Measurements of Distributed Systems and Networks

Measurements

- It has often been stated that “you can’t manage what you can’t measure” ...
- Effective tool to understand, model, test, and improve existing systems …

Let's consider the Internet itself

- Wide area networks are too complex to grasp
  - Many protocols at various levels interact and effect behavior
- Many applications have performance requirements
  - End-to-end delay and loss, reliability
- Its an interesting complex system
  - Has emergent characteristics like many living systems
    - Biological systems
    - Social networks

Internet Measurement Challenges

- Network size [old numbers]
  - 100,000,000s hosts, 1,000,000s routers, ~30,000 ASes
  - E.g., 50 billion devices predicted by 2020 ...
- Network Complexity
  - Interaction between components, protocols, applications, users
- All change over time
  - New applications are added
  - New protocol versions (TCP)
  - New router design (AQM)

Why do we measure the Internet?

- Already mentioned
  - Because it is there!
  - Operational reasons
- We cannot improve the Internet if we don't understand it
  - We cannot understand it if we don't measure
  - We cannot build effective models or simulators if we don't measure

What can we measure in the Internet?

- Structure
  - Topology (router/network) connectivity, link capacities, link loss, available bandwidth, routing, …
- Traffic
  - End-to-end performance, packet arrival process (congestion built-up), …
- Users and applications
  - WWW, peer-to-peer, streaming, …
- Malicious behavior
  - Attack patterns, port scans, …

Where can we measure the Internet?

How to chose representative measurement points?

Example: traffic samples

- LAN traffic vs. WAN traffic,
- Inside an ISP vs. between continents
- Country biases
- Commercial location vs. educational
- More locations is better, BUT most of all, one point is better than no point

How can we measure the Internet?

- Active measurements
  - Probes: Traceroute, ping, packet trains
  - Application simulation
- Passive measurement
  - Logs (WWW)
  - Monitors, sniffers

When should we measure the Internet?

- Diurnal and weekly traffic cycles
- Time scales depend on “what” and “how”
- Passive measurement are typically continuous
  - Can generate **huge** datasets
  - Log access problems
  - Privacy concerns
- Active measurements are typically discrete
  - Important characteristics can be missed
  - Probes can be filtered and/or detected

Who is measuring the Internet?

- Businesses do a great deal of measurement
  - Mostly do not share with the research community
  - examples:
    - Akamai: http delay from server side
    - HP (Mercury): http delay from client side
    - Google: everything
- Academia and Research institutes
  - Publish papers, but data may not always be available
- Internet Statistics and Metrics Analysis (ISMA)
  - CAIDA attempt to create a global meta-data database

Publishing Internet Measurement Studies

- All major networking conferences & journals accept measurement papers; e.g.,
  - ACM SIGCOMM, IEEE INFOCOM, ACM SIGMETRICS
  - IEEE/ACM ToN, IEEE TPDS
- Dedicated meetings
  - ACM Internet Measurement Conf. (IMC)
  - Passive & Active Measurements Conf. (PAM)

Active Measurement Techniques

Active Probes

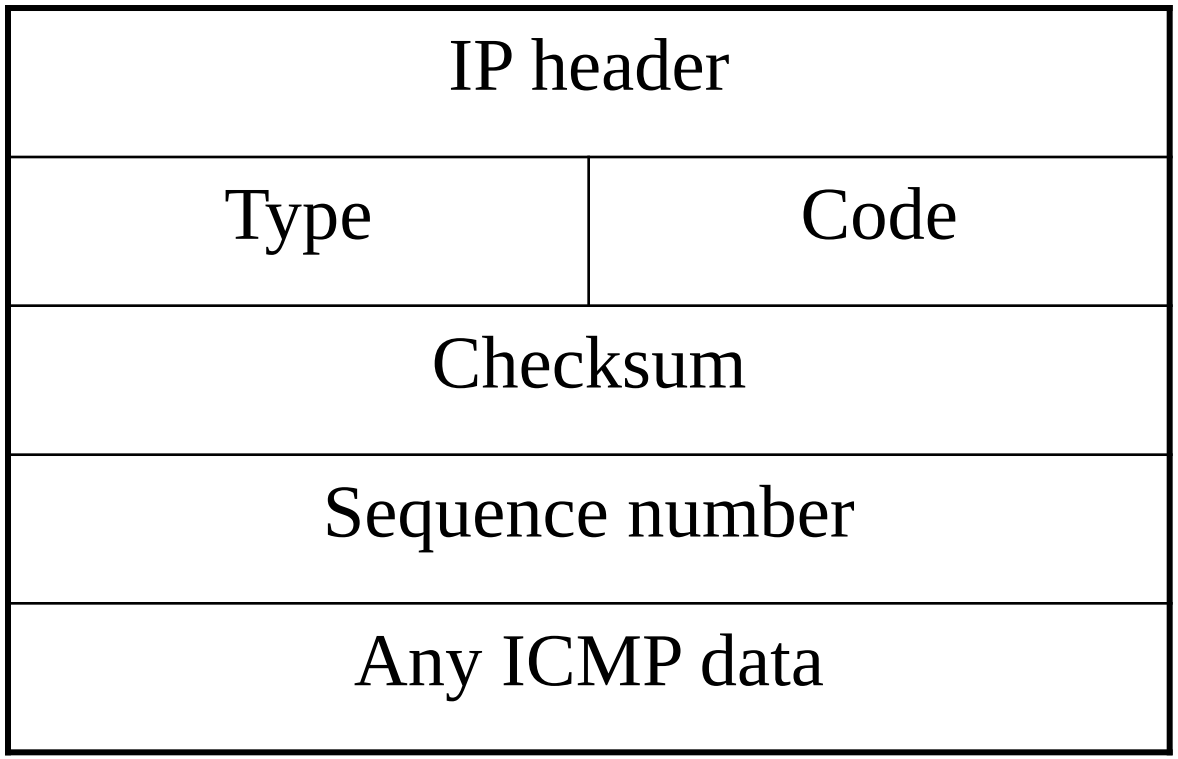
- Active probes send stimulus (packets) into the network and then measure the response
  - Done on network, transport and application layers
- Active probes are useful to measure various things:
  - Delay, delay jitter, and loss
  - Topology and routing behavior
  - Capacity, bandwidth, and throughput

Simple delay/loss probing with ping

```
C:\>ping www.fer.hr

Pinging www.fer.hr [161.53.72.111] with 32 bytes of data:

Reply from 161.53.72.111: bytes=32 time=113ms TTL=49
Reply from 161.53.72.111: bytes=32 time=111ms TTL=49
Reply from 161.53.72.111: bytes=32 time=113ms TTL=49
Reply from 161.53.72.111: bytes=32 time=118ms TTL=49

Ping statistics for 161.53.72.111:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 111ms, Maximum = 118ms, Average = 113ms
```

ICMP

ICMP is the IP error diagnosis protocol.



