COSC 417: Topics in Networking

TOPIC 3: BGP ROUTING

SCHEDULE

1. *Recap*
2. *Path-vector Routing Protocol*
3. *BGP Messaging*
4. *BGP Path Attributes*
5. *eBGP and iBGP*

RECAP – AUTONOMOUS SYSTEMS

- Last week we talked about *autonomous systems*
- An autonomous system is a collection of routers under common administrative control
- Each AS will have routing policies that are under the control of the administrator

RECAP – INTER-AS ROUTING

- Autonomous systems are connected to each other via inter-as routing (*Border Gateway Protocol*)
- BGP allows autonomous systems to share routing information, which in turn allows traffic to flow between autonomous systems

TODAY'S LECTURE

- We're going to dig a bit deeper into how and why the BGP routing protocol works
- Routing tables – why they are important, and how they work

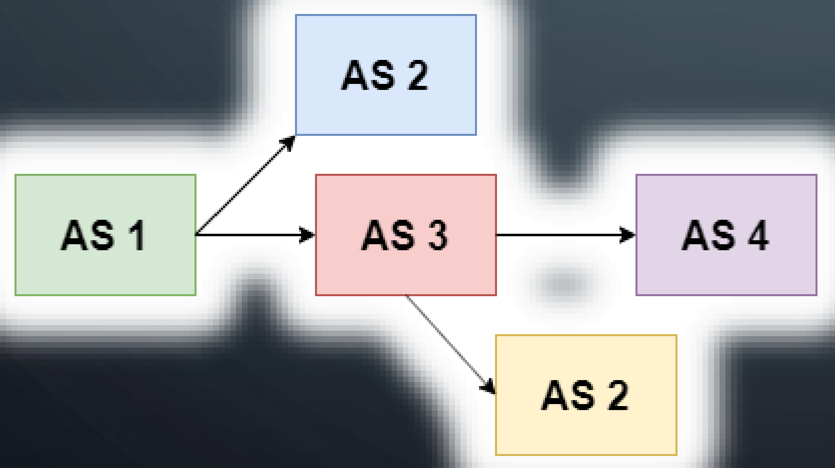
PATH-VECTOR ROUTING PROTOCOL AND THE INTERNET

PATH-VECTOR ROUTING PROTOCOL

- At it's core, the BGP protocol is a type of *path-vector routing protocol*
- Path-vector protocols define a route as the collection of autonomous systems (or any points in a graph, to generalize) that must be travelled through from source to destination

PATH-VECTOR ROUTING PROTOCOL

- Remember the AS-Path variable we looked at previously? This is the path-vector routing protocol in action.



- Traffic between AS1 and AS4 could have an AS-Path like this:

1, 3, 4

WHY IS PATH-VECTOR USED?

- As we discussed in the previous lecture, the primary purpose of the AS-Path is to avoid loops in routing
- If a router sees it’s own AS Number in a route, it won’t use that route, because that implies a loop

NON-PATH VECTOR ROUTING (EGP)

- Interestingly, before BGP became the standard, another routing protocol known as Exterior Gateway Protocol (EGP) existed
- EGP is *not* based on a path-vector algorithm
- As a result, EGP struggles to prevent loops in it’s routes

SO HOW DID THIS COME TO BE?

- Originally, the internet was envisioned as having a more tree-like structure
  - Minimal loops, and with connections always aggregating towards larger and larger "backbone" networks
- Ultimately, the internet developed a much more complicated network structure – full of loops and redundant connections
  - Hence why BGP overtook EGP as the standard routing protocol

IN THE PRESENT DAY

- Currently, BGP is considered *the* standard for doing inter-AS routing
- Other algorithms (like EGP) exist, but if you want to actually connect your AS to the outside world, you're going to need to use BGP
- Just one of many examples of how increasingly complex networking conditions have driven the need for robust, standardized protocols

BGP MESSAGING

THE GENERAL BGP FLOW

- The BGP protocol basically performs the following actions:
  1. Listen for paths from internal (intra-AS) and external (inter-as) BGP sources (neighbouring routers)
  2. Pick the “best” path and put it into the routing table
  3. Advertise that best path to your neighbours

GETTING TO KNOW THE NEIGHBOURS

- When a TCP connection is established between two routers, the first thing sent is an open message (basically, a “hello” between routers)
- They will send to each other:
  - Their own AS Number
  - Their BGP version and router ID
  - The time-out time (heartbeat)

STAYIN' ALIVE

- Assuming the open messages were exchanged successfully, a BGP connection can be established between the two routers
- After the connection has been established, the two routers will regularly send each other *keepalive* messages
- These keepalive messages have no data, and just maintain the BGP connection state

