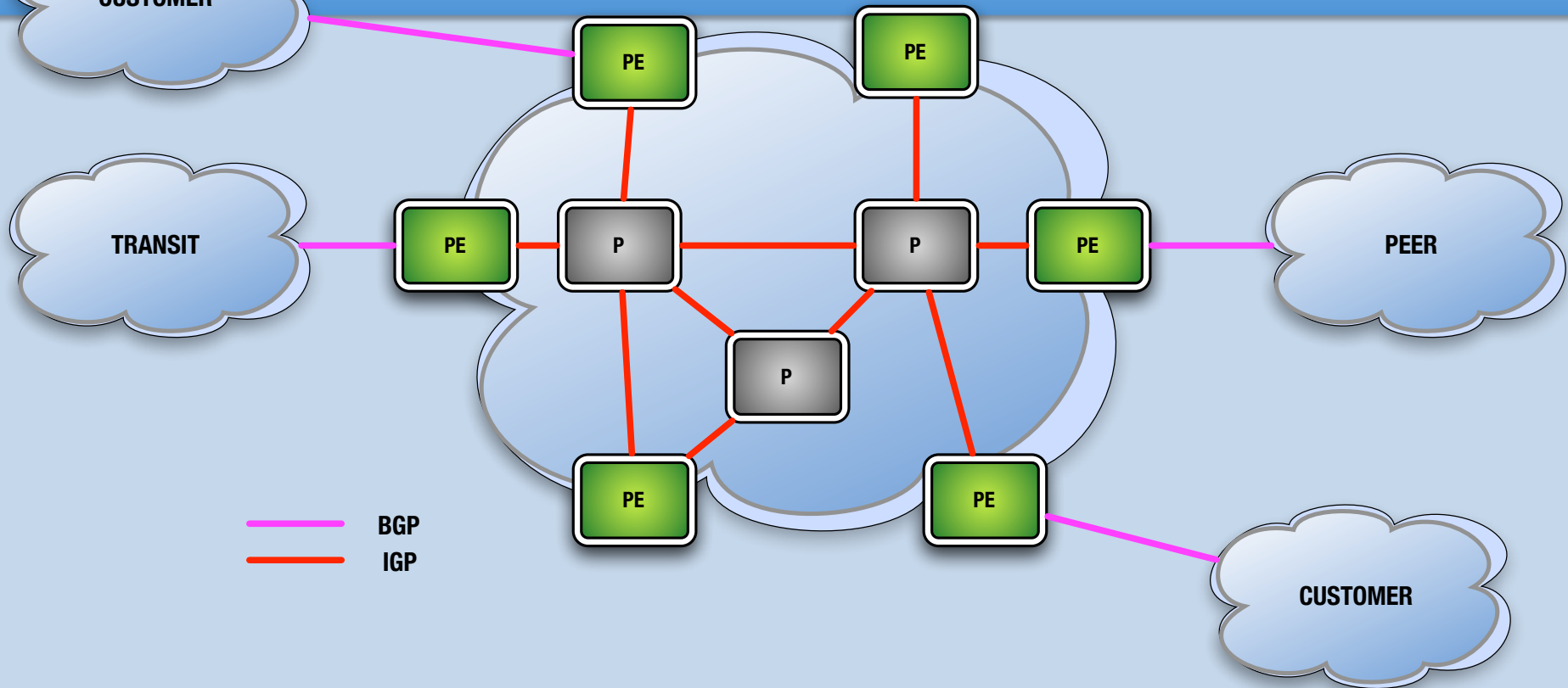


BGP ERROR HANDLING.

DEVELOPING AN OPERATOR-LED APPROACH IN THE IETF.

