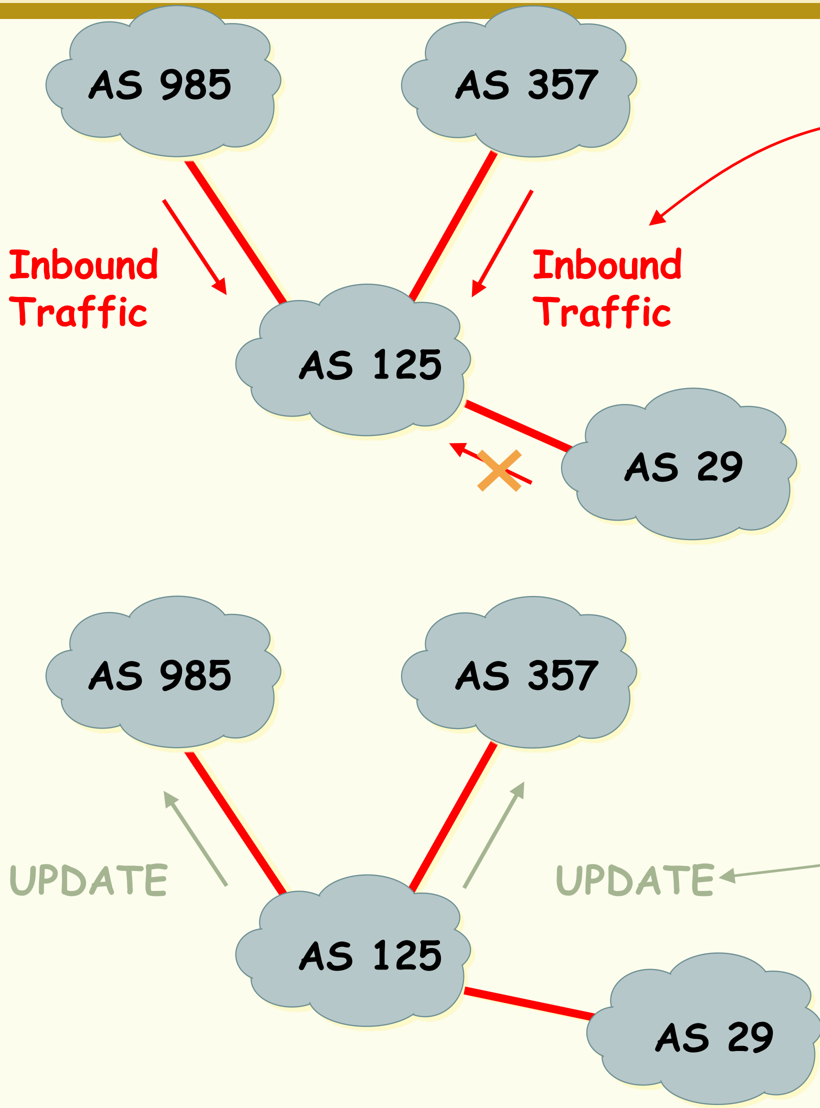
Interdomain routing

BGP-4 decision process

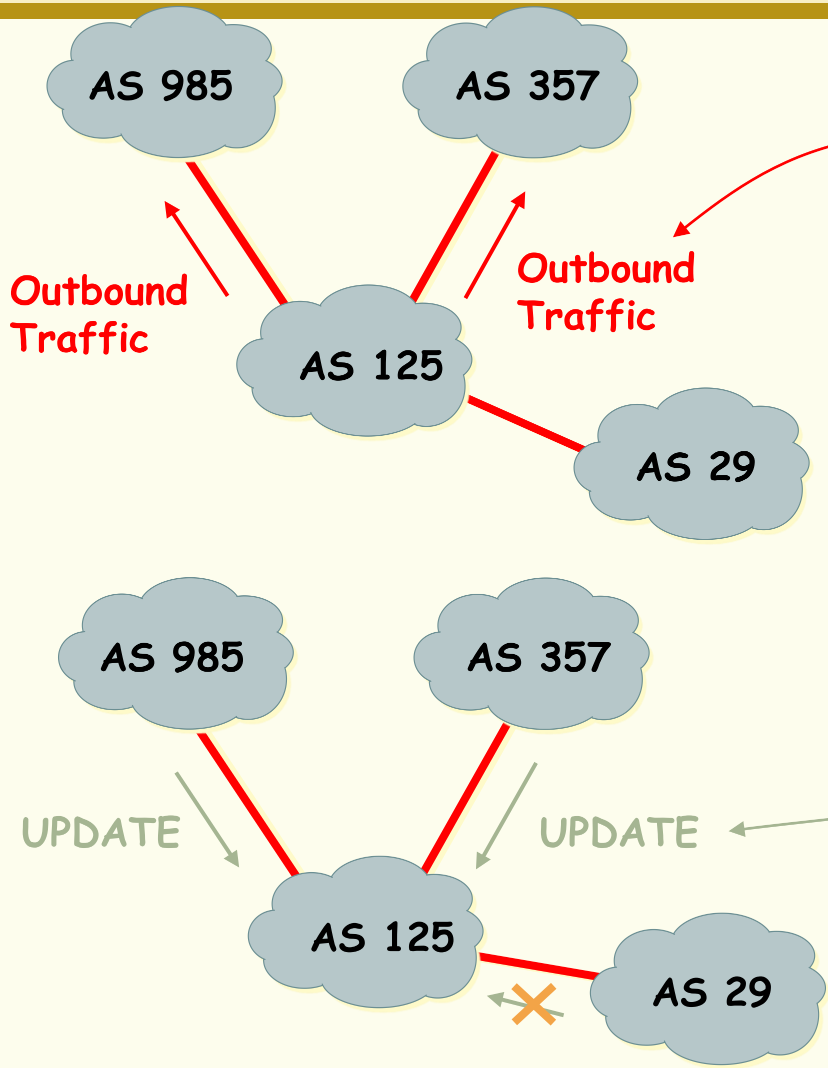
Enzo Mingozzi
Professor @ University of Pisa
enzo.mingozzi@unipi.it

Route Filtering



An autonomous system (AS 125) can identify the inbound traffic it is willing to accept from other neighbors by specifying the list of routes it advertises to its neighbors

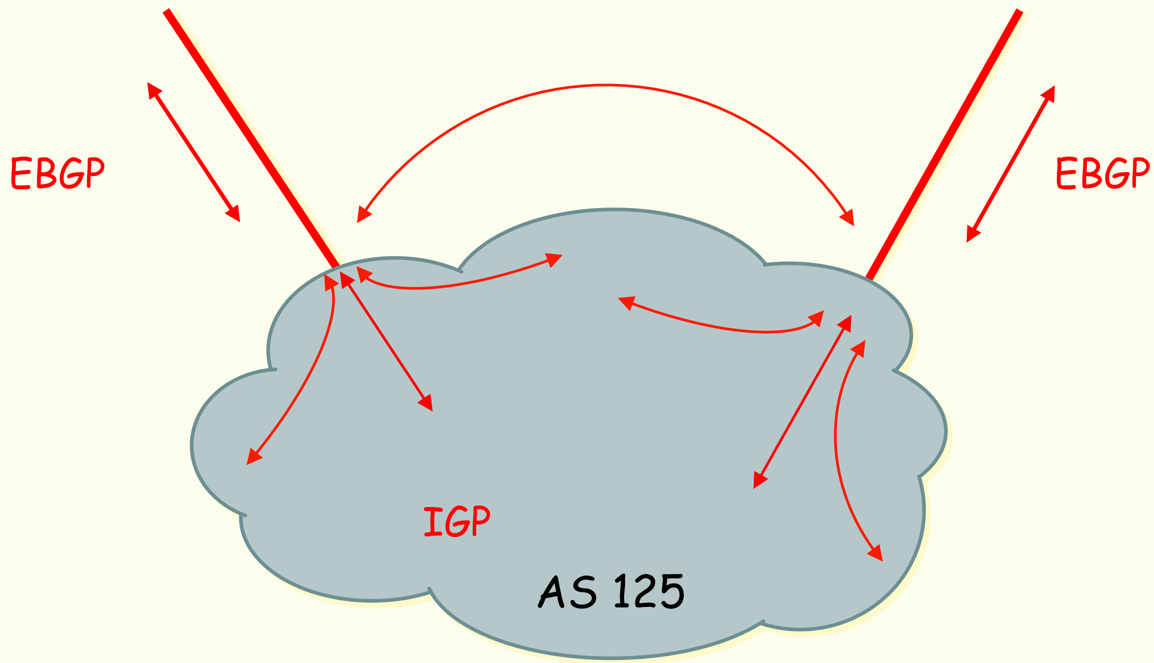
Route Filtering



An AS can control what routes its outbound traffic uses by specifying the routes it accepts from its neighbors

