RFC 4264 : BGP Wedgies ---- Bad Policy Interactions that Cannot be Debugged

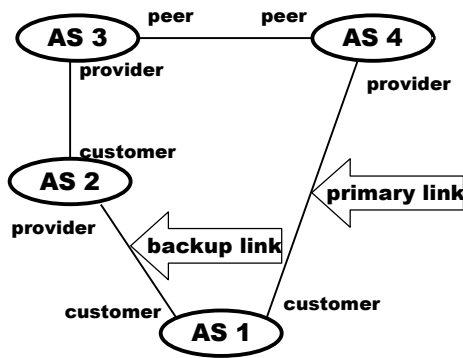**Timothy G. Griffin**

Geoff Huston

What is a BGP Wedgie?

full wedgie { **¾ wedgie** {

- **BGP policies make sense locally**
- **Interaction of local policies allows multiple stable routings**
- **Some routings are consistent with intended policies, and some are not**
  - **If an unintended routing is installed (BGP is "wedged"), then manual intervention is needed to change to an intended routing**
- **When an unintended routing is installed, no single group of network operators has enough knowledge to debug the problem**

¾ Wedgie Example

AS 3 — peer — peer — AS 4

provider (AS 3) — provider (AS 4)

customer (AS 2)

provider (AS 2)

primary link

backup link

customer — customer — AS 1

- AS 1 implements backup link by sending AS 2 a "depref me" community.
- AS 2 implements this community so that the resulting local pref is below that of routes from it's upstream provider (AS 3 routes)
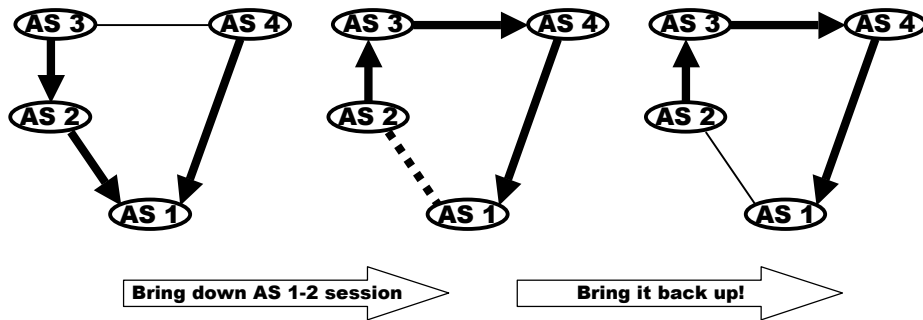
And the Routings are...

AS 3 → AS 4

AS 2

AS 1

AS 3 — AS 4

AS 2

AS 1
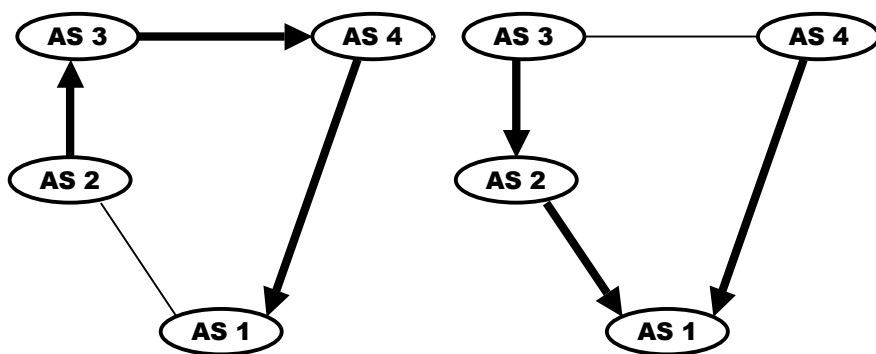
Intended Routing

Note: this would be the ONLY routing if AS2 translated its "depref me" community to a "depref me" community of AS 3