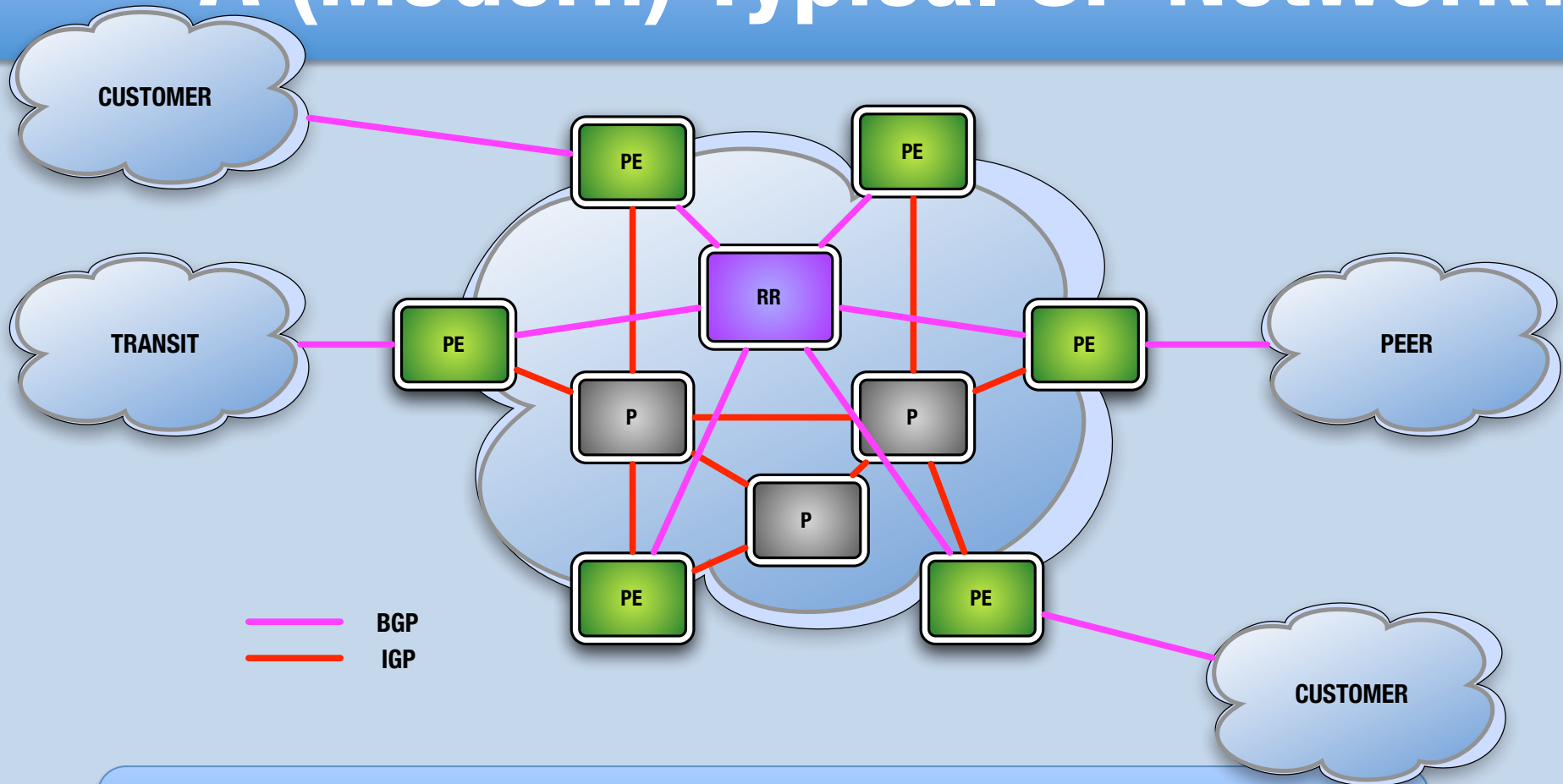
Rob Shakir, Cable&Wireless Worldwide.

A Typical SP Network?



IGP Signals customer/Internal prefixes between PEs
EGP Propagates internal prefixes to neighbouring ASes.

A (Modern) Typical SP Network?



IGP Minimal infrastructure routing information.
BGP Propagate internal routing and service data.

BGP Failures I.

**JAN.
09**

ERRORS IN AS4_PATH

Erroneous data in the AS4_PATH optional transitive attribute causing BGP session failure (JunOS bug).

VERY LONG AS_PATH

Very long AS_PATHs in the global BGP table cause session failure. Not the first time this had been seen.

**FEB.
09**

BGP Failures II.

