Internet AS relationships, Routing policy on Internet paths

Z. Morley Mao

Lecture 5

Jan 20, 2005

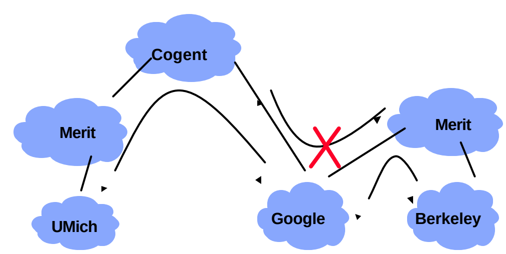
BGP route propagation

- Connectivity does not imply reachability
- Not all possible routes propagate
- Commercial relationships determine policies for
  - Route import
  - Route selection
  - Route export
- Typical relationships
  - Provider-customer: customer pay money for transit
  - Peer-peer: typically exchange respective customers' traffic for **free**

Transit vs. peering

- ISP definition:
  - Internet service provider is an organization tha tsells access to the Internet
- Transit definition:
  - "Business relationship whereby one ISP provides (usually sells) access to all destinations in its routing table".
- Peering is non-transitive relationship
  - A peers with B, B peers with C, does not imply A peers with C
- Peering definition:
  - "An interconnection business relationship whereby ISPs provide connectivity to each others' **transit** customers."
- Hybrid exists
  - Regional transit
  - Paid peering

Example of commercial relationship



Tier1 vs Tier2 peering

- Tier 1 ISPs
  - Buy no transit from any other providers
  - Have only customers and peers
  - Has full mesh peering with other Tier 1's
  - Motivation for peering:
    - Minimize their interconnection costs while providing sufficient interconnection BW to support customer and its growth
- Tier 2 ISPs
  - ISP that purchases (resells) transit within an Internet region

Benefit of tier-2 peering

- Decreases the cost and reliance on purchased Internet transit
- Lowers inter-AS traffic latency
  - Fewer AS hops, AS peering links traversed
- Is peering always better than transit?
- Concerns of peering:
  - Traffic asymmetry
  - No SLAs: less liability or incentive to improve performance
  - "free" rather than getting paid
  - Peers become more powerful

Peering Wars

Peer

- Reduces upstream transit costs
- Can increase end-to-end performance
- May be the only way to connect your customers to some part of the Internet ("Tier 1")

Don't Peer

- You would rather have customers
- Peers are usually your competition
- Peering relationships may require periodic renegotiation

Peering struggles are by far the most contentious issues in the ISP world!

Peering agreements are often confidential.

Where to peer?

- Public peering: at public peering locations
- Private peering
- Exchange-based interconnection model
  - A meet point at which ISPs exchange traffic
  - Can be neutral Internet business exchange
- Direct circuit interconnection model
  - Point-to-point circuit between the exchange parties

Four Types of BGP Messages

- **Open** : Establish a peering session.
- **Keep Alive** : Handshake at regular intervals.
- **Notification** : Shuts down a peering session.
- **Update** : Announcing new routes or withdrawing previously announced routes.

announcement
=
prefix + attributes values

Policy with BGP

- BGP provides capability for enforcing various policies
- Policies are **not** part of BGP: they are provided to BGP as configuration information
- BGP enforces policies by choosing paths from multiple alternatives and controlling advertisement to other AS's
- Import policy
  - What to do with routes learned from neighbors?
  - Selecting best path
- Export policy
  - What routes to announce to neighbors?
  - Depends on relationship with neighbor

