CE693: Adv. Computer Networking

L-20 Measurement

Acknowledgments: Lecture slides are from the graduate level Computer Networks course thought by Srinivasan Seshan at CMU. When slides are obtained from other sources, a a reference will be noted on the bottom of that slide. A full list of references is provided on the last slide.

Motivation

- Answers many questions
  - How does the Internet really operate?
  - Is it working efficiently?
  - How will trends affect its operation?
  - How should future protocols be designed?
- Aren’t simulation and analysis enough?
  - We really don’t know what to simulate or analyze
    - Need to understand how Internet is being used!
  - Too difficult to analyze or simulate parts we do understand

Internet Measurement

- Process of collecting data that measure certain phenomena about the network
  - Should be a science
  - Today: closer to an art form
- Key goal: Reproducibility
- “Bread and butter” of networking research
  - Deceptively complex
  - Probably one of the most difficult things to do correctly

Measurement Methodologies

- Active tests – probe the network and see how it responds
  - Must be careful to ensure that your probes only measure desired information (and without bias)
  - Labovitz routing behavior – add and withdraw routes and see how BGP behaves
  - Paxson packet dynamics – perform transfers and record behavior
  - Bolot delay & loss – record behavior of UDP probes
- Passive tests – measure existing behavior
  - Must be careful not to perturb network
  - Labovitz BGP anomalies – record all BGP exchanges
  - Leland self-similarity – record Ethernet traffic

Types of Data

Active

- traceroute
- ping
- UDP probes
- TCP probes
- Application-level “probes”
  - Web downloads
  - DNS queries

Passive

- Packet traces
  - Complete
  - Headers only
  - Specific protocols
- Flow records
- Specific data
  - Syslogs …
  - HTTP server traces
  - DHCP logs
  - Wireless association logs
  - DNSBL lookups
  - …
- Routing data
  - BGP updates / tables, ISIS, etc.

Overview

- Active measurement
- Passive measurement
- Strategies
- Some interesting observations

Active Measurement

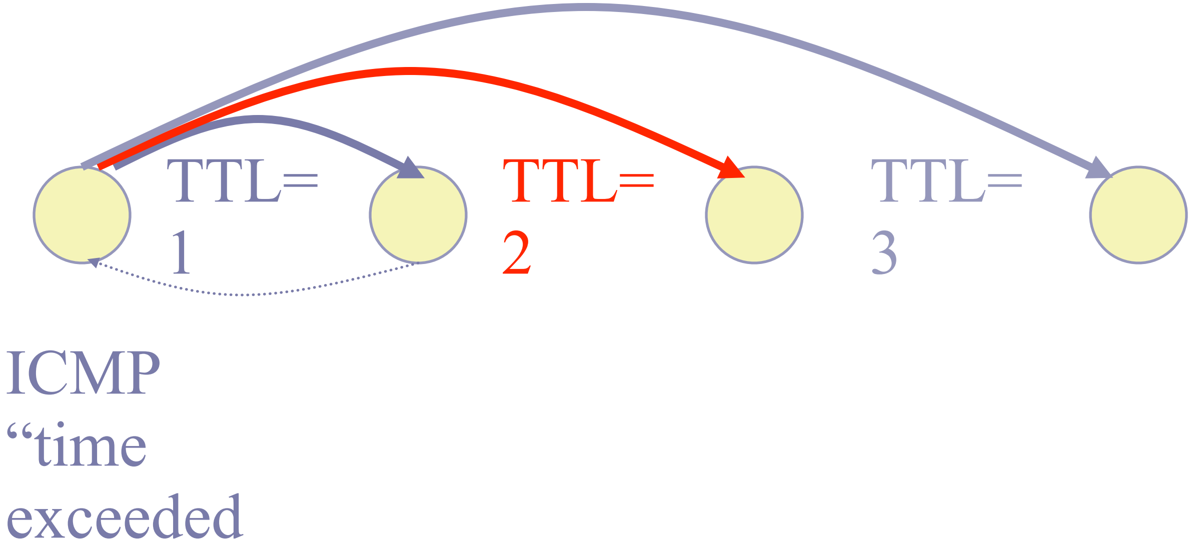
- Common tools:
  - ping
  - traceroute
  - scriptroute
  - Pathchar/pathneck/… BW probing tools