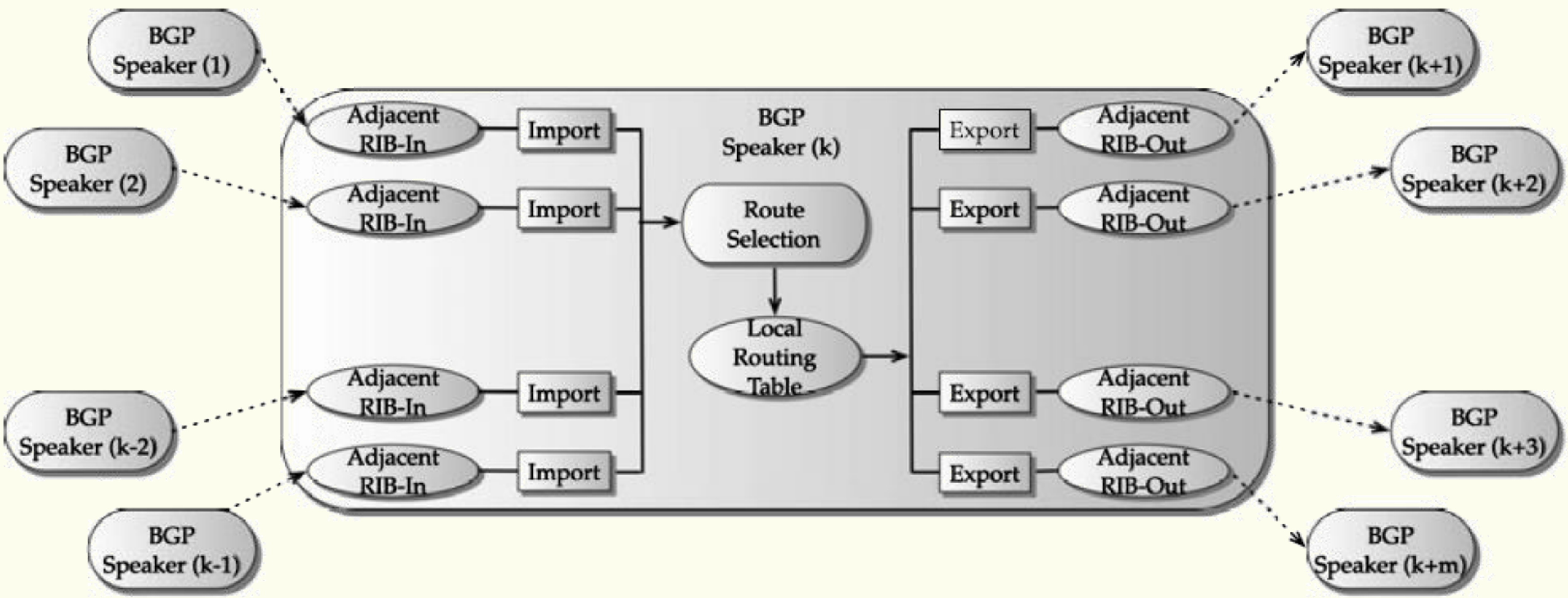
Route Filtering

Filtering can also limit routing updates flowing from one protocol to another
There is the possibility of injecting BGP routes in the IGP as well as injecting the IGP or static routes into BGP



BGP decision process

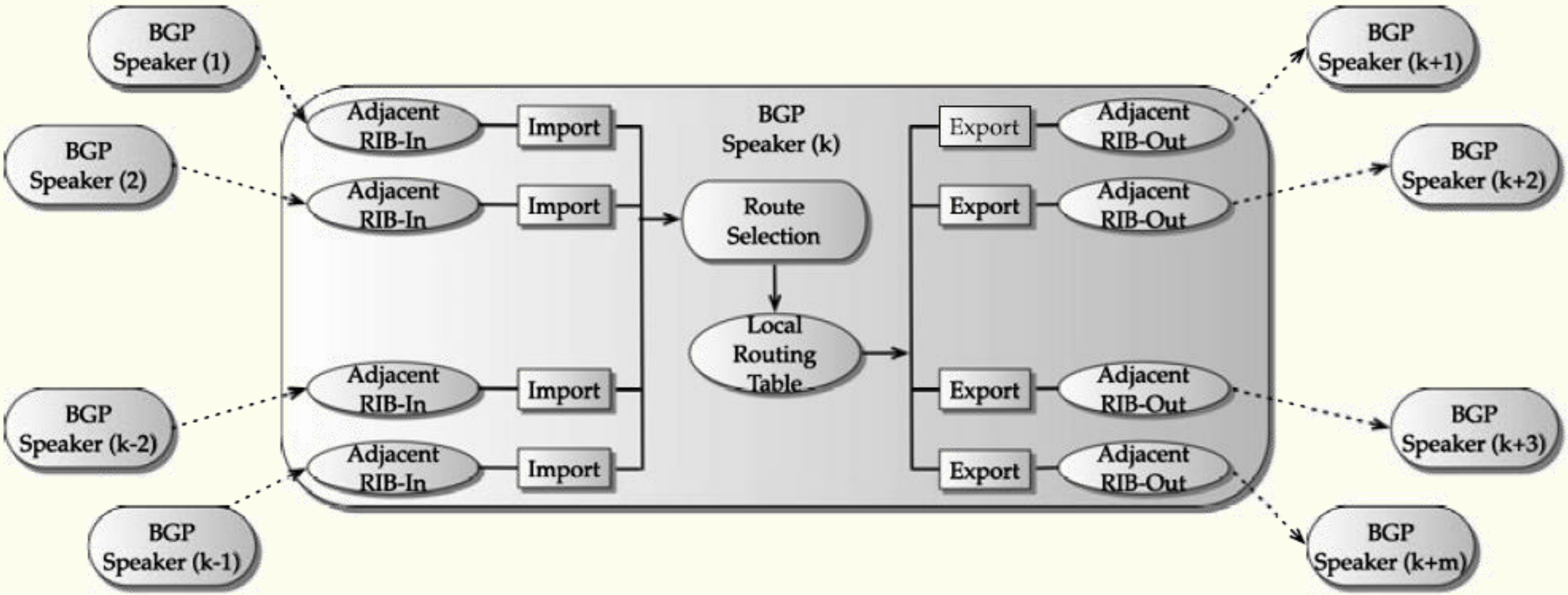
The BGP decision process consists of
1) **path selection**, and
2) **(aggregation and) dissemination**



BGP decision process

Each BGP speaker maintains several **Routing Information Bases**

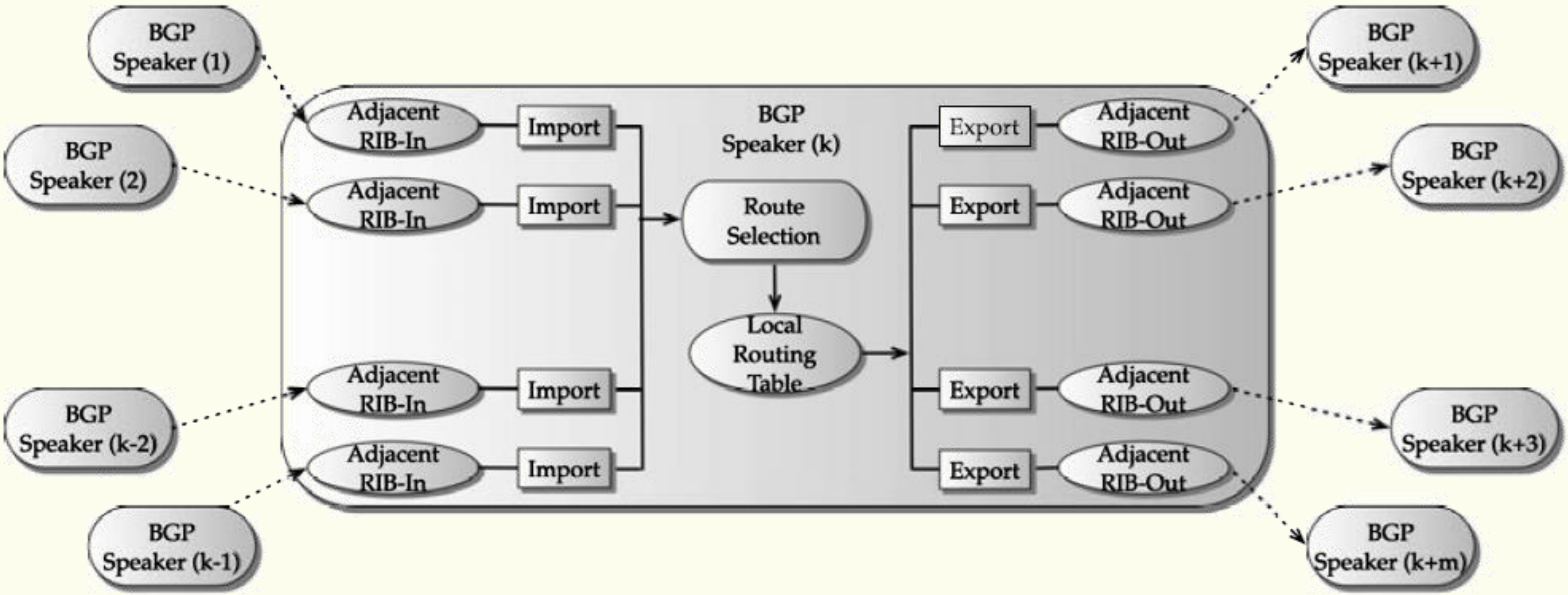
Adjacent RIBs-In (Adj-RIBs-In) stores AS level routing information for each IP prefix it has learned about from its neighbors through inbound UPDATE messages



