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

tim.griffin@intel.com

**http://www.cambridge.intel-research.net/~tgriffin/**

APNIC 18 Routing SIG
September 2, 2004
Fiji

What is a BGP Wedgie?

¾ wedgie

full
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
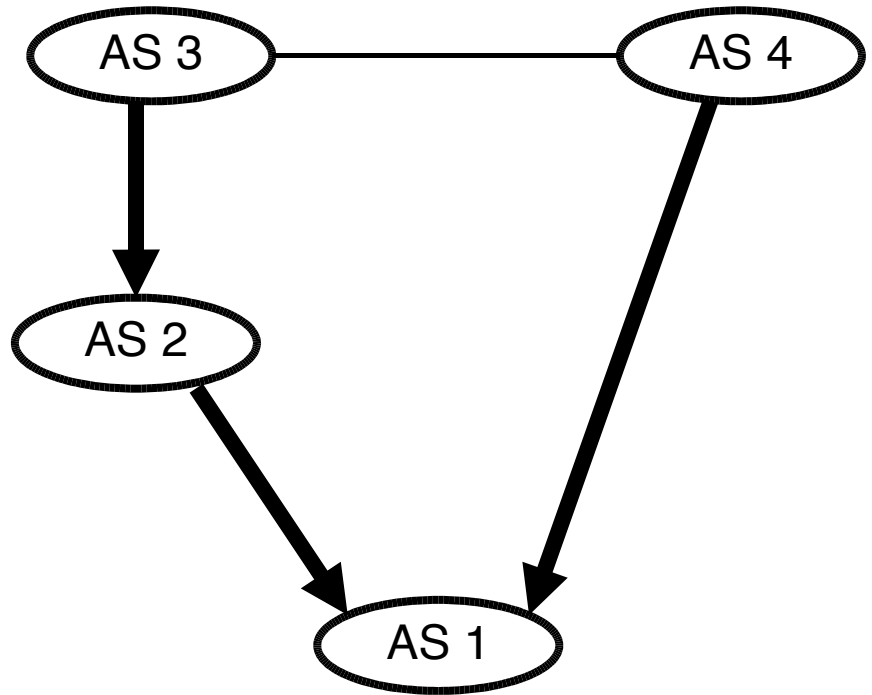
¾ Wedgie Example
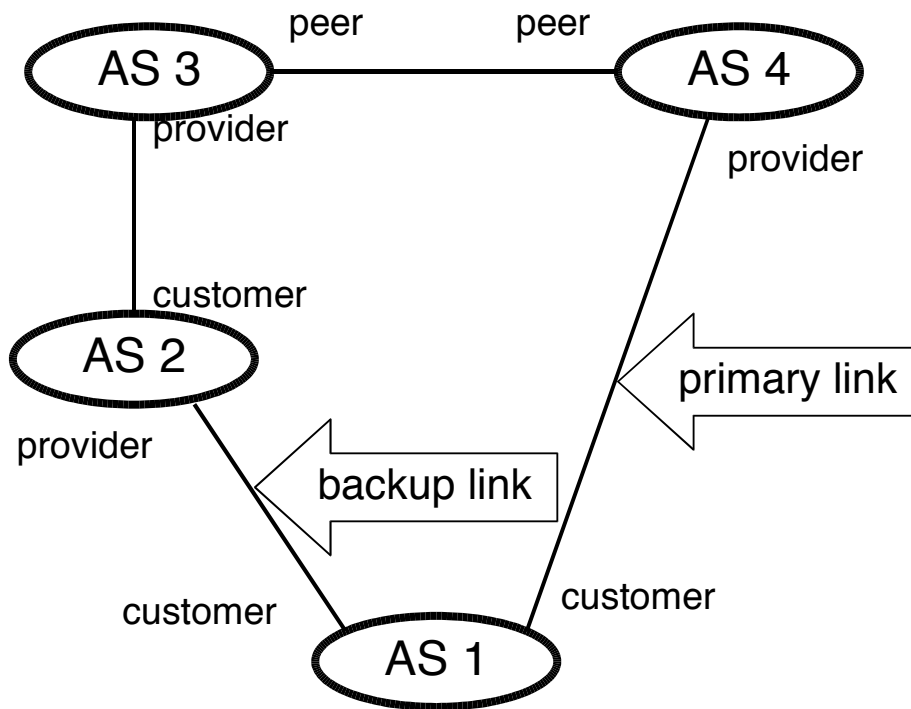
- AS 1 implements backup link by sending AS 2 a "depref me" community.