**AUG.
10**

RIPE NCC RIS EXPERIMENT

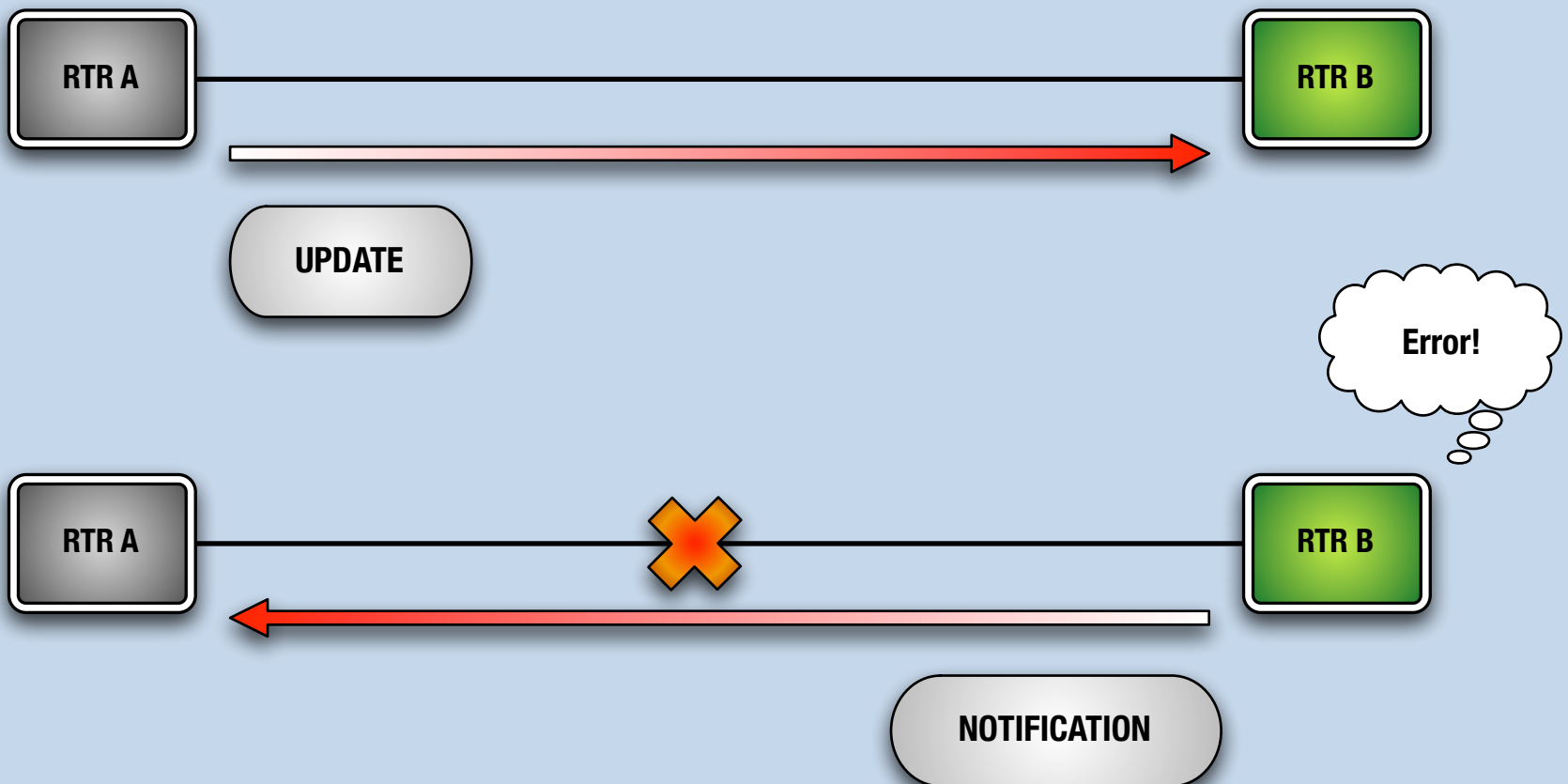
A RIPE NCC RIS/Duke University experiment results in BGP sessions being reset – disrupting global table (IOS XR bug).

iBGP FAILURES

Multiple occurrences within xSP networks.
Likely to cause higher financial impact (L3VPN margin).

**??
??**

Why do we see these events?



Cause/Impact.

**LIMITED
TOOLSET IN STANDARDS.**

Must either DISCARD attributes or
respond with NOTIFICATION.

**SERVICE
IMPACT.**

Transit/Peering failure - although error source may be remote.
iBGP failure – high impact sessions? Route reflectors?

Results in loss of RIB!

Would you tolerate this in your IGP based on one erroneous LSP?

Intent of Work.

DEFINE HOW BGP IS USED.

Document the way xSPs use BGP.
Ensure that critical nature of the protocol is understood.

PROVIDE REQUIREMENTS

Determine how **OPERATORS** think that BGP should fail – and what we'll compromise on.

TIE TOGETHER IETF WORK ITEMS.

Ensure that tools resulting from existing drafts form a useful framework to make BGP robust.

Approach Overview.

01

DON'T SEND NOTIFICATION.

02

RECOVER RIB CONSISTENCY.

