The bridge to possible

Exploring the Inner Workings of OSPF

Elvin Arias Soto
High Touch Engineer @Cisco

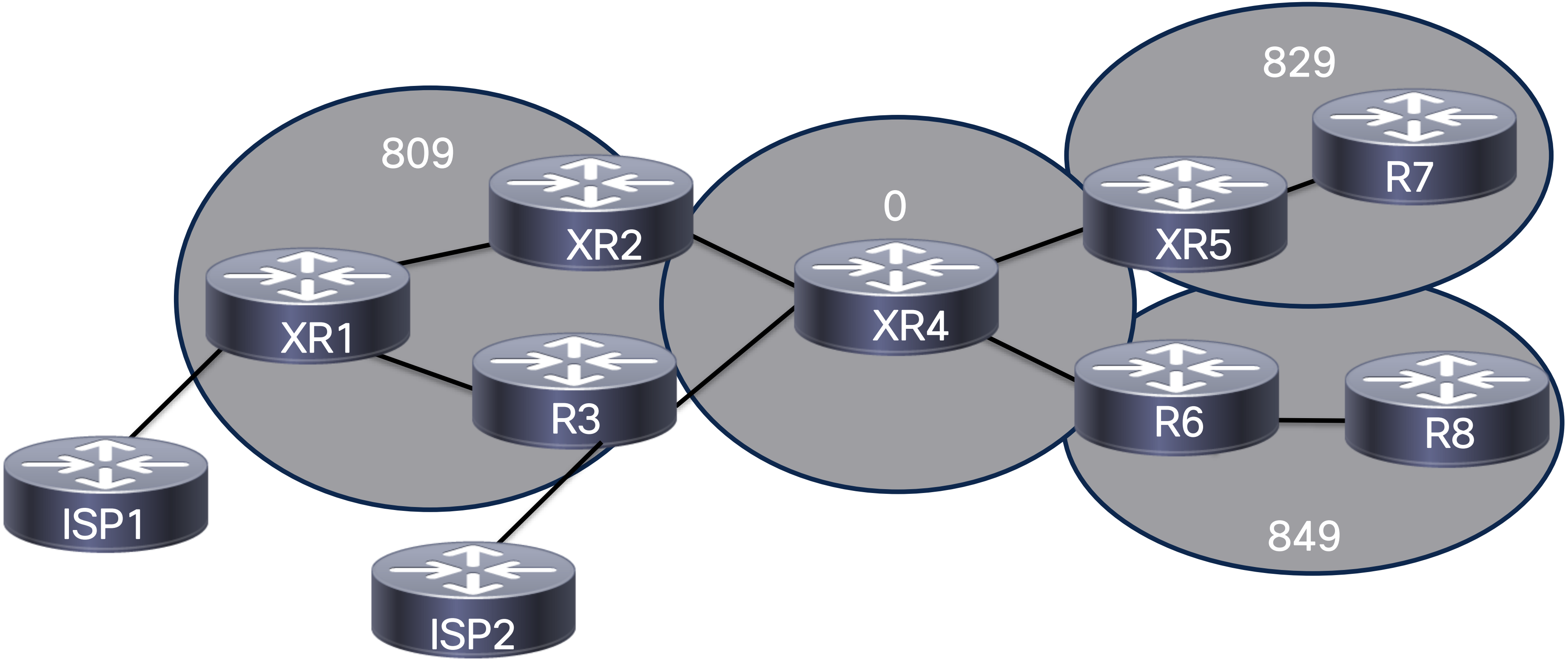
BRKENT-2088

Agenda

- Introduction
- Deep Dive into OSPF Mechanics
  - Router Roles
  - Packet Types
  - LSA Types
  - Network Types, Adjacencies, Designated Router
  - LSDB Synchronization
- Inter-Area Routing
  - Special Areas
- Path Selection
- Security Hardening
- Optimization Features
- Fun: Stupid Routing Tricks!

OSPF Routing Overview

Final Topology



What is OSPF? (1) ☺

- The Open Shortest Path First (OSPF) is an Interior Gateway Protocol (IGP) currently defined in RFC 2328. Offers many benefits such as:
  - High scalability
  - Extensibility
  - Feature Richness
  - Operational flexibility
  - Security

What is OSPF? (2) ☺

- Each router sends information about its own directly connected links to all other routers in the network
- All routers use this information to build a complete map of the network topology
- Routing decisions are then made based on this complete picture of the network, considering link speed, cost, and reliability
- Link-state protocols: OSPF, IS-IS

Areas & Router Roles

What is an area?

