



BGP Route Server Functionality & Design



AusNOG September 2010

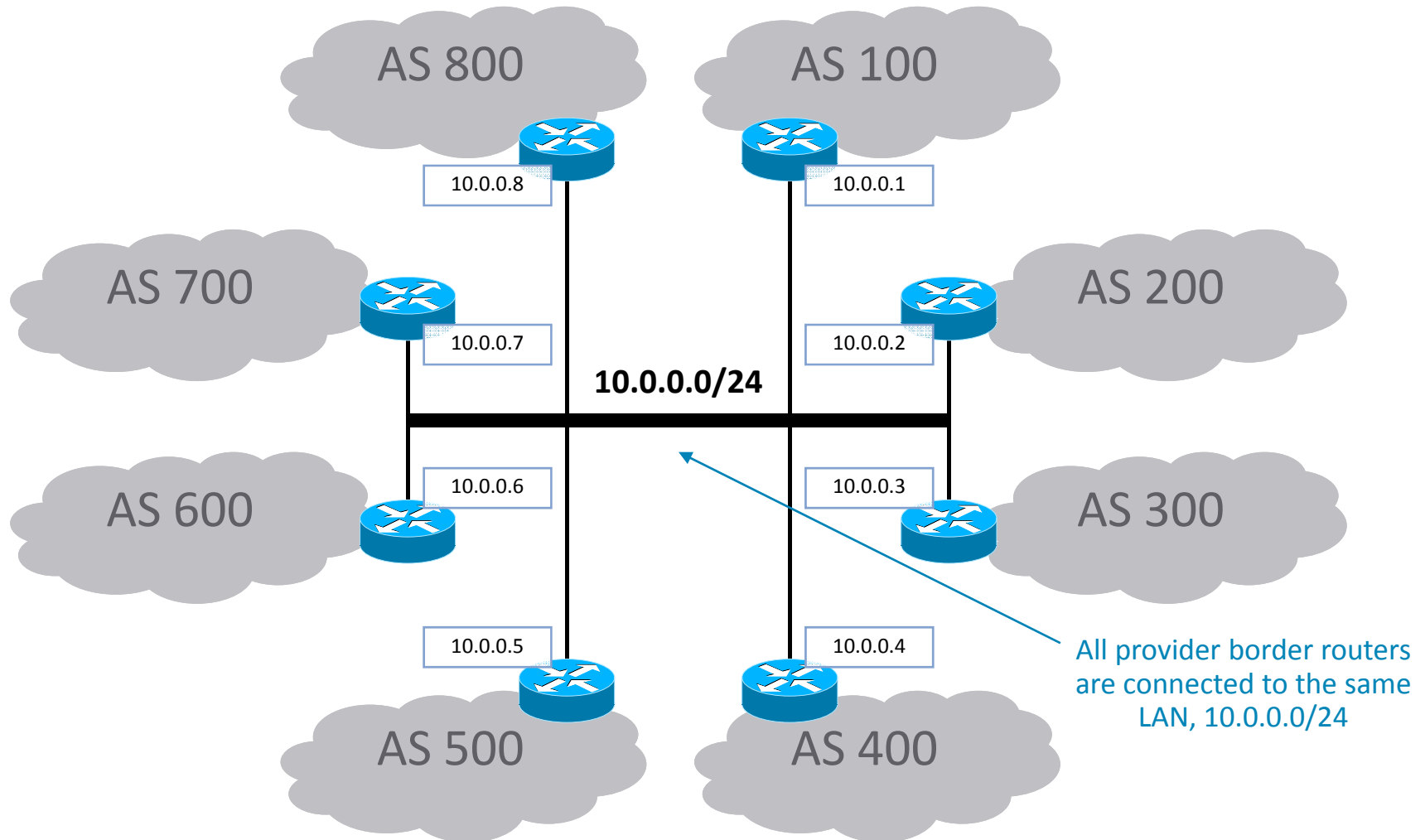
Ryan Bickhart

Robert Raszuk

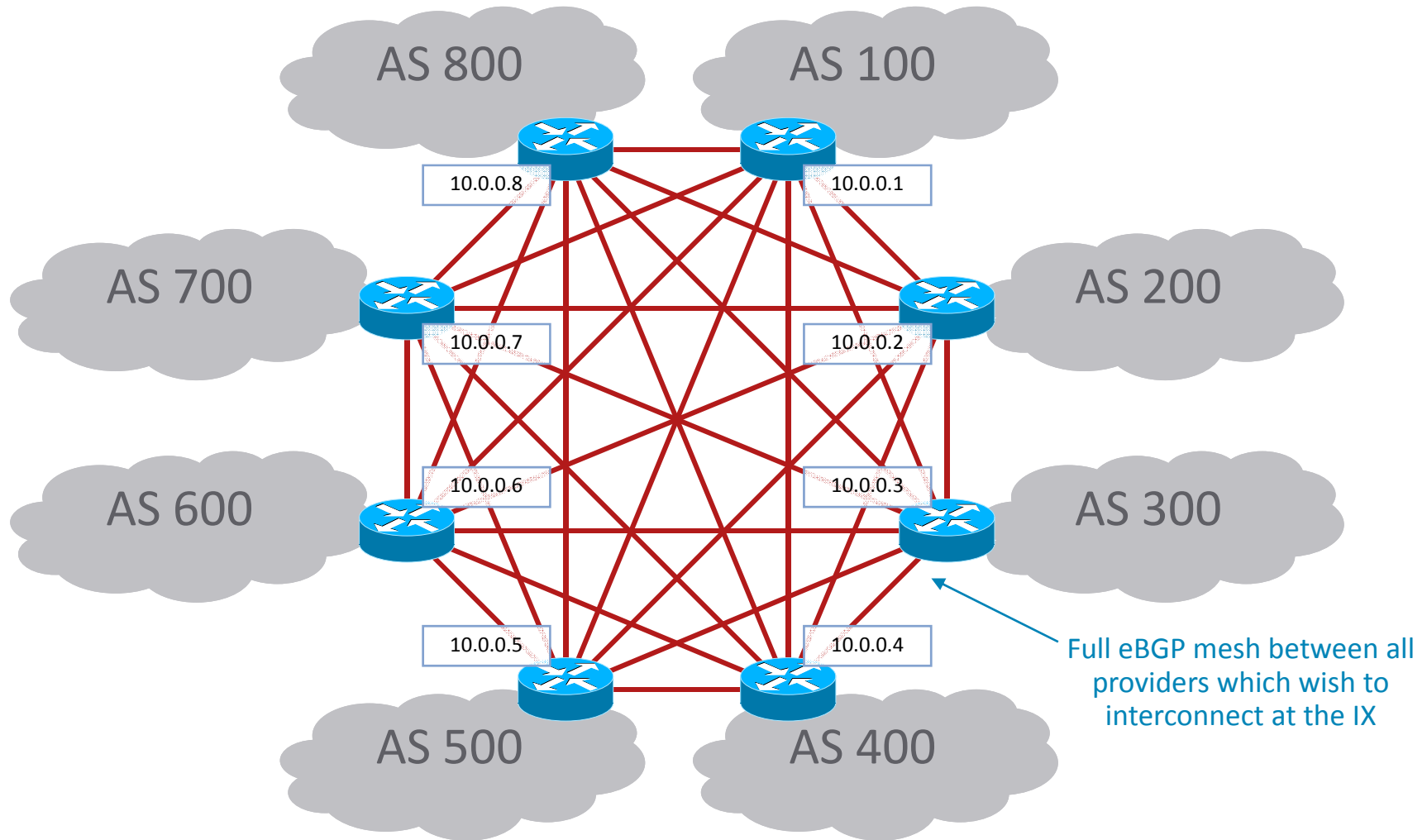
Role of an Internet Exchange (IX)