- AS 2 implements this community so that the resulting local pref is below that of routes from it's upstream provider (AS 3 routes)
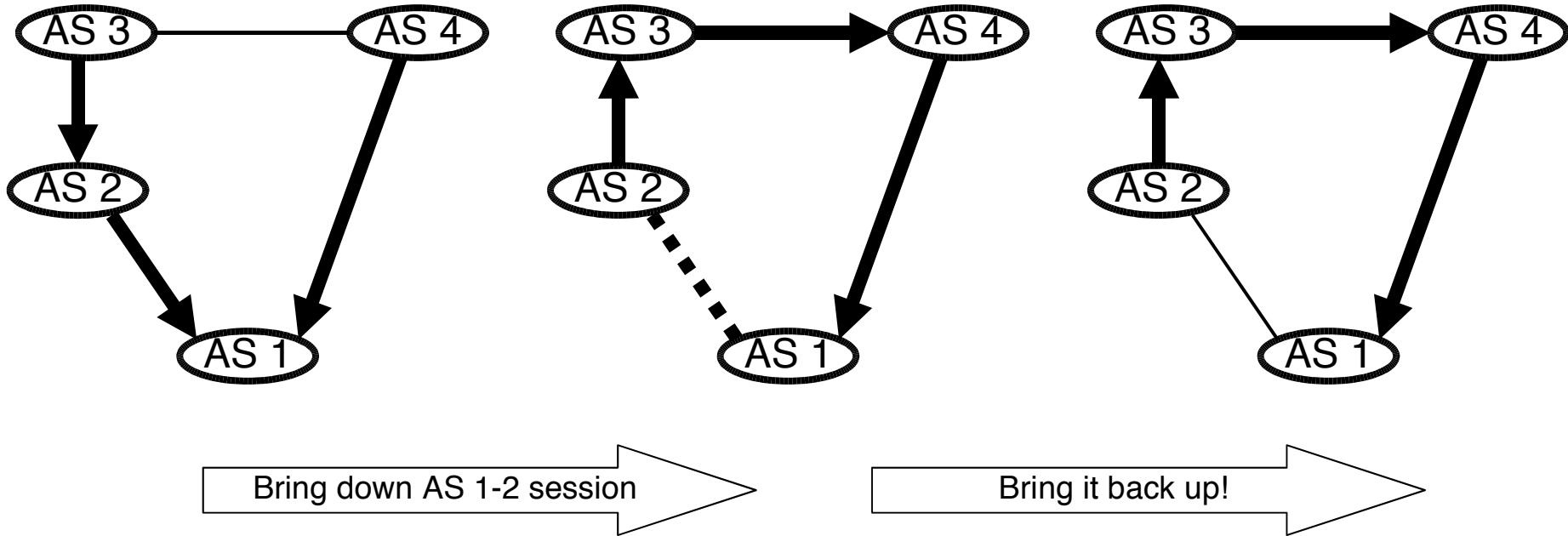
And the Routings are…
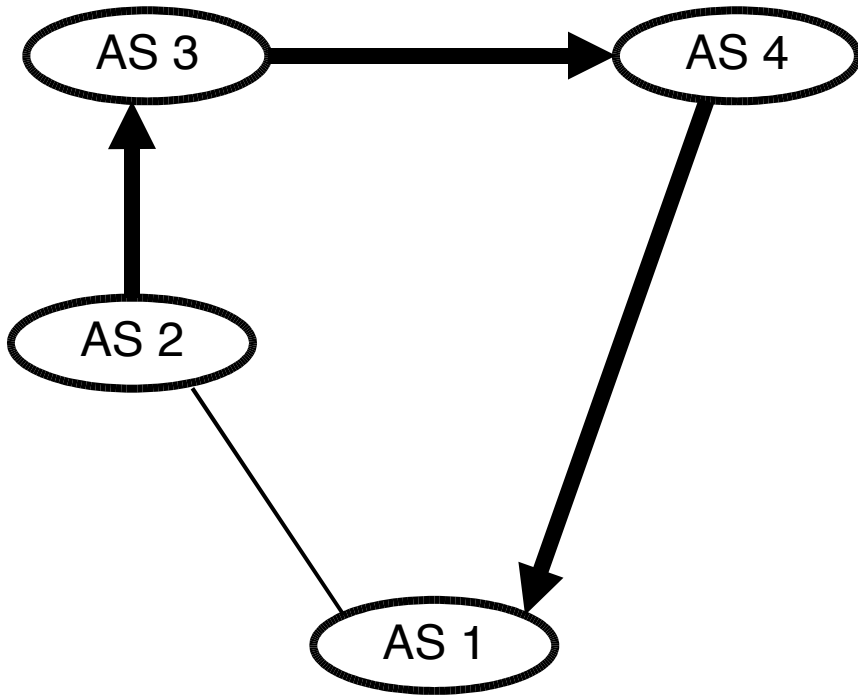
Intended Routing

