Network Metrology: Introduction

WHY MEASURE NETWORKS?

"*In God we trust; all others must bring data*"

William Edwards Deming*

Network monitoring is essential for network operators

- Monitor service-level agreements
  - Performance of traffic through network
- Fault diagnosis
  - Quick detection of faults
  - Root cause analysis
- Security
  - Anomaly detection
  - Intrusion detection

Network monitoring is essential for users

- Verify network performance
  - Am I getting what I paid for?
- Select service provider, applications
  - What is the best provider in my neighborhood?
  - Which VoD service delivers best performance?
- Manage their network services

Network monitoring is essential for application/service developers

- Verify application performance
- Tuning applications to network conditions
  - E.g., adapt video rate to achieved throughput
- Server or path selection
  - Decide the best server for each client

Network monitoring is essential for regulators, policy makers

- Comparison among ISPs
- Verification of compliance to regulatory laws
  - Network neutrality
- Decisions about investments
  - Which areas need better infrastructure?

Network monitoring is essential for scientists

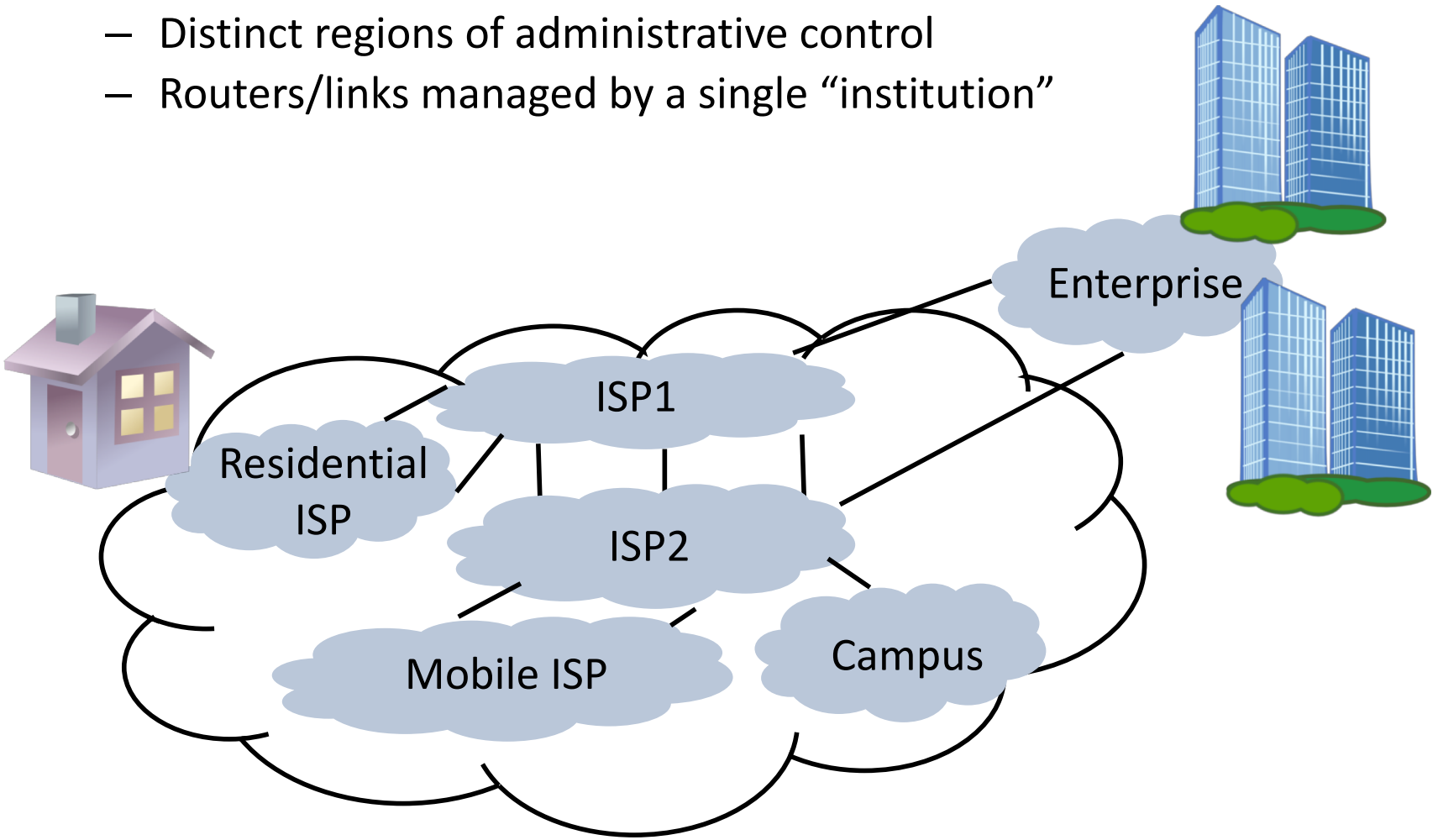
- Networking researchers
  - Evaluate and design of new systems/protocols
- Social/political scientists
  - Internet touches many aspects of society
- Physicists
  - Structure and dynamics of the Internet structure

WHAT ARE WE MEASURING?

The “Internet”

