Using BGP communities to control your transit providers

Martin J. Levy

Hurricane Electric

APRICOT 2013 Singapore

Peering Forum Part 2

February 2013

Introduction

- Goals

  – To show how BGP global routing operates today

  – To show how BGP global routing could improve

- Explaining BGP communities

  – To help your network operate better

BUT FIRST; USEFUL INFORMATION

Traffic Engineering – Andy Davidson vs. Martin Levy

- Andy Davidson's talk

  – Very accurate; but missed some vital information

- Martin Levy's talk

  – Includes the most important information …

The secret menu at In-n-Out burger

Not-So-Secret Menu

Double Meat
two pure beef patties hand-leafed lettuce, tomato, spread with or without onions, stacked high on a freshly baked bun

4 x 4 ®
four 100% pure beef patties, hand-leafed lettuce, tomato, spread, four slices of American cheese, with or without onions, stacked high on a freshly baked bun

Protein® Style
your favorite burger wrapped in hand-leafed lettuce instead of a bun

3 x 3 ®
three 100% pure beef patties, hand-leafed lettuce, tomato, spread, three slices of American cheese, with or without onions, stacked high on a freshly baked bun

Grilled Cheese
two slices of melted American cheese, hand-leafed lettuce, tomato, spread with or without onions on a freshly baked bun

Animal® Style
burger of your choice with hand-leafed lettuce, tomato, a mustard cooked beef patty; add pickle, extra spread with grilled onions
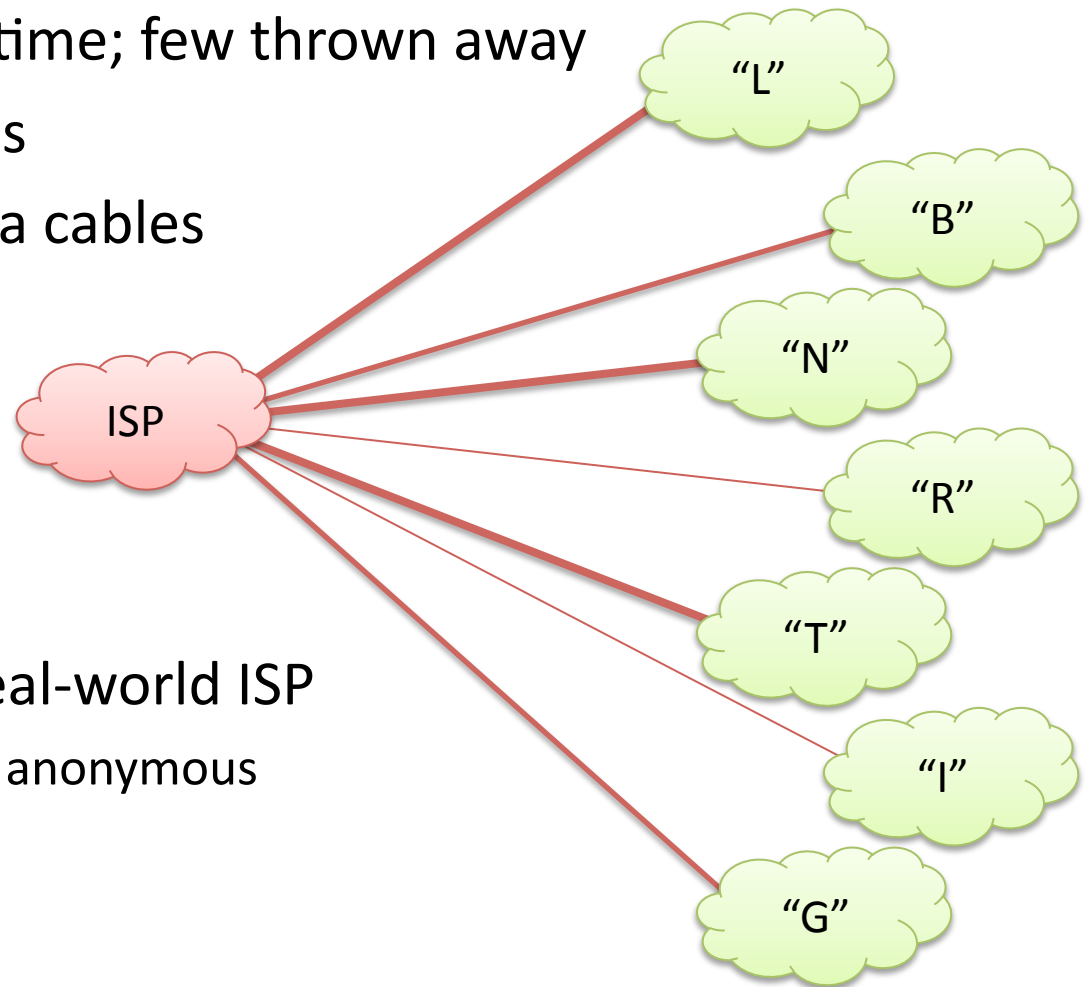
The best option in my opinion

http://www.in-n-out.com/menu/not-so-secret-menu.aspx

A REAL WORLD EXAMPLE

Our sample (real-world) ISP example

- IP transits added over time; few thrown away
- Mixed bandwidth levels
- Mainly tied to undersea cables

- This is all based on a real-world ISP
  - All ASNs are somewhat anonymous

Situation today

- Nothing is wrong today
  - ISP interconnects via many upstream networks
  - Full connectivity is provided to all of ISPs customers

- But; things could be better
  - Cleaner routing makes for easier operations
  - Cleaner routing makes for a better customer experience

- One step at a time
  - Routing improvements will require bandwidth balancing

Basic rules – the ones that work

- Methodology's that are taught:

  - BGP prepends for traffic engineering
  - Route de-aggregation for traffic engineering
  - Announcing a subset of routes to an upstream neighbors for traffic engineering

- Methodology's that don't actually work:
  - BGP prepends for traffic engineering
  - Route de-aggregation for traffic engineering
  - Announcing a subset of routes to an upstream neighbors for traffic engineering

- The new world (not so new world) order:
  - Heavy use of BGP community signaling
  - Careful control of depth of route announcements
  - Announce a full set of your routes to ALL upstream neighbors

- Is this hard to learn?
  - No!