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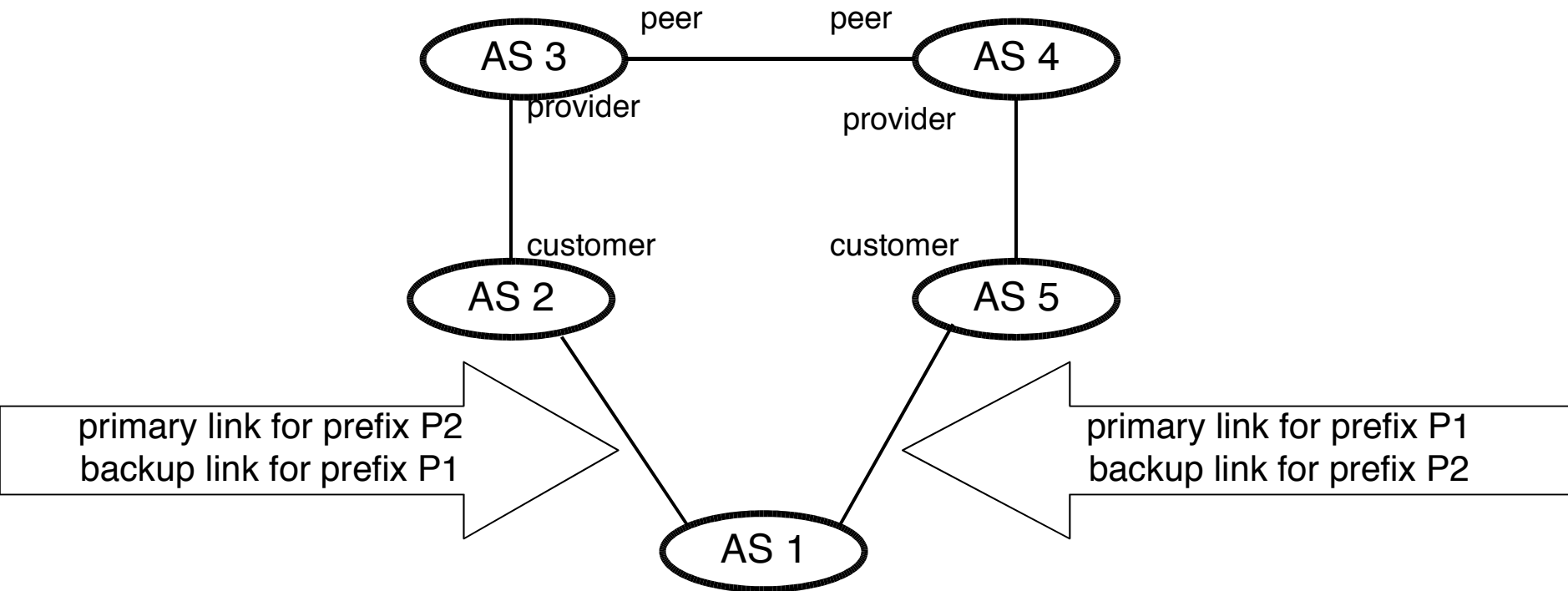
Unintended Routing