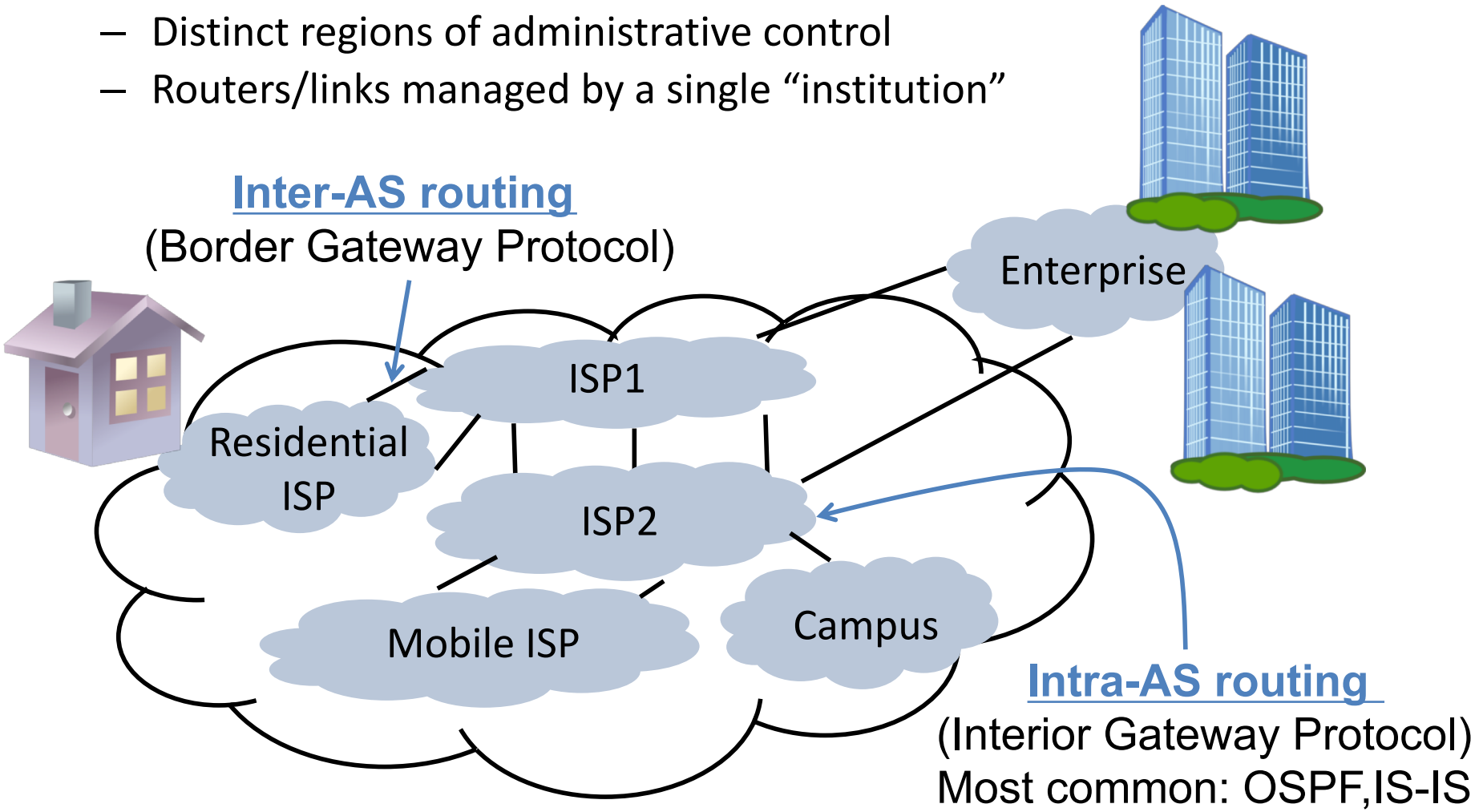
Internet: network of networks*

- Network = Autonomous System (AS)
  - Distinct regions of administrative control
  - Routers/links managed by a single “institution”

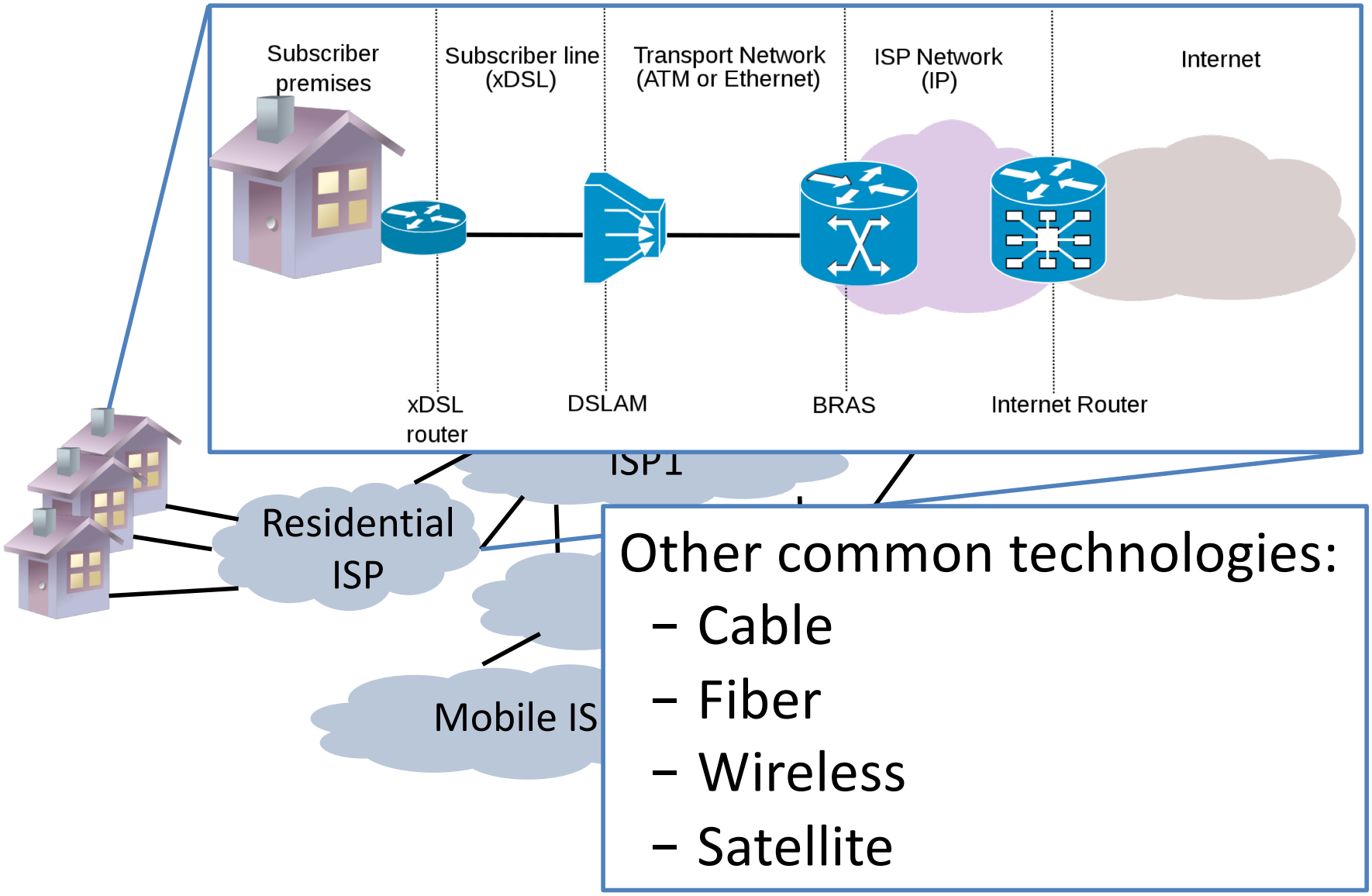


Internet: network of networks

- Network = Autonomous System (AS)
  - Distinct regions of administrative control
  - Routers/links managed by a single “institution”