THE MANY UPSTREAM PROVIDERS

Visualizing routing – the ISP's upstream #routes

- Presently a different list of routes is sent to each upstream
  - No single upstream gets all ~1,700 routes
  - Different aggregated routes to different upstreams
  - Very few routes (read: customers) are actually multihomed
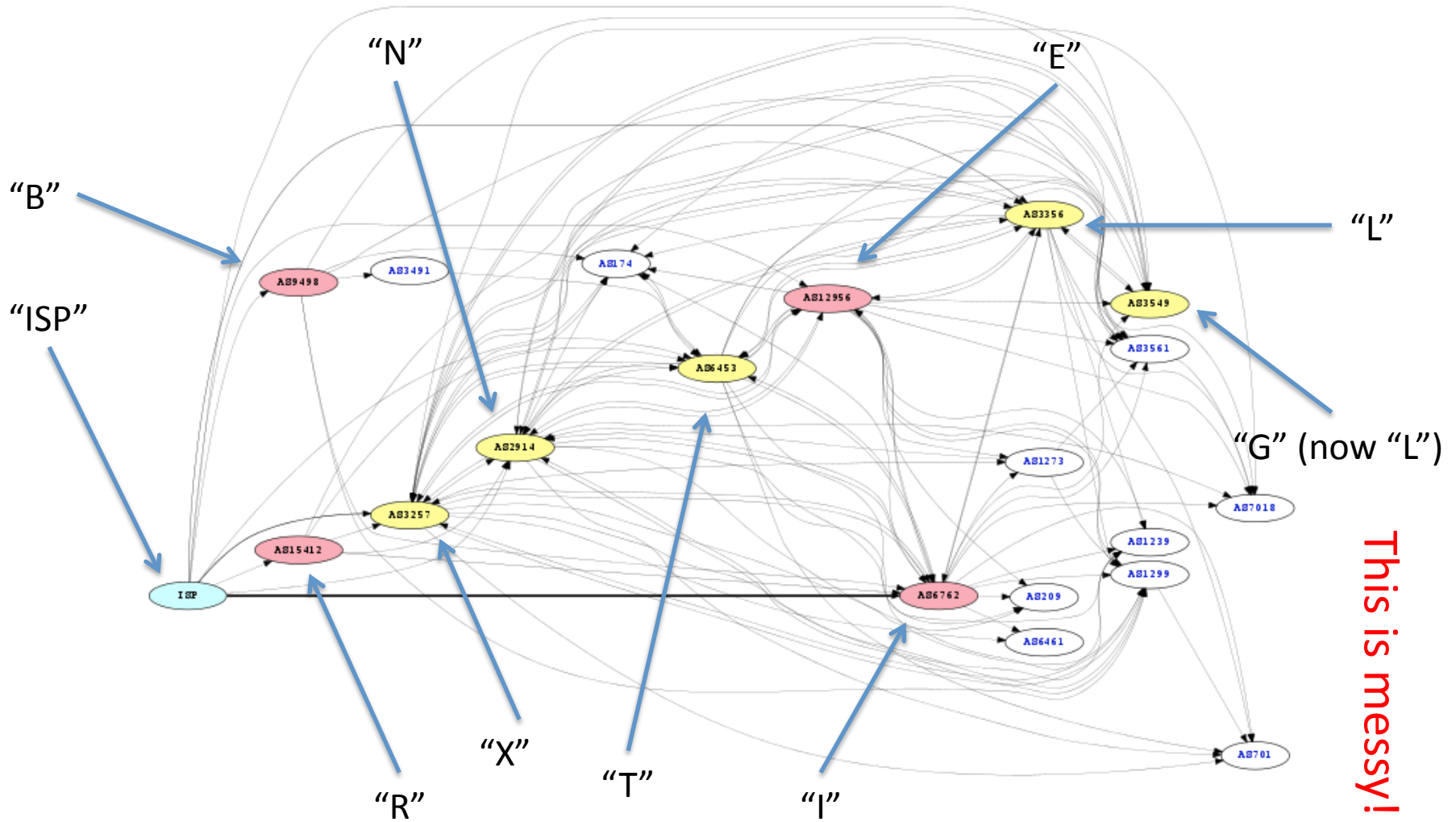 - Very few routes (read: customers) get efficient routing

- See next page

Real world BGP routes

| ASN	Count	Name
AS2***	192	"N"
AS3***	1,080	"X"
AS3***	440	"L"
AS3***	265	"G" (now "L")
AS6***	1	"T"
AS6***	484	"I"
AS9***	8	"B"
AS1****	42	"E"
AS1****	28	"R"

Visualizing routing – the ISP's upstream #routes

AS2*** "N"	AS*** "X"	AS3**** "L"	AS3*** "G"	AS6*** "T"	AS6*** "T"	AS9*** "B"	AS1**** "E"	AS1**** "R"	AS-PATH PREPEND (NETWORK SECTION)	AS-PATH (CUSTOMER SECTION)
38	280	115	96	1	91	8		27	65000	
					3				65000 65000	
					3				65000 65000 65000	
		10			17		1		65000 65000 65000 65000	
		2							65000 65000 65000 65000	
1									65000 65000 65000 65000 65000 65000 65000 65000 65000	
					1				65000	65099
	1	1							65000 65000 65000 65000	65099
	1	1							65000 65000 65000 65000	65099 65099 65099 65099 65099 65099

Is this really multi-homed?

High level TE (traffic engineering) by deaggregation of routes and/or filtering announcements

Various levels of AS path prepend

AS path prepend by customer

VISUALIZING THE ROUTING

Visualizing routing – The ISP via all upstreams

"N"

"E"

"B"

"L"

"ISP"

"G" (now "L")

This is messy!

"X"

"T"

"I"

"R"

Reading the visualized routing graphs

- Blue lines show routes being propagated
  - This is good
- Red lines show routes being propagated backwards
  - This is bad; it shows a route that could be more direct

- The many lines are caused by de-aggregation of routes and varied announcements (and prepends)

- A cleaner diagram is better (with networks connected closer)

- Colored network ovals are considered major networks

- This ISP doesn't have an unusual diagram

Diagrams from http://bgp.he.net/

Visualizing routing – The ISP via "I"

"ISP"

