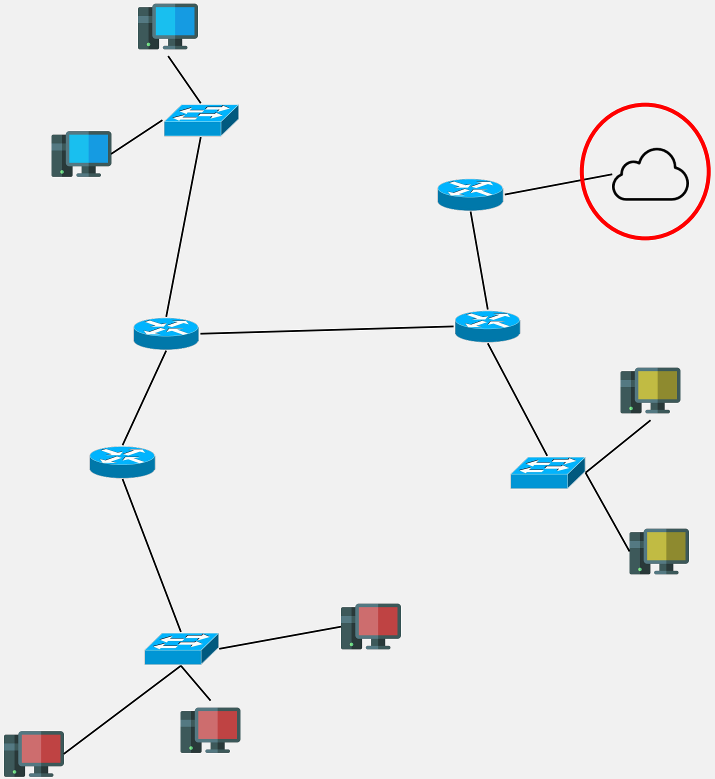
Network Security

Knowing and Taking the Path: Border Gateway Protocol

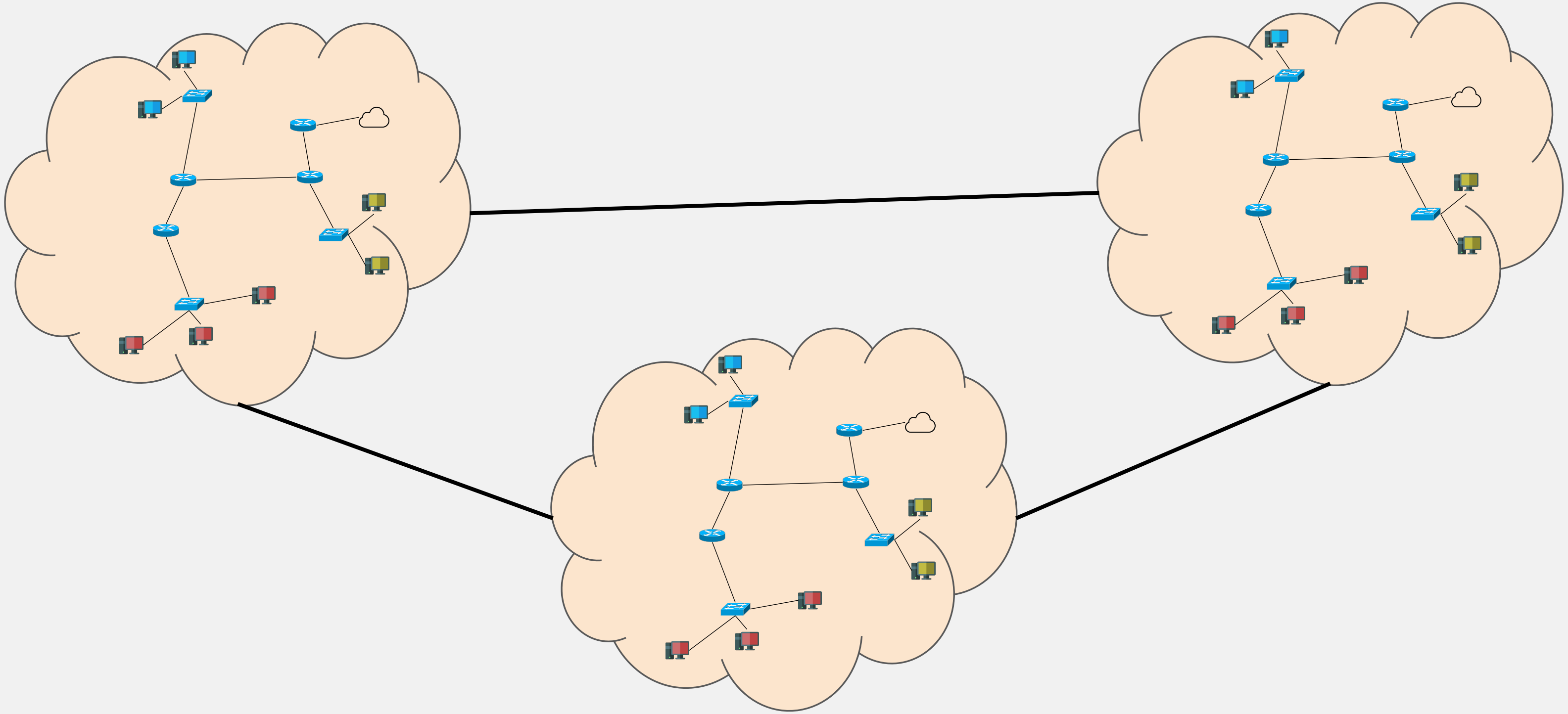
Gwendal Patat
Univ Rennes, CNRS, IRISA

2025/2026

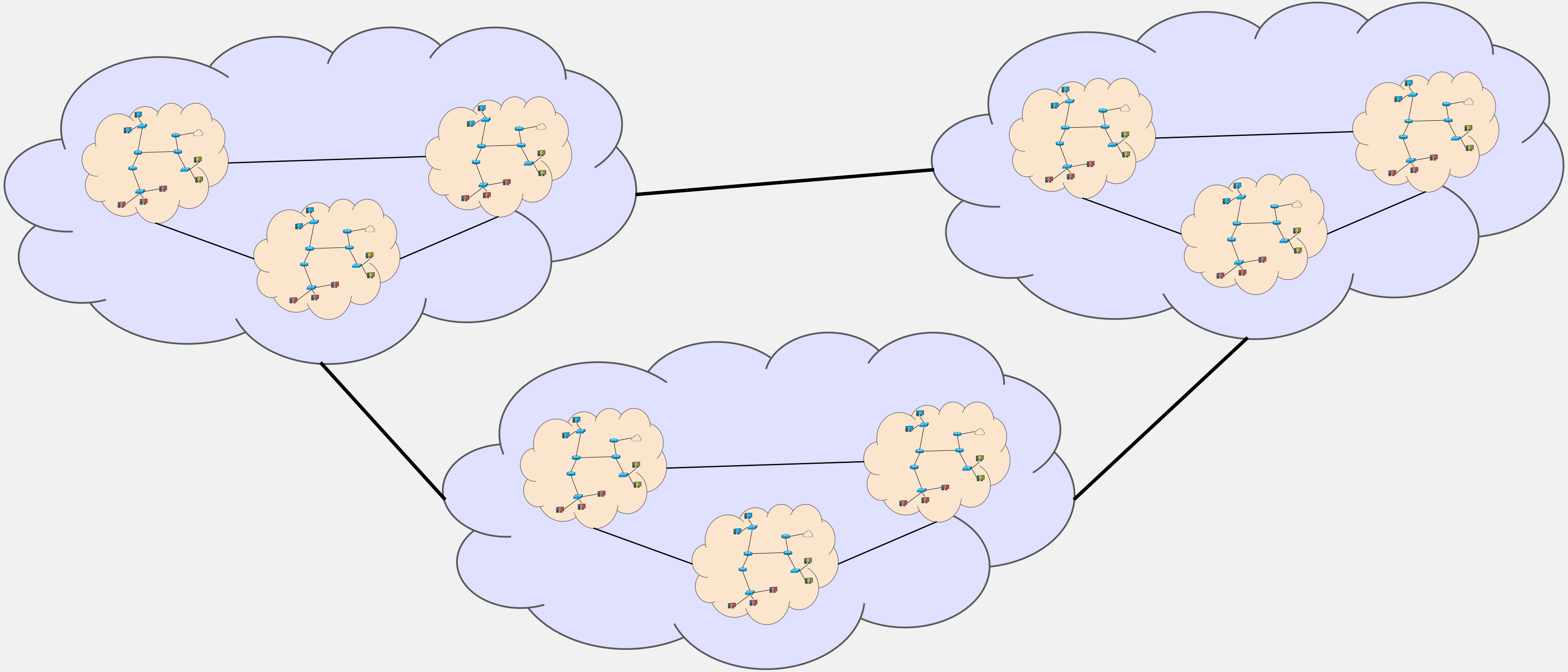
Where are we now?