Sample Question: Topology

- What is the topology of the network?
  - At the IP router layer
  - Without “inside” knowledge or official network maps
- Why do we care?
  - Often need topologies for simulation and evaluation
  - Intrinsic interest in how the Internet behaves
    - “But we built it! We should understand it”
    - Emergent behavior; organic growth

How Traceroute Works

- Send packets with increasing TTL values



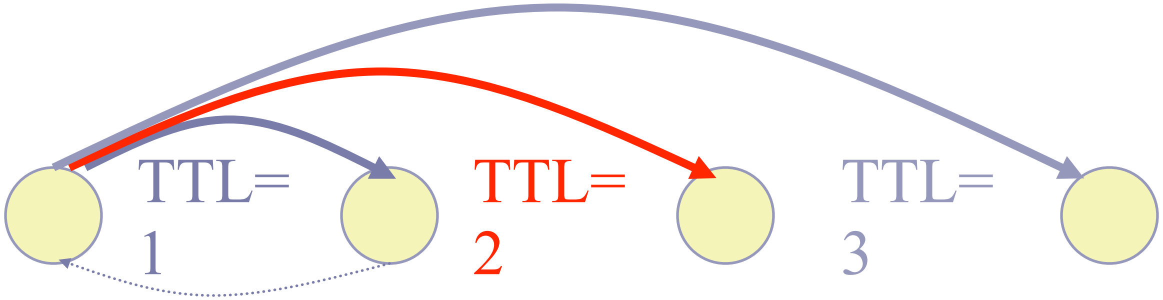
- Nodes along IP layer path decrement TTL
- When TTL=0, nodes return “time exceeded” message

Problems with Traceroute

- Can’t unambiguously identify one-way outages
  - Failure to reach host : failure of reverse path?

- ICMP messages may be filtered or rate-limited

- IP address of “time exceeded” packet may be the outgoing interface of the return packet



Famous Traceroute Pitfall

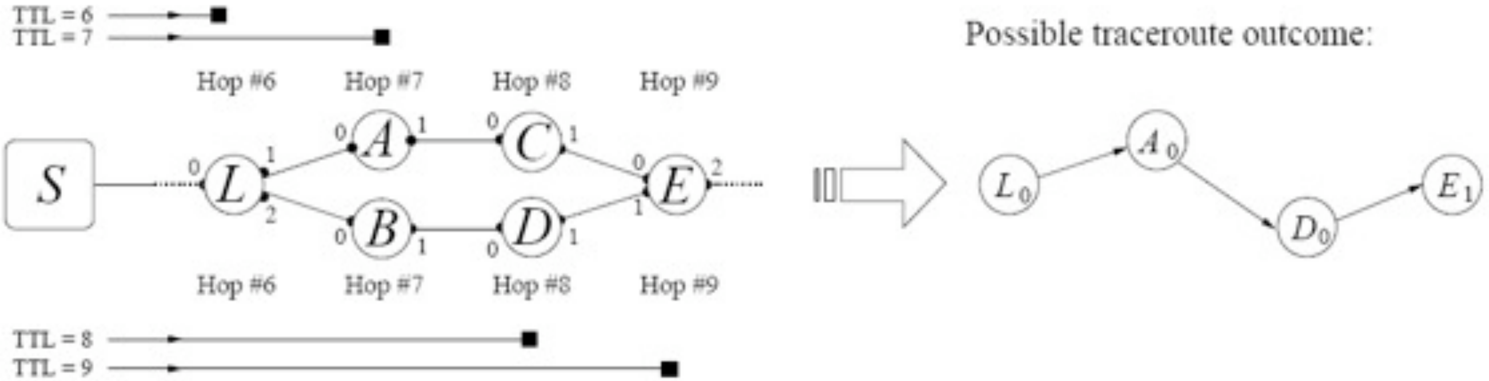
- Question: What ASes does traffic traverse?
- Strawman approach
  - Run traceroute to destination
  - Collect IP addresses
  - Use “whois” to map IP addresses to AS numbers
- Thought Questions
  - What IP address is used to send “time exceeded” messages from routers?
  - How accurate is whois data?

