CS348 Notes
BGP+IGP
Video Numbers: 20, 21

OjMaha

I have prepared these notes by watching the videos from Networks Playlist. The following notes may be asynchronous and irrelevant to what Prof. Vinay teaches in class (cuz I do not pay attention during lectures lol). Further, these notes might not cover *everything* as explained in the video lectures. Consider these to be a supplemental read :). If you find any errors, do notify me so they can be edited.

Border Gateway Protocol (BGP)

It conducts inter-domain routing b/w ASes.

Customer: The one who pays for service. (Multi-home customer: pays multiple providers for service)

Provider : The one who provides service.

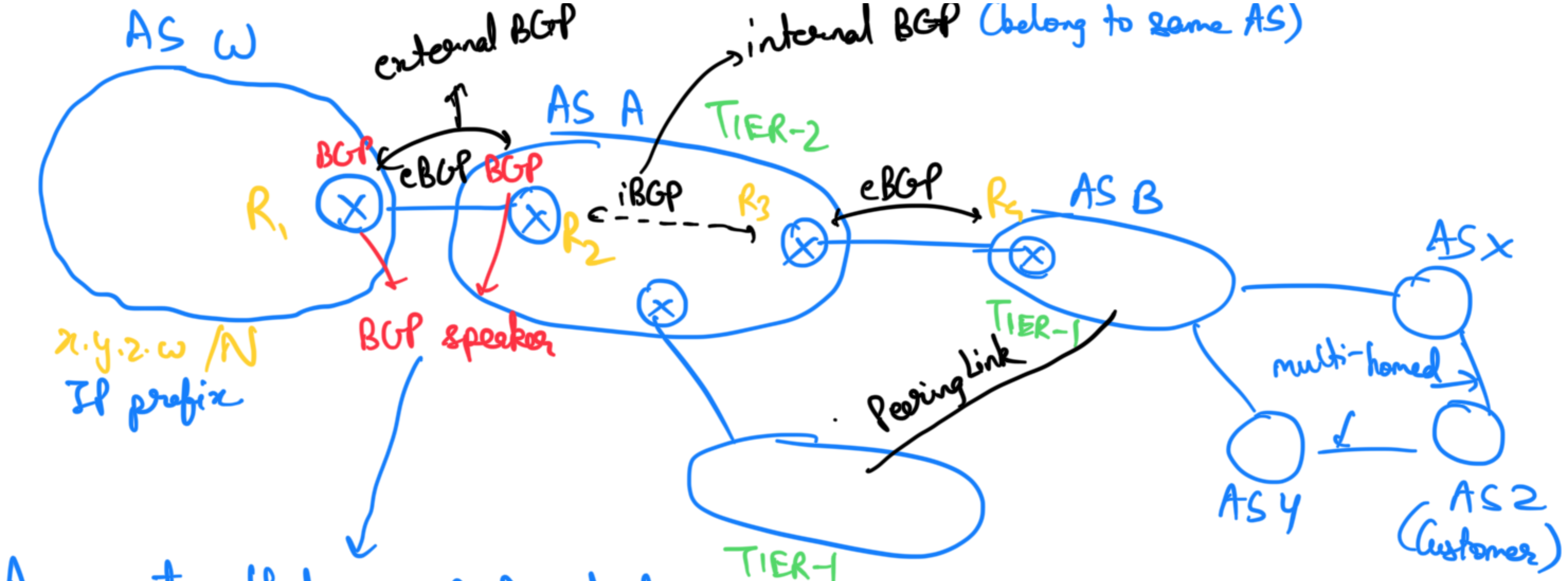
Tier-1 AS: pays nobody. It only acts as a provider. They are widely connected

Tier-2 AS: takes internet service from Tier-1 AS.

Define Tier-n likewise.

Peering Link: Connection b/w two Tier-1 AS for convenient data sharing to others. No one pays the other here. Since this is unpaid, the bandwidth is low to avoid "cheating" by sending huge amts of data to the other Tier-1. It is "under-provisional".

