Fiber Telecom
THE NETWORK PARTNER

Transforming BGP Operations

Serving the best prefix out of 26 million Modern BGP Design in a Tier 2 Network

The network brothers...

Antonio Prado is a technology enthusiast with an IT career spanning back to 1993 and active involvement in the Internet industry since 1995. An open-source advocate and IPv6 early adopter, he served as **CTO** for twenty years across various companies.

Alessandro D'Eliseo is a Network Engineer with over 8 years of experience at Fiber Telecom, specializing in the design and optimization of high-performance, scalable, and resilient core networks.

Agenda

- Fiber Telecom BGP Architecture (Scenario)
- Limits of the Legacy Route Reflector
- Technical Solution
- Conclusions and Takeaways

Scenario

Fiber Telecom BGP Architecture

Fiber Telecom

Tier 2 Wholesale Operator

- Backbone SR-MPLS
- 30+ PoPs across European Data Centers and IXPs
- ~5000 external BGP sessions
- Peer with 2000+ different ASNs
- 3 RR per family inet, inet6 ed evpn

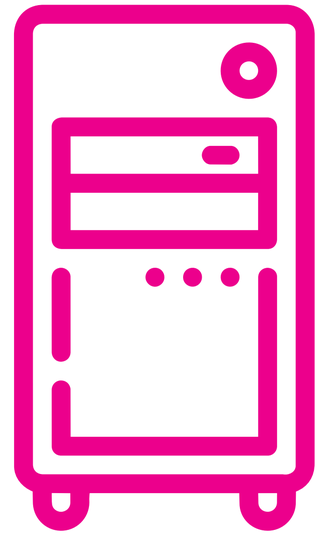
Who do we establish BGP sessions with?

- Tier-1 Upstreams
- IXP
- CDN
- Downstreams

BGP prefix distribution with Route Reflector

Why Route Reflector?