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

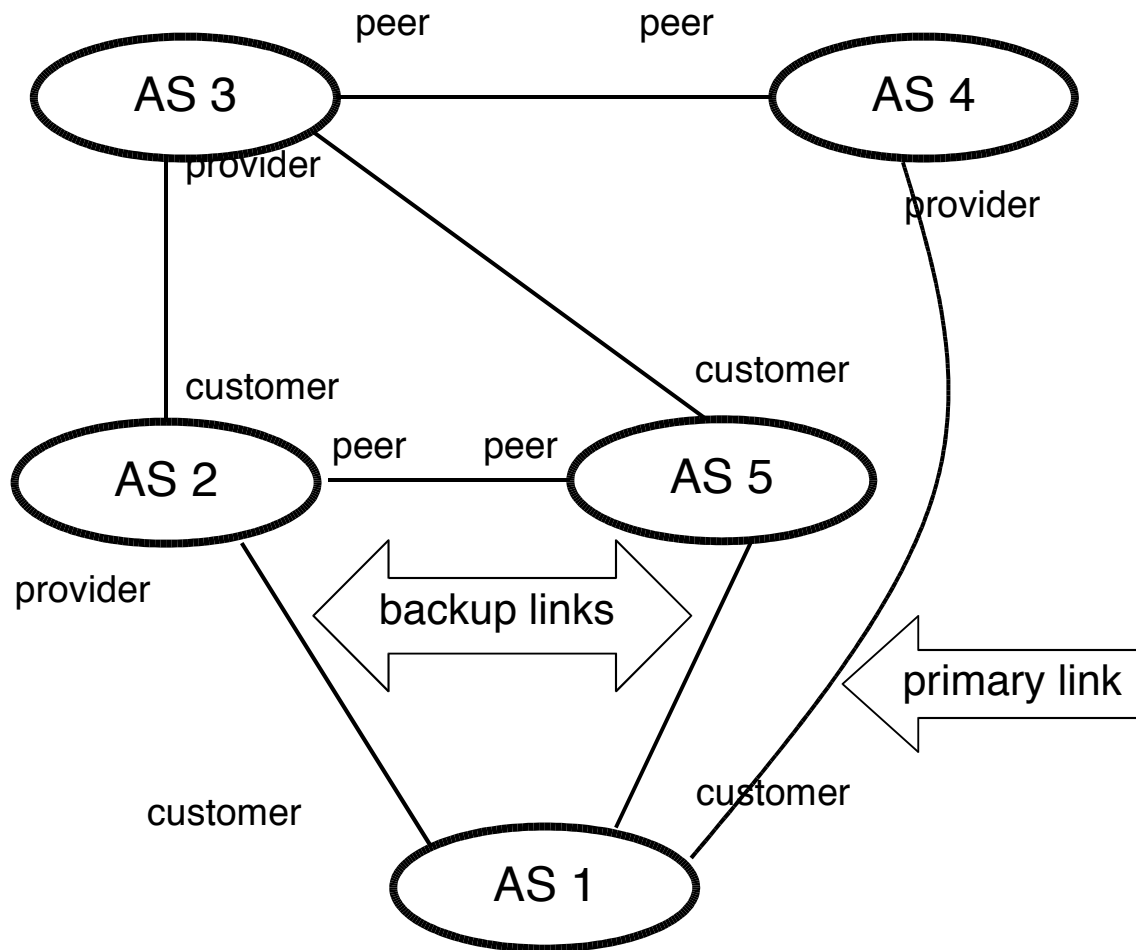# Recovery

Bring down AS 1-2 session

Bring it back up!