ICMP Message Types	
Type No.	Meaning
0	Echo reply
3	Destination unreachable
4	Source quench
5	Redirect
8	Echo
9	Router advertisement
10	Router solicitation
11	Time exceeded
12	Parameter problem
13	Timestamp
14	Timestamp reply
15	Information requeste
16	Information reply

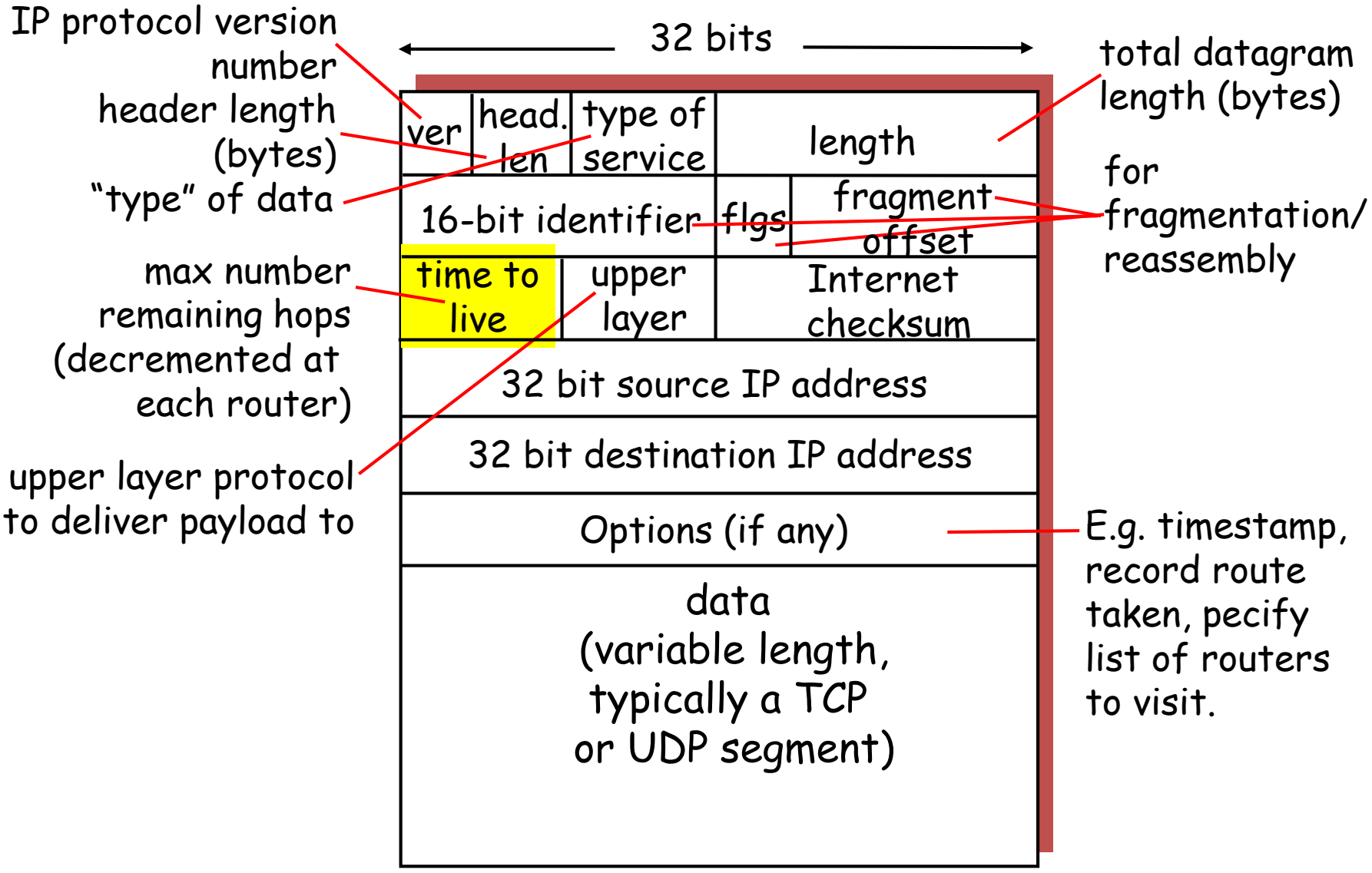
Application layer “ping”

- One can generate application layer messages to test application reaction time
- Most common:
  - TCP SYN message to port 80

traceroute

- Useful to learn the route characteristics between two hosts.
- Sends a series of probes to successive nodes along a route to an intended destination and records the source address and time delay of the message returned by each.
- Based on ICMP “TTL expired” message

IP datagram format



ICMP Message Types	
Type No.	Meaning
0	Echo reply
3	Destination unreachable
4	Source quench
5	Redirect
8	Echo
9	Router advertisement
10	Router solicitation
11	Time exceeded
12	Parameter problem
13	Timestamp
14	Timestamp reply
15	Information requeste
16	Information reply

Type	Code	description
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown

traceroute

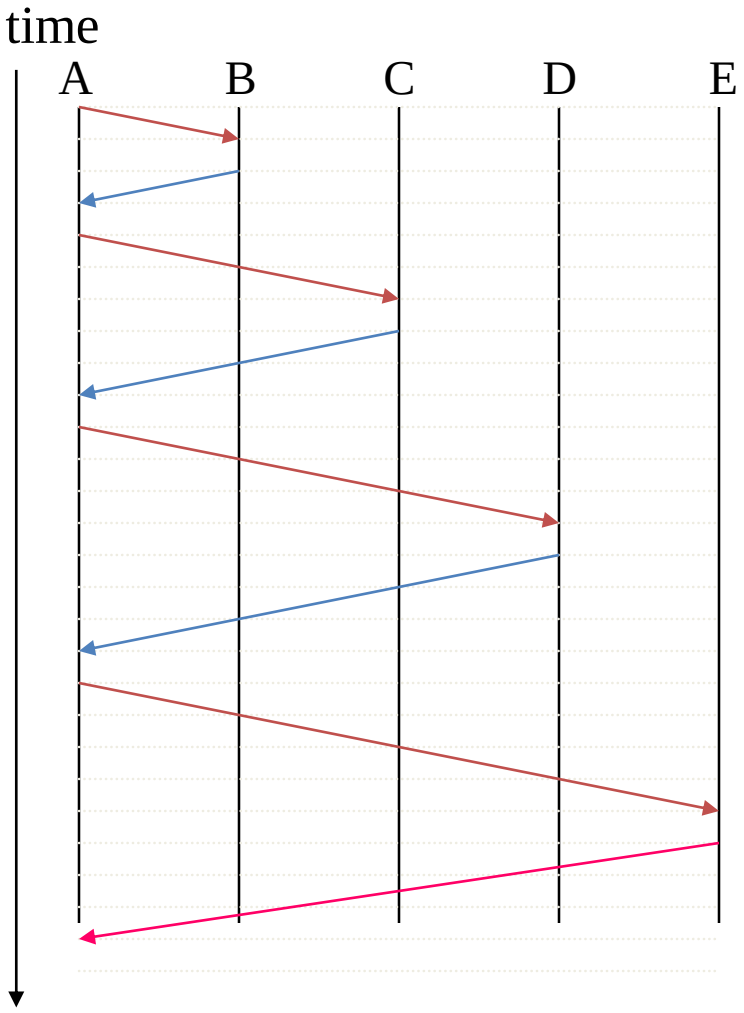
traceroute

Regular UDP packets

- successive TTLs

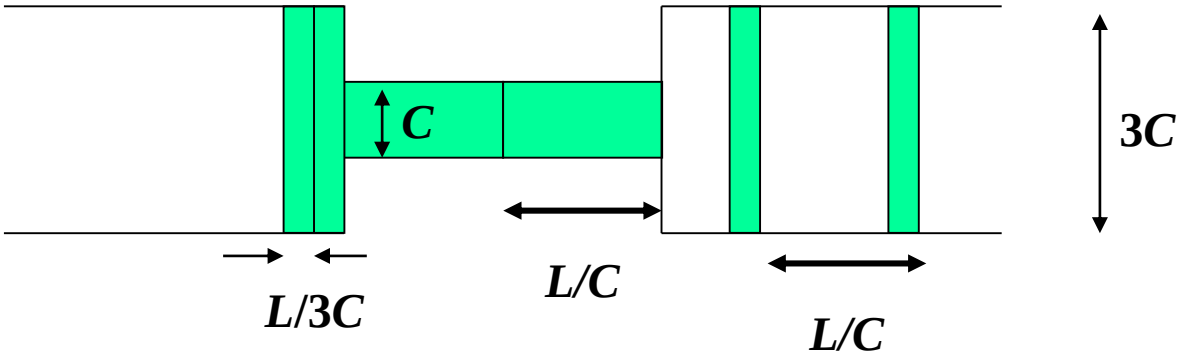
ICMP “TTL expired” message

ICMP “port unreachable” message



Packet Dispersion to Estimate Capacity

- Packet transmission time: $\tau = L/C$
- Send two packets back-to-back
- Measure dispersion Δ at the receiver
- Estimate C as L/Δ



