RIPE NCC
RIPE NETWORK COORDINATION CENTRE

Internet Data Analysis with RIPE RIS

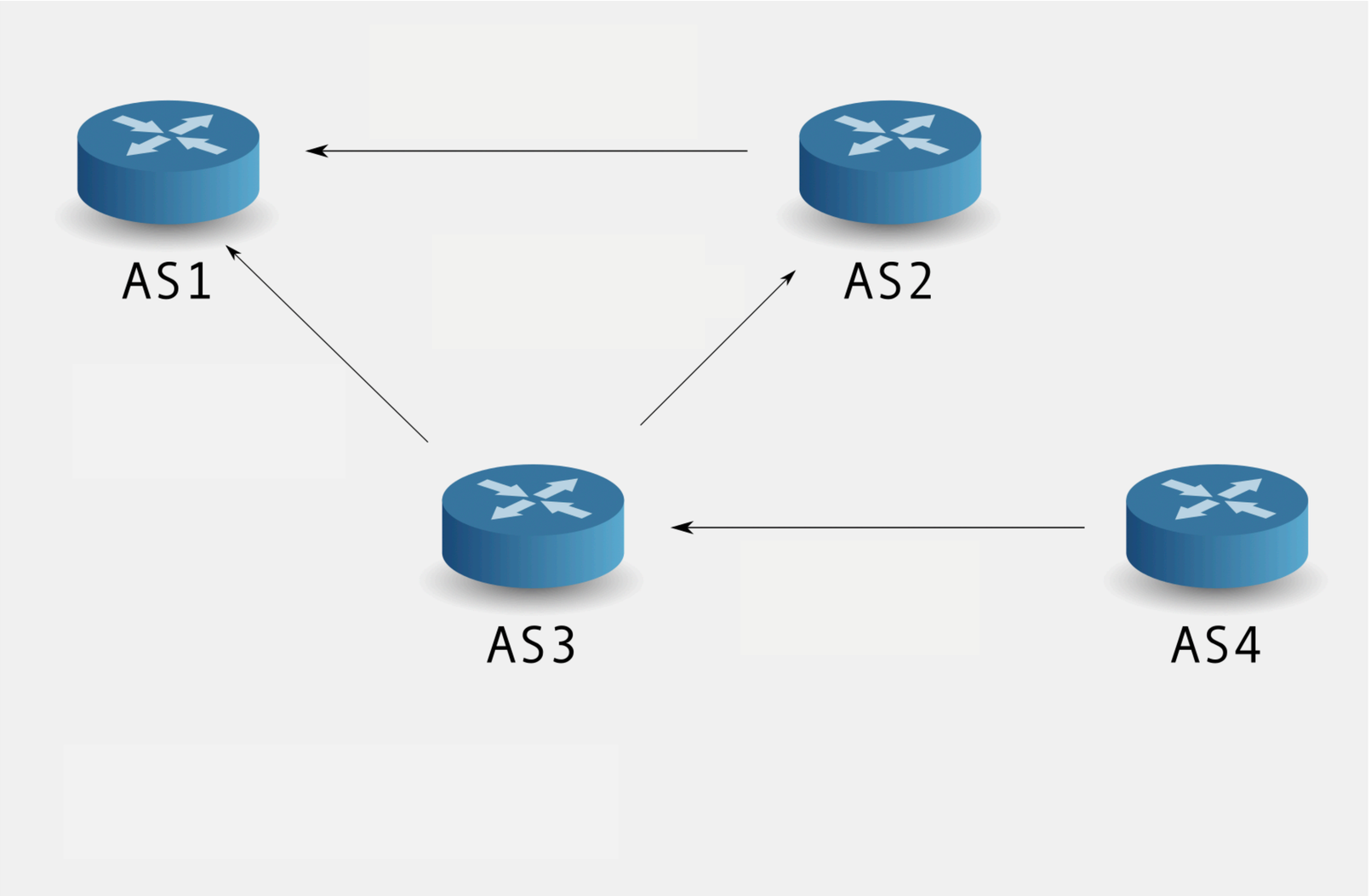
BGP Crash Course

Emile Aben | 2022 | TMA PhD School

What Is BGP?

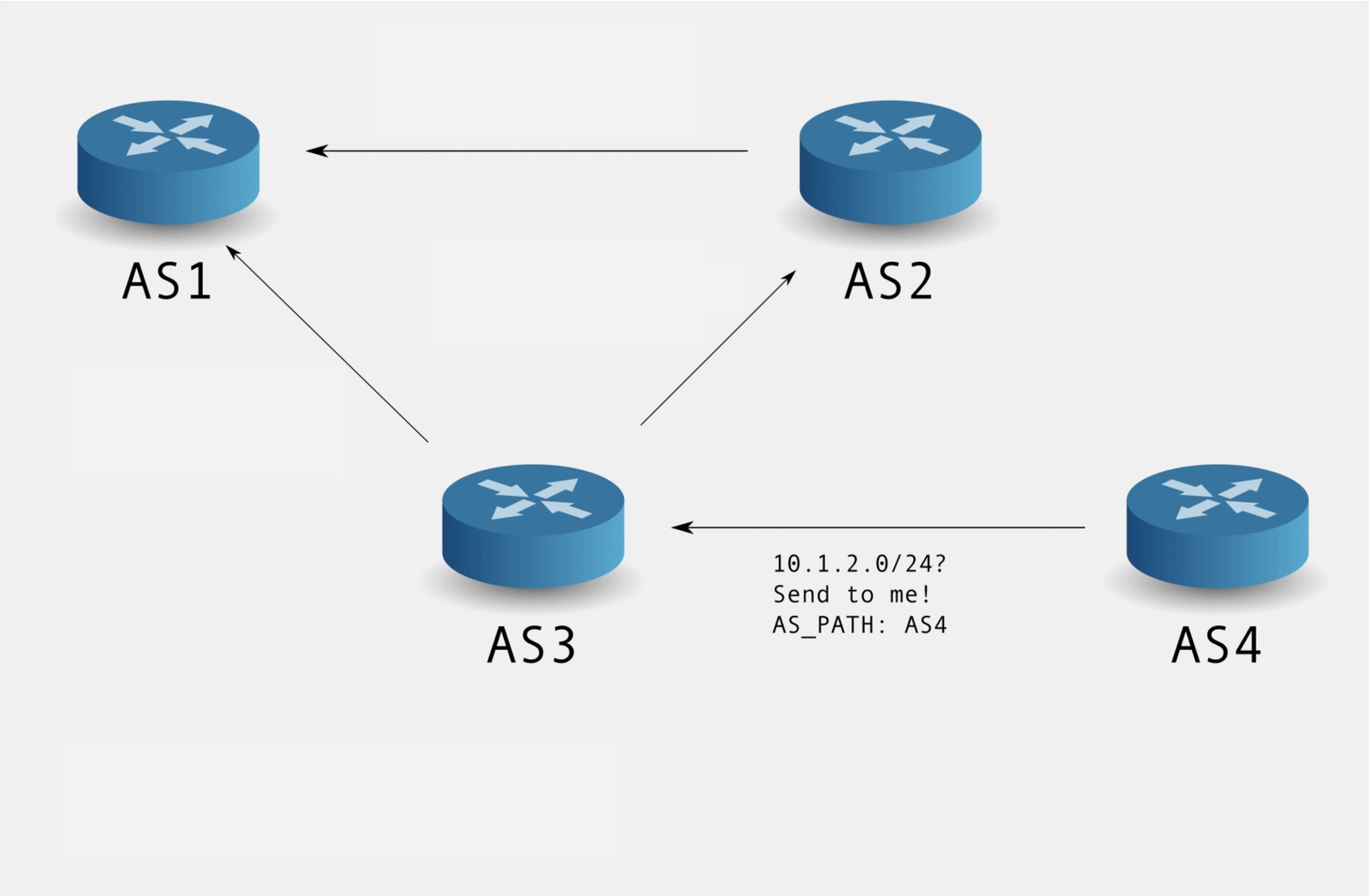
- BGP = Border Gateway Protocol
- The Internet control plane
- Distributes Internet routing information between routers
  - iBGP: Routers within the same administrative domain
  - eBGP: Between different admin domains (Autonomous Systems, ASN)
- BGP deals with external routing, other protocols deal with internal routing
  - Internal Gateway Protocol (IGP)
  - Examples: OSPF, IS-IS