TIMEOUT AND KEEPALIVE

- Something important to note here: the timeout (also known as hold-time) must be longer (usually 3-4x) than the keepalive interval
- If the timeout expires without receiving a keepalive message, it is assumed the neighbour has gone offline
- When sharing timeouts in the opening message, the shorter of the two times (one from each router) is used

UPDATING PATHS

- Once a BGP connection has been established, routers can send update messages to each other
- Each update message updates a single path
- An update message will specify the networks reachable through the path (NLRI), the path attributes, as well as withdrawn (no longer valid) routes

NETWORK LAYER REACHABILITY INFORMATION

- In the BGP Update message, the router shares something known as Network Layer Reachability Information (NLRI)
- NLRI is composed of a length, and a prefix, using CIDR IP notation:
  - 255.255.255.240 /28 (16 addresses)
  - 255.255.255.252 /30 (4 addresses)

UPDATE, IN SUMMARY

- So, each update message describes a path that connects to some set of addresses described in NLRI format, the attributes that apply to these addresses, and any routes to these addresses that are no longer valid
- It is the constant sharing of updates that allows the BGP protocol to modify routing behaviour as network conditions change

WHEN THINGS GO WRONG

- The last major type of BGP message is a notification
- A Notification message is sent when an error has been detected, and contains error codes and information related to the error
- When a notification message is received, the BGP connection is immediately terminated to prevent further errors and/or loss of traffic due to errors

BGP MESSAGES, IN SUMMARY

- There are four major types of message shared between BGP routers:
  - Open – Create a BGP connection
  - Keepalive – Heartbeat monitoring messages
  - Update – Send out path information to your neighbours
  - Notification – Announce errors and close the BGP connection

BGP PATH ATTRIBUTES

PATH ATTRIBUTES

- When a BGP-connected router shares a path to some network (NLRI), it includes certain attributes for that path
- Some of these attributes are mandatory and are part of the BGP specification
- Others are used for specific types of routing policy, or by particular router vendors (such as Cisco)

WELL-KNOWN VS MANDATORY VS TRANSITIVE

- Attributes can be broadly classed into a handful of groups:
  - Well-Known attributes must be recognized by all routers
  - Mandatory attributes must be sent with all BGP updates (or an error will result)
  - Transitive attributes must be passed along when a path is re-advertised to neighbours

WELL-KNOWN MANDATORY ATTRIBUTES

- There are three well-known, mandatory attributes that must be included in each update message
- If these aren't included, a notification message will be sent and the BGP connection will be terminated
- These three attributes are *Origin*, *AS_Path*, and *Next_Hop*

THE ORIGIN ATTRIBUTE

- The *Origin* attribute describes how a path for a given prefix was discovered or “learned” by the router
- Three acceptable values:
  - IGP – path was learnt from another router in the same AS (Internal Gateway Routing Protocol)
  - EGP – Obsolete, no longer used (Exterior Gateway Protocol)
  - Incomplete – path was learnt via indeterminate way (manual entry, etc)

IS ORIGIN IMPORTANT?

- You might wonder if it really matters how we learnt about a given path?
- The answer is that it doesn't, in terms of how the actual path itself works, but it *is* important for choosing a path
- The origin attribute is used to help calculate the best path (IGP is preferable to Incomplete, for example)

THE AS PATH ATTRIBUTE

- The *AS Path* is a mandatory, well-known attribute
- As we've previously discussed, the AS Path contains the AS numbers for each autonomous system in the path
- When a router re-advertises a path, it appends itself to the AS Path
- If a router encounters a path that contains it's own ASN, it ignores that path and will not re-advertise (loop prevention)

AS PATH AS A POLICY DEVICE

- Besides avoiding loops, the AS Path attribute gives us an idea of how long a route is, in terms of autonomous systems that must be traversed
- A longer AS Path indicates the traffic must travel over more autonomous systems
- As a result, AS Path length can also be used as a metric for selecting the best path
- Some networks artificially increase the size of the AS Path to discourage the network being used for transit purposes

THE NEXT HOP ATTRIBUTE

- The last major attribute is the *Next Hop*
- The Next Hop attribute usually contains the IP address of the neighbouring router (in a different AS), that is the next transit point for that path
- When a router sends an update message to another external gateway router via BGP, it updates the next-hop with it’s own IP

WHAT ABOUT THE OTHERS?

- There are many other attributes available, such as local_pref, community, aggregator, etc
- Some of them are transitive (must be passed along if received), others are well-known but optional
- Many serve primarily to express routing policies – to provide weighting or describe preferences for certain paths over others, depending on the routing policies set by the administrator

EBGP AND IBGP

WHAT ARE EBGP AND IBGP?