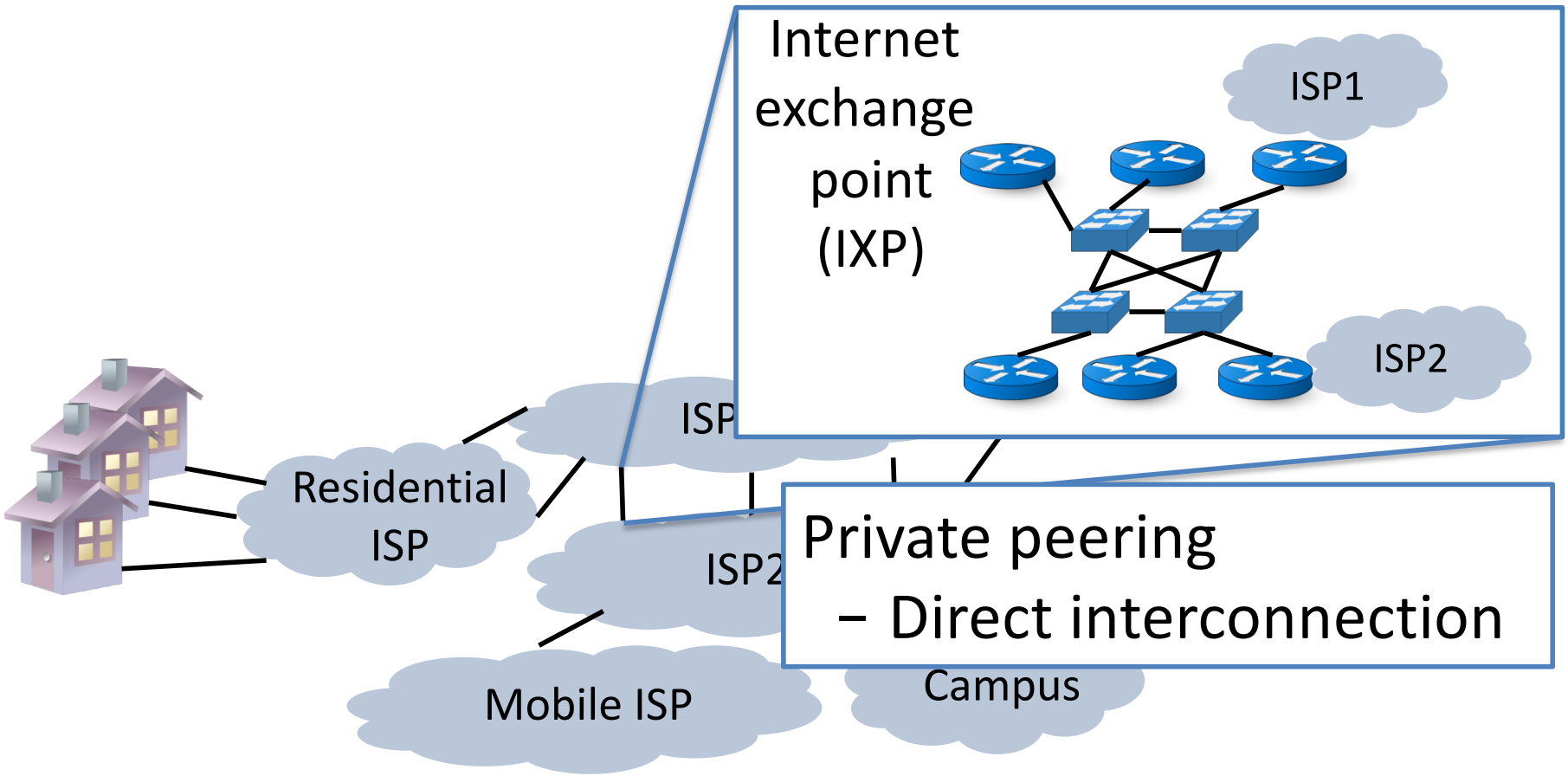
Internet technologies: Access



Other common technologies:

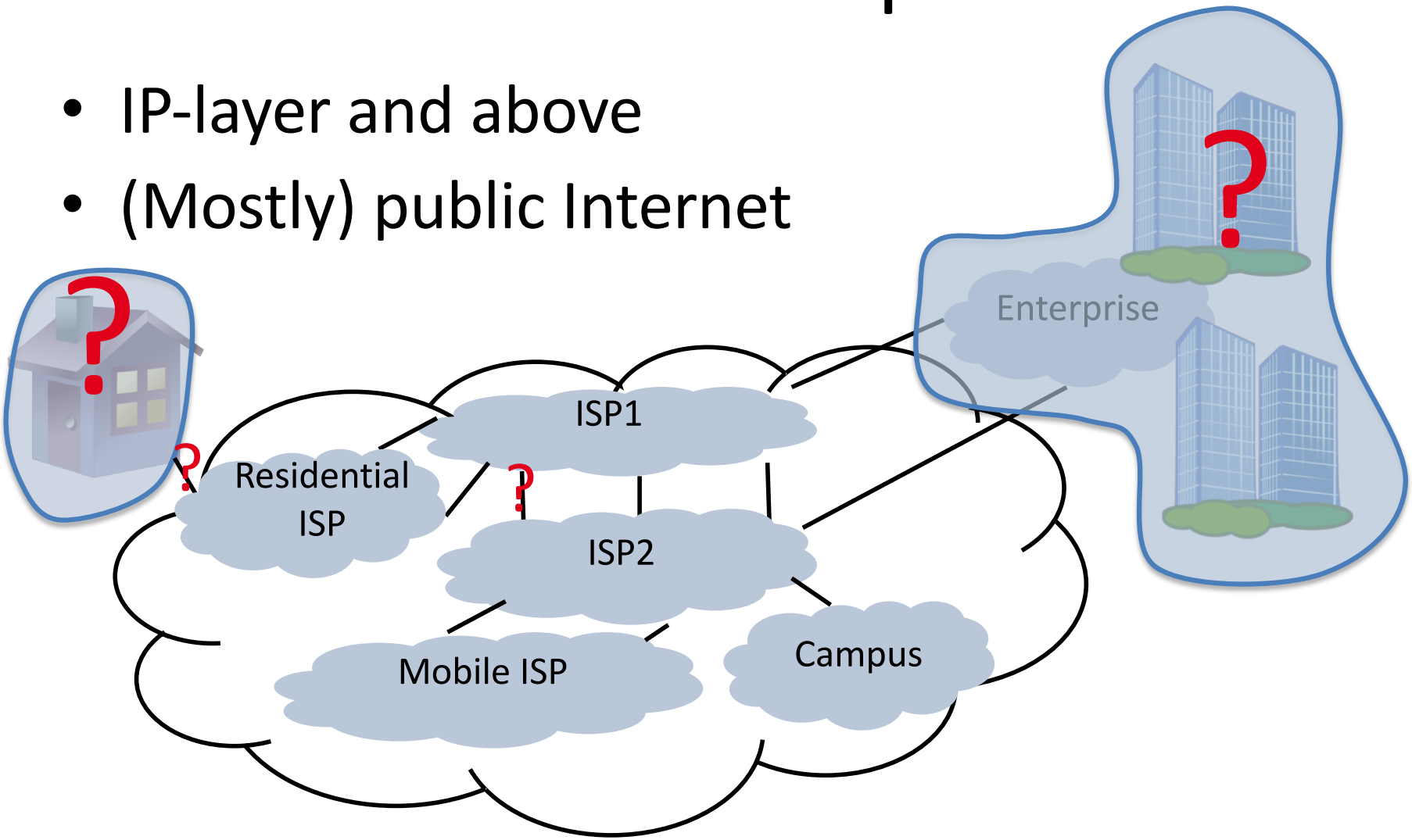
- Cable
- Fiber
- Wireless
- Satellite

Internet technologies: Inter-domain links



Class scope

- IP-layer and above
- (Mostly) public Internet

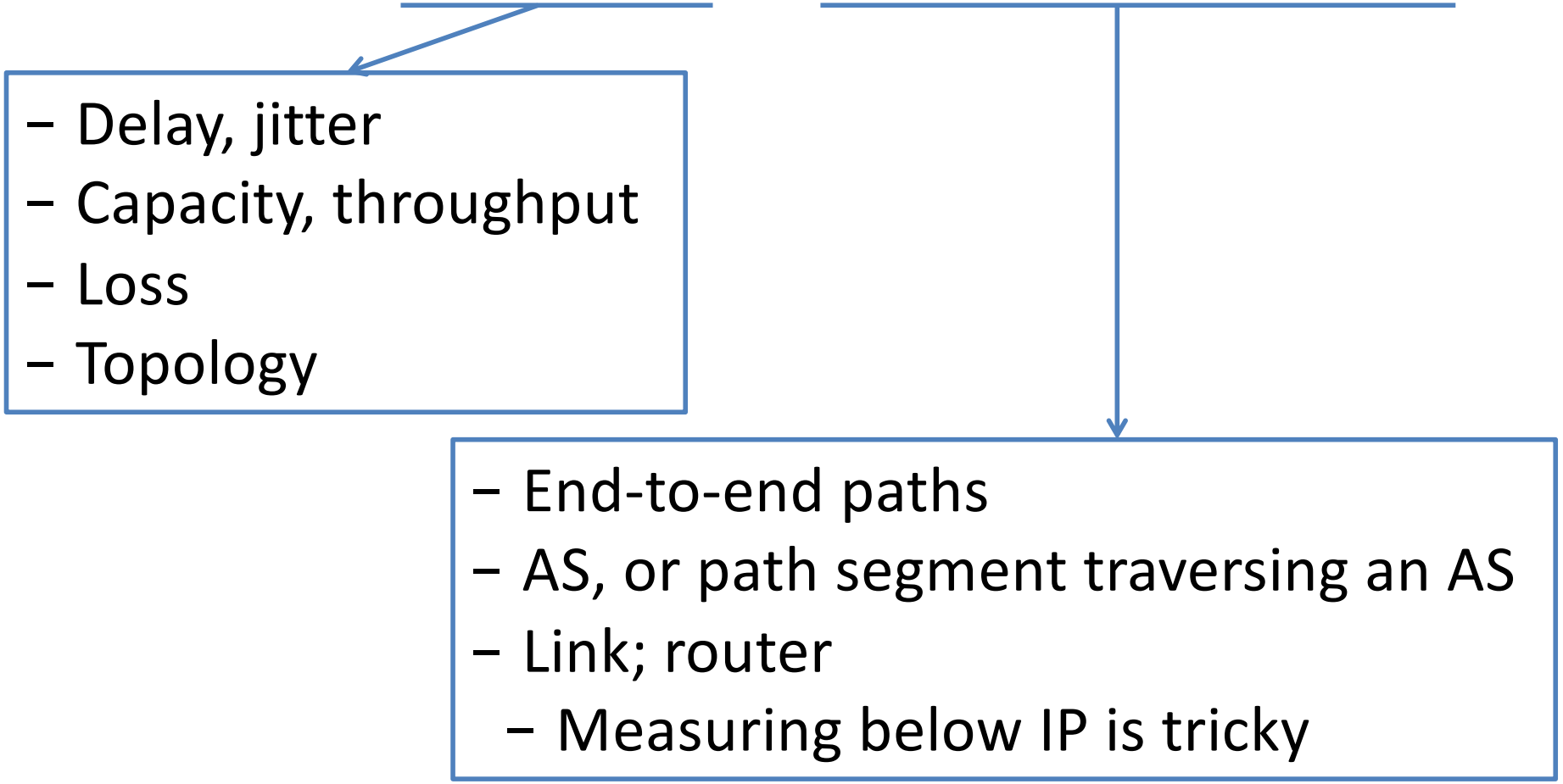


Class scope

- Measuring Internet Infrastructure
- Measuring Internet Traffic
- Measuring Internet Applications