More Caveats: Topology Measurement

- Routers have multiple interfaces
- Measured topology is a function of vantage points

Less Famous Traceroute Pitfall

- Host sends out a sequence of packets
  - Each has a different destination port
  - Load balancers send probes along different paths
    - Equal cost multi-path
    - Per flow load balancing



Soule *et al.*, “Avoiding Traceroute Anomalies with Paris Traceroute”, *IMC 2006*

Designing for Measurement

- What mechanisms should routers incorporate to make traceroutes more useful?
  - Source IP address to "loopback" interface
  - AS number in time-exceeded message
  - ??
- More general question: How should the network support measurement (and management)?

Overview

- Active measurement
- Passive measurement
- Strategies
- Some interesting observations

Two Main Approaches

- Packet-level Monitoring
  - Keep packet-level statistics
  - Examine (and potentially, log) variety of packet-level statistics. Essentially, anything in the packet.
  - Timing

- Flow-level Monitoring
  - Monitor packet-by-packet (though sometimes sampled)
  - Keep aggregate statistics on a flow

Packet Capture: tcpdump/bpf

- Put interface in promiscuous mode
- Use bpf to extract packets of interest
- Packets may be dropped by filter
  - Failure of tcpdump to keep up with filter
  - Failure of filter to keep up with dump speeds

- **Question:** How to recover lost information from packet drops?

Traffic Flow Statistics

- *Flow monitoring* (*e.g.*, Cisco Netflow)
  - Statistics about groups of related packets (*e.g.*, same IP/TCP headers and close in time)
  - Recording header information, counts, and time

- More detail than SNMP, less overhead than packet capture

What is a flow?

- Source IP address
- Destination IP address
- Source port
- Destination port
- Layer 3 protocol type
- TOS byte
- Input logical interface (ifIndex)

Flow Record Contents

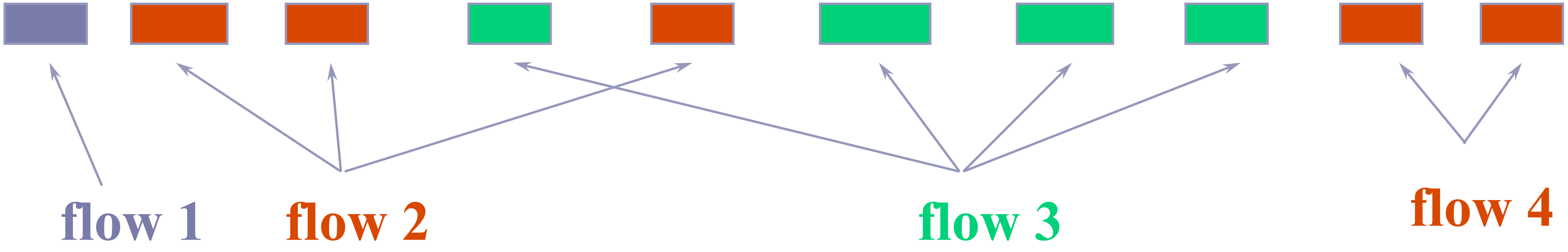
Basic information about the flow…

- Source and Destination, IP address and port
- Packet and byte counts
- Start and end times
- ToS, TCP flags

…plus, information related to routing

- Next-hop IP address
- Source and destination AS
- Source and destination prefix

