Controlling the Impact of BGP Policy Changes on IP Traffic

Nick Feamster
M.I.T. Laboratory for Computer Science
feamster@lcs.mit.edu

Jennifer Rexford
AT&T Labs--Research
jrex@research.att.com

Jay Borkenhagen
AT&T Labs
jayb@att.com

Summary

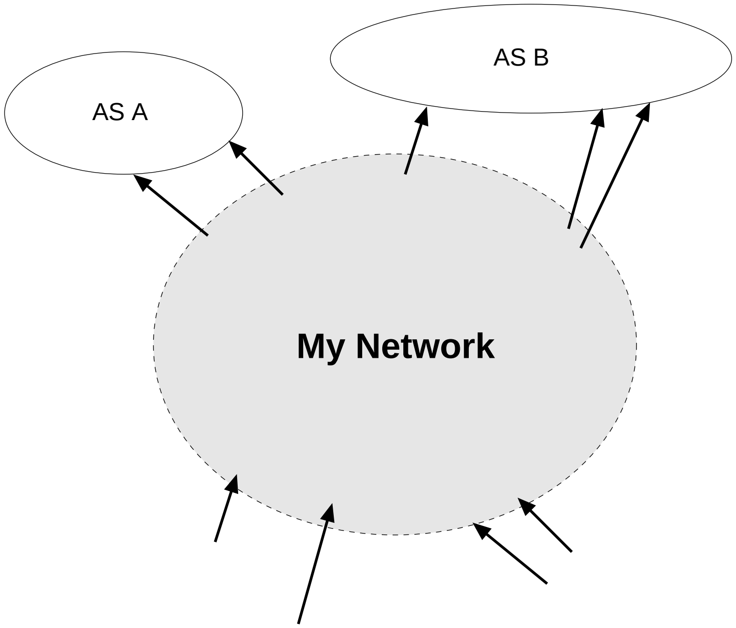
- BGP traffic engineering practices that:
  - Have good scaling properties
  - Result in predictable changes to traffic flows
  - Limit the influence of neighboring domains

- Tool for BGP traffic engineering
  - Model that describes the effect of BGP policies on traffic flows
  - Deterministic, network-wide algorithm to determine best routes

