Computer Networks
CS 552

Badri Nath
Rutgers University
badri@cs.rutgers.edu

1. Measurements
Internet measurements - Why?

Why measure?

- What's the need? Do we need to measure?
- Can we just google it?
- What is the motivation? Who gains?
Why?

- Characterize behavior
  - Performance
  - Understand growth
  - Properties

- Design issues
  - Identify problems, suggest improvements
  - Understand protocol issues
  - Understand behavior

- Manage the network
  - Diagnose and correct
  - Traffic patterns
  - Nielsen rating for network performance
What to measure?
– need to determine X (X - a property of the internet)
What property?
• Performance
  • Latency, throughput, link B/W, jitter, …..
• Structure
  • Topology, relationships, characteristics
• Behavior
  • Stability, failures, pathological, misconfigurations
…
How to measure?

– need techniques and tools to measure, gather data, analyze data, present data
Network Measurement

- Measurement techniques
- Passive techniques
  - Collect traces and analyze or build special tools
  - Monitor application traffic, route announcements and analyze
  - Number of point to collect, single vantage point, several vantage points
  - Across ASes, or within AS (Gao 01, labovitz 97, labovitz 10)
- Active Techniques
  - Measure by probing (ping, traceroute, pathchar or build special tools (e.g., Bolot93, paxson 97, rocketfuel 02)
  - Inject probes/packets or remove packets and measure impact
  - Inject faults at various points BGP beacon
Issues in measurement

- To collect traces need mechanism to capture packets and dump information
- Design packet filters, should be able to take in a spec and collect trace
- Data collection and analysis are independent
- Share traces for different experiments
- Manipulation and storage cost for traces
- How to analyze packet traces?
  - Time-sequence plots
  - Plot sequence number as a function of time
  - Repeat for different time scales
Internet measurements - challenges

- Routing/announcement and data interaction
  - Control plane vs data plane
- Causality: cause and effect analysis
- Coverage and completeness
- Internet is too big, not everything is public
  - Commercial interests conflict
  - Access to data difficult (ASes hide stuff)
- Huge amounts of data
Network measurement Papers

- Jean Bolot SIGCOMM 93
- Craig Labovitz SIGCOMM 2010 Internet Inter-domain Traffic
- Faloutsos et al., Power-law relationships of internet topology SIGCOMM 99
- Additional reading: IXP paper by Anja Feldman SIGCOMM 13
Active Measurement

- Common tools:
  - ping
  - traceroute
  - scriptroute
  - Pathchar/pathneck/… BW probing tools
End-to-end delay and loss behavior in the Internet

- Simple experiments, interesting way of arriving at the results
  - Presentation makes all the difference
- Simple analysis and model fitting
- Use end-to-end measurements to derive bottleneck link behavior
- Use active measurement tool
- Send UDP packets at regular intervals and measure RTT. Repeat experiments for different intervals
- Data analysis using phase plots
Time-series plot

- For every sample, plot the $\text{rtt}_n$
- Not much to gain (except some have zero or packet loss)
- Average, max $\text{rtt}_n$, min $\text{rtt}_n$
- Instead plot $\text{rtt}_n$, $\text{rtt}_{n+1}$ (x,y)
- If constant, point.
- $D$, $\mu$
- RTT = Prop delay + Queuing delay + Transfer delay
- Where does the time go? $\Rightarrow D + w_{n+1} + \text{Service time}$
- Probe size is $P$
- Both see same load, one sees load, interarrival time has a significant impact on probe compression.
Delay analysis

- $s_n =$ time of $n^{th}$ probe, $r_n =$ time at which echo of the $n^{th}$ probe is received, $\text{rtt}_n = r_n - s_n$ and $\delta = S_{n+1} - S_n$
- For various values of $\delta$ determine the distribution of $\text{rtt}_n$
- Need to figure the impact of load on the network
- Assume that there is very light load
  \[ w_{n+1} \approx w_n \text{ or } w_{n+1} = w_n + \epsilon_n \]
- A phase plot should give us points on or around the diagonal
Phase plot