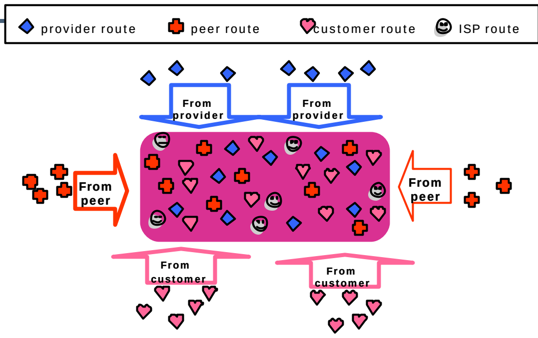
Examples of BGP Policies

- A multi-homed AS refuses to act as transit
  - Limit path advertisement
- A multi-homed AS can become transit for some AS's
  - Only advertise paths to some AS's
  - Eg: A Tier-2 provider multi-homed to Tier-1 providers
- An AS can favor or disfavor certain AS's for traffic transit from itself

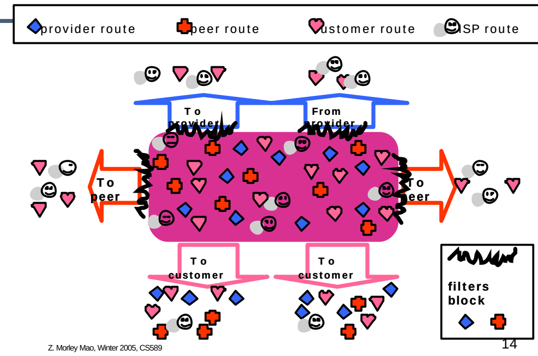
Export Policy

- An AS exports only best paths to its neighbors
  - Guarantees that once the route is announced the AS is willing to transit traffic on that route
- To Customers
  - Announce all routes learned from peers, providers and customers, and self-origin routes
- To Providers
  - Announce routes learned from customers and self-origin routes
- To Peers
  - Announce routes learned from customers and self-origin routes

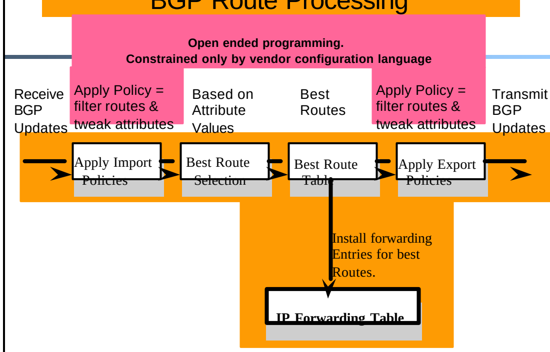
Import Routes



Export Routes



BGP Route Processing



BGP UPDATE Message

- List of withdrawn routes
- Network layer reachability information
  - List of reachable prefixes
- Path attributes
  - Origin
  - Path
  - Metrics
- All prefixes advertised in message have same path attributes

Path Selection Criteria

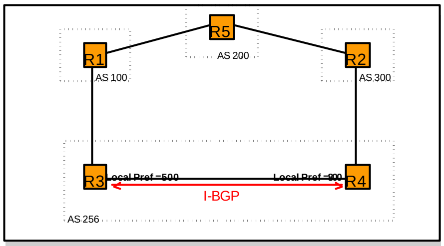
- Information based on path attributes
- Attributes + external (policy) information
- Examples:
  - Hop count
  - Policy considerations
    - Preference for AS
    - Presence or absence of certain AS
  - Path origin
  - Link dynamics

Important BGP Attributes

- Local Preference
- AS-Path
- MED
- Next hop

LOCAL PREF

- Local (within an AS) mechanism to provide relative priority among BGP routers

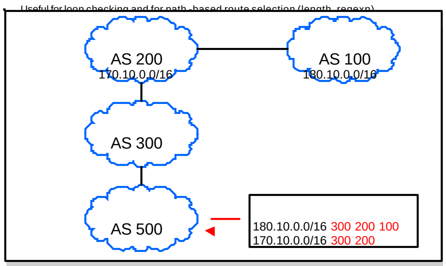


LOCAL PREF – Common Uses

- Handle routes advertised to multi-homed transit customers
  - Should use direct connection (multihoming typically has a primary/backup arrangement)
- Peering vs. transit
  - Prefer to use peering connection, why?
- In general, customer > peer > provider
  - Use LOCAL PREF to ensure this