Aggregating Packets into Flows

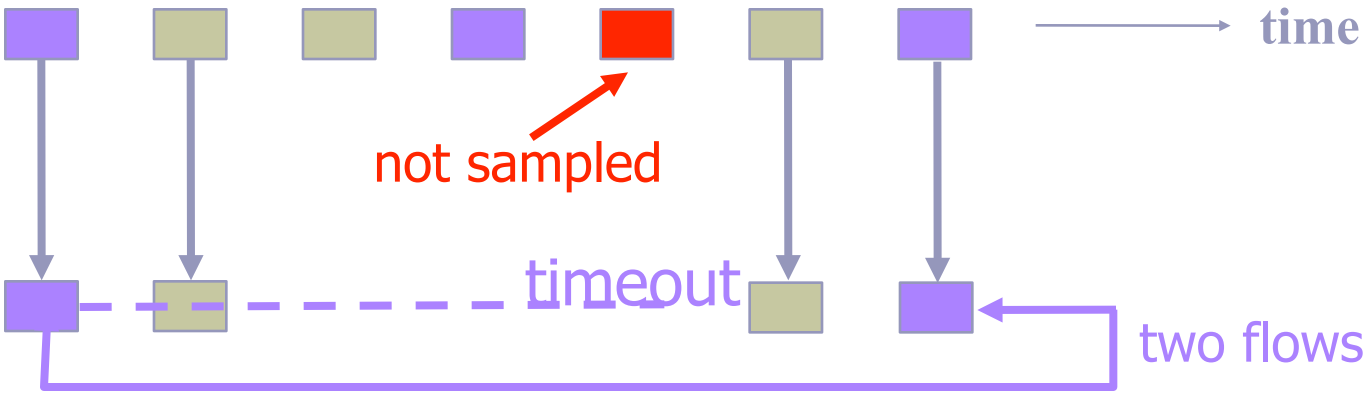


- **Criteria 1:** Set of packets that “belong together”
  - Source/destination IP addresses and port numbers
  - Same protocol, ToS bits, …
  - Same input/output interfaces at a router (if known)

- **Criteria 2:** Packets that are “close” together in time
  - Maximum inter-packet spacing (e.g., 15 sec, 30 sec)
  - **Example:** flows 2 and 4 are different flows due to time

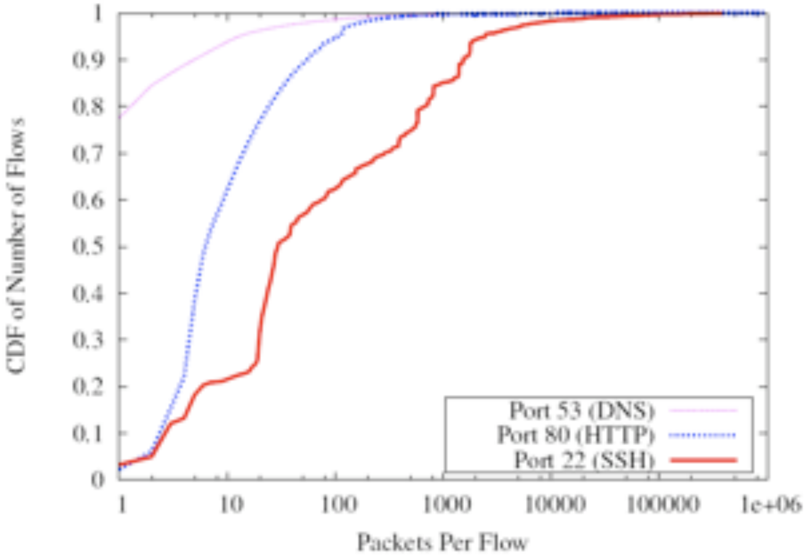
Packet Sampling

- Packet sampling before flow creation (Sampled Netflow)
  - 1-out-of-m sampling of individual packets (e.g., m=100)
  - Create of flow records over the sampled packets
- Reducing overhead
  - Avoid per-packet overhead on (m-1)/m packets
  - Avoid creating records for a large number of small flows
- Increasing overhead (in some cases)
  - May split some long transfers into multiple flow records
  - … due to larger time gaps between successive packets



Problems with Packet Sampling

- Determining size of original flows is tricky
  - For a flow originally of size *n*, the size of the *sampled* flow follows a binomial distribution
  - Extrapolation can result in big errors
  - Much research in reducing such errors
- Flow records can be lost
- Small flows may be eradicated entirely



Overview

- Active measurement
- Passive measurement
- Strategies
- Some interesting observations

Strategy: Examine the Zeroth-Order

- Paxson calls this “looking at spikes and outliers”
- More general: Look at the data, not just aggregate statistics
  - Tempting/dangerous to blindly compute aggregates
  - Time series plots are telling (gaps, spikes, etc.)
  - Basics
    - Are the raw trace files empty?
      - Need not be 0-byte files (e.g., BGP update logs have state messages but no updates)
    - Metadata/context: Did weird things happen during collection (machine crash, disk full, etc.)