Service Level Agreement: A QoS guarantee b/w 2 AS abt latency, bandwidth. Note that there cannot be a full end-to-end latency/bandwidth guarantee since AS do not know the whole path. The agreement is just for within own network.



Any router that runs BGP protocol.

eBGP R_1 tells R_2 : $\underbrace{x.y.z.w/N}_{\text{IP prefix}}$ $\underbrace{W}_{\text{AS Path}}$ $\underbrace{\qquad\qquad}_{\text{attributes}}$ » Advertisement

iBGP

eBGP R_3 tells R_4 : x.y.z.w/N A–W $\underbrace{\qquad\qquad}_{\text{attributed}}$

R_5 tells R_6 . — — B-A-W

The AS path is the path from me to reach the IP prefix. It also means that if R_6 gives R_5 an IP pkt with Dst. IP matching prefix;

it will be able to route it along B-A-W.

☆ B does not necessarily have to tell C that there is a path B-A or B-A-W. B & C can decide to keep some of their connections quiet. Because both are Tier-1 AS, they (say B) do not want the other (say C) to send its traffic to it (B).

Similarly, Z is a small network and doesn't want packets from X to Y to be routed through it. It doesn't advertise the connection.

BGP does not compel any router to share its ads.

e-BGP : protocol b/w speakers of diff. ASes. They use TCP port 179.

no. of routers in AS.

i-BGP : same AS. ∃ pairwise TCP connections. $O(n^2)$ connections

IGP : Interior Gateway Protocol (Intra-domain routing) ⇒ LSR, DV, etc.

Steps to decide e-BGP paths:

(i) e-BGP speakers learn AS-paths from neighboring routers in other ASes.

(ii) e-BGP nodes share learned info via iBGP with other BGP speakers in own AS

(ii) e-BGP nodes share routes info

(iii) BGP speakers select routes to various IP prefixes.

(iv) Insert chosen routes into IGP. We do this cuz there may exist nodes & speakers that do not follow BGP but only IGP. Thus this allows all routers to fwd packets to a destination IP.

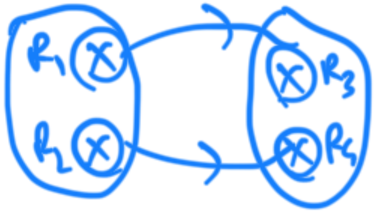
(v) eBGP speakers can now advertise newly created routes to neighboring ASes.

BGP attributes:

(i) Local-preferences

Suppose AS2 receives two ads for same prefix. The network administrator determines what the local pref for these routes. Higher local pref ⇒ more preferred path.
This pref. gets forwarded.

(ii) Multi-exit discriminator (MED)