Where are we now?

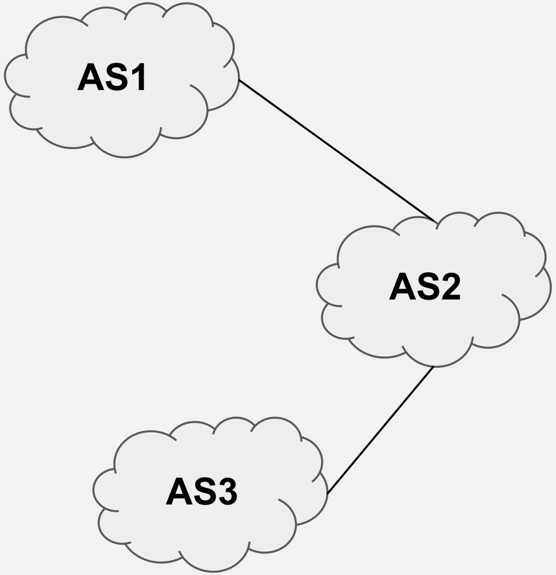


Where are we now?



Autonomous System (AS) 1/2

- AS: A network controlled by a **single entity** such as an ISP, government, company, etc.
- This network uses the same internal logic for routing packages.
- ASes are identified by an unique ID: the **AS number (ASN)**
  - ASN: 16-bit identifier, now extended to 32-bit.
  - Assigned by the IANA (like IP address ranges).

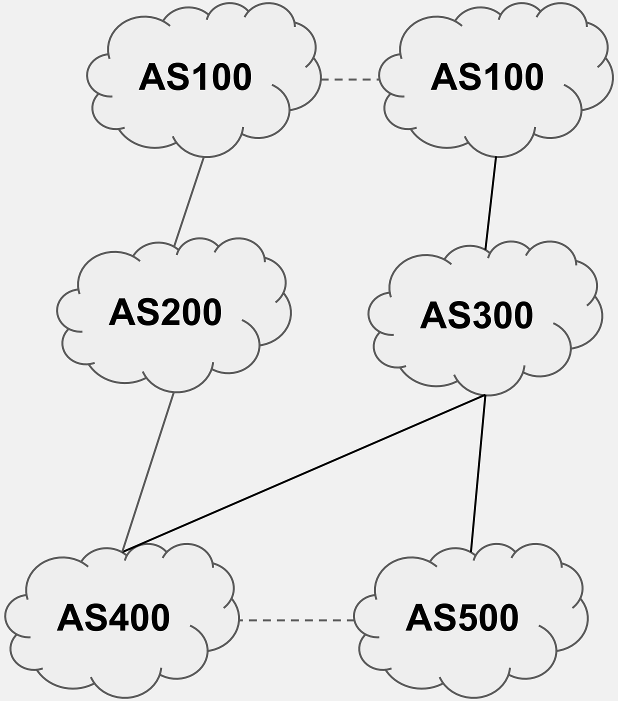


Autonomous System (AS) 2/2

ASes can greatly differ in size or goals, and are organized in a hierarchic way.

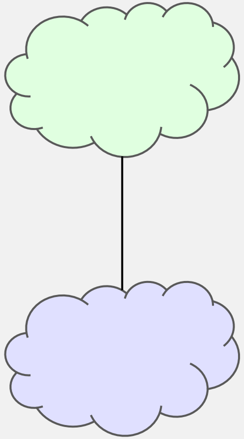
ASes can be either:

- **Client**: pays to gain access to transit to bigger ASes.
- **Provider**: gets paid to transit packets.
- **Peer**: used for transit between neighbors.

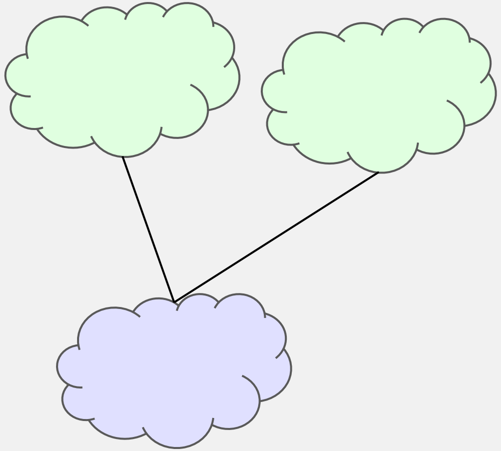


Type of ASes

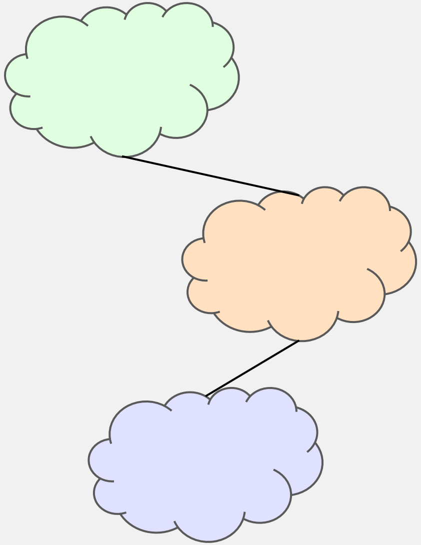
Stub



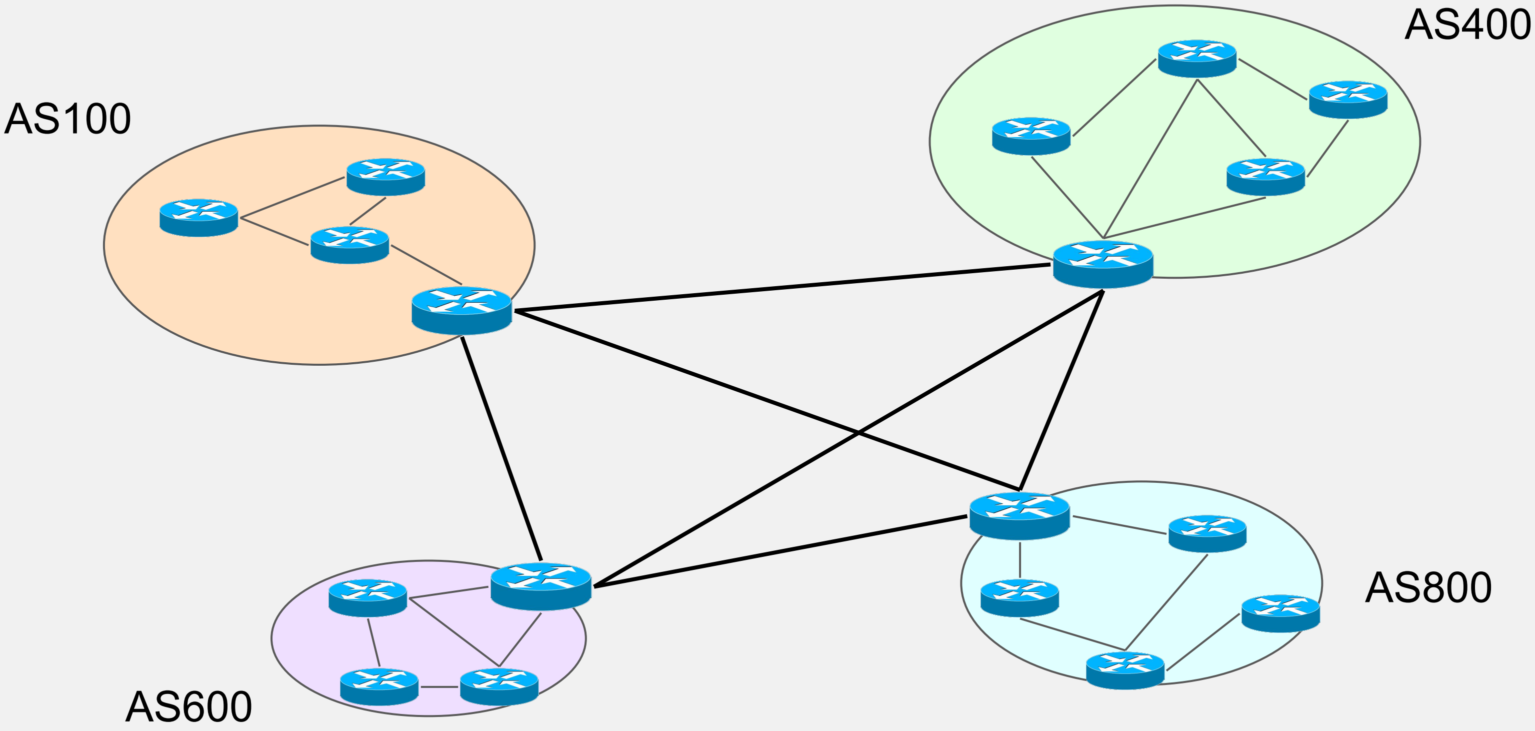
Multihomed



Transit



Interior and Border Routers



Interior and Border Routers

Interior Router:

- The regular routers that we already mentioned before.
- Do not support connection with another AS.
- Used to deliver packages inside the same AS.

