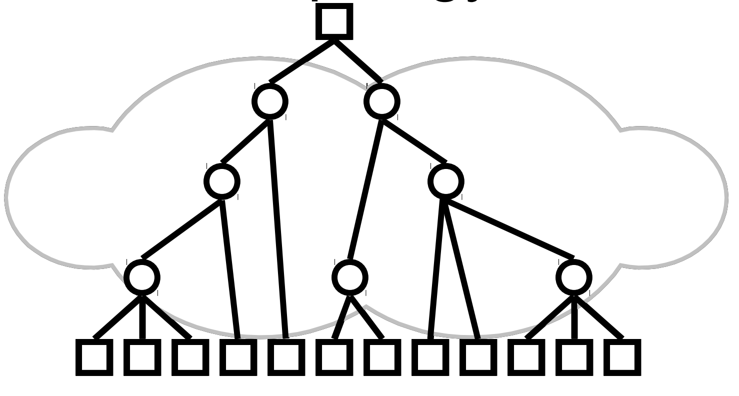
- d : propagation delay
- L/c : transmission delay
- ε : queueing delay + noise

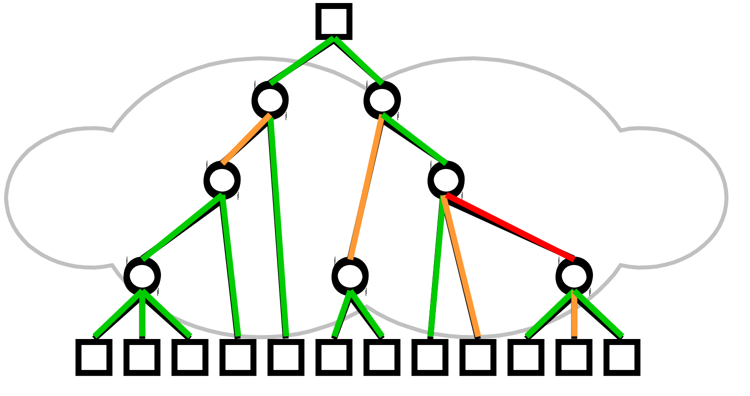


How to infer d,c?

MINC

- MINC (Multicast Inference of Network Characteristics)
- General idea:
  - *A multicast packet "sees" more of the topology than a unicast packet*
  - *Observing at all the receivers*
  - *Analogies to tomography*



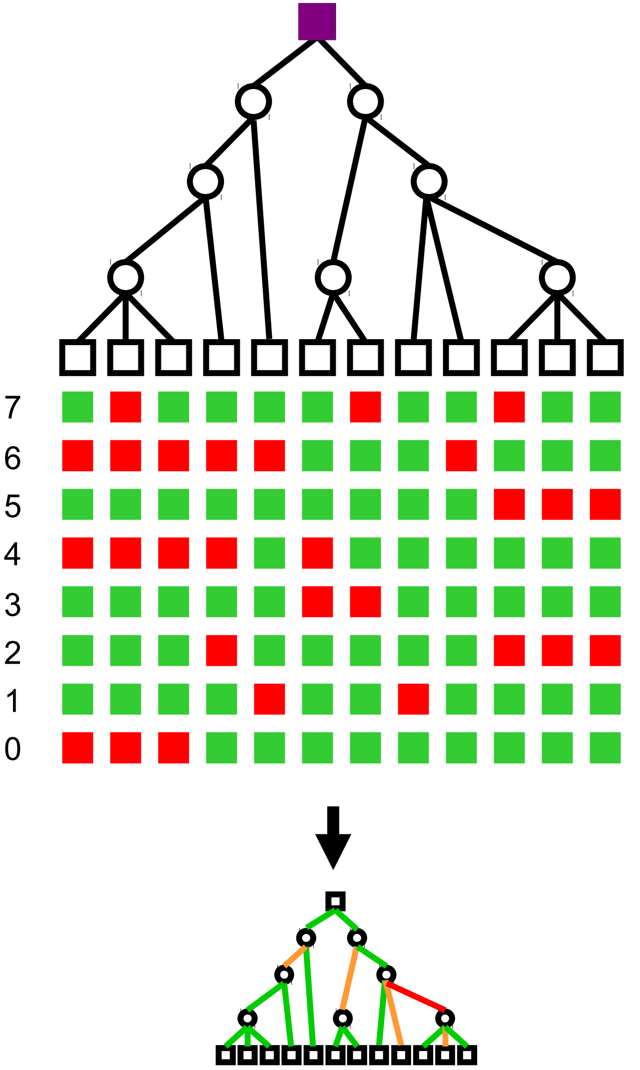


The MINC Approach

1. Sender multicasts packets with sequence number and timestamp

2. Receivers gather loss/delay traces

3. Statistical inference based on loss/delay correlations



Standardization Efforts

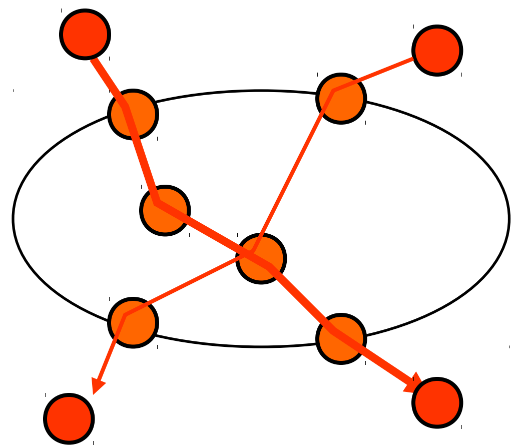
- IETF IPPM (IP Performance Metrics) Working Group
  - Defines standard metrics to measure Internet performance and reliability
    - connectivity
    - delay (one-way/two-way)
    - loss metrics
    - bulk TCP throughput (draft)