- Up until now, we've been discussing BGP entirely in the context of inter-AS communication, with routers on the edges of an autonomous system conversing with each other
- This is actually known as *eBGP*, or *external BGP*
- BGP can also be used internally within an AS, in which case it is known as *iBGP*, or *internal BGP*

HOW ARE THEY DIFFERENT?

- In broad strokes, eBGP and iBGP follow the same rules – same attributes, same message types
- In the finer details, there are some differences
  - Differences in what information is shared
  - Differences in how the Next-Hop attribute works

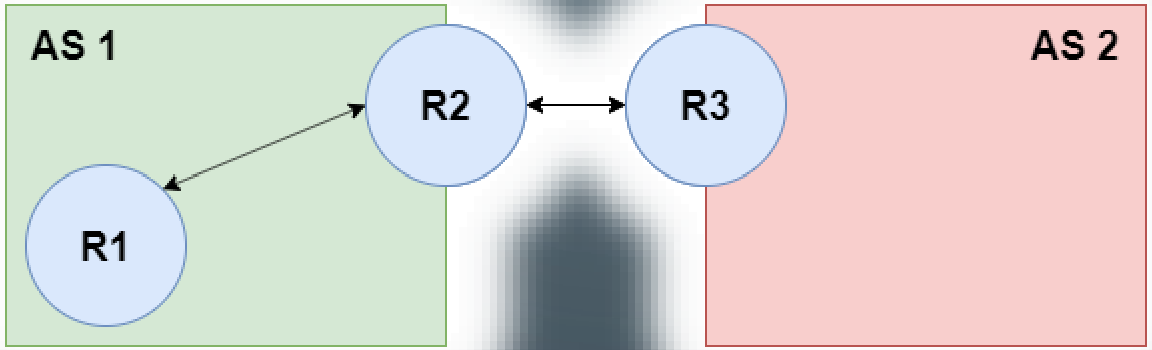
EBGP VS IBGP, AS PATH

- Recall that a router will prepend it's own ASN to the AS Path when it re-advertises a path
- This is *only* done when performing eBGP, i.e. only when a router is a gateway router advertising to a gateway router in a different AS (external connection)
- Internal routing via iBGP does not prepend the ASN!
- If it did, it would be impossible to share routes internally, due to the anti-looping mechanism

EBGP VS IBGP, NEXT HOP

- In eBGP, a router sets the next-hop to it's own IP address (usually) when advertising a path
- This works because the external gateway routers have direct connections to each other, and as such are reachable
- Problem: How can a path be shared internally, when the internal router probably doesn't have a direct connection to the router specified in the next-hop?

THE PROBLEM



- Consider that R3 advertises a path to AS2 to R2 via eBGP
- R3 sets the next-hop to it's own IP address (the IP of R3)
- R1 knows of this route, but the next-hop is still the IP of R3, which R1 is not connected to!

THE SOLUTION

- First, as a rule, routes shared via iBGP will not update the next-hop attribute, i.e. the next hop remains the same as it was when the path was first shared with the AS via the external router
- An Internal Gateway Protocol (IGP) is used to distribute internal routing information instead, allowing internal routers to determine where traffic needs to go

HOW THE IBGP/IGP SOLUTION WORKS

- A gateway router learns of a path, and then advertises that path internally in the AS via iBGP
- An internal router learns the path from iBGP – it knows traffic destined for a given prefix goes to a certain external IP in a different AS
- The internal router uses an IGP-based routing table to determine the path to the gateway router, allowing communication between the internal router, the gateway router, and the external destination router

A FINAL, IMPORTANT CAVEAT

- Remember how iBGP doesn't prepend to the AS Path?
- Well, while it avoids traffic being thrown out due to the ASN appearing multiple times, it means that iBGP lacks a proper loop-detection system
- As a result, certain constraints must be put on iBGP that aren't placed on eBGP, in order to maintain the integrity of the routing system

IBGP ADVERTISING RULES AND FULL MESH

- The major constraints are:
  - Routes learnt via iBGP cannot be re-advertised via iBGP (RFC 4271).
  - As a result, the internal BGP network in the AS must be *full mesh* – all nodes are connected to every other node
  - Every iBGP router is networked with every other iBGP router, advertisements only make a single trip between router-router, no re-advertising necessary

SCHEDULE

1. *Recap*
2. *Path-vector Routing Protocol*
3. *BGP Messaging*
4. *BGP Path Attributes*
5. *eBGP and iBGP*

SO LONG, FOLKS!