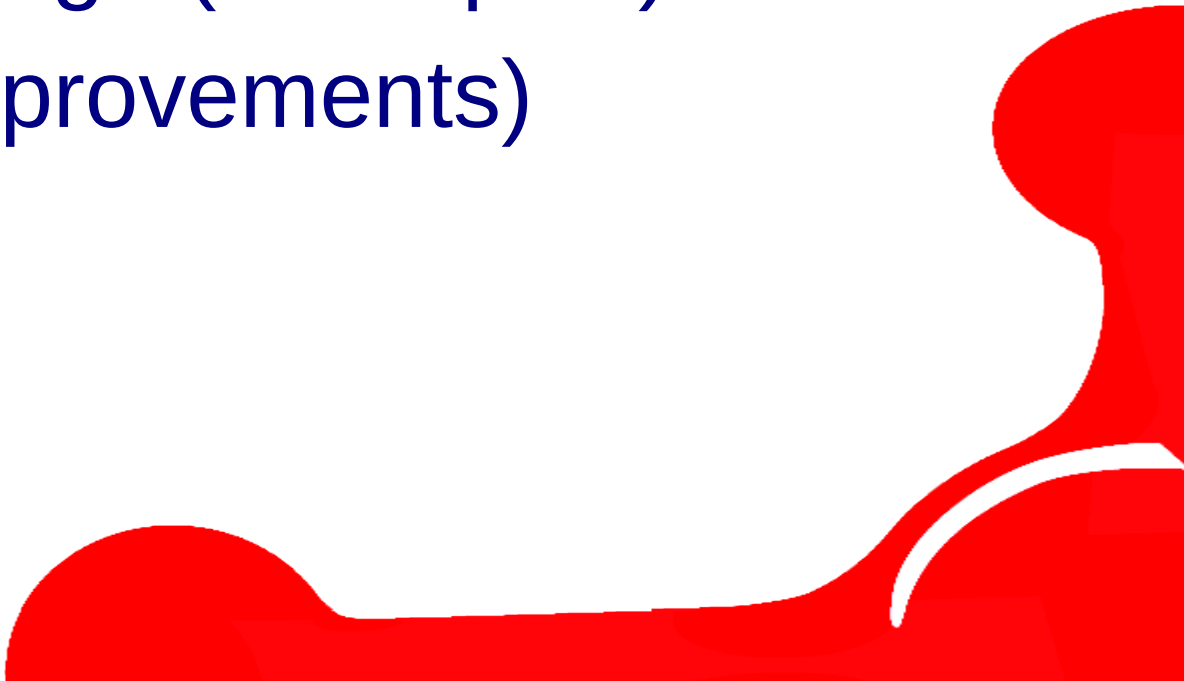
Anycast Routing Challenges

Aleksi Suhonen

European Peering Forum
Vienna 2024

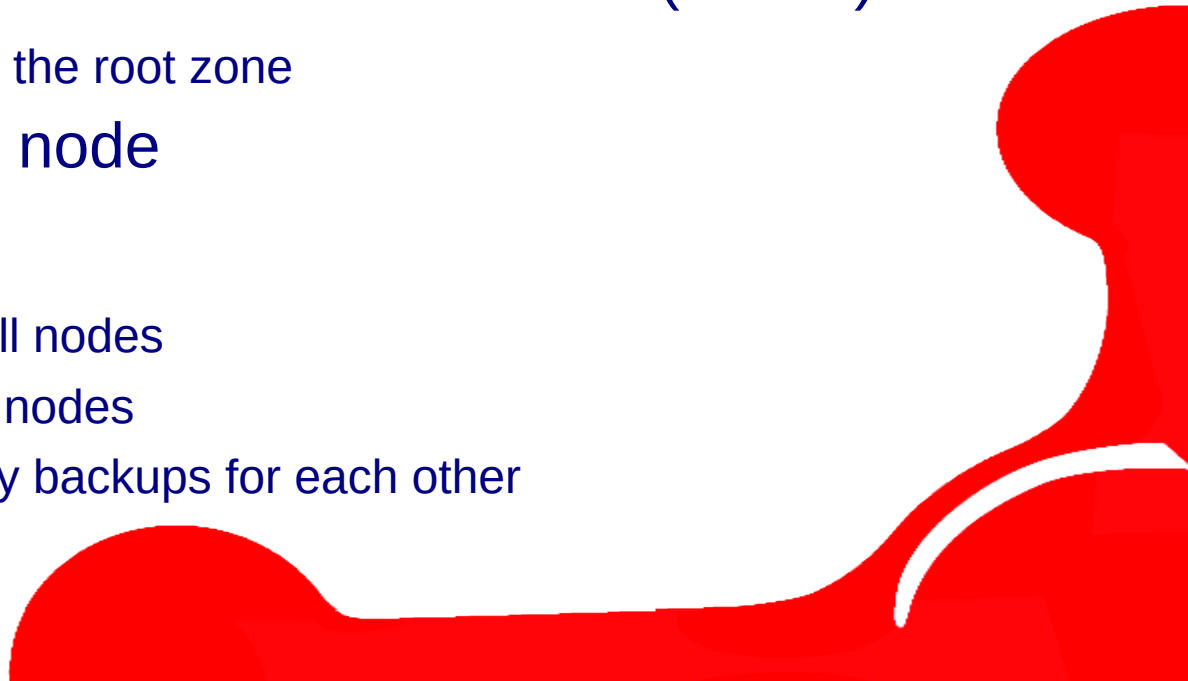
What's This All About?

- What is anycast routing?
- What's going wrong? (Examples)
- How to fix it? (Improvements)



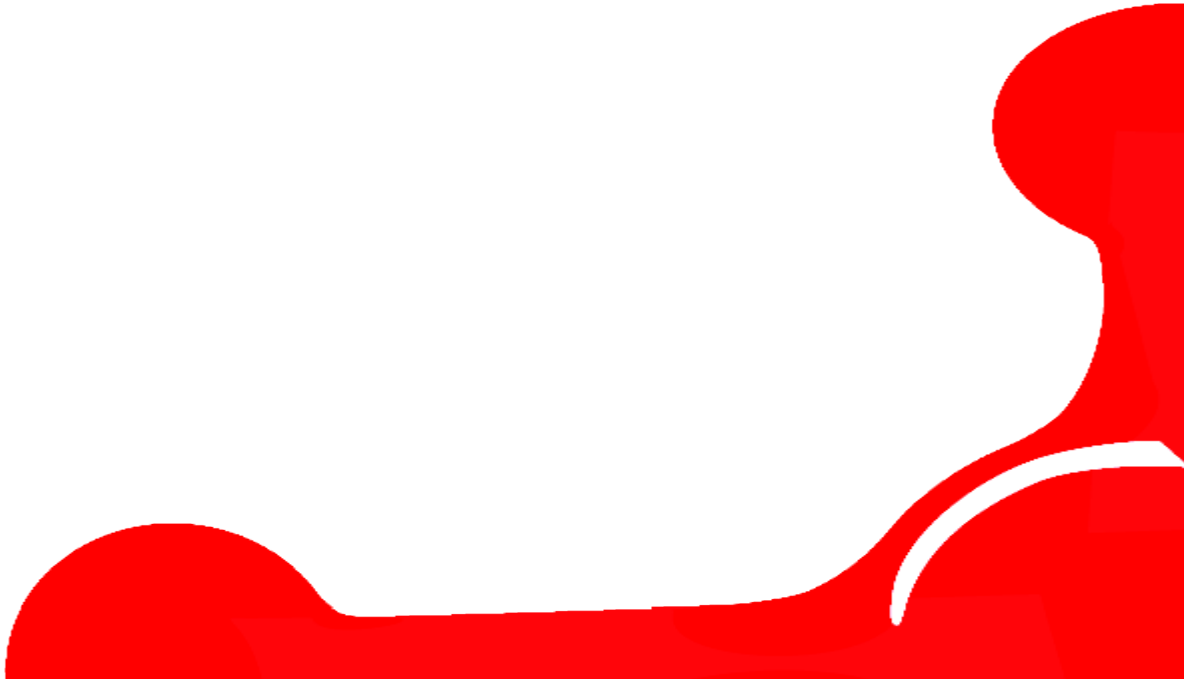
Anycast Routing in a Nutshell

- The same prefix is announced in lots of different locations, usually disconnected
- The same service is cloned at all these locations (nodes)
  - Typically DNS for TLDs or the root zone
- Traffic flows to the closest node
- Benefits include:
  - Load is shared between all nodes
  - DDoS doesn't take out all nodes
  - The nodes are hot standby backups for each other