BGP decision process

Each BGP speaker maintains several **Routing Information Bases**

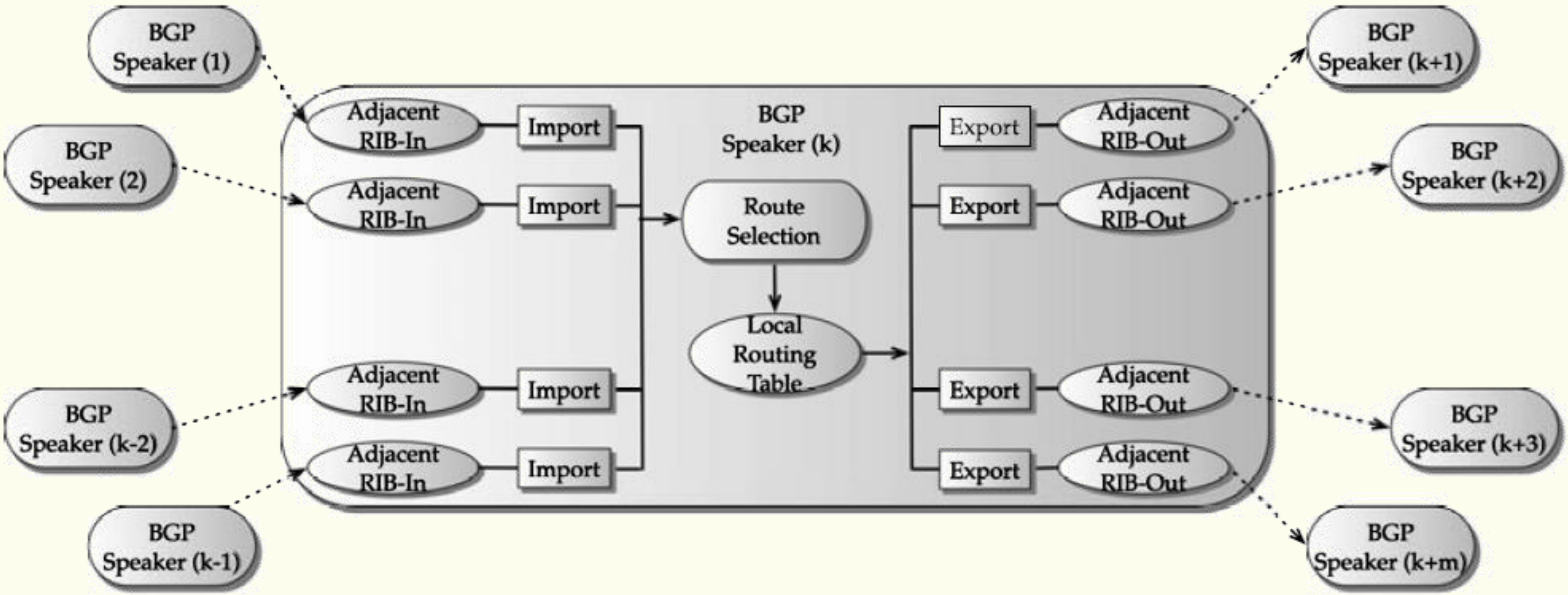
Loc-RIB stores the routes that have been determined locally by the BGP speaker decision process, used for updating the forwarding table



BGP decision process

Each BGP speaker maintains several **Routing Information Bases**

Adjacent RIBs-Out (Adj-RIBs-Out) stores the routes for advertisement to its neighboring BGP speakers through outbound UPDATE messages

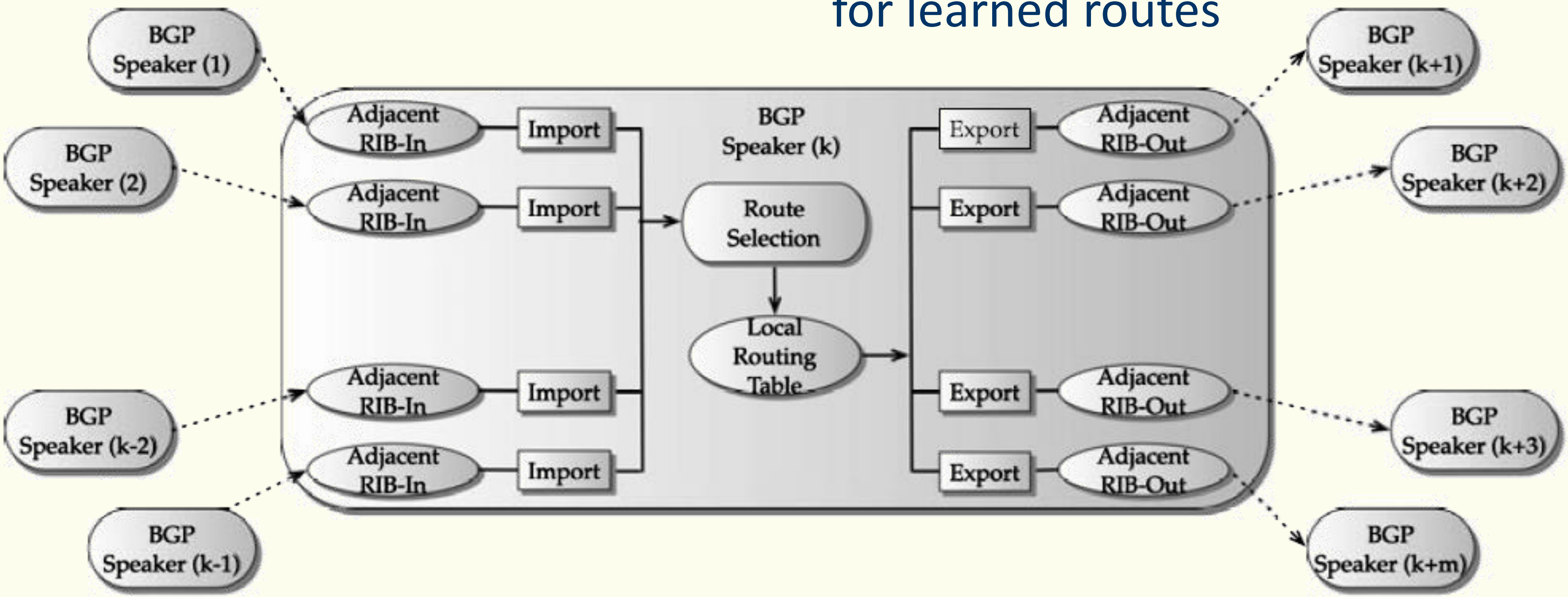


BGP path selection

Two phases

1) **Import policy and filtering**
2) Best route determination

- Filter out IP prefixes that are not allowed or that should not be reached via that peer
- Assess the degree of preference for learned routes