- But cross-traffic 'noise' can effect Δ.
- E.g., patchar "allows any user to find (estimate) the bandwidth, delay, average queue and loss rate of every hop between any source & destination on the Internet"

Passive Measurement Techniques

Passive packet measurement

- Capture packets as they pass by
  - Packet capture applications (`tcpdump`) on hosts use packet capture filter
    - Requires access to the wire
    - Promiscuous mode or mirror ports to see other traffic
  - Hardware-based solutions
    - Endace, Inc.'s DAG cards .OC12/48/192 (0.622/2.5/10Gbps)
    - Programmable NIC cards (<$100)
- Issues:
  - Timestamps
  - Data volumes
  - Privacy

tcpdump

- Can capture entire packet or *n* first bytes
- Timestamps each packet
- Can filter based on any combination of header field

Passive IP flow measurement

- An IP flow is defined by the five-tuple:
  - src addr, src port, dst addr, dst port, protocol
- Cisco's NetFlow
  - Part of the IOS
  - Provide template based flow records
- Many tools can manipulate NetFlow data

HTTP Logs

- Have data about the client IP, transaction time, command (GET/POST), return code, bytes transferred, referrer, metadata (browser type, OS, languages, etc.)
- Tools are available to analyze HTTP logs
  - Webalizer

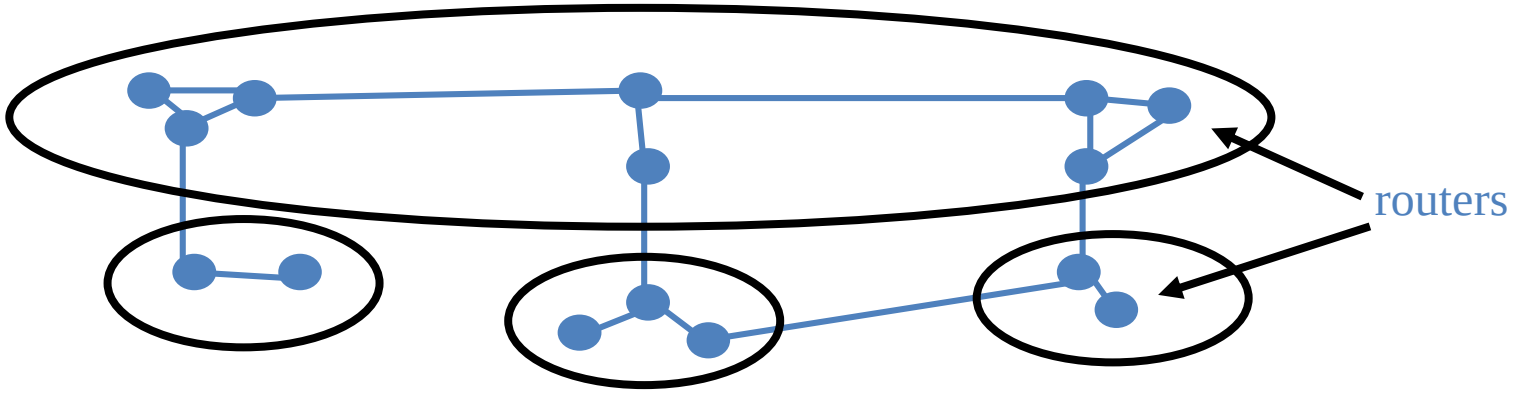
HTTP Log Example

24.77.192.99 - - [15/May/2005:23:54:59 +0300] "GET /science_down.gif HTTP/1.1" 200 1138 "http://www.netdimes.org/science.html
" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.7.7) Gecko/20050414 Firefox/1.0.3"
68.231.117.28 - - [15/May/2005:23:52:05 +0300] "GET /ipmap.png HTTP/1.1" 200 4874697 "http://slashdot.org/" "Mozilla/5.0 (X11
; U; Linux i686; en-US; rv:1.7.7) Gecko/20050414 Firefox/1.0.3"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /home_up.gif HTTP/1.1" 200 1096 "http://www.netdimes.org/" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /AboutUs_up.gif HTTP/1.1" 200 1169 "http://www.netdimes.org/" "Mozilla/4
.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.77.192.99 - - [15/May/2005:23:55:00 +0300] "GET /Install_down.gif HTTP/1.1" 200 1219 "http://www.netdimes.org/science.html
" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.7.7) Gecko/20050414 Firefox/1.0.3"
69.141.103.137 - - [15/May/2005:23:54:50 +0300] "POST /DIMES/server HTTP/1.1" 200 3 "-" "Java/1.4.1_03"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /news_up.gif HTTP/1.1" 200 1086 "http://www.netdimes.org/" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /community_up.gif HTTP/1.1" 200 1199 "http://www.netdimes.org/" "Mozilla
/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /datastat_up.gif HTTP/1.1" 200 1233 "http://www.netdimes.org/" "Mozilla/
4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /science_up.gif HTTP/1.1" 200 1126 "http://www.netdimes.org/" "Mozilla/4
.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
71.106.2.53 - - [15/May/2005:23:55:00 +0300] "GET /favicon.ico HTTP/1.1" 200 5694 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.
1; en-US; rv:1.7.8) Gecko/20050511 Firefox/1.0.4"
62.179.197.156 - - [15/May/2005:23:54:02 +0300] "GET /ipmap.png HTTP/1.1" 200 4874697 "http://slashdot.org/" "Mozilla/5.0 (Wi
ndows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511 Firefox/1.0.4"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /Install_up.gif HTTP/1.1" 200 1219 "http://www.netdimes.org/" "Mozilla/4
.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
24.236.177.187 - - [15/May/2005:23:55:00 +0300] "GET /EVERGROW40.gif HTTP/1.1" 200 4089 "http://www.netdimes.org/" "Mozilla/4
.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
195.252.52.155 - - [15/May/2005:23:55:00 +0300] "GET /science_down.gif HTTP/1.1" 200 1138 "http://www.netdimes.org/science.ht
ml" "Mozilla/5.0 (Windows; U; Windows NT 5.1; sv-SE; rv:1.7.6) Gecko/20050318 Firefox/1.0.2"