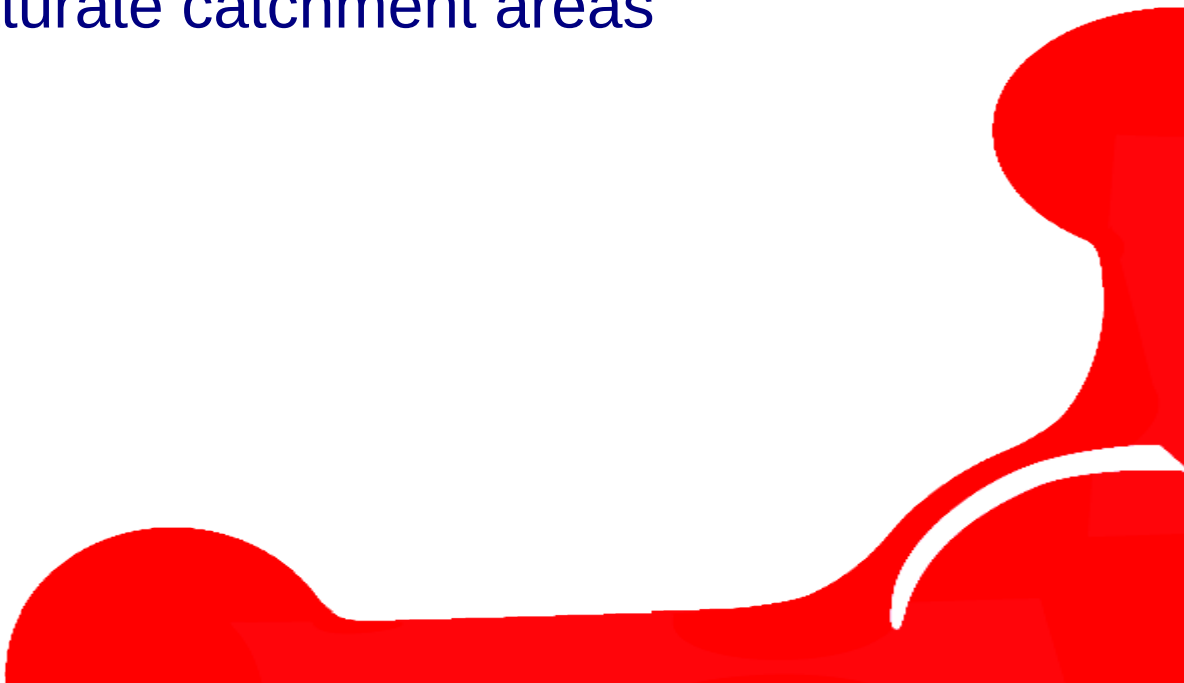
The Anycast Challenge

- Trouble is traffic *doesn't always* flow to the closest node



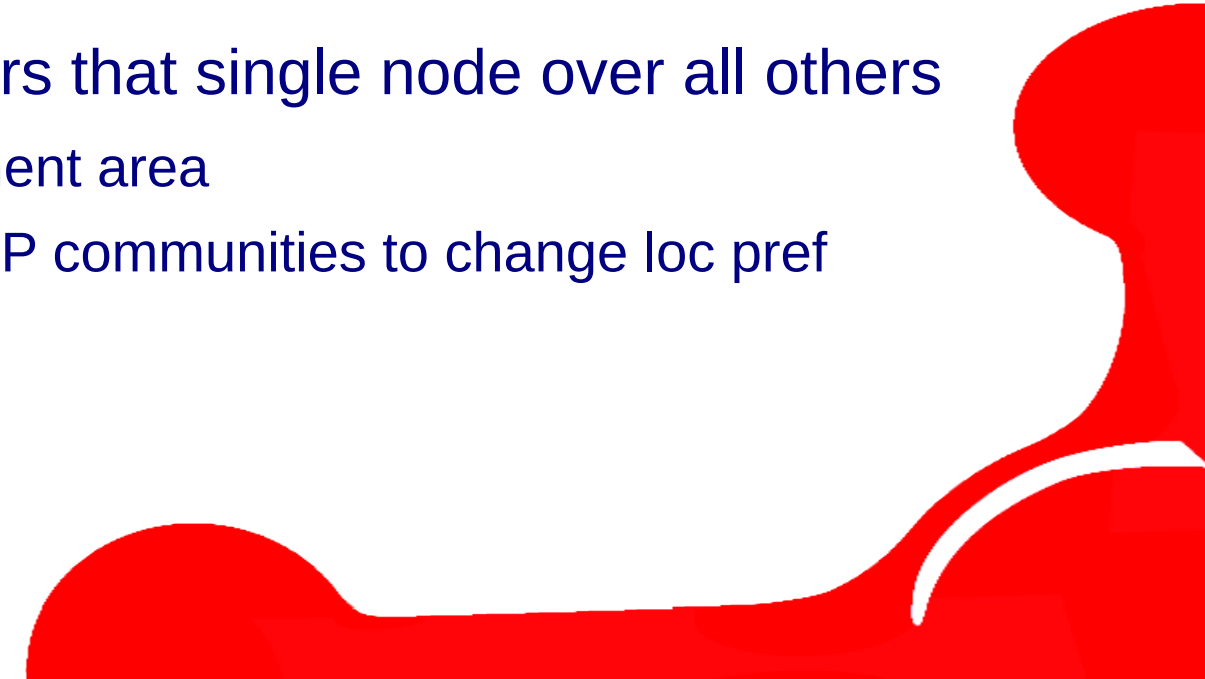
Example: Nodes Behind IXP Routers

- ICANN L-root nodes are connected directly to many older IXPs
- All their recent deployments are behind IXP's ASNs
- The older deployments saturate catchment areas



Example: Anycast Nodes with Transit

- Some nodes will have transit
  - Intentionally or
  - Unintentionally
- The transit provider prefers that single node over all others
  - Makes for a large catchment area
  - Some providers have BGP communities to change loc pref