Unintended Routing

Note: This is easy to reach from the intended routing just by "bouncing" the BGP session on the primary link.
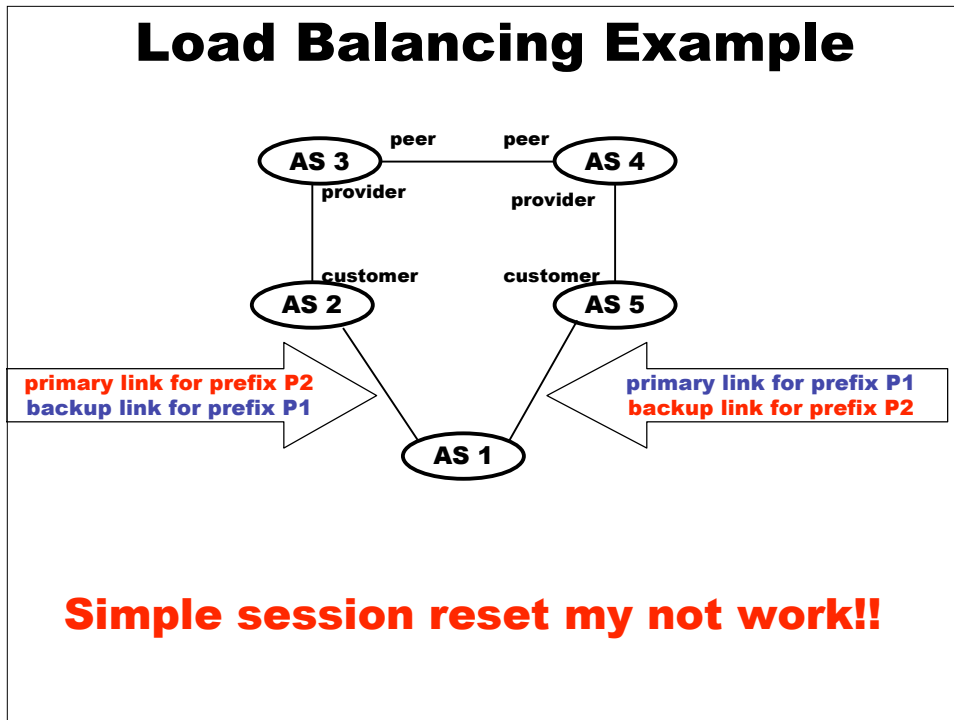
Recovery

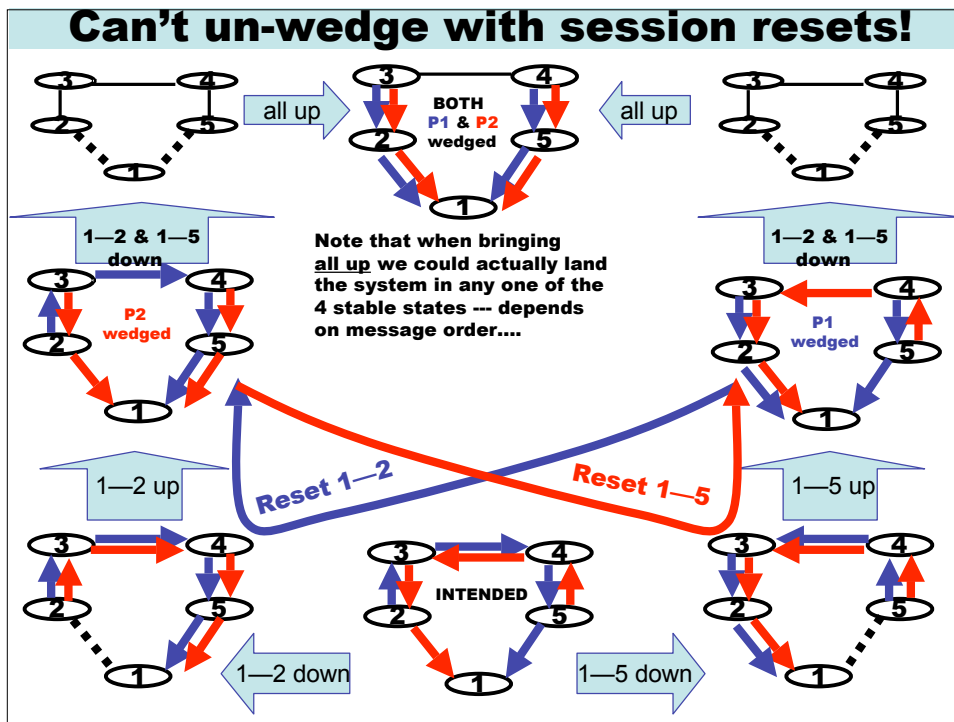- **Requires manual intervention**
- **Can be done in AS 1 or AS 2**

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**
  - **Lab demonstrations of BGP configs never converging**
- **Complex policies (weights, communities setting preferences, and so on) increase chances of routing anomalies.**
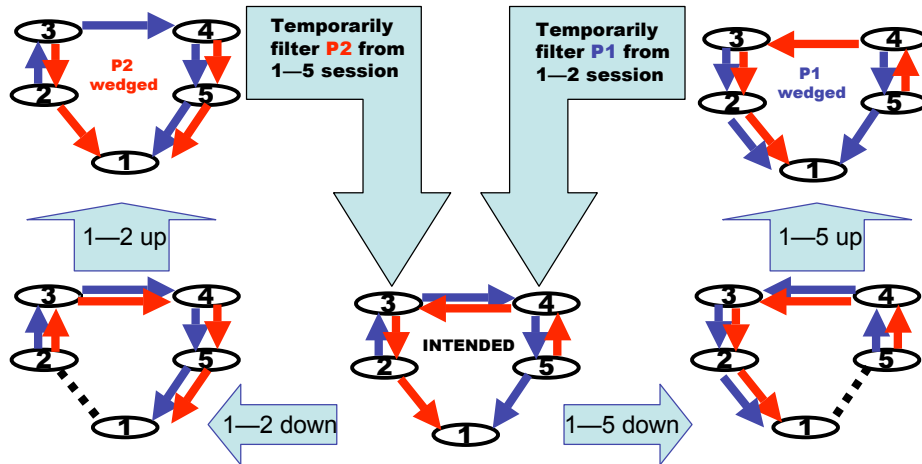 - **… yet this is the current trend!**