AS_PATH

- List of traversed AS's
- Useful for loop checking and for path-based route selection (length, regexp)

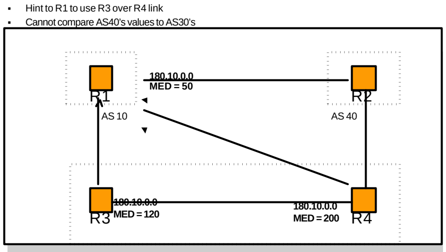


Multi-Exit Discriminator (MED)

- Hint to external neighbors about the preferred path into an AS
  - Non-transitive attribute
  - Different AS choose different scales
- Used when two AS's connect to each other in more than one place

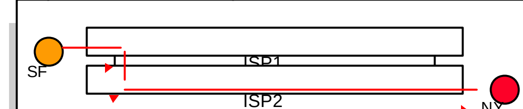
MED

- Typically used when two ASes peer at multiple locations
- Hint to R1 to use R3 over R4 link
- Cannot compare AS40's values to AS30's



MED

- MED is typically used in provider/subscriber scenarios
- It can lead to unfairness if used between ISP because it may force one ISP to carry more traffic:



- ISP1 ignores MED from ISP2
- ISP2 obeys MED from ISP1
- ISP2 ends up carrying traffic most of the way

Other Attributes

- ORIGIN
  - Source of route (IGP, EGP, other)
- NEXT_HOP
  - Address of next hop router to use
- Check out http://www.cisco.com for full explanation
- Question: Too many choices/ attributes how to select routes !

Route Selection Process

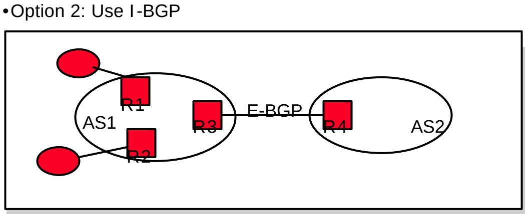
Highest Local Preference — **Enforce relationships**

Shortest ASPATH
Lowest MED
i-BGP < eBGP
Lowest IGP cost to BGP egress — **traffic engineering**

Lowest router ID — **Throw up hands and break ties**

Internal vs. External BGP

•BGP can be used by R3 and R4 to learn routes
•How do R1 and R2 learn routes?
•Option 1: Inject routes in IGP
  •Only works for small routing tables
•Option 2: Use I-BGP



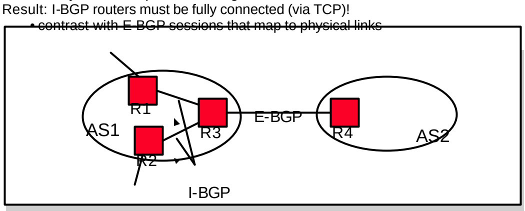
Internal BGP (I-BGP)

- Same messages as E-BGP
- Different rules about re-advertising prefixes:
  - Prefix learned from E-BGP can be advertised to I-BGP neighbor and vice- versa, but
  - Prefix learned from one I-BGP neighbor cannot be advertised to another I-BGP neighbor
  - Reason: no AS PATH within the same AS and thus danger of looping.

