Traceroute and BGP AS Path Incongruities

Young Hyun, Andre Broido, kc claffy

CAIDA, San Diego Supercomputer Center
University of California, San Diego
{**youngh, broido, kc**}**@caida.org**

Internetworking 2003, San Jose, June 2003

Background

Motivation

Methodology

- data collection
- initial processing

Analysis

- exchange point ASes
- ASes under common ownership
- remaining causes

Conclusions

Background

- AS-level Internet topology is very useful . . .
  - for studying growth, performance, resiliency, convergence times
  - for designing routing protocols
- complete, up-to-date topology not available
  - only two practical sources of *partial* topology:
    - BGP tables (e.g., at RouteViews and RIPE)
    - AS paths derived from traceroute paths
- most analysis/modeling of Internet topology based on BGP AS paths

Motivation

Are Internet topologies based on BGP AS paths valid?

- answer by comparing two topology sources: BGP tables and traceroute paths
- simplistically:

  BGP AS path = *specified* (by policy)

  traceroute AS path = *actual* (by per-hop forwarding decisions)
- expect specified and actual paths to agree, but they differ in practice
- want to know the extent and causes of incongruities

Methodology

1. collect data at three sites worldwide
   - BGP table from router near the host performing traceroutes
2. convert traceroute IP paths to AS paths
3. match up traceroute AS paths with BGP AS paths
4. compare pairs of AS paths

Source of traceroute paths

- CAIDA's skitter monitors
  - around two dozen deployed worldwide
  - TTL-based like `traceroute` but using ICMP `ECHO_REQUEST`
  - probe predetermined set of addresses ("destination list")
- chose three monitors based on geographical diversity and availability of BGP table nearby

monitor	location	network
`sjc`	San Jose, CA	MFN/AboveNet
`k-peer`	Amsterdam	RIPE, near AMS-IX
`m-root`	Tokyo	WIDE, near NSPIXP

Destination lists used

- **IPv4** with 302k dests: `sjc`
  - broad cross-section of Internet hosts
  - e.g., web servers, backbone routers, business desktops, consumer dial-up/broadband desktops
- **DNS** with 143k dests: `k-peer`, `m-root`
  - clients of DNS root servers
- IPv4 and DNS lists have 24k dests in common

Data collected

- on Apr 1, 2002
- keep only *complete* traceroute paths—destination and all intermediate hops responded

	sjc	**k-peer**	**m-root**
complete paths	220k	90k	89k
% all paths	73%	63%	62%
BGP prefixes	108k	116k	116k

Pairing of AS paths

- pair up traceroute and BGP AS paths based on prefix of traceroute destination
- can have several destinations per prefix ⇒ several traceroute IP paths per prefix
  - reduce to *distinct* traceroute AS paths *per prefix* to avoid overrepresentation of any one prefix
    - avg. 97% of prefixes have only one distinct traceroute AS path

	sjc	**k-peer**	**m-root**
distinct traceroute AS paths	60,271	36,950	38,527
BGP prefixes with paths	58,037	36,170	37,292
% all prefixes	54%	31%	32%

Analysis

Terminology

- **traceroute path** for traceroute *AS* path
    - no more discussion of *IP* paths
- **BGP path** for BGP AS path

Incongruent paths

- Def: A traceroute path is **incongruent** to a BGP path if the paths don't have the same sequence of ASes.

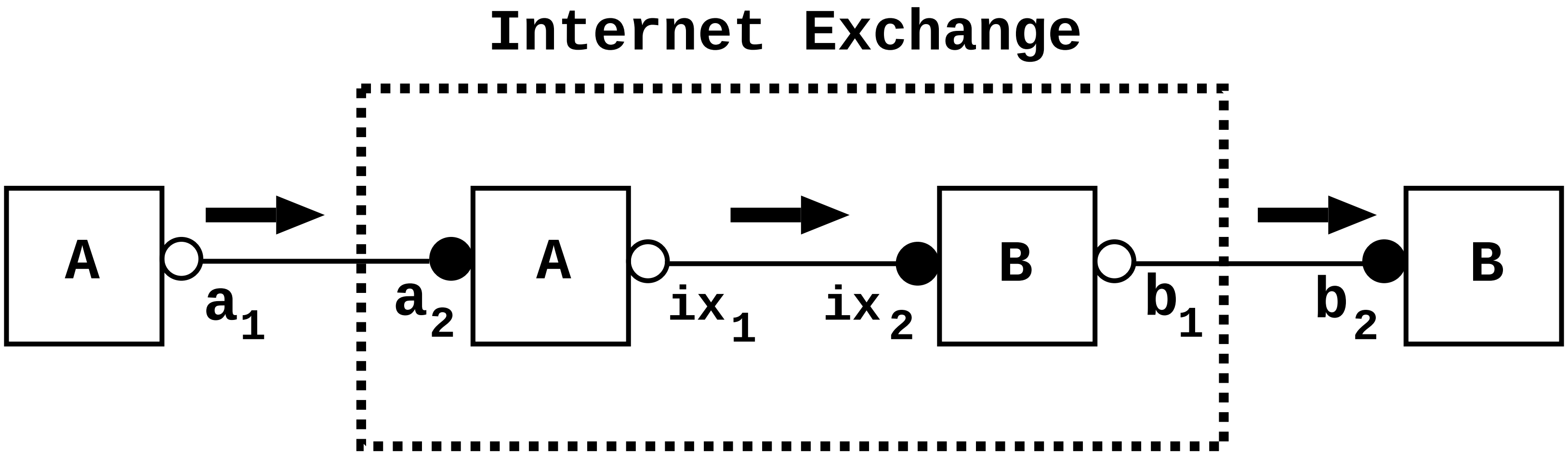
	sjc	**k-peer**	**m-root**
all distinct paths	60,271	36,950	38,527
incongruent paths % of all distinct	11,297 19%	36,888 99.8%	38,460 99.8%