How It Works (Simplified)



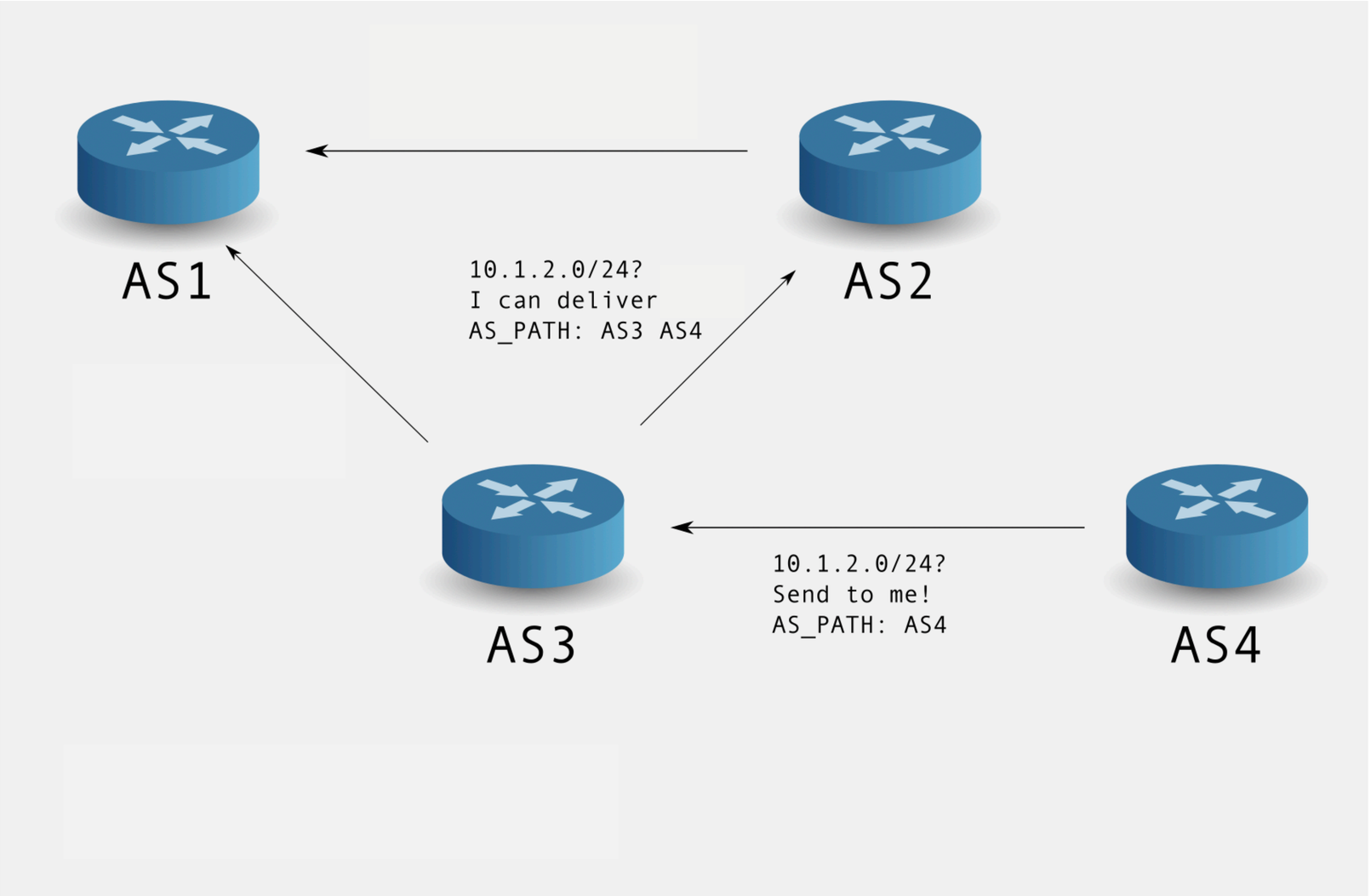
10.1.2.0/24

How It Works (Simplified)