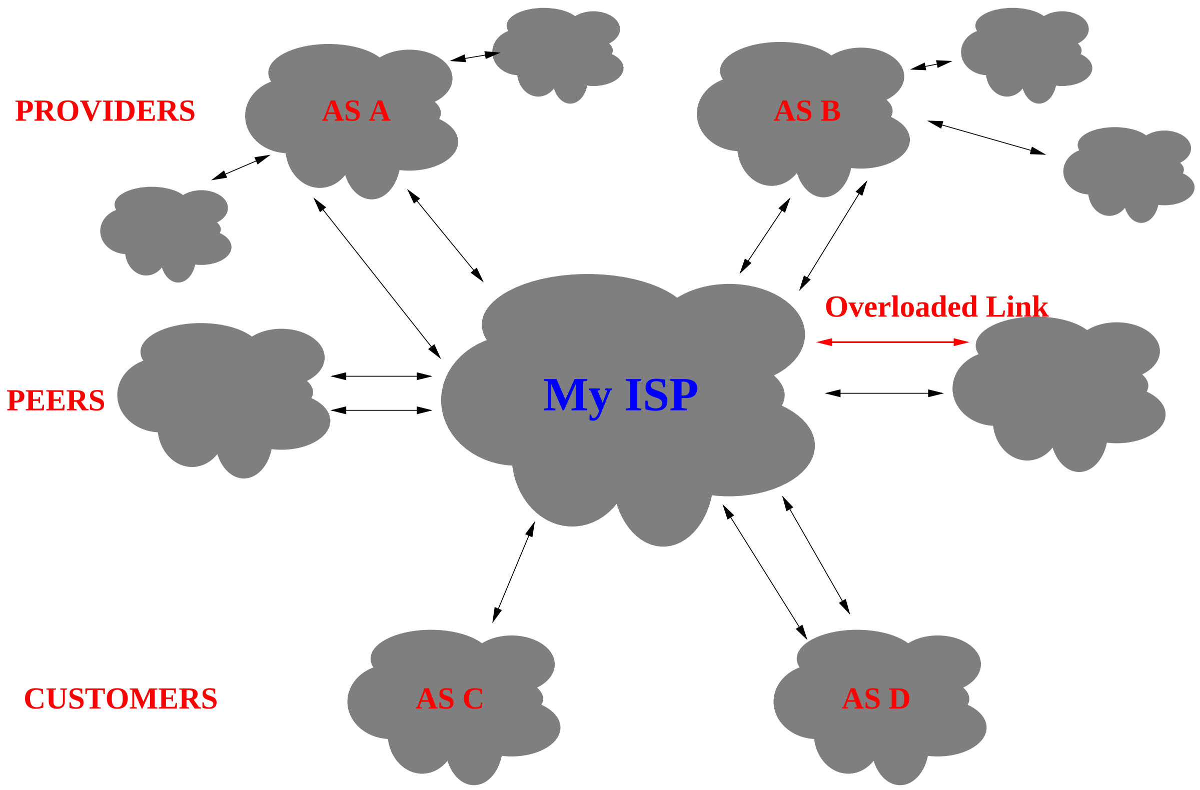
Interdomain Traffic Engineering

- Why?
  - Alleviating congestion on edge links
  - Adapting to provisioning changes (e.g., link capacity)
  - Achieving good end-to-end performance

- How?
  - Directing traffic to a different neighbor AS
  - Directing traffic to different links to the same neighbor



Many Breeds of Networks

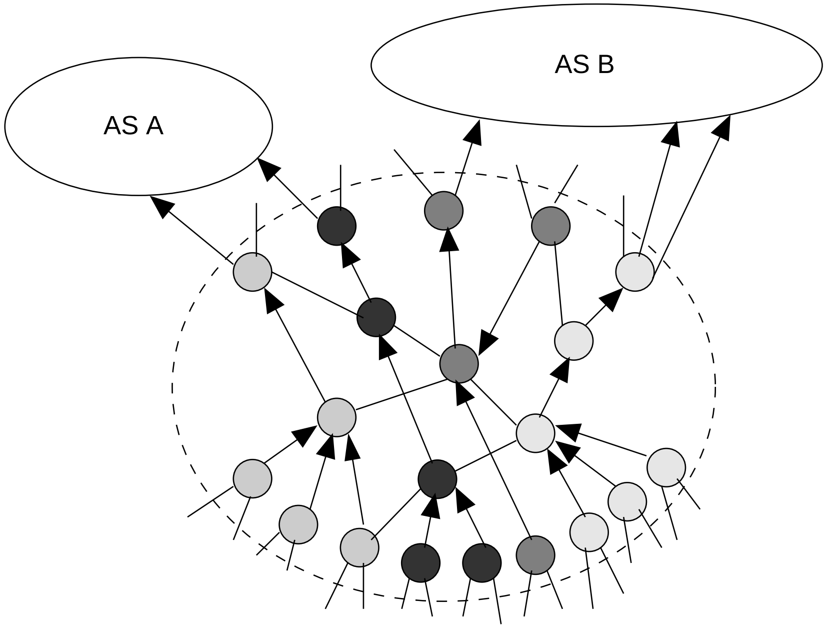


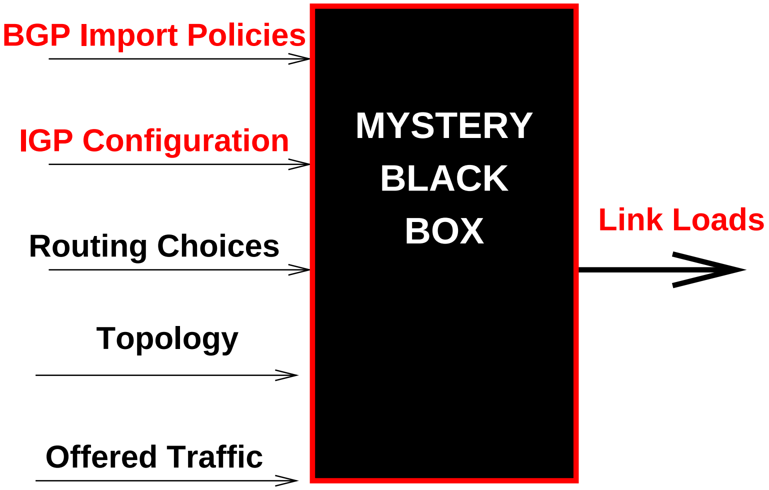
- Where should we offload traffic?
- We have to be careful about the impact of policy changes!