Load Balancing Example

AS 3 — peer — peer — AS 4

provider   provider

customer   customer

AS 2   AS 5

AS 1

primary link for prefix P2
backup link for prefix P1

primary link for prefix P1
backup link for prefix P2

Simple session reset my not work!!

Can't un-wedge with session resets!

all up   BOTH P1 & P2 wedged   all up

1—2 & 1—5 down

Note that when bringing __all up__ we could actually land the system in any one of the 4 stable states --- depends on message order....

1—2 & 1—5 down

P2 wedged

P1 wedged

Reset 1—2   Reset 1—5

1—2 up   1—5 up

INTENDED

1—2 down   1—5 down

4

Recovery

Temporarily filter P2 from 1—5 session

Temporarily filter P1 from 1—2 session

P2 wedged

P1 wedged

1—2 up

1—5 up

INTENDED

1—2 down

1—5 down
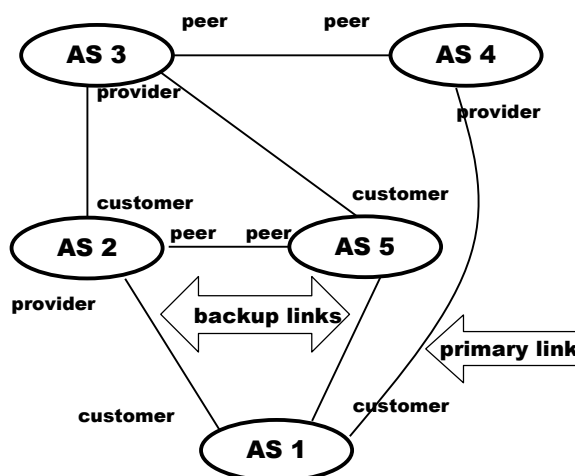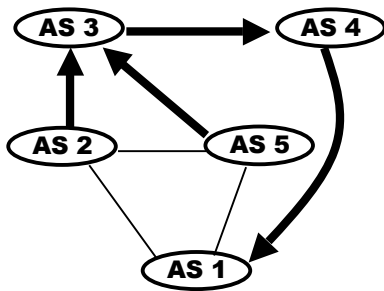
Who among us could figure this one out?
When 1—2 is in New York and 1—5 is in Tokyo?

Full Wedgie Example

peer   peer

AS 3   AS 4

provider

provider

customer   customer

AS 2   peer   peer   AS 5

provider

backup links

primary link

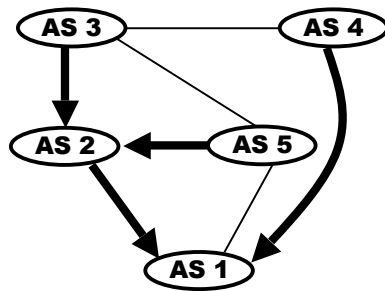customer   customer

AS 1

- AS 1 implements backup links by sending AS 2 and AS 3 a "depref me" communities.
- AS 2 implements its community so that the resulting local pref is below that of its upstream providers and it's peers (AS 3 and AS 5 routes)
- AS 5 implements its community in the SAME WAY