- Simplified network design
- Resource Efficiency
- Hierarchy Support

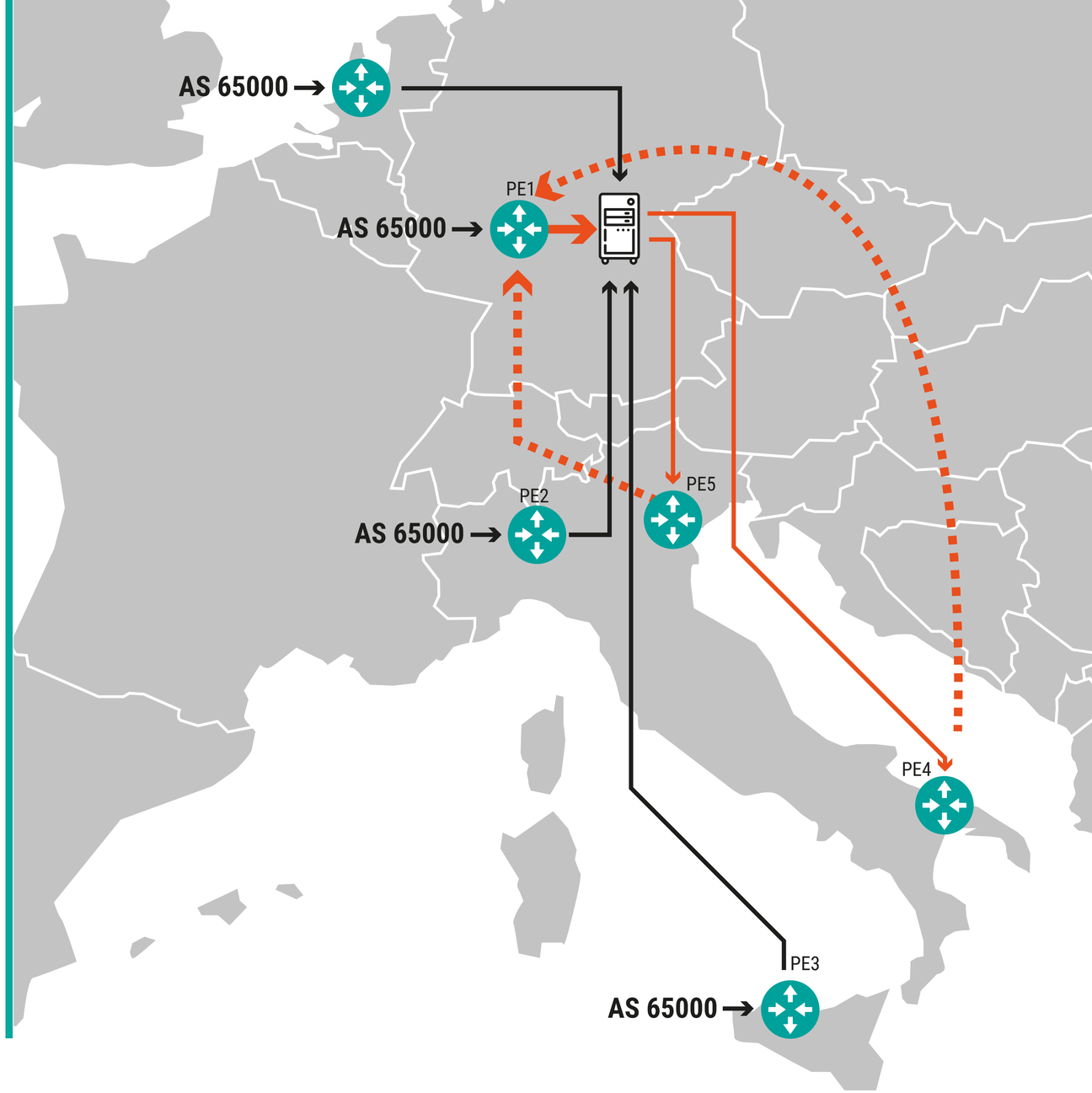


RR Limits

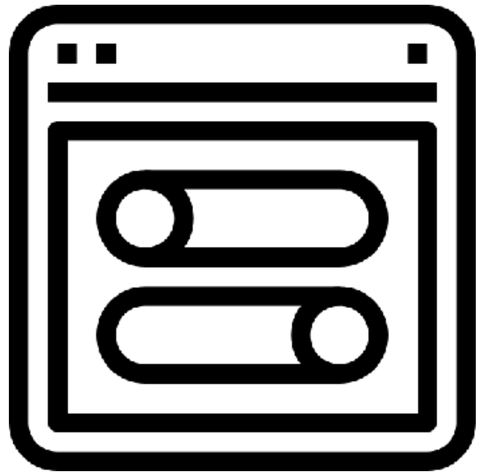
It reflects the best path according to its topological position*:

- Sub-optimal Routing
- Path Hiding

(*) While multiple RRs reduce path hiding, high POP density in Tier-2 networks often makes iBGP Full-Mesh preferable for optimal visibility.



Can we do better?



Routing Optimization

- Advertise the best path to each client.
- Balance traffic across multiple peering points
- Improve network resilience
- Egress Traffic Engineering
- Active monitoring of All Alternatives