BGP Traffic Engineering Overview

- Change outbound traffic using BGP import policy.
- Why not scrap BGP and start over?
  - No flag days
  - Perhaps...ideas for improving BGP (?)
- "Good" choices? Adjustments should...
  - Impose minimal management and message overhead
  - Result in predictable changes in traffic volumes
  - Not affect neighboring AS’s routing decisions

Model: Effect of Import Policies on Traffic





BGP Import Policies

IGP Configuration

Routing Choices

Topology

Offered Traffic

MYSTERY BLACK BOX

Link Loads

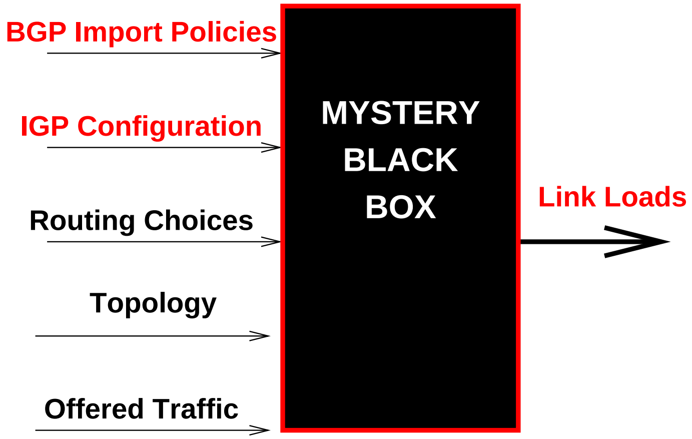
- Predict link loads when certain inputs are unstable?
  - Routing choices (e.g., neighbor’s BGP advertisements)
  - Inbound traffic

How can we adjust BGP import policies to affect outbound traffic and maintain stable/predictable inputs?

Traffic Engineering with BGP?!

- Protocol Difficulties
  - No performance metrics in advertisement attributes.
- Configuration Difficulties
  - Can’t express conjunction between attributes.
  - Indirect influence on route selection.
- Decision Process Difficulties
  - At most one best route per prefix per router.
    - Egress router cannot "split" traffic across multiple links to different neighbors.
    - Limits granularity at which we can shift traffic.
  - Can’t split traffic to a prefix over paths of different lengths.
  - Interaction with Interior Gateway Protocols (IGPs)
- Commercial relationship constraints

Guidelines: Playing with the Black Box



- Deterministic Output:
  - bgp deterministic-med
  - Disable tiebreaking based on age of advertisement (use router ID instead).
- Minimal Overhead:
  - Minimize the frequency of changes.
  - Enable soft reconfiguration or route refresh options.

What types of constraints should we impose on BGP policy changes?

Challenges

- *Scale:* 100k+ Prefixes, can’t set independent policy for every one!
  - Configuration overhead
  - Traffic instability
- *Predictability:* Policy-based adjustments are indirect
  - So many things can happen when a change is made!
  - Is there a way to tell what’s going to happen?
- *Control:* Neighbors’ behavior can affect traffic volumes in ways we can’t control.
  - What if our neighbors change the inbound traffic?
  - Neighbors announce "strange advertisements".

Data from AT&T's Commercial Backbone

- BGP Routing Tables
  - Received paths for each prefix at each peering point
  - Best guess at what future updates will look like
  - Aggregate traffic statistics by prefix
- Cisco Netflow data
  - Medium-grained traffic statistics
  - Used in conjunction with tables to:
    - Determine popular prefixes
    - Assess significance of events w/respect to traffic
- Router Configuration Files
  - Who our "peers" are
  - Which import policies apply to which eBGP sessions

We focus on outbound traffic over peering links;
examples are from March 1, 2002.