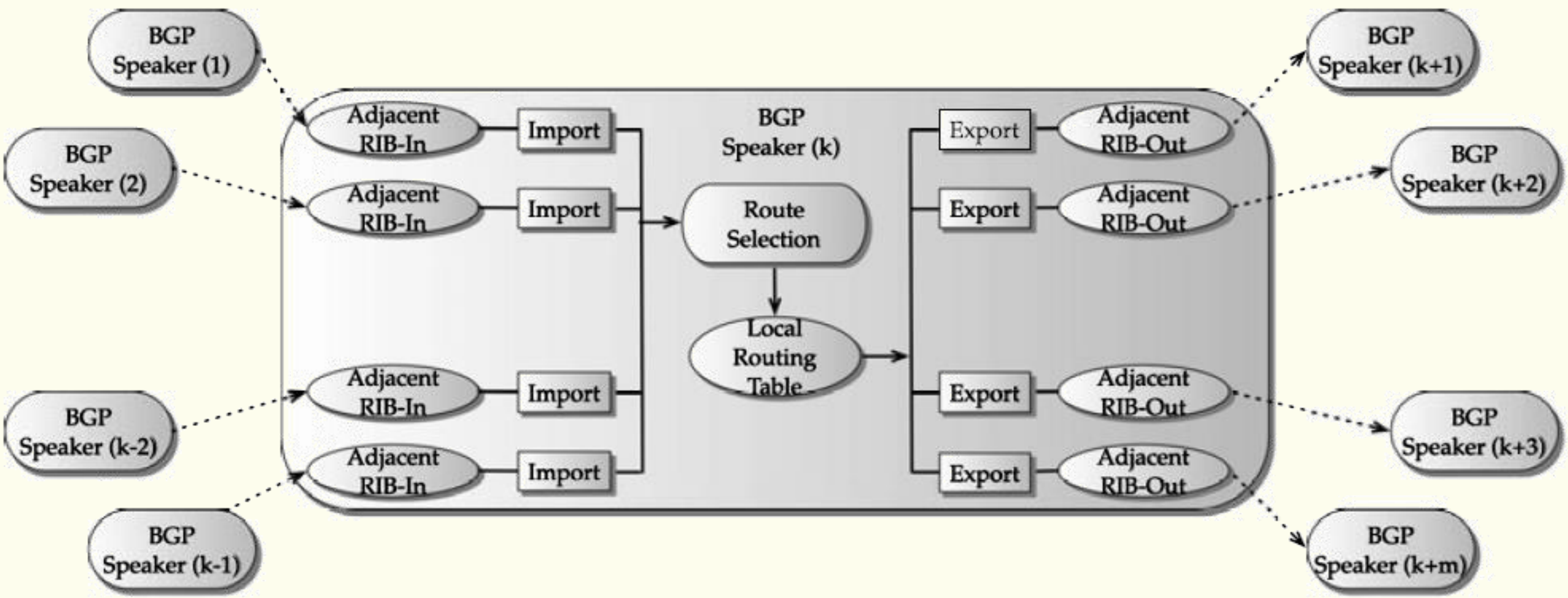


BGP path selection

Two phases

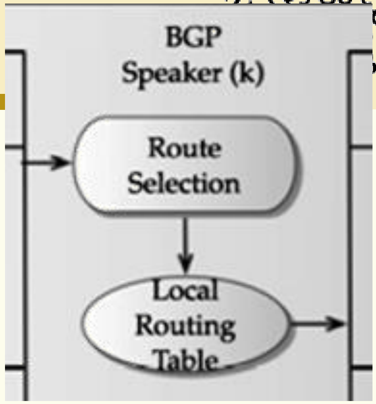
1) Import policy and filtering
2) **Best route determination**

- Select the best route for each separate **imported** IP prefix



BGP path selection

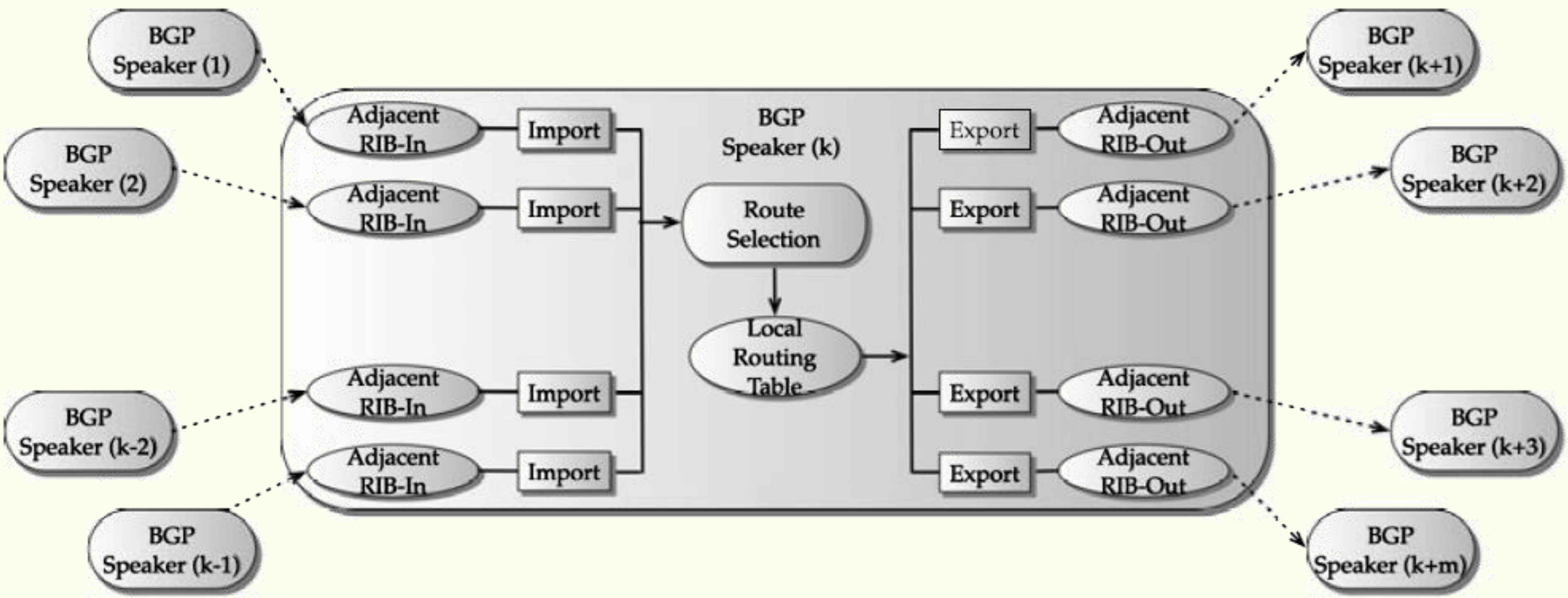
- **Tie-breaking rules** when multiple routes are available to the same **imported** IP prefix



1. Ignore routes for which the **NEXT–HOP attribute is not resolvable**
2. Apply the **degree of preference** assessed during the *import policy and filtering* phase (either on LOCAL–PREF if the announcement is received from an iBGP speaker, or any locally pre-configured decision)
3. Select the route that **originated locally** at the BGP speaker
4. Select the route with the **shortest AS path**
5. Select the one with the **lowest ORIGIN** attribute (IGP, then EGP, then Incomplete)
6. Select the route with the **lowest MED** for eBGP routes (learned from the same AS)
7. Select the route received from **eBGP** over iBGP
8. Select the route with **shortest (internal) path to the NEXT–HOP router** (as determined by IGP)
9. Select the route learned from the eBGP neighbor with the **lowest BGP identifier**
10. Select the route from the iBGP neighbor with the **lowest BGP identifier**