- $\text{rtt}_{n+1} = \text{rtt}_n \pm \varepsilon$
Delay analysis

- Assume a bottleneck link
- \( rtt_n = \text{prop delay} + \text{queuing delay} \)
- \( rtt_n = D + \text{wait time} + \text{service time} \)
  - \( rtt_n = D + w_n + P/\mu \)
- Heavy cross traffic
  - \( w_{n+1} = w_n + B/\mu \) or \( rtt_{n+1} - rtt_n = B/\mu - \delta \) or approx \( B/\mu \)
- During this time \( k \) probes arrive
- For these probes the waiting times are \( P/\mu \)
  - \( rtt_{n+k} - rtt_{n-(k-1)} = P/\mu - \delta \)
Delay analysis

- $\delta > \frac{P}{\mu}$, RHS is negative
  - $rtt_{n+1} - rtt_n < 0$, probe compression (traffic light analogy)
  - The line $y = x + \frac{P}{\mu} - \delta$ with a large number of points shows this effect
  - Find the x intercept, we know $P$ and $\delta$ hence calculate $\mu$
  - Figure 2, $\delta = 50$ msec
  - Bottleneck link = 128 Kbps
- $\delta >> \frac{P}{\mu}$, RHS is also negative
  - But two probes never get behind each other
  - Say $\delta = 500$ msec, then $\delta - \frac{P}{\mu} = 498$ msec
  - Find how many points on the line $y = x - 498$
  - Figure 4, $\delta = 500$ msec
- $\delta < \frac{P}{\mu}$ probe saturates the queue
Repeated experiments for different values of $\delta$ over a different path (umd, pittsburgh)

For $\delta = 8$ msec, y intercept = -7.8 msec, $P/\mu = 0.2$
  - Given $P = 32$ bytes, we get $m = 1.28$ Mbps (Figure 5)

For $\delta = 50$ msec, phase plot consists of points scattered around $y=x$ line. No probe compression
  - Figure 6
Assume cross traffic between two probes δ units apart

\[ w_{n+1} = \text{service time for } b_n \text{ and } w_n + \text{wait time for } w_n \]

\[ = \frac{(B + P)}{\mu} - \delta + w_n \]

\[ B = \mu(w_{n+1} - w_n + \delta) + P \]

\[ w_{n+1} - w_n + \delta = \frac{(B+P)}{\mu} \]

Interarrival time between successive probe returns is \((r_{n+1} - r_n)\) is \(w_{n+1} - w_n + \delta\)

which is the same as \(\text{rtt}_{n+1} - \text{rtt}_n + \delta\)

Plot CDF of this function
Finding packet size

Figure 8: Distribution of $w_{n+1} - w_n + \delta$ for $\delta = 20$ ms
Three cases

- Probes that were subject to probe compression and were serviced according to probe packet size
  - Difference in Wait times should be around $P/\mu$
  - First peak
- Probes that were not subject to probe compression
  - Difference in Wait times should be $\delta$
  - Second peak
- Probes that were behind one internet packet
  - Difference in Wait times should $(B + P)/\mu$
  - Third peak
- Use the value of third peak in Equation 6. Knowing bottleneck link, probe size, we can calculate internet packet size
Packet loss

- $rtt_n$ was considered to be zero for a lost packet
- Look at the number of packets that were lost for various values of $\delta$
  - Smaller the value of $\delta$ larger the loss
- Find probability of correlated losses (burst size)
  - Further, conditional loss probability $>$ loss probability
  - Because loss depends on queue length
- Burst loss size is about 1 for most cases
- Interesting observation
- Implications for FEC and Go-back-N
Discussion