- so that the resulting local pref is below its peers (AS 2) but above that of its providers (AS 3)
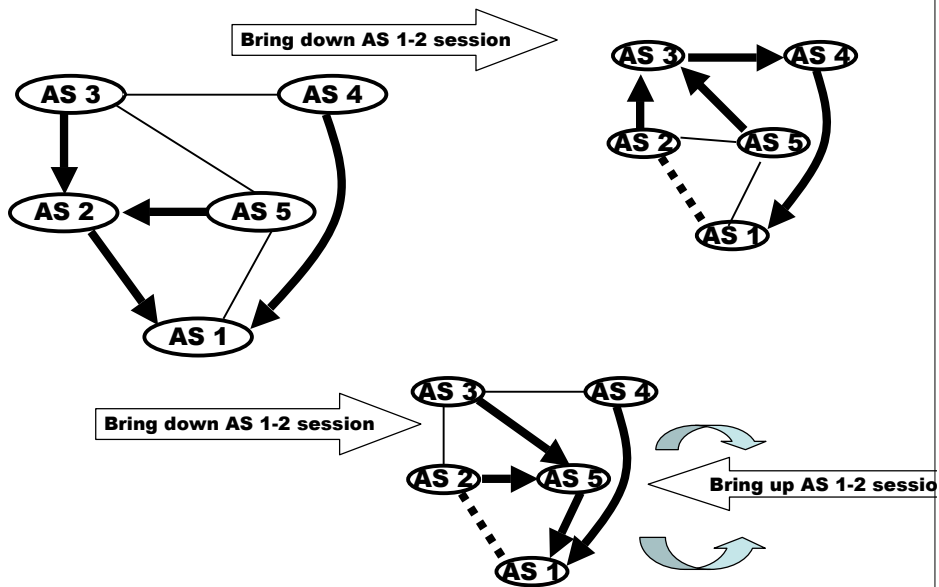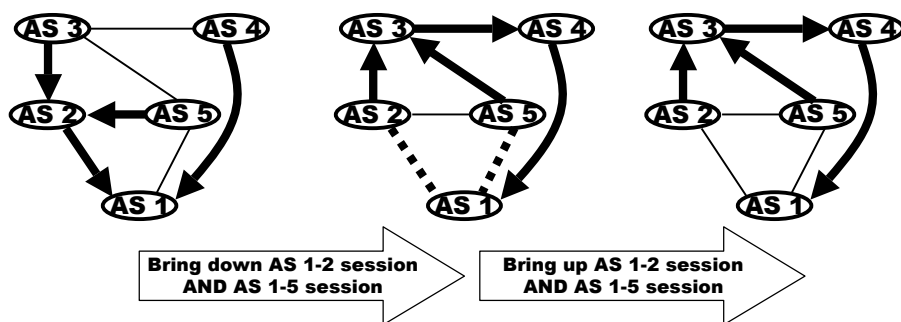
And the Routings are...

Intended Routing

Unintended Routing


Resetting 1—2 does not help!!

Bring down AS 1-2 session

Bring down AS 1-2 session

Bring up AS 1-2 session

Recovery

Bring down AS 1-2 session AND AS 1-5 session → Bring up AS 1-2 session AND AS 1-5 session →

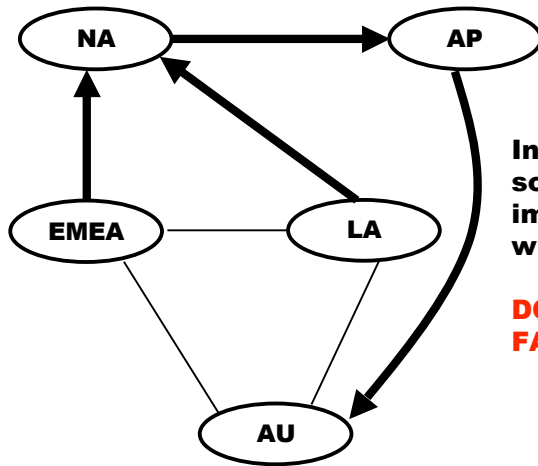A lot of "non-local" knowledge is required to arrive at this recovery strategy!

Try to convince AS 5 and AS 1 that their session has be reset (or filtered) even though it is not associated with an active route!

That Can't happen in MY network!!
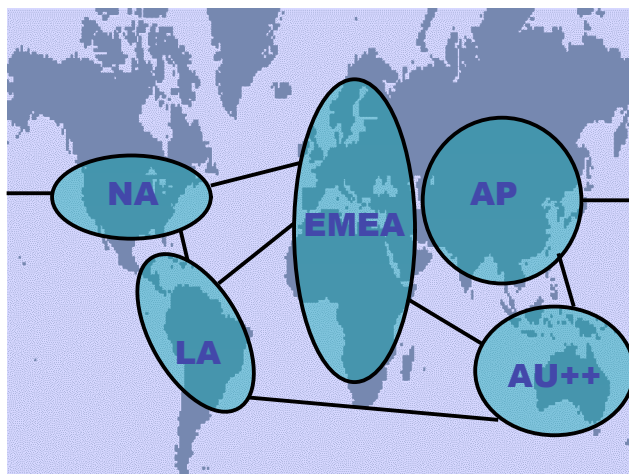
An "normal" global global backbone (ISP or Corporate Intranet) implemented with 5 regional ASes

The Full Wedgie Example, in a new Guise

Intended Routing for some prefixes in AU, implemented with communities.

DOES THIS LOOK FAMILIAR??

Message: Same problems can arise with "traffic engineering" across regional networks.