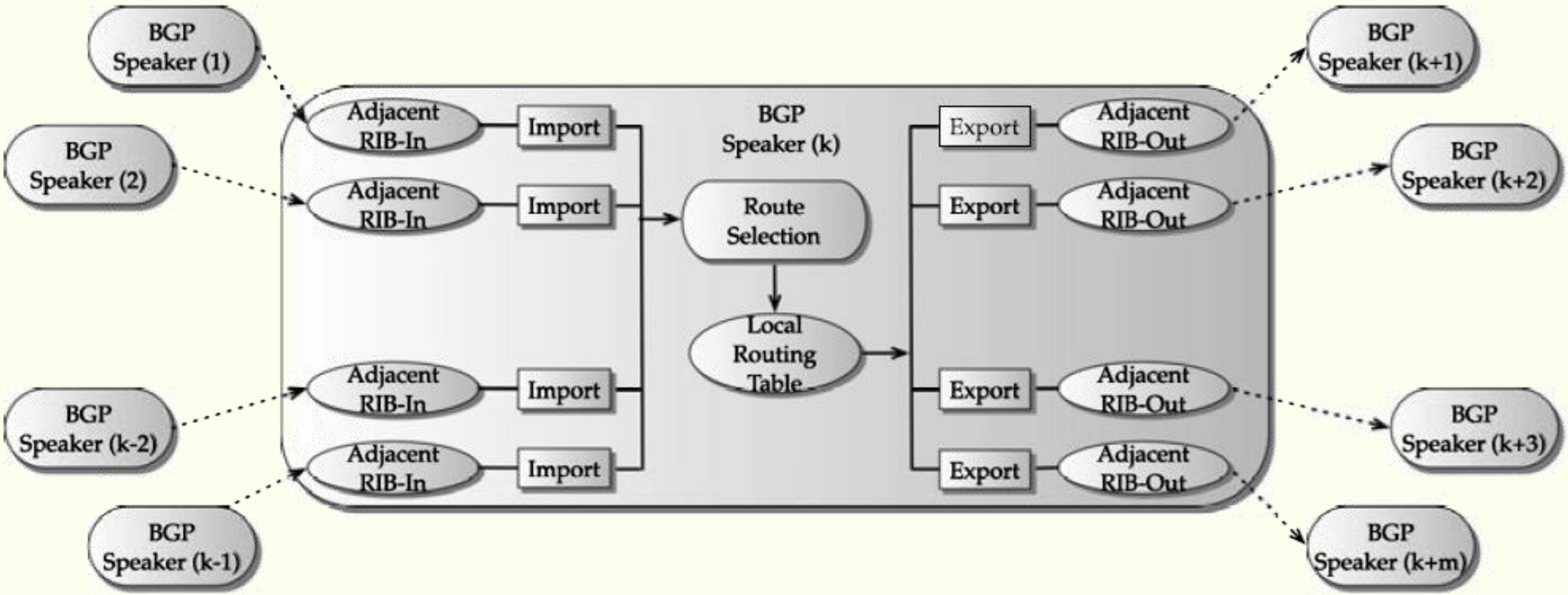
BGP route aggregation and dissemination

- *Optional* **route aggregation** based on CIDR: combine IP prefixes (*supernetting*) to reduce the number of networks announced to a downstream AS



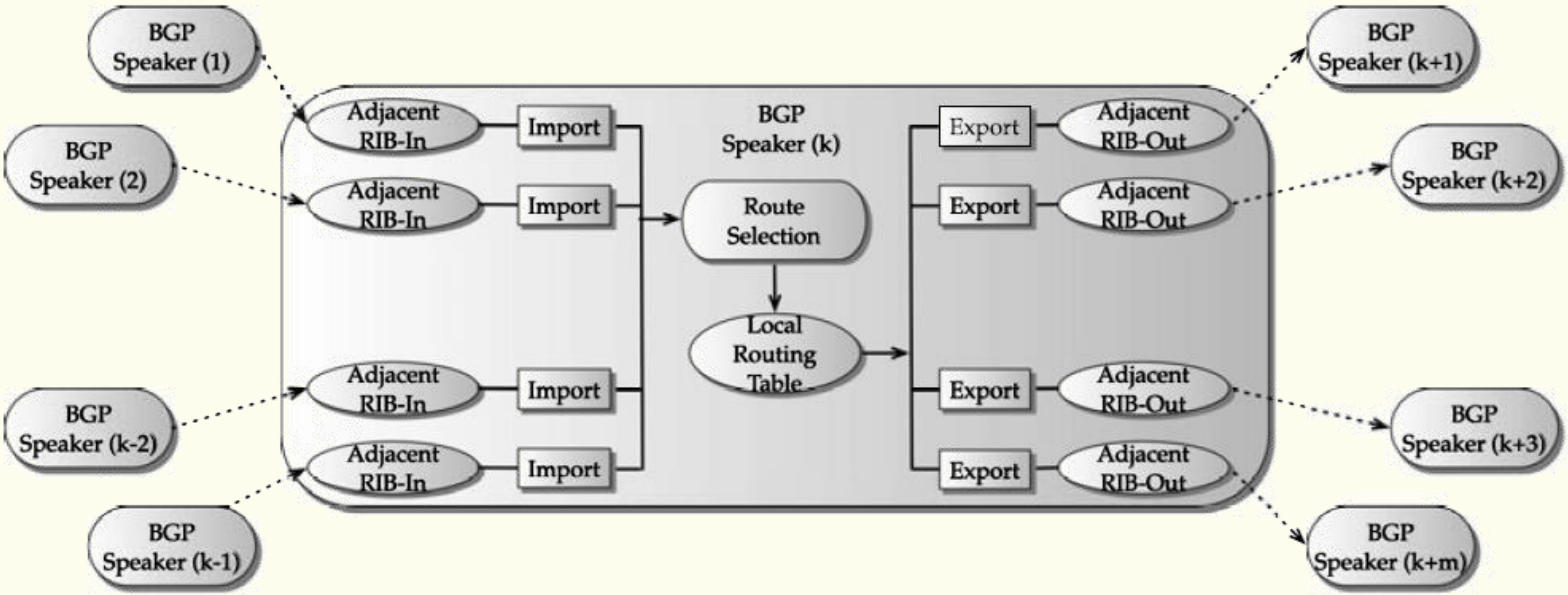
BGP route aggregation and dissemination

- A BGP speaker applies an **export policy** before propagating routes to other BGP speakers
- Export policies are separate per neighboring BGP speaker