- Facilitate interconnection of providers' networks
- Provide facilities for operation and connection of BGP border routers:
 - Rack space
 - Electricity
 - Cooling
 - Switching infrastructure
- Alternative to private direct peerings for customers and smaller providers

IX: Physical Topology



IX: Direct BGP Peering

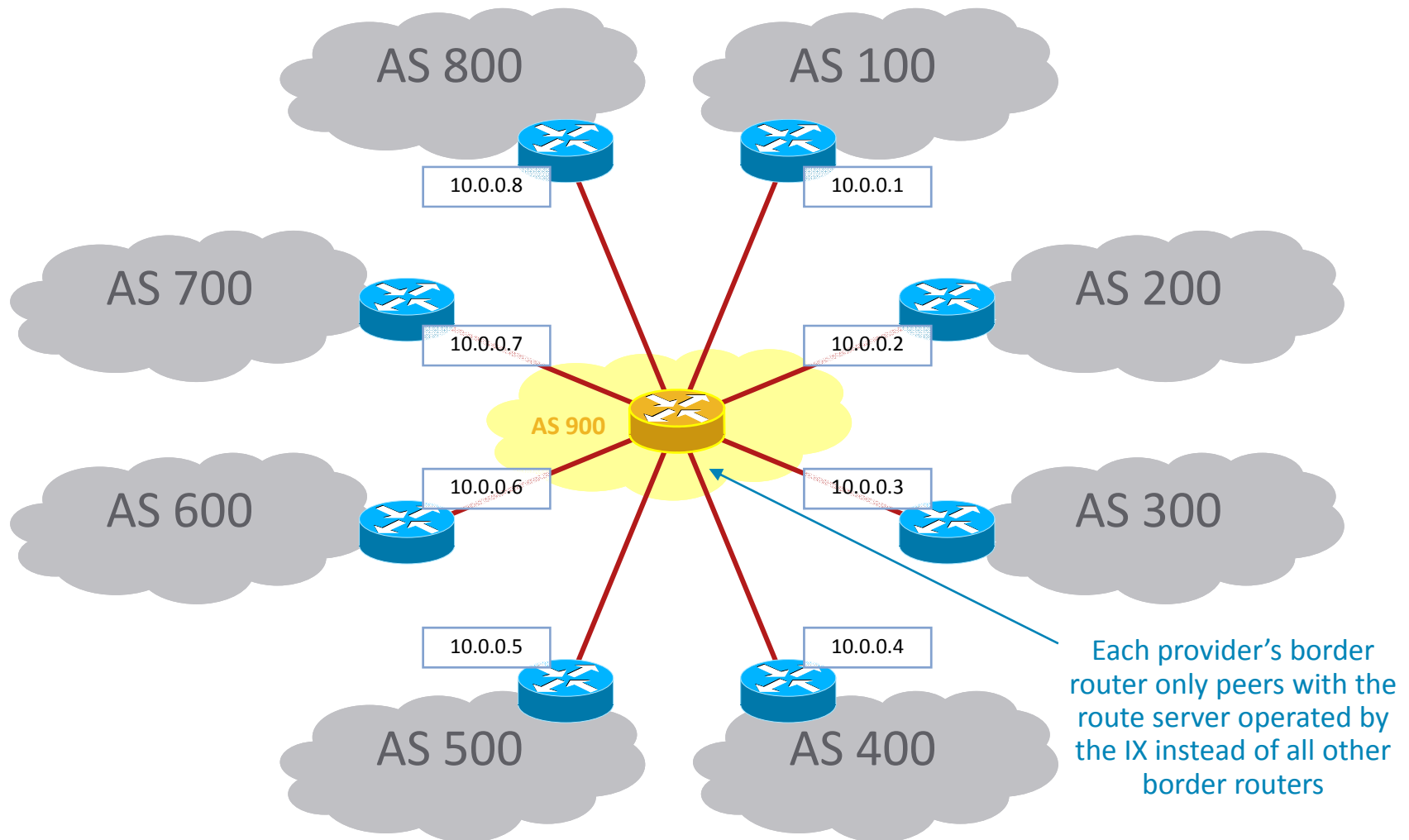


IX: Direct Peering Drawbacks

- Does not scale – drawbacks similar to iBGP full mesh
 - Individual config for every peer on every border router
 - Dozens to hundreds of potential connections at every IX
 - Not economically viable for large providers to connect to hundreds of smaller providers – large operational cost to maintain thousands of peering agreements worldwide

- Solution?
 - For iBGP: Route Reflector
 - For eBGP: ...