But I Don’t Have That Data!   :(

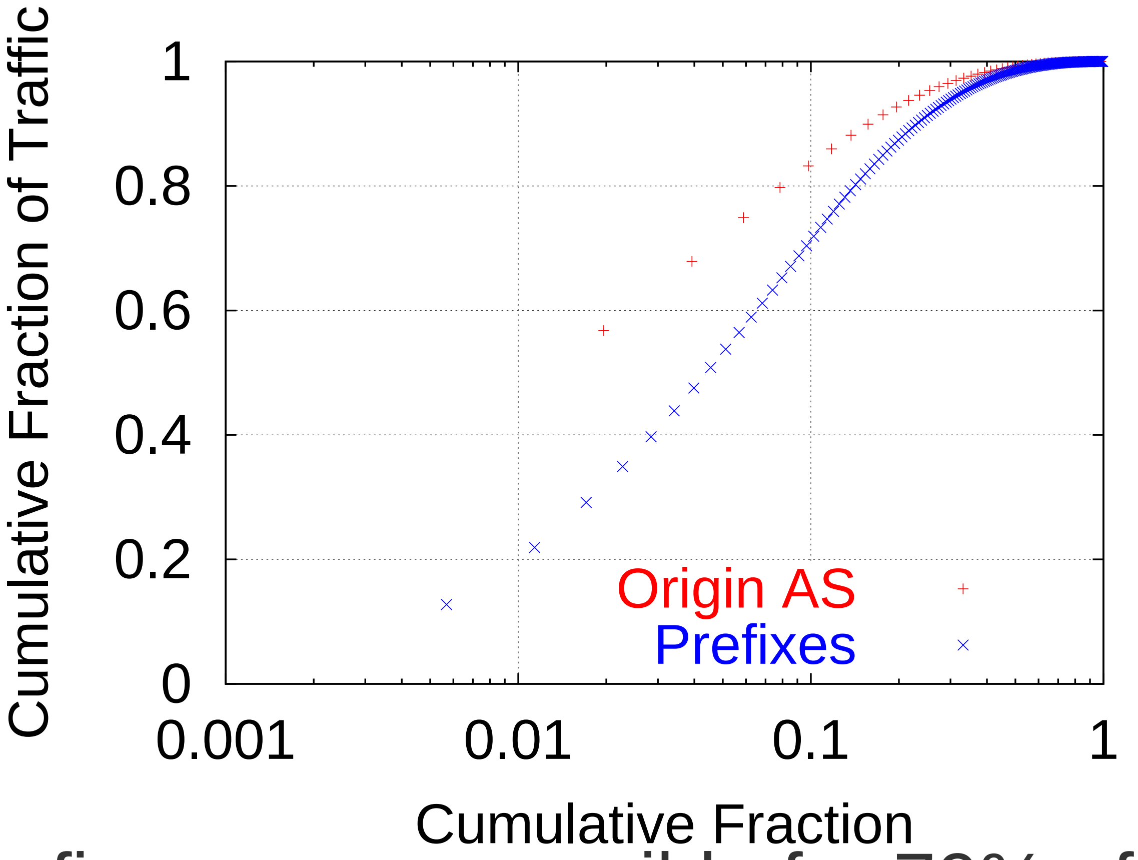
- BGP Advertisements
  - iBGP monitors can be used to determine at least the best routes
  - Juniper support for outputting a feed of all BGP routes
- Traffic Measurement
  - Netflow
  - Policy-based accounting
  - Packet sampling/monitoring
- Our analysis also applies with limited traffic data...

Managing Scale

- *Problem:* Large number of prefixes preclude setting import policy on every one.

- *Solution:* Change policies for the small fraction of groups of prefixes that are responsible for the majority of traffic.

Scale: Heed Traffic Characteristics



- 10% of prefixes are responsible for 70% of traffic
- Focus: small number of popular prefixes/origin AS's.
  - Per-prefix tweaking is tractable
  - Hopefully, more predictable offered loads...

Predictability: Changes in Inbound Traffic

- *Problem:* Inbound traffic volumes change over time.
- *Solution:* Change policies for the groups of prefixes that have more stable traffic volumes.

Which prefixes are those?

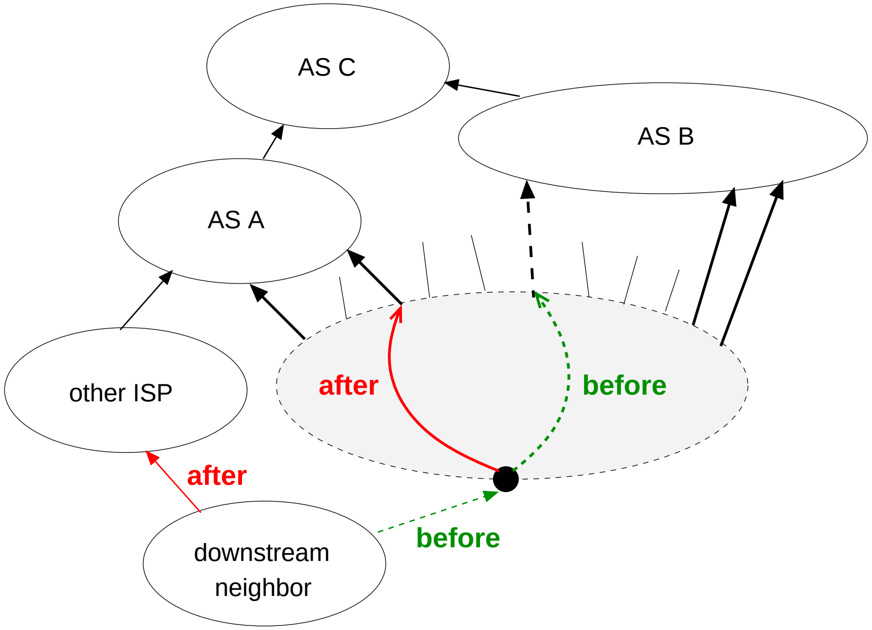
Predictability: Focus on Stable Prefixes

- Origin AS's responsible for top 1% of outbound traffic in one week experienced a 10% change in traffic over a one-week period.
- Most origin AS's that are responsible for more than 10% of outbound traffic do not change by more than a factor of 2 from week-to-week.

