Large-scale Networks: Some Open Challenges

D. Towsley

U. Massachusetts

Outline

- background
- challenges
  - network tomography
  - modeling, analysis, control
  - network information theory
- summary

Overview

- millions of end hosts
- 100,000s of routers
- 10,000 autonomous systems (ASes)
  - independently administered
  - heterogeneous
- applications:
  - Web (text, images, audio, video)
  - IP telephony
  - peer2peer appl. (e.g.,Napster)
  - ...

Overview

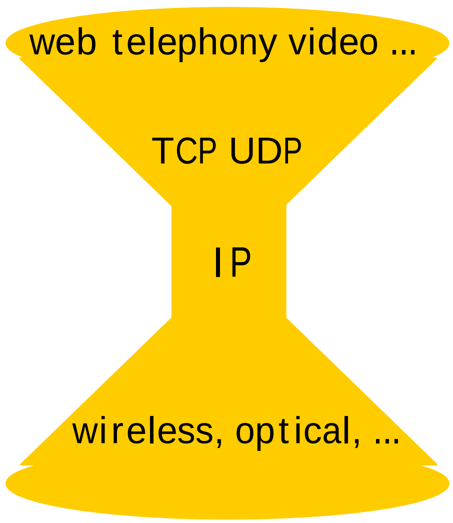
- … nd hosts
- 100, … ers
- 10,000 a … stems (ASes)
  - independen …
  - heterogeneous
- applications:
  - Web (text, images, audio, …
  - IP telephony
  - peer2peer appl. (e.g.,Napster)
  - ...

continuous, unpredictable change

Exponential growth,

Background

- Internet work in progress
  - no centralized authority
- built on IP architecture
  - flexibility/robustness vs. efficiency/optimization
- IP hourglass
  - common IP protocol
  - many protocols, applications
  - many low-level technologies



TCP

- transport protocol over IP
- carries >90% of data in Internet
- provides
  - reliable, in-order delivery to application
  - congestion control to network
    - additive increase, multiplicative decrease (AIMD)
    - expects no network support
      - indication of available resources
      - congestion notification
      - slowly changing (explicit congestion notification)

TCP

- window-based algorithm
  - self-clocking
- packet loss signals congestion
- additive increase multiplicative decrease
  - halve congestion window at loss
  - increase window by one per RTT
- heavy congestion produces exponentially increasing timeouts
- slow start: double window every RTT

BGP (Border Gateway Protocol)

- used to set up routes between ASes (agreed upon by all)
- BGP (Border Gateway Protocol)
- distance vector algorithm
- ~100,000 entries in table; exponential growth
- 10+ sec convergence time (and growing)

Challenges: understanding network behavior

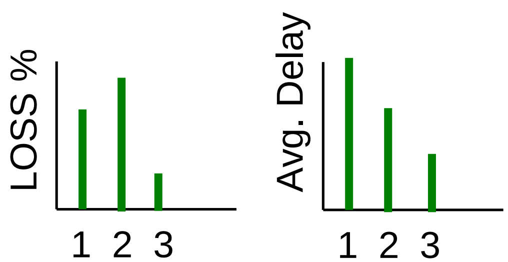
- network tomography
  - internal from end-end observations
  - end-end from internal observqtions
- modeling, analysis, control