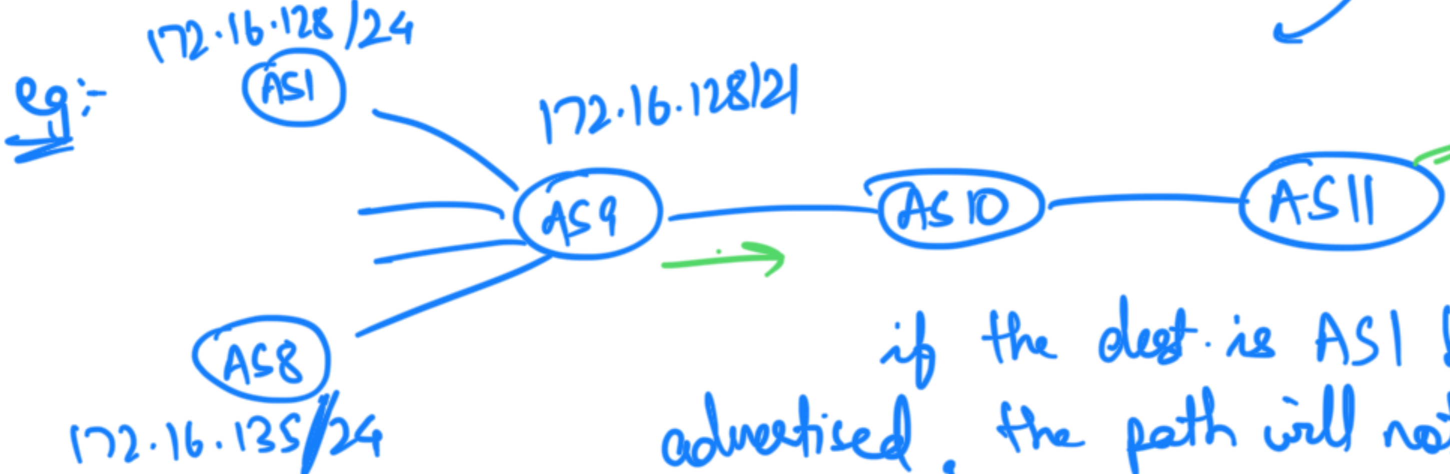
Suppose $R_1 \rightarrow R_3$ & $R_2 \rightarrow R_4$ advertise abt. same prefix but diff. MEDs.
↓ MED: 2000 ↓ MED: 1000 ⇓ say P

This means; AS1 is telling AS2 that it prefers packets with dest. P to be routed via $R_4 \rightarrow R_2$. Lower MED ⇒ higher pref. The value may be the IGP distance to the next BGP speaker.

(iii) AS Path : List of all AS nos all the way to the dest. with that IP prefix.

whoever first advertises that prefix.

eg:-

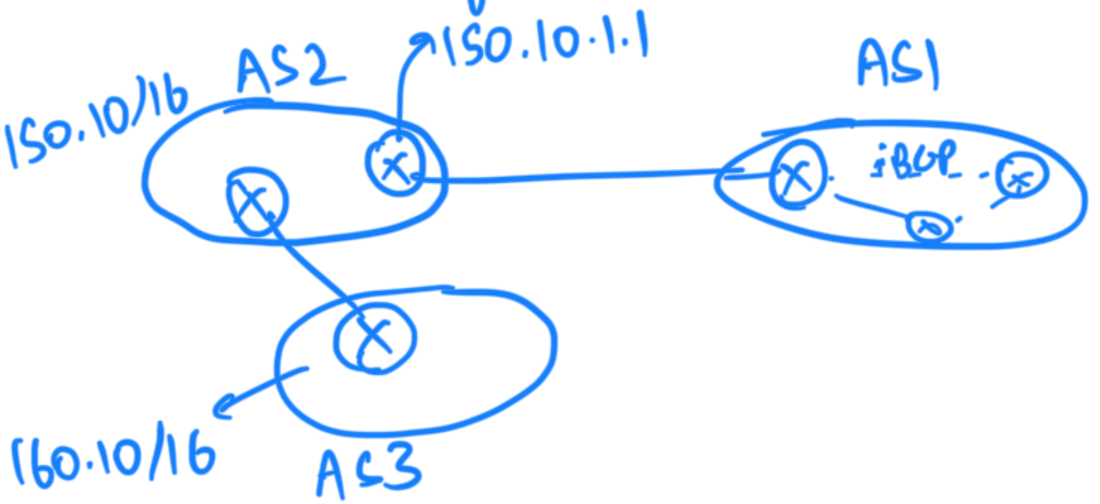


Path: AS11, AS10, AS9

if the dest. is AS1 but AS9 is the first who advertised, the path will not contain AS1 but end at AS9. After AS9 receives info; it forwards to AS1.

