BGP Session Culling / BCP214

Will Hargrave // LONAP

will@lonap.net

INEX
2018-03-15

What is an IXP?

- A switched fabric for interconnecting networks
  – usually 1->100 interconnected ethernet switches
- Operators connect their routers and exchange traffic with each other
- BGP is the control plane for network operators to signal reachability

How to arrange a maintenance

From: Will Hargrave <will@lonap.net>
Subject: [LONAP Members] LONAP Maintenance, Telehouse North - 20170123 2300Z -> 20170124 0300Z
Date: 16 January 2017 at 17:27
To: Members of LONAP <members@lonap.net>

LONAP Scheduled Maintenance - Telehouse North

From: Mon 20170123 2300Z (UK Time)
To: Tue 20170124 0300Z (UK Time)

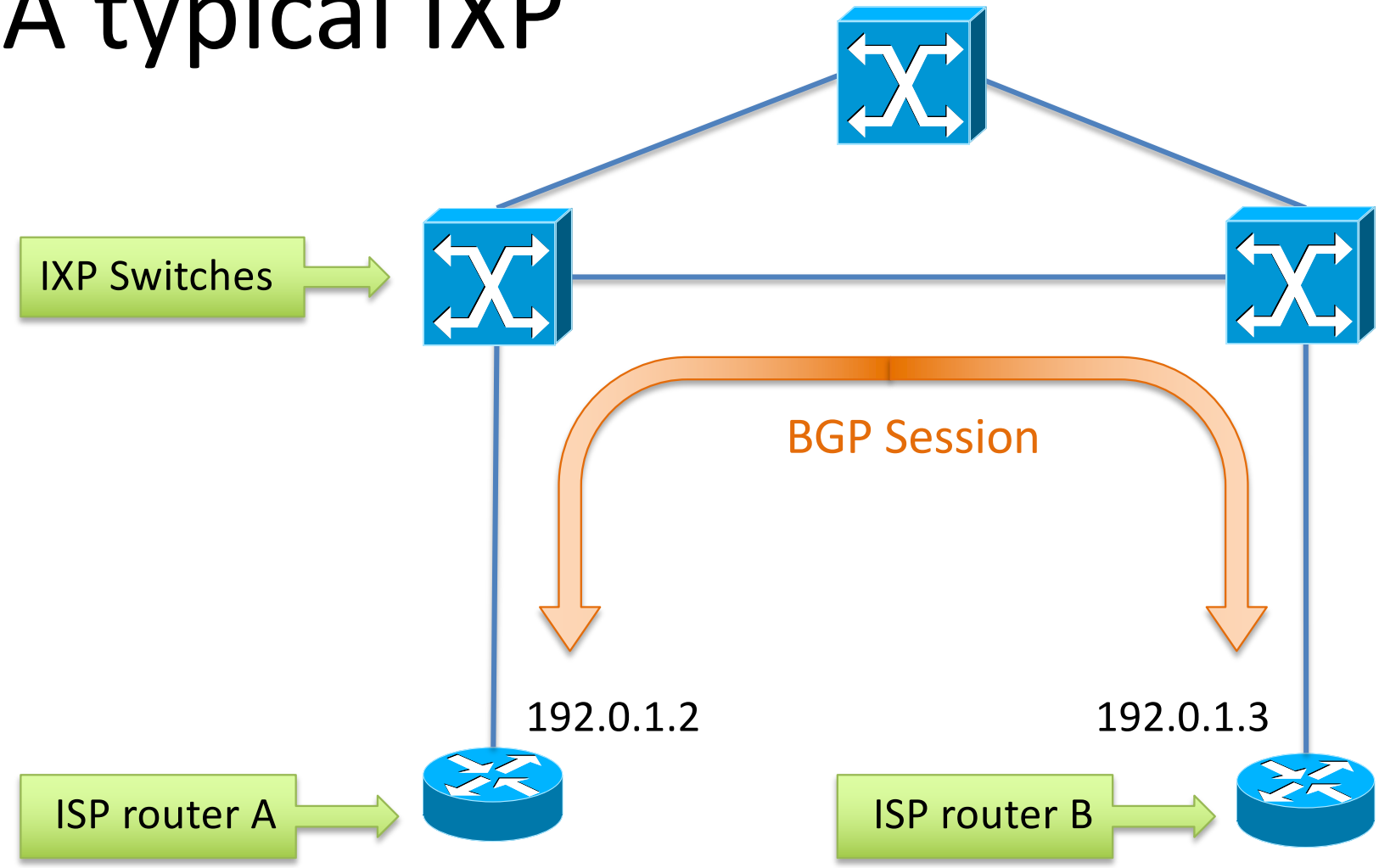
During this period LONAP will carry out maintenance as follows:

- Ideally members would shut their BGP sessions ☺
- But who actually bothers? ☹

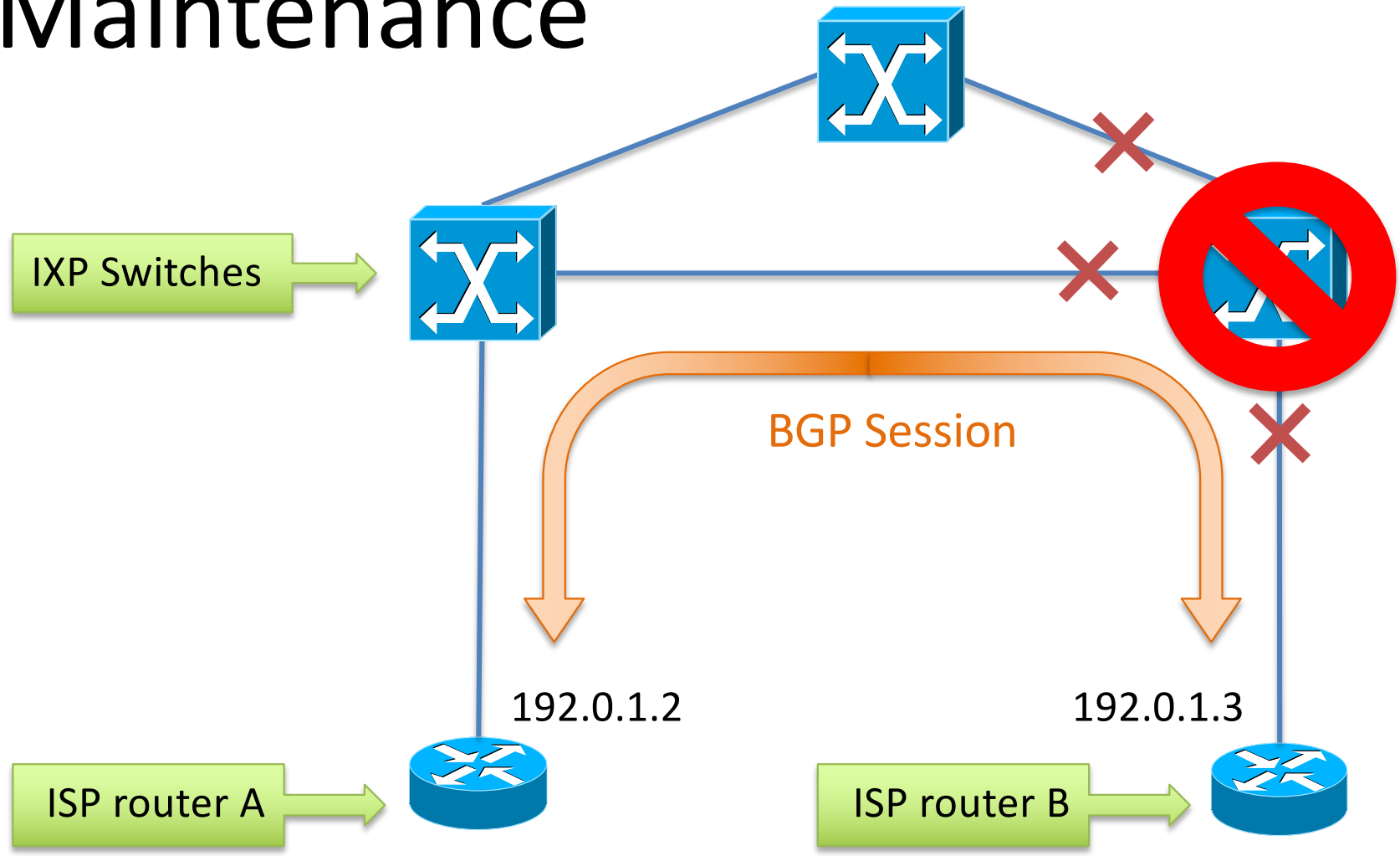
IXP Maintenance

- IXP operators need to carry out works:
  - reboot switch
  - repatch the ODF
  - replace equipment
- What happens to production traffic when an IXP operator does maintenance?