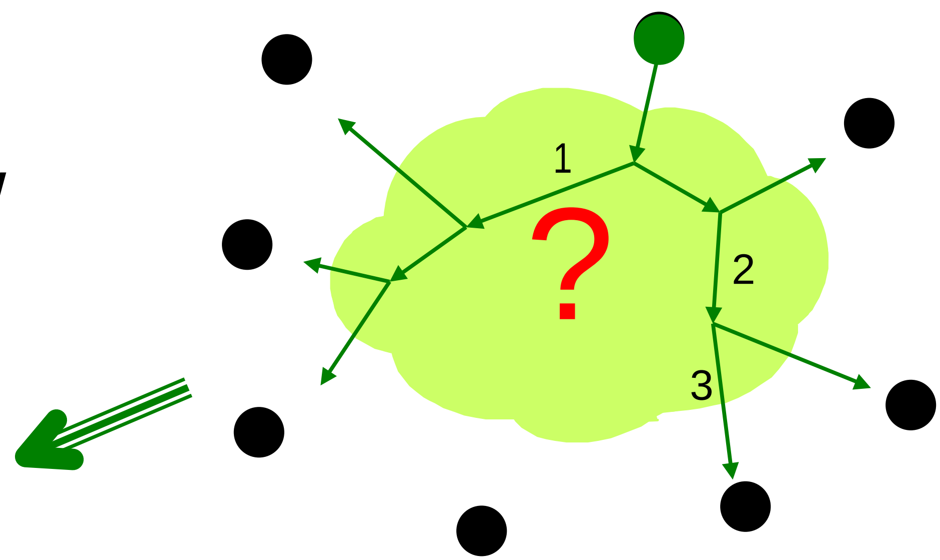
Challenge: network tomography

Goal 1: obtain spatial-temporal picture of a network/internet from end-to-end views

Infrastructure

cross section view



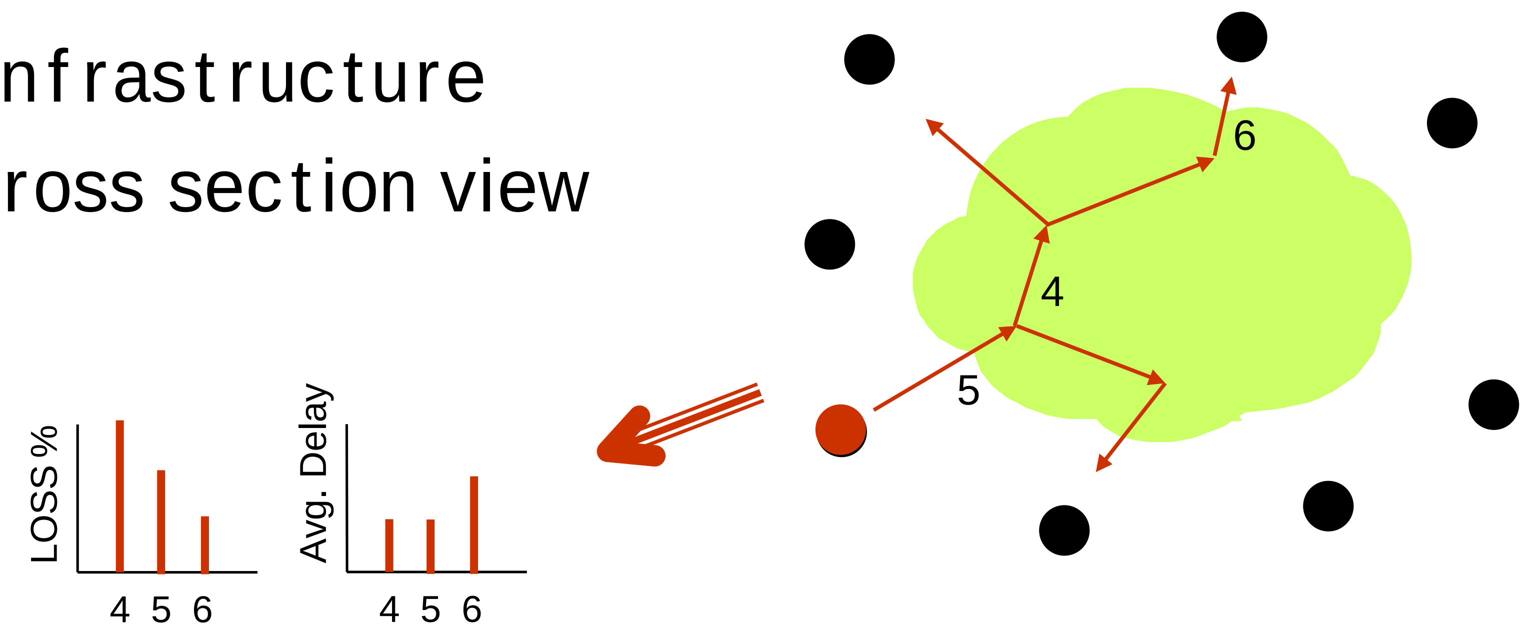


Challenge: network tomography

Goal 1: obtain spatial-temporal picture of a network/internet from end-to-end views