(iv) Next-HOP :

IP address of the first router in the next AS.



info that AS1 has : Next-HOP

dest^n	Path	Next-HOP
150.10/16	AS2	150.10.1.1
160.10/16	AS2, AS3	150.10.1.1

Rules to choose Routes:

[illegible] AS-route to use among many available

Each BGP speaker must decide which AS-route to use among many available for the same prefix. (tie-breaks along the way)

if no local pref was set

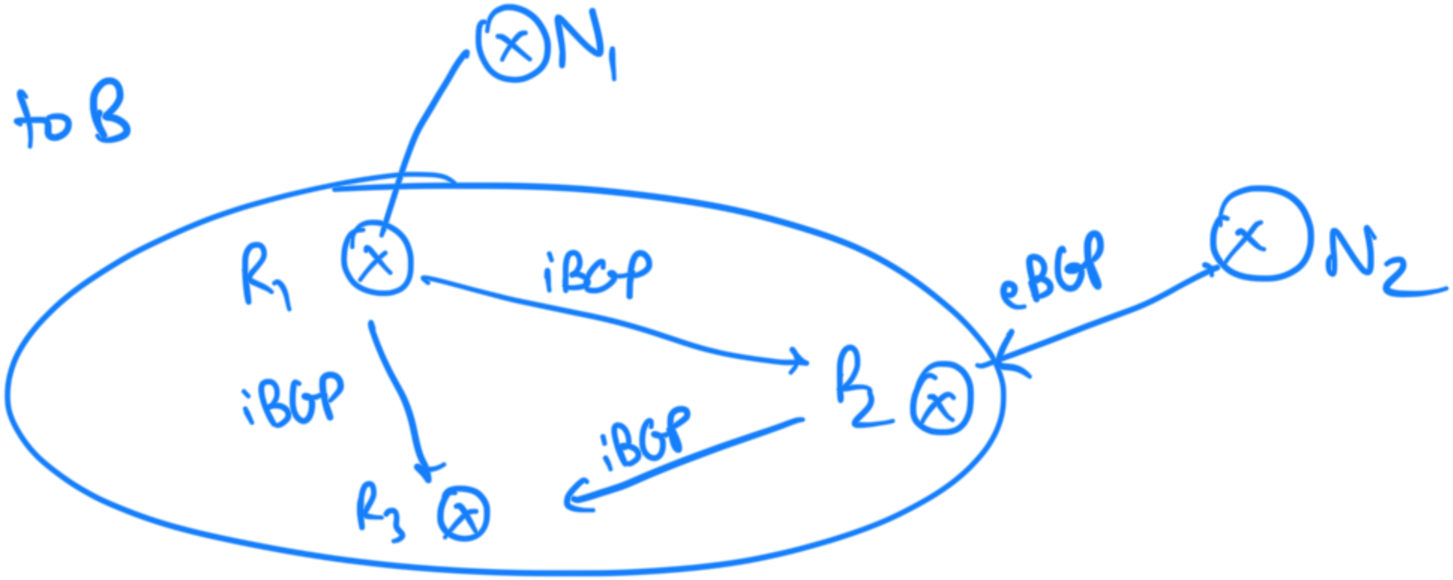
tie-breaker for $a_1 == a_2$ && $b_1 == b_2$

a) Use largest local-pref. b) Use shortest AS-path. c) choose lowest MED path

d) Choose path learned by eBGP over path learned by iBGP.

e) Choose path with lowest IGP metric to next-hop. → HOT-Potato routing

A→B means A advertises to B



Here, R_2 chooses to route over N_2 (eBGP) rather than R_1 (iBGP) in case #AS, and all other metrics are the same.

Here, R_3 recd only iBGP from R_1 & R_2. Say dist.$(R_3, N_2) >$ dist.(R_3, N_1).
Then; R_3 chooses to route via R_2. (nearest exit) (garam aloo routing)

