BGP and R&E Networks

Brenna Meade

International Networks at Indiana University

EPOC

EPOC
Engagement and Performance Operations Center

Agenda

- Introduction of EPOC
- BGP Overview
- R&E routing
  - What it is
  - Why it matters
- BGP Steering mechanisms and real world examples
  - Examples
  - Localpref
  - AS Path Padding
  - Communities
  - Multi Exit Discriminators (MEDs)
- Questions / Discussion

EPOC
Engagement and Performance Operations Center

Engagement and Performance Operations Center (EPOC)

- Joint project between Indiana University and ESnet
  - PI Dr. Jennifer Schopf, co-PI Jent (IU GlobalNOC) and Zurawski (ESnet)
- Partnerships with regional, infrastructure, and science communities that span the NSF and DOE continuum of funding.
- 5 Focus Areas: Roadside Assistance and Consultation, Application Deep Dives, Network Analysis (Netsage), Services “in a box” (DMZ, perfSONAR, etc), Training

EPOC
Engagement and Performance Operations Center

What is BGP

- BGP or Border Gateway Protocol is protocol used between routers to exchange routing information and reachability information between or inside AS on the Internet.
- BGP makes the Internet work, and in most cases it just works
  - Needs to be tuned for best performance
- BGP makes routing decisions based on paths, network policies and rule-sets, etc.

EPOC
Engagement and Performance Operations Center

AS number

- AS stands for Autonomous System - used to identify networks / systems
- A collection of prefixes (routes) under the control of one or more network operators under one administrative entity with similar/common routing policies
- Used to be a 2-byte number (1 to ~65k)
- Some reserved for private use
- Ran out of numbers, increased to 4-bytes
- Currently over 100k in use. TransPAC is AS22388

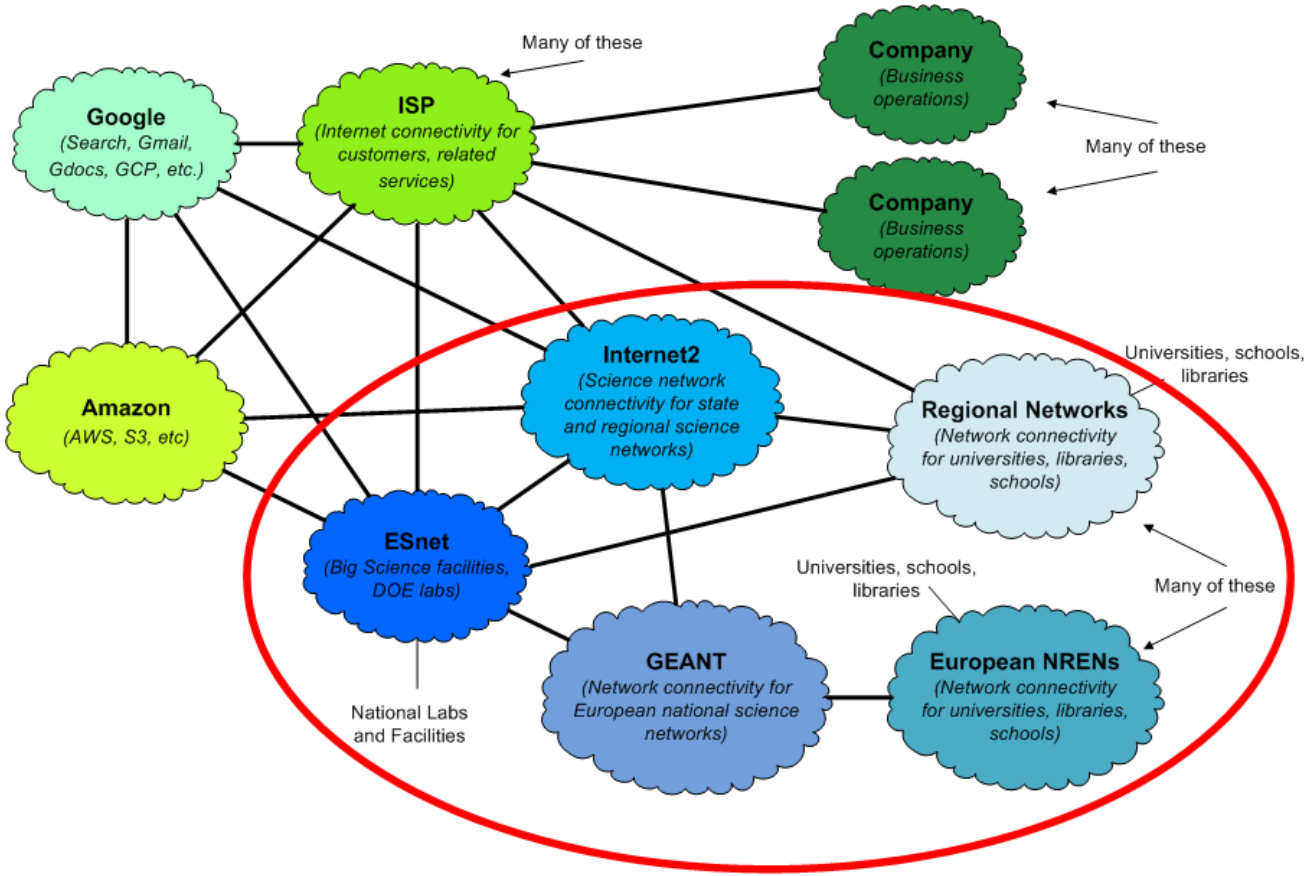
EPOC
Engagement and Performance Operations Center