Strategy: Cross-Validation

- Paxson breaks cross validation into two aspects
  - Self-consistency checks (and sanity checks)
  - Independent observations
    - Looking at same phenomenon in multiple ways
- What are some examples?

Example Sanity Checks

- Is time moving backwards?
  - **Typical cause:** Clock synchronization issues

- Has the the speed of light increased?
  - *E.g.*, 10ms cross-country latencies

- Do values make sense?
  - IP addresses that look like 0.0.1.2 indicate bug

Cross-Validation Example

- Telnet connection arrivals should follow a poison distribution (human induced)
- **Puzzle**
  - Every time a call comes in to the modem, the host launched a telnet connection
  - Data shows an unusual spike
  - So no poison distribution?
- Why?
  - Collection bugs … or
  - Broken mental model
    - It was assumed that human behavior was being measured, where as the modem was faulty

Longitudinal measurement hard

- Accurate distributed measurement is tricky!
- Lots of things change:
  - Host names, IPs, software
- Lots of things break
  - hosts (temporary, permanently)
  - clocks
  - links
  - collection scripts

Anonymization

- Similar questions arise here as with accuracy
- Researchers always want full packet captures with payloads
  - …but many questions can be answered without complete information
- Privacy / de-anonymization issues

PlanetLab for Network Measurement

- Nodes are largely at academic sites
  - Other alternatives: RON testbed

- Repeatability of network experiments is tricky
  - Proportional sharing
  - Work-conserving CPU scheduler means experiment could get more resources if there is less contention

Overview

- Active measurement
- Passive measurement
- Strategies
- Some interesting observations

Traces Characteristics

- Some available at http://ita.ee.lbl.gov
  - E.g. tcpdump files and HTTP logs
  - Public ones tend to be old (2+ years)
  - Privacy concerns tend to reduce useful content
- Paxson’s test data
  - Network Probe Daemon (NPD) – performs transfers & traceroutes, records packet traces
  - Approximately 20-40 sites participated in various NPD based studies
  - The number of “paths” tested by NPD framework scaled with (number of hosts)^2
    - 20-40 hosts = 400-1600 paths!

Observations – Routing Pathologies

- Observations from traceroute between NPDs
- Routing loops
  - Types – forwarding loops, control information loop (count-to-infinity)
  - Routing protocols should prevent loops from persisting
  - Fall into short-term (< 3hrs) and long-term (> 12 hrs) duration
  - Some loops spanned multiple BGP hops! → seem to be a result of static routes
- Erroneous routing – Rare but saw a US-UK route that went through Israel → can't really trust where packets may go!

Observations – Routing Pathologies

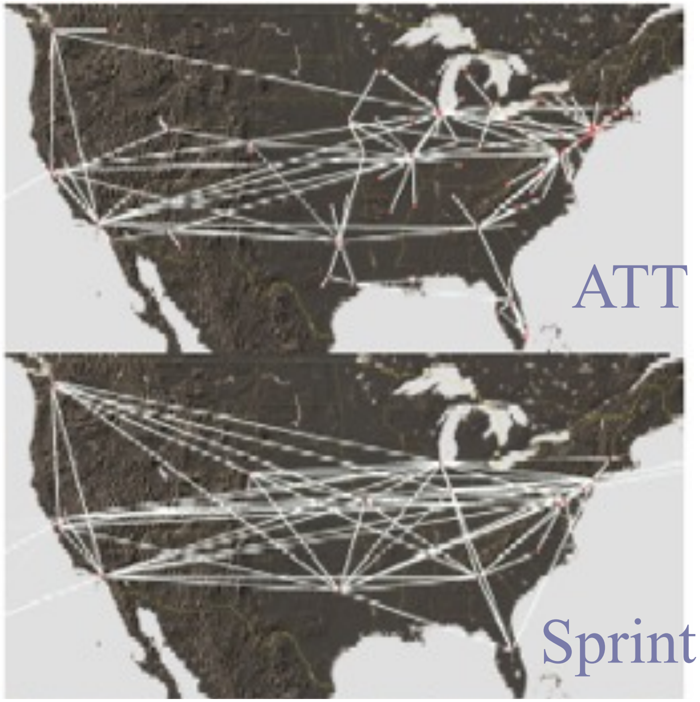
- Route change between traceroutes
- Temporary outages
  - Traceroute probes (1-2%) experienced > 30sec outages
  - Outage likelihood strongly correlated with time of day/ load
- Most pathologies seem to be getting worse over time