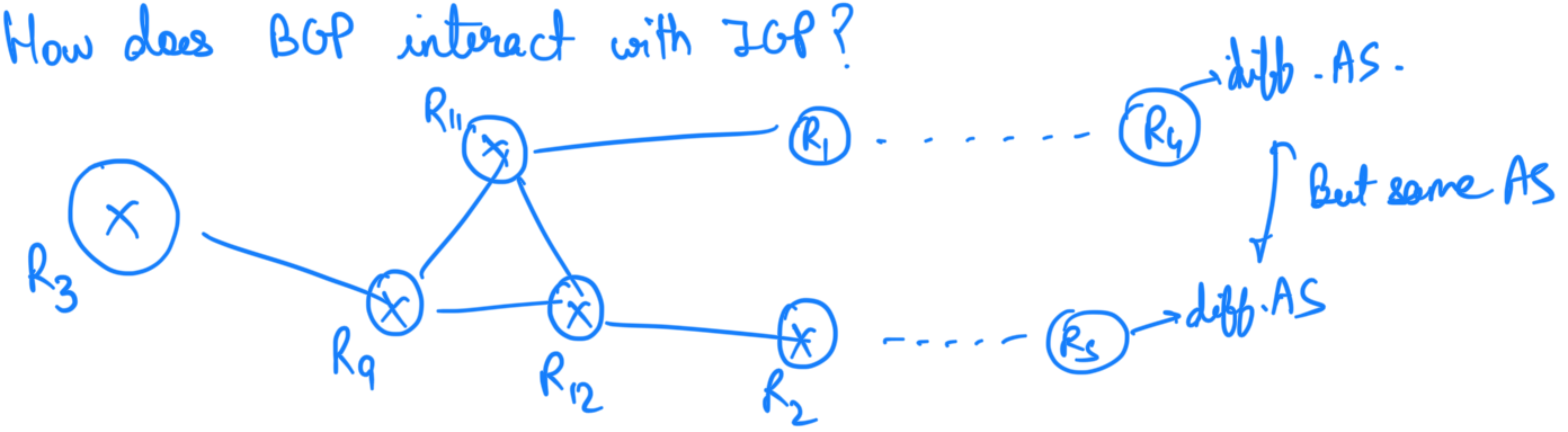
f) Choose lowest router ID among all BGP speakers who have sent there ads.

router ID: highest IP on router. This is the final tie-breaker.

Why do routers have diff. IP addresses? Because the router may be a part of several networks, each of which requiring a diff. prefix. To satisfy all the cond^n for those nets, it has diff. IPs allowing flexibility.

Watch vid lec 21 till 22 mins for complete example.

How does BGP interact with IGP?



(i) Encapsulation:

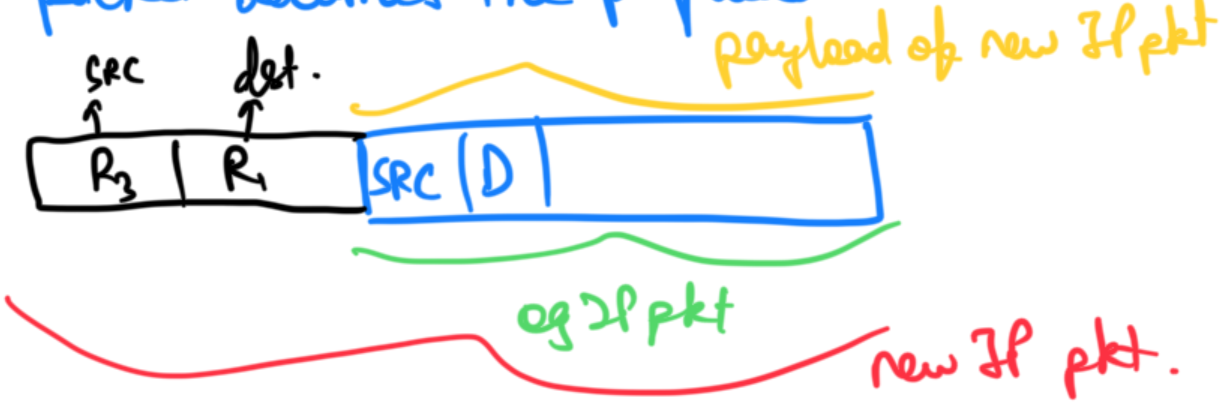
Suppose R_9, R_{11}, R_{12} have no BGP info. It only has info on how to route pkt to routers of own AS.

in R_q table (IGP)

Dest.	Next Hop
R_1	R_{11}
R_2	R_{12}
R_3	R_3
R_{11}	R_{11}
R_{12}	R_{12}

Sidenote: each interface in a router has diff. IP. (Here, we ignore that)

Suppose dest is some router D with IP 142.32.6.21. It is not in routing table of R_3. When R_3 receives this IP packet, it creates another IP pkt where the recd packet becomes the payload.



