BGP Wedgries ---- Bad Policy Interactions that Cannot be Debugged

Timothy G. Griffin
Intel Research,
Cambridge UK
tim.griffin@intel.com
http://www.cambridge.intel-research.net/~tgriffin/

NANOG 31
May 23-25, 2004
BGP Wedgie

• BBP policies make sense locally
• Interaction of local policies allows multiple global solutions
• Some solutions are consistent with intended policies, and some are not
• Manual intervention is required to kick the system back to an intended solution
• When unintended solutions are installed, no single AS has enough global knowledge to effectively debug the problem
Shedding Inbound Traffic with ASPATH Prepending

Prepending will (usually) force inbound traffic from AS 1 to take primary link.

Yes, this is a Glorious Hack ...
... But Padding Does Not Always Work

Padding in this way is often used as a form of load balancing. But, AS 3 will send traffic on the "backup" link because it prefers customer routes and local preference is considered before ASPATH length!
COMMUNITIES to the Rescue!

AS 1
provider

AS 3
provider

192.0.2.0/24
ASPATH = 2

primary

backup

Customer

192.0.2.0/24

AS 2

Customer import policy at AS 3:
If 3:90 in COMMUNITY then set local preference to 90
If 3:80 in COMMUNITY then set local preference to 80
If 3:70 in COMMUNITY then set local preference to 70

AS 3: normal customer local pref is 100, peer local pref is 90
Don’t Celebrate Just Yet....

Now, customer wants a backup link to C....
Customer installs a “backup link” ...

customer sends community that lowers local preference below a provider’s
Disaster Strikes!

Provider A (Tier 1) → Provider B (Tier 1) → Provider C (Tier 2) → customer

provider/customer

primary

backup

peering

customer is happy that backup was installed ...
The primary link is repaired, yet routing does repair!

One “solution” --- reset BGP session on backup link!

Better --- C should translate its customer depref communities to those of Provider A when re-exporting routes to A (not DT!)
Using communities for per prefix load balancing
What the heck is going on?

- **There is no guarantee that a BGP configuration has a unique routing solution.**
  - When multiple solutions exist, the (unpredictable) order of updates will determine which one is wins.

- **There is no guarantee that a BGP configuration has any solution!**
  - And checking configurations NP-Complete

- **Complex policies (weights, communities setting preferences, and so on) increase chances of routing anomalies.**
  - ... yet this is the current trend!
More fun with communities ....

- Provider A (Tier 1)
- Provider B (Tier 1)
- Provider C (Tier 2)
- Provider D (Tier 2)

- customer sends to D
  community that lowers preference below D’s peers but above D’s providers
- customer sends to C
  community that lowers preference below C’s peers and C’s providers

backup I
backup II
Primary goes down!

Provider A (Tier 1)

Provider B (Tier 1)

Provider C (Tier 2)

Provider D (Tier 2)

backup I

backup II

customer
Primary repaired

Provider A (Tier 1)

Provider B (Tier 1)

Provider D (Tier 2)

Provider C (Tier 2)

backup I

backup II

customer
Reset Backup I session?
First, take it down....
Now Bring it up ... THE FULL WEDGIE!

“solution” --- reset BGP session on BOTH backup links simultaneously!

Customer has no reason to think D is involved!
BGP Wedgie

- BBP policies make sense locally
- Interaction of local policies allows multiple global solutions
- Some solutions are consistent with intended policies, and some are not
- Manual intervention is required to kick the system back to an intended solution
- When unintended solutions are installed, no single AS has enough global knowledge to effectively debug the problem
Recommendations

• Interdomain communities that can tweak a route’s preference should be defined with care and consistently implemented.

• Communities should be translated and transited