"I"
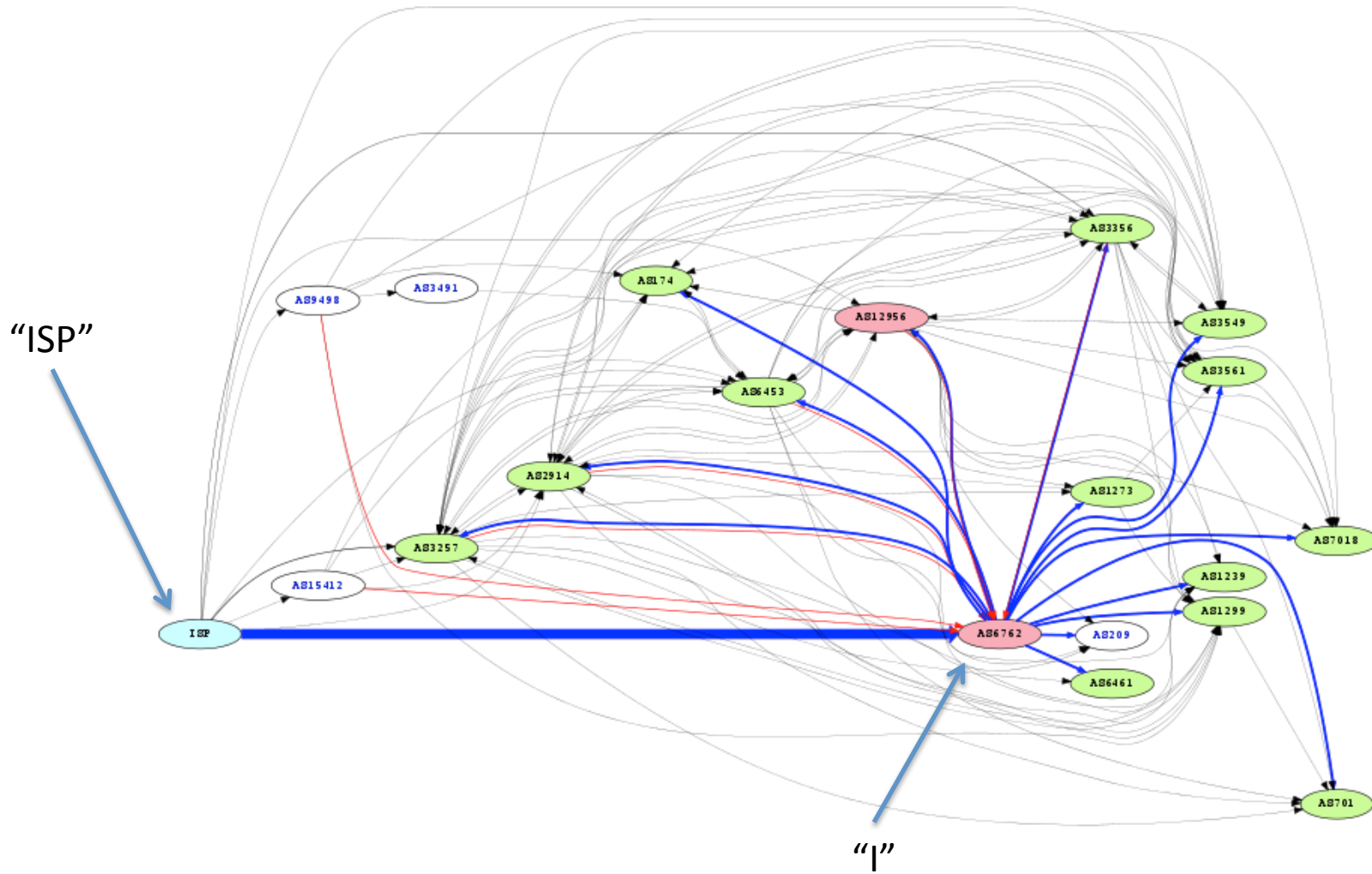
Visualizing routing – The ISP via "X"
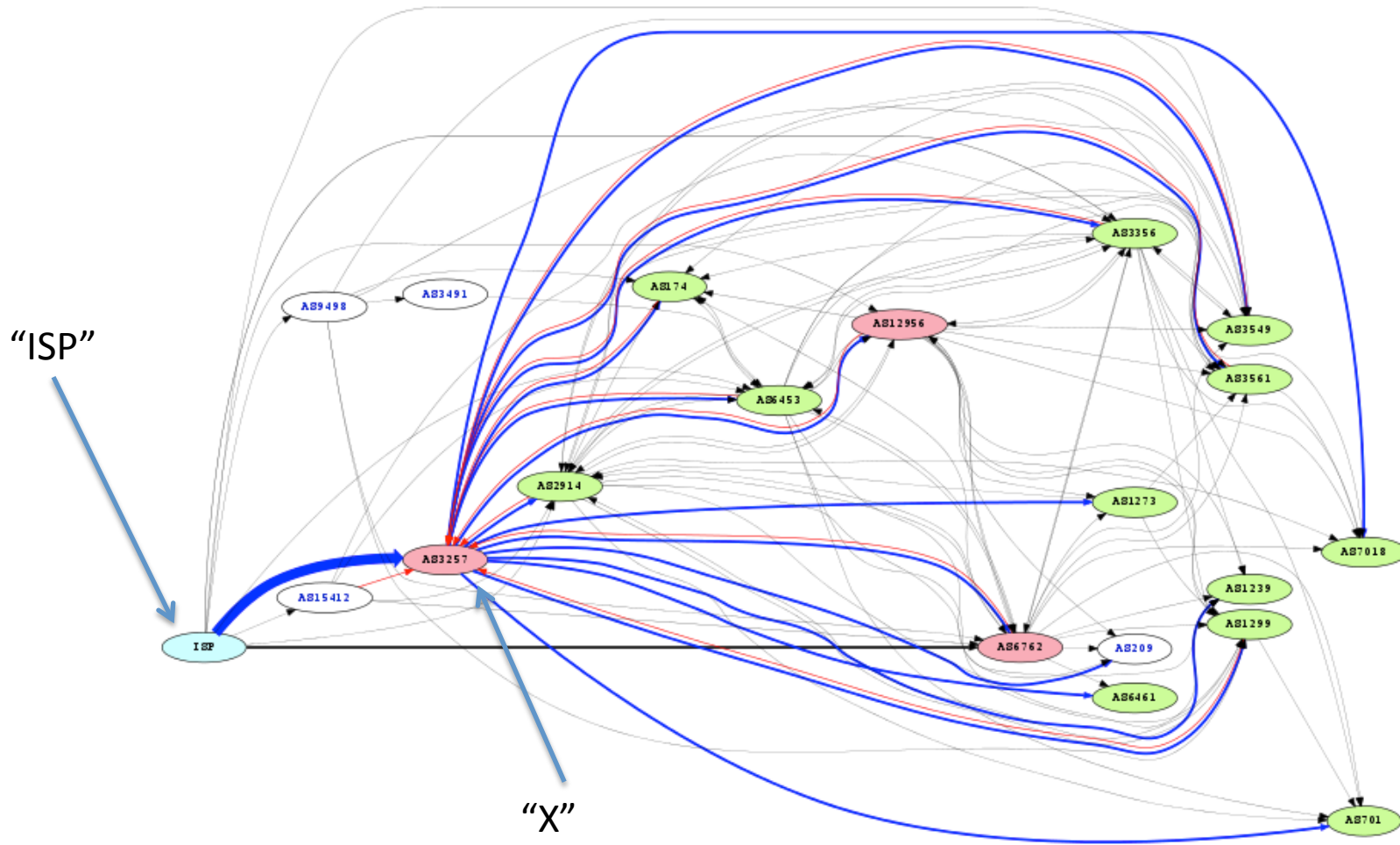
Visualizing routing – The ISP via "R"

Visualizing routing – The ISP via "E"

"E"

"ISP"

Visualizing routing – The ISP via "N"

"N"