Networks that terminate more traffic are more likely to have stable offered load from week-to-week.

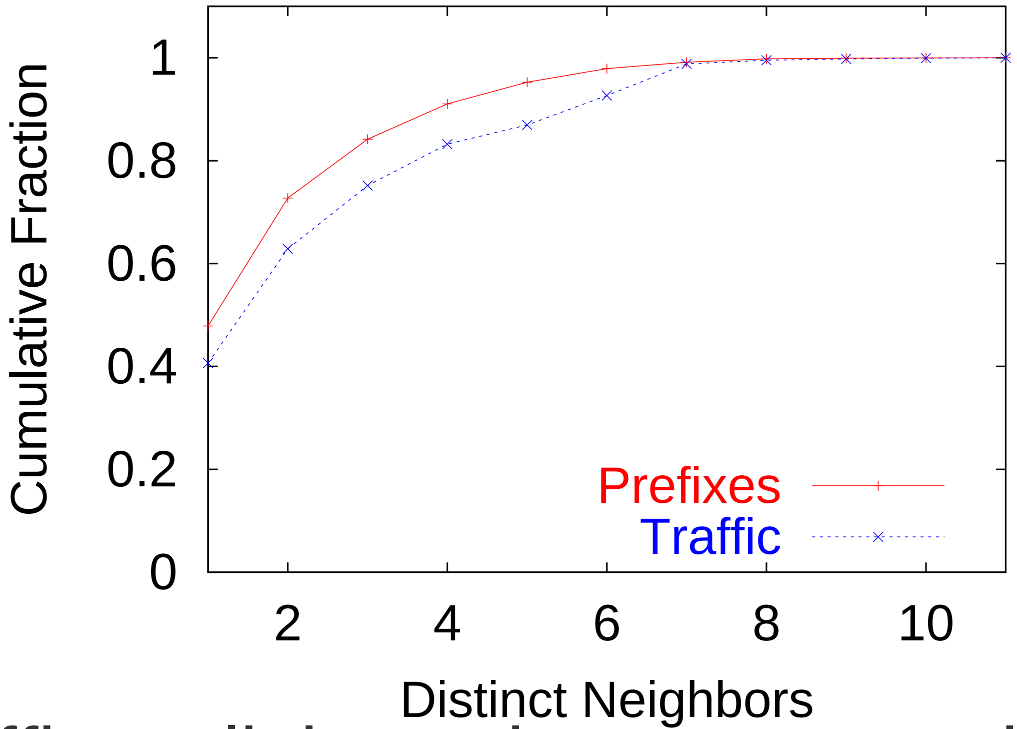
Predictability: Big Changes, Fickle Neighbors

- *Problem:* Internal changes that are externally visible can change inbound traffic volumes.



- *Solution:* Shift traffic among paths
  - to the same AS
  - to different AS, but with the same path length

Predictability: Shift to the Same AS



- Shifting traffic on links to the same peer keeps inbound traffic more predictable.
- ~70% of outbound traffic to peers has shortest-path advertisements for only one next hop AS

Predictability: Advertisement Changes

- *Problem:* Want to shift traffic aggregates
  - On a finer granularity than per AS
  - On a more coarse granularity than per path