R_1 receives this pkt, strips off the outer IP layer and forwards the internal pkt to R_4. (R_1's table has info of R_4 via eBGP)

(ii) Pervasive BGP:

All routers must speak BGP.

Suppose there is a unique exit for 142.32.6 /24

BGP table:

Prefix	Exit/Gateway
142.32.6/24	R_1

true for all

IGP table:

IP addr	Next Hop
⋮	⋮

142.32.6/24

→ Recursive Lookup.

On receiving a packet, it (say R_9) finds out dest. prefix; looks for gateway in BGP table (R_1) & then looks for next hop in the IGP table (R_{11}). R_{11} does a similar thing and so on.

(iii) Tagged IGP;

Internal routers may not be BGP speakers here. But IGP allows add^n of "tags" in the ads.

R_1 inserts into its own IGP table:

gateway router

142.32.6/24 R_1 → gateway router.

prefix tag

This info is propagated to all routers in AS using IGP

Now @ R_9:

IGP table:

	Dest	Next	Tag	Cost
	R_1	⋮	⋮	5
	R_2			10
	⋮			⋮
announced by R_1 ←	142.32.6/24		R_1	
announced by R_2 ←	142.32.6/24		R_2	

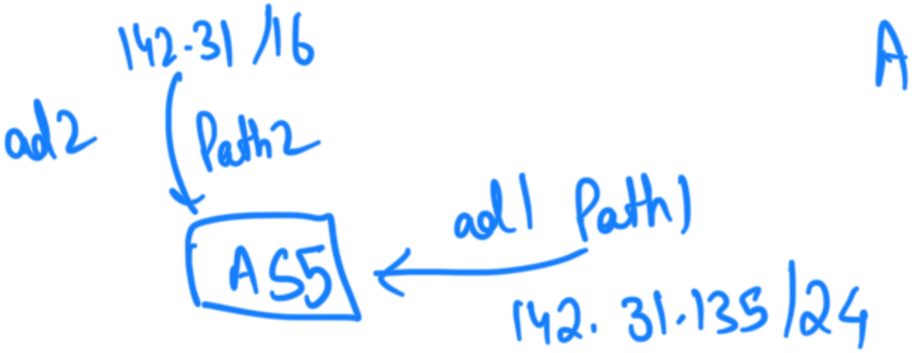
Thus, R_9 looks up for the prefix & finds closest match. (cost-wise)

hot-potato routing.

Longest Prefix Matching:

We do supernetting to decrease the size of routing tables.

Now say there is an AS that receives 2 ads in which one IP is the subset of the other. eg:



AS5 decides to use Path1 for 142.31.135/24 & Path2 for 142.31/16 excluding 142.31.135/24

Now say dest. IP of a packet rcd @ AS5 is 142.31.135.20

This particular addr matches both the prefixes as mentioned earlier in the routing table. So, longest prefix is chosen to route.

Why? Because longer prefix match would mean that that particular path is shorter. If we used shorter prefix, there woulda been more number of hops in order to aggregate /supernet into the one with shorter prefix.

Further, longest prefix ensures pkt reaches dest?. If we used shorter prefix, there is a possibility that a pkt keeps hopping b/w 2 routers who each had advertised to each other IP_1 and IP_2 s.t. either $IP_1 \subseteq IP_2$ or $IP_2 \subseteq IP_1$.

CAM: Content addressable memory. Allows quick look up in the table.