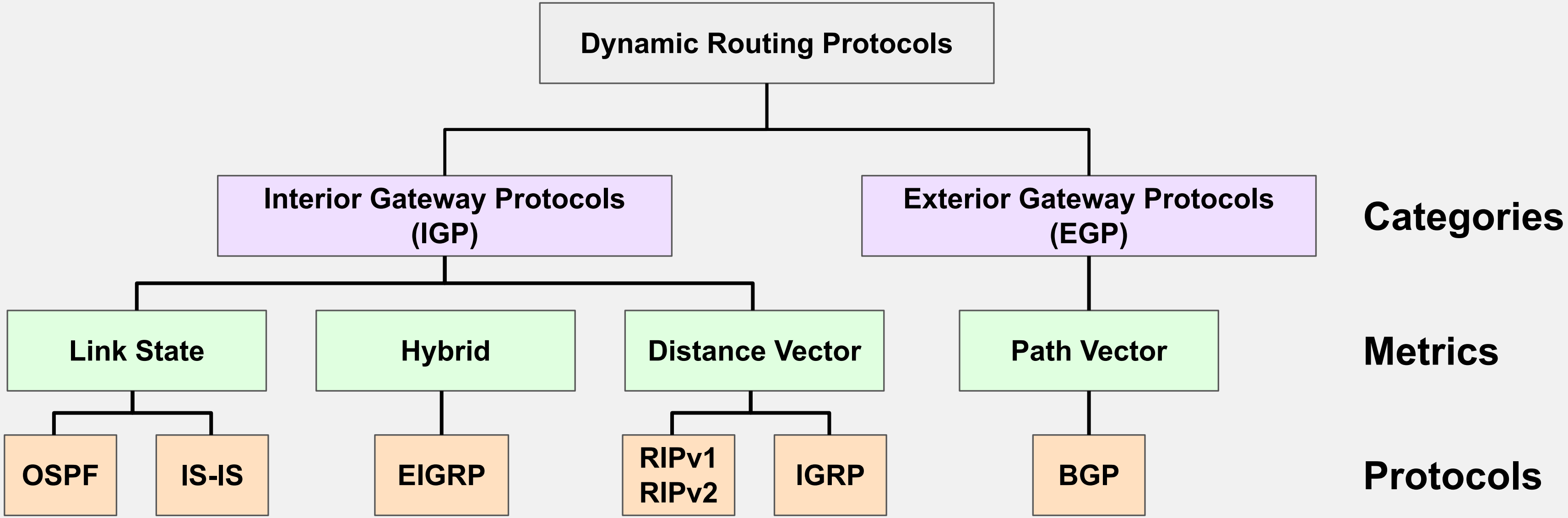
Border Router:

- Used to deliver packages between AS.
- They need to know about their ASN and AS concept to work properly.
- Use a different routing protocol then interior routers.

Dynamic Routing Protocols

Interior Gateway Protocols (IGP): Routing protocols used within an AS.

Exterior Gateway Protocols (EGP): Routing protocols used to connect ASes together.



Border Gateway Protocol

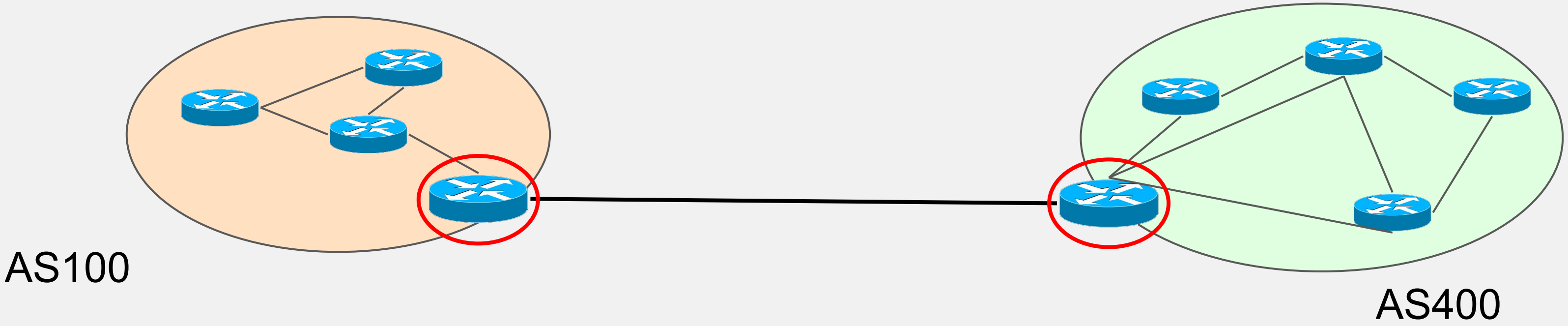
Border Gateway Protocol

Border Gateway Protocol (BGP):

- Introduced in 1989 with RFC 1105 and used since 1994.
- Current version is **BGP-4** from 1994, updated with the RFC 4271 from 2006.
- Protocol to exchange routing information between ASes.
  - **Path Vector**: Each router has a list of paths to other ASes (ASes route to a specific AS).
    - ***We are not here to find the shortest path!***
  - **Application Layer** protocol since it is above the pure IP routing of layer 3 and works over TCP.
  - IPv4 and IPv6 compatible.
  - Can be used within the same AS: In this case, called iBGP (internal (or interior) Border Gateway Protocol).

BGP Peers and Speakers

- Connected border routers are called **Peers.**
  - **Static** configuration of the neighbors' IP addresses.
  - The peering operation is configured as a **TCP** connection on port 179.
    - This is called a BGP session.
  - Every 30 seconds, the connection is kept alive with a 19-byte long keepalive message from one of the router.
- In BGP, a router sending protocol messages is called a **BGP speaker**.

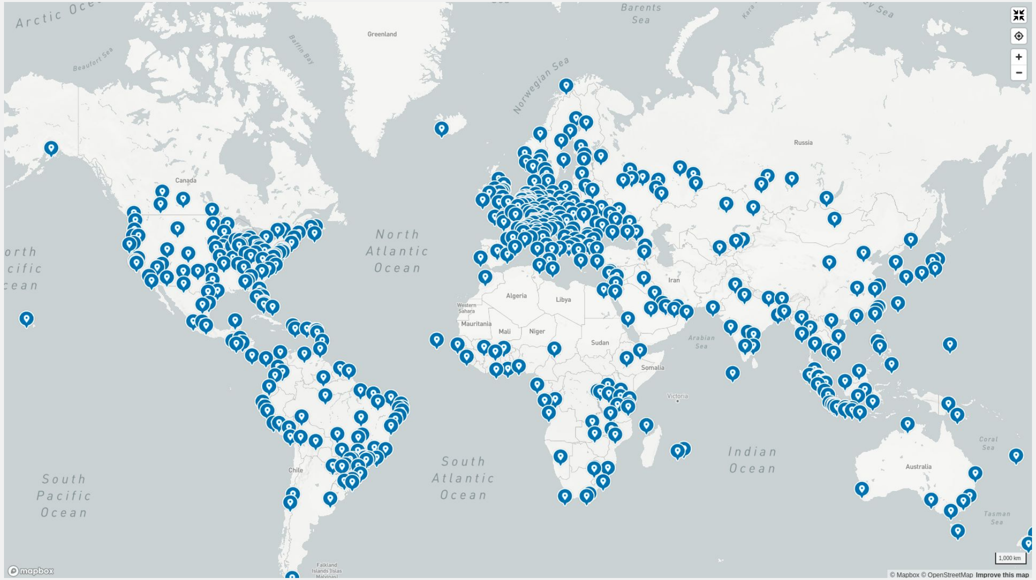


Peering and Internet Exchange Point

The peering process happens either:

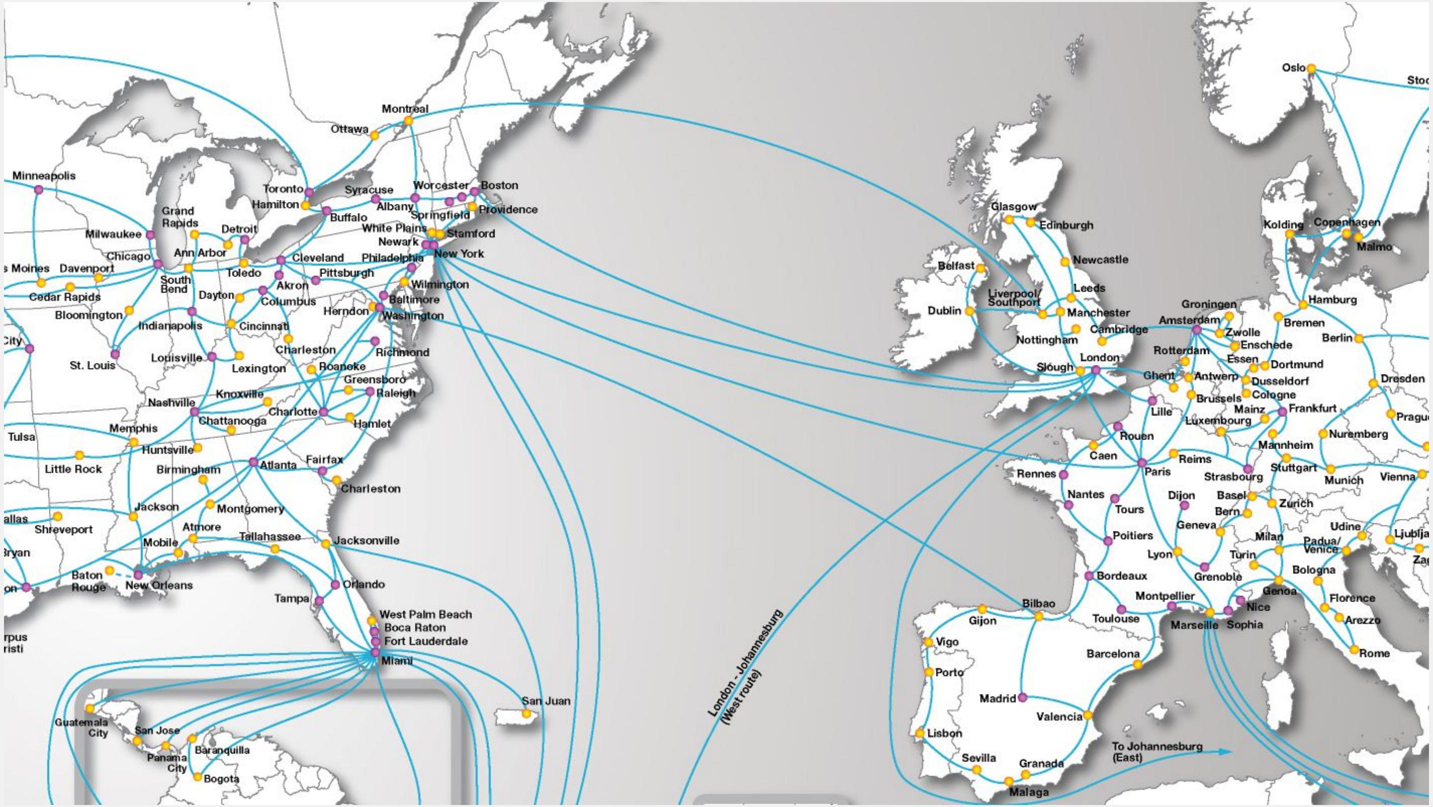
□ **In Private within Data Centers.**

□ **Using Internet Exchange Point (IX/IXP)**



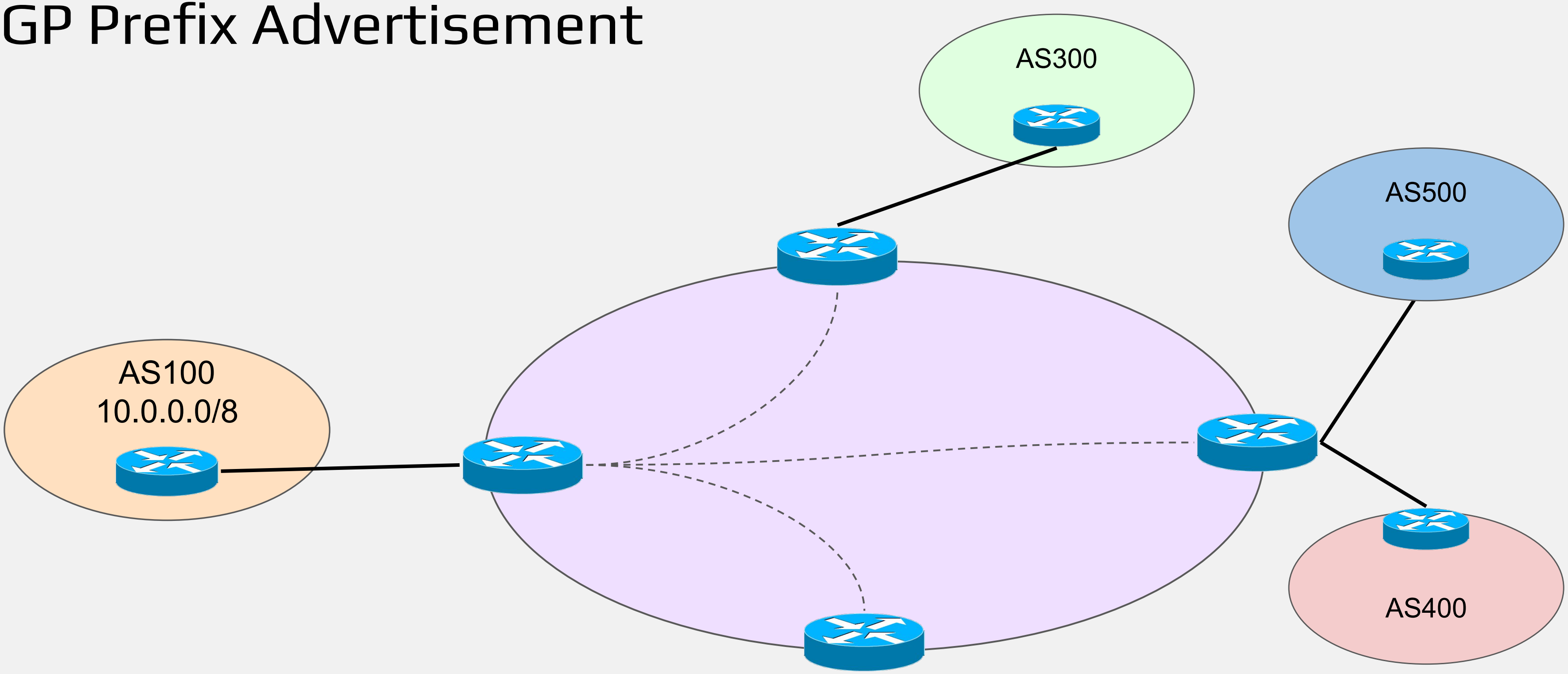
https://www.datacentermap.com/ixp/

Example: AS174 Cogent



https://www.cogentco.com/fr/network/network-map

BGP Prefix Advertisement

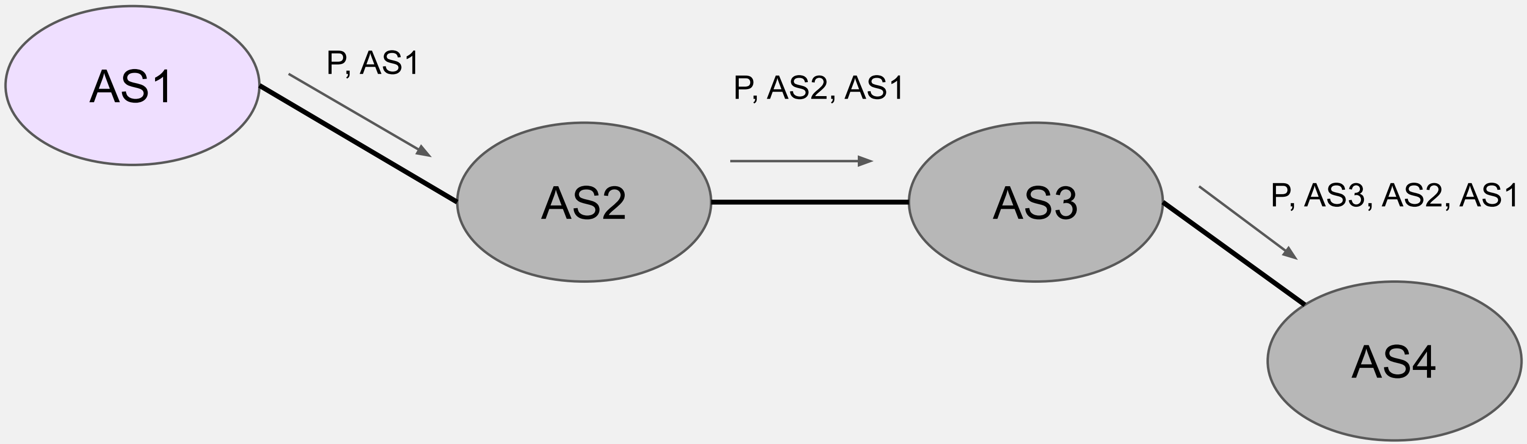


BGP Prefix Advertisement

Using a BGP Update message:

- The AS sends to all its neighbors its IP prefixes and its ASN.
    - This create an **AS Path** of one node.
- The next AS will disseminate this information by adding their ASN to the AS Path.
    - If multiple AS paths are received for the same prefix, the AS selects **only one for dissemination** but keeps the others in its BGP table.