Technical Solution

Legacy suboptimal routing mitigation

FULL-MESH

Pros: Elimination of path hiding by intermediaries

Cons: Poor scalability

BGP ADD Path RFC 7911

Pros: Sends multiple paths for the same prefix

Cons: Sends its own best path, not the client's

BGP Optimal Route Reflector RFC 9107

Pros: Sends the client's best path

Cons: Sends only one path, so reconvergence occurs in the event of a fault

INSPIRATION

Modern BGP Design (ITNOG6 Nicola Modena)*

BGP ADD PATH + ORR

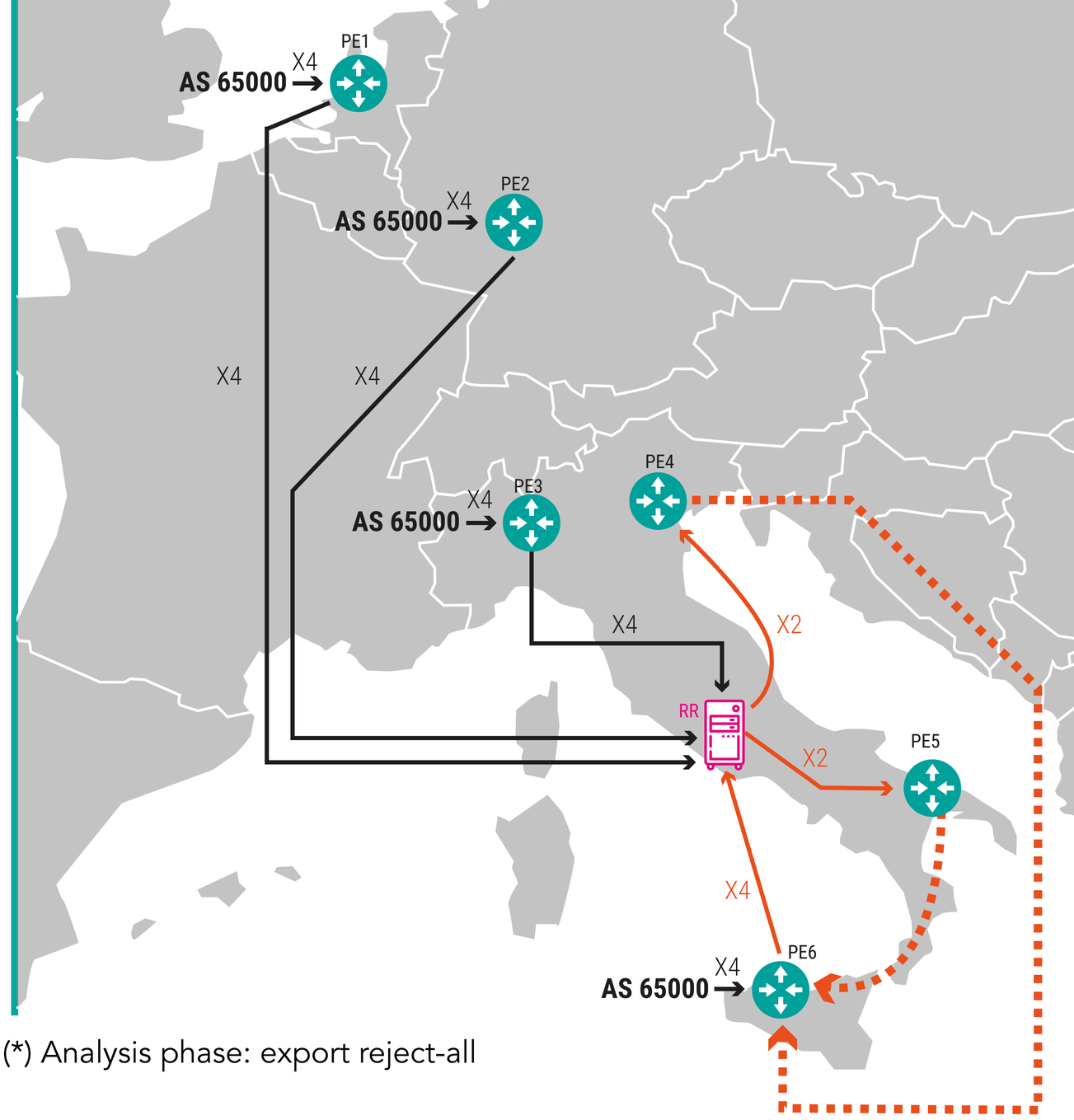
*https://github.com/nmodena/blog/blob/master/202209-Modern_BGP_Design_nmodena_itnog6.pdf

Analysis

BGP ADD-PATH RFC 7911

Device Role	Path Policy	Reason
Provider Edge	Add-Path Send 4	Full topology awareness & path diversity
Route Reflector	Add-Path Send 2*	Optimized FIB usage & Backup path (PIC) on PE

26M paths identified—revealing 18M routes
previously hidden by legacy RR Best Path selection