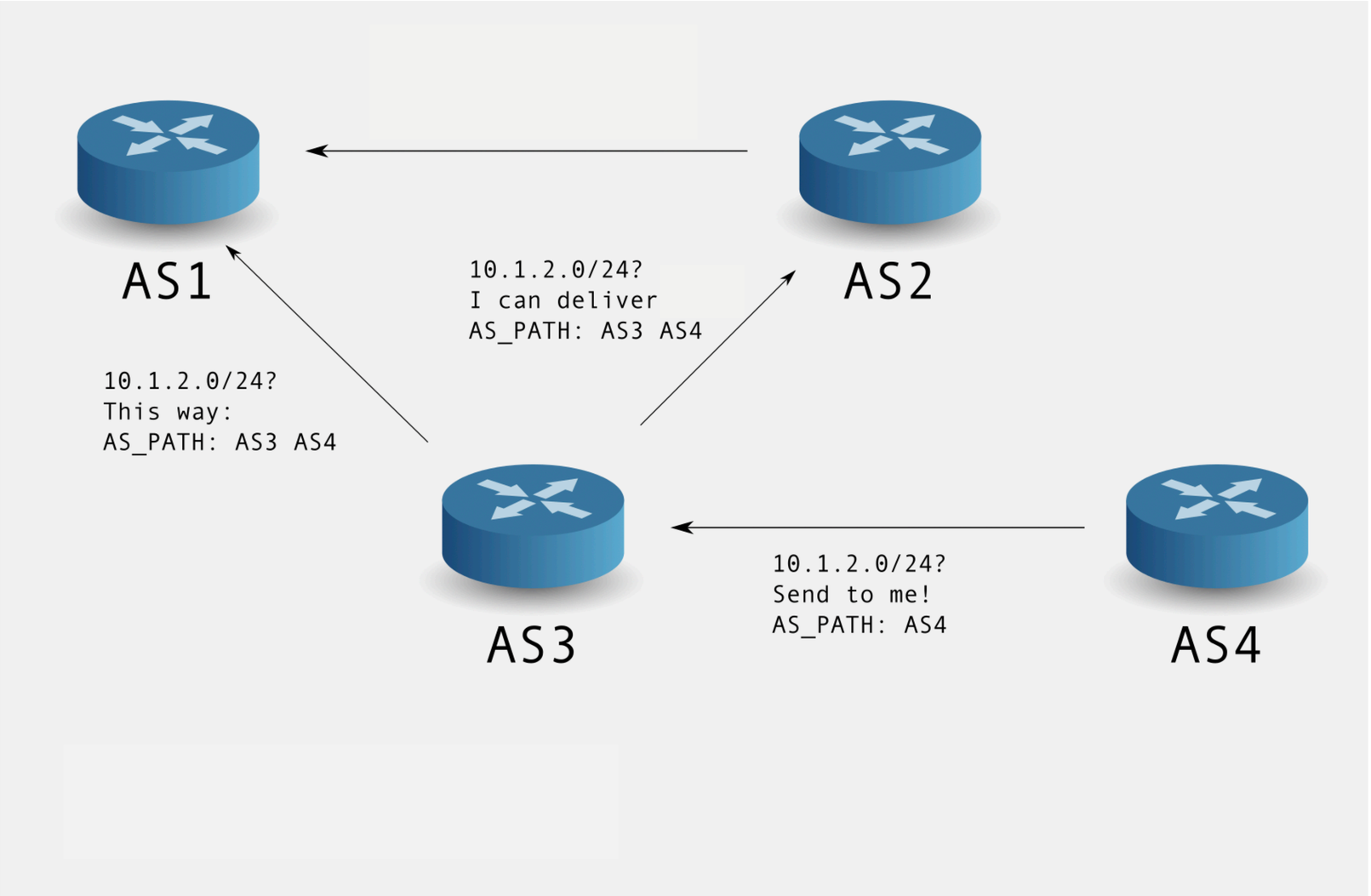
10.1.2.0/24

How It Works (Simplified)



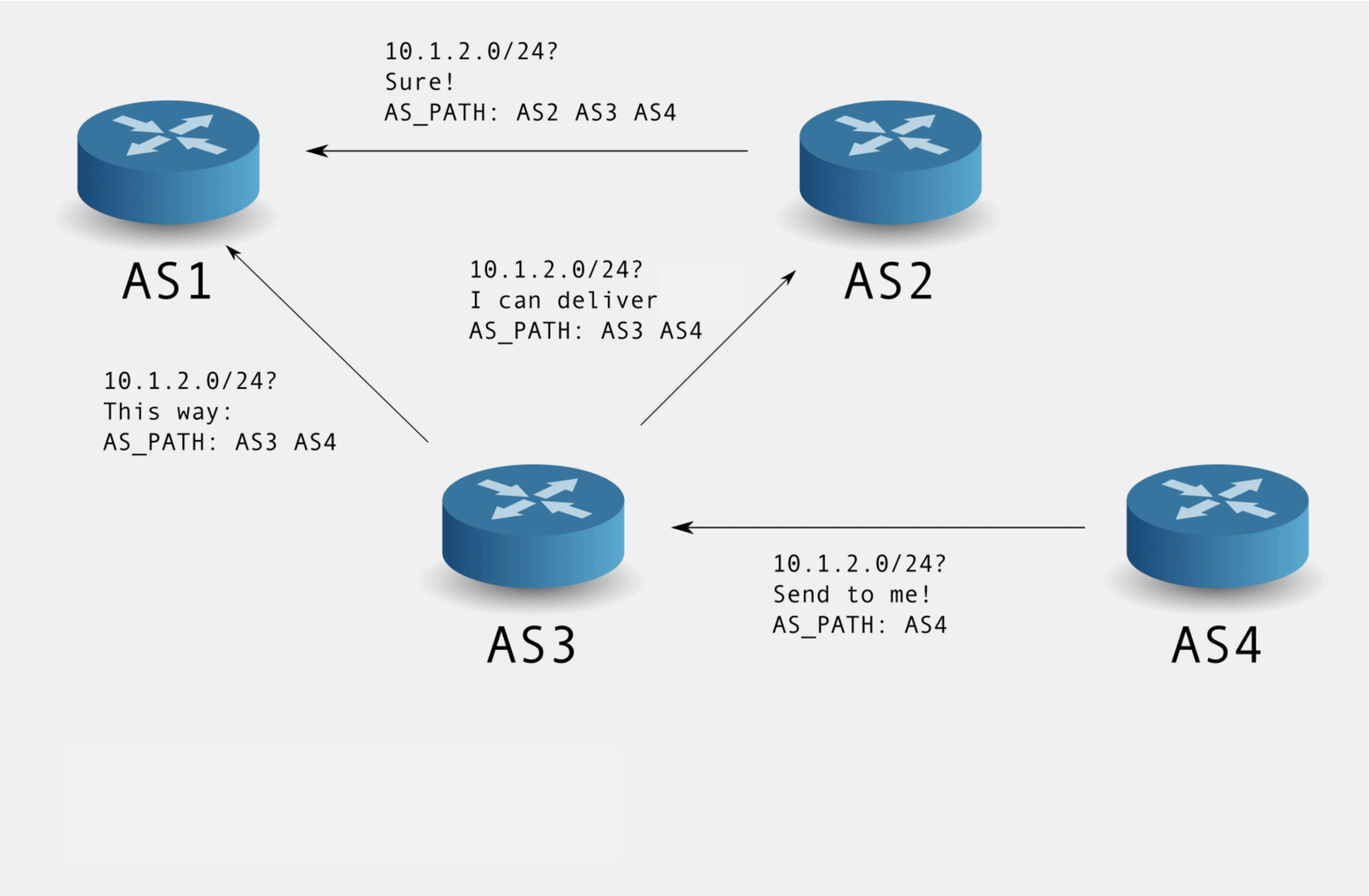
10.1.2.0/24

How It Works (Simplified)