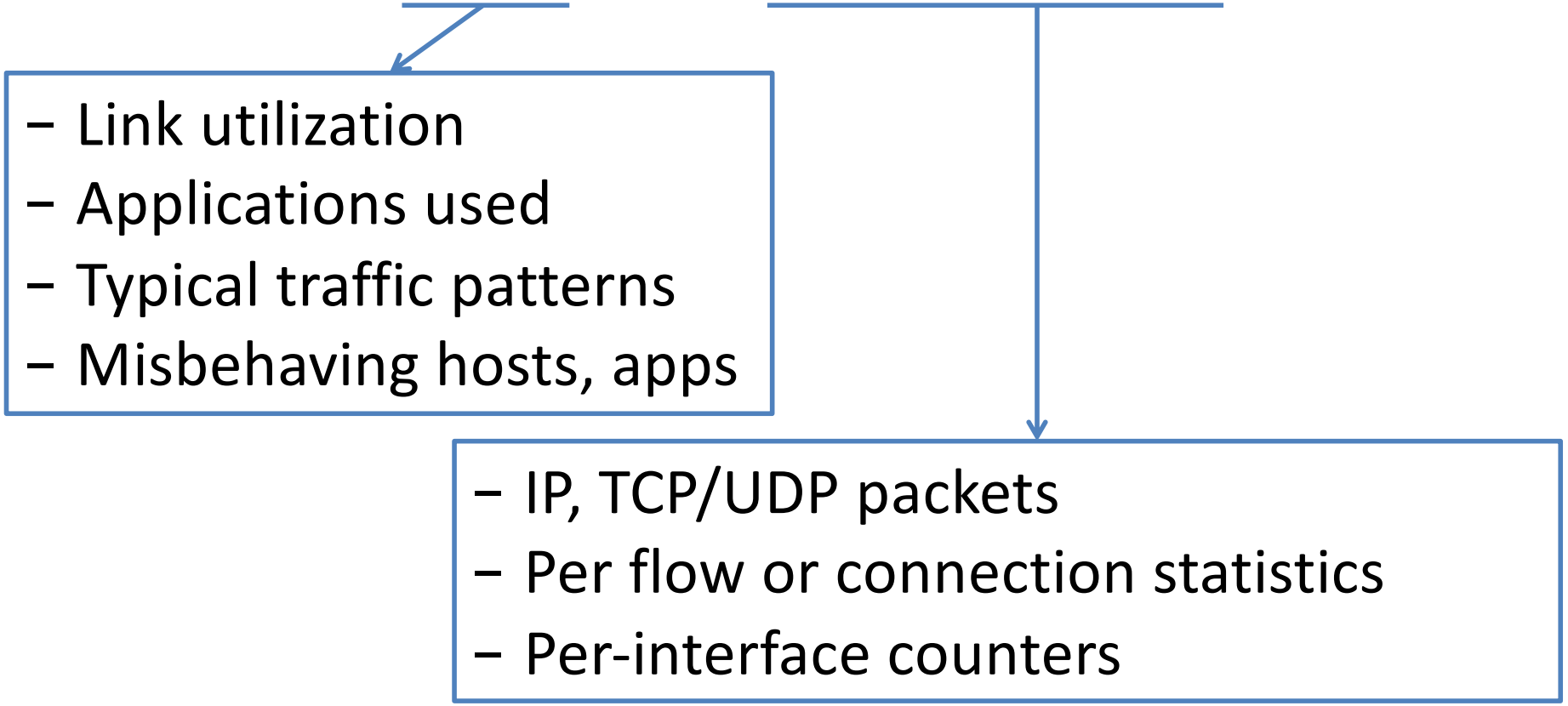
Measuring Internet infrastructure

- Goal: infer properties of parts of the network



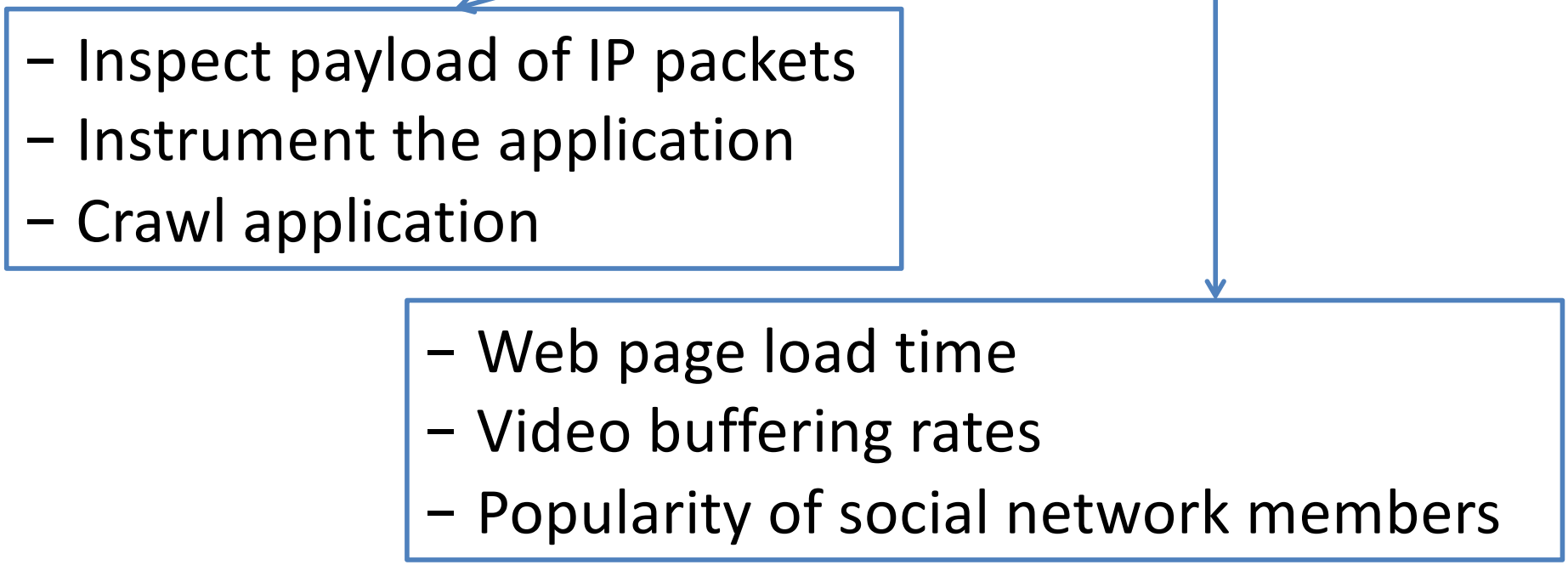
Measuring Internet traffic

- Goal: infer usage from network traffic



Measuring Internet applications

- Goal: infer application performance and usage from network traffic or application