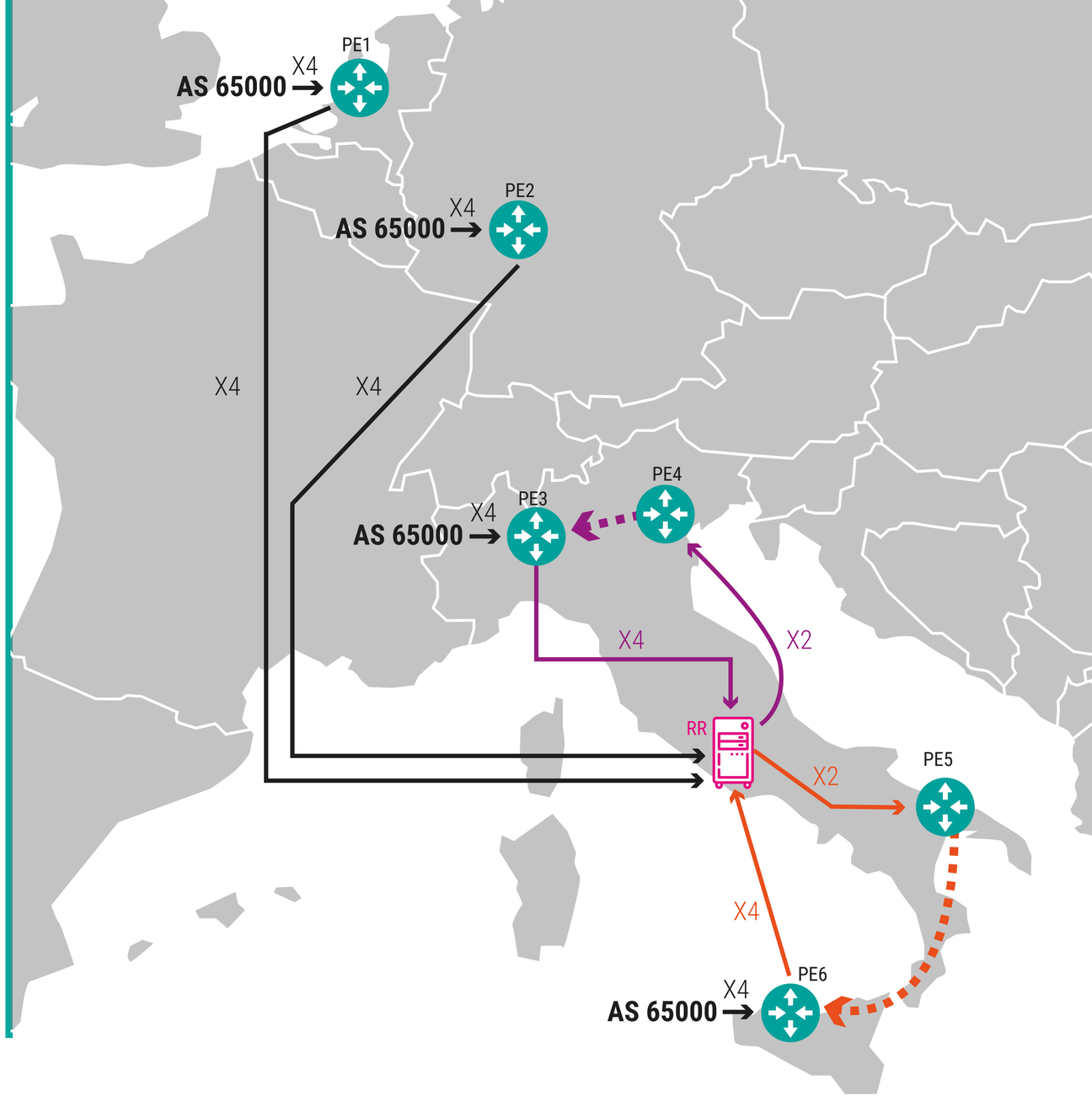
(*) Analysis phase: export reject-all

Analysis

BGP Optimal Route Reflector (ORR) RFC 9107

RESULT

- Full Network Visibility: Handling 26M+ Paths
- Suboptimal routing elimination
- Optimized Traffic Balancing Across Multiple Peering Points
- Improve network resilience



IMPROVING CONVERGENCE

Initial convergence time: 25 minutes*

REASONS:

- High number of paths to process
- Single-threaded nature of the BGP process

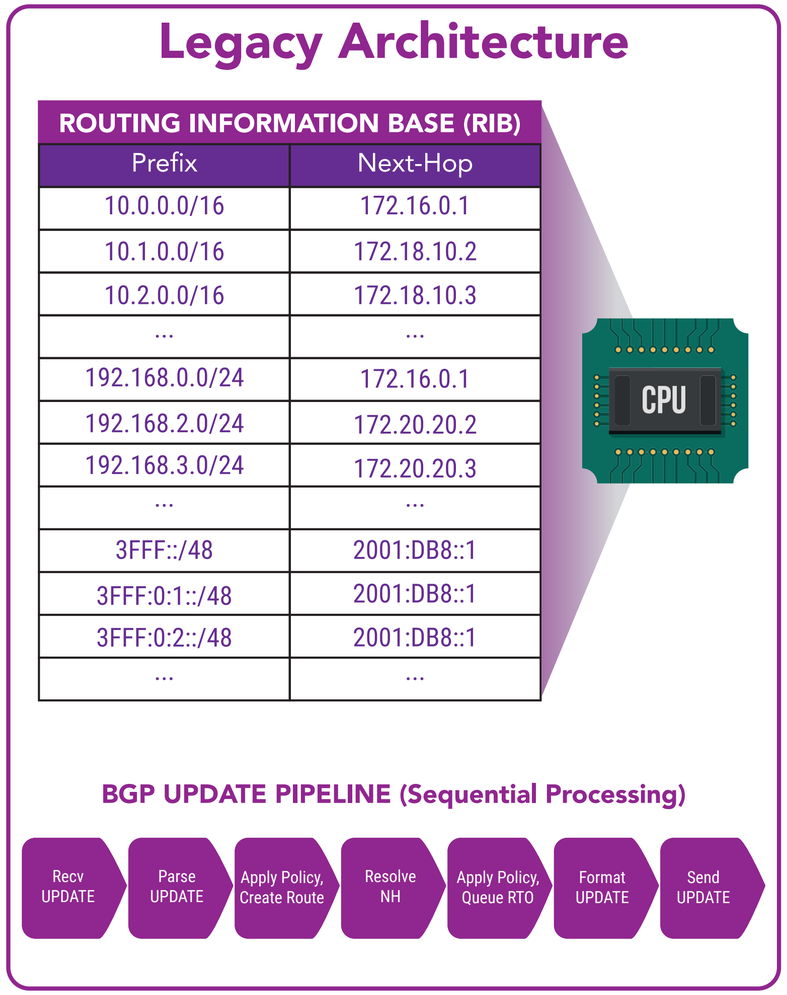
(*) Zero packet loss because PEs are connected to multiple RRs.

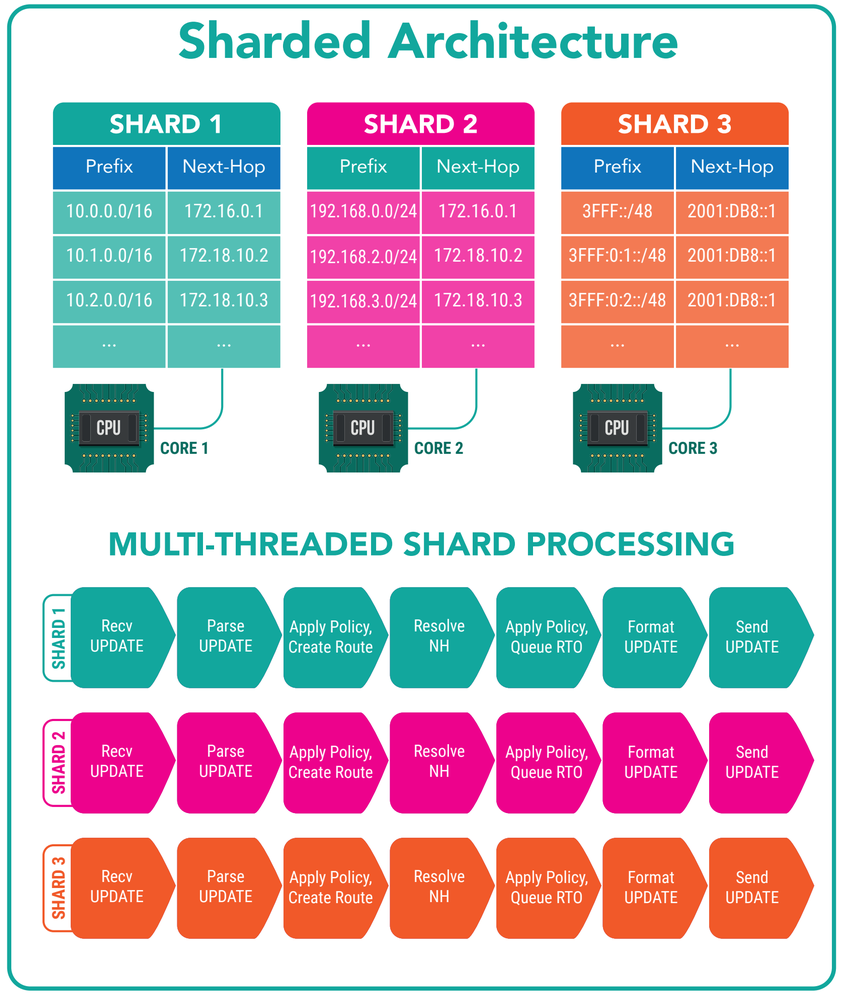
How to resolve it?