Infrastructure

cross section view

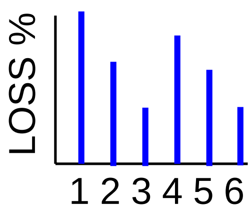


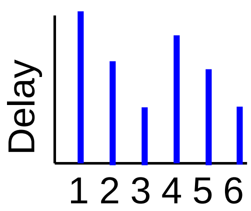
Challenge: network tomography

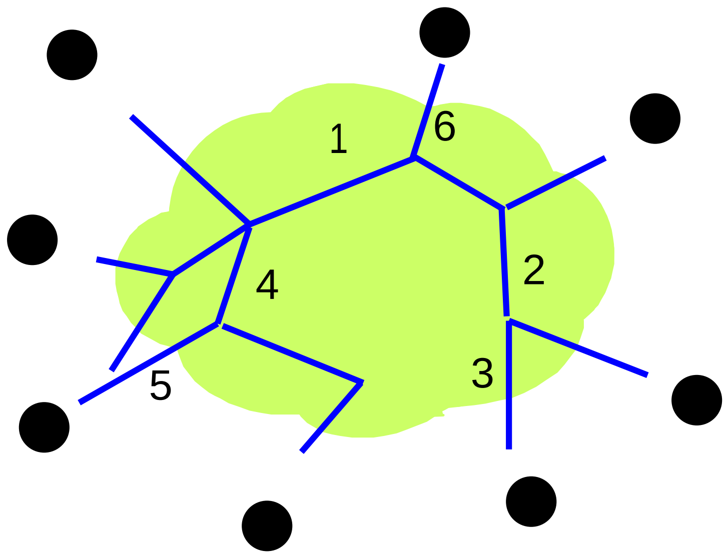
Goal 1: obtain spatial-temporal picture of a network/internet from end-to-end views

Infrastructure

composition of views



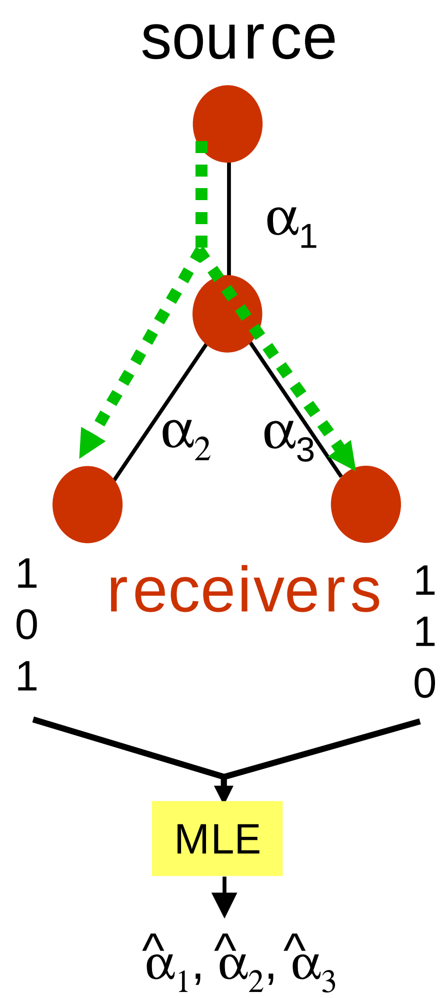




Example: MINC

Uses end-end observations to peer inside network

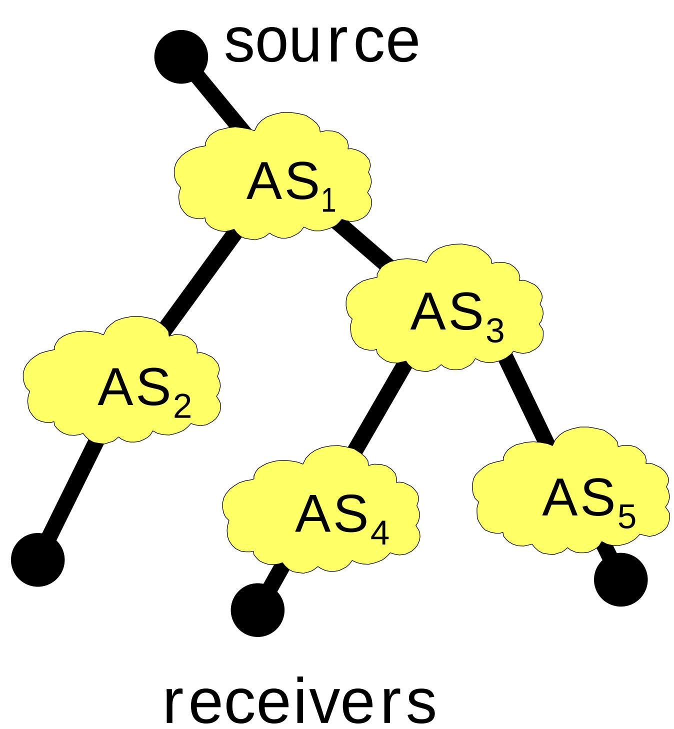
- multicast probes
- correlated performance observed by receivers
- maximum likelihood estimators for
  - loss rates
  - delay statistics
  - tree topology