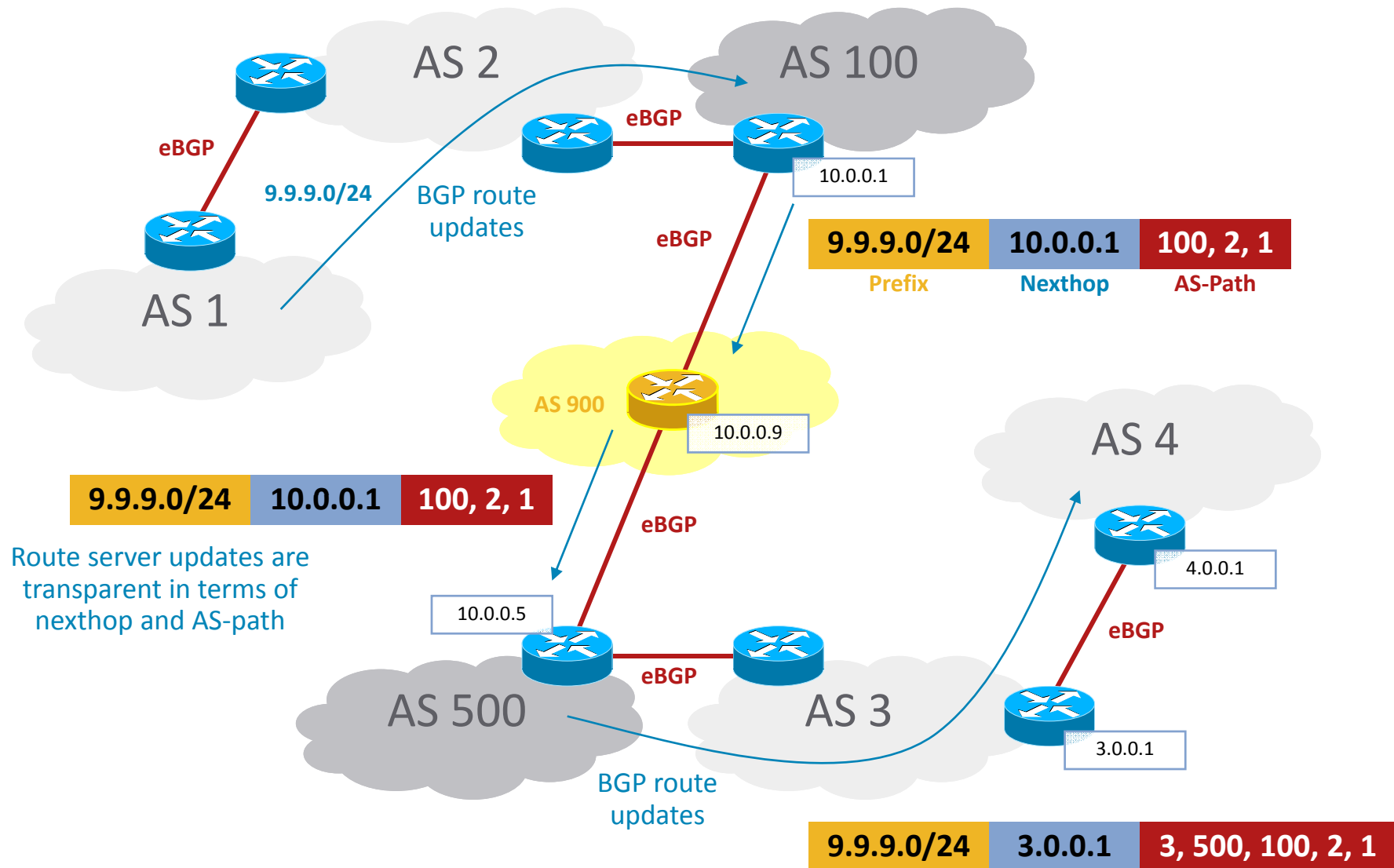
IX: BGP Route Server



Route Server Benefits

- Simplifies interconnection at IX
 - Each provider operates one eBGP session between border router and the route server
 - Reduced config on border routers
 - Reduced memory/CPU requirements on border routers
 - One peering agreement with IX rather than individual peering agreements with every directly connected provider
- Presence of the route server is invisible outside of the IX
 - AS-Path transparency
 - MED transparency
 - Nexthop transparency

IX Using BGP Route Server



Main Route Server Requirements (1)