- Rib-Sharding
- Update Threading

RIB Sharding

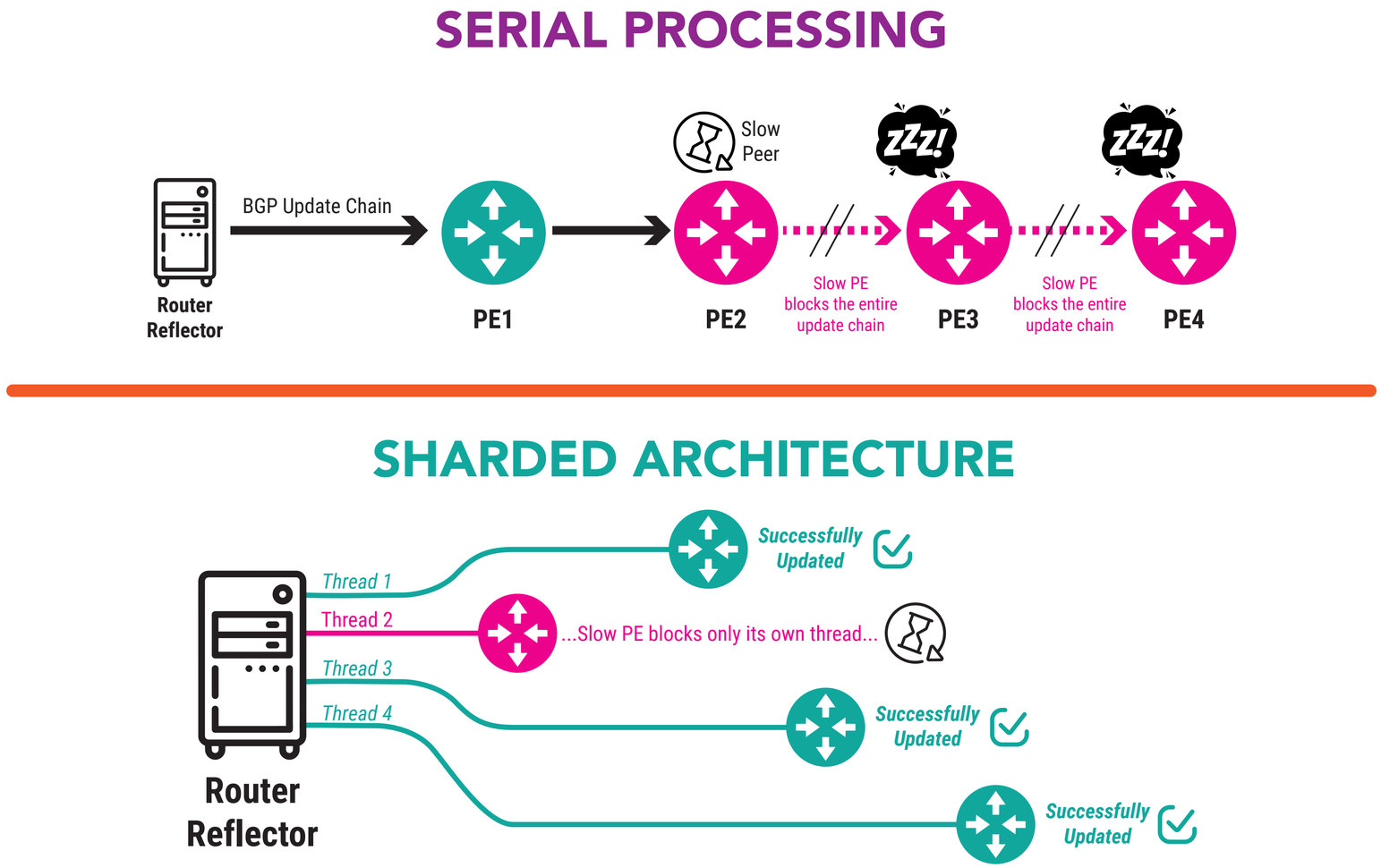
- RIB divided into shards
- Each shard is assigned to a CPU core