- Simple experiments, clever analysis leads to interesting results
- Phase plots, very creative
- Assumptions: single bottle neck
  - Needed for simple analysis
- How about internet of today
  - Asymmetric links
  - Traffic patterns may not pan out as in Jean Bolot’s paper
- What about packet size distributions?
Active Measurement (2)

- Using traceroute
- Every packet has TTL
- At each hop, TTL is decremented
- When TTL becomes 0, packet is dropped and an ICMP error message is sent back to the source
- Use this to determine routers along the path from source to destination
- www.traceroute.org
Traceroute

tracing path from www.net.princeton.edu to 128.6.25.18 ...
traceroute to 128.6.25.18 (128.6.25.18), 30 hops max, 40 byte packets
1 gigagate1 (128.112.128.114) 0.530 ms 0.300 ms 0.343 ms
2 vgate1 (128.112.12.22) 0.360 ms 0.328 ms 0.348 ms
3 local1.princeton.magpi.net (216.27.98.113) 23.504 ms 2.484 ms 2.490 ms
4 remote2.njedge.magpi.net (216.27.98.82) 4.624 ms 4.217 ms 4.180 ms
5 130.156.251.206 (130.156.251.206) 28.162 ms 4.967 ms 4.706 ms
6 198.151.130.133 (198.151.130.133) 5.180 ms 22.034 ms 5.067 ms
7 sr02-hill012-svcs.runet.rutgers.net (128.6.1.22) 5.391 ms 5.343 ms 5.316 ms
8 * * *
9 * * *
10 * * *
11 * * *
12 rags.rutgers.edu (128.6.25.18) 6.534 ms 5.813 ms 5.650 ms
Traceroute

Warning: multiple IP addresses found for www.iitb.ac.in, using 203.78.217.179
tracing path from www.net.princeton.edu to 203.78.217.179 ...
traceroute to 203.78.217.179 (203.78.217.179), 30 hops max, 40 byte packets
1 gigagate1 (128.112.128.114) 0.826 ms 0.329 ms 0.343 ms
2 vgate1 (128.112.12.22) 0.373 ms 0.587 ms 0.349 ms
3 gi-6-0-226.hse1.phlapa02.paetec.net (209.92.27.33) 2.188 ms 2.616 ms 3.321 ms
4 ge-1-1-0-311.core02.phlapa02.paetec.net (169.130.105.19) 8.179 ms 2.550 ms 2.501 ms
5 ***
6 ***
7 as0.core02.asbnva01.paetec.net (169.130.81.233) 5.919 ms 6.093 ms 6.013 ms
8 157.130.49.209 (157.130.49.209) 5.946 ms 6.119 ms 5.699 ms
9 0.so-1-0-3.XL4.IAD8.ALTER.NET (152.63.41.26) 6.560 ms 6.263 ms 6.234 ms
10 0.so-3-0-0.IL4.NYC9.ALTER.NET (152.63.23.177) 14.967 ms 15.097 ms 14.756 ms
11 0.so-0-0-1.IR2.LND17.ALTER.NET (210.80.51.141) 87.187 ms 86.995 ms 85.641 ms
12 so-0-0-2.XT2.BOM1.ALTER.NET (210.80.37.198) 209.186 ms 210.845 ms
13 so-3-0-0.GW3.BOM1.ALTER.NET (210.80.37.114) 209.359 ms 209.795 ms 211.322 ms
14 iitb.ac.in-gw.customer.alter.net (203.78.210.150) 212.256 ms 211.991 ms 210.394 ms
15 *** 16 *** 17 *** 18 *** 19 ***
Known issues with Traceroute

- Not all routers accept ICMP messages
  - Filtered or given low priority
- Can’t identify one-way failures (* * *)
  - Forward link or reverse link
- IP address of “time exceeded” packet may be the outgoing interface of the return packet
Extend to multiple links