Traffic Engineering

- Goal: domain-wide control & management to
  - Satisfy performance goals
  - Use resources efficiently
- Knobs:
  - Configuration & topology: provisioning, capacity planning
  - Routing: OSPF weights, MPLS tunnels, BGP policies,…
  - Traffic classification (diffserv), admission control,…
- Measurements are key: closed control loop
  - Understand current state, load, and traffic flow
  - Ask what-if questions to decide on control actions
  - Inherently coarse-grained

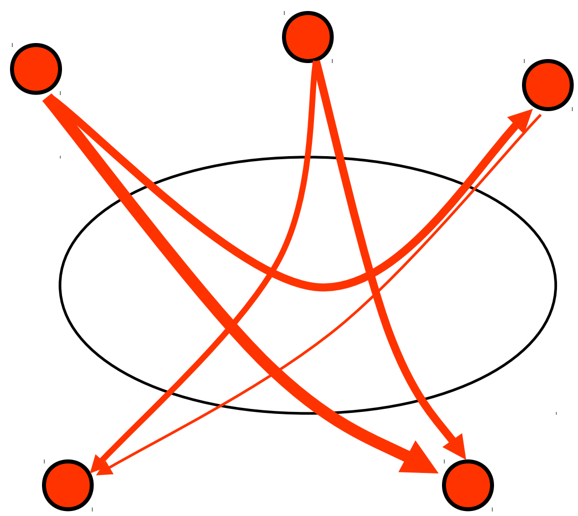
End-to-End Traffic & Demand Models

Ideally, captures all the information about the current network **state and behavior**



path matrix = bytes per path

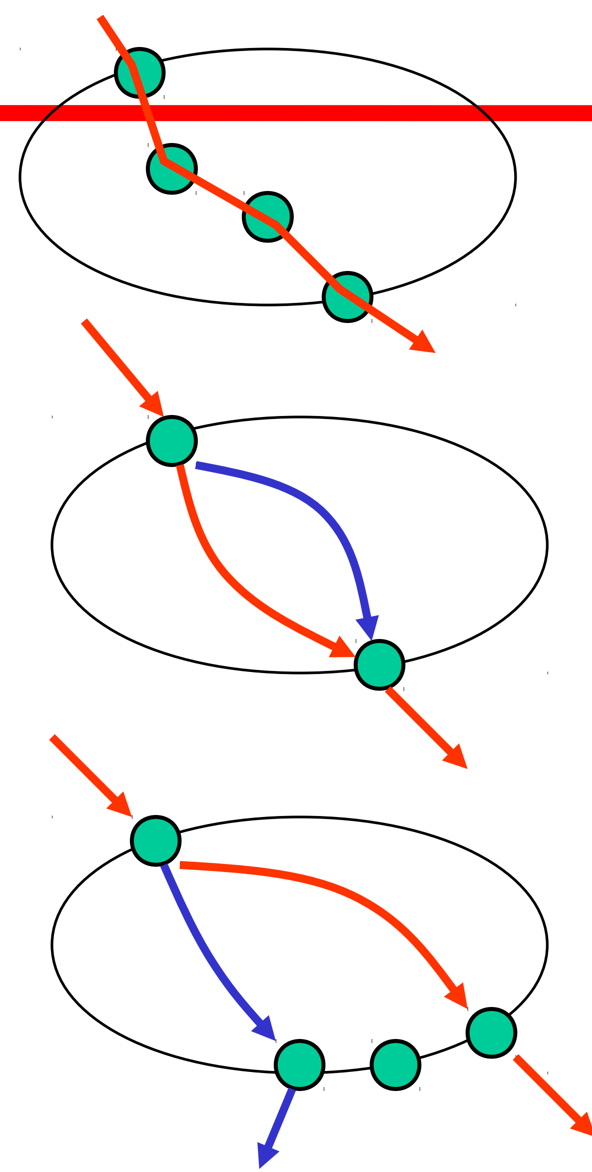
Ideally, captures all the information that is **invariant** with respect to the network state



demand matrix = bytes per source-destination pair

Domain-Wide Traffic & Demand Models

current state & traffic flow



fine grained: path matrix = bytes per path

predicted control action: impact of intra-domain routing

intradomain focus: traffic matrix = bytes per ingress-egress

predicted control action: impact of inter-domain routing

interdomain focus: demand matrix = bytes per ingress and set of possible egresses

Traffic Representations

- Network-wide views
  - Not directly supported by IP (stateless, decentralized)
  - Combining elementary measurements: traffic, topology, state, performance
  - Other dimensions: time & time-scale, traffic class, source or destination prefix, TCP port number
- Challenges
  - Volume
  - Lost & faulty measurements
  - Incompatibilities across types of measurements, vendors
  - Timing inconsistencies
- Goal
  - Illustrate how to populate these models: data analysis and inference
  - Discuss recent proposals for new types of measurements