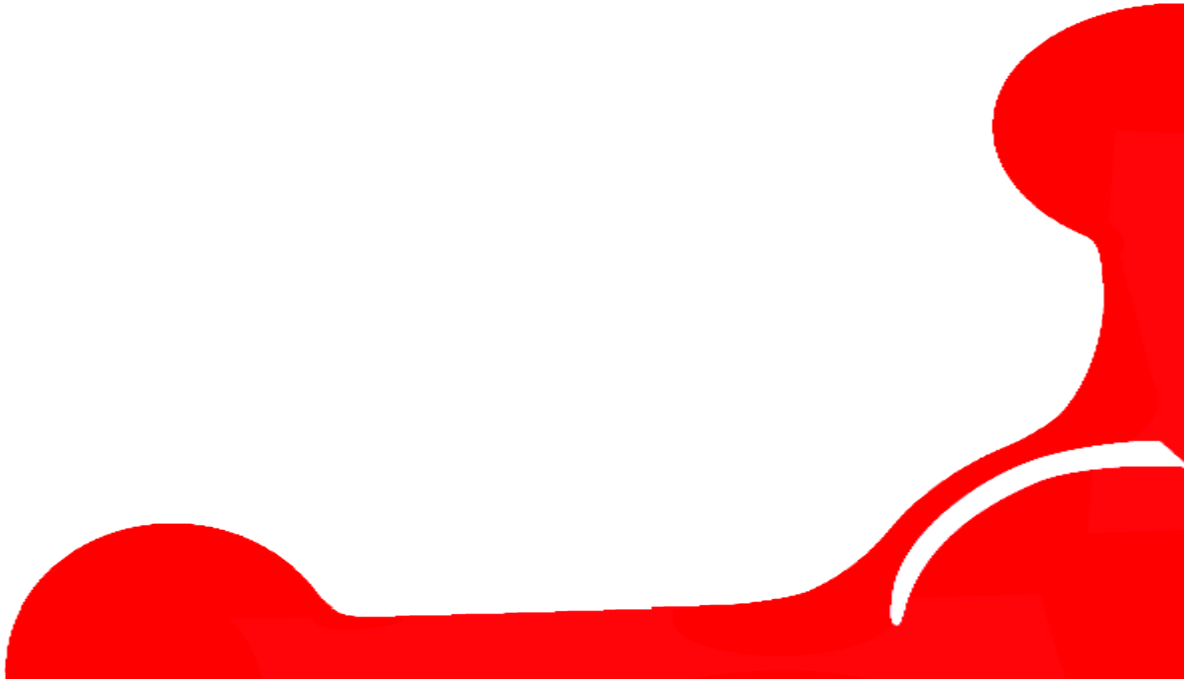


Example: 6to4 and Teredo

- 6to4 is the anycast experiment with the poorest track record
  - Any source ASN
  - Asymmetric routing over several oceans
  - Direct assignment from IETF/IANA --> RPKI not possible
- Teredo fixes asymmetric routing
  - Still best effort volunteer based announcements
  - No coordination between nodes

Improvement: Try to Balance AS Paths

- One approach that nic.at uses is to send routes with up to 5 prepends to everybody
- (nic.at slide 35)



Prepending is needed

- To control traffic