"ISP"
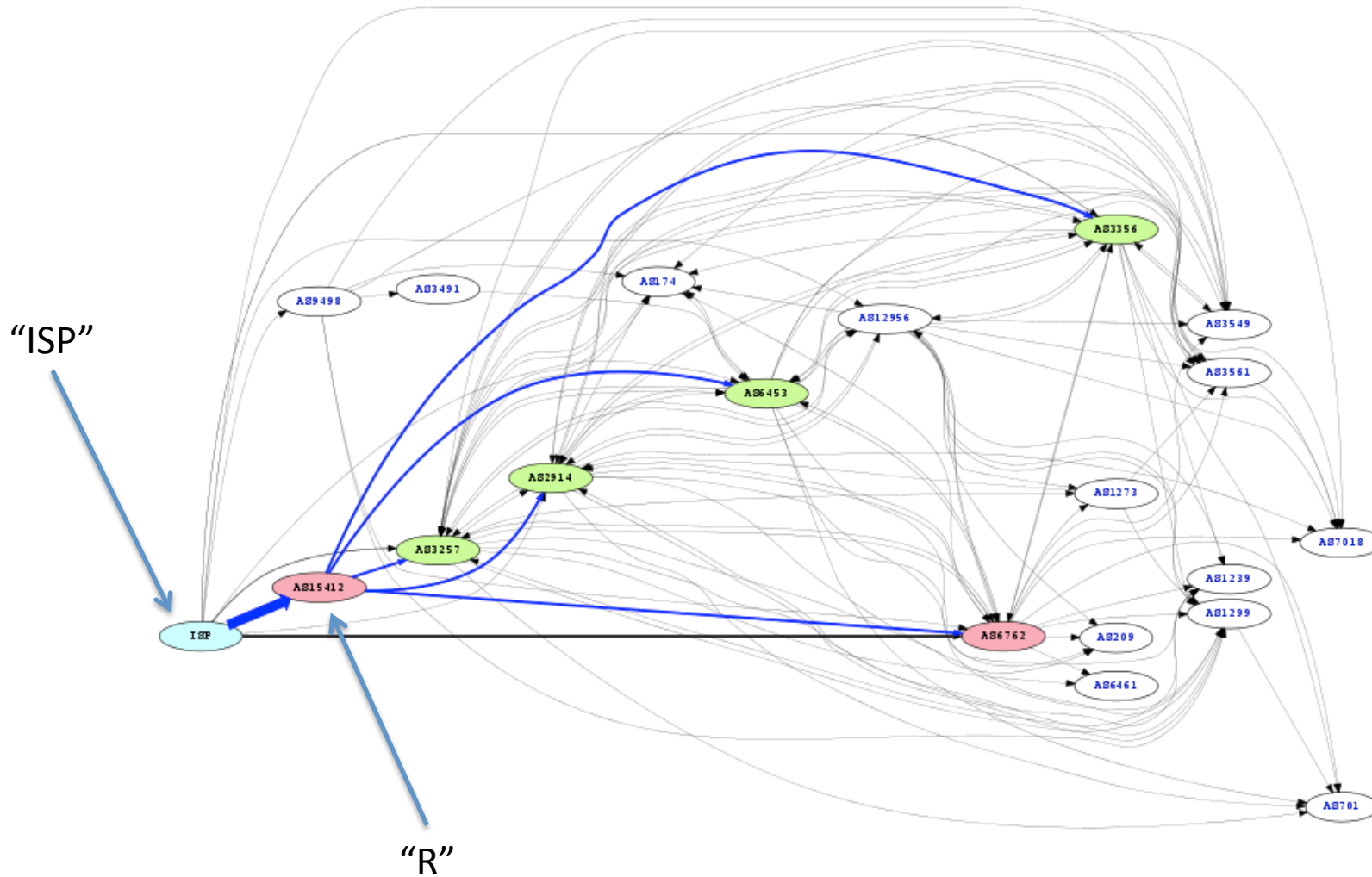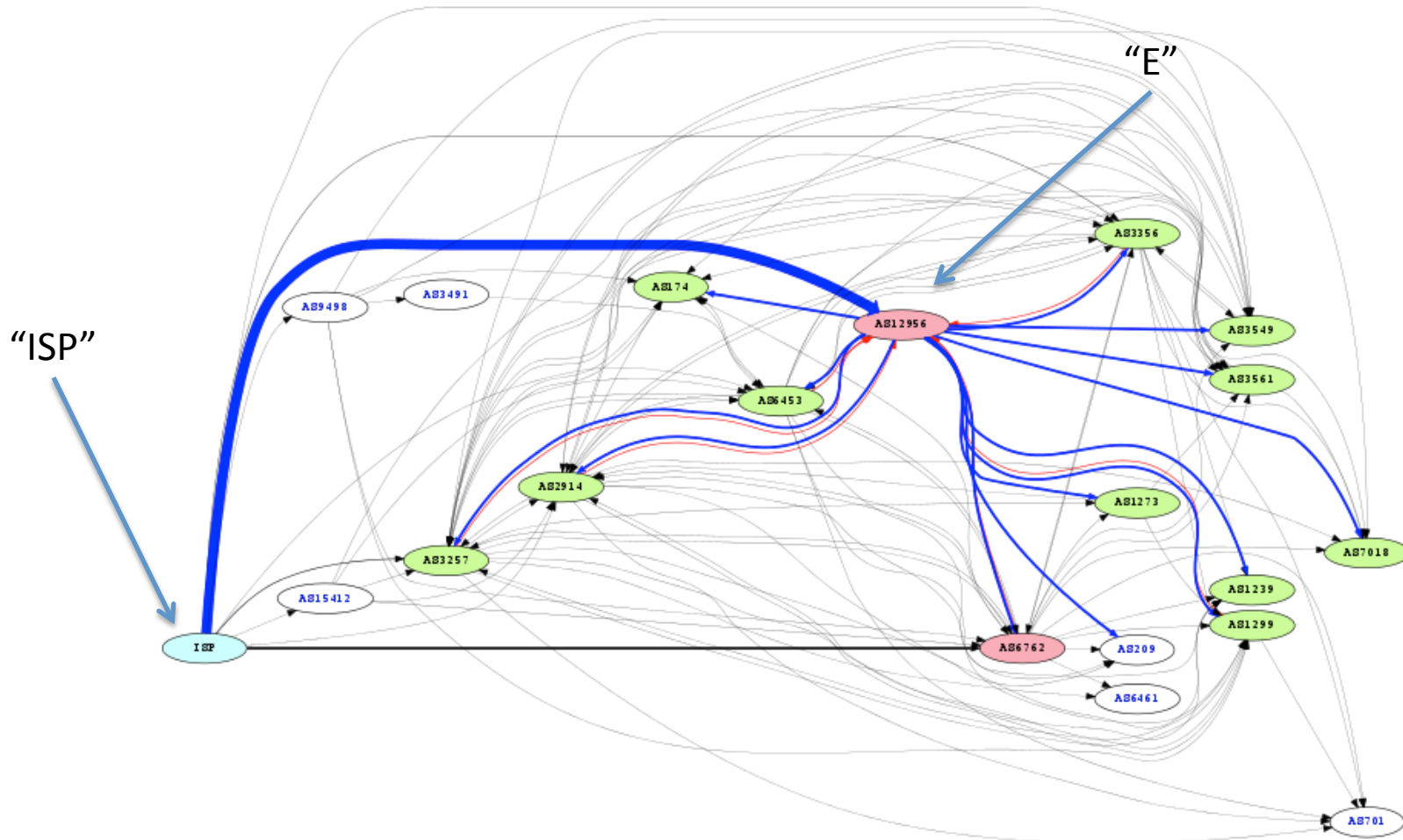
Visualizing routing – The ISP via "L"

"L"