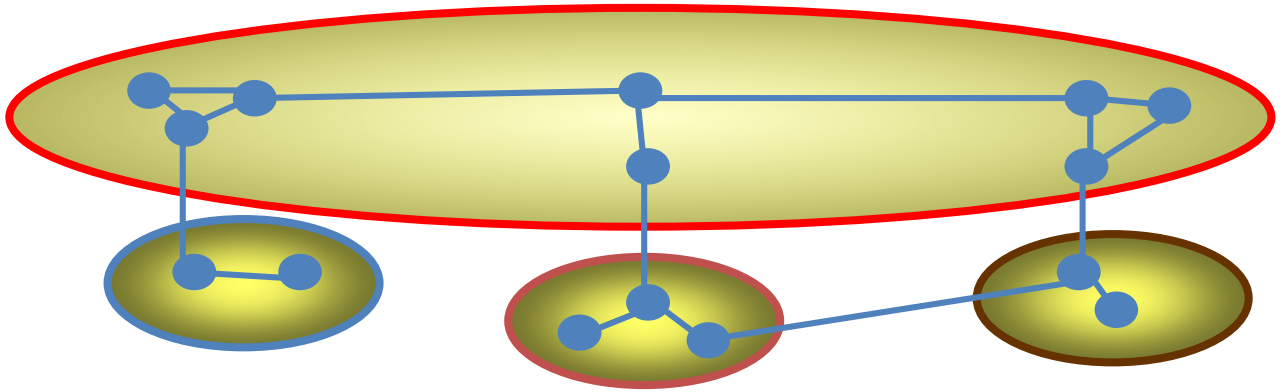
```
[root@jupiter httpd]# grep "GET / " access_log |tail -10
68.54.223.47 - - [19/May/2005:12:36:20 +0300] "GET / HTTP/1.1" 200 14067 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322)"
132.76.80.118 - - [19/May/2005:12:49:44 +0300] "GET / HTTP/1.1" 304 - "http://www.eng.tau.ac.il/~shavitt/" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322)"
24.169.148.213 - - [19/May/2005:13:06:58 +0300] "GET / HTTP/1.1" 200 14067 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511 Firefox/1.0.4"
84.170.181.64 - - [19/May/2005:13:07:14 +0300] "GET / HTTP/1.1" 200 14067 "http://www.google.de/search?hl=de&q=dimes&meta=" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
130.240.136.220 - - [19/May/2005:13:07:25 +0300] "GET / HTTP/1.1" 304 - "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)"
81.72.13.30 - - [19/May/2005:13:11:00 +0300] "GET / HTTP/1.1" 200 14067 "http://www.miranet.it/php/Articolo.php?id=708" "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)"
194.78.199.123 - - [19/May/2005:13:13:44 +0300] "GET / HTTP/1.1" 200 14067 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0; .NET CLR 1.1.4322)"
82.152.182.12 - - [19/May/2005:13:23:10 +0300] "GET / HTTP/1.1" 200 14067 "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)"
80.119.126.44 - - [19/May/2005:13:38:08 +0300] "GET / HTTP/1.1" 200 14067 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511 Firefox/1.0.4"
80.250.186.101 - - [19/May/2005:13:46:14 +0300] "GET / HTTP/1.1" 200 14067 "http://distributed.ru/forum/?a=topic&topic=583" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.8) Gecko/20050511 Firefox/1.0.4"
```