    ...and remain resilient to changes in neighbor’s advertisements

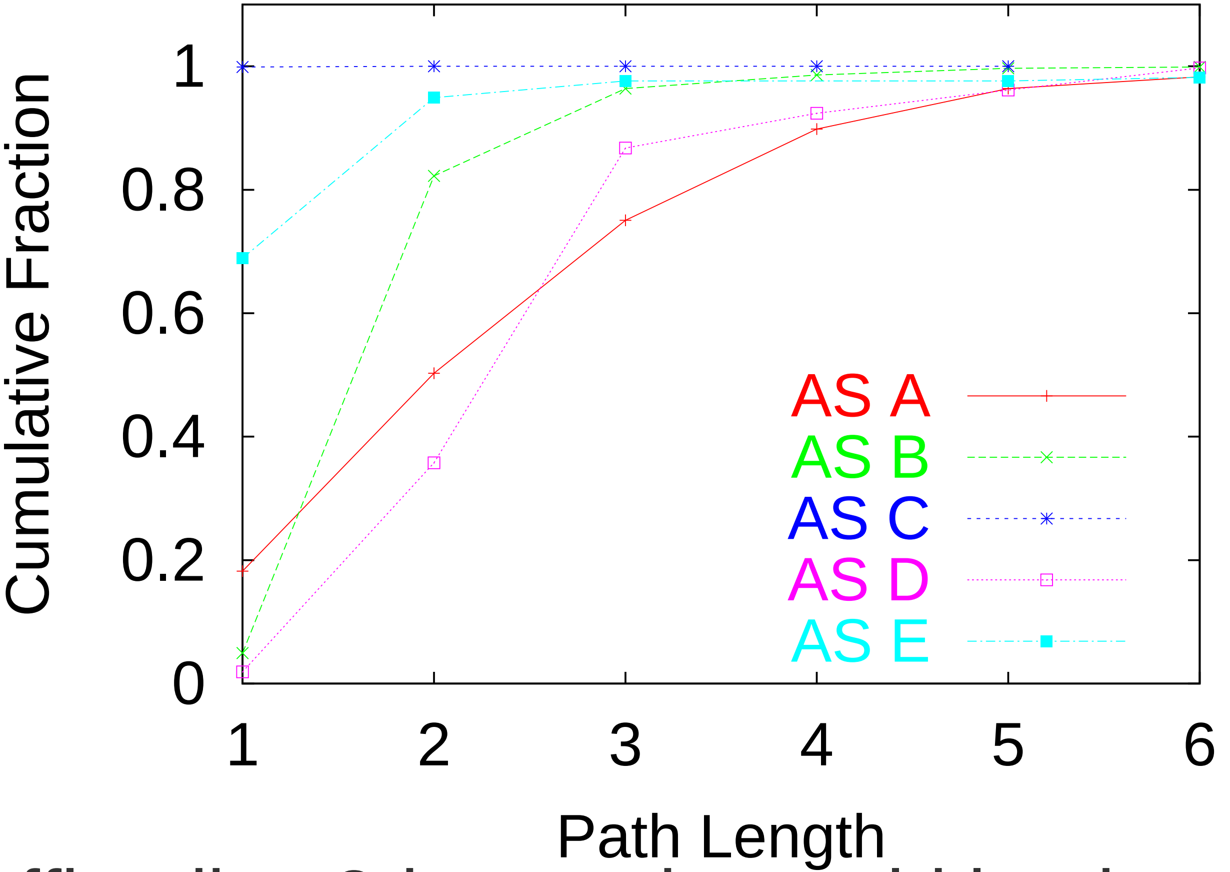
- *Solution:* Assign policies using regular expressions.

```
ip as-path access-list 1 permit ^701$
ip as-path access-list 1 permit ^701_[0-9]+_$

route-map IMPORT permit 5
 match as-path 1
 set local-preference 100
!
route-map IMPORT permit 10
 set local-preference 105
!
```

But be careful...

Predictability: AS’s are Not Created Equal



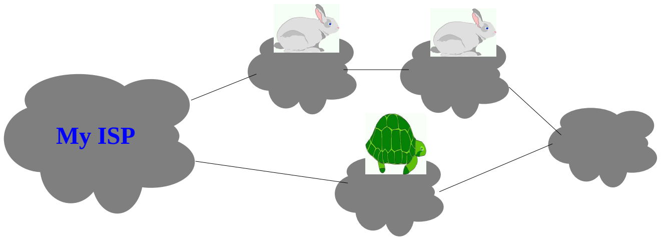
- Blindly offloading 2-hop paths could lead to trouble!
- Pay attention to the type of AS when making policy changes.

Control: Why AS Path Length Doesn’t Fit In

- *Problem:* AS path length comes early in the decision process, is controlled by neighbors, and doesn’t often reflect a short path.