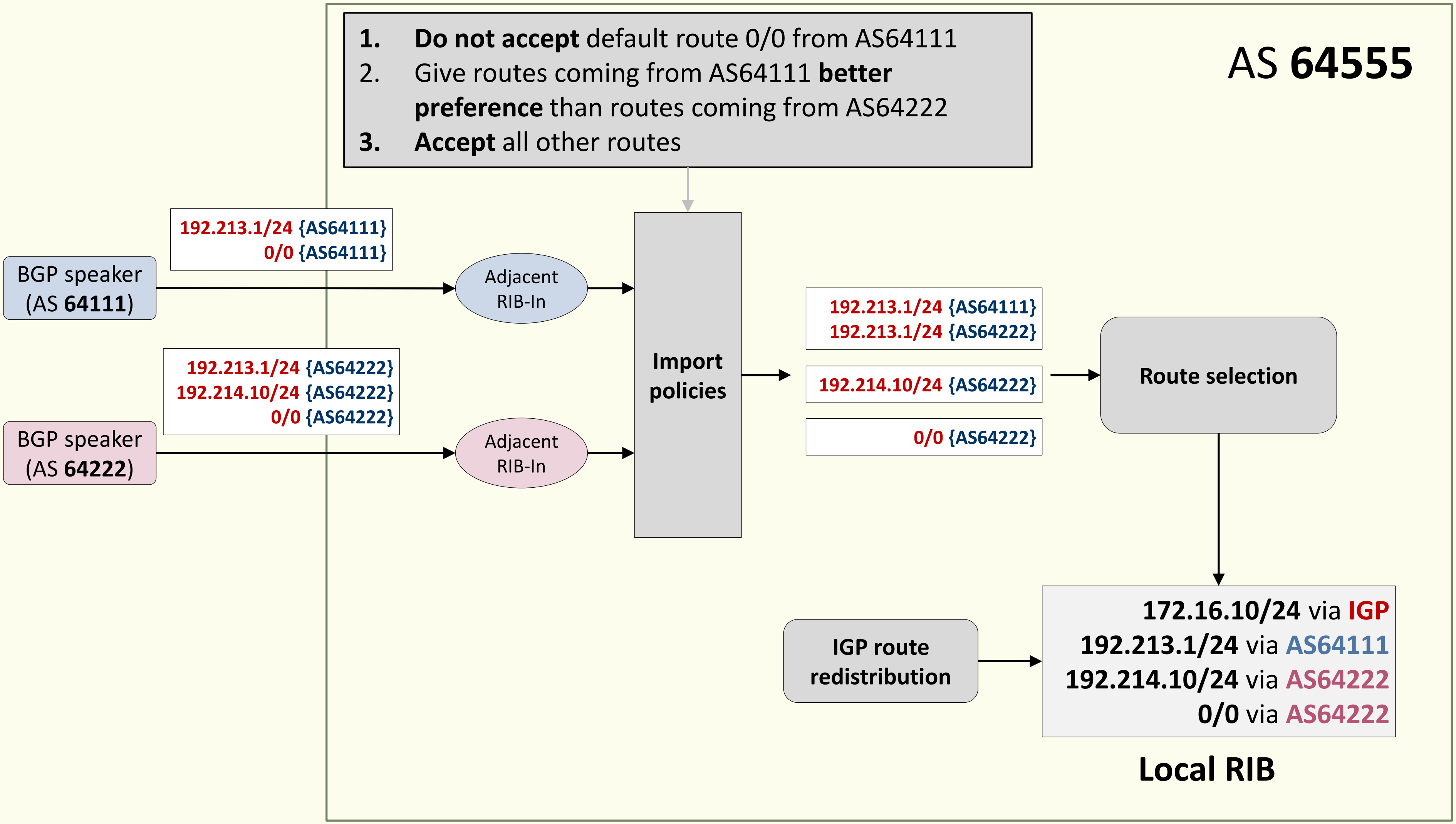


BGP decision process

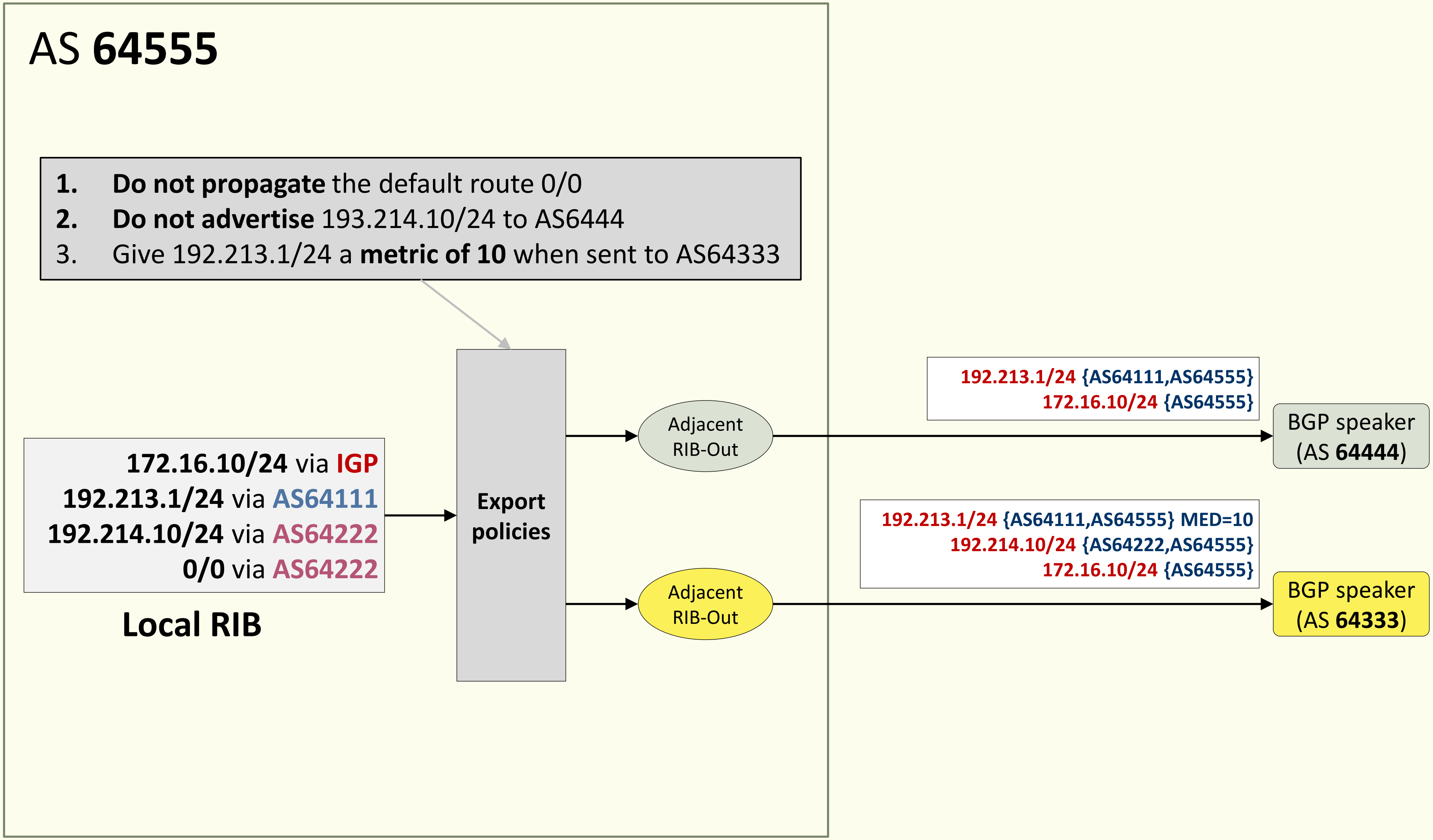
- **Policy-based routing**: import and export policies are placed at a BGP speaker by a network administrator due to business relations or peering arrangement, i.e., external factors



Example


1. **Do not accept** default route 0/0 from AS64111
2. Give routes coming from AS64111 **better preference** than routes coming from AS64222
3. **Accept** all other routes
AS **64555**
192.213.1/24 {AS64111}
0/0 {AS64111}
BGP speaker (AS **64111**)
Adjacent RIB-In
192.213.1/24 {AS64222}
192.214.10/24 {AS64222}
0/0 {AS64222}
BGP speaker (AS **64222**)
Adjacent RIB-In
Import policies
192.213.1/24 {AS64111}
192.213.1/24 {AS64222}
192.214.10/24 {AS64222}
0/0 {AS64222}
Route selection
IGP route redistribution
172.16.10/24 via IGP
192.213.1/24 via AS64111
192.214.10/24 via AS64222
0/0 via AS64222
Local RIB

Example


AS 64555
1. Do not propagate the default route 0/0
2. Do not advertise 193.214.10/24 to AS6444
3. Give 192.213.1/24 a metric of 10 when sent to AS64333
172.16.10/24 via IGP
192.213.1/24 via AS64111
192.214.10/24 via AS64222
0/0 via AS64222
Local RIB
Export policies
Adjacent RIB-Out
Adjacent RIB-Out
192.213.1/24 {AS64111,AS64555}
172.16.10/24 {AS64555}
BGP speaker (AS 64444)
192.213.1/24 {AS64111,AS64555} MED=10
192.214.10/24 {AS64222,AS64555}
172.16.10/24 {AS64555}
BGP speaker (AS 64333)

Internal BGP scalability

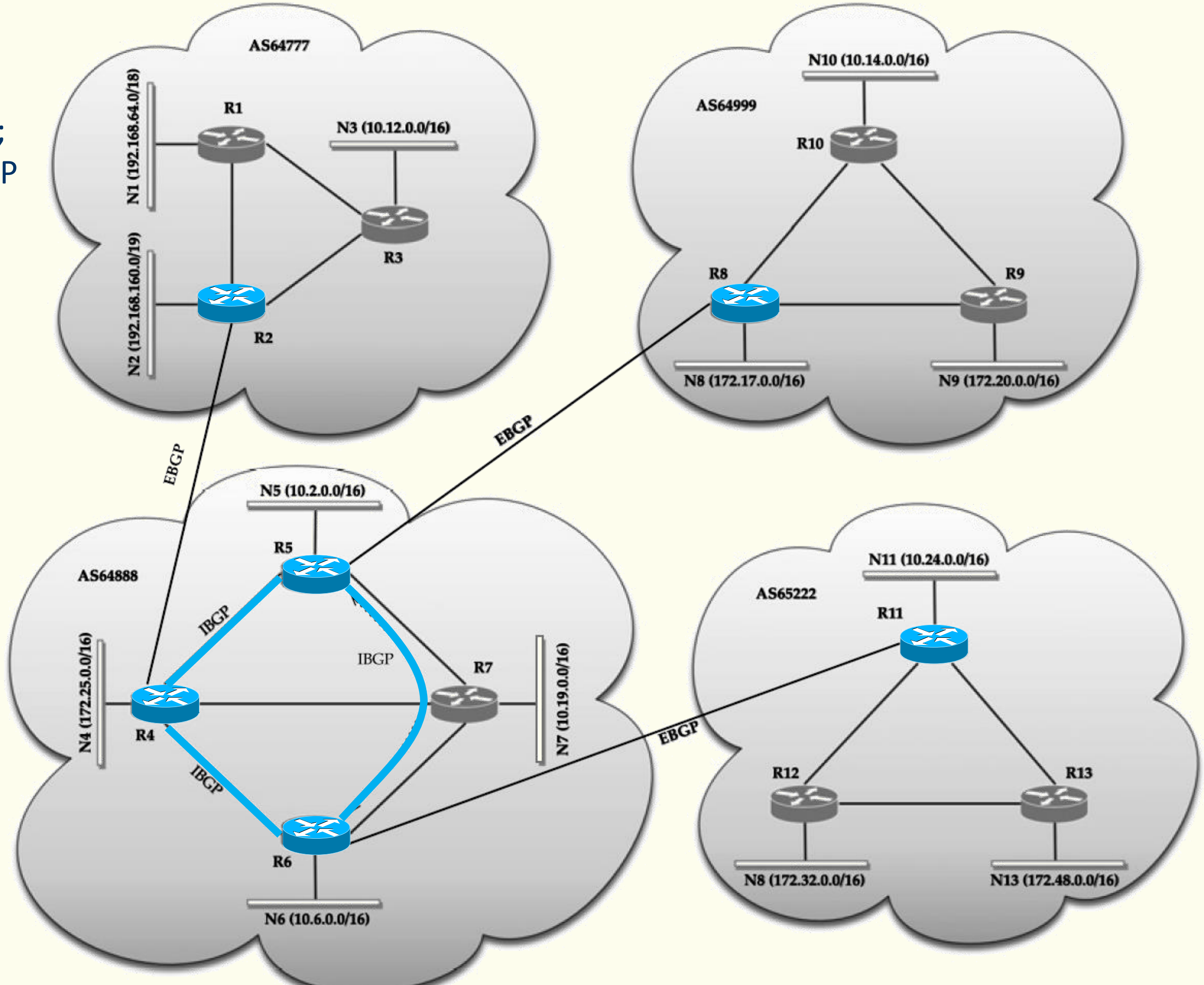
<u>Rule 1</u> A BGP speaker can advertise IP prefixes it has learned from an eBGP speaker to a neighboring iBGP speaker; similarly, a BGP speaker can advertise IP prefixes it has learned from an iBGP speaker to an eBGP speaker