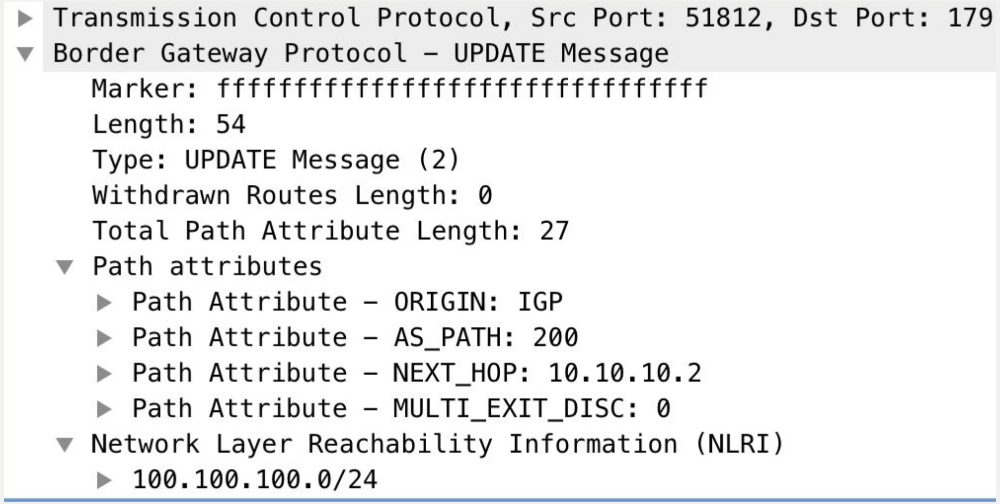
AS Path



An AS could also add its ASN multiple times to the AS Path, doing what we call AS Path Prepending, to make the path less desirable: e.g., for more expensive routes, backup link.

BGP Update Message

- **Withdraw**: Delete the route.
- **Path Attributes:**
  - Origin: interior or exterior.
  - AS Path.
  - Next Hop: src IP of the BGP speaker.
- **NLRI:**
  - IP prefix at the end of the path.



```
▶ Transmission Control Protocol, Src Port: 51812, Dst Port: 179
▼ Border Gateway Protocol - UPDATE Message
    Marker: ffffffffffffffffffffffffffffffff
    Length: 54
    Type: UPDATE Message (2)
    Withdrawn Routes Length: 0
    Total Path Attribute Length: 27
  ▼ Path attributes
    ▶ Path Attribute - ORIGIN: IGP
    ▶ Path Attribute - AS_PATH: 200
    ▶ Path Attribute - NEXT_HOP: 10.10.10.2
    ▶ Path Attribute - MULTI_EXIT_DISC: 0
  ▼ Network Layer Reachability Information (NLRI)
    ▶ 100.100.100.0/24
```

For a given prefix and AS Path, the dissemination will be done to all neighbors **not already in the AS Path**.

BGP Path Selection

BGP Tables include multiple criteria for path selection in the following order:

- Weight
- Local Preference
- Locally Generated or Aggregated
- Shortest AS Path (in terms of ASes, not router hops)

The selection can be due to economic agreements, load balancing, geopolitics, etc.

Security in BGP

- BGP links are authenticated using a share secret used within a TCP-MD5 digest (an HMAC-MD5 hash within TCP packets).
- IPsec can also be used for the connection between BGP routers.

Nevertheless, there is no real built-in security in BGP:

- Security was not a priority/a concerne in 1989.
- Based on the trust everyone model of the early internet.
- No prefix verification or path validation.

BGP Attacks

BGP Hijack - Pakistan Telecom and YouTube

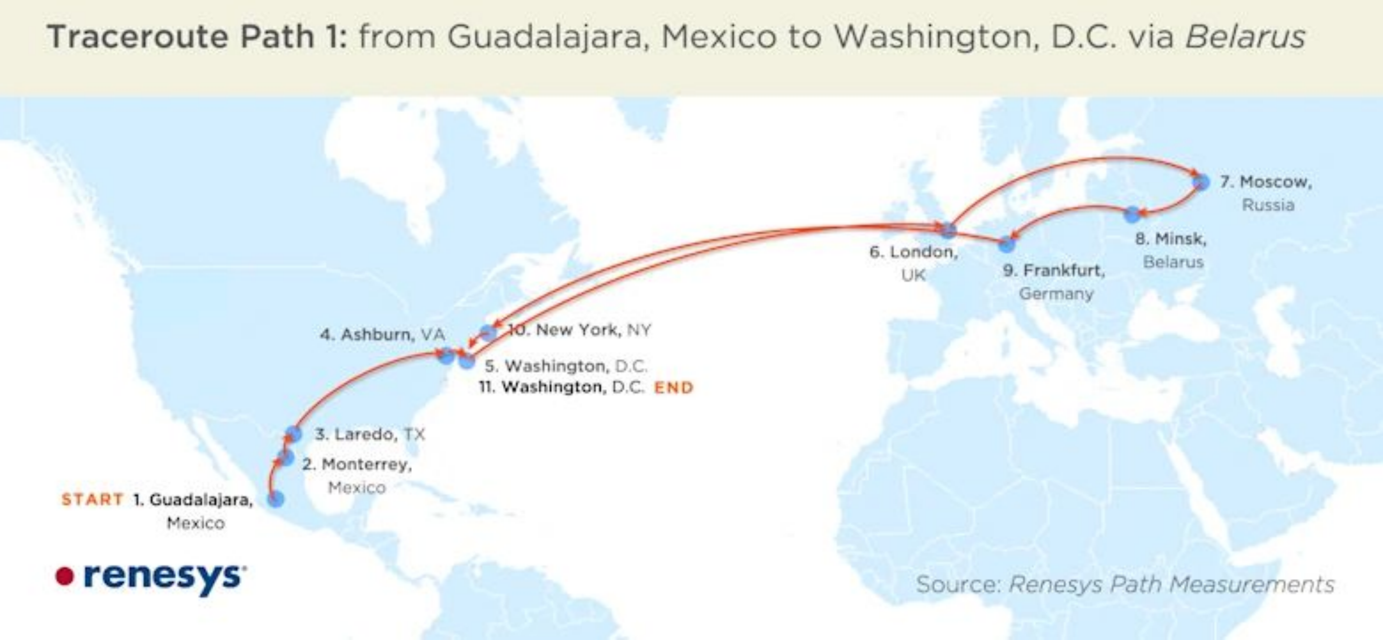
In 2008, during a national exam period, the Pakistan government asked the national telecom to block any access to YouTube by announcing a more specific prefix to the youtube IPs than the original.

By announcing a /24 instead of the /22 of YouTube at the time, all traffic was hijacked and dropped by Pakistan Telecom.

The problem is: we are not talking about a national black hole but **a worldwide one with path propagation!**

MitM Attack on BGP

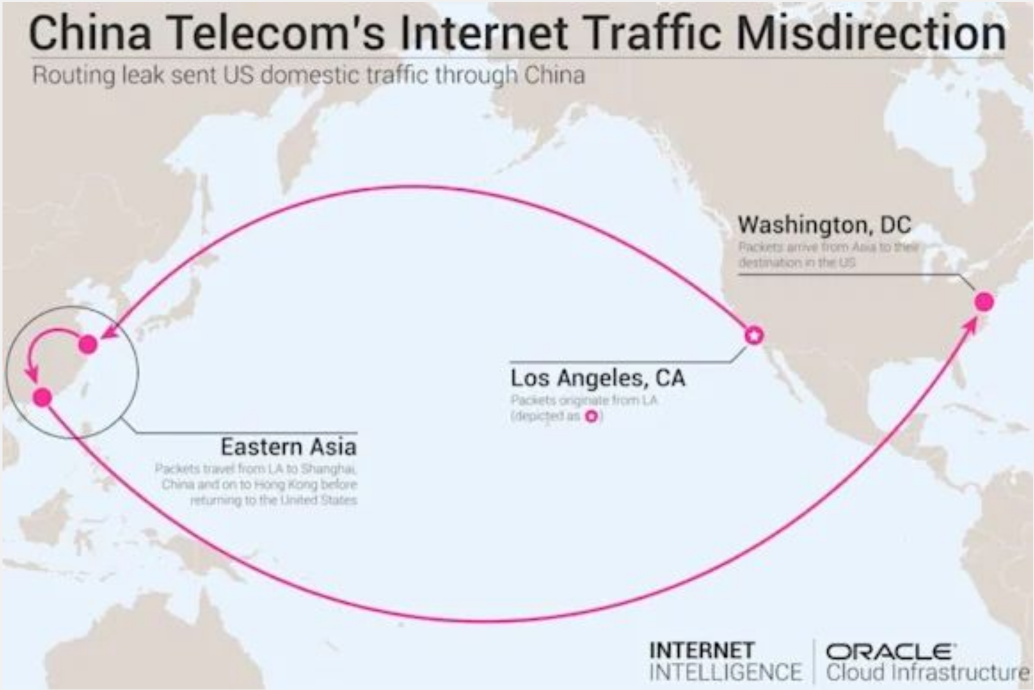
In 2013, a MitM attack has been performed to redirect US credit card companies and government traffics to Belarus.