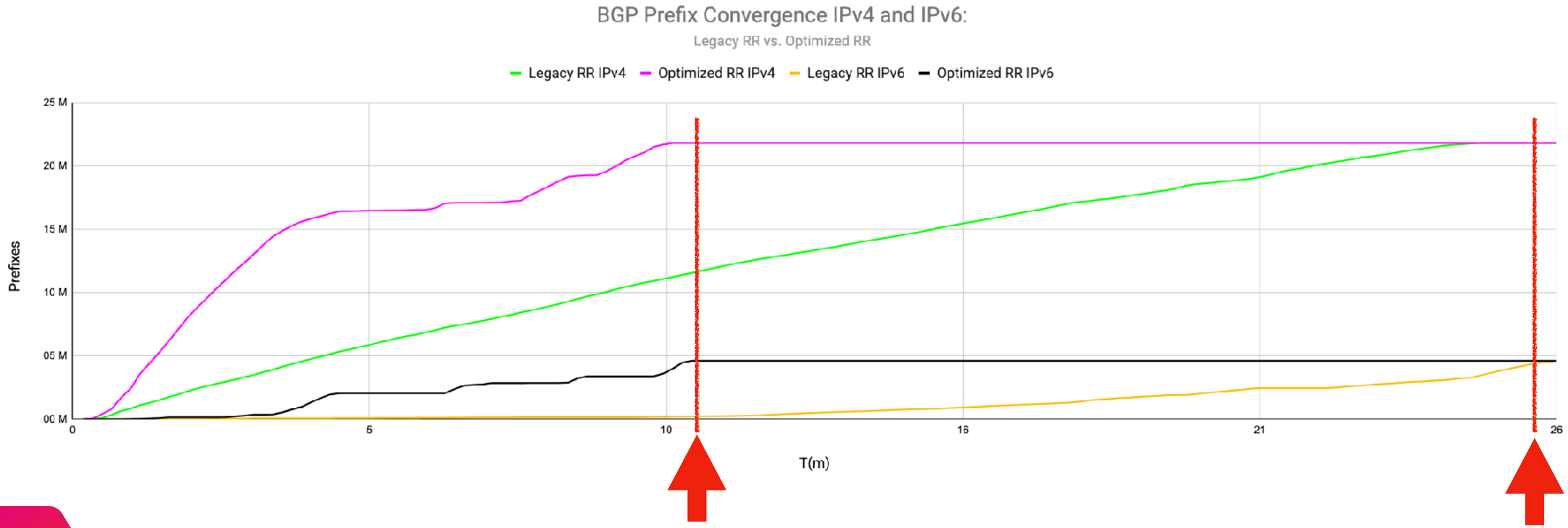
Update Threading

- Each BGP session (or group of sessions) has its own dedicated update thread.
- If one peer is slow it doesn't slow down the others



Comparative Analysis: Convergence

Convergence time reduced by 56%



OPTIMIZED
11 minutes

LEGACY
25 minutes

ARE WE ALL THE SAME?

BGP process has route prioritization capability we use for:

1. Infrastructure (Next-hop, Mgmt networks, etc.)
2. Customers
3. Top ASes by traffic volume
4. Everything else

Platform selection, RR placement strategy

What hardware to use?

- Multicore CPU
- RAM
- no forwarding (out-path)