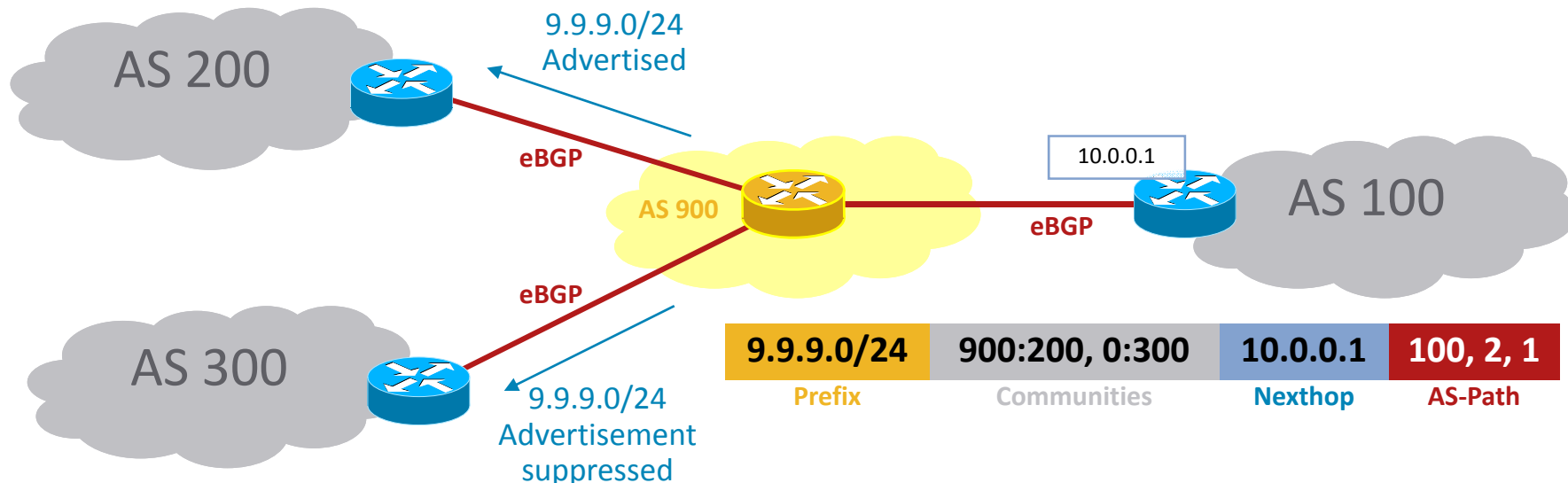
- Support IPv4 and IPv6 with full transparency and customized policy, multicast with transparency
- Transparency for all attributes: Nexthop, MED, AS-path
- Community-based policies for setting local preference; examples:

Community	Local Preference
No-Advertise	0
No-Export	0
0:<IX-ASN>	0
0:x, x != <IX-ASN>	50
<IX-ASN>:<IX-ASN>	100
No Community	100

Main Route Server Requirements (2)

- Community-based policies for prefix propagation; examples:

Community	Action
0:<peer-ASN>	Block announcement of route to AS <peer-ASN>
<IX-ASN>:<peer-ASN>	Announce route to AS <peer-ASN>
0:<IX-ASN>	Block announcement of route to all ASes
<IX-ASN>:<IX-ASN>	Announce route to all ASes



Main Route Server Requirements (3)

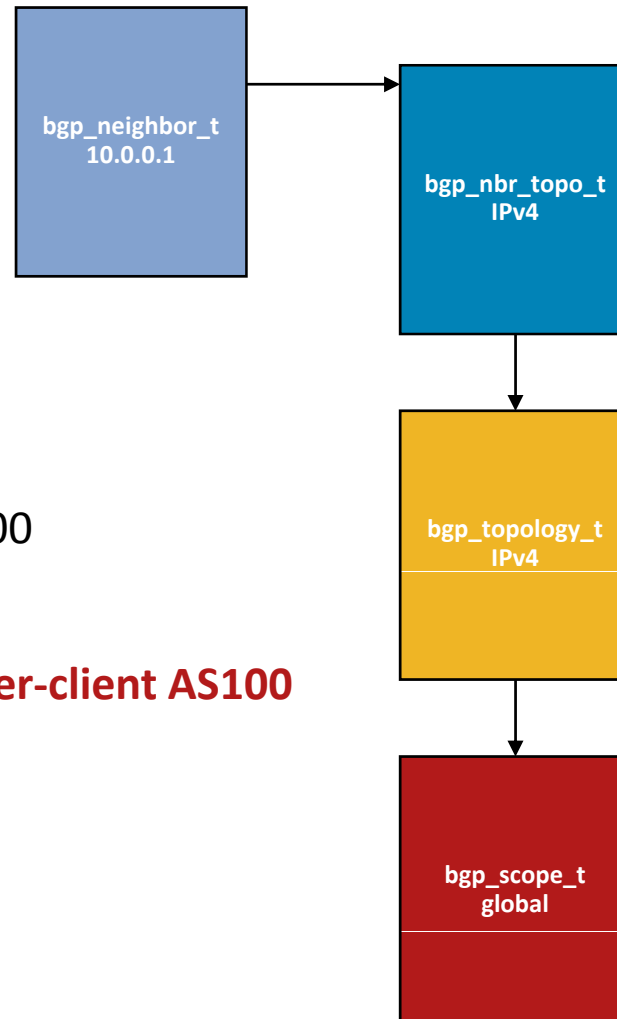
- Primary and secondary route servers connect via iBGP VPNv4
- Flexible inbound and outbound policies
 - Route Targets
 - Route Maps
- Scalable per-route-server-client/AS bestpath support
 - Main BGP table holds master copy of all nets and paths
 - Per-client/AS virtual BGP table contains nets/paths which require customized policy handling based on RTs/route maps

Route Server CLI

Route Server Configuration:

```
router bgp 900
  bgp route-server
  neighbor 10.0.0.1 remote-as 100

  address-family ipv4 unicast
    neighbor 10.0.0.1 route-server-client AS100
    neighbor 10.0.0.1 activate
```