Source: *Renesys Path Measurements*

Misredirection

In 2017, a routing leak caused some **US-to-US traffic** and, for more than two years, part of the global traffic destined for Verizon’s Asia-Pacific network to be **misdirected through China Telecom** via a shared South Korean peer.


China Telecom's Internet Traffic Misdirection
Routing leak sent US domestic traffic through China
Washington, DC
Packets arrive from Asia to their destination in the US
Los Angeles, CA
Packets originate from LA (depicted as ✪)
Eastern Asia
Packets travel from LA to Shanghai, China and on to Hong Kong before returning to the United States
INTERNET INTELLIGENCE
ORACLE Cloud Infrastructure

Prefix Hijacking Attack

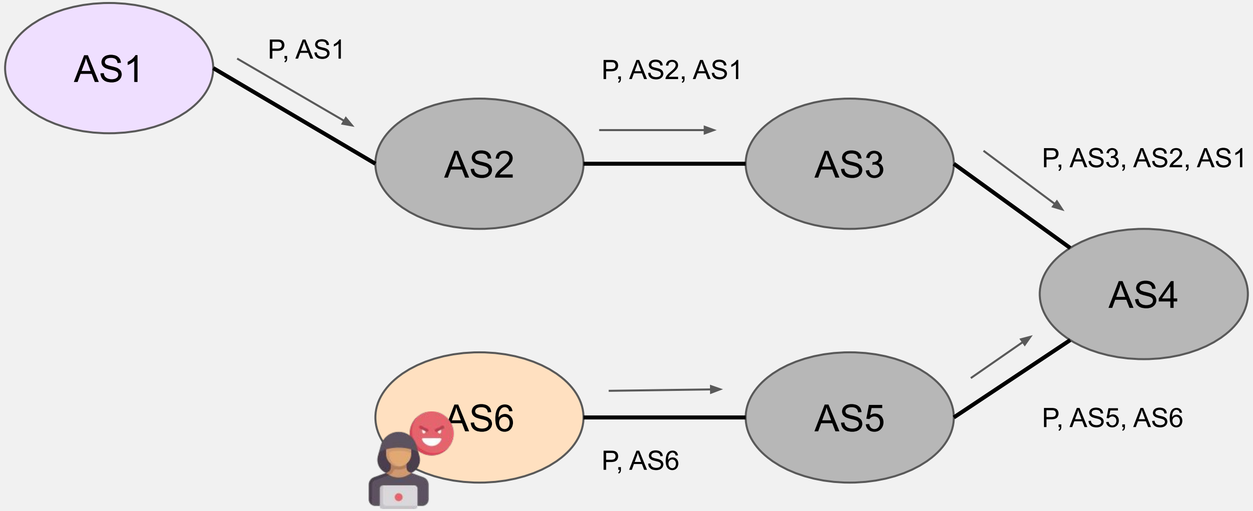
Using a forged BGP Update Message:

- A BGP speaker can advertise prefix it does not own to redirect traffic.
- **Subprefix attack:**
  - A speaker can advertise a more specific prefix to match a more precise route entry.
    - e.g., 10.0.0.0/16 hijacked with 10.0.0.0/17 and 10.0.128.0/17.
- If an AS disseminates a malicious prefix, the hijacking can propagate exponentially.

Hard to debug:

- Victim AS may not receive hijacking route messages.
- Victim AS may not notice traffic differences if the hijack is a non-destructive MitM.

Origin Hijacking



The attacker announces the **same prefix** as the victim, pretending to be the origin. Differs from the subprefix attack by claiming the exact same IP range.

Path Hijacking Attack

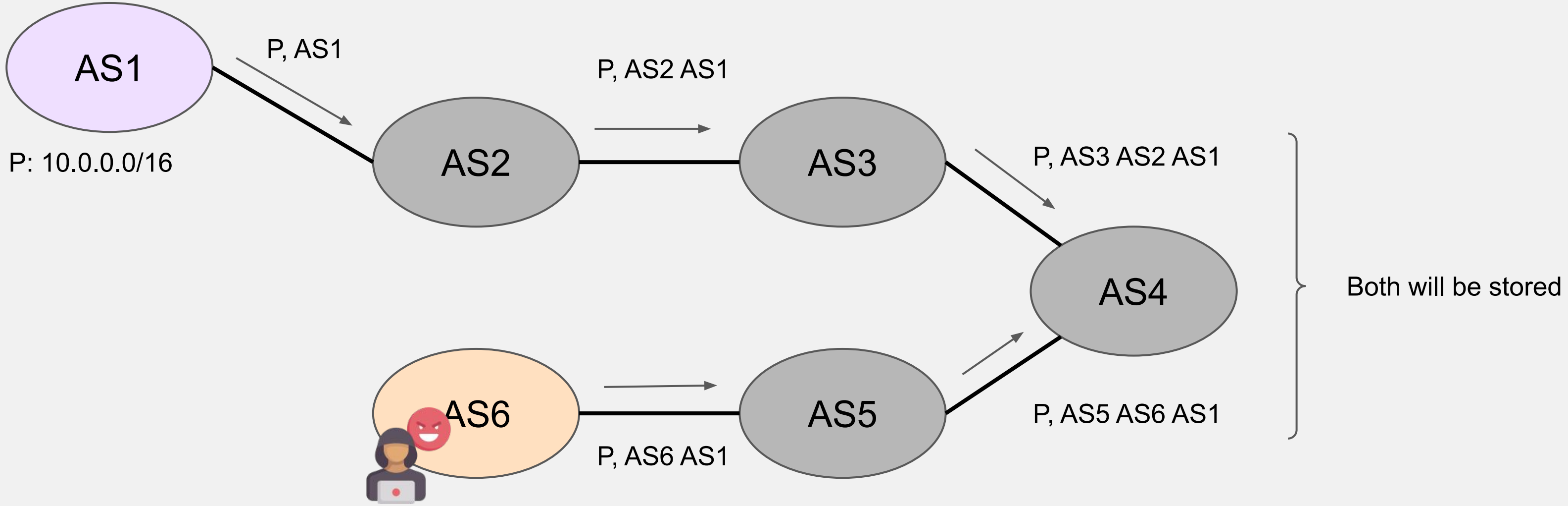
Path hijacking:

- An attacker does not claim the prefix itself, but forges an AS Path to to insert themselves in the path.

Problems:

- Easy to launch, you just need a BGP speaker
- Hard to debug, we need many probes to distinguish inconsistency.

Path Hijacking Attack



In practice, the path might be less attractive then the original, for instance longer, but can still "win" the routing decision based on local preferences, down links, etc.

Route Leakage

Unintentional or Intentional propagation of path advertisement against BGP policies

- Legitimate routes are propagated beyond their intended scope, leading to misrouting.
    - For instance, customers should not advertise internet provider’s routes to other providers to avoid becoming transit AS themselves.
    - E.g., China Telecom incident.

BGP Best Common Practices

BGP security is weak, so best practices exist for speakers:

- Reject prefixes that are more specific than /24.
- Do not accept address blocks that have not been officially assigned.
- Filter out routes that incorrectly advertise your own address space.
- Exclude private or special-use ranges.
- Ignore the default route (0.0.0.0/0), unless your AS operates as a stub.
- Decline any unallocated address space announced by a customer.

The problem is: not all BGP speakers apply these carefully, misconfiguration can happen, and verification can be cumbersome.

BGP Security

Thankfully smart people did smart things:

- **Origin Authentication (RPKI)**
  - Confirms that the announcing AS is authorised to originate the prefix.
  - Helps detect and reject invalid or mis-originated announcements.
  - Supports incremental deployment and works with existing BGP infrastructures.
- **Path Validation (BGPsec)**
  - Provides cryptographic validation of each AS hop along the AS Path.
  - Prevents path manipulation such as AS Path shortening or forging.
  - Requires universal deployment for full effectiveness.

Resource Public Key Infrastructure (RPKI)

Origin Validation using Internet Registers

Distributed routing databases

- Used to check which AS is associated with a given prefix.
  - IRR – https://irr.net/registry/
  - RADb – https://www.radb.net/

Limitations

- Poor data quality not well verified.
- Many entries are outdated or even incorrect.
- Multiple sources exist, leading to inconsistencies and incomplete information.