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

Load Balancing Example

- Recovery for prefix P1 may cause a BGP wedgie for prefix P2 …

Full Wedgie Example

AS 3 —peer— peer— AS 4

AS 3 —provider

AS 4 —provider

AS 3 —customer— customer— AS 5

AS 2 —peer— peer— AS 5

AS 2 —provider

backup links

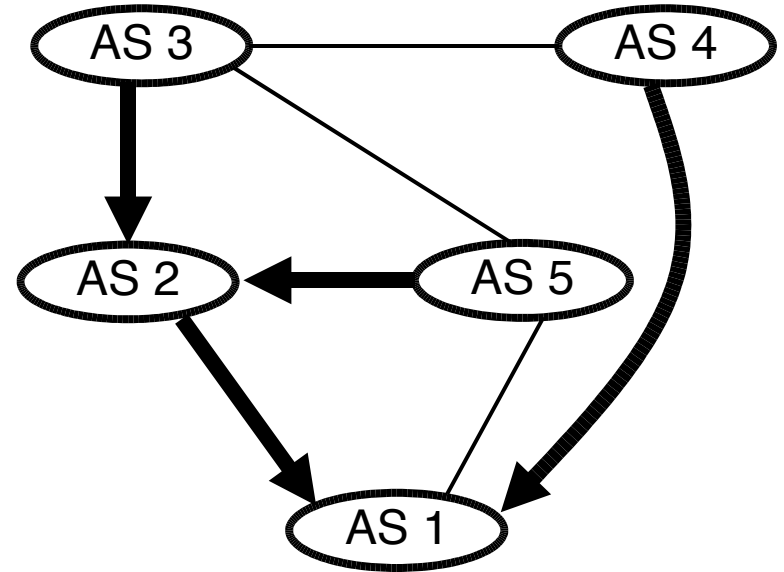primary link