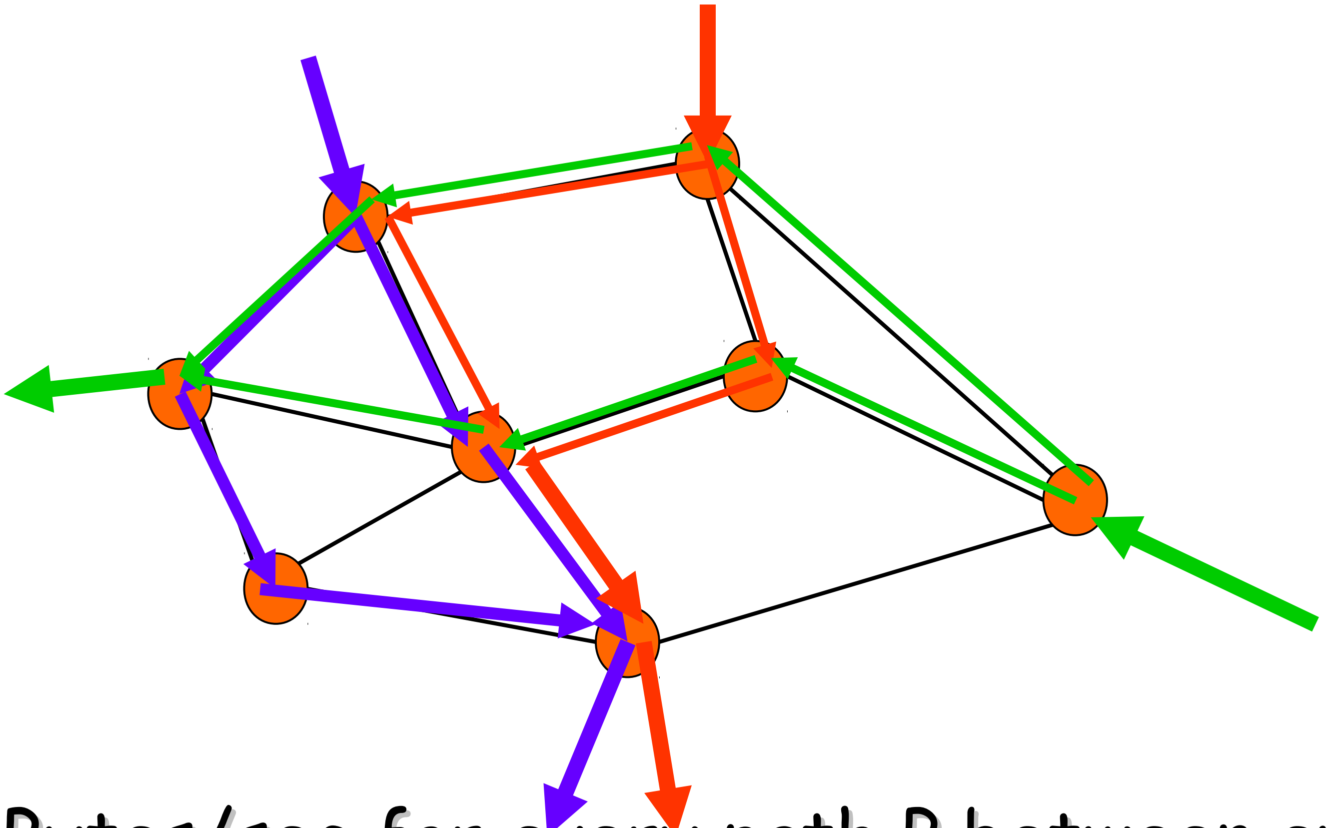
Outline

- Path matrix
- Traffic matrix
  - Network tomography
- Demand matrix
  - Combining flow and routing data

Path Matrix: Operational Uses

- Congested link
  - Problem: easy to detect, hard to diagnose
  - Which traffic is responsible?
  - Which customers are affected?
- Customer complaint
  - Problem: customer has insufficient visibility to diagnose
  - How is the traffic of a given customer routed?
  - Where does it experience loss & delay?
- Denial-of-service attack
  - Problem: spoofed source address, distributed attack
  - Where is it coming from?