"ISP"
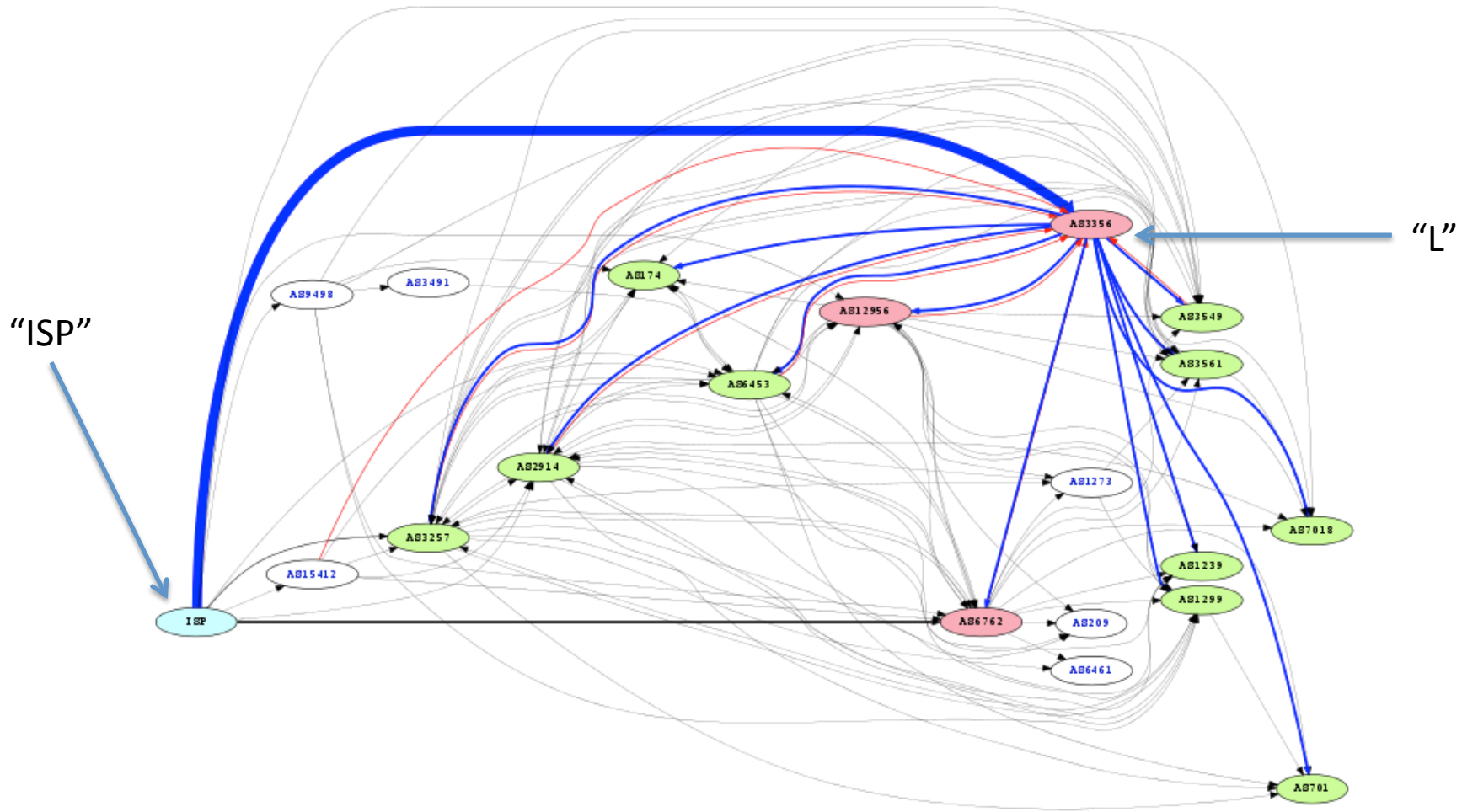
Visualizing routing – The ISP via "B"

Visualizing routing – The ISP via "T"
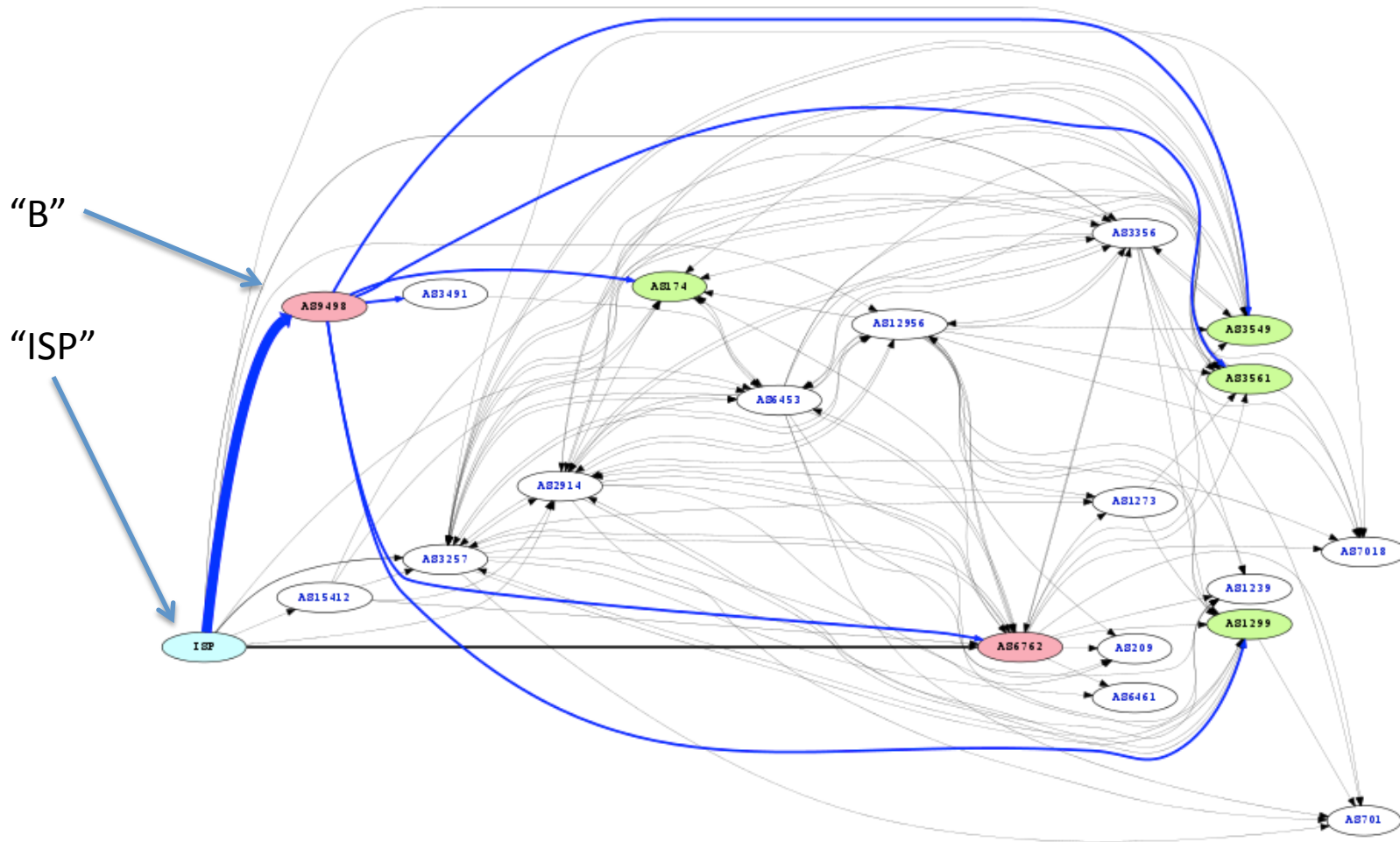
Visualizing routing – The ISP via "G"