AS 2 —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 so that the resulting local pref is below its peers (AS 2) but above that of its providers (AS 3)
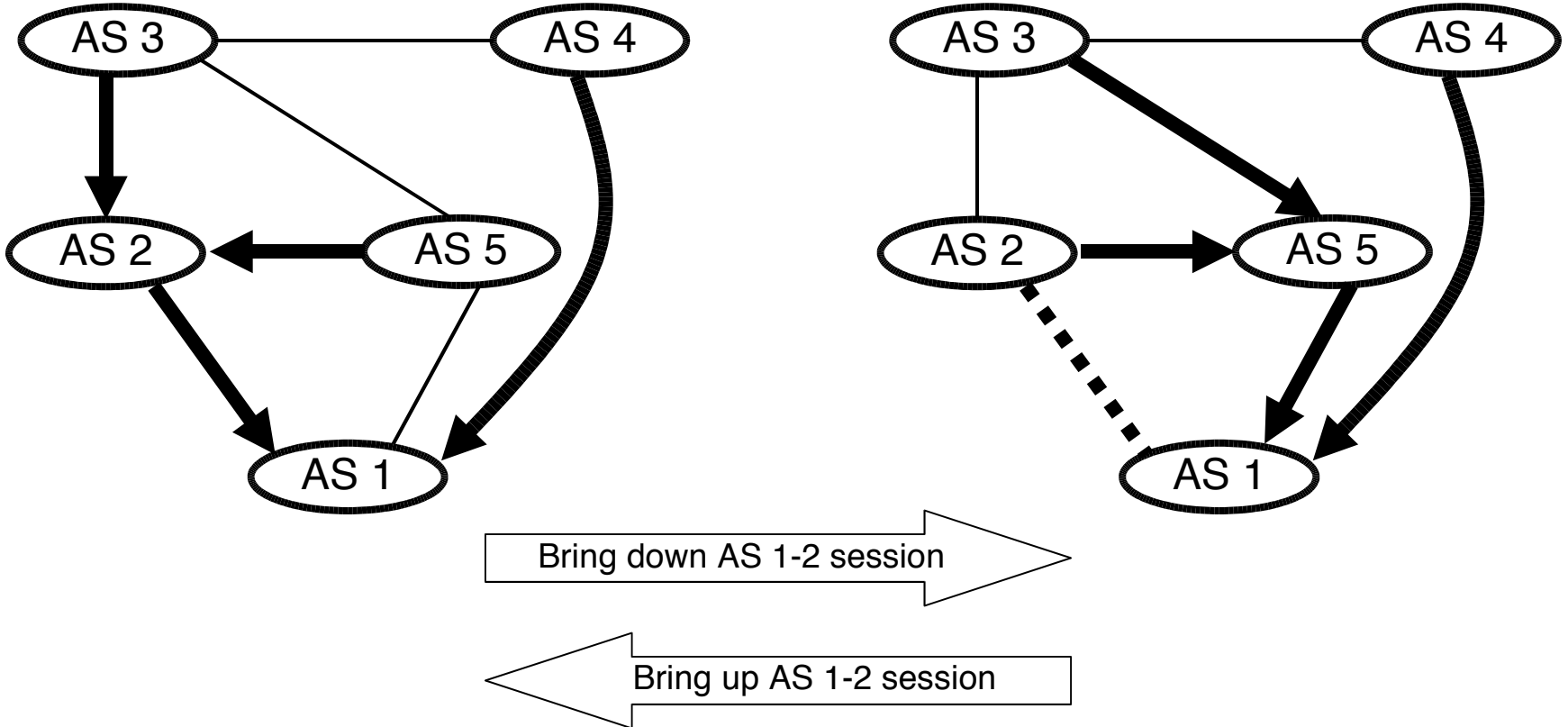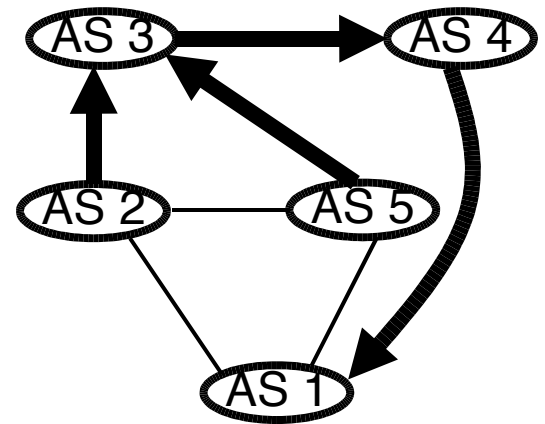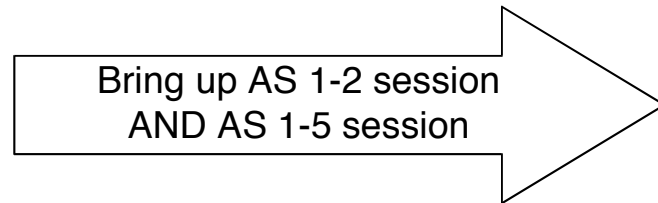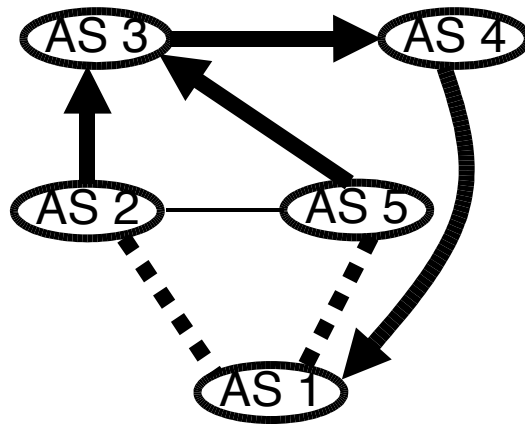
And the Routings are…
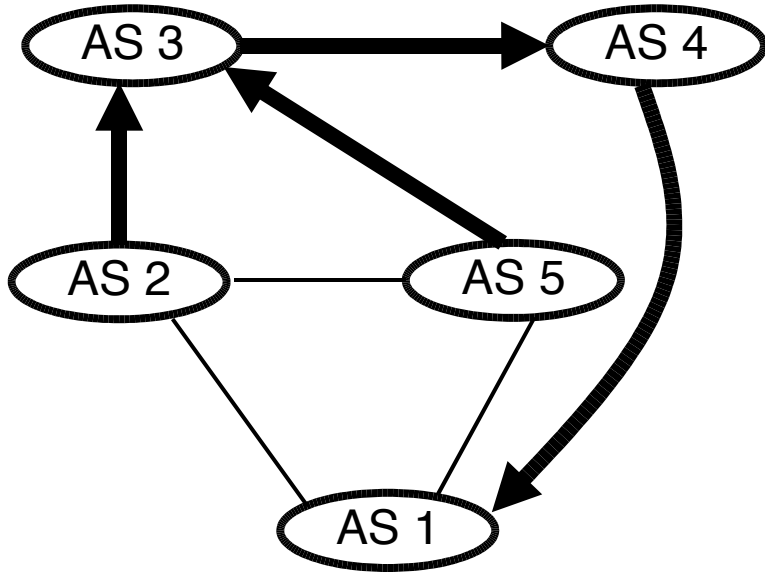
Intended Routing

Unintended Routing

Recovery??