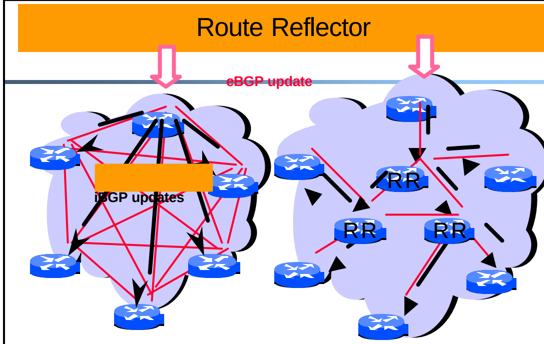
Internal BGP (I-BGP)

• R3 can tell R1 and R2 prefixes from R4
• R3 can tell R4 prefixes from R1 and R2
• R3 cannot tell R2 prefixes from R1

R2 can only find these prefixes through a *direct connection* to R1
Result: I-BGP routers must be fully connected (via TCP)!
  • contrast with E-BGP sessions that map to physical links



Route Reflector



Mesh does not scale

Each RR passes only best routes, no longer N^2 scaling problem

Policy Impact

- Different relationships – Transit, Peering
- Export policies → selective export
- "Valley -free" routing
  - Number links as (+1, 0, -1) for customer-to-provider, peer and provider-to-customer
  - In any path should only see sequence of +1, followed by at most one 0, followed by sequence of -1

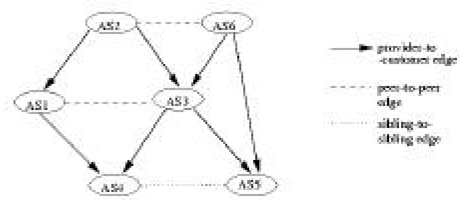
Why is it useful to infer AS relationships?

- Identify the AS-level hierarchy of Internet
  - Not shortest path routing
- Predict AS-level paths
- Traffic engineering
- Understand the Internet better
- Correlate with and interpret BGP update
- Identify BGP misconfigurations
  - E.g., errors in BGP export rules

AS relationships translate into BGP export rules

- Export to a provider or a peer
  - Allowed: its routes and routes of its customers and siblings
  - Disallowed: routes learned from other providers or peers
- Export to a customer or a sibling
  - Allowed: its routes, the routes of its customers and siblings, and routes learned from its providers and peers
- Valley-free:
  - After traversing a provider-customer or peer-peer edge, cannot traverse a customer-provider or peer-peer edge
  - Invalid path: >= 2 peer links, downhill-uphill, downhill-peer, peer-uphill

Example



- [1 2 3], [1 2 6 3] are valley -free
- [1 4 3], [1 4 5 3] are not valley free

Related work in the area of inferring AS relationships

- On inferring Autonomous Systems Relationships in the Internet [Gao]
  - Find the highest degree AS node to be the top provider of the AS path
  - Left to the top node: customer-provider or sibling-sibling links
  - Right to the top node: provider-customer or sibling-sibling links
  - Sibling-sibling: if providing mutual transit service for each other
  - Peer-peer: with top provider and of comparable degree value

What are siblings?

- Mutual transit agreement
  - Provide connectivity to the rest of the Internet for each other
- Typically between two administrative domains such as small ISPs or universities located close to each other, cannot afford additional Internet services for better connectivity

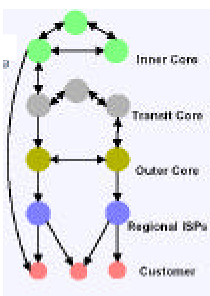
Assumptions of the Gao algorithm

- Provider is typically larger than its customers
- Two peers are typically of comparable size

Follow up work by Subramanian et al.

- Use BGP tables from multiple vantage points
  - More complete
  - Exploit uniqueness of each point
- Build AS-level hierarchy of Internet
  - Relationship based, not degree based
  - 5 level classification of AS's
- Relationship inference rules
  - Position of AS in AS graph gives rank
  - Combine ranks from multiple tables
  - Compare ranks:
    - Peer-peer with similar ranks
    - Provider-customer: provider with higher ranks

Hierarchy inference



- Internet hierarchy inference
- Based on relationships
- Not degree [Gao]

Level	# of AS's	E.g.
Inner core	20	AT&T
Transit core	129	UUNet Europe
Outer core	897	Turkish Telcom
Regional ISPs	971	PacBell
Customers	8898	UC Berkeley