"ISP"

"G" (now "L")
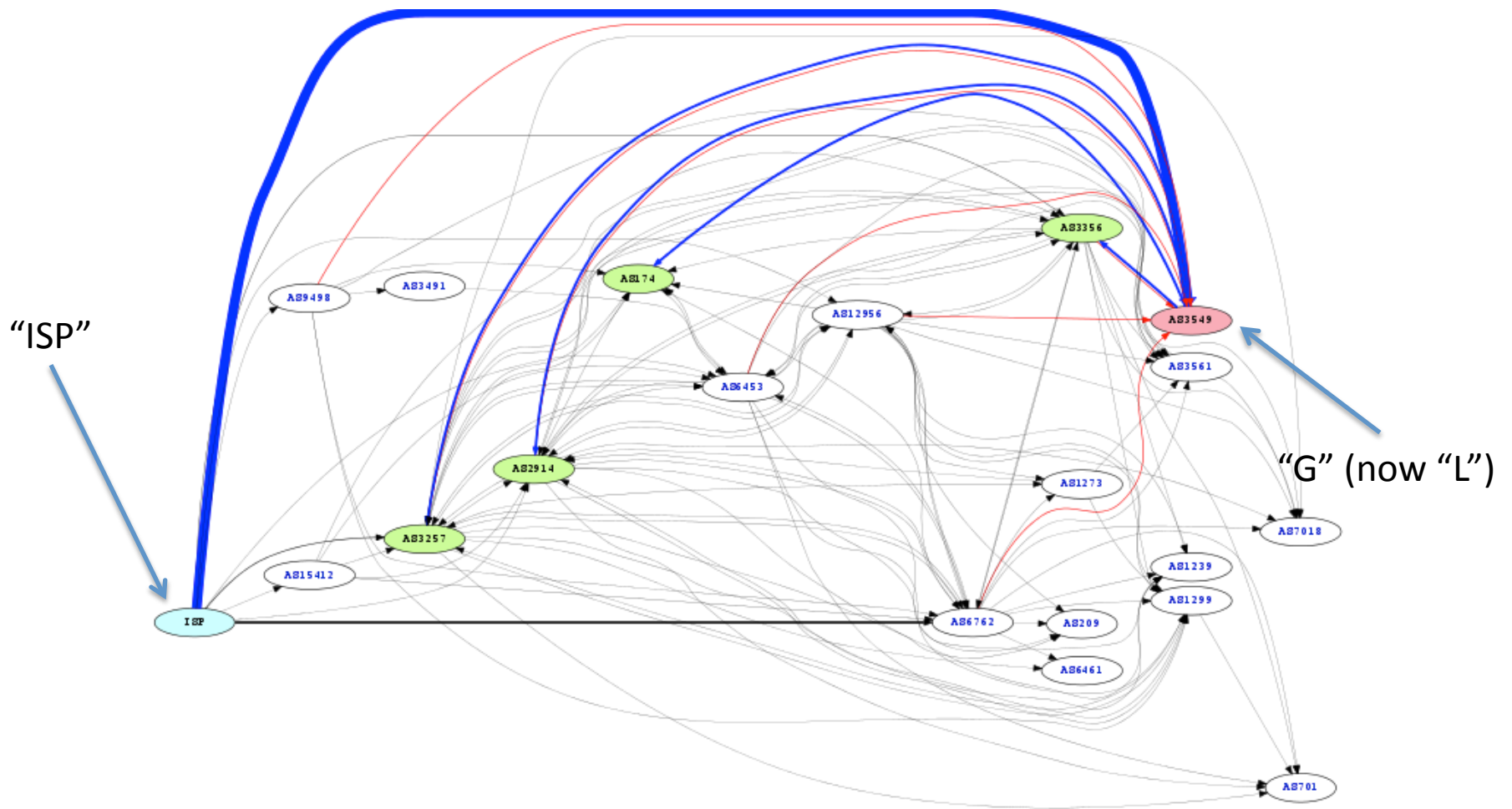
HOW TO IMPROVE GLOBAL ROUTING USING COMMUNITIES

Sending all routes to all upstreams

- In theory; all routes would be sent to all upstreams
  - This could cause excessive bandwidth over limited pipes

- Controlling the transit level of a route could help
  - A route not re-advertized would provide better routing

- BGP communities help with this process
  - A community can be set on a route to control depth

- This is a well documented process; but seldom used
  - It's NEVER in the interest of the transit provider to explain this
  - It works very well
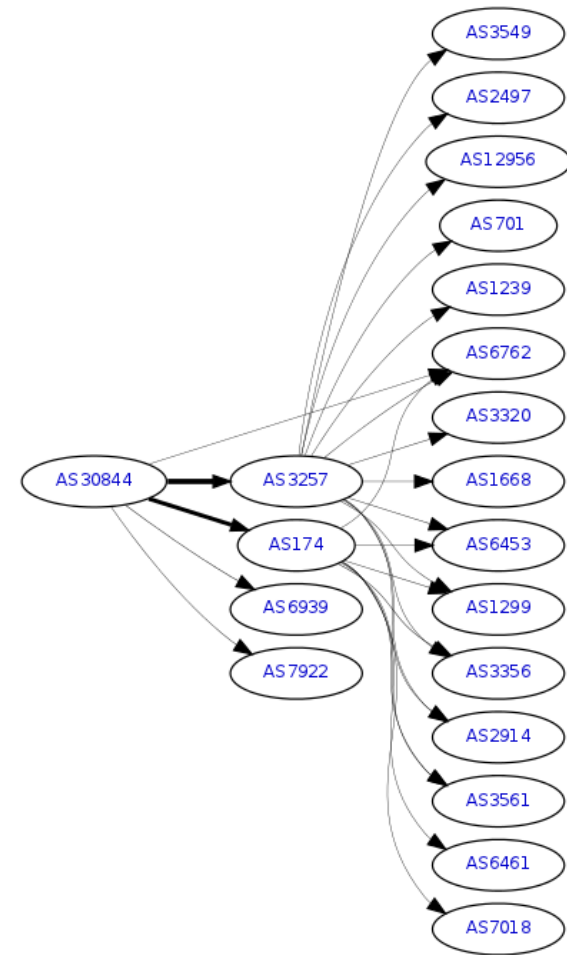 - It potentially reduces the upstream Renesys ratings