R&E Routing Architecture

- R&E networks engineered for science traffic
- Keep R&E traffic on R&E paths if possible
  - Bandwidth
  - Performance Engineering
  - Deterministic behavior
- We all have to do our part
  - All routing decisions made locally
  - Emergent behavior is important

EPOC
Engagement and Performance Operations Center

R&E vs. Commodity



EPOC
Engagement and Performance Operations Center

So what do we do?

- High level : We need to use BGP policy to keep R&E traffic on R&E networks
  - Announcements attract traffic
  - Routing determines the path the traffic takes through the network - BGP gives us the tools
- BGP is a path vector protocol
  - For a given prefix, the shorter AS path is preferred
  - If AS path length is the same, then other criteria are used, in order (“BGP path selection algorithm”)
- Override BGP’s use of AS path length when choosing between R&E and commodity paths
  - R&E path will be longer in the general case (more organizations involved)
  - Use normal BGP route selection between R&E routes, and between commodity routes
  - Remember - hop count is a legacy metric

EPOC
Engagement and Performance Operations Center

Why does this matter? Example - OSC <> NERSC

- Data transfers between Ohio Supercomputer Center and NERSC were slow (below 60mbps)
- Out of date bgp policy was causing traffic to pass over commodity instead of R&E paths.
  - Updating the policy improved performance
- Commodity networks often throttle high-speed flows
  - In their world a traffic spike means a DoS attack
  - In our world it’s another scientist doing their normal thing

BGP Management

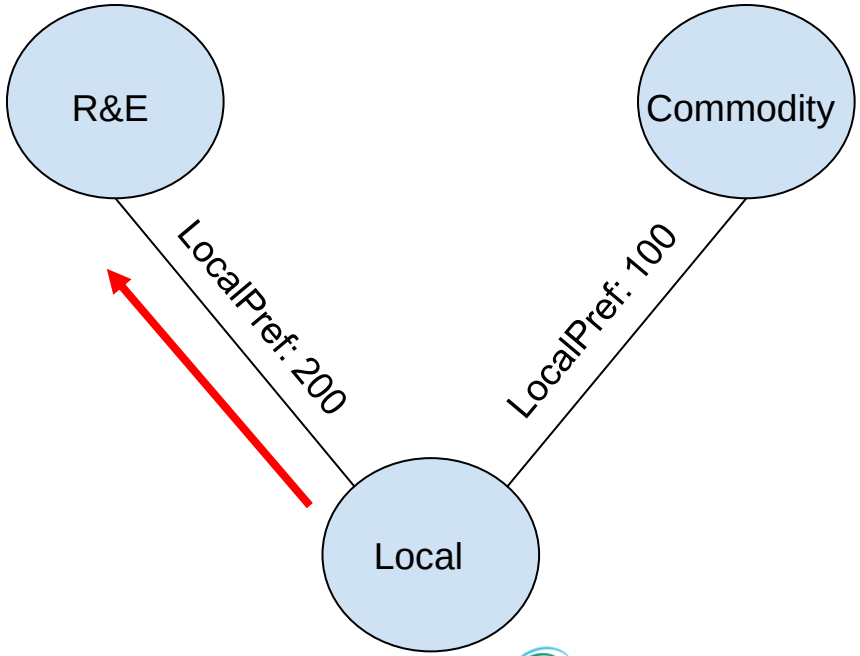
- BGP just works but needs tuned for best performance
- Best path selection is a 10+ step process!
- Common steering mechanisms
  - Localpref
  - Communities
  - AS Padding
  - MEDs