<u>Rule 2</u> An iBGP speaker cannot advertise IP prefixes it has learned from an iBGP speaker to another peer iBGP speaker

<u>Two reasons:</u>

1. Avoid looping of BGP route updates within the AS
2. No need to advertise internal routes

A full mesh iBGP connectivity is needed



Internal BGP scalability

<u>**Rule 1**</u> A BGP speaker can advertise IP prefixes it has learned from an eBGP speaker to a neighboring iBGP speaker; similarly, a BGP speaker can advertise IP prefixes it has learned from an iBGP speaker to an eBGP speaker

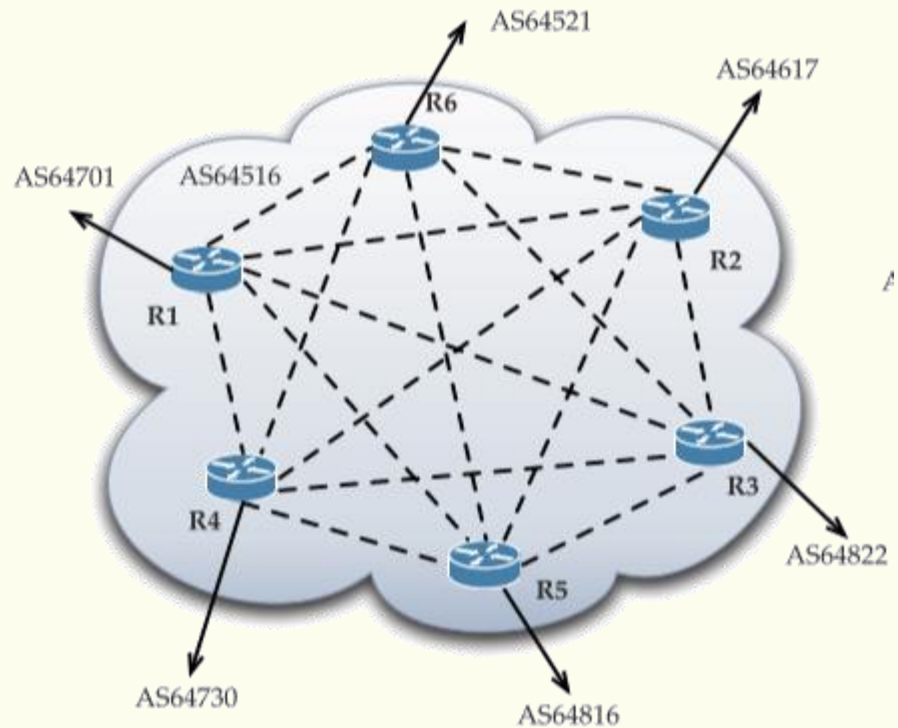
<u>**Rule 2**</u> An iBGP speaker cannot advertise IP prefixes it has learned from an iBGP speaker to another peer iBGP speaker

<u>Two reasons:</u>

1. Avoid looping of BGP route updates within the AS
2. No need to advertise internal routes

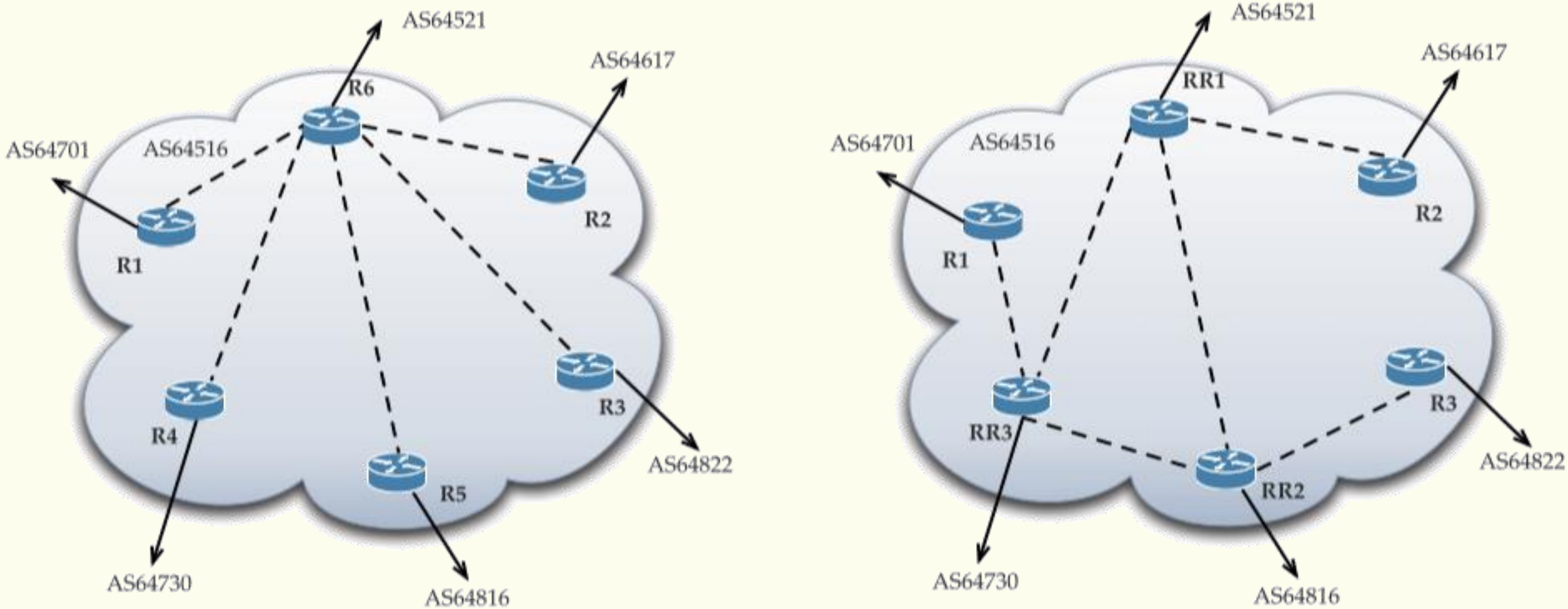
A full mesh iBGP connectivity is needed

<u>***n* iBGP speakers → *n*(*n* – 1)/2 iBGP sessions each speaker handling *n*–1 sessions**</u>