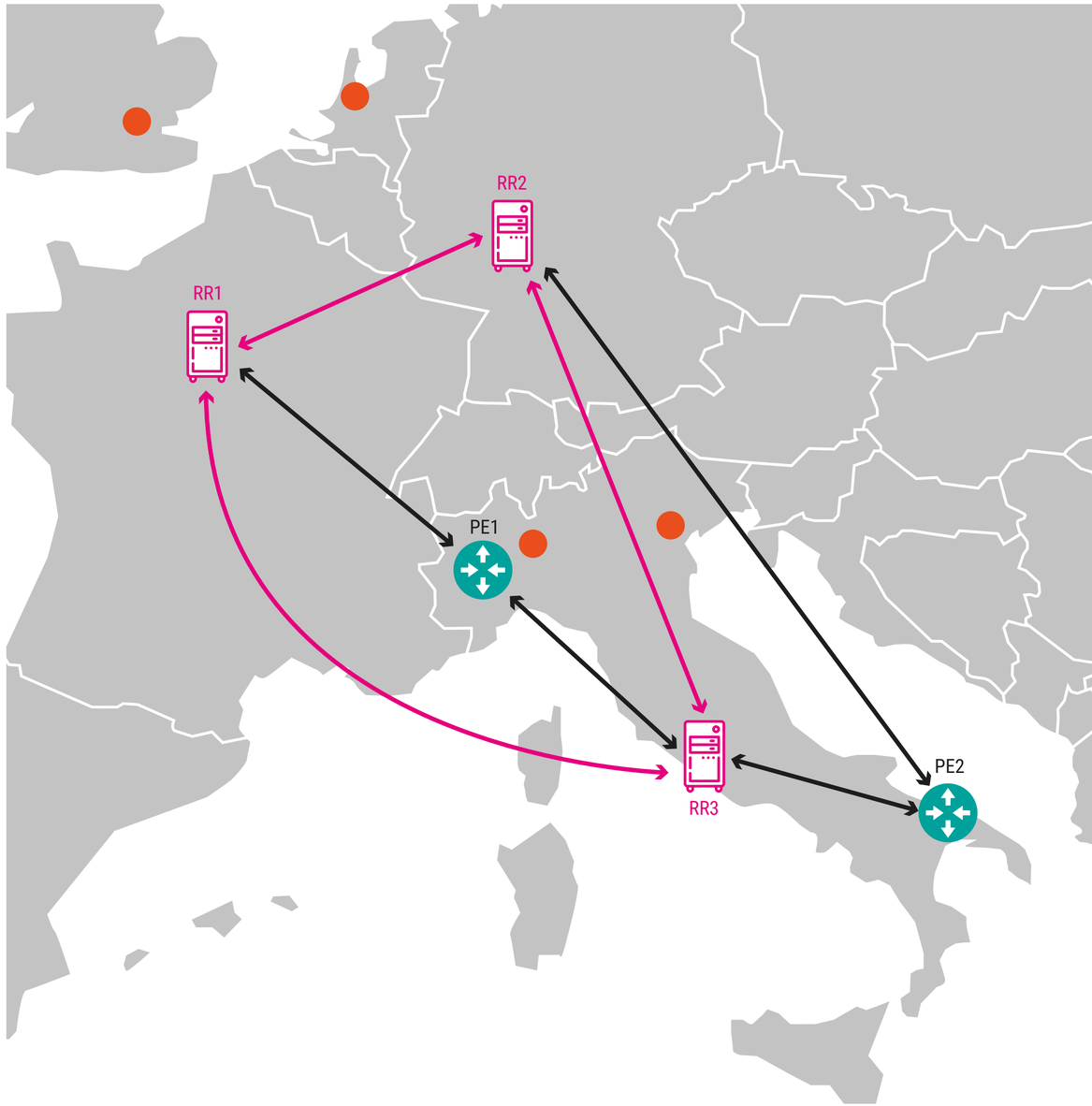
Answer: VM

How do we place them?

- Geographic Segmentation: 3 Different Regions
- Zonal Route Reflectors (1 per Region)
- Full-Mesh iBGP Core between RRs
- Redundant PE Peering: Each Edge Node connected to 2 RRs

What services?

- Unicast routing IPv4/IPv6 and BGP Flowspec*



(*)EVPN services, in DC and Man, on decoupled Legacy RR Architectures:

- EVPN natively provides path diversity
- Stringent Convergence Requirements

Migration Process

1. **Setup & Preparation**
   - Parallel Infrastructure: Deployment of 3 new Route Reflectors (RRs) and peering with existing RRs.

2. **Migration Strategy**
   - Gradual Replacement: Sequential "one-by-one" swap of legacy RRs with the new units on each Client.
   - Phased Rollout & Monitoring: * Constant monitoring of traffic volume and BGP convergence.

3. **Reliability & Results**

   - Zero Interruption: The migration was completed with **zero packet loss** and no service disruption.
   - Loop Prevention: Routing integrity was guaranteed by **SR-MPLS**, which natively prevented routing loops during the topology change.

Conclusions and Takeaways

Current optimization

- Tier 1 and Tier 2 have challenging BGP scalability needs.
- Better traffic distribution
- Modern BGP Design (ORR+AddPath) if you want to always provide the optimal path and improve convergence times
- Leverage multicore architecture (Sharding/Threading) recommended for large tables

The solution is not a single feature.

It is the result of analysis, method and design.

Thank you

FiberTelecom
The Network Partner