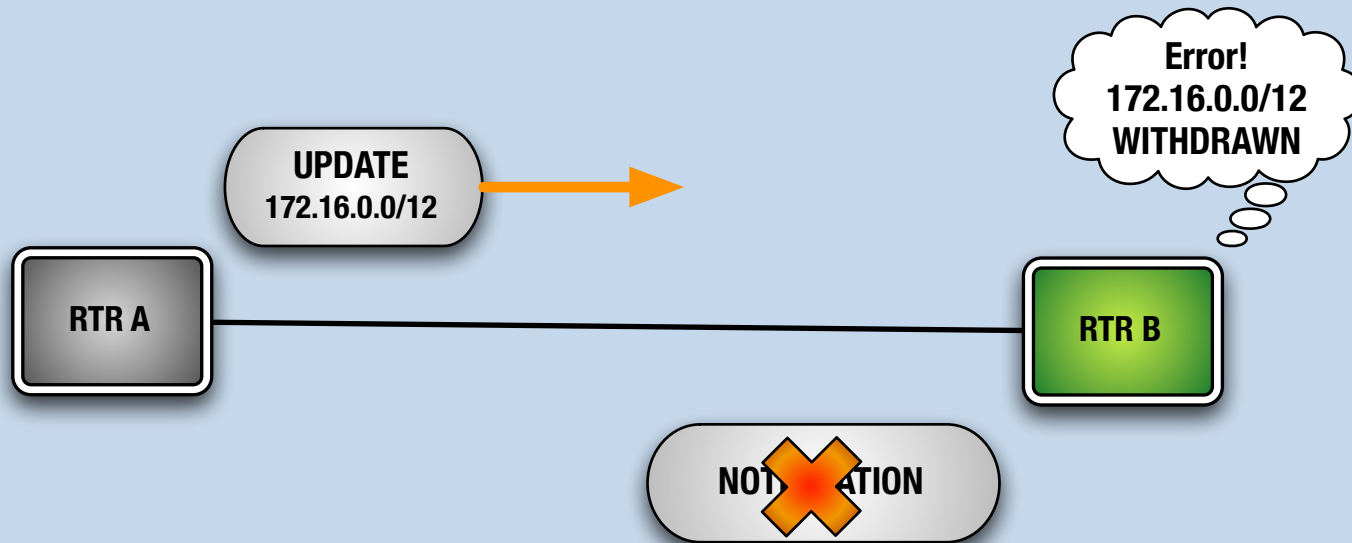
03

RESTART BGP HITLESSLY.

04

MONITORING

Avoid sending NOTIFICATION.



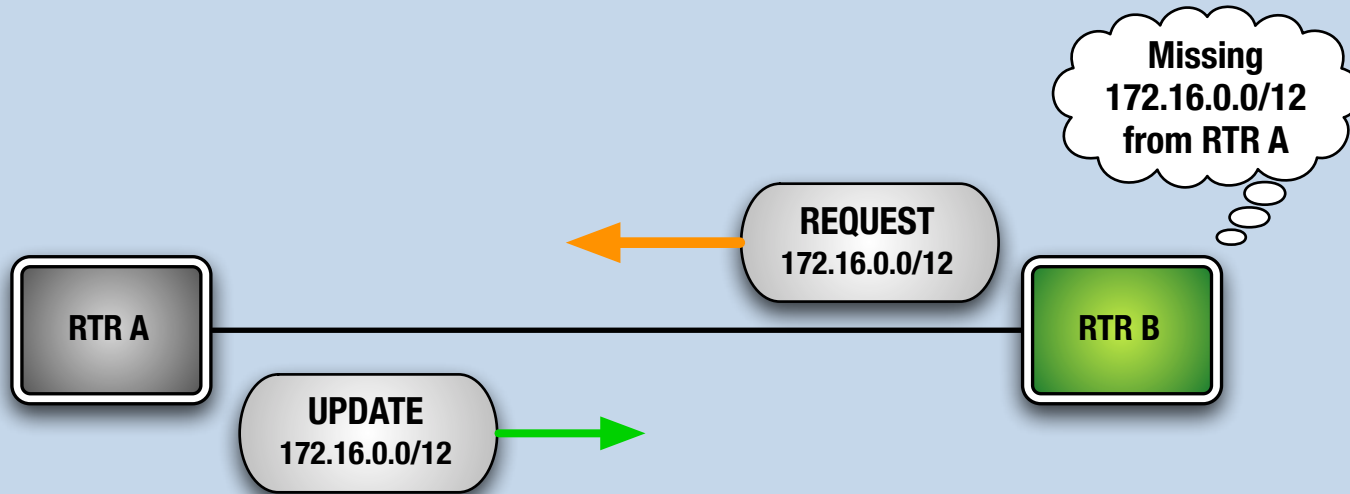
**WHAT DO WE
COMPROMISE ON?**

“treat-as-withdraw” mechanism can result in routing inconsistency (possible loops!).