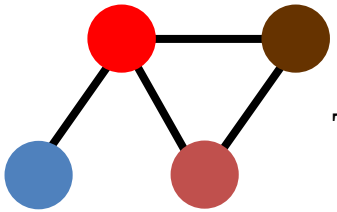
Measuring the Internet’s topology

The Internet Structure



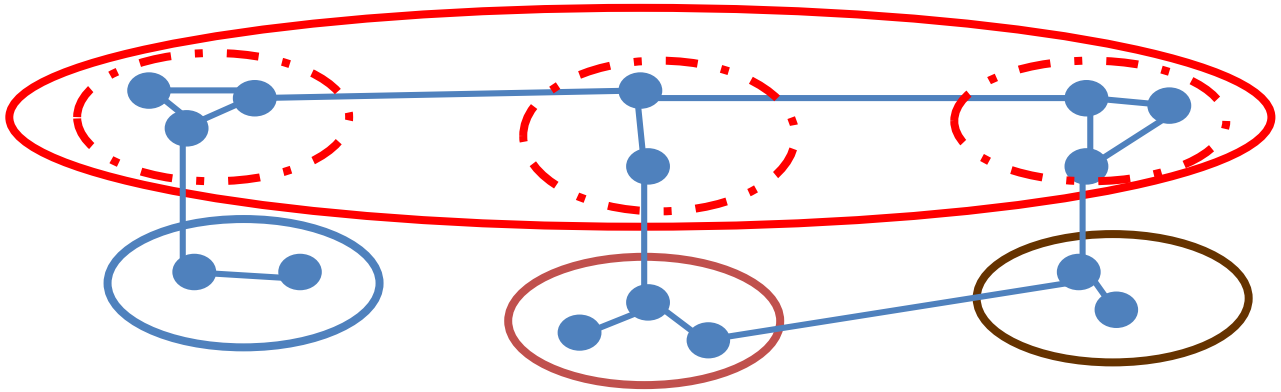
The Internet Structure

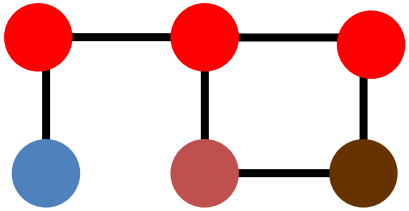


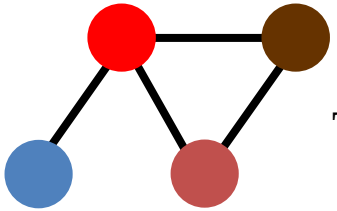


The **AS** graph

The Internet Structure





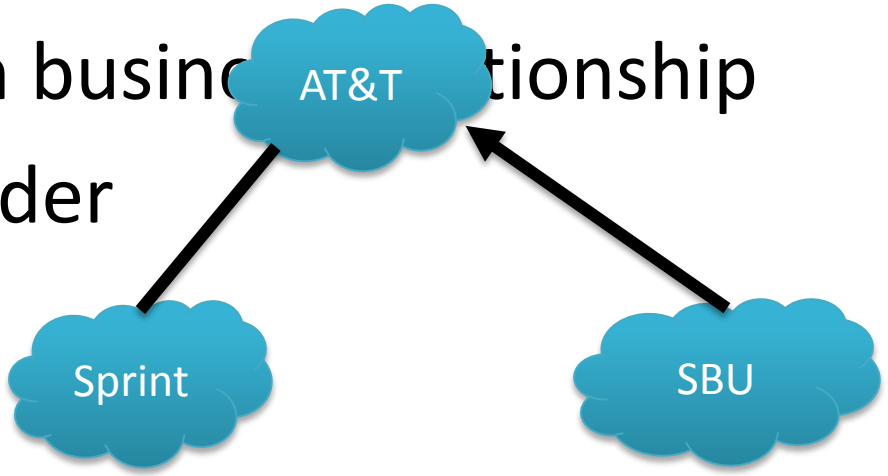


The **AS** graph

The **PoP level** graph

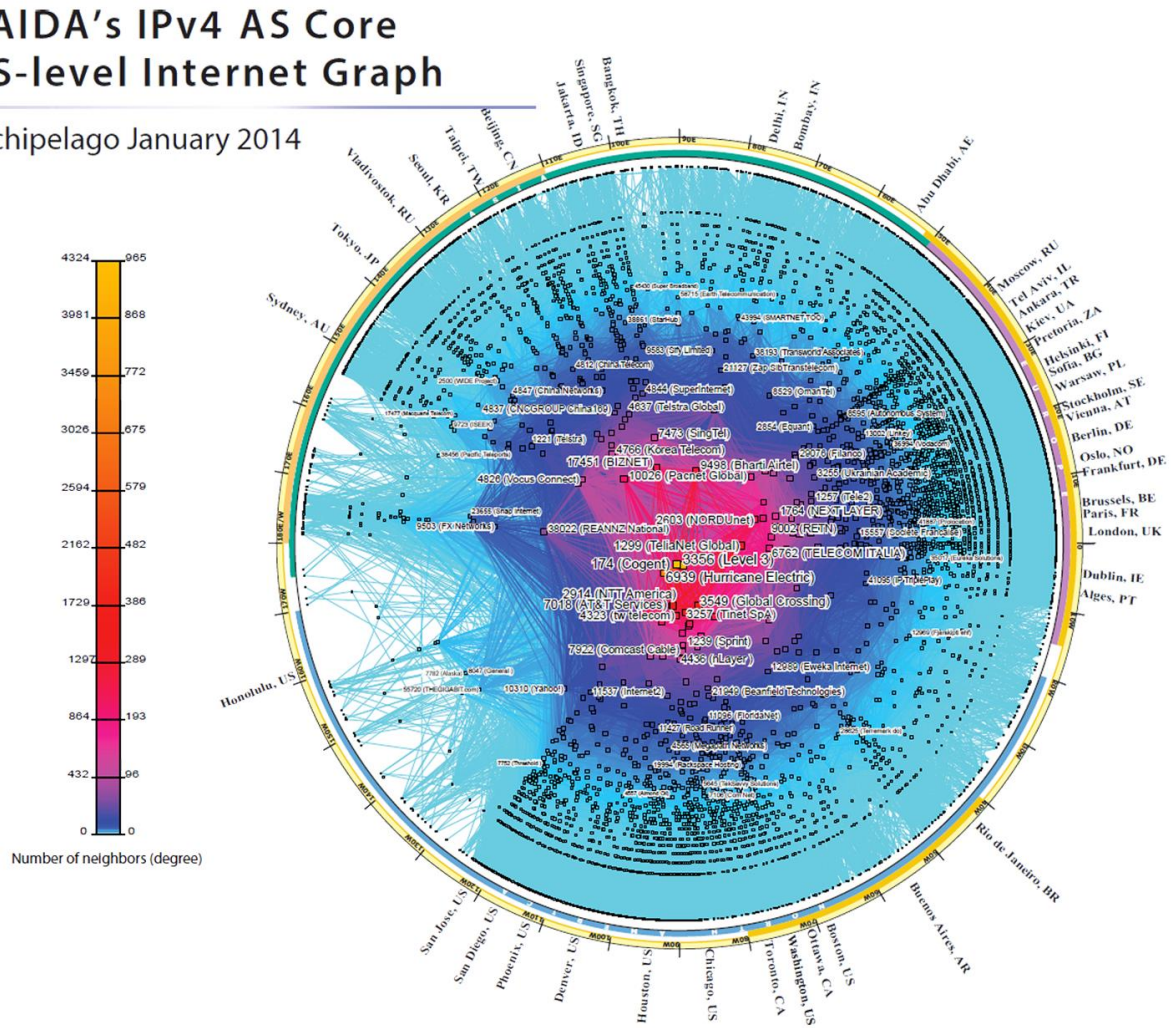
Measuring the Internet's topology

- What do we mean by topology?
  - Internet as graph
  - Edges? Nodes?
  - Node = Autonomous System (AS); edge = connection.
- Edges labeled with busin tionship
- Customer → Provider
- Peer -- Peer



The outputs ….

15412	12041	p2c
15412	12486	p2c
15412	12880	p2c
15412	13810	p2c
15412	15802	p2c
15412	17408	p2c
15412	17554	p2c
15412	17709	p2c
15412	18101	p2c
15412	19806	p2c
15412	19809	p2c
15413…		

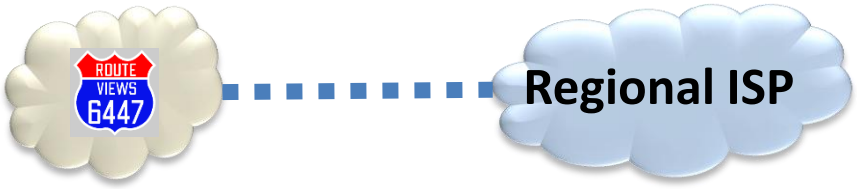


So how do we measure this graph?

- Passive approach: BGP route monitors
  - Coverage of the topology
  - Amount of visibility provided by each neighbor
- Active approach: Traceroute
  - From where?
  - Traceroute gives series of IP addresses not ASes

Passive approach: BGP Route Monitors

- Receive BGP announcements from participating ASes at multiple vantage points



www.routeviews.org

"originally motivated by interest on the part of operators in determining how the global routing system viewed their prefixes and/or AS space"

www.routeviews.org

Going from BGP Updates to a Topology

- Example update:
- TIME: 03/22/11 12:10:45
- FROM: 12.0.1.63 AS7018
- TO: 128.223.51.102 AS6447
- ASPATH: 7018 4134 9318 32934 32934 32934
- 69.171.224.0/20

AT&T (AS7018) it telling Routeviews (AS 6447) about this route.

This /20 prefix can be reached via the above path

Going from BGP Updates to a Topology

- Key idea
  - The business relationships determine the routing policies
  - The routing policies determine the paths that are chosen
  - So, look at the chosen paths and infer the policies
- Example: AS path “7018 4134 9318” implies
  - AS 4134 allows AS 7018 to reach AS 9318
  - China Telecom allows AT&T to reach Hanaro Telecom
  - Each “triple” tells something about transit service

Why are peering links hard to see?

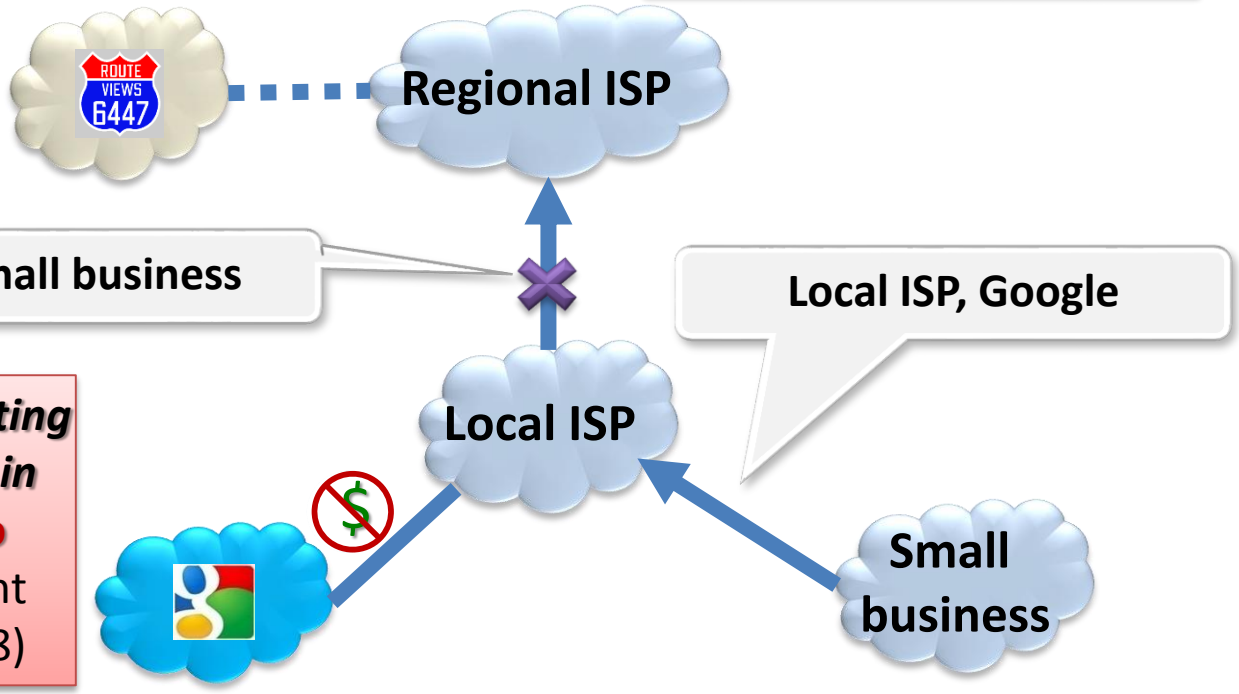
- **The challenge:**
  - BGP announcements ***do not reflect complete connectivity*** information
  - They are an agreement to transit traffic for the AS they are advertised to…

Regional ISP won't see the peering edge!

Neither will Routeviews

Local ISP will **only tell its customers** about the peering link.

(ASes only transit traffic if it generates revenue!)