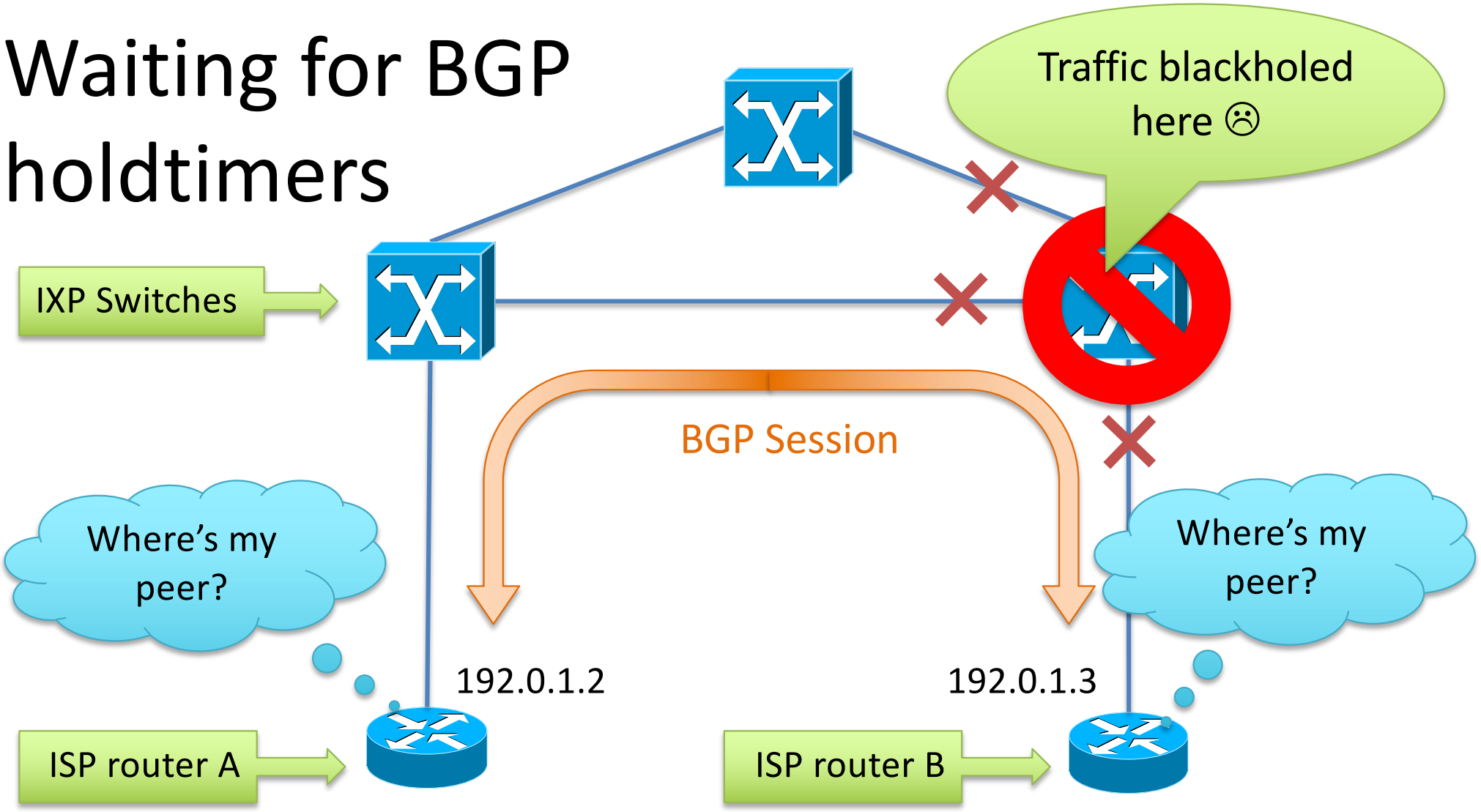
A typical IXP



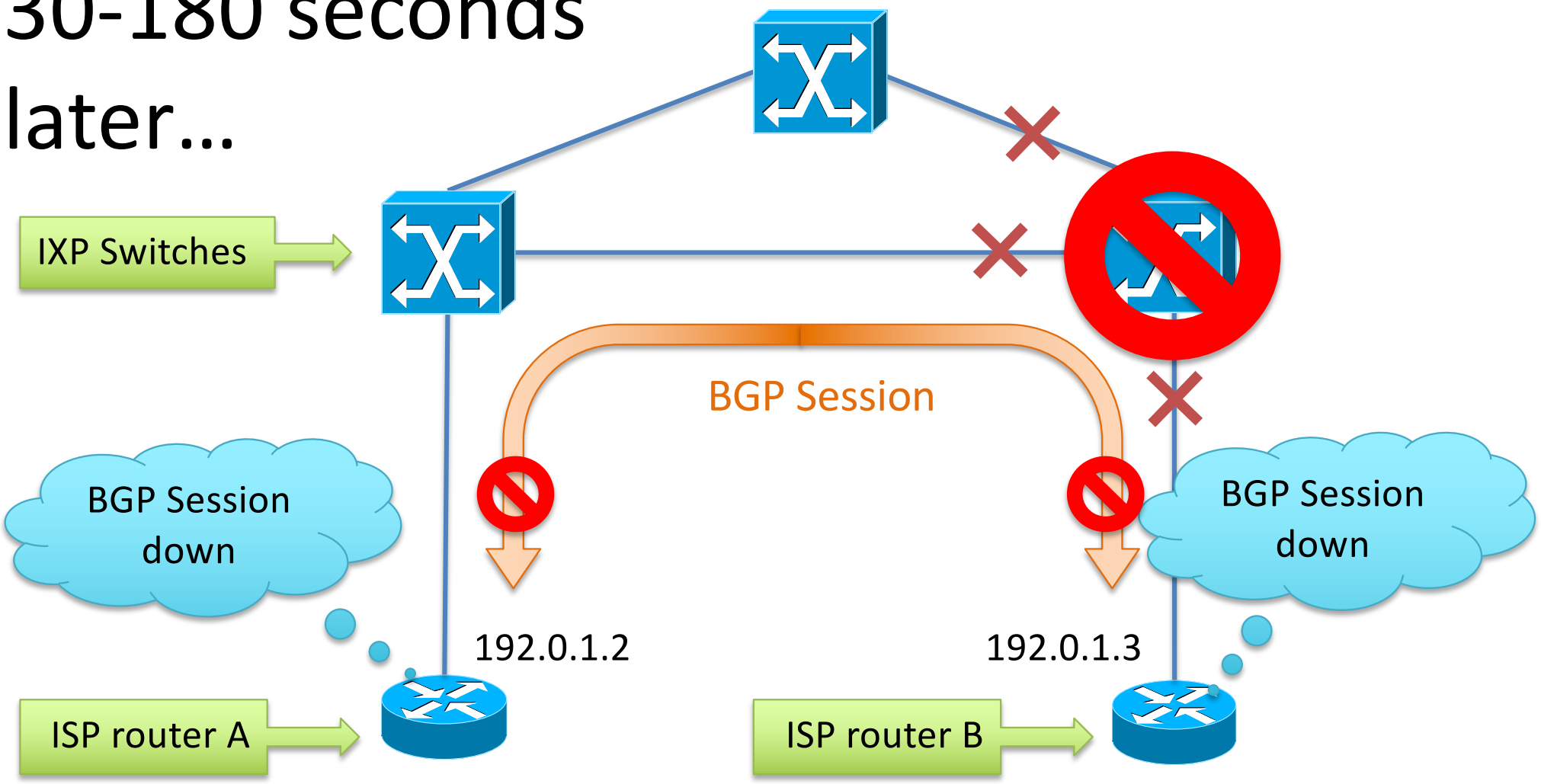
Maintenance



Waiting for BGP holdtimers



30-180 seconds later…



Improving the experience

In many cases, IXP switch maintenance causes 90+ seconds blackholing of production traffic **– making a terrible internet** ☹

Solution:

1. At the start of the maintenance window, cull the BGP sessions
2. Wait for traffic to diminish (3-5 minutes)
3. Now do your maintenance

How do we cull BGP sessions when we don't control the endpoints?

Answer: L4 ACLs on IXP port!

L4 BGP ACLs on IXP

```
ipv6 access-list acl-ipv6-permit-all-except-bgp
    10 deny tcp 2001:db8:2::/64 eq bgp 2001:db8:2::/64
    20 deny tcp 2001:db8:2::/64 2001:db8:2::/64 eq bgp
    30 permit ipv6 any any
!
ip access-list acl-ipv4-permit-all-except-bgp
    10 deny tcp 192.0.2.0/24 eq bgp 192.0.2.0/24
    20 deny tcp 192.0.2.0/24 192.0.2.0/24 eq bgp
    30 permit ip any any
!
interface Ethernet33
    description IXP Participant Affected by Maintenance
    ip access-group acl-ipv4-permit-all-except-bgp in
    ipv6 access-group acl-ipv6-permit-all-except-bgp in
!
```

Your IXP subnet

Block in both directions, otherwise sessions will re-establish

Obviously repeat this for IPv4 too

Results

- First tested this during several LONAP maintenances in 2013 with good success.
  - Subsequently used at INEX☘, SIX, STHIX, TORIX, NETNOD
- We now have 4+ years of good experience
- Traffic removal in < 3mins
- Restore traffic when you are ready
  - even after multiple reboots – think microcode upgrades, mistakes