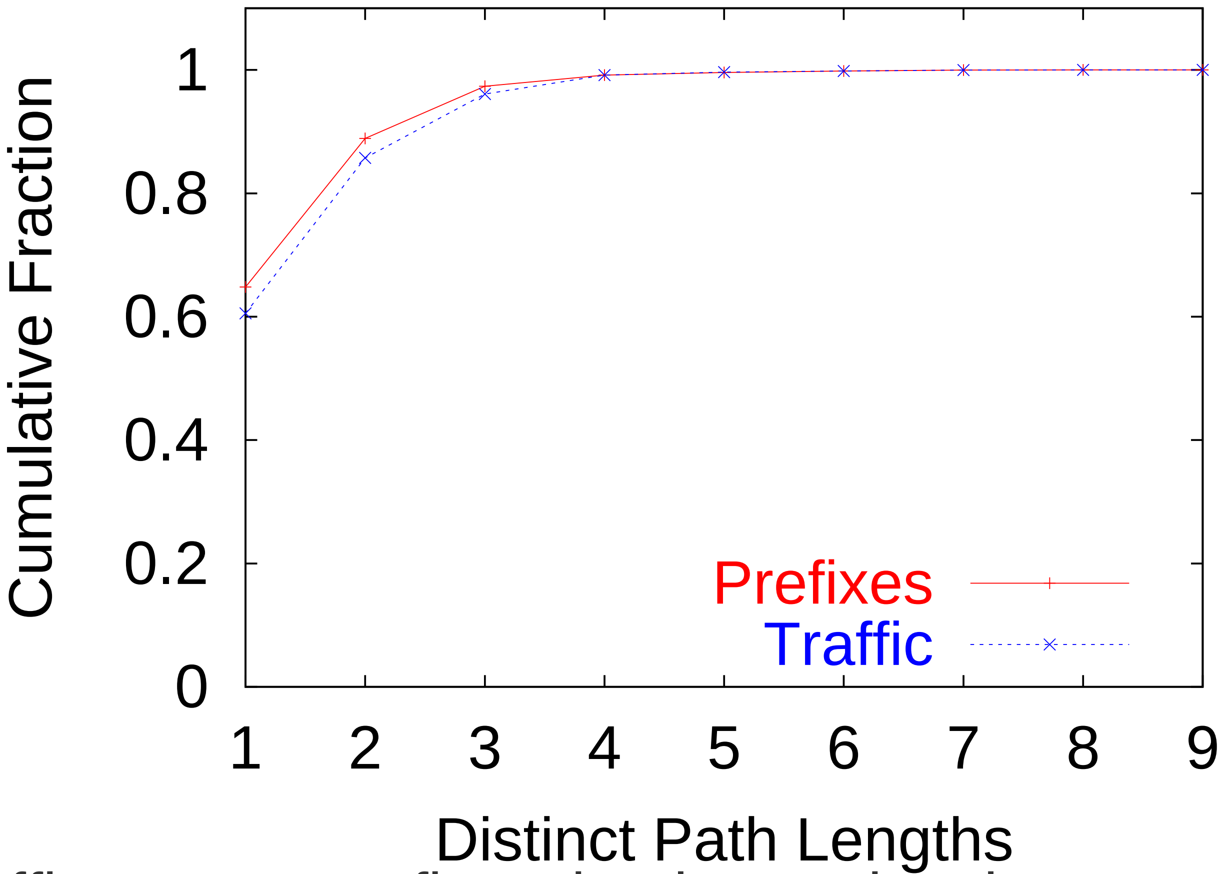
Step 1:	Highest Localpref	Operator-Controlled
Step 2:	AS Path Length	*Neighbor-Controlled*
Step 3:	Origin Type	Operator-Controlled
Step 4:	Lowest MED	Operator-Controlled

...