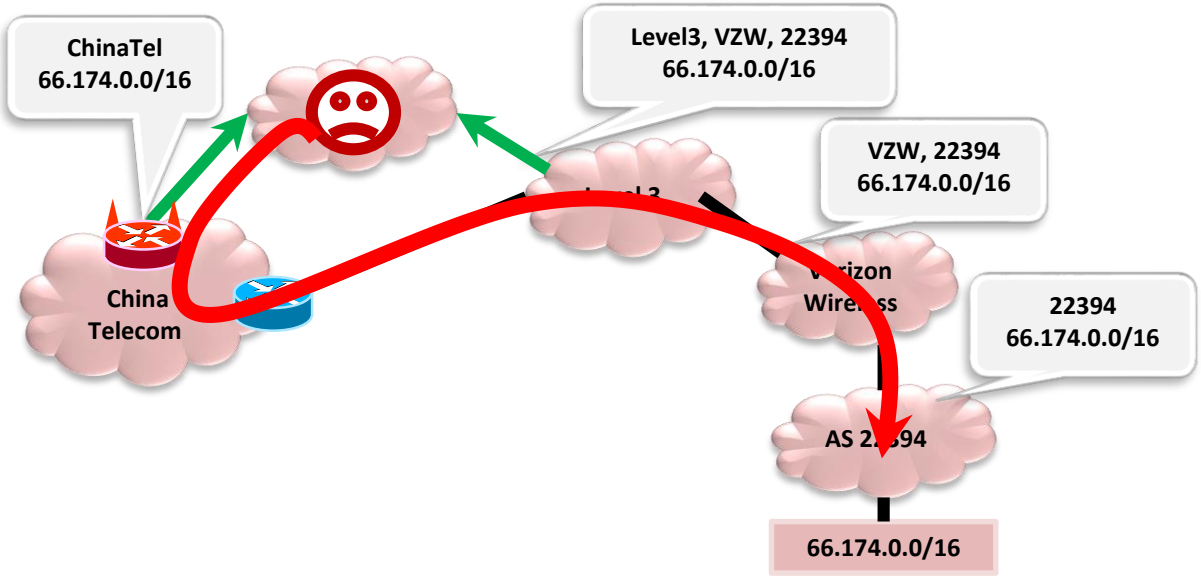


Combination of ***no valley routing policy*** and a ***lack of monitors in stub ASes*** mean missing **up to 90%** of peering links of content providers! (Oliveria et al. 2008)

Active approach: Traceroute

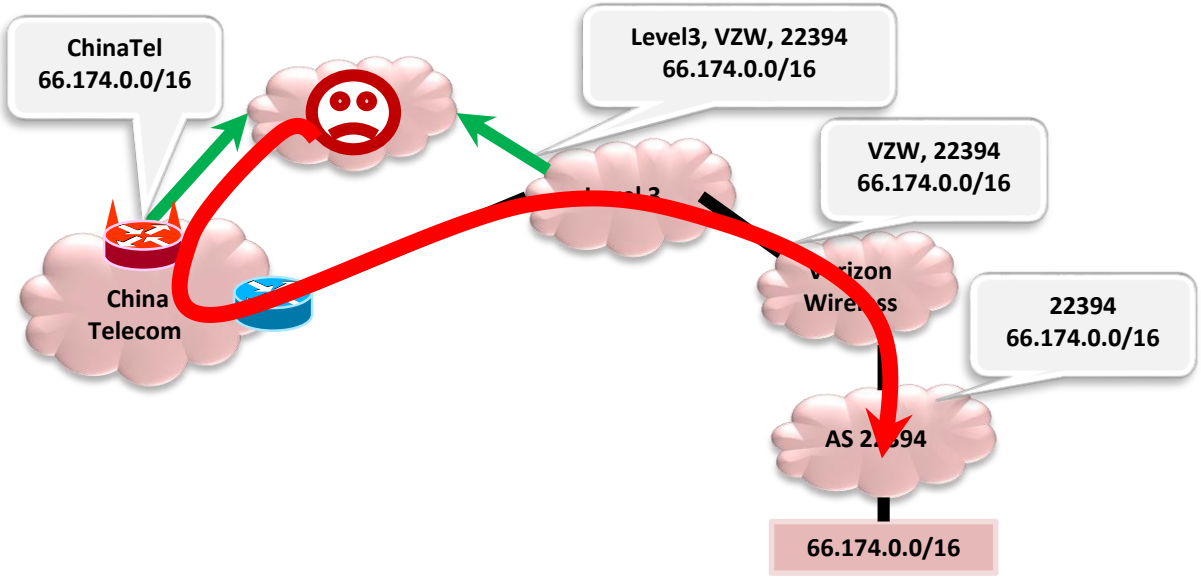
- Issue: Need control over end hosts to run traceroute
  - How to get VPs?
- http://www.traceroute.org/
  - Collection of O(100) servers that will run traceroute
  - Hosted by ISPs/other network operators (e.g. universities)
- RIPE Atlas
  - Distribute specialized hardware to volunteers
  - O(1000s) of probes
- Dasu
  - Bittorrent plug in that does measurements
  - O(200) ASes with Dasu clients

Traceroute vs Announced Path



Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident, R. Hiran et al., Proc. PAM 2013.

Traceroute vs Announced Path



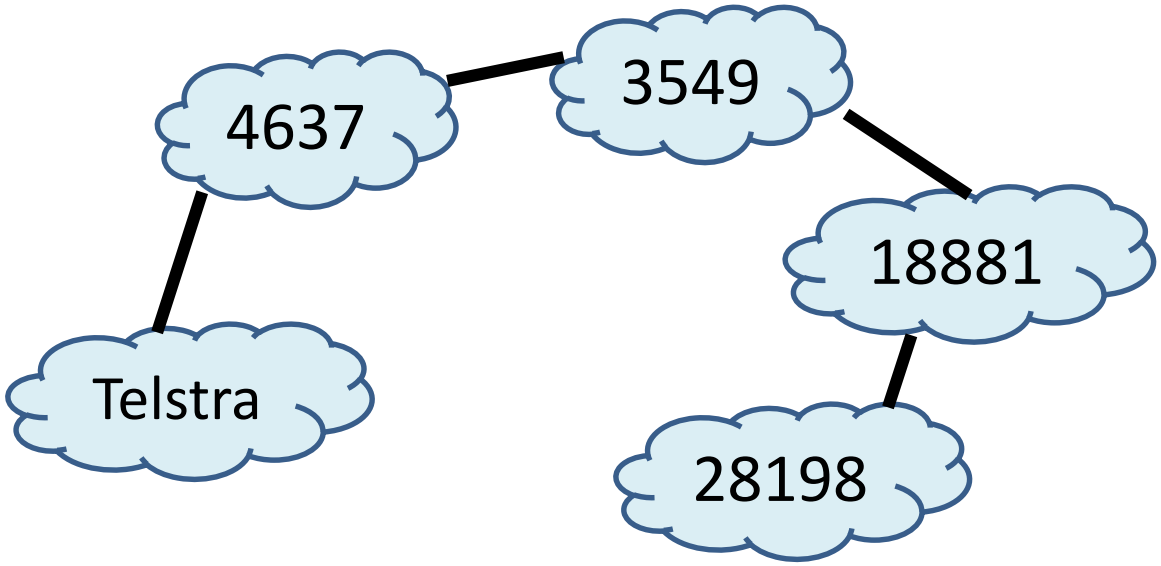
Interception typically results in differences between

- Announced AS-PATH
- Data path (traffic)

Policy checks if legit reason(s)

Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident, R. Hiran et al., Proc. PAM 2013.