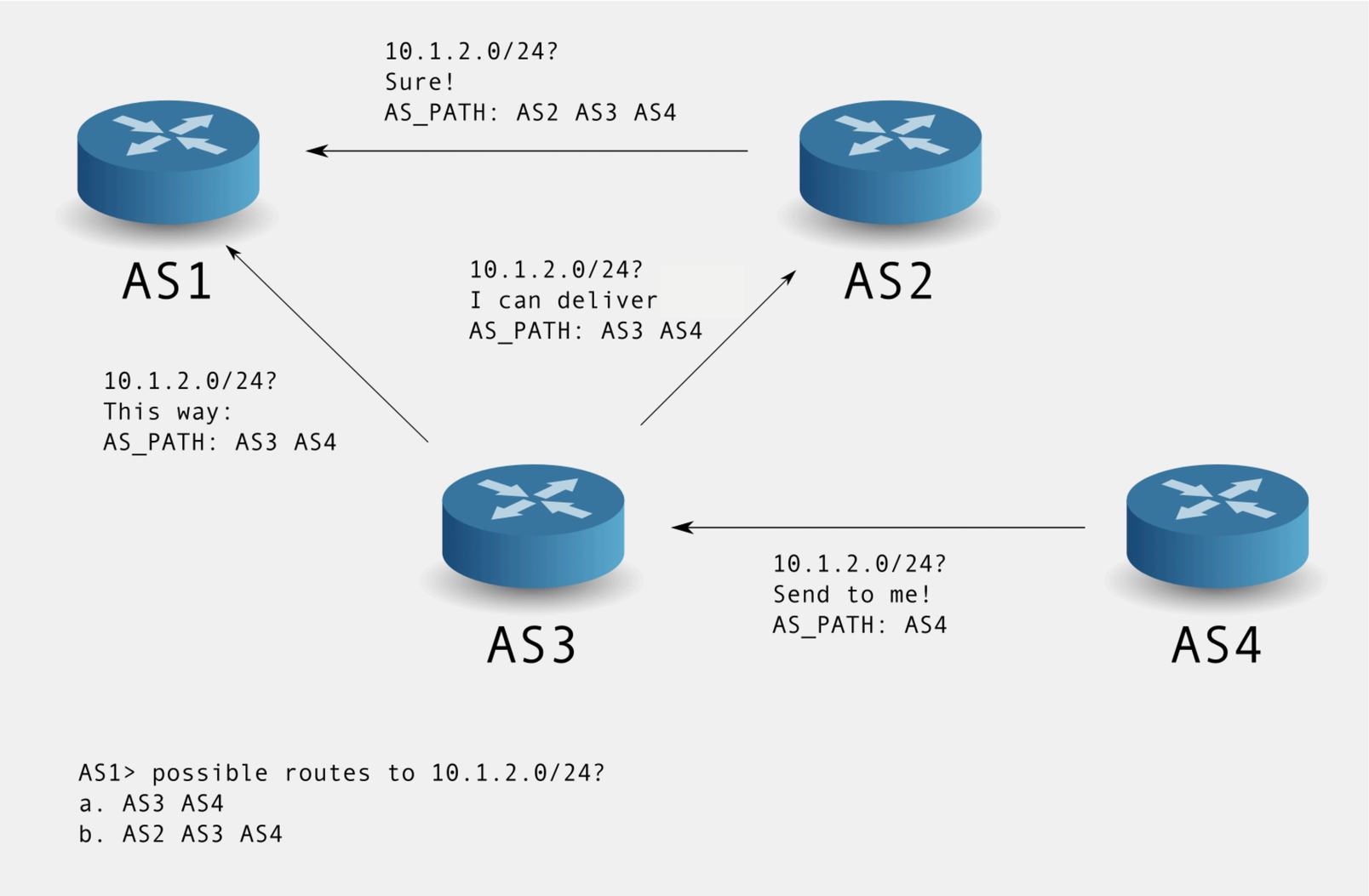
10.1.2.0/24

How It Works (Simplified)



10.1.2.0/24

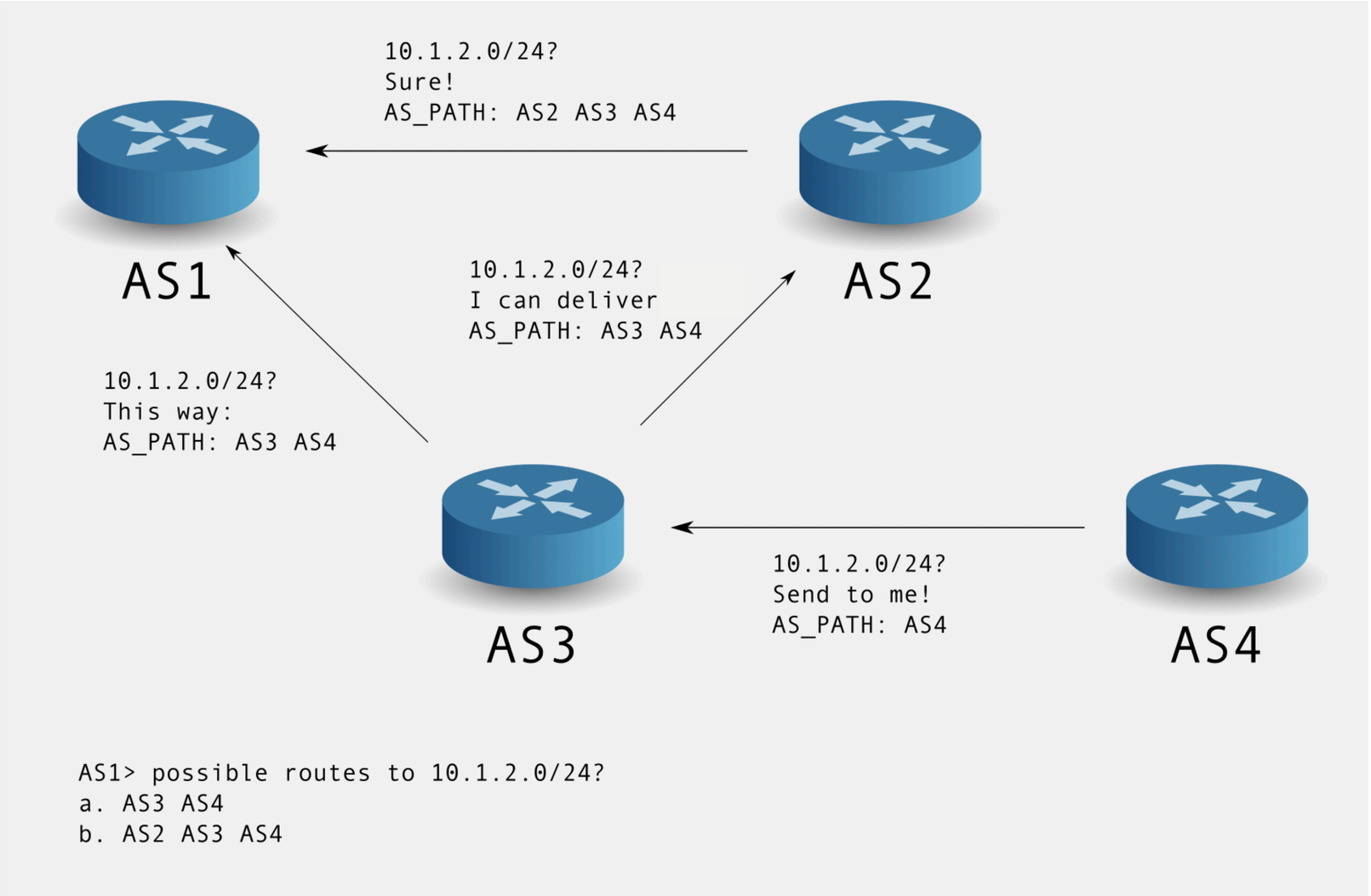
How It Works (Simplified)