Open Problems

- data massaging
  - delay measurements exhibit many artifacts
- scalability
  - works for 10 - 50 routers; 1000? different ASes?
- layout/composition of views
- use (partial) information from network

AS-level tomography

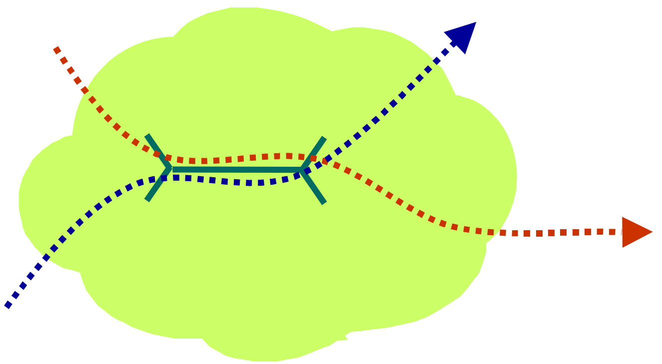


The *Real* Internet

- most traffic point-point
- point-to-point behavior differs from multicast behavior

Q: how to infer internal network behavior, structure from point-to-point measurements?

- introduce/exploit correlation



Network Tomography

Goal II: obtain detailed picture of end-end behavior from internal views

- network design
- traffic engineering

From internal observations, infer

- end-end application-level behavior
- traffic flows
- workload characterization
- spatial/temporal correlation

Network Traffic

Invariants:

- 24/7 cycles
- log-normal/pareto connection size distributions
- Poisson arrivals for sessions
- complex correlation structures at fine/coarse time scales

Traffic Flows

- correlation structure
  - what and how due to: protocols? workloads
  - how to test/validate hypotheses?
- spatial correlations?

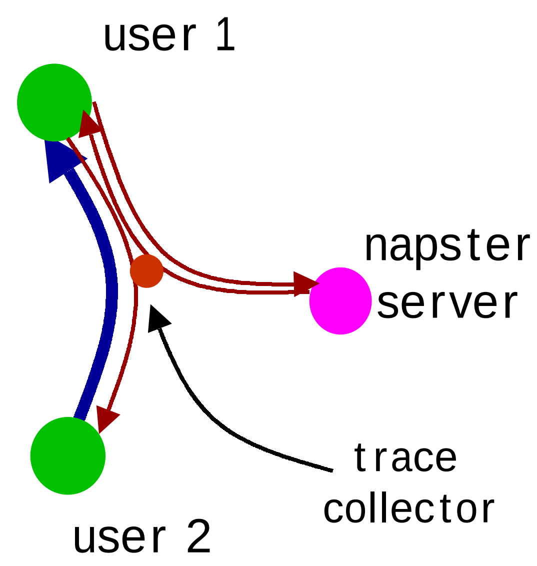
- parameter estimation?
  - at gigabit speeds
    - sampling? inferencing?

Traffic Analysis

- Many applications have well known protocol/port (in packet header)
  - 21 - ftp, 23 - telnet, 80 - http
- Some dynamically select ports
  - napster, gnutella
  - some games
  - denial of service attack

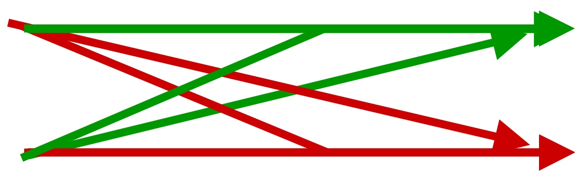
Traffic Analysis

- application signatures
  - napster, gnutella
    - service location/data transfer pattern?
  - games
- denial of service attack signatures



Summary

- observations
  - internal
  - end-end



- views
  - internal
  - end-end
- theory needed to make this happen
- correlation, correlation!
- scalability, scalability!
- robustness, robustness!