- Areas are a logical partition of an autonomous system (AS)
- Areas are introduced to put a boundary on the explosion of link-state updates. Floods and calculation of the Dijkstra algorithm on a router is limited to changes within an area
- All routers within an area have the exact link-state database. Routers that belong to multiple areas, and connect these areas to the backbone area are called area border routers (ABR)

Router Roles

- Internal Router
- Backbone Router
- Area Border Router (ABR)
- Autonomous System Boundary Router (ASBR)
- Designated Router (DR)
- Backup Designated Router (BDR)

Area Role Verification: IOS-XR

```
RP/0/0/CPU0:r2#show ospf
Routing Process "ospf 1" with ID 2.2.2.2
 Role: Primary Active
 NSR (Non-stop routing) is Enabled
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
 It is an area border and autonomous system boundary router
<snip>
```

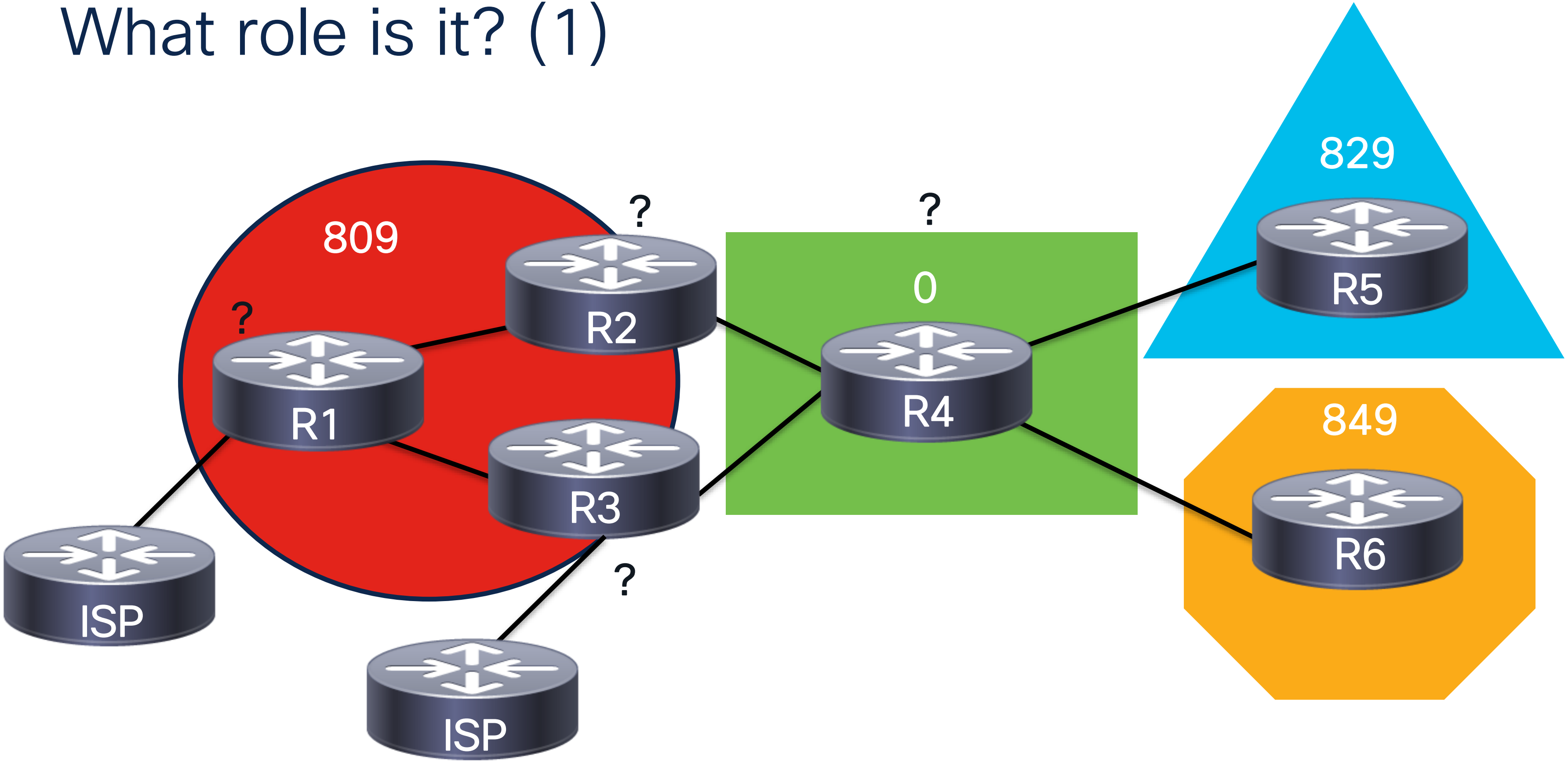
```
RP/0/0/CPU0:r2#show ospf database router self-originate
OSPF Router with ID (2.2.2.2) (Process ID 1)
                Router Link States (Area 0)
  LS age: 5
  Options: (No TOS-capability, DC)
  LS Type: Router Links
  Link State ID: 2.2.2.2
  Advertising Router: 2.2.2.2
  LS Seq Number: 80000002
  Checksum: 0xad68
  Length: 36
  Area Border Router
  AS Boundary Router
```