Traceroute vs Announced Path

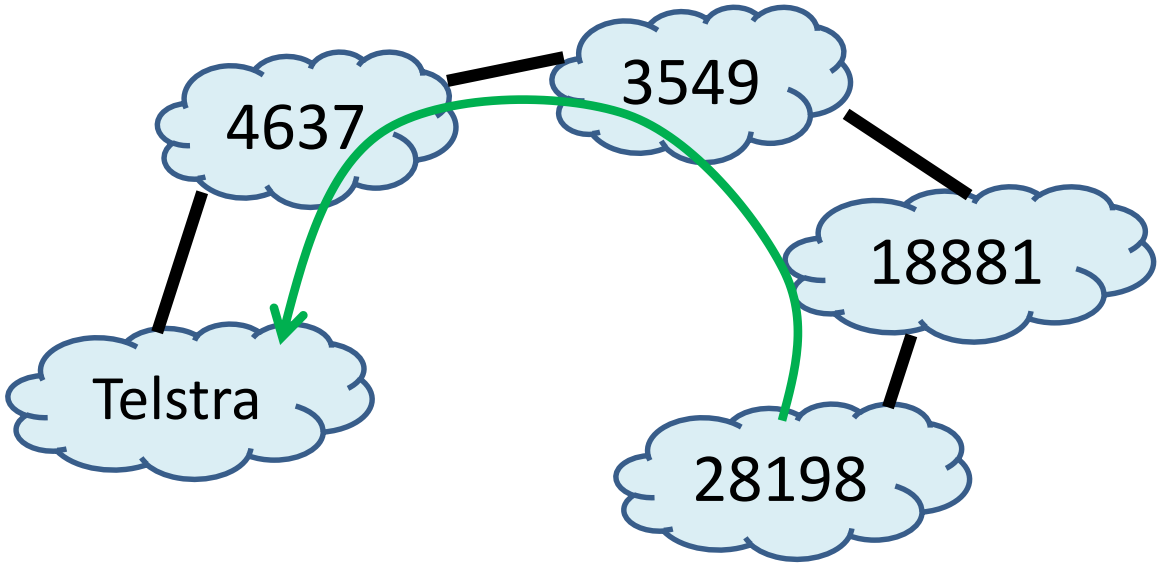


Sometimes differences

- Announced AS-PATH
- Data path (traffic)

Many legit reason(s)

Traceroute vs Announced Path



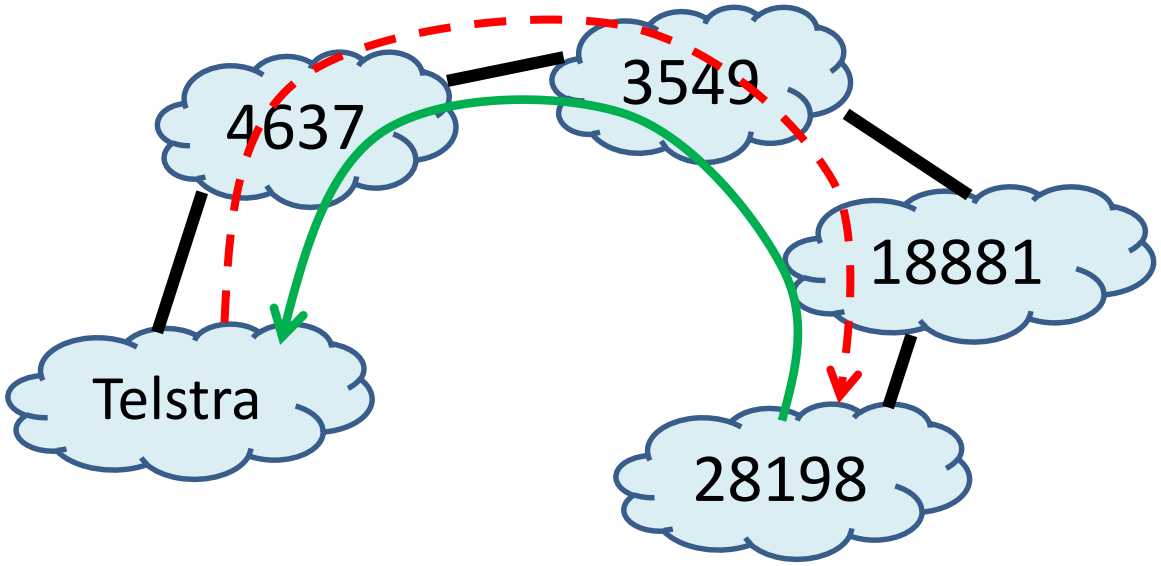
Sometimes differences

- Announced AS-PATH
- Data path (traffic)

Many legit reason(s)

AS-PATH: 177.52.48.0/21|1221 4637 3549 18881 28198

Traceroute vs Announced Path



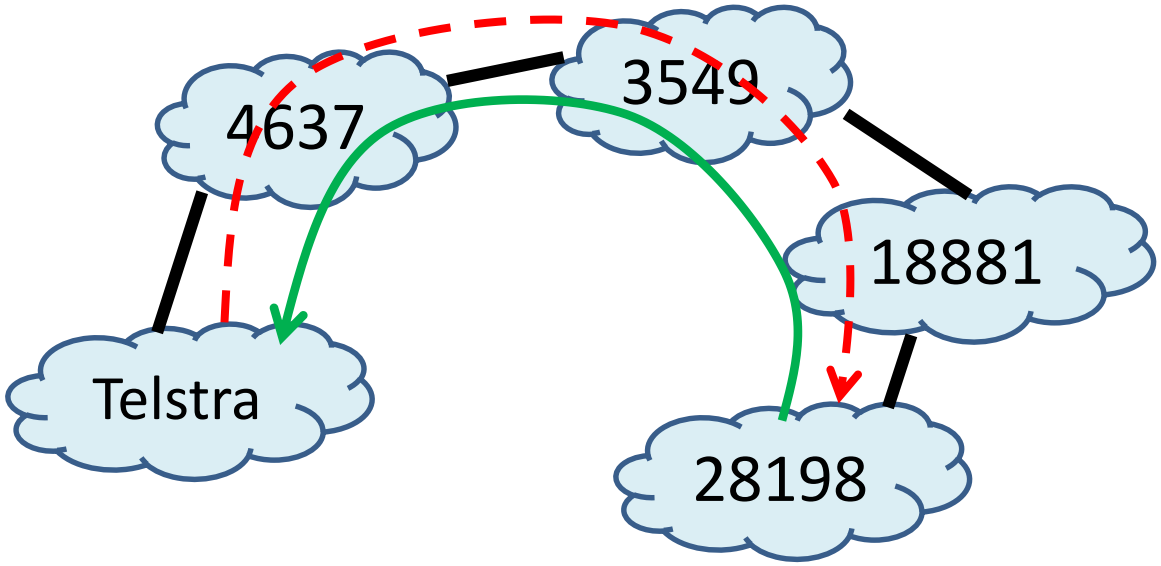
Sometimes differences

- Announced AS-PATH
- Data path (traffic)

Many legit reason(s)

AS-PATH: 177.52.48.0/21|1221 4637 3549 18881 28198

Traceroute vs Announced Path



Sometimes differences

- Announced AS-PATH
- Data path (traffic)