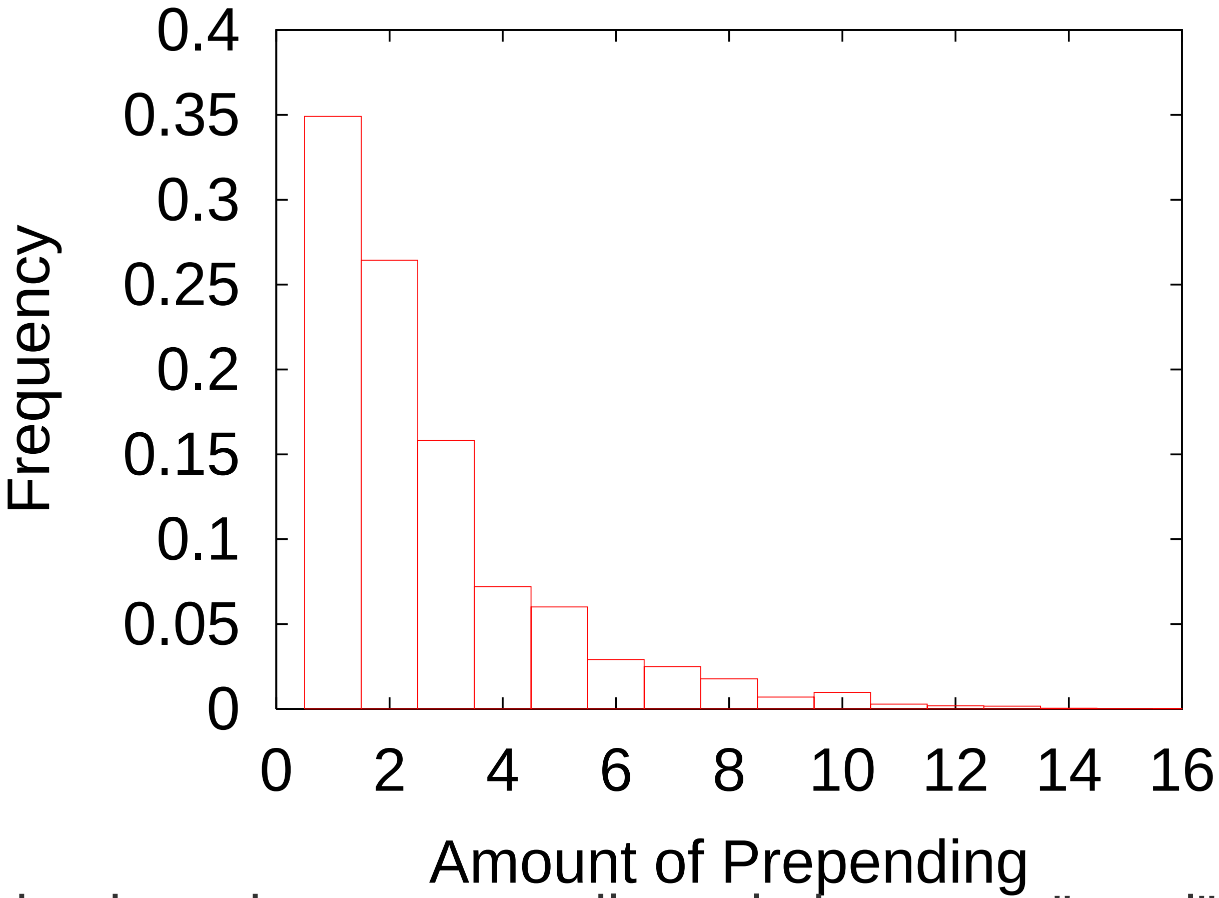
- *Solution:* Assign coarse-grained localpref based on path length, rather than using path length metric.

Control: AS Paths Can Be Deceiving



- >35% of traffic goes to prefixes that hear advertisements of more than one distinct length.
  - Prepending often used to indicate a backup route.
  - Many backup links could be used to offload traffic, but AS path length metric limits this possibility.

Control: Prepending Limits Choices



- Coarse-grained metric unnecessarily excludes some "good" routes.
- Difference between 7 and 8 prepends?

Solution:

Ignore AS path length as an absolute metric.
Use it as an attribute to assign localpref!

Control: Eternal Vigilance

- *Problem:* Neighbors can play the following games that limit a network's ability to do traffic engineering:
  - Filtering on some peering points but not others.
  - Advertising different paths to different peering points.
    - Different path lengths.
    - Same path lengths, different paths.
  - Advertising next-hop different from BGP session IP address.

- *Solution:* Pay attention. :)

These don't happen that often in the AT&T network,
but they're good to watch out for...

Conclusions

- BGP not designed for TE, but it is here to stay!
  - Language is indirect and inflexible
  - Restrictive decision process
  - Limited control, many interactions with neighbors
- We can have BGP traffic engineering practices that
  - Have good scaling properties
  - Result in predictable changes to traffic flows
  - Control the influence of neighboring domains
  - Operate within the existing BGP infrastructure
- A tool for network-wide routing prediction
  - Model that describes the effect of BGP on traffic flows
  - Algorithm to determine best routes, without simulating BGP messag passing

http://nms.lcs.mit.edu/~feamster/paper-nanog25.pdf

Shameless Plea for Network Presence

- Resilient Overlay Networks (RON) Project
  - http://nms.lcs.mit.edu/projects/ron/
  - 15 active nodes
- Research Questions
  - How are BGP announcements and end-to-end path failures correlated?
  - What are fate-sharing relationships between prefixes? (looking at prefixes that are announced/withdrawn together)
  - Where along the path are failures occurring, and why?
- We need network presence
  - iBGP Monitor
  - Place to send active probes (low-traffic)
  - Thanks to Randy Bush for volunteering!

feamster@lcs.mit.edu