Resource Public Key Infrastructure (RPKI)

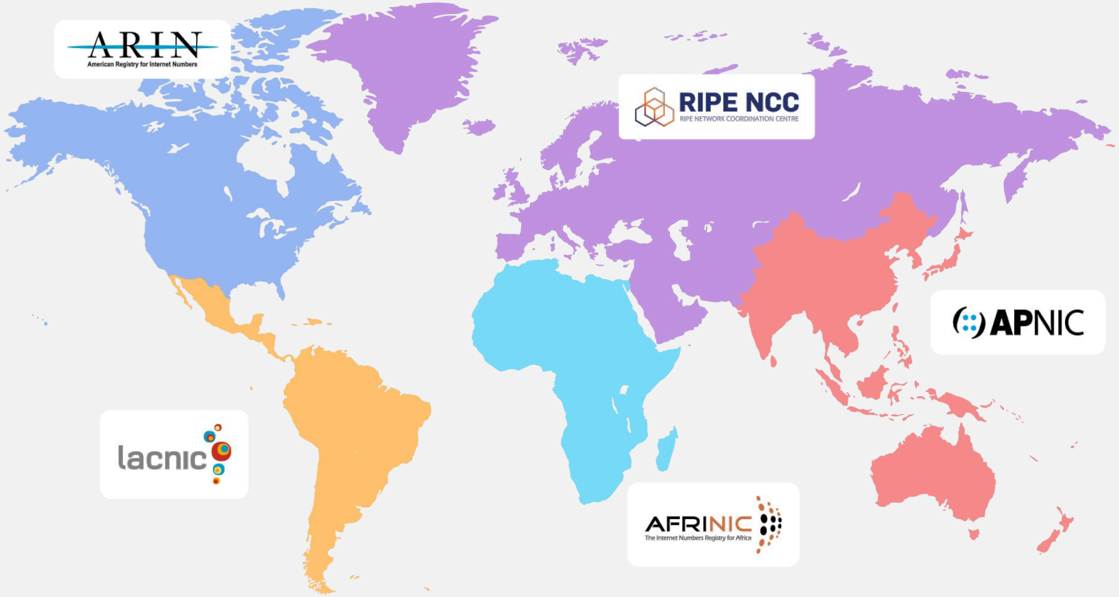
RPKI:

- Specialized **Public Key Infrastructure (PKI)** for BGP introduced with RFC 6480 in 2012 by the SIDR Working Group.
- Security framework for BGP Route Origin Validation (ROV).
- Allows resources (IP prefix) holders to prove their ownership and deliver announcement authorization to specific ASes.
- These authorizations can be used by ASes to verify the origin of a BGP announcement.
- Independent from BGP: no need to modify equipment.

Prevents:

- **Route Hijacking**: Unauthorized intentional prefix announcement.
- **Mis-origination**: Unintentional leakage of prefix.

Trust Anchors 1/2


ARIN
American Registry for Internet Numbers
RIPE NCC
RIPE NETWORK COORDINATION CENTRE
APNIC
lacnic
AFRINIC
The Internet Numbers Registry for Africa

As in all PKI, Certificate Authorities (CA) are needed to trust the upper layer of the chain. In RPKI, these are called **Trust Anchor (TA)** and correspond to the five Regional Internet Registries (RIR).

Trust Anchors 2/2

TAs are responsible for:

- Providing an infrastructure for resource holders to sign their prefix and ASN.
- Providing a public **RPKI repository,** so that others can verify them.

Each RIR operates its own repository to store objects (prefixes / ASNs) and certificates.

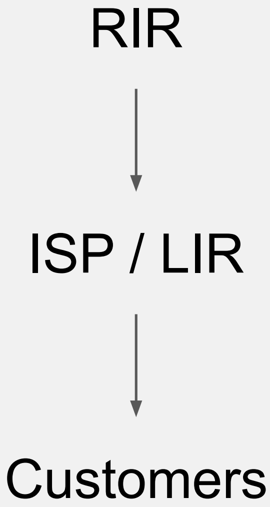
To allow an AS to announce a prefix, the prefix owner must first acquire a resource certificate from its RIR, confirming the address space it has been assigned.

Resource Certificate (RC)

A Resource Certificate (RC) certifies possession of a prefix or ASN.

A prefix holder can delegate it prefix or part of it to another by creating another RC.

In this case the prefix holder uses its own key pair to generate a new X.509 certificate to be stored in a RPKI repository.



End-Entity Certificate (EEC)

An **End-Entity certificate (EEC)** is simply a certificate that does not issue further certificates:

- A CA certificate can delegate resources and issue child certificates.
- An EEC cannot issue further certificates but can be used to sign a routing object.
- In RPKI:
  - RIRs → CA certificates
  - LIRs/ISPs → CA certificates (they delegate resources to others)
  - Customers → can be either
    - CA if they sub-delegate to someone else, or
    - EE if they only want to sign ROAs and do not manage further delegations.

Route Origin Authorizations (ROAs)

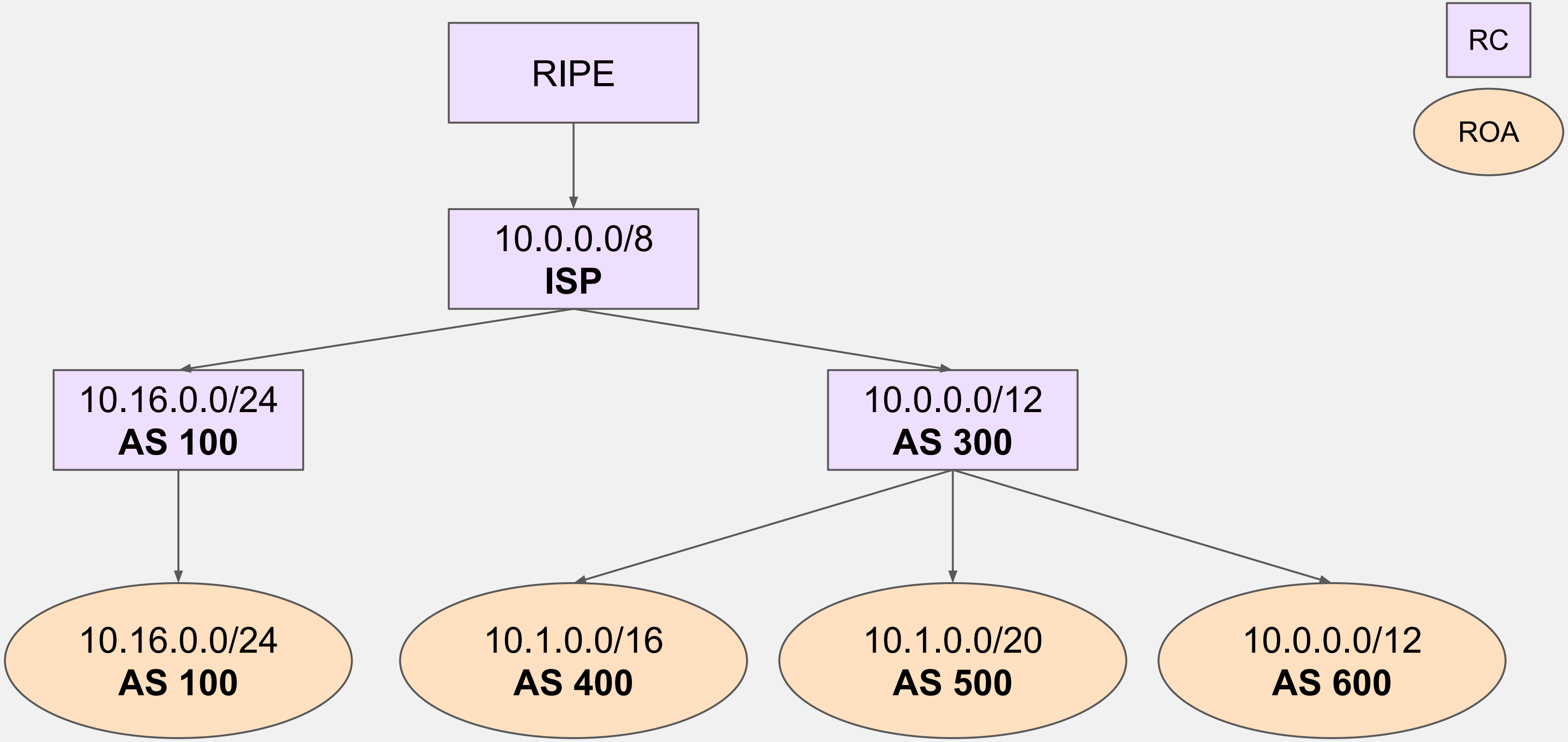
Route Origin Authorizations (ROAs) are the routing object at the core of RPKI.

They are used to authorize an **ASN to announce a specific prefix**.

An ROA contains:

- An ASN (can be 0 for reserved prefix or allocated ones that should not be announced).
- A prefix.
- A maximum authorized length (for prefix subdivision if needed).