Next hop reachable?	continue if "yes"
Local Preference	higher wins
AS path	shorter wins
Origin Type	IGP over EGP over incomplete
MED	lower wins
eBGP, iBGP	eBGP wins
Network exit	nearest wins
Age of route	older wins
Router ID	lower wins
Neighbor IP	lower wins

EPOC
Engagement and Performance Operations Center

LocalPref

- Per prefix
- Modifies path for outbound traffic
- Higher preferred



EPOC
Engagement and Performance Operations Center

BGP Community Strings

- A community string is a number value that the peer uses like a tag.
- Tagging prefixes with communities tells the peer to handle the prefixes in a special way.
- Can make changes to routing policy based on per prefix strings
- Prefixes can have multiple community strings
- Can provide useful information about the prefix
- Communities that might be useful to external networks should be made public
  - Provides a mechanism for peers to affect a network's internal behavior
  - Common uses: change local preference, DDoS mitigation
- Look for upstream networks published communities
  - Regional?
  - National?

EPOC
Engagement and Performance Operations Center

Public BGP Community Strings offered by Internet2

- Set LocalPref on your advertised prefixes
  - Default - 100
  - 11537:40 - Low
  - 11537:160 - High
- Prefix identification?
  - 11537:5004 - Amazon
- Where does the prefix enter the network?
  - 11537:242 New York
- Emergency!
  - 11537:911 - Discard all traffic destined to these prefixes!
- AS Path Padding?
  - 65001:65000 - prepend x1

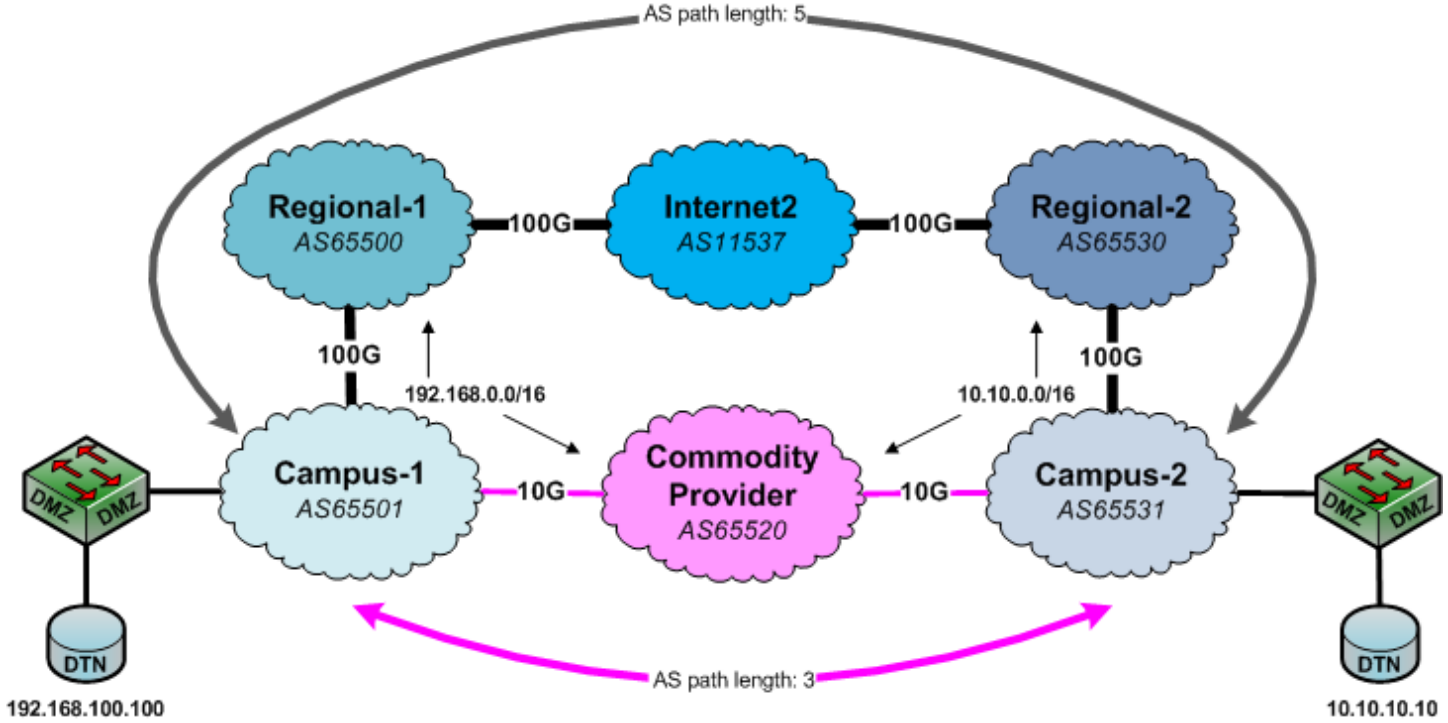
For more information about community strings offered by Internet2

- https://noc.net.internet2.edu/i2network/maps-documentation/documentation/bgp-communities.html

EPOC
Engagement and Performance Operations Center

BGP AS Path Length Illustrated

- Hop count is a legacy metric!