10.1.2.0/24

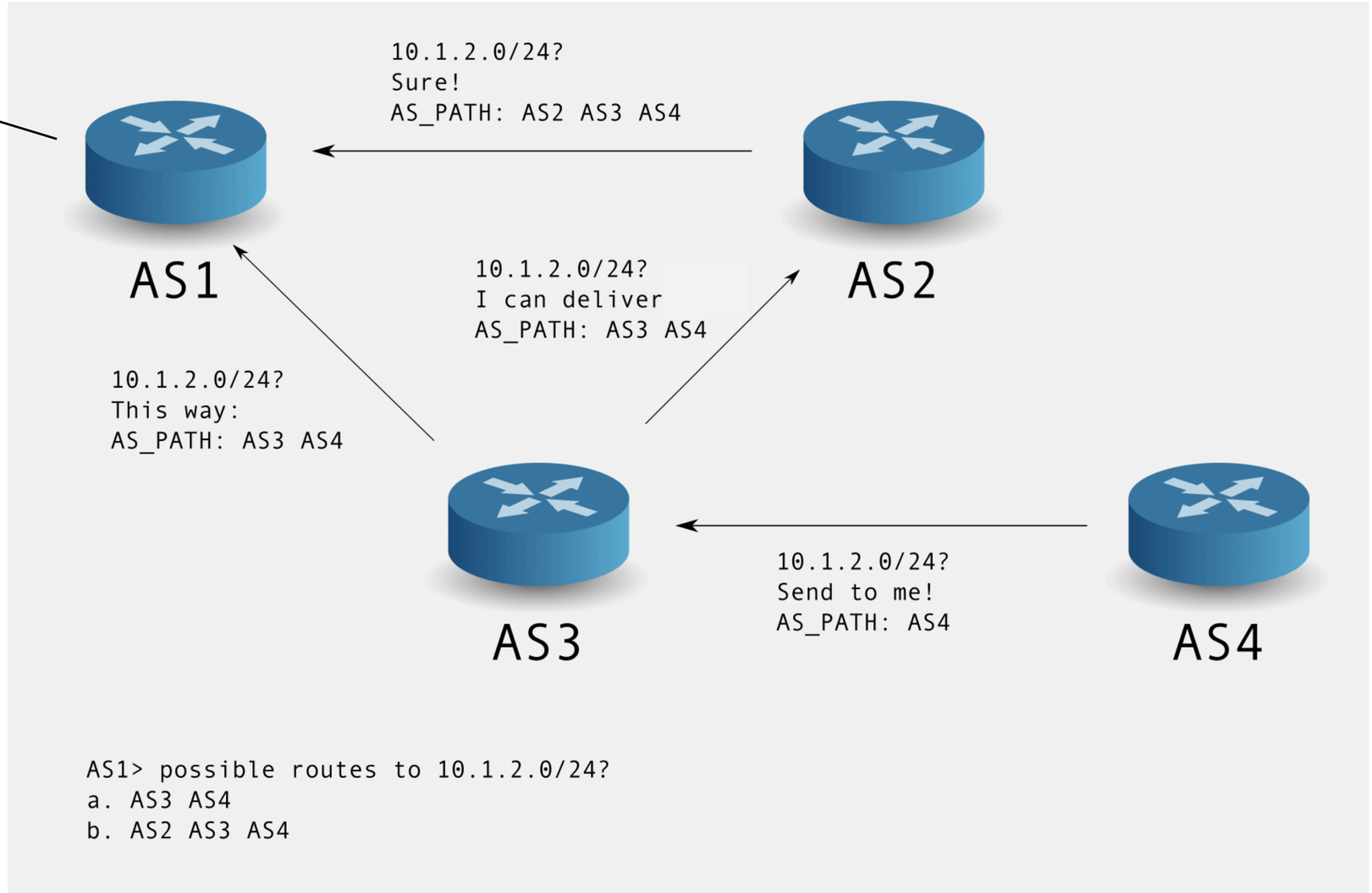
How It Works (Simplified)

?



10.1.2.0/24

How It Works (Simplified)



10.1.2.0/24

BGP Route Selection

- Best Path Selection Algorithm
  - How does AS1 select the best path for 10.1.2.0/24?

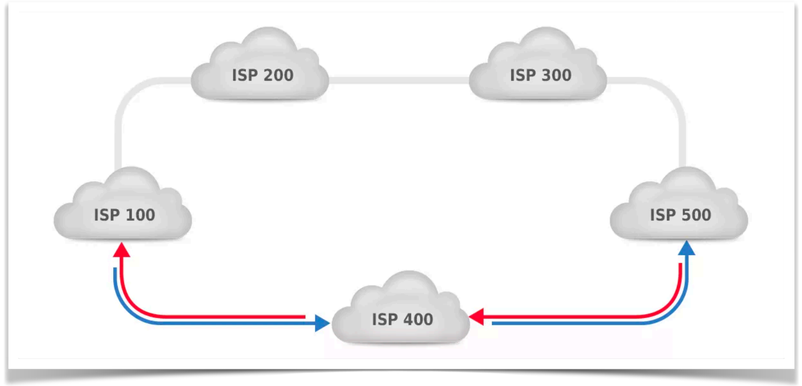
- Selection criteria
  1. Local preference (stays internal to a network)
  2. Shortest AS path
  3. Closest next hop router (aka hot potato routing)
  4. Additional criteria, breaking ties

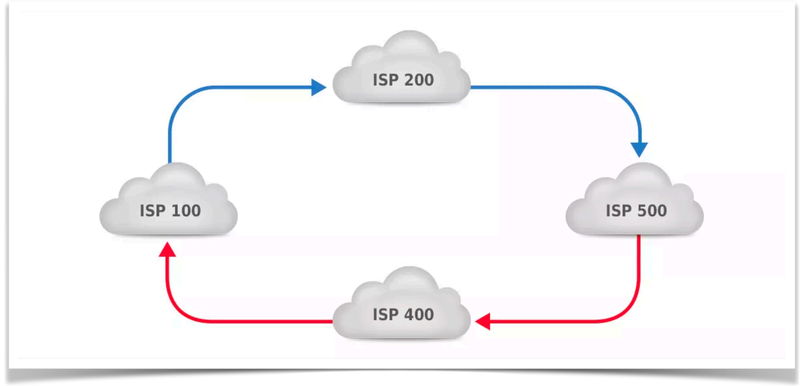
RIB vs. FIB

- Each router has a
  - Routing Information Base (RIB): Database of all routes + metadata (AS_PATH, etc.)
  - Forwarding Information Base (FIB): Database on how to forward packets
- Best route for a prefix gets installed into FIB
- RIB
  - 10.1.2.0/24 AS1 AS2 AS3 AS4 AS5 AS6 AS7
  - 10.1.2.0/24 AS3 AS4 AS5 AS6 AS7
  - 10.1.2.0/23 AS6 AS7
- How will traffic for 10.1.2.1 be routed?
- FIB
  - 10.1.2.0/24 via AS3
  - 10.1.2.0/23 via AS6

Routing Symmetry?