**EXISTING WORK
ITEMS IN IETF?**

draft-chen (eBGP errors) – includes Opt Trans.
Needs to be extended to cover iBGP.

Recover RIB Consistency.



**HOW CAN THIS
BE ACHIEVED?**

Mechanisms to re-request missing NLRI.
One prefix at once, or whole RIB.

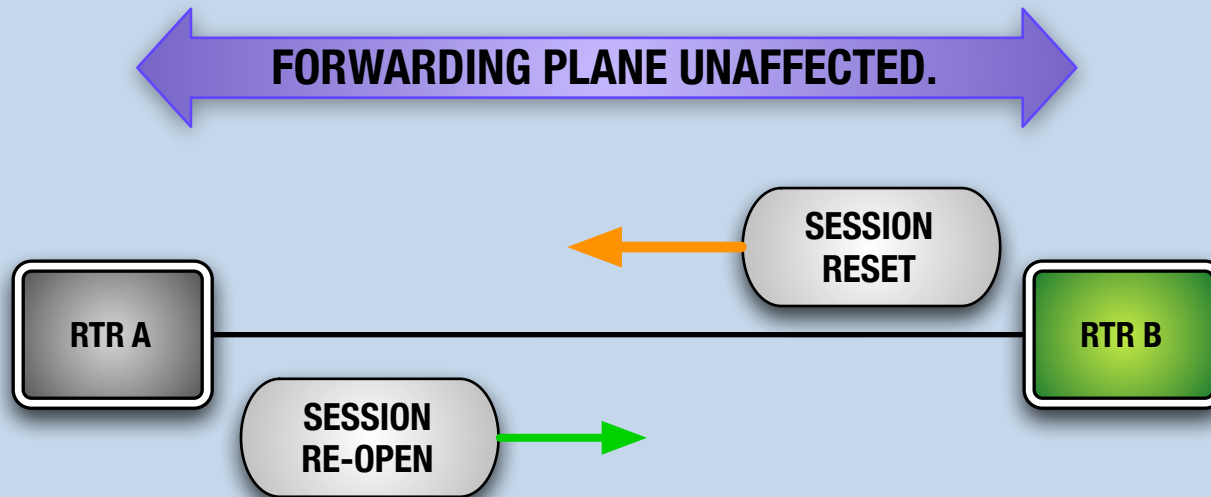
**EXISTING WORK
ITEMS?**

“One-Time Prefix ORF”.
Enhanced ROUTE REFRESH.

Reduce Impact of Session Reset.

**SESSION RESETS,
CAN WE AVOID THEM?**

NOTIFICATION has utility for resetting state.
Consider that sometimes it is unavoidable.



**EXISTING WORK
ITEMS IN IETF?**

(Expired) “SOFT-NOTIFICATION”.
Further work required to revive!

Introduce Further Monitoring.

EXISTING ERRORS ARE VERY VISIBLE.

NOCs can see session failures very easily – both via session monitoring and forwarding outage!