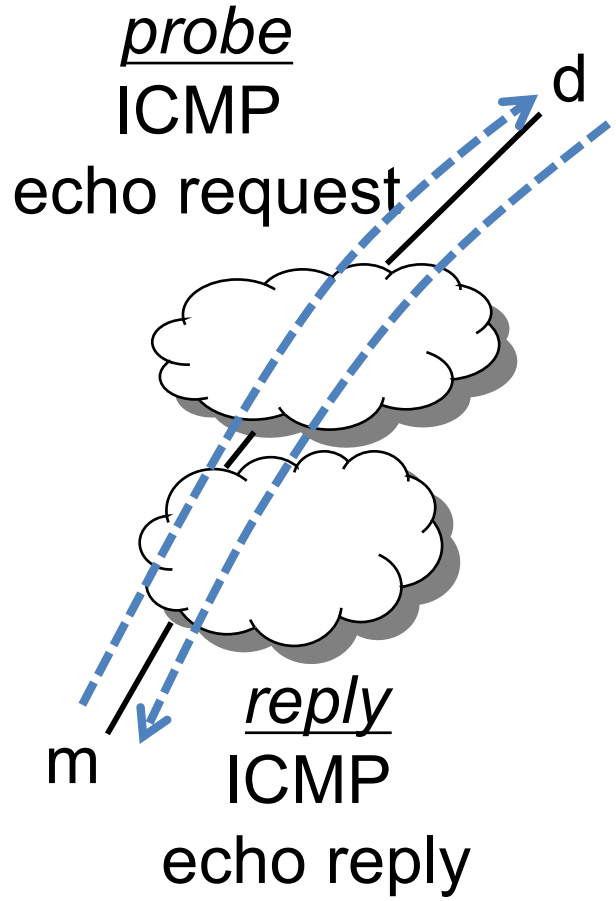


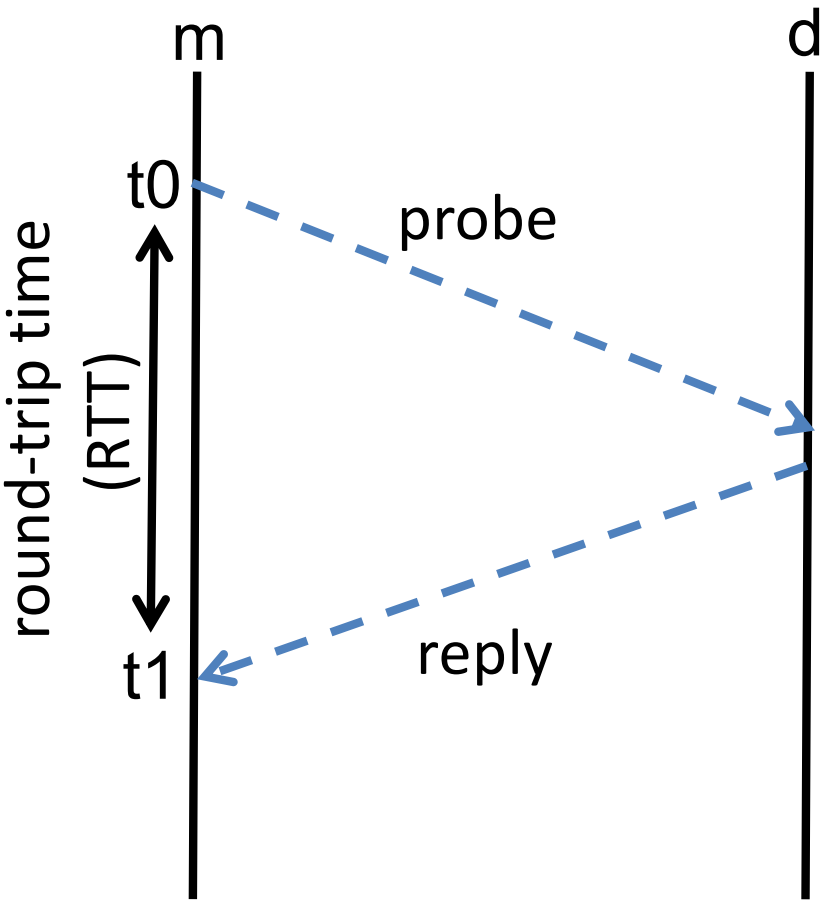
WHAT TYPES OF MEASUREMENTS EXIST?

Measurement techniques

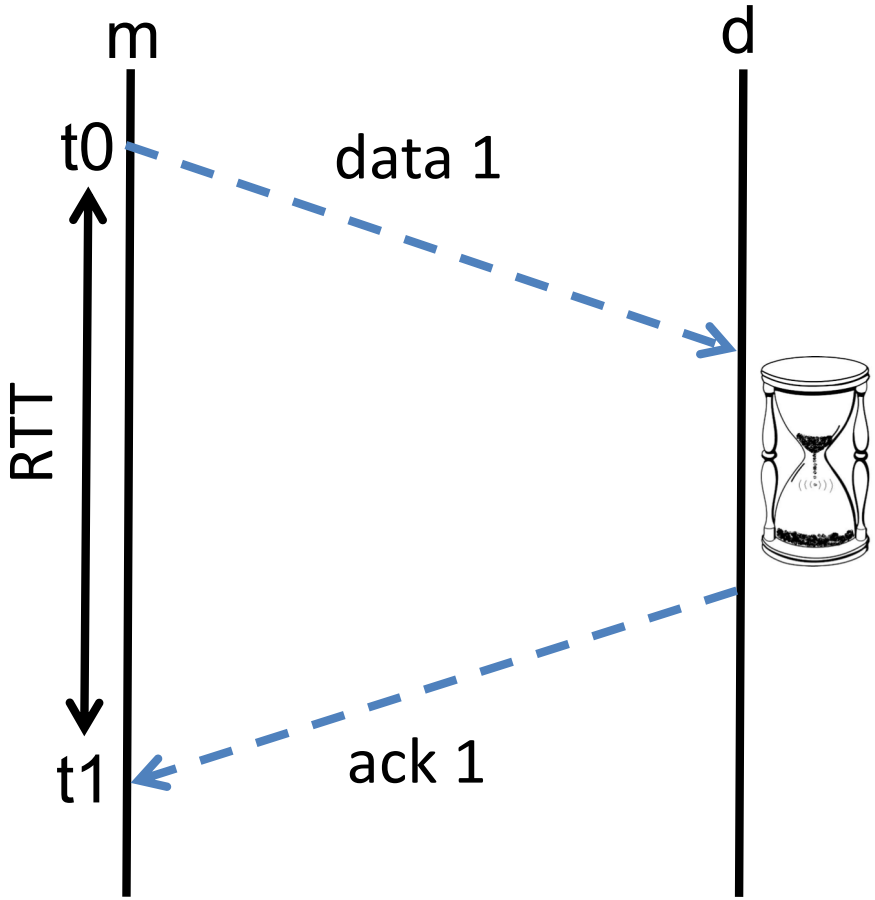
- Active
  - Based on issuing probes, analyzing response
- Passive
  - Observe existing traffic
    - E.g., IP packets, routing messages

