ECMP considerations for BGP

Petr Lapukhov,
Facebook,
petr@fb.com
ECMP never standardized

• General rule of thumb
• Stop tie-breaking after IGP metric
• Best-path still unique
• Documented by vendors
• No standard reference
Number of ECMP paths

• Vendor/platform specific
• Different for iBGP / eBGP
• Configuration knob
• How to report (e.g. for BMP)
eBGP multipath

- Same AS_PATH
- AS_PATH relax
- Same Peer AS
- Next-hop cost
eBGP + iBGP multipath

- eBGP > iBGP
- However ECMP could apply
- MPLS/other tunnels needed
- AS_PATH constraints?
Ignoring IGP Cost

• Blasphemy!
• ...Possible with tunneling
• Configuration Only?
• Per next-hop?
iBGP: ECMP without IGP

• IGP cost was used for loop prevention
• Need to compare CLUSTER_LIST length
• Often ignored (routing loops!)
• Same for BGP Confederations
DMZ Link-Bandwidth

• Weighted ECMP
• Currently: iBGP only (per prefix)
• eBGP case?
• How to implement? (RFC 2992)
• Next-hop DMZ link bandwidth?
Other stuff

- AIGP and ECMP
- iBGP + Confed
- Next-hop with IGP cost vs no-IGP cost
- Non-deterministic tie-breaking
- Interop effects
Proposal/Question

- Adopt Informational document
- Provide common guidelines
- Document caveats
Questions