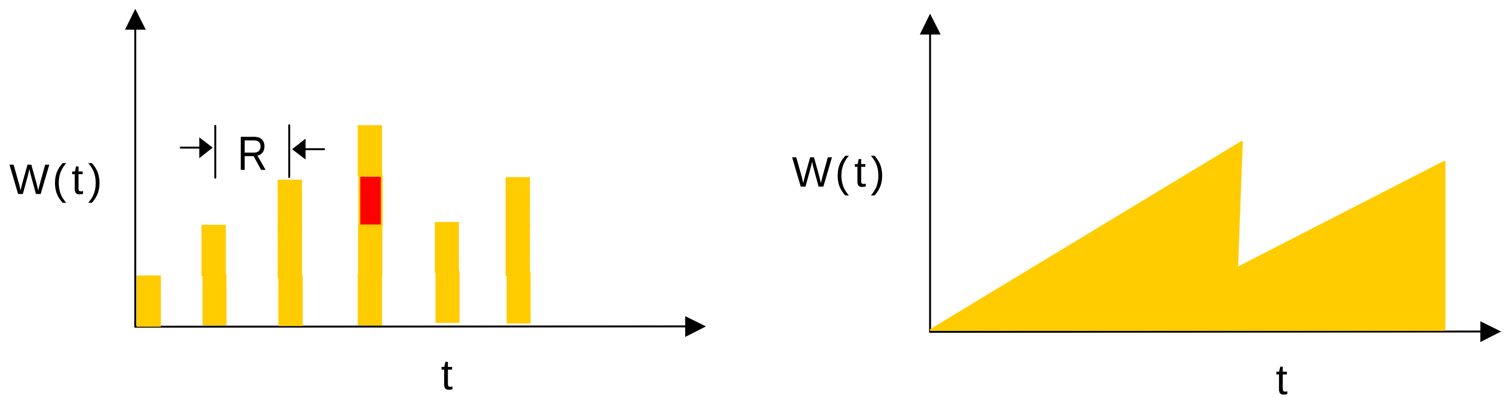
Challenge: performance evaluation

How to deal with

- protocol complexity
- interaction between protocols
  - routing, TCP
  - closed loop, open loop
- network size, heterogeneity
- workload characteristics
  - mice, elephants

Example: Fluid Model

- fluid traffic model for TCP
  - window size W(t); round trip time R



Example:
Stochastic Differential Equations

Long duration TCP flow (elephant)

Window size: $$dW_i = \frac{dt}{R_i(t)} - \frac{W_i}{2} dN_i(t - R_i(t))$$

Additive increase **Mult. decrease**

$W_i(t)$: window size for flow i

$N_i(t)$: counting process (losses)

$R_i(t)$: feedback delay (round trip time)

SDEs

Window size: $$dW_i = \frac{dt}{R_i(t)} - \frac{W_i}{2} dN_i(t - R_i(t))$$

Queue length: $$dq = -1\{q(t) > 0\} C\, dt + \sum \frac{W_i(t)}{R_i(q(t))} dt$$

Outgoing traffic

Incoming traffic

C : link capacity

SDEs

- how to solve?
  - closed form? numerically
- timeouts?
- large networks $\Rightarrow$ 100,000s of SDEs
  - solution techniques?
  - solution representation?
  - other approaches?
- mice as well as elephants?

... to ODEs

- what are the ODEs for average quantities?
- in what sense are they limits of the SDEs?
- convergence rates? error bounds?
- large networks $\Rightarrow$ 100,000s of ODEs
  - how to solve?
  - how to represent solution?
  - other techniques? mean field approaches?
- mice?
- Kelly, LeBoudec, Srikant, ...

... to Fixed Point Problems

- fixed point problem comes from
  - limits of ODEs
  - heuristics
- existence/uniqueness of solutions?
- solutions for 100,000s entities?
- mice?
  - partial results Roberts, Massoulie, ...