EPOC
Engagement and Performance Operations Center

AS Path Padding

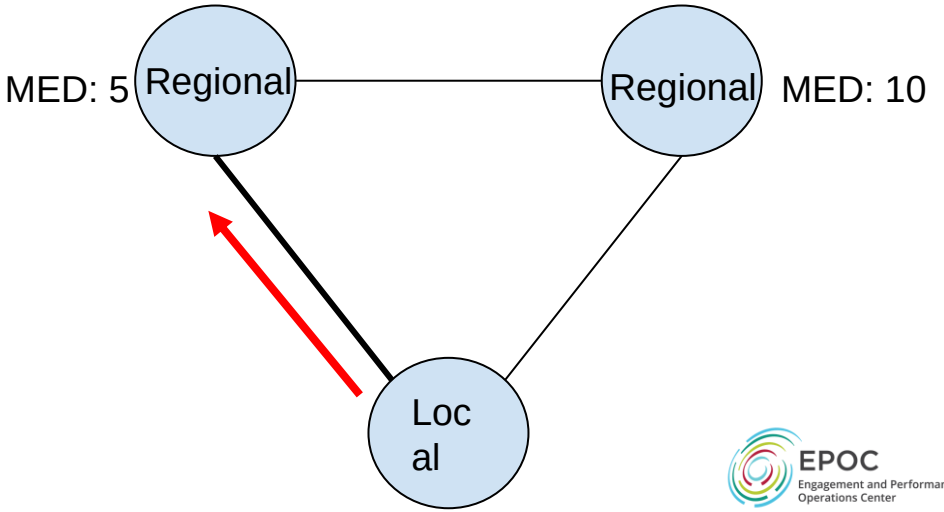
- BGP will choose shortest AS Path
- Add one or more copies of your AS# to prefixes advertised to specific neighbors.

* 180.208.59.0/24 202.112.61.57 - - - 4538 4538 24364 **133465 133465 133465** 65300 i

EPOC
Engagement and Performance Operations Center

Multi Exit Discriminator (MED)

- Useful when you have N+1 connections to a network
- Indication to external peers of the preferred path into network
- Lowest number preferred



EPOC
Engagement and Performance Operations Center

Takeaways!

Routing will not take care of itself

- Old routes may not work well with new networks
- New routes may not work as planned

How do we address routing, and performance as a community?

The Routing Working Group!

EPOC
Engagement and Performance Operations Center

Routing Working Group - What are the goals?

- **Engineering focus**
  - Document possible erroneous routes
  - Identify teams to address them
  - Check in together as we work through them
- **Policy Focus**
  - Detail routing policies for paths
    - Including preferred backup paths!
  - Verify if policy is being followed

EPOC
Engagement and Performance Operations Center

Routing Working Group

- Asymmetrical routing - meaning a source to a destination takes one path and takes a different path when it returns to the source
- R&E data takes a less efficient route around the world - affecting performance
  - Europe to Asia routes traversing the US
  - Africa to Europe routes traversing the US
- R&E data takes a commodity route when an R&E path is available
- New R&E links are removed or added but routing does not adjust appropriately

EPOC
Engagement and Performance Operations Center

Goals for 2022

- Introduce more cases
- Continue to provide tool talks
  - perfSONAR
  - RIPE atlas
- Create a series of community driven BGP best practice documents
  - Best practices for R&E network providers specifically and how do they differ from other networks?
  - Scalability
  - Load balancing
  - R&E Peering Agreement best practices
  - Communities - how to use this with a focus on R&E networking

EPOC
Engagement and Performance Operations Center

Submit your cases!

Email the Chairs!

meadeb@iu.edu

addlema@iu.edu

warrick.mitchell@aarnet.edu.au

Join the routing working group!

Mailing list routing-wg@gna-g.net

- Contact Brenna to be added meadeb@iu.edu

Slack

- APAN Slack Instance, Channel: Routing

Web

- https://www.gna-g.net/join-working-group/gna-g-routing-wg/

Contact any of the co-chairs for more information!