Observations – Routing Stability

- Prevalence – how likely are you to encounter a given route
  - In general, paths have a single primary route
  - For 50% of paths, single route was present 82% of the time
- Persistence – how long does a given route last
  - Hard to measure – what if route changes and changes back between samples?
  - Look at 3 different time scales
    - Seconds/minutes→ load-balancing flutter & tightly coupled routers
    - 10’s of Minutes → infrequently observed
    - Hours → 2/3 of all routes, long lived routes typically lasted several days

ISP Topologies

- Rocketfuel [SIGCOMM02]
  - Maps ISP topologies of specific ISPs
    - BGP → prefixes served
    - Traceroute servers → trace to prefixes for path
    - DNS → identify properties of routers
    - Location, ownership, functionality



- However…
  - Some complaints of inaccuracy – why?

Network Topology

- Faloutsos^3 [SIGCOMM99] on Internet topology
  - Observed many “power laws” in the Internet structure
    - Router level connections, AS-level connections, neighborhood sizes
  - Power law observation refuted later, Lakhina [INFOCOM00]

- Inspired many degree-based topology generators
  - Compared properties of generated graphs with those of measured graphs to validate generator
  - What is wrong with these topologies? Li et al [SIGCOMM04]
    - Many graphs with similar distribution have different properties
    - Random graph generation models don’t have network-intrinsic meaning
    - Should look at fundamental trade-offs to understand topology
      - Technology constraints and economic trade-offs
    - Graphs arising out of such generation better explain topology and its properties, but are unlikely to be generted by random processes!

Observations – Re-ordering

- 12-36% of transfers had re-ordering
- 1-2% of packets were re-ordered
- Very much dependent on path
  - Some sites had large amount of re-ordering
  - Forward and reverse path may have different amounts
- Impact → ordering used to detect loss
  - TCP uses re-order of 3 packets as heuristic
  - Decrease in threshold would cause many “bad” rexmits

Observations – Packet Oddities

- Replication
  - Internet does not provide “at most once” delivery
  - Replication occurs rarely
  - Possible causes → link-layer rexmits, misconfigured bridges
- Corruption
  - Checksums on packets are typically weak
    - 16-bit in TCP/UDP → miss 1/64K errors
  - Approx. 1/5000 packets get corrupted
  - 1/3million packets are probably accepted with errors!

Observations – Bottleneck Bandwidth

- Typical technique, packet pair, has several weaknesses
  - Out-of-order delivery → pair likely used different paths
  - Clock resolution → 10msec clock and 512 byte packets limit estimate to 51.2 KBps
  - Changes in BW
  - Multi-channel links → packets are not queued behind each other
- Solution – many new sophisticated BW measurement tools
  - Unclear how well they really work ☹

Observations – Loss Rates

- Ack losses vs. data losses
  - TCP adapts data transmission to avoid loss
  - No similar effect for acks → Ack losses reflect Internet loss rates more accurately (however, not a major factor in measurements)
- 52% of transfers had no loss
- 2.7% loss rate in 12/94 and 5.2% in 11/95
  - Loss rate for “busy” periods = 5.6 & 8.7%
  - Has since gone down dramatically…
- Losses tend to be very bursty