Path Matrix

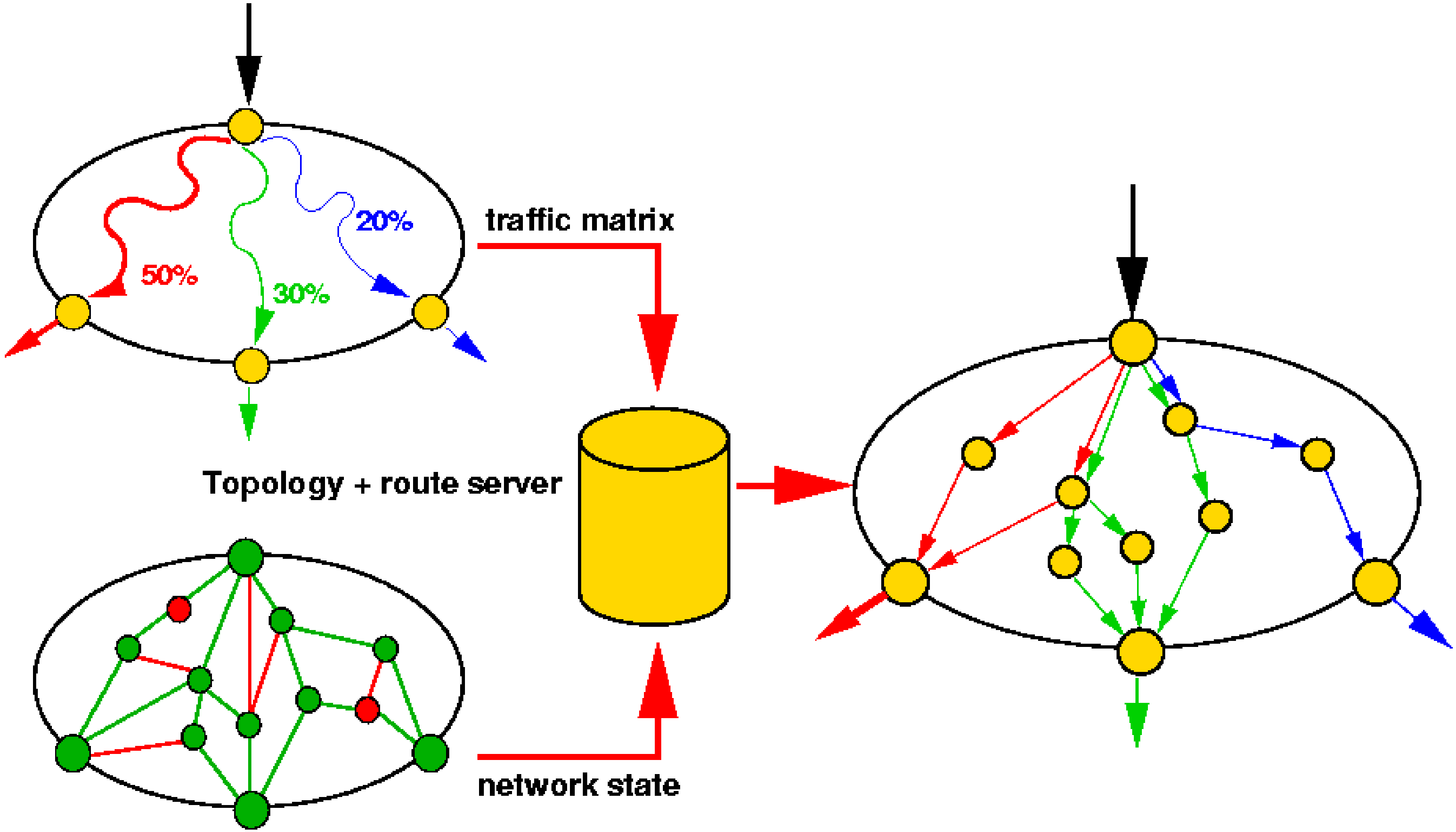


- Bytes/sec for every path P between every ingress-egress pair
- Path matrix ⇒traffic matrix

Measuring the Path Matrix

- Packet or flow measurement on every link
  - Combine records to obtain paths
  - Drawback: excessive overhead, difficulties in matching up flows
- Combining packet/flow measurements with network state
  - Measurements over cut set (e.g., all ingress routers)
  - Dump network state
  - Map measurements onto current topology

Path Matrix through Indirect Measurement



Outline

- Path matrix
  - Traffic matrix
    - Network tomography
- Demand matrix
  - » Combining flow and routing data