- BGP routing does not have to be symmetric!

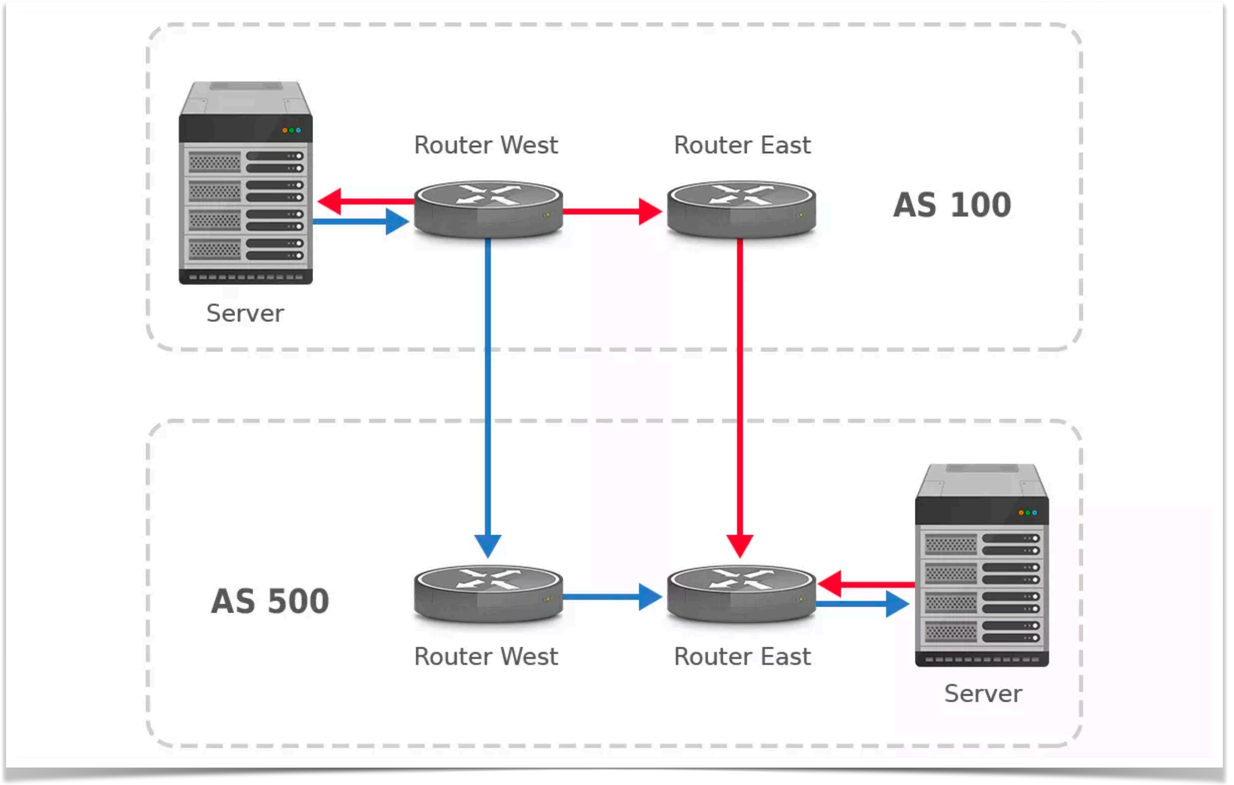




Images from: https://www.noction.com/blog/bgp-and-asymmetric-routing

Hot Potato Routing

- Get traffic off your network as quickly as possible
  - In BGP you only see the BGP adjacency
- Some networks do “cold potato”
  - Costs more
  - You have better control of the quality of the path (packet drop, latency)

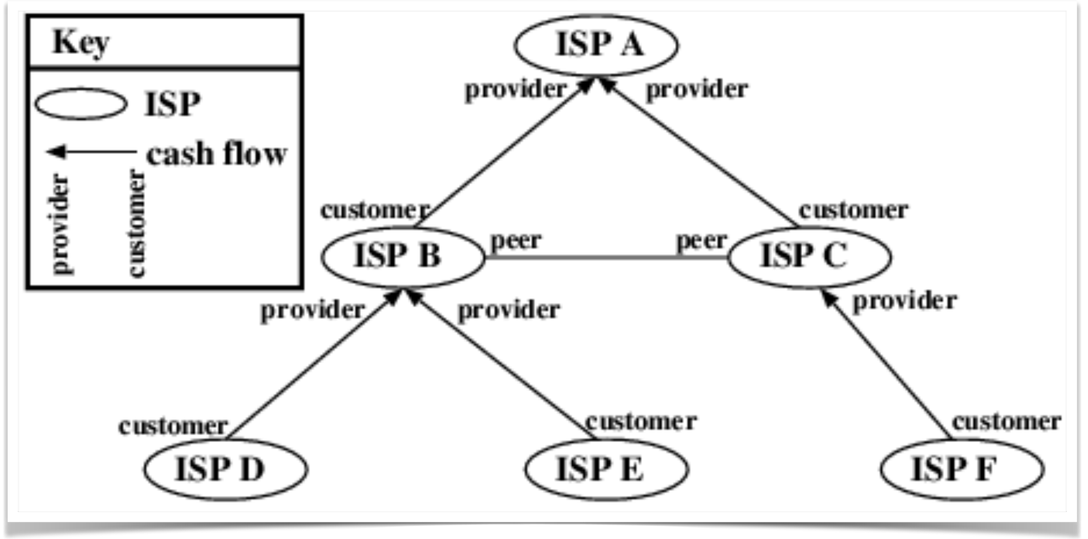


Images from: https://www.noction.com/blog/bgp-and-asymmetric-routing

Local Policy Matters

- Customer - Provider
  - Money in exchange for access to (typically) all of the Internet
- Peer (“bypass”)
  - (Typically) no money, mutual exchange of a limited set of Internet routes

Received From	Advertised To
Customer	Everyone
Peer	Customer
Provider	Customers



“BGP Is An Information Hiding Protocol”