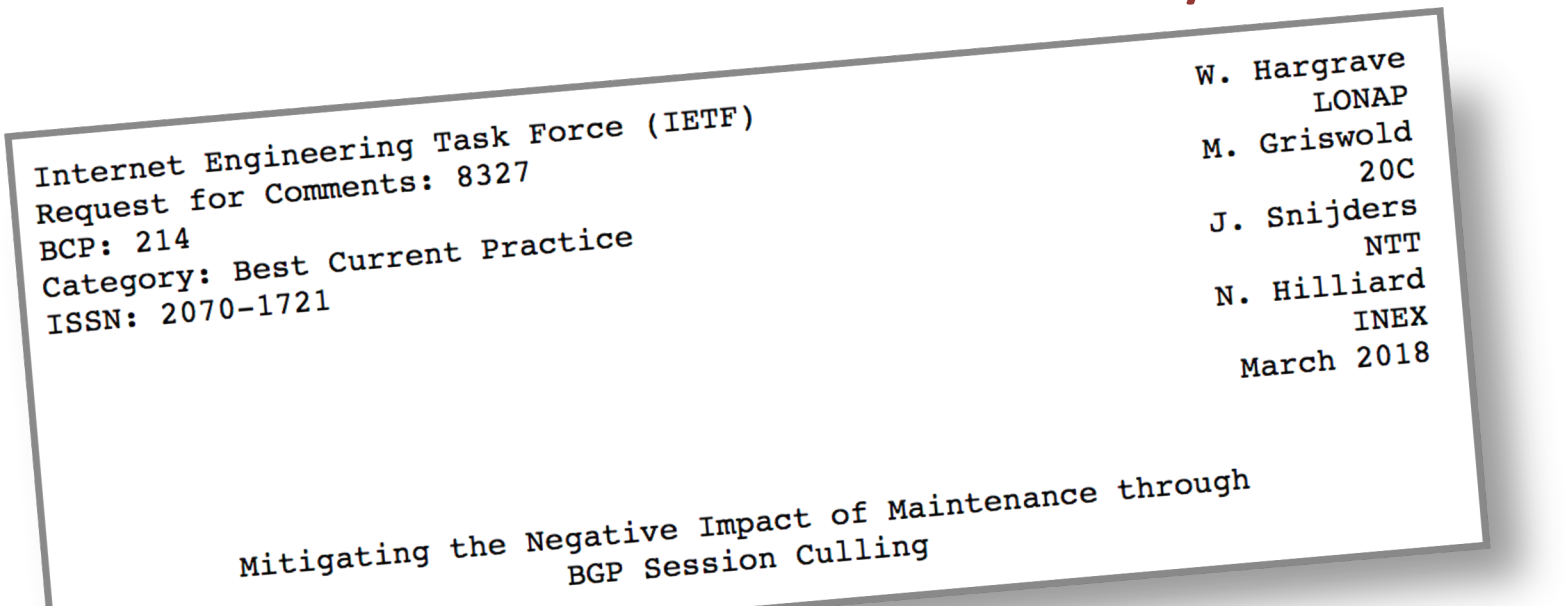
Further Experience

- The ACL works in both directions
  - it doesn't matter if you apply on ingress or egress
- You must have sufficient TCAM space
  - Conflicts with existing ACLs
  - Conflicts between v4 and v6
  - Consider ASIC port groups
- On Broadcom Tomahawk boxes, we ended up deploying IPv4 ACL on ingress and IPv6 ACL on egress

Time to write an RFC

"Mitigating Negative Impact of Maintenance through BGP Session Culling"

March 2018: Published as RFC8327 / BCP214



Internet Engineering Task Force (IETF)
Request for Comments: 8327
BCP: 214
Category: Best Current Practice
ISSN: 2070-1721

W. Hargrave
LONAP
M. Griswold
20C
J. Snijders
NTT
N. Hilliard
INEX
March 2018

Mitigating the Negative Impact of Maintenance through
BGP Session Culling

Summary of RFC8327 / BCP214

“Mitigating Negative Impact of Maintenance through BGP Session Culling”

1. Ideally BGP operators shut their own BGP sessions
2. IXP operator should use BGP Session Culling where the operator is too lazy

End

- This benign technique makes a better internet
- Ask **your** IXP operator about BGP Session Culling today!

Thanks to RFC co-authors Job Snijders, Matt Griswold, Nick Hilliard

Questions? Comments?

Will Hargrave
will@lonap.net