Causes of incongruities

- exchange point ASes
- ASes under common ownership
- other causes

Exchange point ASes

- Def: An **exchange point (IX) AS** is an AS number belonging to an IX that is used to announce prefixes assigned to the routers at the IX.
  - e.g., 6695 ⇒ DE-CIX; 5459 ⇒ LINX; 1200 ⇒ AMS-IX
- appear in traceroute paths:



expect AS path	$A\ B$
get IP path get AS path	$a_2\ ix_2\ b_2$ A *IX* B

Exchange point ASes cont'd

- IX ASes are significant cause of incongruity

cause of incongruity	sjc	k-peer	m-root
involving IX ASes	4,461 (40%)	36,884 (100%)	31,701 (82%)
● only IX ASes	3,749 (33%)	30,163 (82%)	20,601 (54%)
● IX & non-IX ASes	712 (6%)	6,721 (18%)	11,100 (29%)
only non-IX ASes	6,818 (60%)	4 (0%)	6,759 (18%)
total: incongruent paths	11,279	36,888	38,460

- most paths of `k-peer` and `m-root` cross nearby IX; hence greater impact
  - but see IX ASes in paths regardless of traceroute source location
  - e.g., IXes neer `k-peer` and `m-root` have been excluded below:

# IX ASes per path	sjc	k-peer	m-root
1+	5,725 (9%)	1,070 (3%)	4,198 (11%)
1	5,648 (9%)	1,052 (3%)	4,060 (11%)
2	77 (0%)	18 (0%)	118 (0%)
3	0	0	20 (0%)
total: distinct paths	60,271	36,950	38,527

ASes under common ownership

- many organizations have several AS numbers
  - after merger or acquisition
  - for convenience implementing routing policy, such as segregating:
    - academic vs. commercial traffic
    - transit vs. customer traffic

- some closely related organizations
  - MCI/WorldCom/UUNET/AlterNet/ANS/Bertelsmanns
  - SBC/Pacific Bell/Nevada Bell/Southwestern Bell
  - C&W/Exodus/PSI
  - Qwest/US West/SuperNet/Touch America

- impacts topology analysis
  - e.g., want "peering between *organizations*", not "peering between AS numbers"

- different concept than "sibling ASes"—organizations under *separate* ownership that provide mutual transit

Common ownership cont'd

- during comparison, two AS numbers match if
  1. numerically equal
  2. under common ownership
- incongruities due to common ownership ($B \equiv B'$):

```
BGP            A  B     C
Traceroute     A  B  B′ C

BGP            A  B  C
Traceroute     A  B′ C
```

- breakdown of incongruities by cause:

cause of incongruity	**sjc**		**k-peer**		**m-root**	
common ownership & IX ASes	2,711	(24%)	1,464	(4%)	932	(2%)
only IX ASes	3,749	(33%)	30,163	(82%)	20,601	(54%)
other causes	4,819	(43%)	5,261	(14%)	16,927	(44%)
total: incongruent paths	11,279		36,888		38,460	

Analysis of remaining incongruent paths

- compared paths in terms of editing distance
  - minimal amount of change needed to convert BGP path to traceroute path (cf. Unix `diff` program)
  - *insertions*, *deletions*, and *substitutions* of one or more ASes
- delete 11422, insert 1

```
BGP            207.99.128.0/17    6461 209 11422 2151   2920
Traceroute     207.99.161.1       6461 209       2151 1 2920
```

- substitute (3549 701 1) for (209)

```
BGP            216.152.160.0/20   6461      209   11081
Traceroute     216.152.163.248    6461 3549 701 1 11081
                                       ----------
```

Analysis cont'd

- examined incongruent paths not caused *entirely* by IX ASes or common ownership
- most traceroute paths longer than corresponding BGP paths

traceroute path	sjc		k-peer		m-root	
longer	3,125	(65%)	3,673	(70%)	15,765	(93%)
equal	474	(10%)	1,533	(29%)	1,126	(7%)
shorter	1,220	(25%)	103	(2%)	36	(0%)
total: remaining paths	4,819		5,216		16,927	