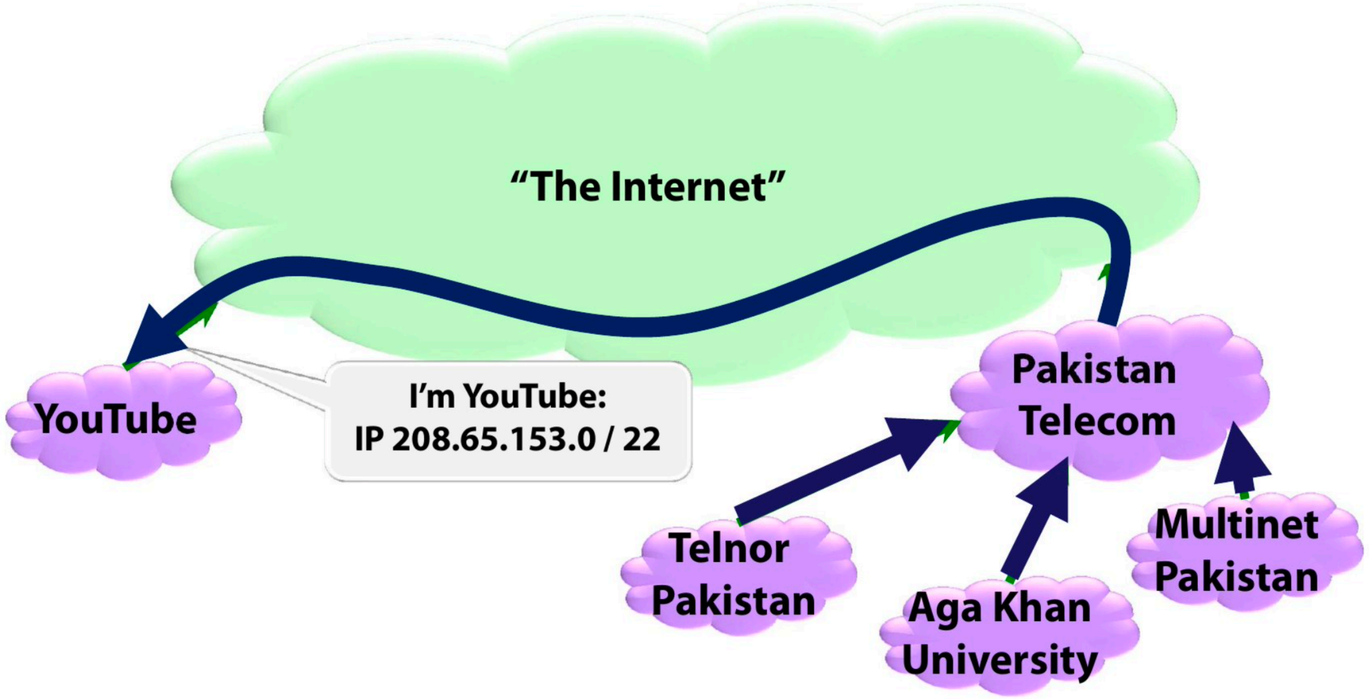
- BGP will not provide you with a full picture of all available routes!
  - You don’t see default routes, static routes
  - BGP speakers do not always forward paths for a prefix they know
    - Local policy ($$)
    - Filtering (RPKI, IRR)
  - If a BGP speaker forwards, it forwards only best path, not all
    - The further you are from a BGP router, the more chance you miss available routes when collecting BGP data
- BGP scales well

BGP Is Simple and Complex

- Protocol itself is relatively simple
- Internet: 900k+ IPv4 prefixes, 150k+ IPv6 prefixes, 70k+ ASNs
- Routing policy that can be created with it is complex
- Large ASNs can (and will!) have policies that are different between different parts of their networks

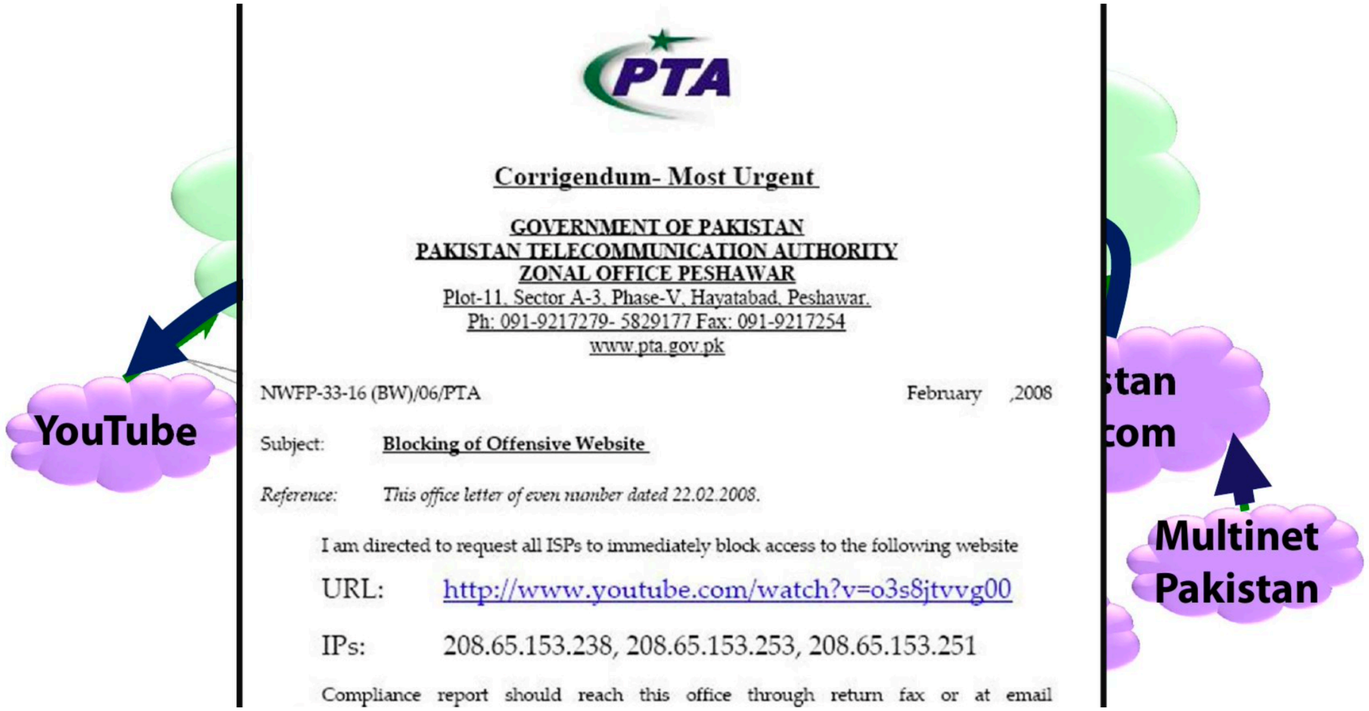
Example: Youtube Hijack

- Usual route to Youtube (AS36561)