Visualizing perfect (or near perfect) routing
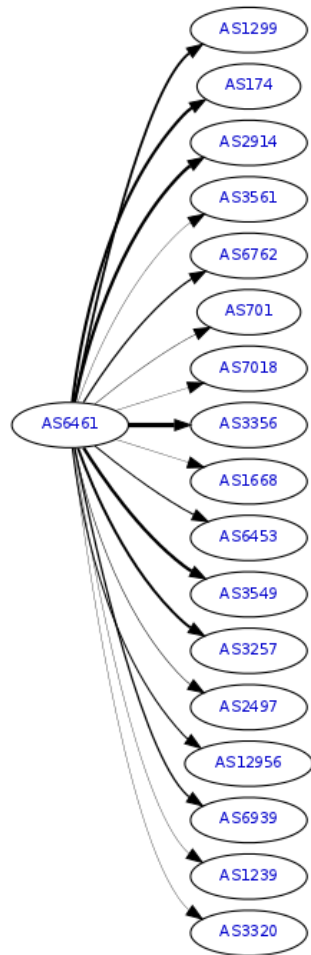
Perfect "Tier 1" routing
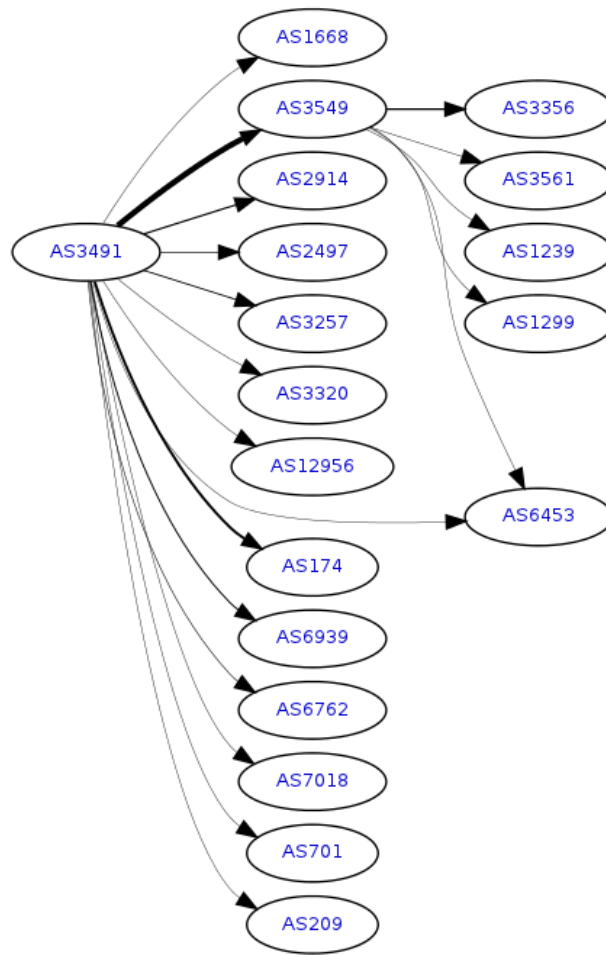
Pretty good routing (with many peers; one transit)

Classic backbone with two transits (well balanced)

BGP COMMUNITIES PRACTICAL EXAMPLES

BGP communities – and how they help

- BGP backbones publish their information (One Step has collected many of them)
  - http://www.onesc.net/communities/
  - etc …

- Specific IP backbone information can be found via specific URLs
  - Telecom Italia AS6962 @ http://etabeta.noc.seabone.net/communities.html
  - Level3 AS3356 @ http://www.onesc.net/communities/as3356/
  - Tinet AS3257 @ http://www.as3257.net/communities.txt
  - NTT/Verio AS2914 @ http://www.us.ntt.net/support/policy/routing.cfm
  - Flag/Reliance AS15412 @ https://apps.db.ripe.net/whois/lookup/ripe/aut-num/AS15412.html
  - etc …

- Setting a community on a route is very powerful
  - Can be used to control re-announcement
  - Can be used to control geography
  - Can be used to control style of route (peering/transit)

- Communities can be set on a per-route basis
  - Hence an upstream could receive all routes; but only re-announce some routes

BGP communities – an example

- Fixing Level3 routing
  - Announce all routes; set some routes to peering

- The goal
  - Bring all Level3 connectivity closer
  - Don't overrun bandwidth on Level3 connection
  - Don't produce a black holed route

- Customer experience?
  - Vastly improved!

- Relationship with upstream? (Level3)
  - Unchanged; but they now know you're smarter at operating your routing!

- Complexity level?
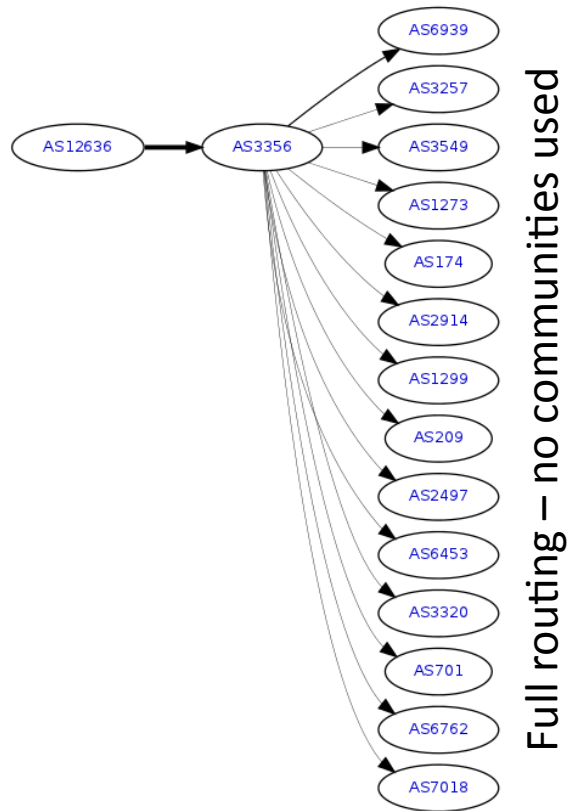 - Low; once understood

BGP communities – an example

- Example
  - Three routes are setup as peering
  - Two routes as peering + AS3320 DTAG
  - Remainder - full transit (the default)

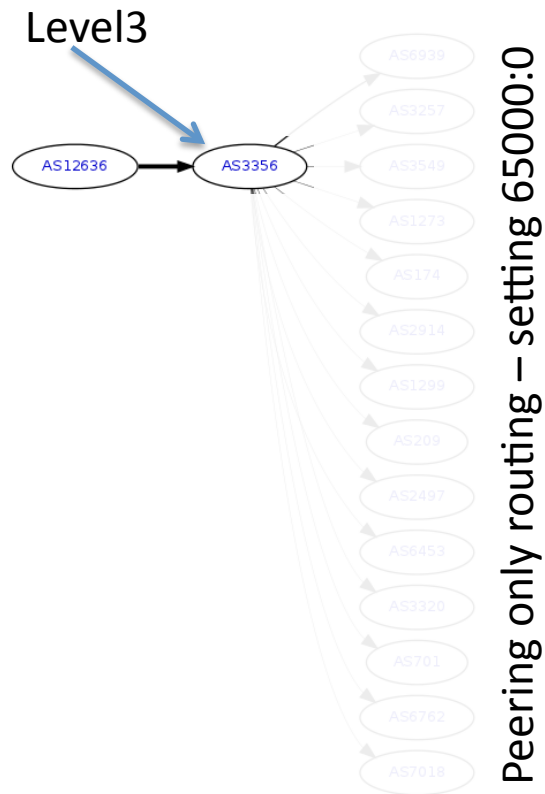- access-list methodology
  - Could be as-path filter