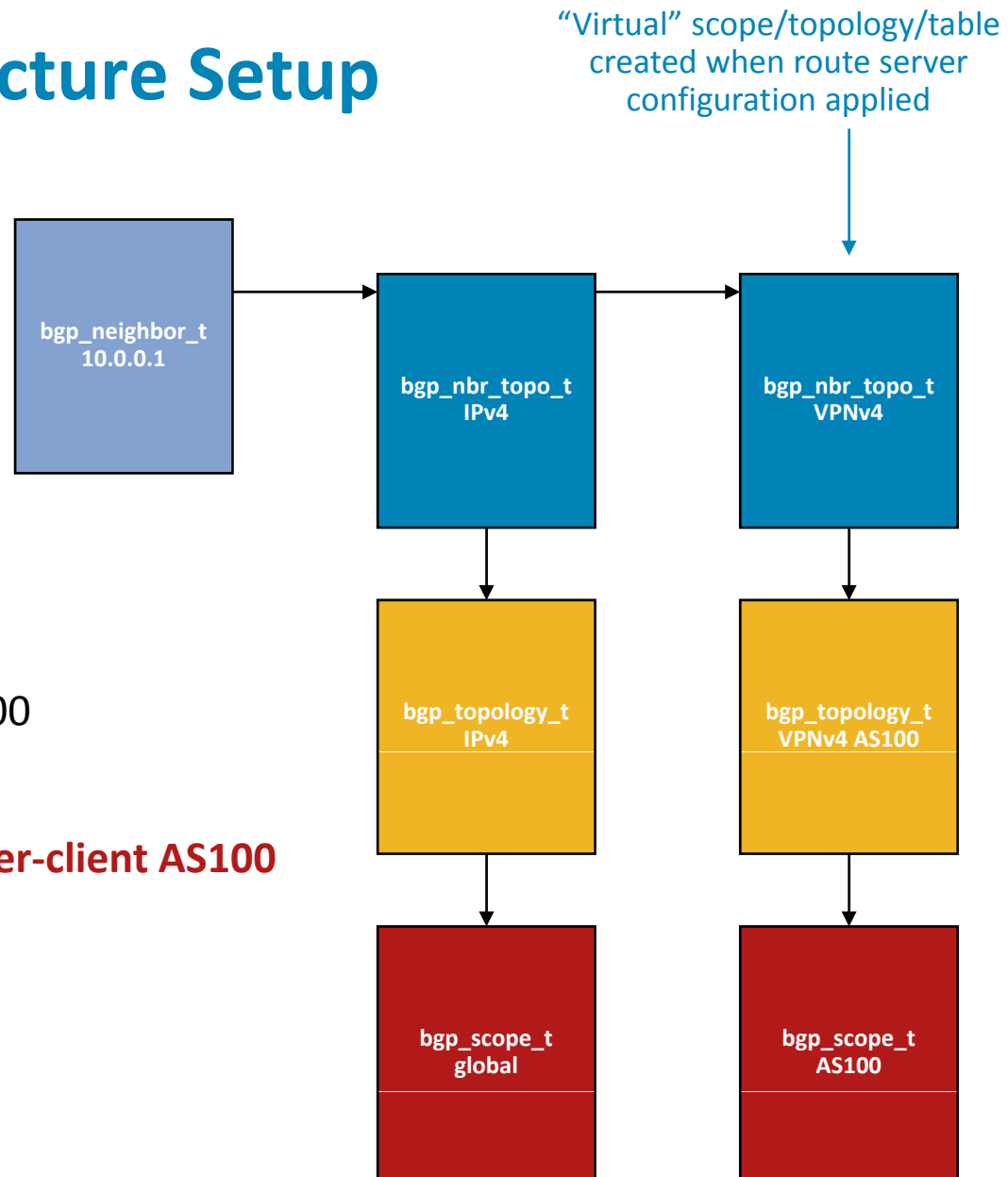


Virtual Data Structure Setup

Route Server Configuration:

```
router bgp 900
  bgp route-server
  neighbor 10.0.0.1 remote-as 100

address-family ipv4 unicast
  neighbor 10.0.0.1 route-server-client AS100
  neighbor 10.0.0.1 activate
```



Route Server Client Unique RD

- For each route server client, allocate a unique RD value
- Unique RD maps to per-client/AS “virtual” BGP VRF table

Existing RD types:

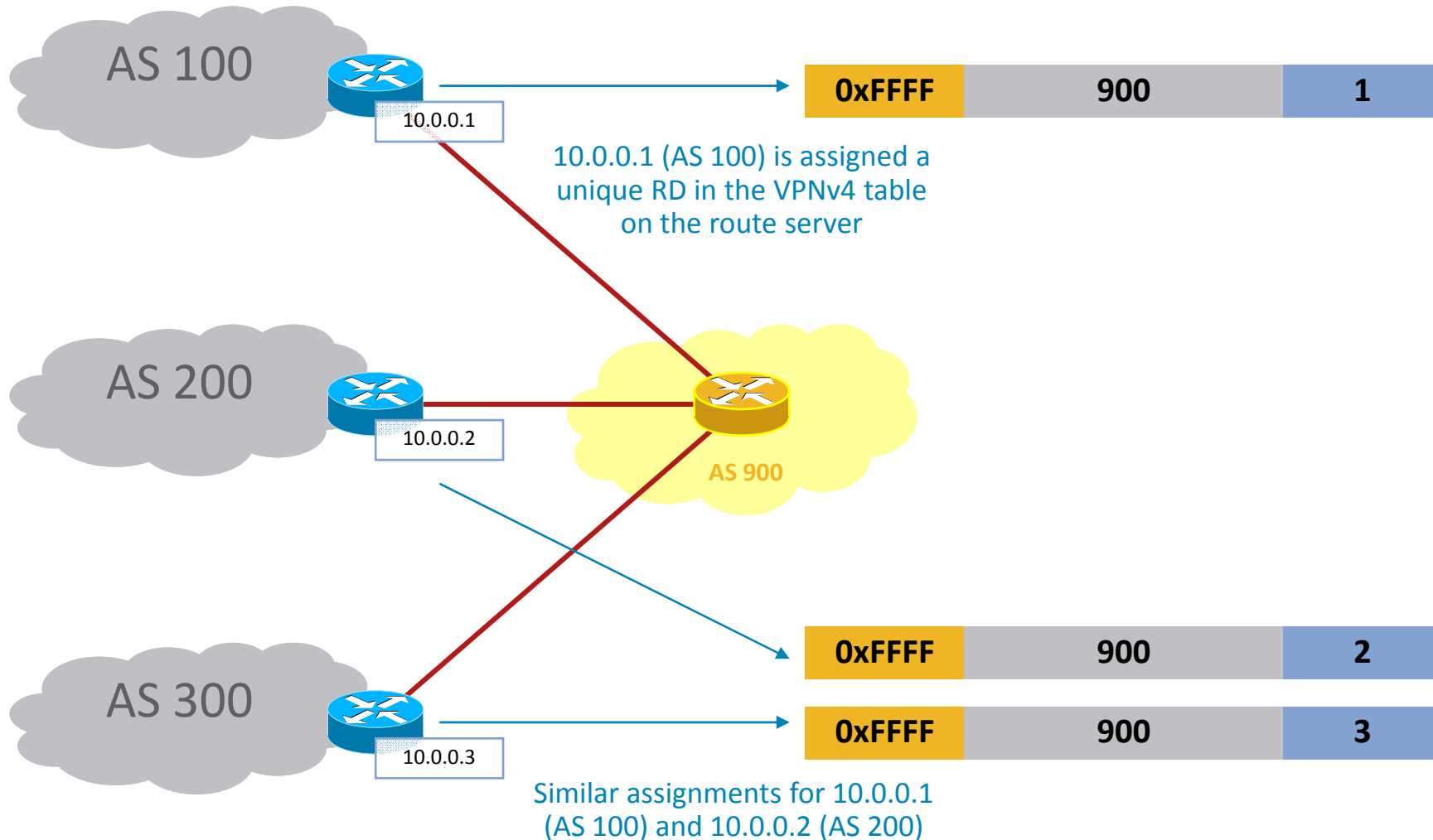
0x0	ASN(2B)	Number
0x1	A.B.C.D	Number
0x2	ASN(4B)	Number