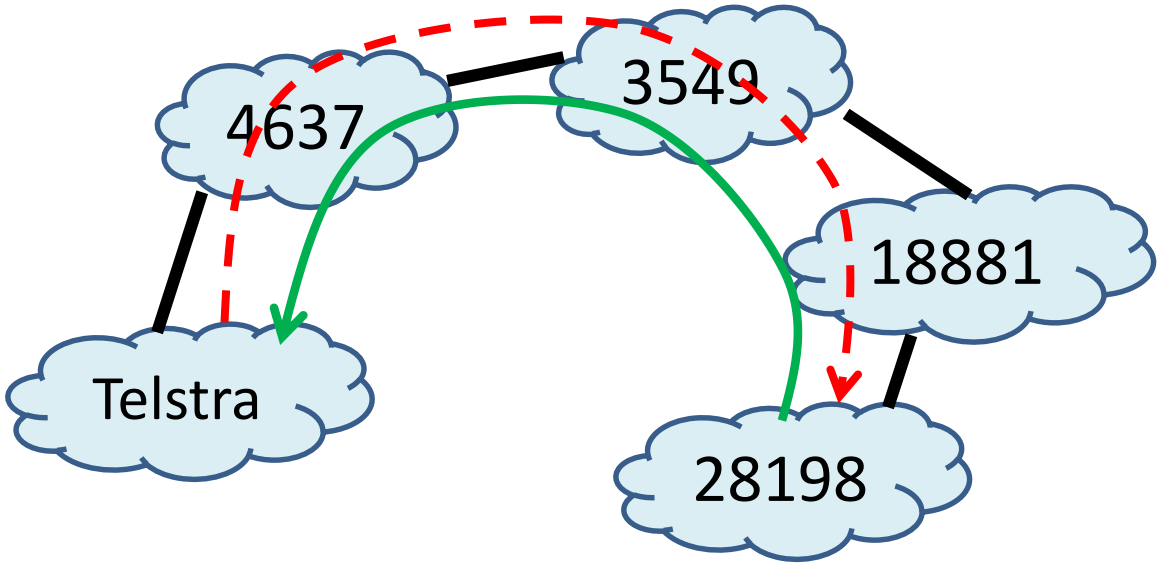
Many legit reason(s)

AS-PATH: 177.52.48.0/21|1221 4637 3549 18881 28198

Traceroute:

```
... (initial hops)
9.  telstraglobal.net (134.159.63.202)  164.905 ms
10  impsat.net.br (189.125.6.194)  337.434 ms
11  spo.gvt.net.br (187.115.214.217)  332.926 ms
12  spo.gvt.net.br (189.59.248.109)  373.021 ms
13  host.gvt.net.br (189.59.249.245)  343.685 ms
14  isimples.com.br (177.52.48.1)  341.172 ms
```

Traceroute vs Announced Path



Sometimes differences

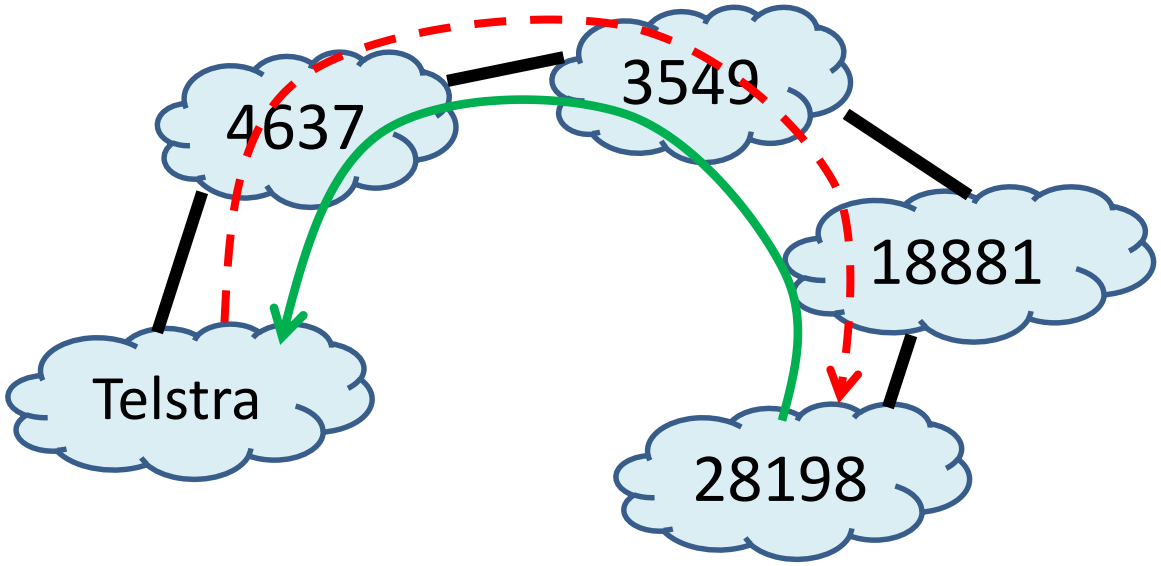
- Announced AS-PATH
- Data path (traffic)

Many legit reason(s)

AS-PATH: 177.52.48.0/21|1221 4637 3549 18881 28198

AS HOPS in traceroute: 1221 1221 1221 1221 4637 4637 4637 4637 4637 3549 3549 3549 18881 18881 18881 18881 28198

Traceroute vs Announced Path



Sometimes differences

- Announced AS-PATH
- Data path (traffic)

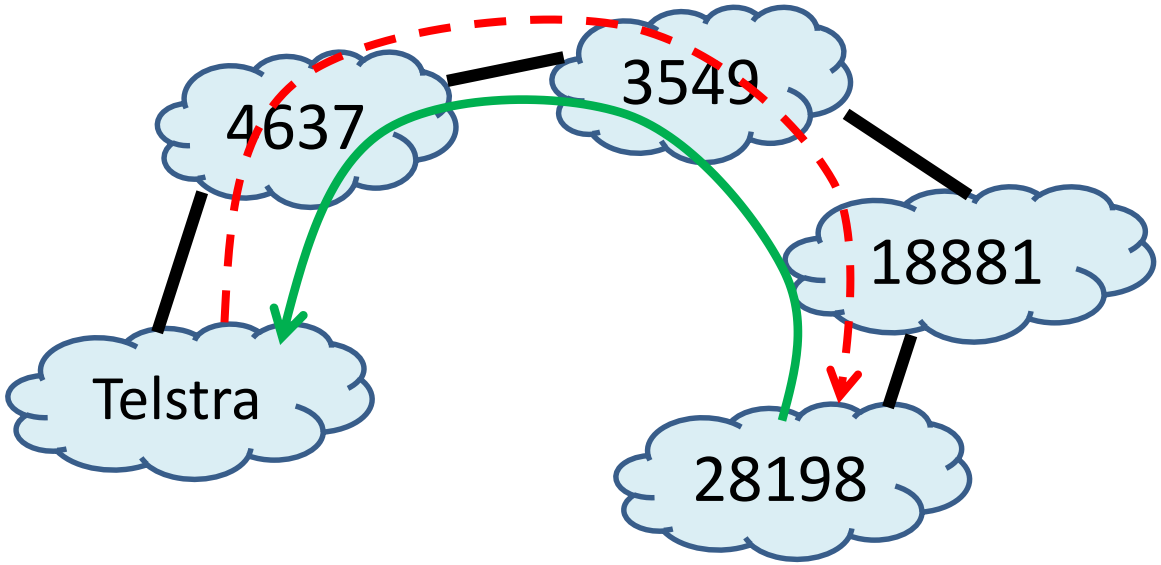
Many legit reason(s)

AS-PATH: 177.52.48.0/21|1221 4637 3549 18881 28198

AS HOPS in traceroute: 1221 1221 1221 1221 4637 4637 4637 4637 4637 3549 3549 3549 18881 18881 18881 18881 28198

Traceroute-PATH: 1221 4637 3549 18881 28198

Traceroute vs Announced Path



Sometimes differences

- Announced AS-PATH
- Data path (traffic)

Many legit reason(s)

AS-PATH: 177.52.48.0/21| 1221 4637 3549 18881 28198

AS HOPS in traceroute: 1221 1221 1221 1221 4637 4637 4637 4637 4637 3549 3549 3549 18881 18881 18881 18881 28198

Traceroute-PATH: 1221 4637 3549 18881 28198