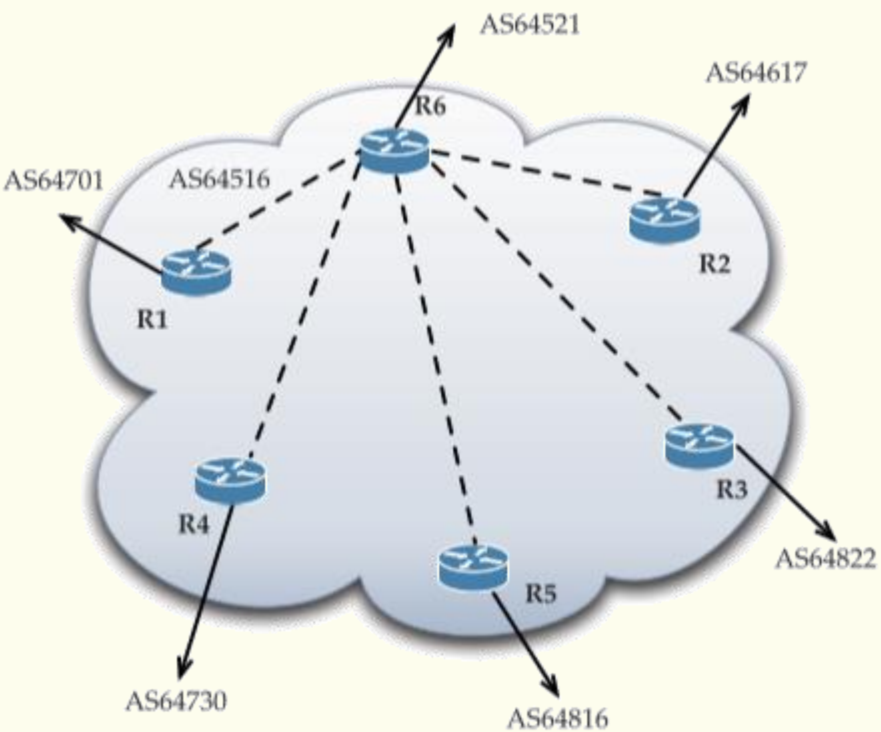
Route reflector

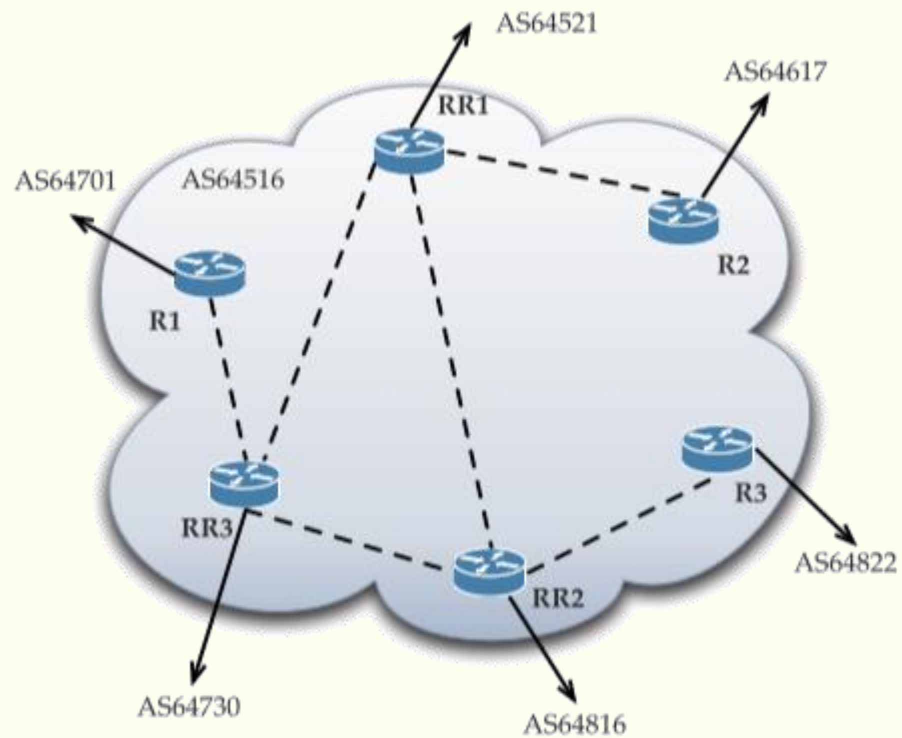
- One or more iBGP speakers act as **concentration routers**
- The other iBGP speakers establish only one BGP session to a route reflector (route reflector **clients**)
- Each route reflector with its clients form a **cluster**, identified by a **CLUSTER-ID**



Route reflector

- Announcement received **from another route reflector** → reflect/pass it to its clients
- Announcement received **from a route reflector client** → reflect to another route reflector
- Announcement received **from an eBGP speaker** → reflect to all other route reflectors and clients

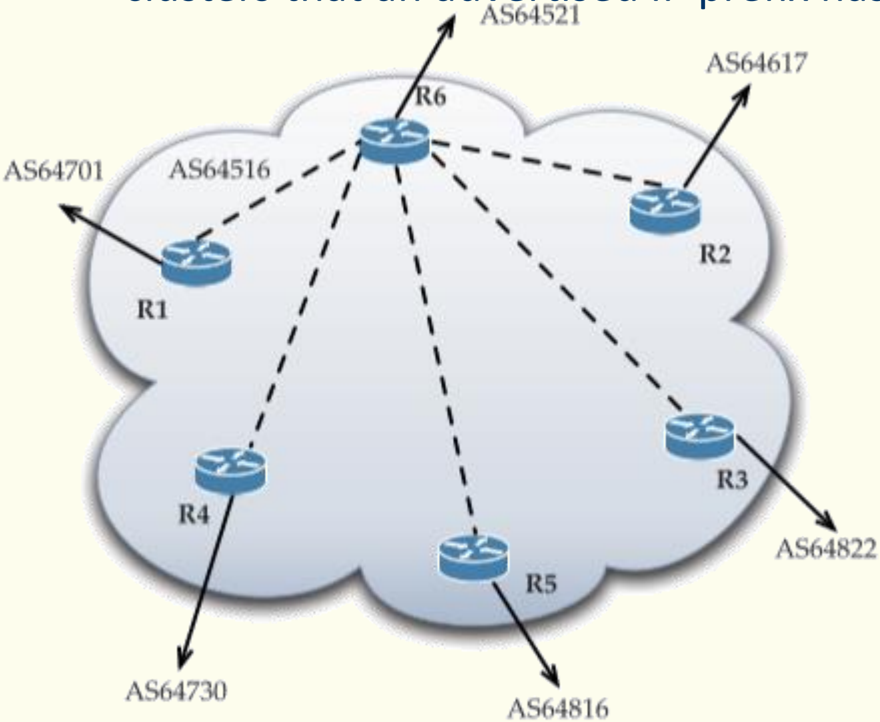


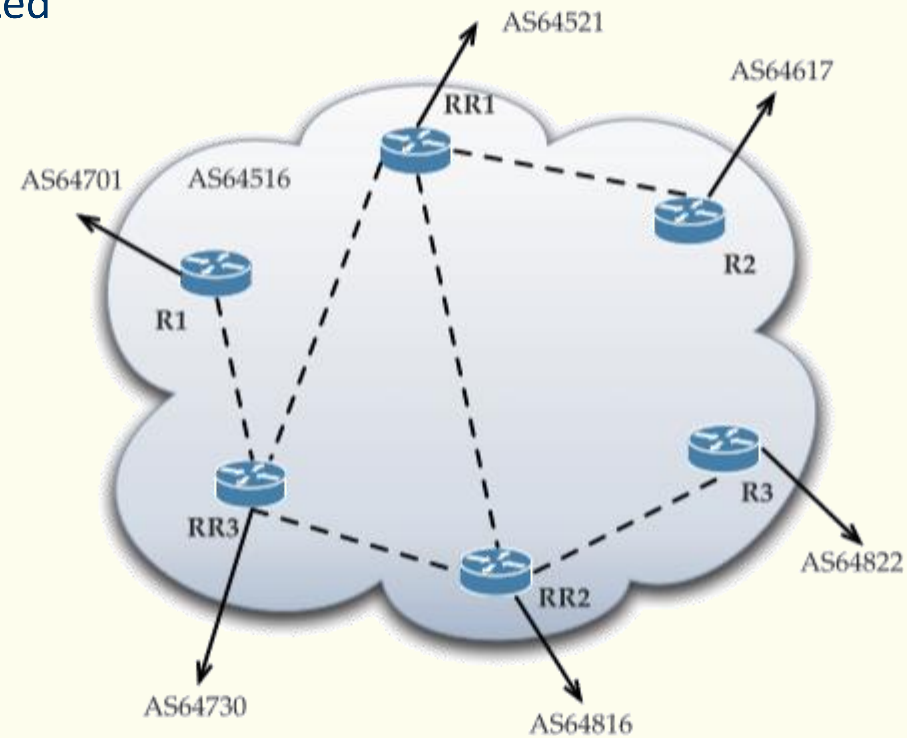


Route reflector

- Route reflectors must form a **full mesh connectivity among themselves**!
- How to avoid routing loops? Two additional attributes

1. **ORIGINATOR–ID**: identifies a route reflector through its 4-byte router ID, added only by the originating route reflector
2. **CLUSTER–LIST**: stores a sequence of 4-byte CLUSTER–ID values to indicate the path of clusters that an advertised IP prefix has visited





References

- D. Medhi, K. Ramasamy, **Network Routing: Algorithms, Protocols, and Architectures**, 2nd/ed. Morgan Kaufmann, ©2018
- RFC
  - **RFC4271**, A Border Gateway Protocol 4 (BGP-4), Jan. 2006
  - **RFC4360**, BGP Extended Communities Attribute, Feb. 2006