Traffic Matrix: Operational Uses

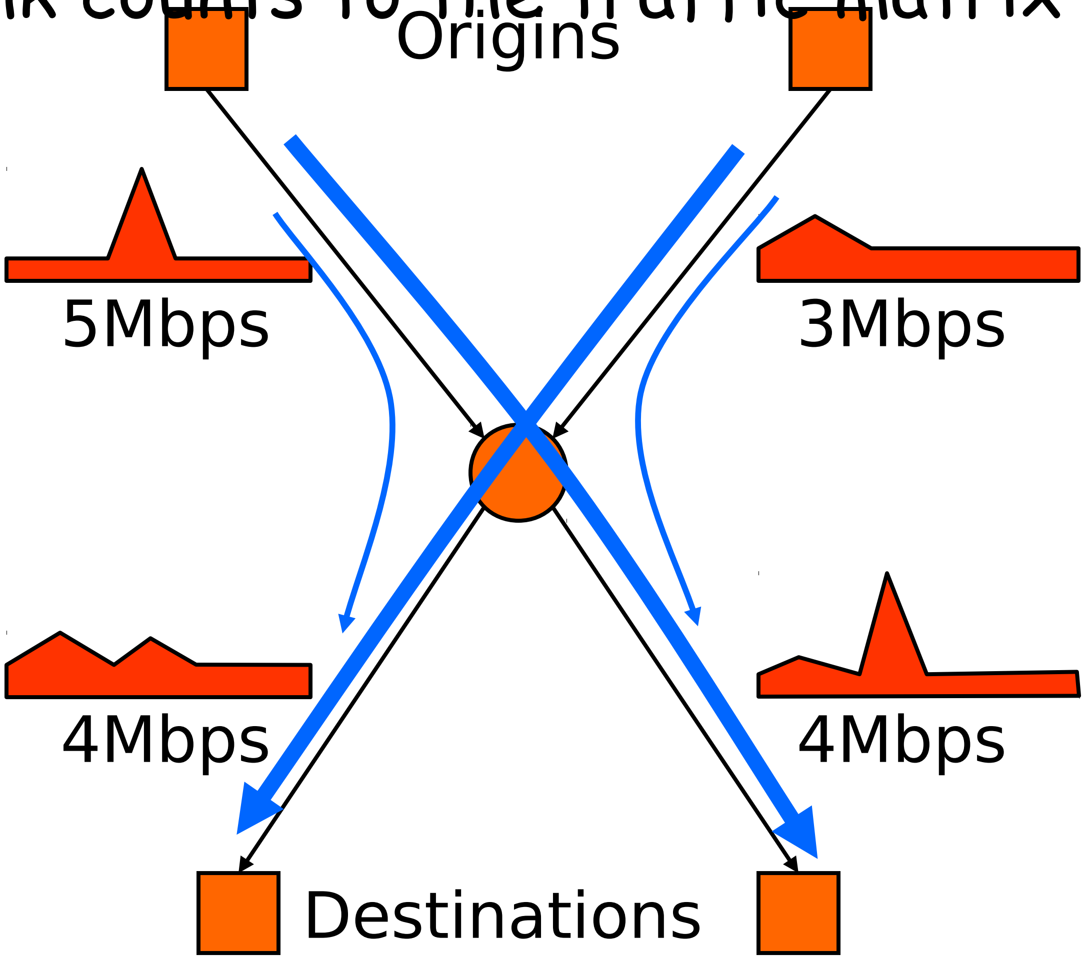
- Short-term congestion and performance problems
  - Problem: predicting link loads and performance after a routing change
  - Map traffic matrix onto new routes
- Long-term congestion and performance problems
  - Problem: predicting link loads and performance after changes in capacity and network topology
  - Map traffic matrix onto new topology
- Reliability despite equipment failures
  - Problem: allocating sufficient spare capacity after likely failure scenarios
  - Find set of link weights such that no failure scenario leads to overload (e.g., for "gold" traffic)

Obtaining the Traffic Matrix

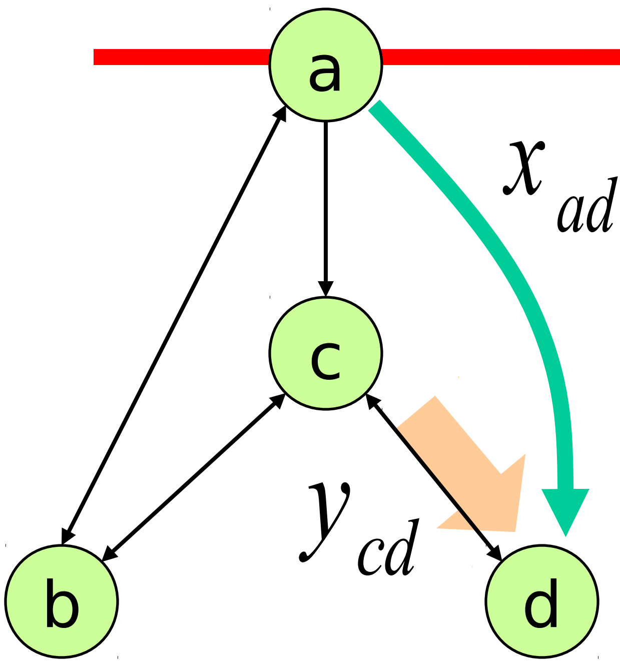
- Tomography:
  - » Assumption: routing is known (paths between ingress-egress points)
  - » Input: multiple measurements of link load (e.g., from SNMP interface group)
  - » Output: statistically inferred traffic matrix

Network Tomography

From link counts to the traffic matrix



Matrix Representation



$$A=\begin{bmatrix}
1 & & & & & & & & & & & \\
 & 1 & 1 & & & 1 & & & & & & \\
 & & & 1 & & 1 & 1 & & & 1 & & \\
 & & & & 1 & & & & & & & \\
 & & & & & & 1 & 1 & & 1 & 1 & \\
 & & 1 & & & 1 & & & 1 & & & \\
 & & & & & & & & & 1 & 1 & 1
\end{bmatrix}$$