A more recent work

- "*Computing the Types of the Relationships between Autonomous Systems*"
- Giuseppe Di Battista, Maurizio Patrignani, Maurizio Pizzonia (University of Rome III)
- Infocom 2003
- Cast it as an optimization problem to find provider-cutomer relationships that minimize the number of conflicts
- Shows the problem is NP-hard
- Do not deal with peer-peer relationships well

Quantifying the causes of path inflation [Spring03]

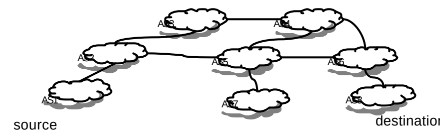
- Path inflation:
  - End-to-end paths are significantly longer than necessary
- Trace-driven study of 65 ISPs to characterize the root causes of path inflation
  - Topology and routing policy choices within an ISP, between pairs of ISPs, and across the global Internet
- High-level conclusion:
  - Peering policies and interdomain routing lead to significant inflation
  - Interdomain path inflation is due to lack of BGP policy to provide convenient engineering of good paths across ISPs

Findings

- Intra-domain traffic engineering is commonplace, but has minimal impact on path inflation
- There is significant cold-potato or non-early-exit routing between adjacent ISPs
  - To avoid poor routes, load-balance across multiple peering links
- Many early-exit paths are inflated
- Topology insensitive load balancing can cause significant path inflation
- Half of the path inflation is due to interdomain routing – using AS-path length as a routing metric

Internet Path Inflation

What path inflation is?



- To go from AS1 to AS8 instead of taking the shortest path:
  1-> 2 -> 5 -> 6 -> 8
  take a longer path like:
  1-> 2-> 3 -> 4 -> 6 -> 8

Outline

- *H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin. The impact of routing policy on Internet paths. In IEEE INFOCOM, 2001*
- *H. Tangmunarunkit, R. Govindan, S. Shenker. Internet path inflation due to policy routing. In SPIE ITCom, 2001*
- *L. Gao and F. Wang. The extent of AS path inflation by routing policies. In IEEE Global Internet Symposium, 2002*
- *N. Spring, R. Mahajan, and T. Anderson. Quantifying the causes of path inflation. In ACM SIGCOMM, 2003*

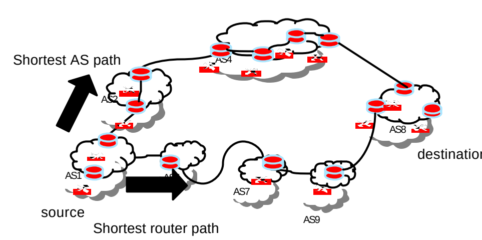
"The impact of routing policy on Internet paths"

H. Tangmunarunkit,
R. Govindan,
S. Shenker,
D. Estrin

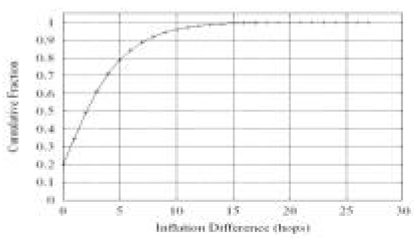
Methodology

- Create [Mar-Apr 2000] a router level map of the Internet using Mercator.
  "Heuristics for Internet Map Discovery" INFOCOMM 2000.
- Create an AS overlay map by assigning routers to ASs.
  - Use RouteViews BGP tables and RADB to find ASs.
- Compare router level path induced by shortest AS path routing with shortest router level path.

Shortest AS path inflation



Results



20% of the node pairs have a path 5 hop longer then the shortest path

Results

- Quantified the contribution of shortest AS path routing to path inflation.
- They also found that longer paths are more inflated.
- Shortcomings:
  - Overlooked policies applied between ASs.
  - Assumed shortest path intradomain routing.
  - Map size (2662 ASs is very small).