  - by making the shortest AS path as long as the longest
    - on all our anycast locations
    - to peers
    - global transit providers
- To consider special routing situations
  - DDoS mitigation provider is activated
    - Traffic should be routed via DDoS mitigation provider, not directly to us
    - Extend AS path even one more time
- So we ended up having an AS path length of 5 towards IX/Peers and Tier1 transit providers.

Routing challenge (real life)

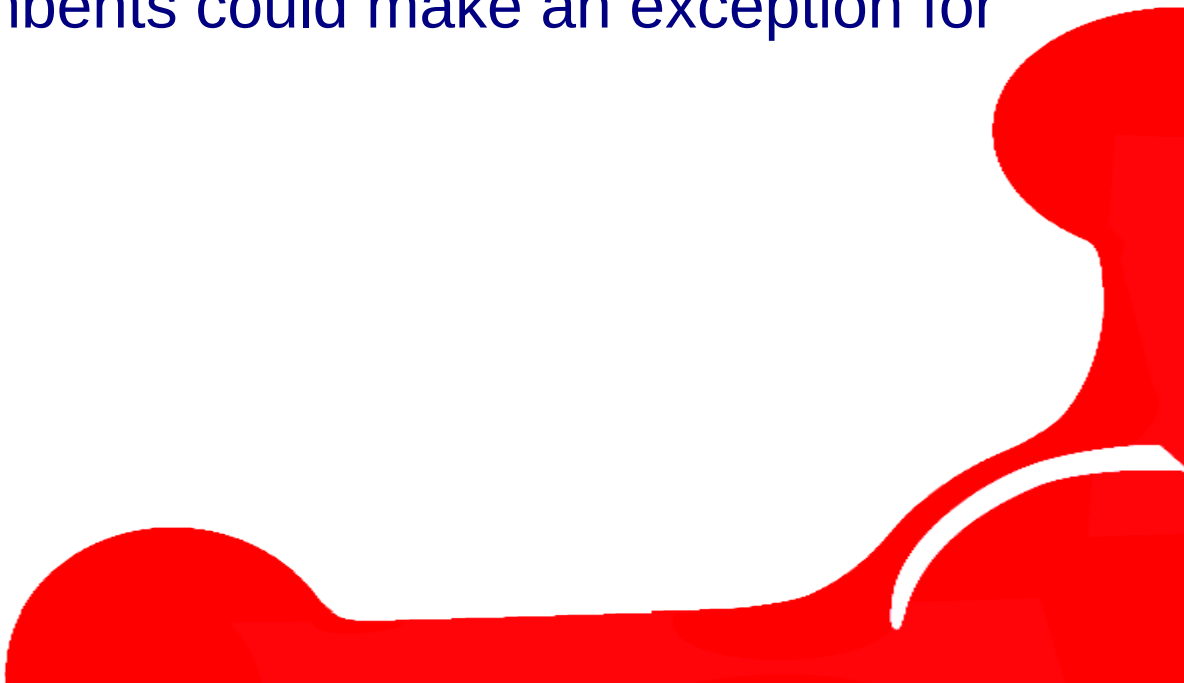
- An incumbent is the local key player. At home they peer with nobody - they want to sell transit.