$x=(x_1,\ldots,x_c)^T$:OD counts

$y=(y_1,\ldots,y_r)^T$:link counts

$$y=Ax$$

Single Observation is Insufficient

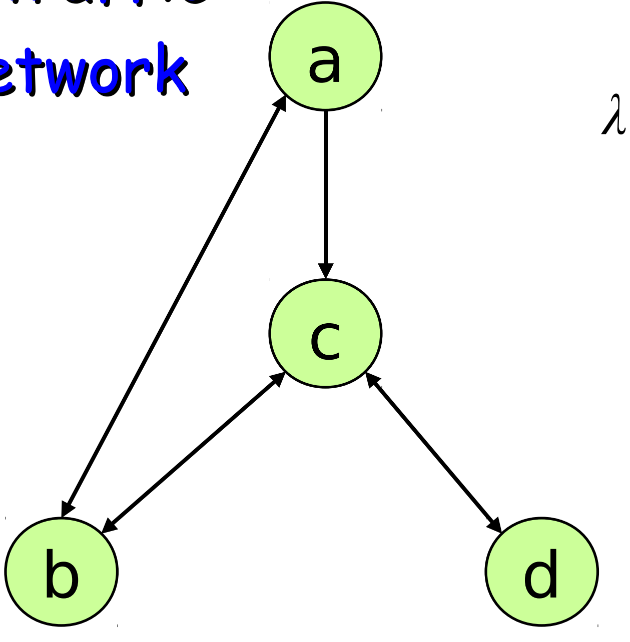
- Linear system is underdetermined
  - Number of links $r \approx O(n)$
  - Number of OD pairs $c \approx O(n^2)$
  - Dimension of solution sub-space at least $c - r$
- Multiple observations are needed
  - Stochastic model to bind them

Network Tomography

- [Y. Vardi, Network Tomography, JASA, March 1996]
- Inspired by road traffic networks, medical tomography
- Assumptions:
  - OD counts: $X_j^{(k)} \equiv \text{Poisson}(\lambda_j)$
  - OD counts i.i.d.
  - *K* independent observations $Y^{(1)}, \ldots, Y^{(K)}$
- MLE Estimators
- Method of Moments

How Well does it Work?

- Experiment [Vardi]:
  - K=100
- Limitations:
  - Poisson traffic
  - Small network

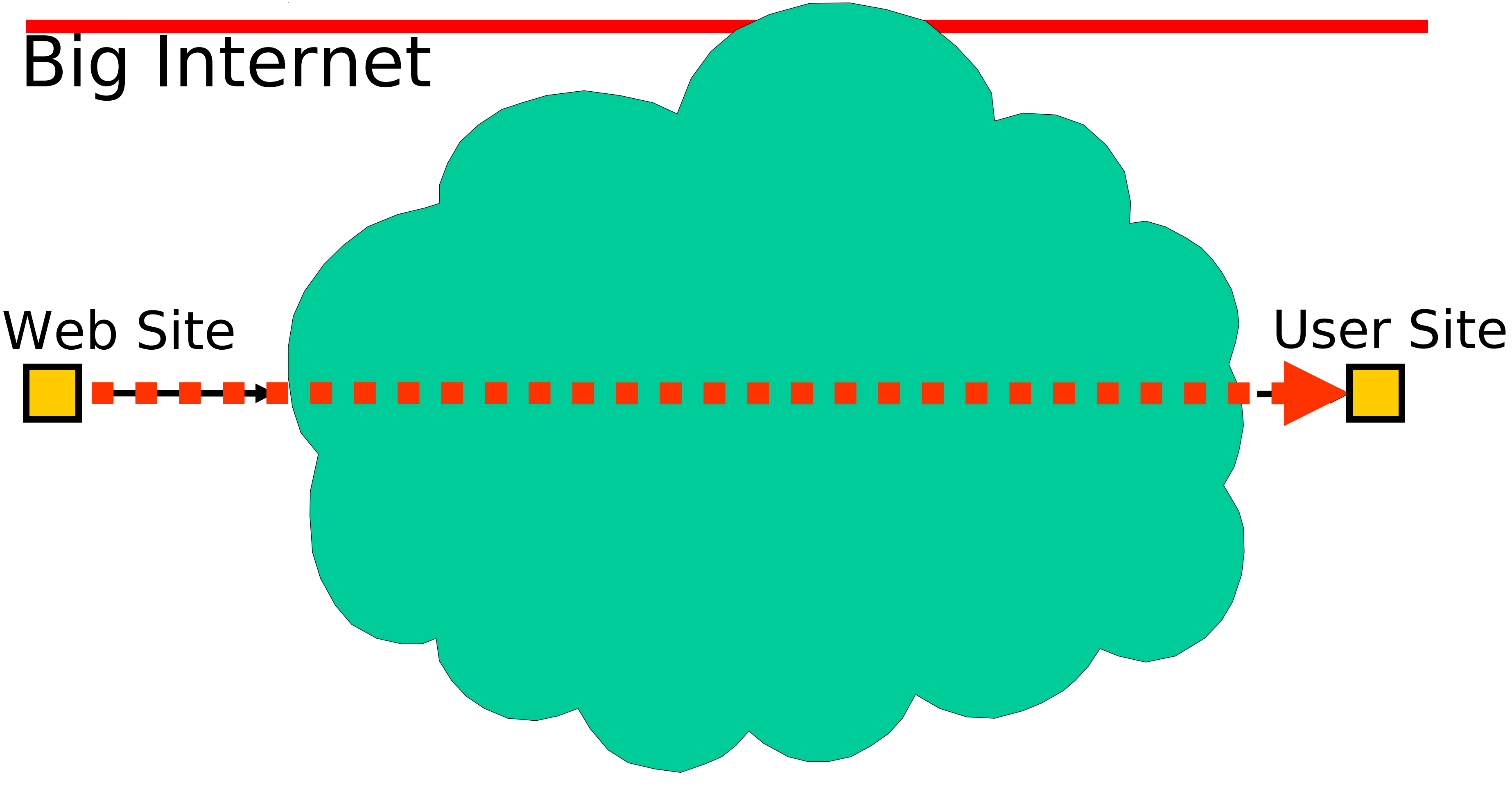


$$\lambda = EX = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{bmatrix}, \hat{\lambda} = \begin{bmatrix} 1.01 \\ 2.37 \\ 2.68 \\ 4.72 \\ 5.06 \\ 5.79 \\ 6.84 \\ 7.92 \\ 9.25 \\ 9.87 \\ 11.33 \\ 12.14 \end{bmatrix}$$