"Internet Path Inflation due to policy routing"

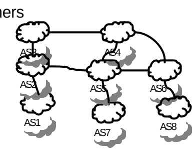
H. Tangmunarunkit,
R. Govindan,
S. Shenker

Methodology

- Re-examination of the previous work using a larger map; also consider interdomain policies.
- Create a router map using Mercator and an AS overlay.
- Infer policies between ASs.
- Assume a routing model and compare router level paths induced by the routing model and shortest router level paths.

Policies

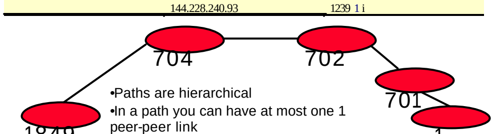
- Three types of peering relationships:
  - Provider-customer:
    - customer pays its provider for transit services
  - Peer-peer:
    - exchange traffic between customers
    - no money exchange
  - Sibling-sibling:
    - have mutual transit agreement
    - merging ISPs



"Interconnection, Peering and Settlements" G. Huston Internet Protocol Journal 1999.

"On inferring autonomous systems relationships in the Internet" L. Gao ACM IEEE Transactions on Networking 2001.

Network	Next hop	AS Path
4.2.24.0/21	134.24.127.3	1740 1 i
	194.68.130.254	5459 5413 1 i
	158.43.133.48	1849 704 702 701 1 i
	193.0.0.242	3333 286 1 i
	144.228.240.93	1239 1 i



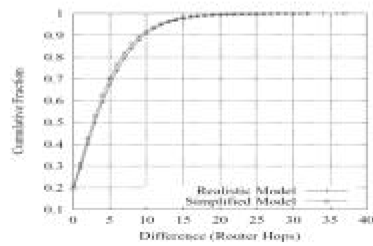
- Paths are hierarchical
- In a path you can have at most one 1 peer-peer link
- You go up the hierarchy through customer provider links (or sibling-sibling) and down the hierarchy through provider customer links (or sibling-sibling),

Figure taken from Lixin Gao

Routing Model

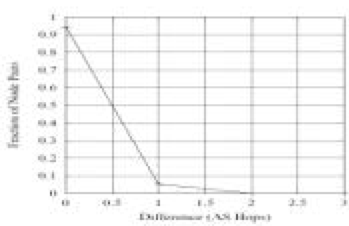
- A path transverses:
  - up the hierarchy through customer-provider links
  - down the hierarchy through provider-customer links
  - across the hierarchy through peer-peer links
- If more than one possible paths randomly pick one.

Results



Inflation difference by realistic and simplified routing policy model.

Results



AS path inflation caused by realistic routing model
95% of the paths have the same AS length

Conclusions

- Shortest AS path routing induces inflation
- Interdomain policies do not induce inflation

“The extent of AS path inflation by routing policies”

L. Gao
F. Wang

Methodology

- Create an AS map of the Internet from RouteViews data.
- Measure the extent of AS path inflation seen by RouteViews.
- Assume a routing model and measure AS path inflation.

Results

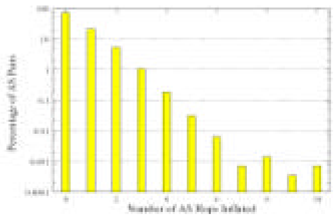
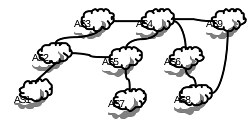


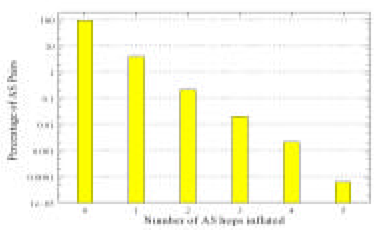
Fig. 7. Percentage of AS pairs whose chosen paths are visible from the Route View server are inflated by a fixed number

Routing Model

- No-valley routing policy:
  - An AS does not provide transit between any two of its providers or peers.
- Prefer Customer routing policy:
  - Prefer the free of charge customer route over the peer or provider route.