- mostly insertions in traceroute paths

operation	sjc		k-peer		m-root	
insertions only	2,788	(58%)	2,764	(53%)	13,661	(81%)
deletions only	1,132	(23%)	1	(0%)	0	(0%)
substitutions only	813	(17%)	1,813	(34%)	2,648	(16%)
mixture	86	(2%)	683	(13%)	618	(4%)
total: remaining paths	4,819		5,216		16,927	

Analysis cont’d

- case: ASes appended only

```
BGP            A  B  C
Traceroute     A  B  C  D  E
```

  - 1,357 paths in `sjc`, 0 in `k-peer`, 2 in `m-root`
  - speculate DNS clients located at provider (not customer) premises

- case: entire path differs, except source and destination

```
BGP            A  B  C  D
Traceroute     A  X  Y  D
```

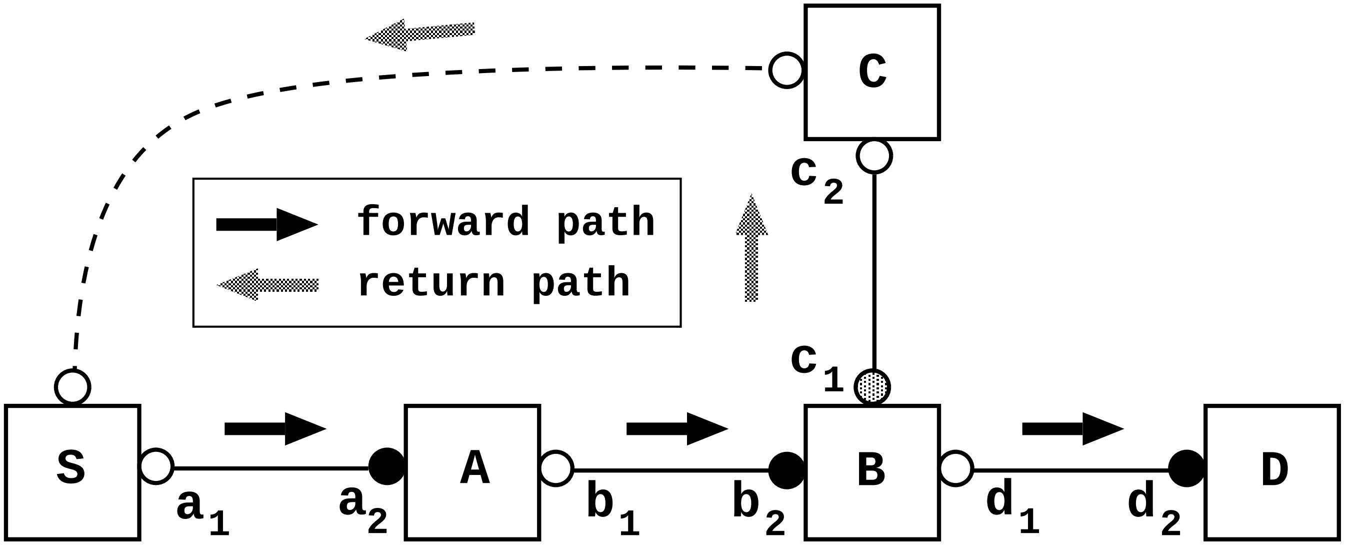
  - 563 paths in `sjc`, 233 in `k-peer`, 251 in `m-root`
  - speculate routing change

Suspected causes of remaining incongruities

1. inaccurate conversion of traceroute paths to AS paths:
   - fundamentally difficult to identify the AS owning the routers seen in traceroute paths
   - made worse by:
     - IP addresses without matching BGP prefixes
     - IX prefixes announced by IX participants
     - less precise mapping due to BGP prefix aggregation/filtering
2. mid-path routing change:
   - single traceroute path reflecting more than one path due to route change or load balancing
3. third-party addresses:
   - traceroute path containing hops not in the actual forward path
   - related work (see below) suggests impact is minimal
4. use of BGP table snapshot rather than BGP updates:
   - BGP route may have changed during the 7–9 hours needed to perform traceroutes

What are third-party addresses?

- addresses in *return* path, not forward path
  - RFC1812: ICMP response packet should have source address set to *outgoing* interface.
- can cause incorrect AS path:



expect IP path expect AS path	$a_2\ b_2\ d_2$ $A\ B\ D$
get IP path get AS path	$a_2\ c_1\ d_2$ $A\ C\ D$

Conclusions

- **IX ASes** and **common ownership** are significant causes of incongruity
  - treating each AS number separately can
    1. miss relationships between *organizations*
    2. lead to incorrect topology models
- analysis of remaining incongruities suggests a diversity of causes
- topologies derived from traceroute and BGP paths differ

Resources

- “Traceroute and BGP AS Path Incongruities”,
  `<www.caida.org/outreach/papers/2003/ASP/>`
- “On Third-party Addresses in Traceroute Paths”, PAM2003,
  `<www.caida.org/outreach/papers/2003/3rdparty/>`