Network Control: congestion control

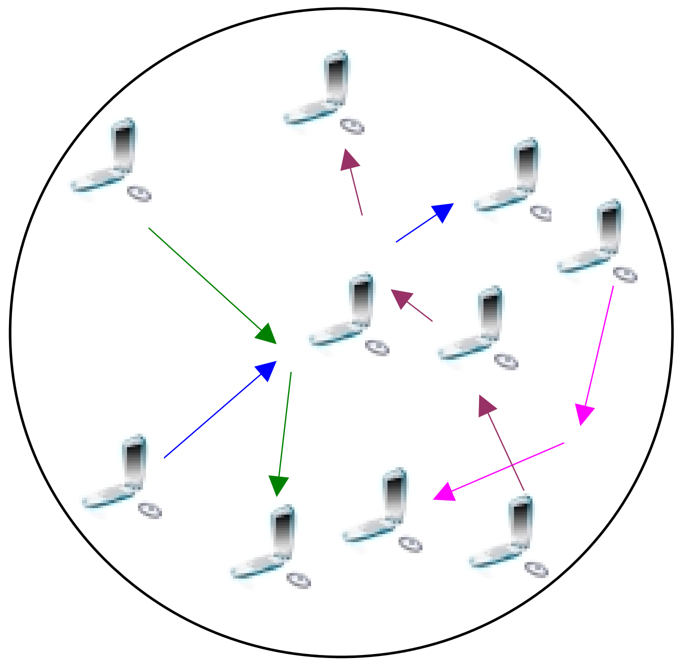
- ODEs - language of control theory
- ODEs associated with distributed optimization problem
  - utility functions
    - delay sensitive?
  - fairness criteria
- stochastics?
- mice?
- Kelly, Low, Misra, Doyle, ...

Challenge: wireless network information theory

- n wireless nodes in area A
- physical model (signal2noise ratio)

Q: maximum thruput

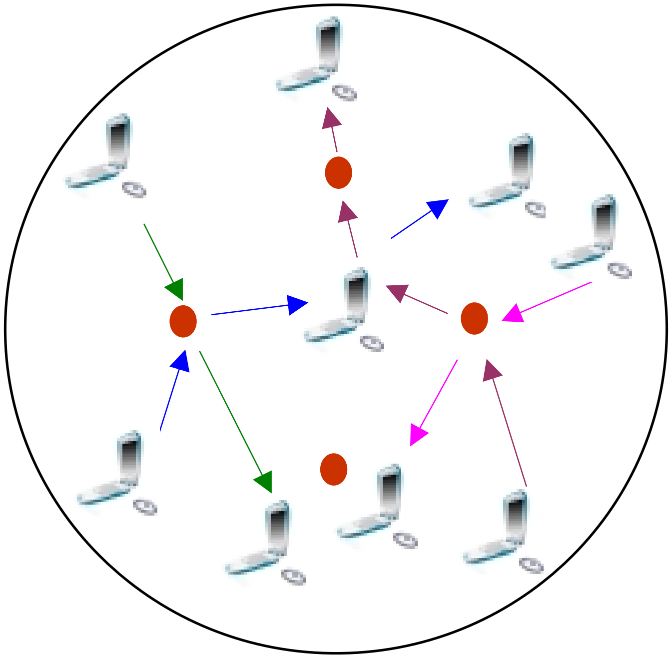
- per node
- per unit area
- scaling with n and A



some results for random, stationary case, n→∞ (Gupta, Kumar)

Wireless Networks

- non-uniform random?
  - node placement, obstacles
- $n \to \infty$, $A \to \infty$, n/A constant?
  - phase transition at some n/A; addition of fixed routers?



Wireless Networks

- non-uniform random?
  - node placement, obstacles
- $n \to \infty$, $A \to \infty$, n/A constant?
  - phase transition at some n/A; addition of fixed routers?
- mobile nodes?
  - store/carry/forward
    - some results for infinite delay (Grossglauser, Tse); delay constraint?
    - addition of fixed routers? mobile routers?
- how to realize bounds?
  - routing algorithm
  - capacity vs. delay

Other Challenges

- interdomain routing
- security
- overlay networks

Summary

- size, complexity of network poses significant challenges
  - tomography
  - modeling, analysis and control
  - network information theory
- provides research opportunities in applied and theoretical mathematics, statistics
- partnerships between mathematicians and networkers can be very fruitful