- Communities can be expanded

```
!
ip classless
ip bgp-community new-format
!
access-list 100 permit ip 10.7.0.0/12
access-list 100 permit ip 10.9.0.0/13
access-list 100 permit ip 10.22.0.0/17
!
access-list 110 permit ip 172.16.0.0/19
access-list 110 permit ip 172.17.0.0/19
!
route-map isp-out-as3356 permit 10
  match ip-address 100
  set community 65000:0
route-map isp-out-as3356 permit 20
  match ip-address 110
  set community 65000:0 64960:3320
!
router bgp ****
  network ...
! Level3
  neighbor x.x.x.x remote-as 3356
  neighbor x.x.x.x send-community
  neighbor x.x.x.x route-map isp-out-as3356 out
!
```
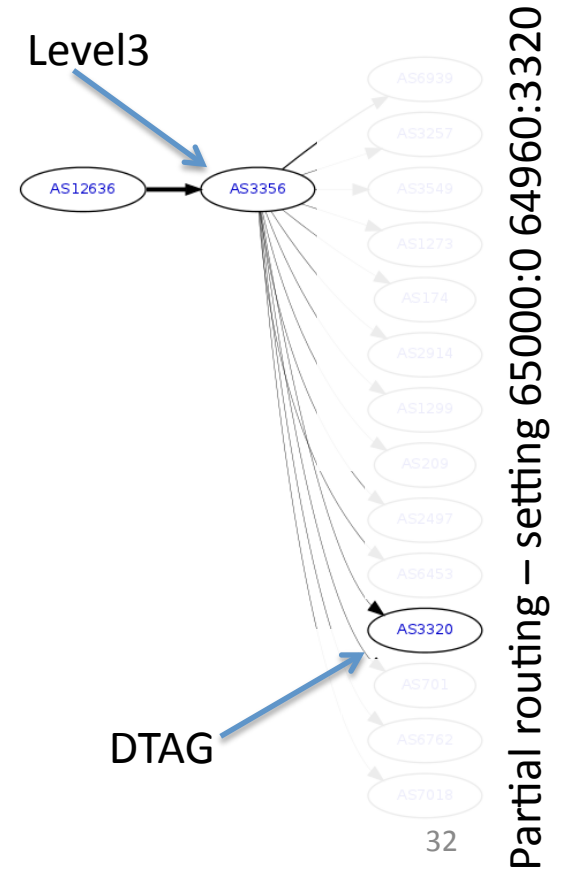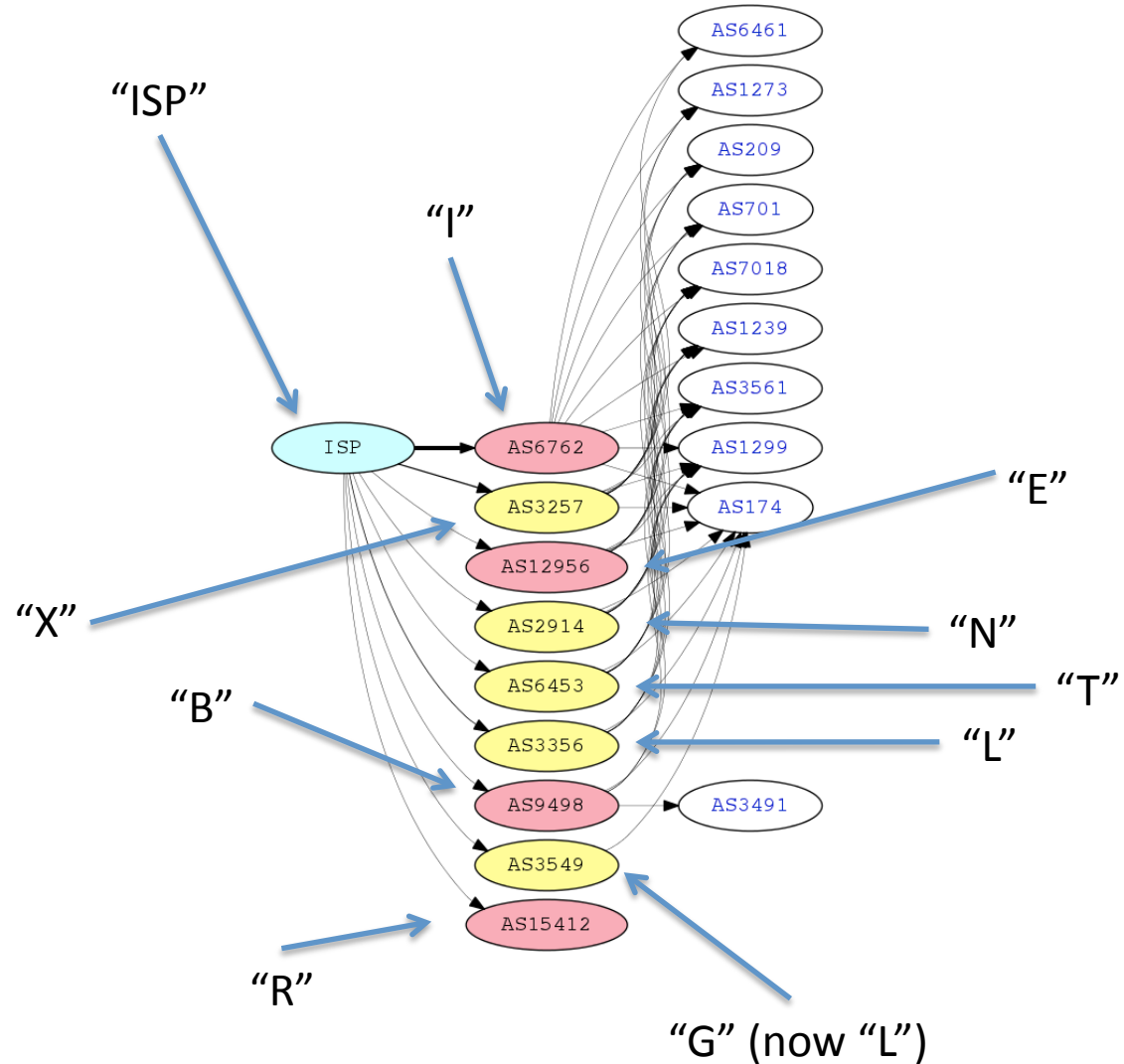
BGP communities – a visual example

Full routing – no communities used

Level3

Peering only routing – setting 65000:0

Level3

DTAG

Partial routing – setting 65000:0 64960:3320

BGP communities – The ISP cleaned up

SUMMARY

Summary

- BGP communities have been around for a long time
  - Well understood
  - Provide many controls; these examples show depth control

- BGP communities can be dangerous
  - Could restrict a routes global reachability
  - Harder to debug (use the transit's looking glass)

- Could lower your #routes announced
  - Reduce router configuration complexity

- Try it!

HURRICANE ELECTRIC
INTERNET SERVICES

?

Contact:

Martin J. Levy
Director, IPv6 Strategy
Hurricane Electric
760 Mission Court
Fremont, CA 94539, USA
http://he.net/

    martin at he dot net
    +1 (510) 580 4167