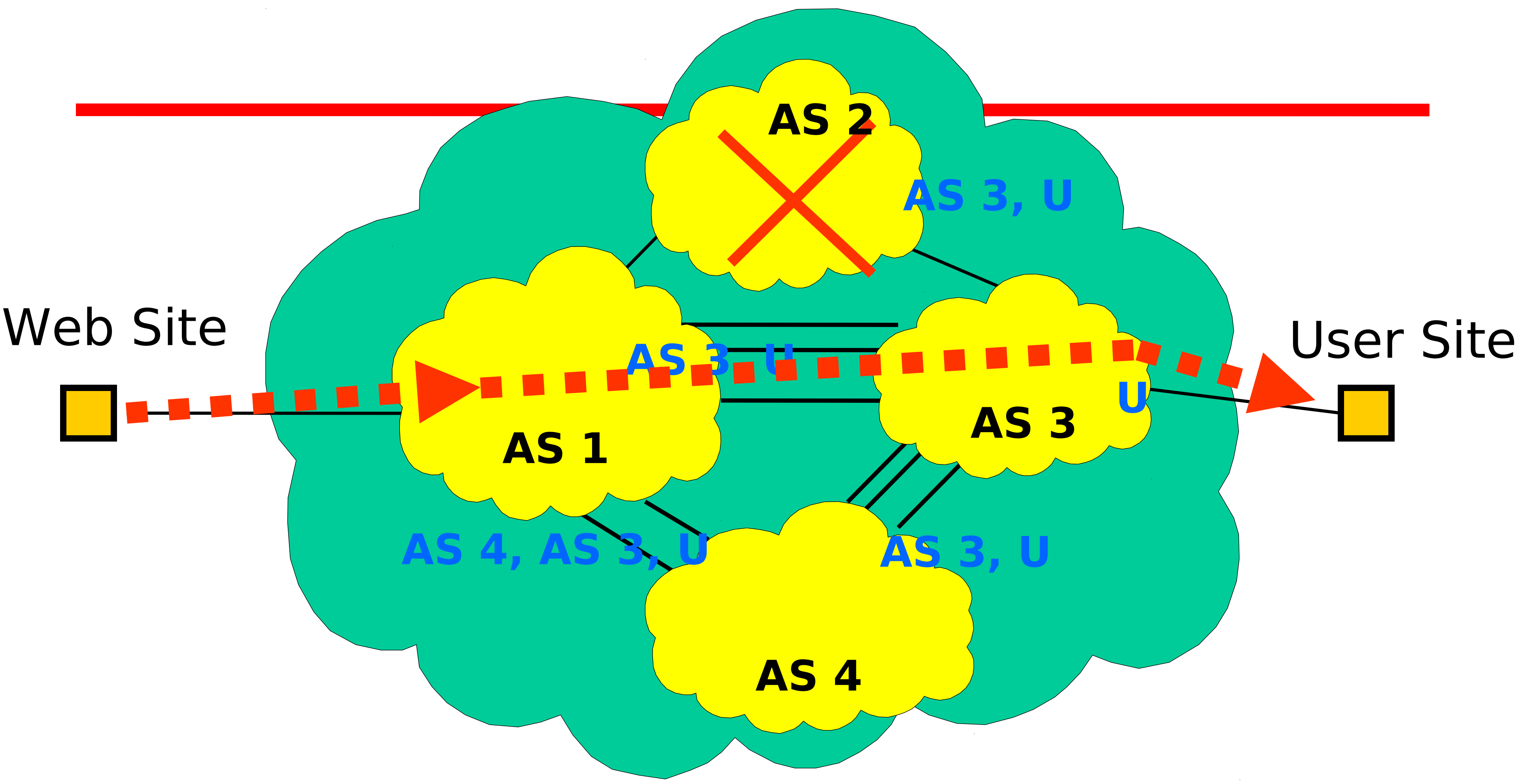
Outline

- Path matrix
- Traffic matrix
  - Network tomography
- Demand matrix
  - Combining flow and routing data