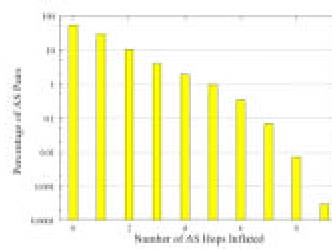


Results



Path inflation using a no-valley routing model.

Results



Path inflation using a no-valley and prefer customer routing model.

Conclusion

- Prefer customer routing model induces significant inflation.
- 45% of the paths are inflated by at least one AS hop.

“Quantifying the Causes of Path Inflation”

Neil Spring
Ratul Mahajan
Thomas Anderson

SIGCOMM 2003

Approach

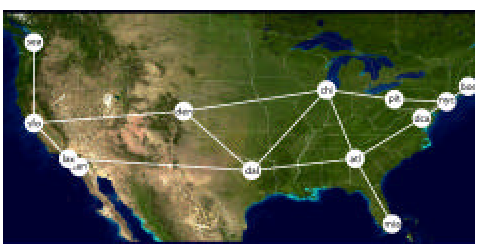
- Quantify Internet Path inflation in 3 layers.
- For each layer find topology and policy triggered inflation.

Interdomain Routing	⇨	Policies Shortest AS path routing
Peering	⇨	Routing between adjacent networks. “Hot potato routing” and MEDs.
Intradomain Routing	⇨	Intradomain routing protocol

Methodology

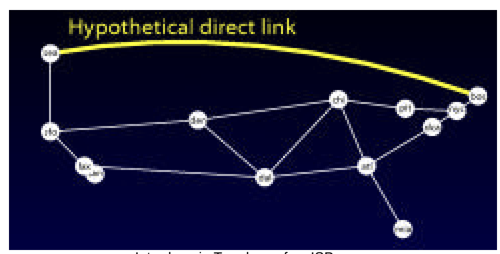
- Infer intradomain topology of 65 ISPs using Rocketfuel, use traceroutes measured from 42 vantage points.
- Choose mainly large ISPs to have interesting topologies and some smaller ISPs for diversity.
- Extract a PoP level map from the router level map.

Metric: Additive Latency



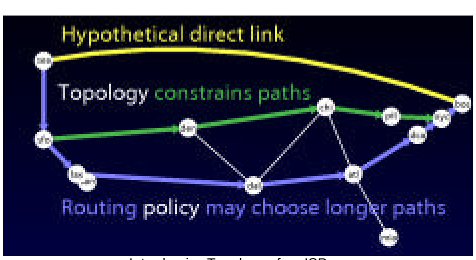
Intradomain Topology of an ISP,
Figure taken from N. Spring

Metric: Additive Latency



Intradomain Topology of an ISP,
Figure taken from N. Spring

Metric: Additive Latency

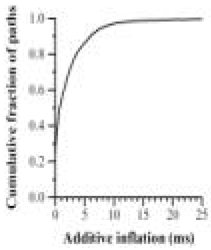


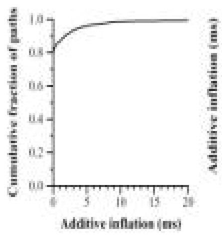
Intradomain Topology of an ISP,
Figure taken from N. Spring

Intradomain layer

- Infer intradomain policies using a constraint based approach.
- Intradomain topology does not cause lot of inflation pointing to well connected topologies.
- Intradomain policies do not cause lot of inflation, meaning that intradomain traffic engineering is not inconsistent with link latencies.