- Somewhere else they are a small players and (needs to) peer with everybody.

-> Traffic is going round the world instead of going to a node close by.

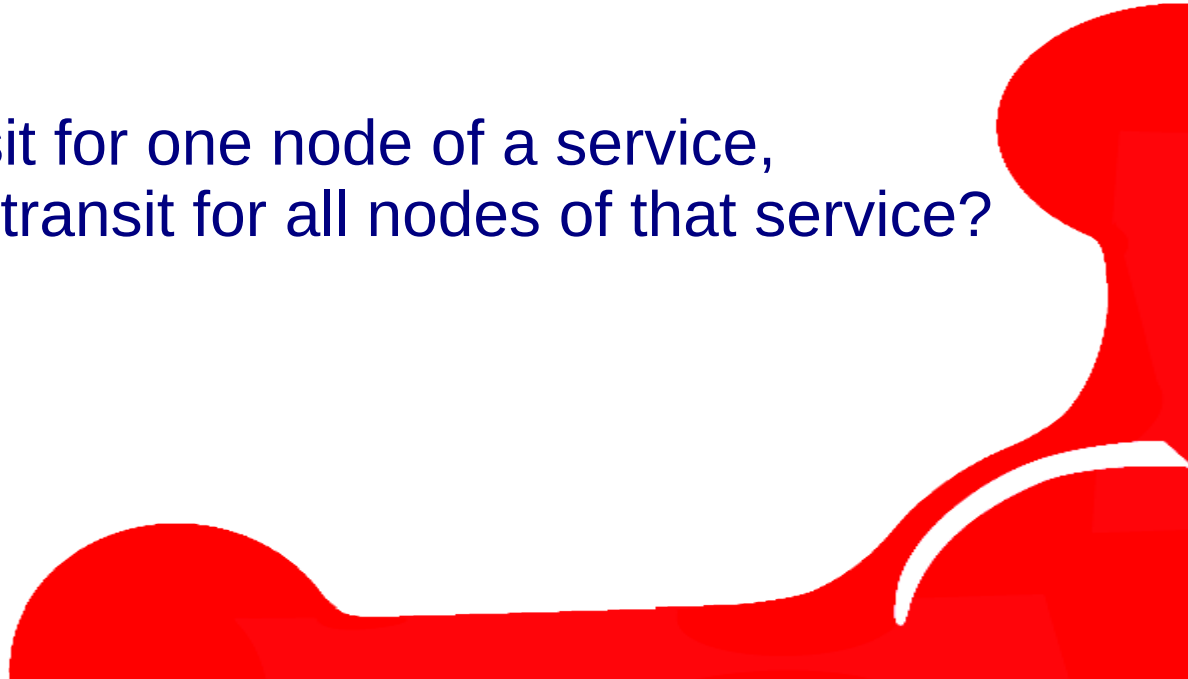
Example: Incumbents Don't Peer At Home

- But they do peer elsewhere
- Anycast traffic doesn't stay local
- Fix could be simple: Incumbents could make an exception for anycast services...



Improvement: Better Coordination

- Some anycast services don’t have a backing organisation
- There is a lack of peering/transit coordination
  - 6to4 and Teredo
  - AS112
- If someone provides transit for one node of a service, perhaps it should provide transit for all nodes of that service?



Improvement: BGP Well Known Community

- This is just an idea
- Mark anycast routes with a community that changes best path selection **locally**:
  - EBGP > IBGP before everything else
  - But only if all alternatives have the community
- Could possibly be a new optional transitive attribute instead

Thank you!

Any(cast) questions?