- Extend analysis and methodology of Bolt’s paper
- Send probes of varying sizes, across several links
- Use traceroute over N hops; repeat expt for various packet sizes
- Traceroute includes Cumulative performance
- By subtracting successive link char you can get
- Subtract Y intercepts (RTT, RTT/2 gives one way delay)
- Subtract successive slopes (P/μ) you get slope which is 1/ link bandwidth
- Note, X-axis is packet size (y=P/μ)

Reference: Allen Downey Using pathchar to estimate internet link characteristics- SIGCOMM 1999
Active measurements

- Multiple vantage points
- Send probes from multiple destinations
- Cover different paths from various S, D pairs
Questions on internet topology

- What does the internet connectivity graph look like?
  - Stub, tree, mesh

- What are the patterns
  - Invariants, growth (edges, core)

- Growth trends
  - Scale

- Realistic topologies for simulation
  - Is it random? Uniformly distributed links realistic?
Why do we need the topology?

- Understand the macroscopic properties of the Internet physical structure
- Network management
- Topology-aware algorithms
- Simulation and topology generation tools
Internet Inter-Domain topology

- Internet Inter-domain traffic by Craig Labovitz, et.al., SIGCOMM 2010
- Anatomy of a large European IXP by Anja Feldmann, et.al., SIGCOMM 2013
- Pictures/graphs from slides of SIGCOMM presentations
Change in carrier traffic demands

<table>
<thead>
<tr>
<th>Rank</th>
<th>2007 Top Ten</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ISP A</td>
<td>5.77</td>
</tr>
<tr>
<td>2</td>
<td>ISP B</td>
<td>4.55</td>
</tr>
<tr>
<td>3</td>
<td>ISP C</td>
<td>3.35</td>
</tr>
<tr>
<td>4</td>
<td>ISP D</td>
<td>3.2</td>
</tr>
<tr>
<td>5</td>
<td>ISP E</td>
<td>2.77</td>
</tr>
<tr>
<td>6</td>
<td>ISP F</td>
<td>2.6</td>
</tr>
<tr>
<td>7</td>
<td>ISP G</td>
<td>2.24</td>
</tr>
<tr>
<td>8</td>
<td>ISP H</td>
<td>1.82</td>
</tr>
<tr>
<td>9</td>
<td>ISP I</td>
<td>1.35</td>
</tr>
<tr>
<td>10</td>
<td>ISP J</td>
<td>1.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>2009 Top Ten</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ISP A</td>
<td>9.41</td>
</tr>
<tr>
<td>2</td>
<td>ISP B</td>
<td>5.7</td>
</tr>
<tr>
<td>3</td>
<td>Google</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Comcast</td>
<td>3.12</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Based on analysis of anonymous ASN (origin/transit) data (as a weighted average % of all Internet Traffic). Top ten has NO direct relationship to study participation.

- In 2007, top ten match “tier-1” ISPs (e.g., Wikipedia)
- In 2009, global transit carry significant traffic volumes
  - But Google and Comcast join the list
  - And a significant percentage of ISP A traffic is Google transit
Impact of cloud

- In 2007, thousands of ASNs contributed 50% of content
- In 2009, 150 ASNs contribute 50% of all Internet traffic
Traditional Internet model

Diagram showing the traditional Internet model with layers of National Backbone Operators, Regional Access Providers, Local Access Providers, and Customer IP Networks, leading down to Consumers and business customers.
A new Internet model

- Flatter and much more densely interconnected Internet
- Disintermediation between content and “eyeball” networks
- New commercial models between content, consumer and transit
Even more recent model

- IXPs central component
- Lots of local peering – rich fabric
- Even flatter AS topology than assumed
IXP Architecture
IXP diurnal traffic patterns

- Traffic Volume: Same as Tier-1 ISPs
- Clear daily patterns
Flatter topology
Difficulties in topology discovery

- Physical links
- Backup links
- Mobile access
- Router Identification and annotation
- Alias resolution
- Completeness Validation

Currently, none of them is completely solved!