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
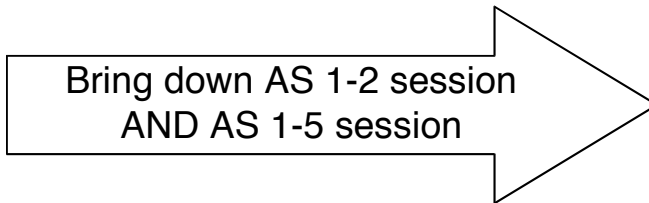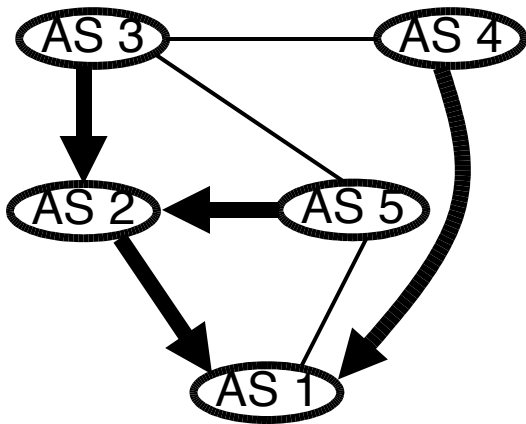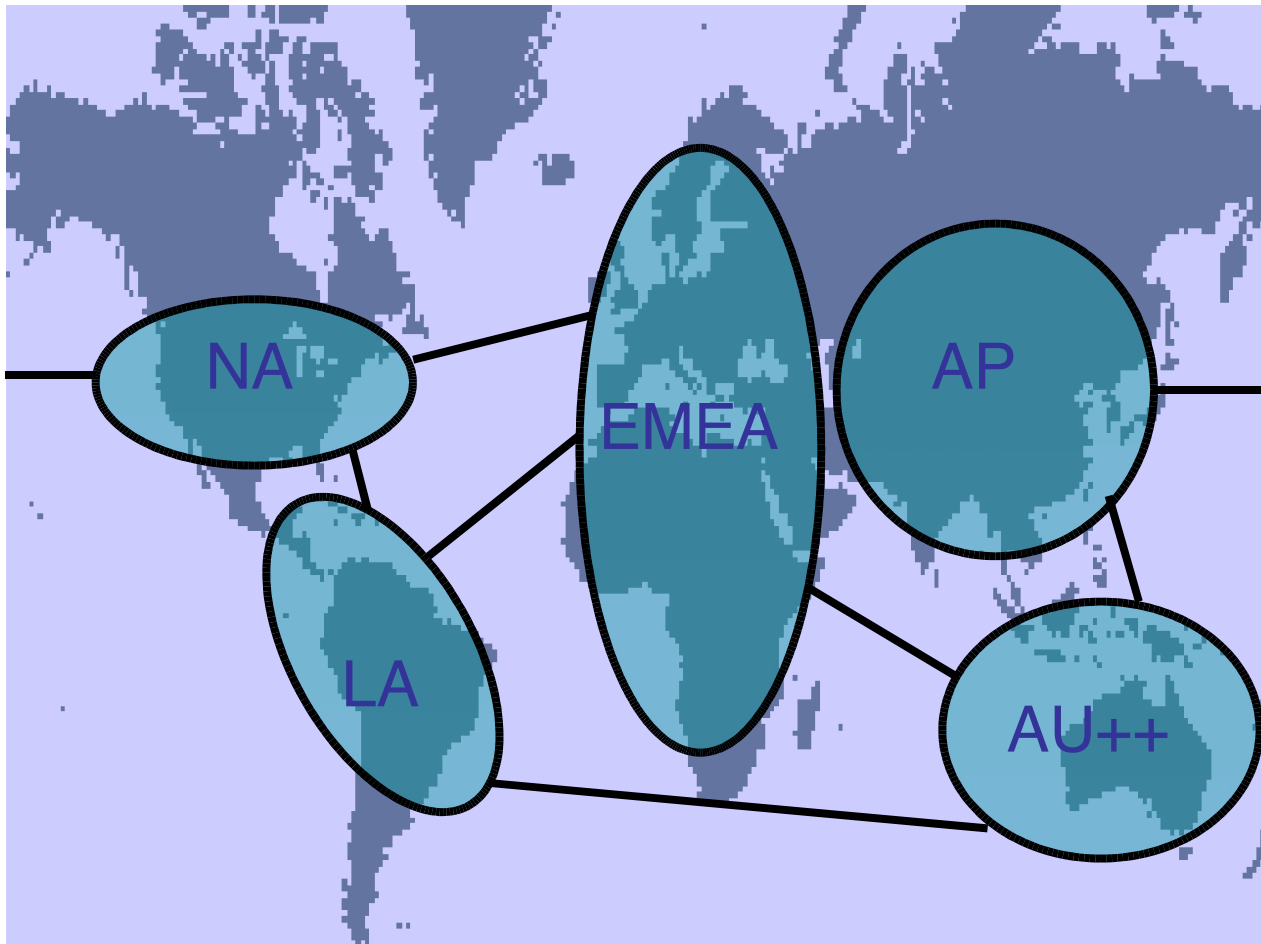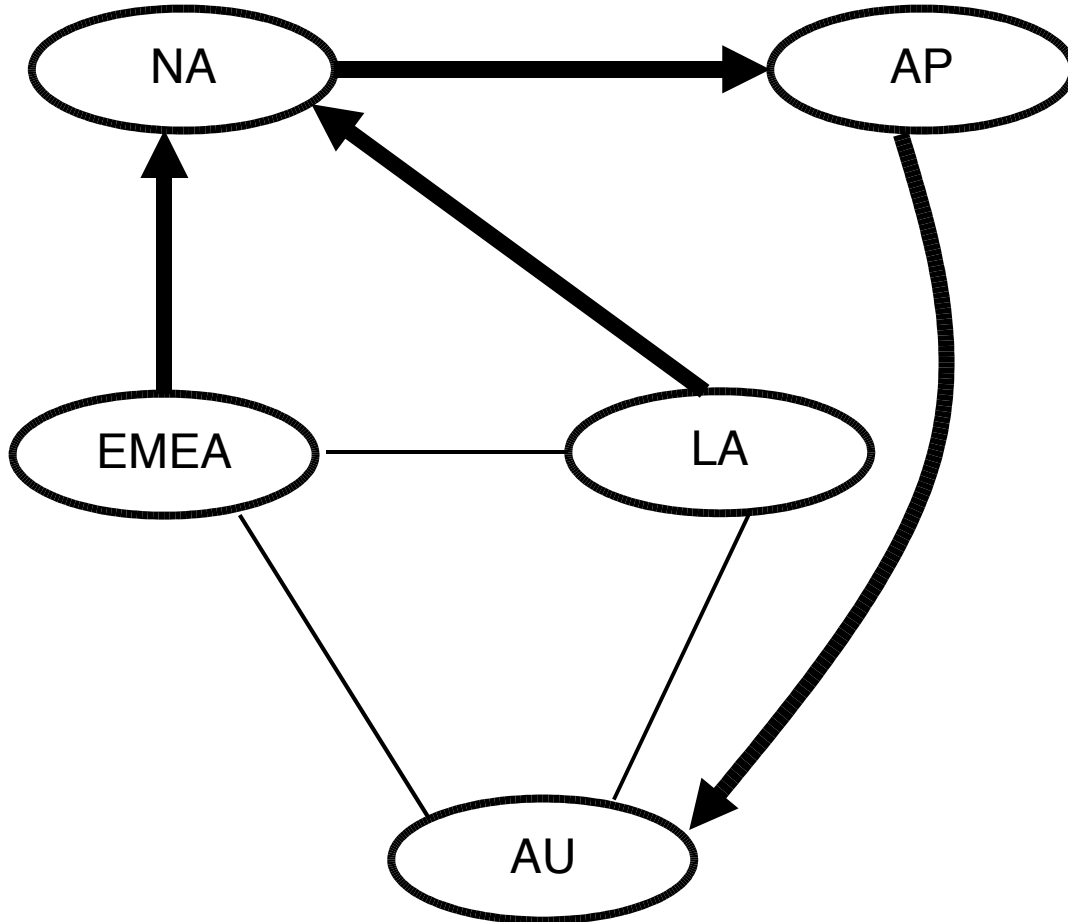
Try telling AS 5 that it has
to reset a BGP session that is
not associated with a BEST route!

A Global ISP (or Corporate Intranet) Implemented with 5 ASes

Full Wedgie Example, in a new Guise

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

Intended Routing for some prefixes in AU

Recommendations

- Be aware of BGP Wedgies

- Interdomain communities that can tweak a route's preference should be defined with care and consistently implemented

- Tools to enumerate all stable routings would be useful
  - inherently exponential in theory, but may not be that bad in practice (on instances much smaller than global Internet!)
  - I'm currently attempting an implementation on top of http://nms.lcs.mit.edu/bgp/rcc/

Comments Appreciated!

tim.griffin@intel.com