RC and ROA Example



RC

ROA

RIPE

10.0.0.0/8
ISP

10.16.0.0/24
AS 100

10.0.0.0/12
AS 300

10.16.0.0/24
AS 100

10.1.0.0/16
AS 400

10.1.0.0/20
AS 500

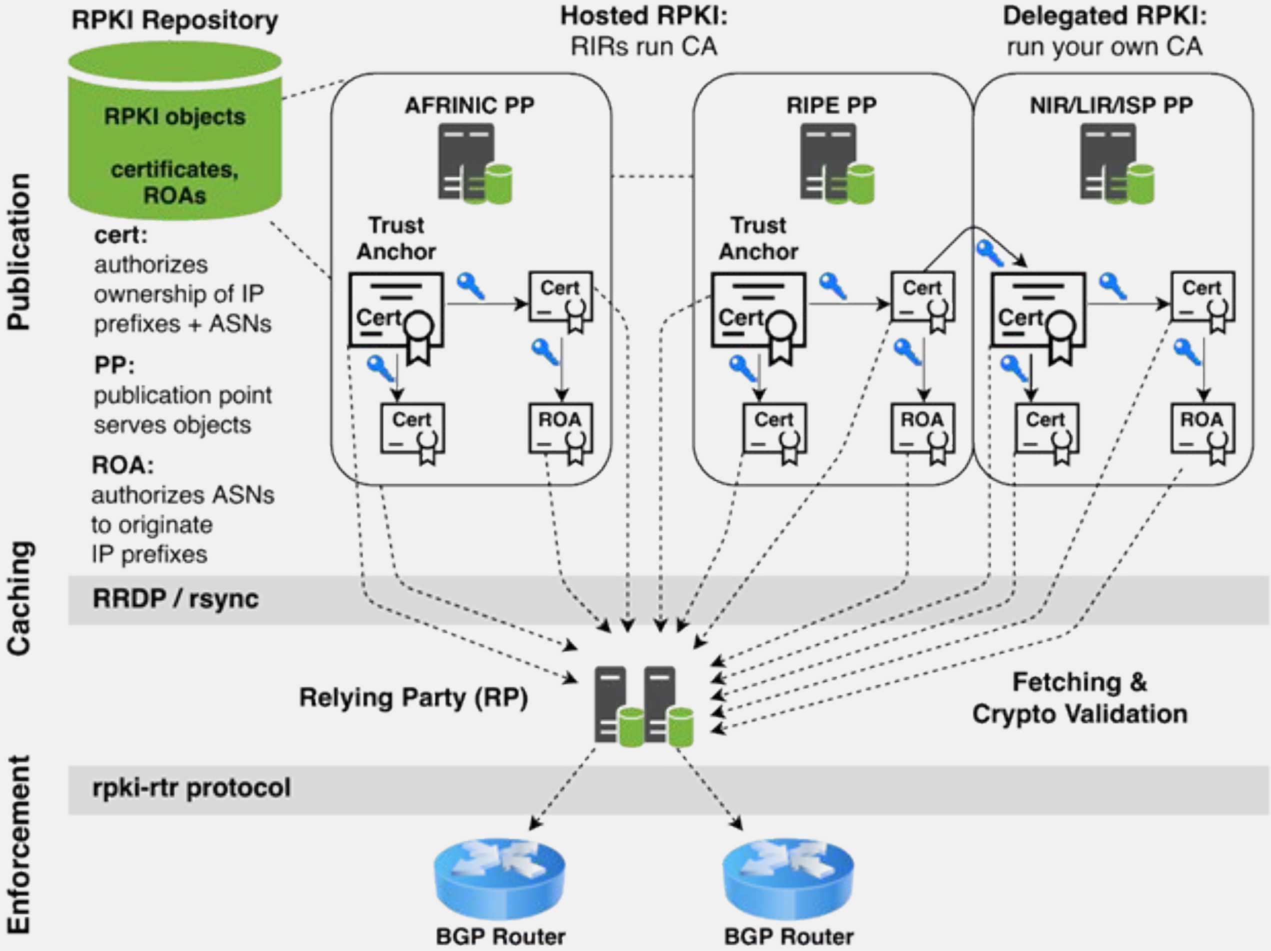
10.0.0.0/12
AS 600

Relying Parties (RP)

Relying Parties (RP) refer to any devices in charge of verifying the RPKI object used to decide of routing.

They are in charge of:

- Downloading the RPKI data from **publication points (PP)** linked to the RPKI repositories.
- Verifying the resource certificates and ROAs.
- Sending the **Validated ROA Payload (VRP)** to BGP routers using the **RTR protocol (RPKI to Router)**.



https://blog.apnic.net/2021/03/22/rpki-relying-party-synchronization-behaviour/

RPKI Takeaways

- RPKI allows origin validation using asymmetric cryptography with a PKI.
  - RC proves prefix/ASN ownership.
  - ROA authorizes an AS to be the origin of a given prefix.
- BGP routers are not the one handling the validation:
  - Relying Parties are the one fetching data from the PP.
  - Relying Parties do the cryptographic verification before sending the results to the BGP routers.
- BGP routers can just check their table entries with the verified data received from the RTR protocol.

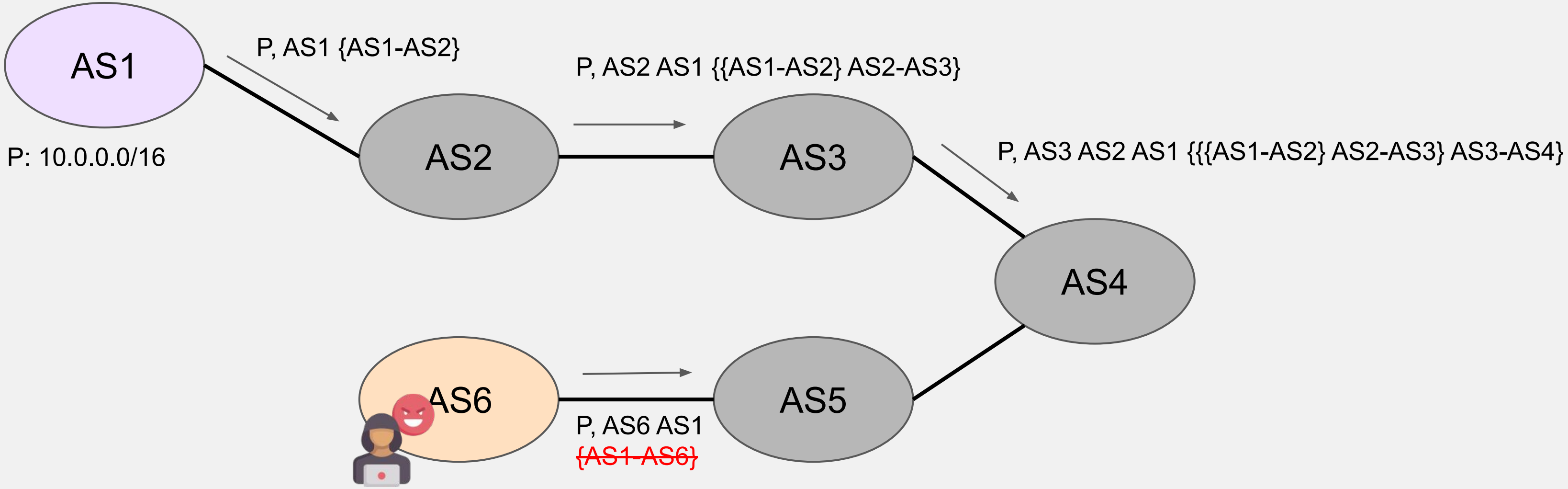
BGP Security
(BGPsec)

BGP Security (BGPsec)

BGPsec:

- **Goal:** Verify the AS Path.
- Extension to BGP providing security to BGP routing defined in RFC 8205 in 2017.
- Adds a BGP attribute containing a digital signature to validate path.
    - The sender signs the prefix announced and the ASN of the receiver as well as all prior signatures.

Path Hijacking Mitigation with BGPsec


AS1
P: 10.0.0.0/16
P, AS1 {AS1-AS2}
AS2
P, AS2 AS1 {{AS1-AS2} AS2-AS3}
AS3
P, AS3 AS2 AS1 {{{AS1-AS2} AS2-AS3} AS3-AS4}
AS4
AS6
AS5
P, AS6 AS1
~~{AS1-AS6}~~

An attacker cannot forge a signed path from AS1 to AS6 since they do not know AS 1 private key.

BGPsec in Practice

Unfortunately, in the wild BGPsec is not deploy at all.[1]

Hard to deploy:

- Need to change equipment to enable cryptographic operations.
- Cryptographic processes might be too slow to handle the speed of equipments.
- There is no benefit in being first on deploying BGPsec:
  - You need others to perform BGPsec, and no one will understand you.

[1] *https://blog.apnic.net/2025/05/23/bgpsec-could-you-run-it-if-you-wanted-to/*

Resources and Acknowledgements

- *Internet Security: A Hands-on Approach, 3rd Edition,* Du Wenliang
- APNIC Documentation