Route server unique RD:

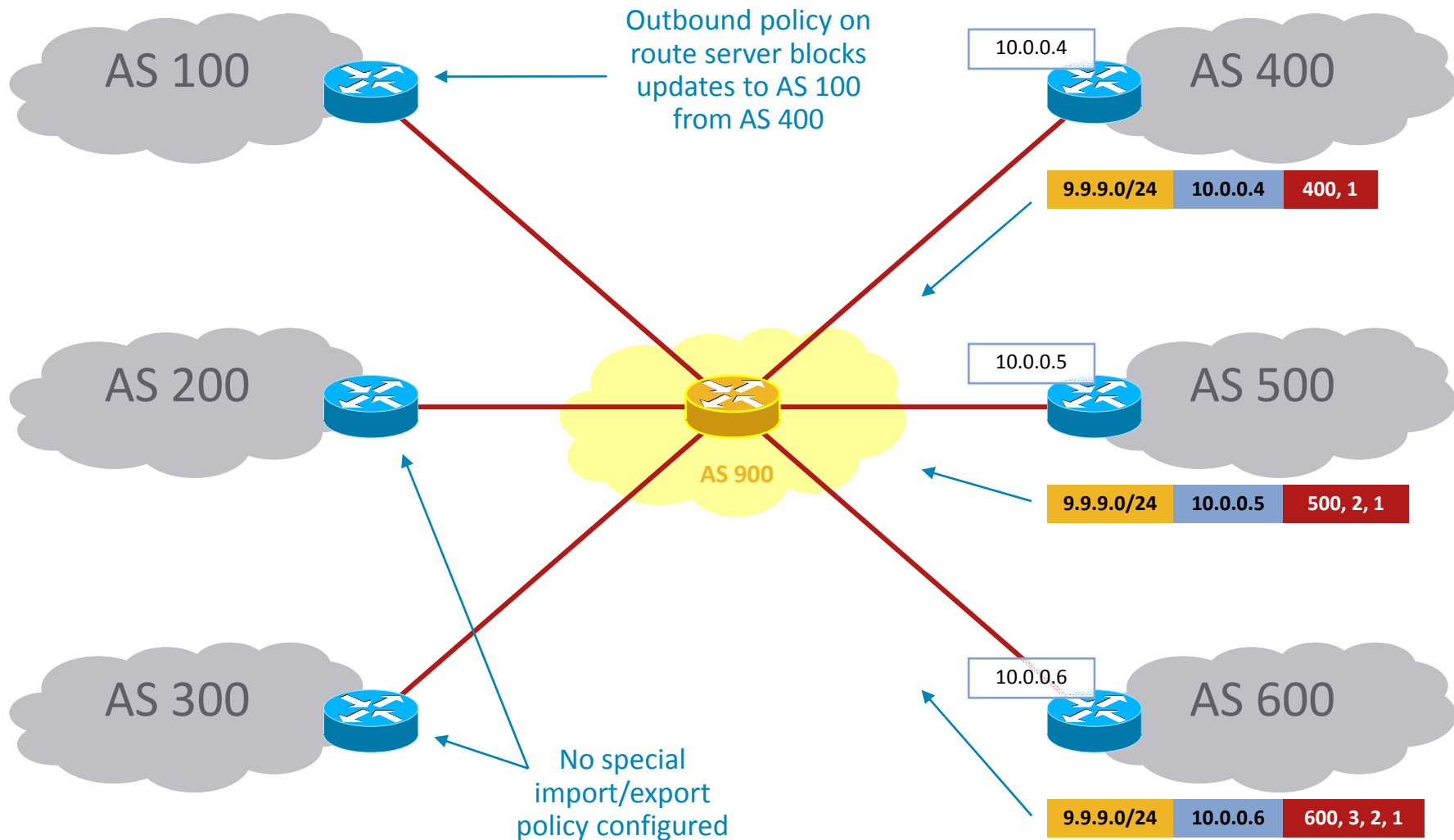
0xFFFF	IX-ASN	Number
---------------	--------	--------

- Virtual table (unique RD in VPNv4 table) and scope/topology have no associated RIB/FIB – control plane only