Traffic Demands



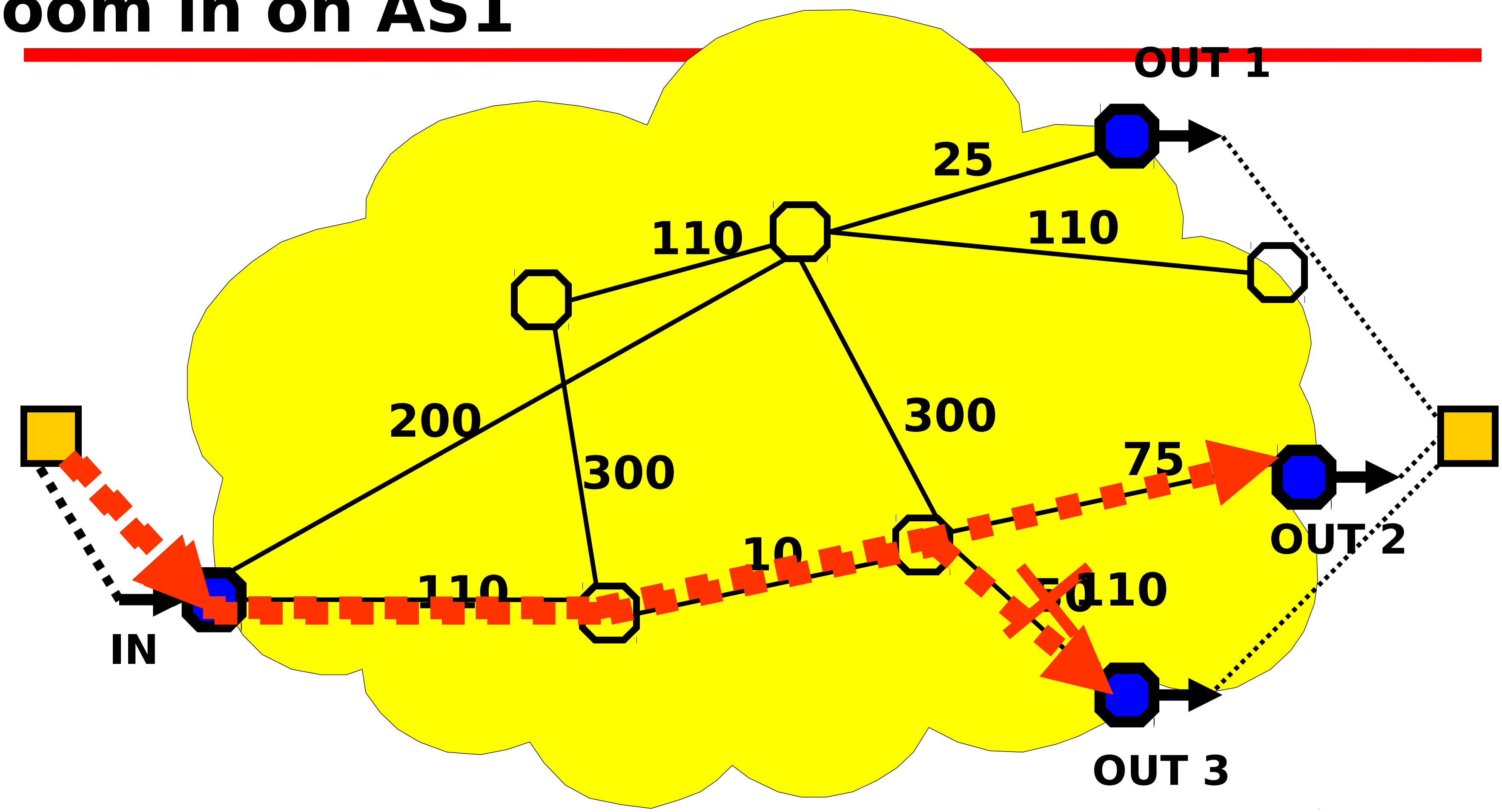
Coupling between Inter and Intradomain



- IP routing: first interdomain path (BGP), then determine intradomain path (OSPF,IS-IS)

Intradomain Routing

Zoom in on AS1



Change in internal routing configuration changes flow exit po
(hot-potato routing)

Demand Model: Operational Uses

- Coupling problem with traffic matrix-based approach:

Traffic matrix
↓
Traffic Engineering
↓
Improved Routing
↗
Traffic matrix
↓
Traffic Engineering
↓
Improved Routing
↗

  - traffic matrix changes after changing intradomain routing!

- Definition of demand matrix: # bytes for every (in, {out_1,...,out_m})
  - » ingress link (in)
  - » set of possible egress links ({out_1,...,out_m})

Demand matrix
↓
Traffic Engineering
↓
Improved Routing

Ideal Measurement Methodology

- Measure traffic where it enters the network
  - » Input link, destination address, # bytes, and time
  - » Flow-level measurement (Cisco NetFlow)
- Determine where traffic can leave the network
  - » Set of egress links associated with each destination address (forwarding tables)
- Compute traffic demands
  - » Associate each measurement with a set of egress links

Traffic Engineering: Summary

- Traffic engineering requires domain-wide measurements + models
  - Path matrix (per-path): detection, diagnosis of performance problems; denial-of-service attacks
  - Traffic matrix (point-to-point): predict impact of changes in intra-domain routing & resource allocation; what-if analysis
  - Demand matrix (point-to-multipoint): coupling between interdomain and intradomain routing: multiple potential egress points

Conclusion

- IP networks are hard to measure by design
  - Stateless and distributed
  - Measurement support often an afterthought → insufficient, immature, not standardized
- Network operations critically rely on measurements
  - Short time-scale: detect, diagnose, fix problems in configuration, state, performance
  - Long time-scale: capacity & topology planning, customer acquisition, ...
- There is much left to be done!
  - Instrumentation support; systems for collection & analysis; procedures