Example active RTT measurement: ping





Example passive measurement
RTT inference: tcptrace*



Comparison

Passive

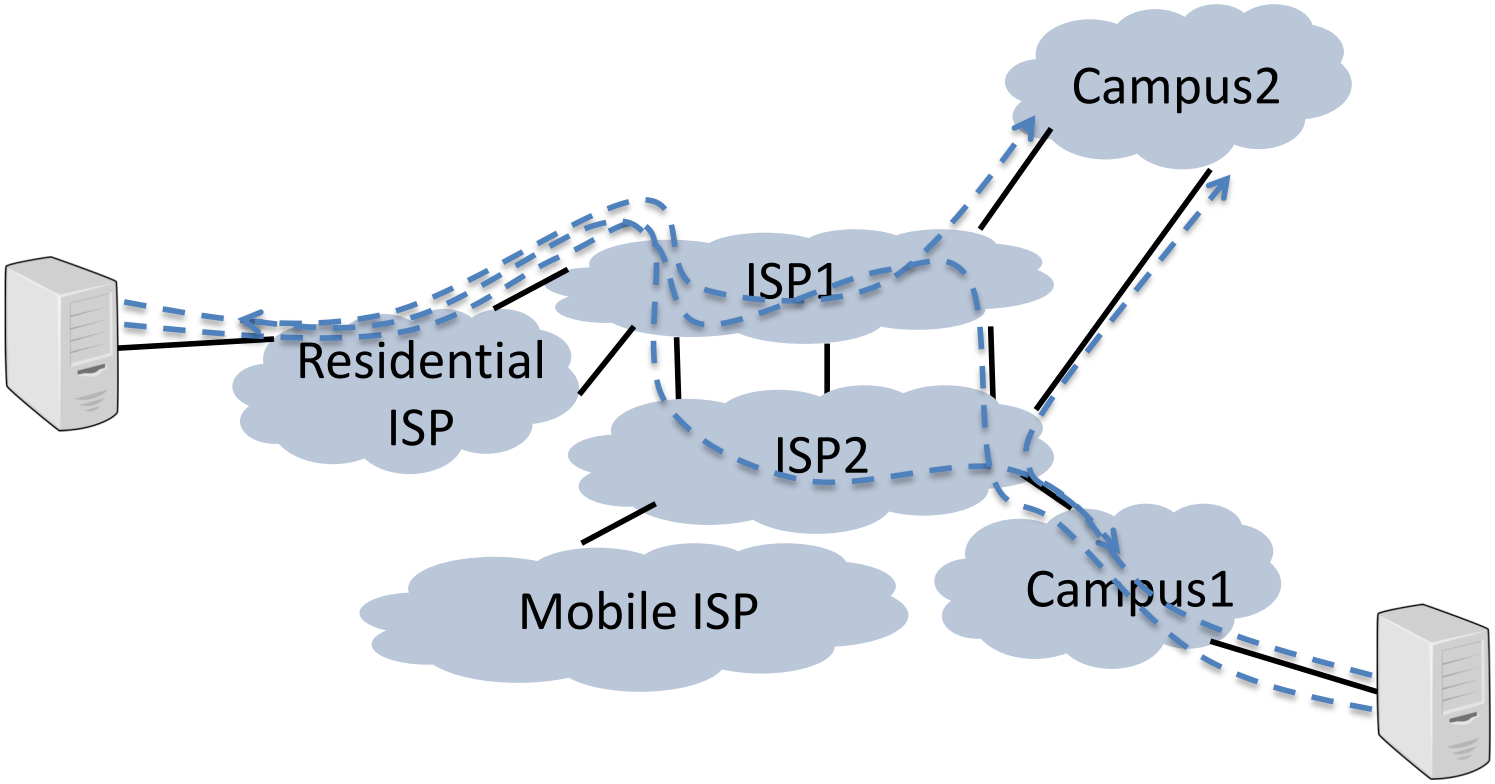
- Only way to measure traffic
- Measures user experience, behavior
- Measures protocol exchanges
- Raise privacy concerns

Active

- Measurements even when taping traffic is not possible
- Measures network, application performance
- Probing extra load
  - Overload network
  - Bias inferences

Measurement vantage point

- Point where measurement host connects to network
  - Observations often depend on vantage point



Possible vantage points

- End-hosts connected to the Internet
  - Active measurements of end-to-end paths
    - Better if control of both ends of the path
  - Passive measurements of host's traffic/apps
- Routers/Measurement hosts in network
  - Active measurements of network paths
  - Passive measurements of traffic, protocol exchanges, configuration

SOUND MEASUREMENT PRACTICES

Goal: build trust on measurement results

- Know what to expect
- Know the measurement tool
- Know where the data comes from

Know what to expect*

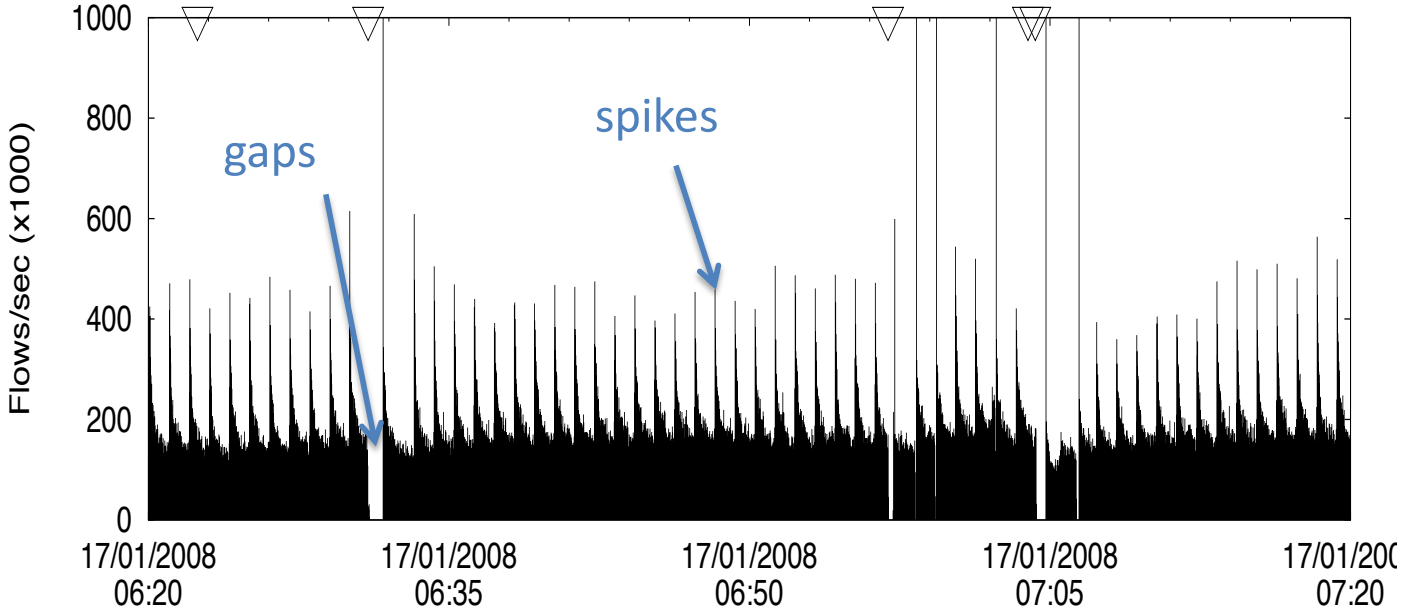
- Identify properties that must hold
  - E.g., RTTs > delay of speed of light
  - E.g., number of bytes in a TCP connection < duration * max capacity
- When properties fail to hold
  - Incorrect assumption: improve mental model
  - Measurement error