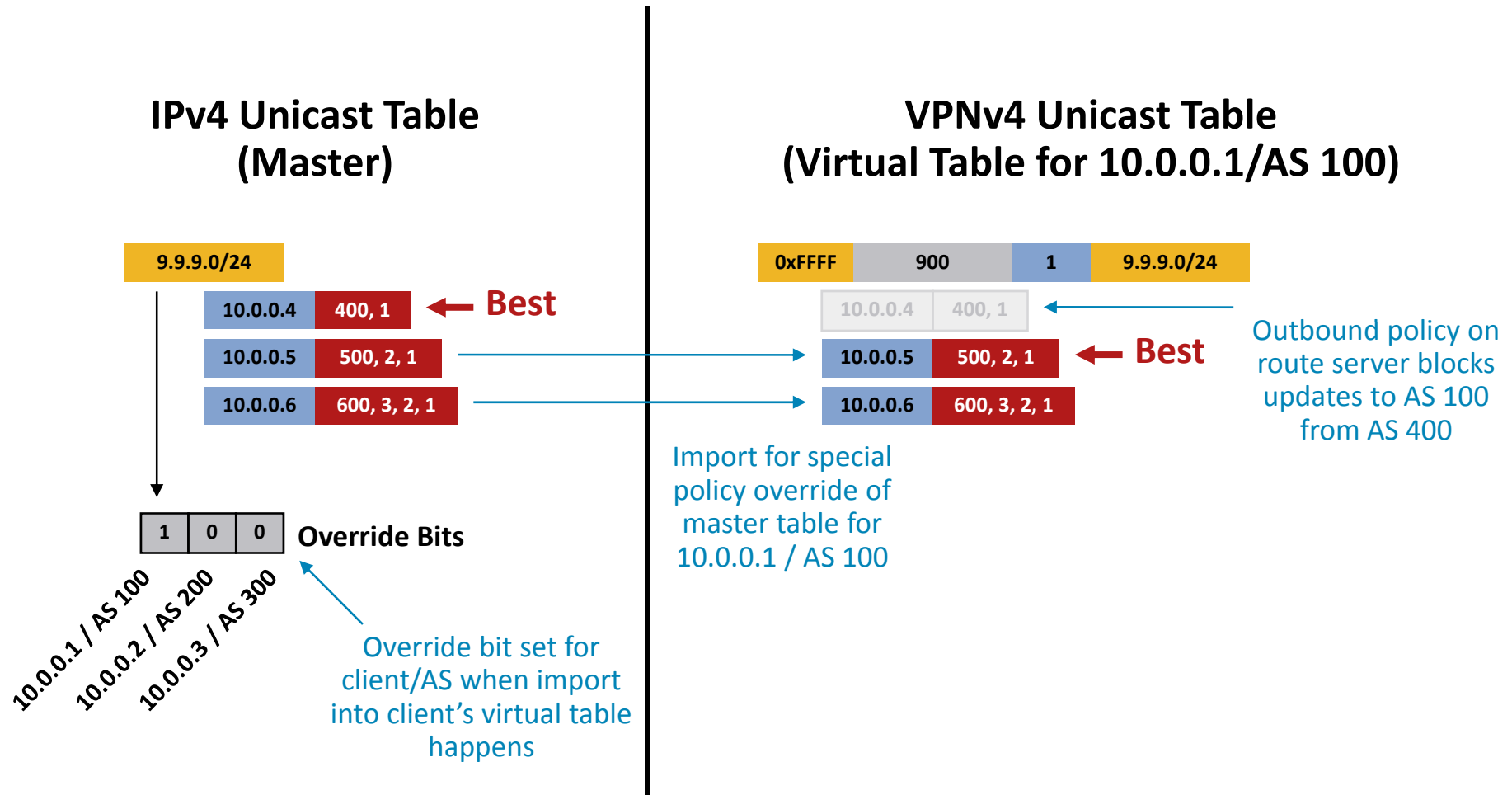
Unique ID Auto-Assignment to Clients



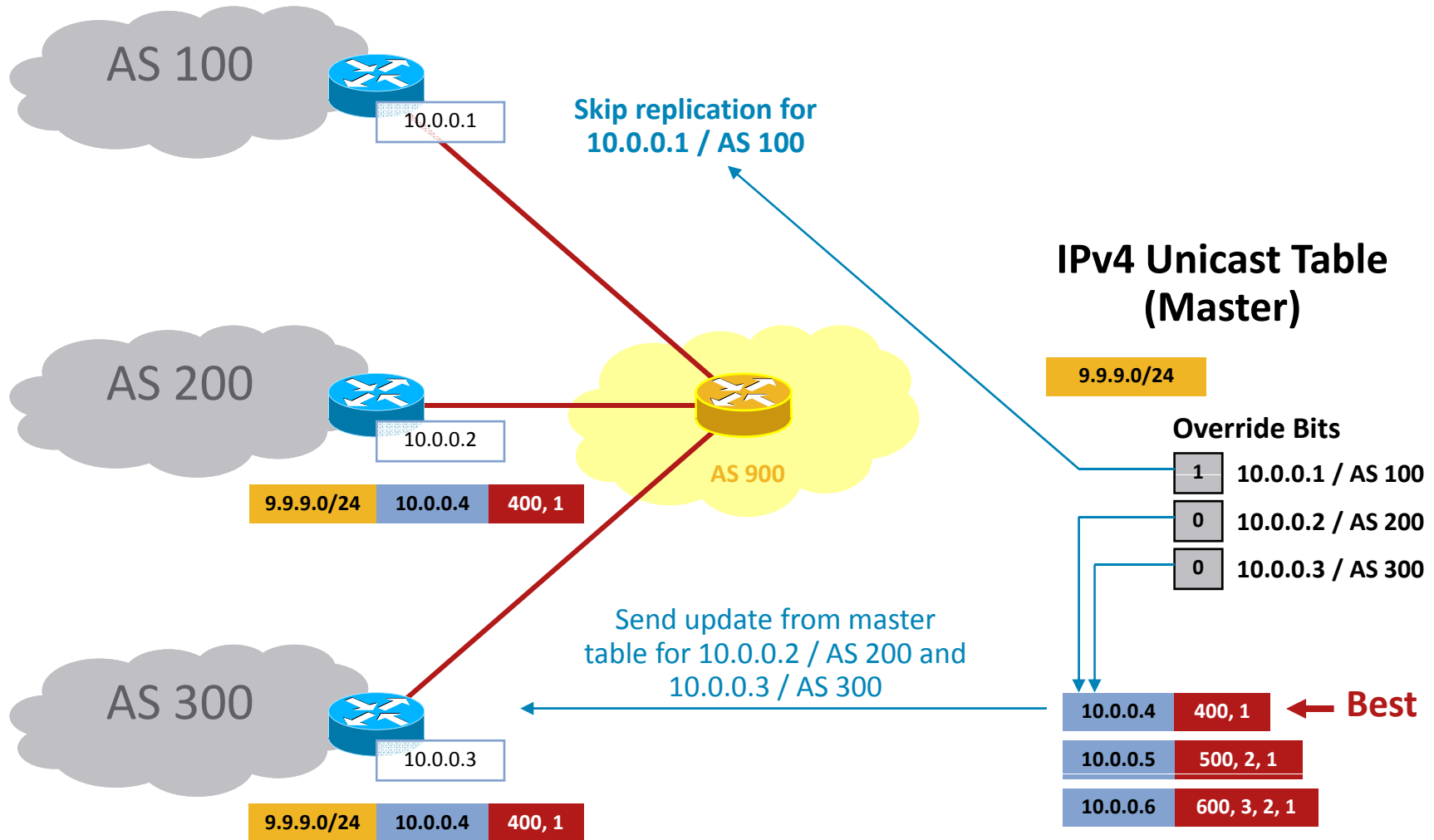
Per-Client/AS Bestpath Support



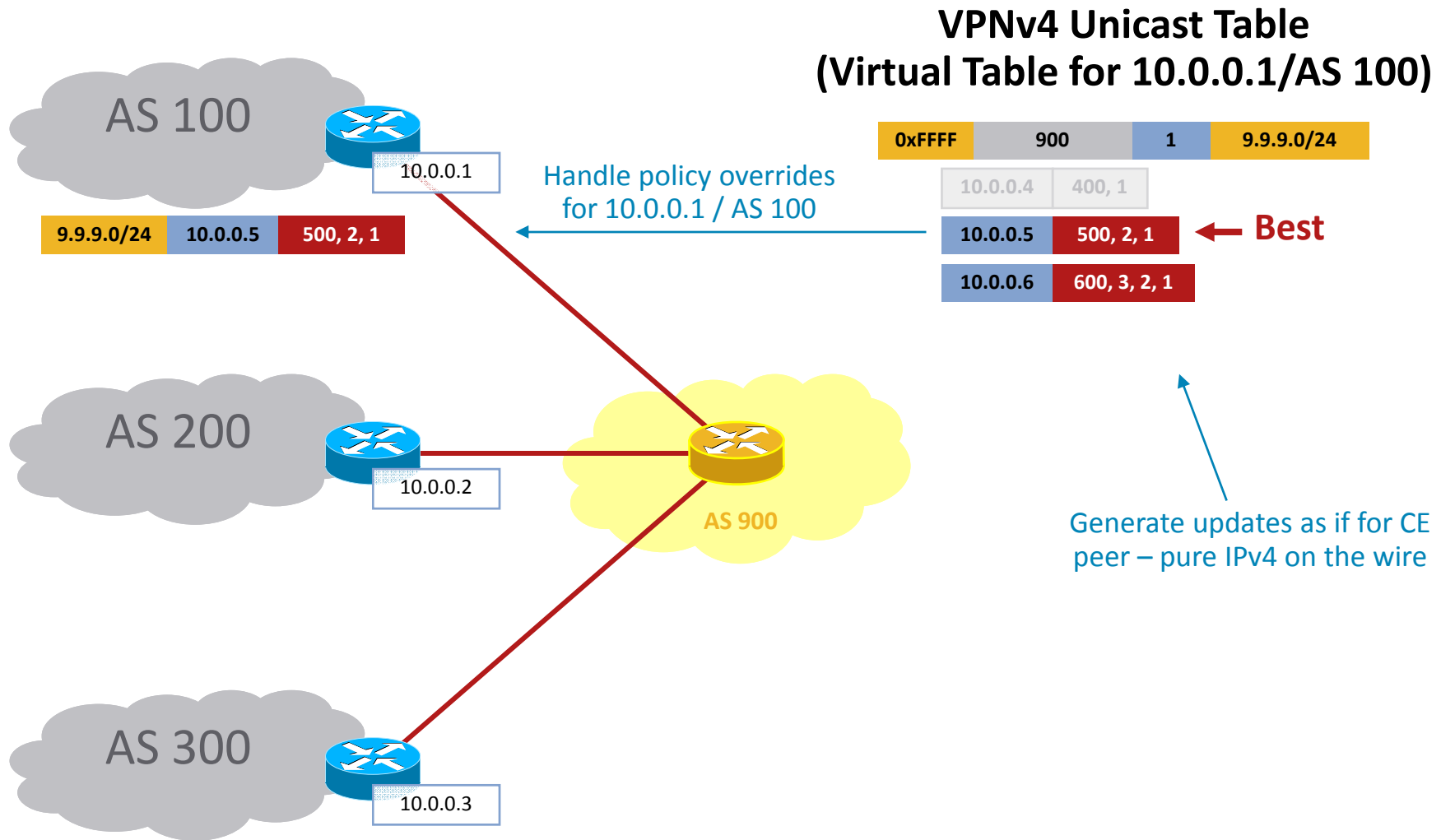
Master and Per-Client/AS BGP Tables



Master Table Update Generation



Virtual Table Update Generation



Route Server Design Summary (1)

- Each route server client is active in two AFs – IPv4 and VPNv4
- Virtual table/RD/scope assigned in VPNv4 table for each route server client
- The VPNv4 bgp_nbr_topo and associated scope and topology structures are created route server CLI parsed

Route Server Design Summary (2)

- Update generation on IPv4 table (master) handles updates when no special per-client/AS override policy applies
 - Replication skipped for route server clients that have policy override on a per prefix basis
- Update generation on VPNv4 table (virtual tables) handles updates that vary depending on per-client/AS policies
 - Updates for prefixes with special policy “fill in the gaps” skipped for certain route server clients in the IPv4 update generation run
- 1 update group with all route server clients for IPv4; 1 update group per-client/AS for VPNv4

Further RS Reading

- Provider BGP Community Examples

<http://www.onesi.net/communities/>