Intradomain layer results

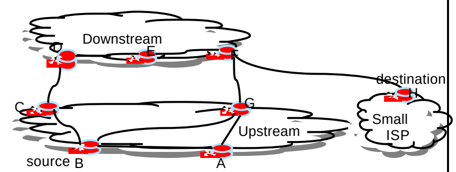




Intradomain Topology Inflation
mean = 3ms

Intradomain Policy Inflation
mean = 1ms

Characterize Peering Policies

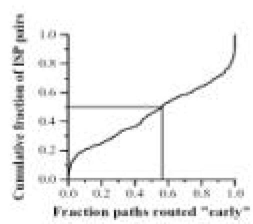


- BGP uses MEDs to indicate preferred links.
- Late exit (Cold potato routing): Use the link indicated (MED) from your neighbor (B->G->F->H).
- Early exit (Hot potato routing): Use the link closest to the source (B->C->D->E->F->H).

Characterize Traces

- Early exit: If one peering link point is seen from each ingress.
- Characterize as *late* if the path length in the downstream ISP from peering point to destination is less than from the early exit to destination. Use this metric to classify traces in three categories:
  - Late exit, often (late exit for most paths)
  - Late exit, sometimes (late exit for the minority of the paths)
  - Engineered, but not late (downstream carries traffic over longer paths)

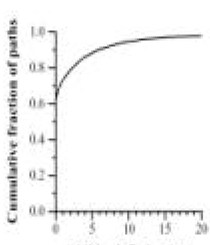
Characterization results



Tier 1 ISPs:
- Late exit, often (15%)
- Late exit, sometimes (10%)
- Early-exit (19%)
- Single peering point (42%)
- Engineered but not late (13%)

Median is 57% meaning that most ISPs use early exit most of the time.

Peering Policies Inflation



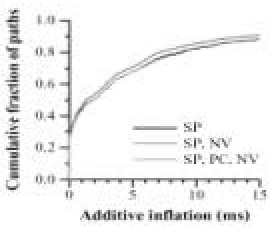
- They compare inflation caused by using early exit routing relative to an ideal optimal exit policy.

The top 5% of the paths suffers an inflation of more than 12 ms

Interdomain layer methodology

- Infer policies using Lixin Gao Heuristics
- Assume No-valley and Prefer customer routing model.
- Find inflation caused by shortest AS path routing, no valley and no valley + prefer customer.

Interdomain layer



SP: Shortest AS path routing
NV: valley free paths
PC: Prefer Customers

Shortest AS path can be much longer than shortest latency paths. Prefer customer and no-valley policies cause little inflation.

Cumulative Results

		Median	Mean	95%
Intra-domain	Topology	1.0 ms	2.4 ms	8.4 ms
	Policy	1.4 ms	3.2 ms	11.5 ms
Peering	Topology	2.0 ms	5.0 ms	17.7 ms
	Policy	3.0 ms	6.5 ms	24.5 ms
Inter-domain	Topology	3.0 ms	7.3 ms	34.1 ms
	Policy	6.9 ms	13.9 ms	60.3 ms

Figure 17: Cumulative path inflation caused by each of the six factors, computed with reference to a hypothetical direct link.

Conclusion

- Path Inflation is caused by BGP shortest AS path routing and by inefficient peering.
- It is not clear if policies contribute to path inflation.
- Propose an informed BGP that carries location of egress links.
- Shortest AS path routing alternative?

Discussion topics

- Alternative routing models
  - Game theory
  - Auction based routing
  - Multipath routing
- What can end-user do given restricted routing policies?
  - Overlay routing
- Security implications
  - How robust is internet routing

Some research project suggestions

- Analyze a new attack against routing protocols and devise a defense mechanism
  - Route flap damping attack
- Design router primitives to defend against DDoS, Worm, infrastructure attacks
  - Push back for DDoS
- How to exploit topology information to launch routing attacks
  - Variations of link-cutting attacks
- Attack detection
  - Exchange of information among ISPs
  - Signature, behavior based
  - Routing protocol analyzers (Bro)
- Intradomain topology design considerations
  - Route reflector vs. AS confederations or hybrid
  - Robustness, ease of configuration, security/trust

Projects ideas continued…

- Dynamic installation of route filters to protect against DDoS attacks