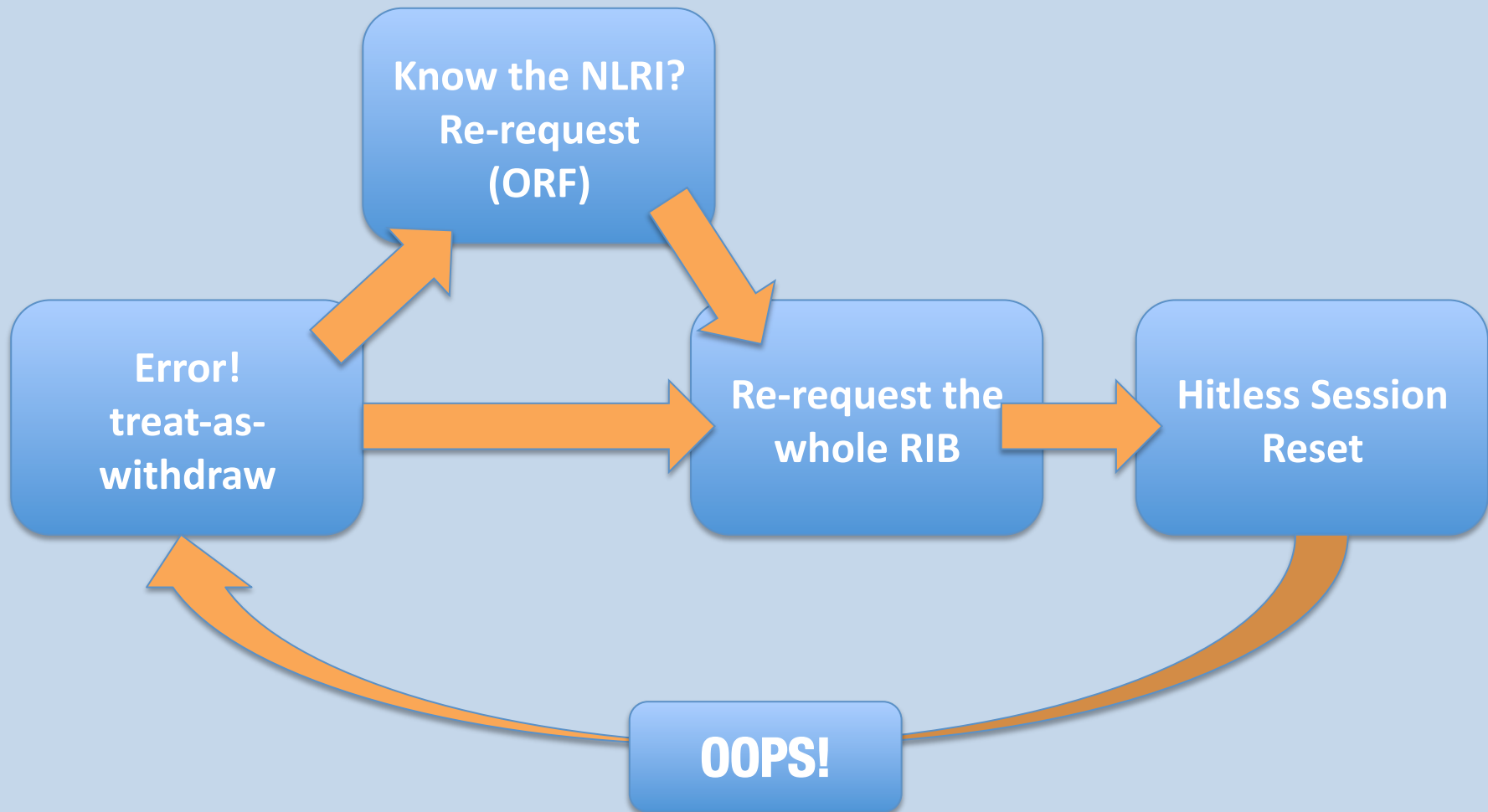
FURTHER COMPLEXITY MEANS LESS MANAGEABLE

Mechanisms are required to make error handling visible to both BGP speakers.

EXISTING WORK ITEMS IN IETF?

(In-band) ADVISORY and DIAGNOSTIC.
(Out-of-Band) BGP Monitoring Protocol.

Complexities of Approach.



Why am I standing here?

NANOg

As Operators, we deal with the fall-out of protocol issues!

SO... an agreed, operator-recommended approach is required.

Questions, comments, review...

ALL MUCH APPRECIATED!

Feedback form at <http://rob.sh/bgp>

<http://tools.ietf.org/html/draft-shakir-idr-ops-reqs-for-bgp-error-handling-00>

rob.shakir@cw.com // +44(0)207 100 7532 // RJS-RIPE

ADDITIONAL SLIDES.

Q&A AND FURTHER BACKGROUND.

Receiver side only?

UTILITY FOR RX-SIDE BUGS?

Yes – we can still use these mechanisms.
Although utility of RIB consistency refresh is reduced.

Error!
treat-as-
withdraw



Hitless Session
Reset



Repeat errors –
NLRI remains
withdrawn?

REQUIREMENTS

Must ensure that this is flagged to Operator.
Implementation bugs cannot be recovered by **any** protocol mechanism.

Multi-session BGP

IS MULTISESSION ANOTHER SOLUTION?

No – it helps, but is not a complete solution.
Many topologies in one AFI.

DOES HAVE UTILITY

Can achieve AFI error separation.
e.g. IPv4 and IPv6 errors can be independent of each other.

SOLUTION DOES NOT SCALE

For complete solution, we would need one session per topology – control-plane does not scale to this!