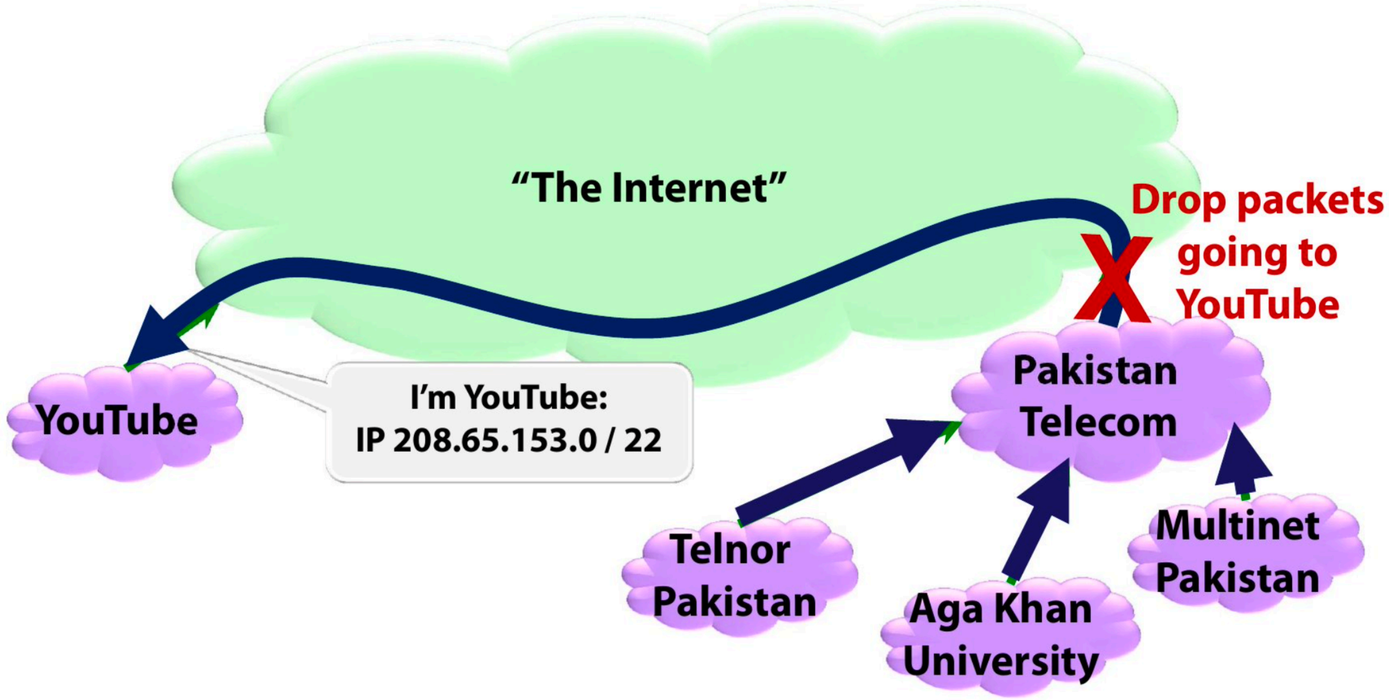
Example: Youtube Hijack

- Pakistan Government orders block of a YouTube video



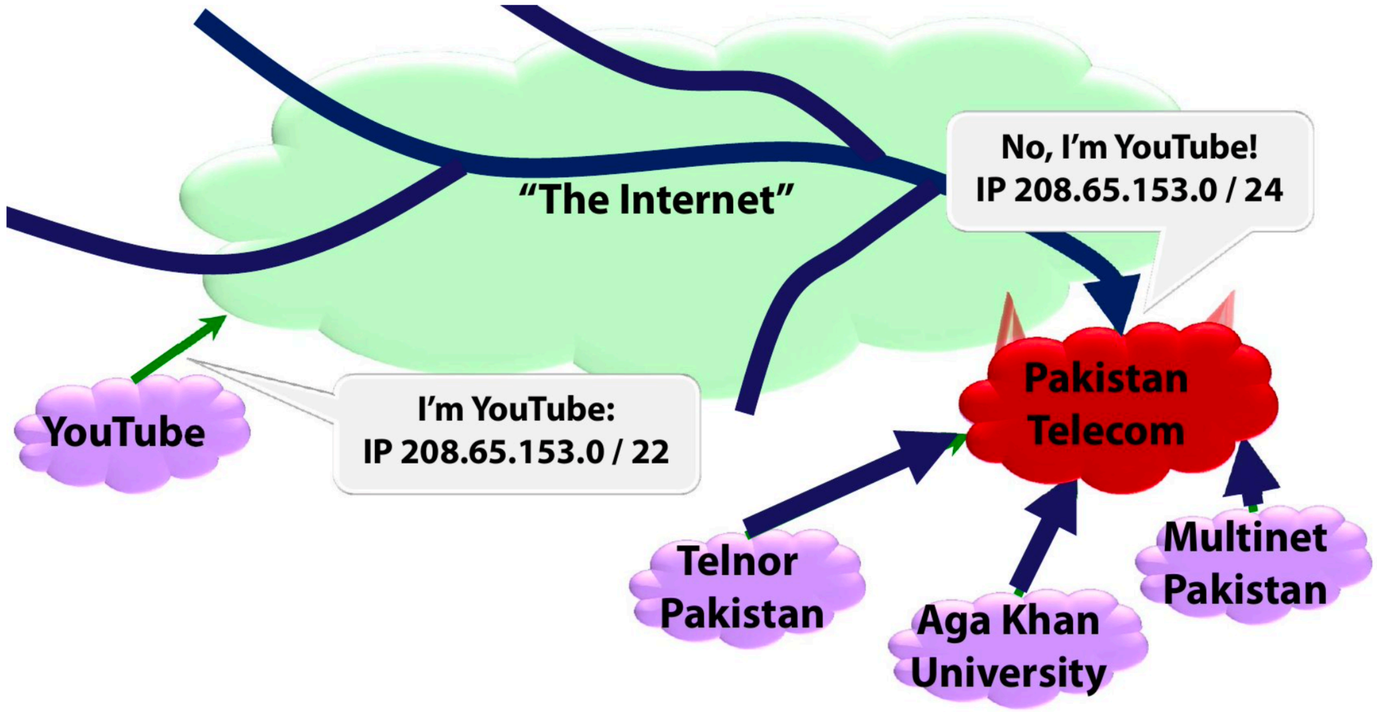
Example: Youtube Hijack

- What probably should have happened



Example: Youtube Hijack

- What PK telecom ended up doing



Timeline

Timeline: Sunday, 24 February 2008

- Before, during and after the event: AS36561 (YouTube) announces 208.65.152.0/22 and other prefixes
- **18:47** AS17557 (Pakistan Telecom) announces 208.65.153.0/24. Routers around the world redirect YouTube traffic to Pakistan.
- **20:07** (YouTube) announces 208.65.153.0/24. BGP policy rules, such as preferring the shortest AS path, determine which route is chosen. (Pakistan Telecom) continues to attract some of YouTube's traffic.
- **20:18** (YouTube) announces 208.65.153.128/25 and 208.65.153.0/25. Every router that receives these announcements will send the traffic to YouTube.
- **21:01** AS3491 (PCCW Global) withdraws all prefixes originated by AS17557 (Pakistan Telecom), thus completely stopping the hijack of 208.65.153.0/24.

1 hour 30min. of downtime reported by users in Germany, China, US, Russia, the UK, and Australia