Know the measurement tool: Study precision and accuracy

- E.g., timestamp precision: 17:21:10.154379
  - resolution = 1 μ second?
  - Some clocks only advance every 10 millisecond

Know the measurement tool: examine outliers and spikes

- Measurement errors are often “corner cases”
  - Outlier = unusually low/high values
  - Spike = values that occur a lot
- Further analysis may identify measurement artifacts



Know the measurement tool

- Study precision and accuracy
- Examine outliers and spikes
- Monitor confounding factors
  - Monitor's CPU, memory, traffic
- Evaluate synthetic data, controlled settings
- Compare multiple methods
- Re-calibrate as needed
  - E.g., changing environments

Know where data comes from

- Log meta-data with traces
  - Any information required to fully understand measurements
  - Remember data often used for unexpected purposes
- Examples of meta-data
  - Version of measurement tool and parameters
  - When, where trace was recorded
  - Clock precision
  - Drops, missing data

ETHICAL ISSUES

Avoid disruption

- Active probing can overload network/hosts
  - "Denial of Service" attack
- Good practices
  - Embed contact info in probes
  - Throttle probing
  - Spread load
  - Keep blacklists of networks/hosts

Respect privacy

- Passive measurements can get personal info
- Good practices
  - Get user informed consent when possible
  - Comply with local data protection laws
  - Anonymize data when possible
    - Caveat: anonymization is not bullet-proof

Do no harm

- Measurement studies can harm
  - Individual/organization privacy, reputation, well-being
- Good practices
  - Identify potential harms/risks
  - Maximize benefits and minimize risks
  - Menlo report
    - https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf

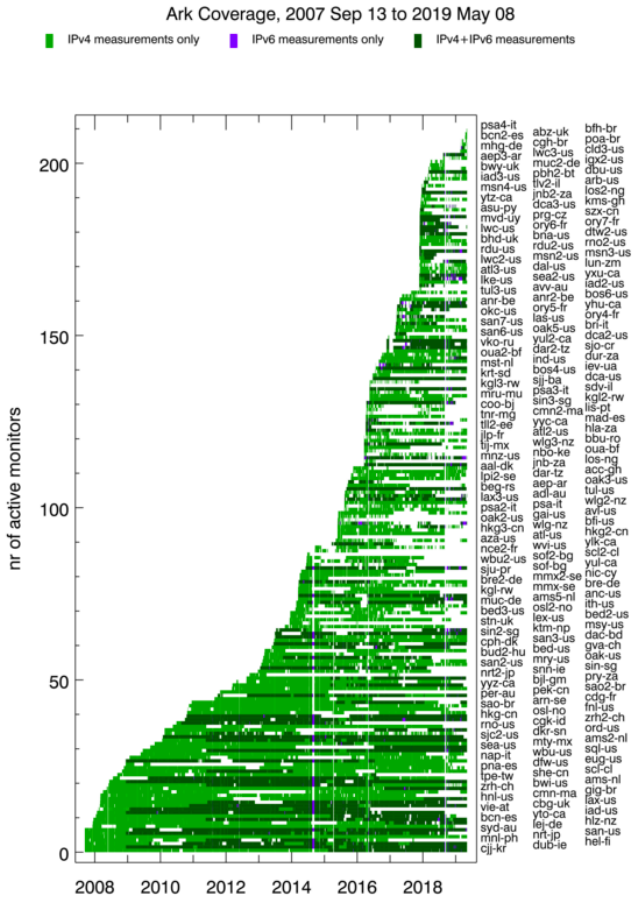
INFRASTRUCTURES: SOURCES OF INFORMATION

Types of data about infrastructure

- Routing monitors
  - BGP
  - OSPF/IS-IS
- Active measurements
  - Path topology
  - Performance (delay, throughput, etc.)
- Passive measurements
  - SNMP counters
  - Wireless metrics (PHY rate, RSSI, etc.)

Public sources of data

- BGP data
  - RouteViews, RIPE RIS
- Topology
  - CAIDA's Ark, RIPE Atlas
- Access, path performance
  - M-Lab, FCC/SamKnows
- Wireless data
  - Crawdad



Caida ARK Nodes
Source: https://www.caida.org/projects/ark/coverage/

Measurement platforms: closer to core

- Looking glass servers
  - Connected to major ISPs, IXPs
  - Allow interactive queries
  - BGP, ping, traceroute
- Distributed servers
  - E.g. PlanetLab, M-Lab
  - Deployed in university campus, data centers
  - Well-connected, powerful machines
  - Support running measurement scripts

Measurement platforms: at the edge

- Low cost monitors
  - E.g., RIPE Atlas (Plug computers), SamKnows/Bismark (access points)
  - Deployed close users (homes, offices)
  - More diverse connectivity, constrained machines
- Software platforms
  - E.g., Dasu (Bittorrent), Fathom (browser/Firefox)
  - Easier to deploy, large number of users
  - Not always on

Things to keep in mind when using measurement platforms

- Platforms require credits to conduct measurements
  - Ensure resource consumption is within limits
- Probing may trigger security alerts
  - Use blacklists/whitelists for probing
- Monitors' load may bias inferences
  - Monitor load on measurement nodes
- Timing issues in different platforms
  - Clocks across monitors often not synchronized
  - Check precision/accuracy of timestamps

Summary

- Focus of this class: Internet measurements
  - Infrastructure
  - Traffic
  - Applications
- Measurement techniques
  - Active probing
  - Passive observation
- Guidelines for sound measurements
- Measurements raise ethical issues
  - Evaluate risk versus benefit
- Sources of information on infrastructure

Recommended reading

- V. Paxson, “Strategies for Sound Internet Measurement”, IMC’04.
  - http://www.icir.org/vern/papers/meas-strategies-imc04.pdf

References

- CAIDA's DatCat
  - http://www.datcat.org/
- BGP datasets
  - RouteViews: http://www.routeviews.org/
  - RIPE-RIS: http://www.ripe.net/data-tools/stats/ris/routing-information-service
  - Cyclops: http://cyclops.cs.ucla.edu/

References

- Topology datasets
  - CAIDA's Ark: http://www.caida.org/projects/ark/
  - Dimes: http://www.netdimes.org
  - Northwestern's EdgeScope: http://aqualab.cs.northwestern.edu/projects/86-edgescope-sharing-the-view-from-a-distributed-internet-telescope
- Path performance/topology datasets
  - iPlane: http://iplane.cs.washington.edu/
  - M-Lab: http://www.measurementlab.net/
  - FCC data: https://www.fcc.gov/measuring-broadband-america/2012/raw-data-2012

References

- Platforms
  - RIPE Atlas: https://atlas.ripe.net/
  - PlanetLab: https://www.planet-lab.org/
  - Bismark: http://projectbismark.net/