Area Role Verification: IOS-XE

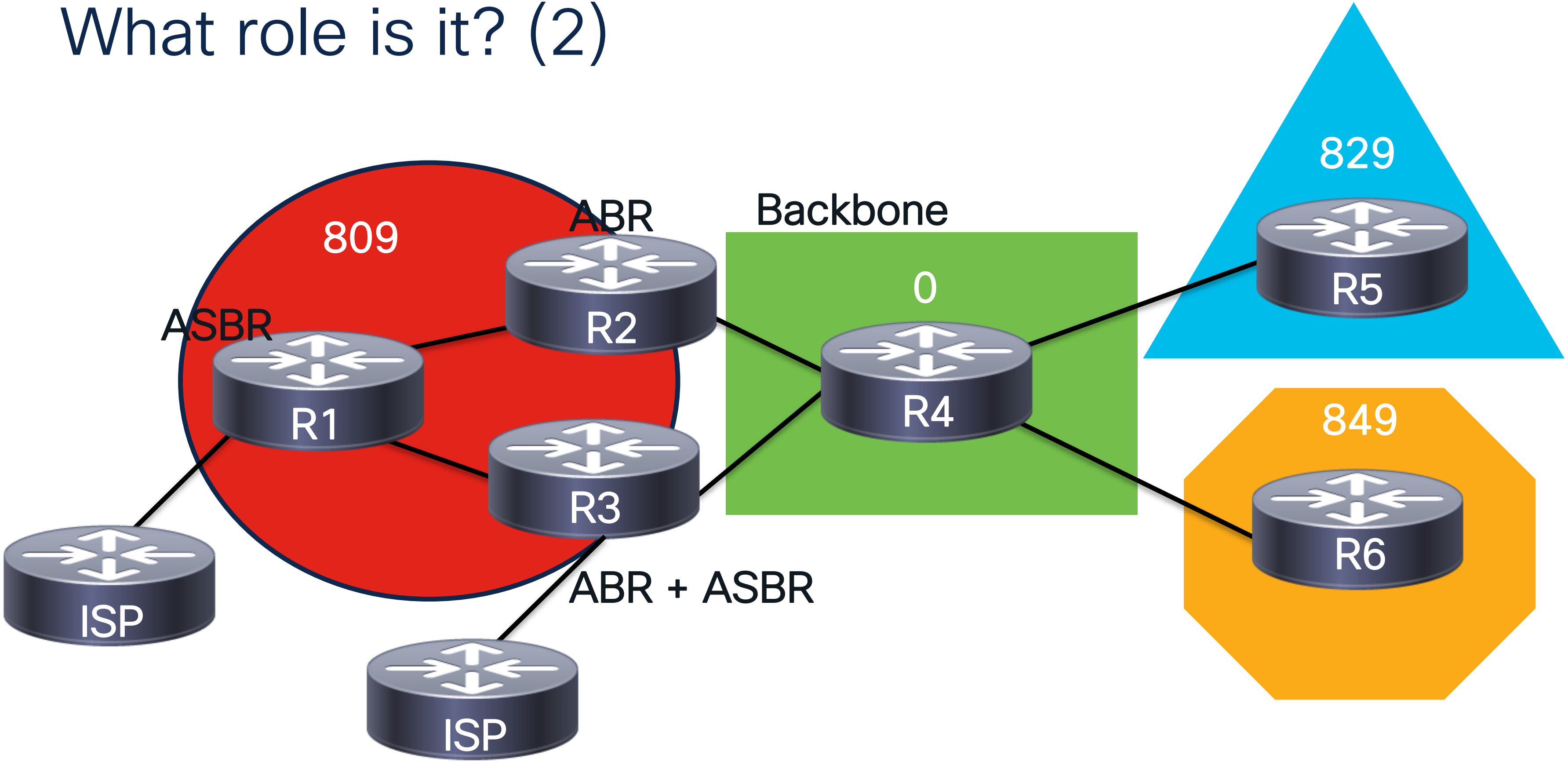
```
r1#show ip ospf
 Routing Process "ospf 1" with ID 1.1.1.1
 Start time: 00:03:52.589, Time elapsed: 3d22h
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
It is an area border and autonomous system boundary router
```

```
r1#show ip ospf database router self-originate
            OSPF Router with ID (1.1.1.1) (Process ID 1)
                  Router Link States (Area 0)
  LS age: 28
  Options: (No TOS-capability, DC)
  LS Type: Router Links
  Link State ID: 1.1.1.1
  Advertising Router: 1.1.1.1
  LS Seq Number: 80000001
  Checksum: 0xCF3D
  Length: 48
  Area Border Router
  AS Boundary Router
```

What role is it? (1)



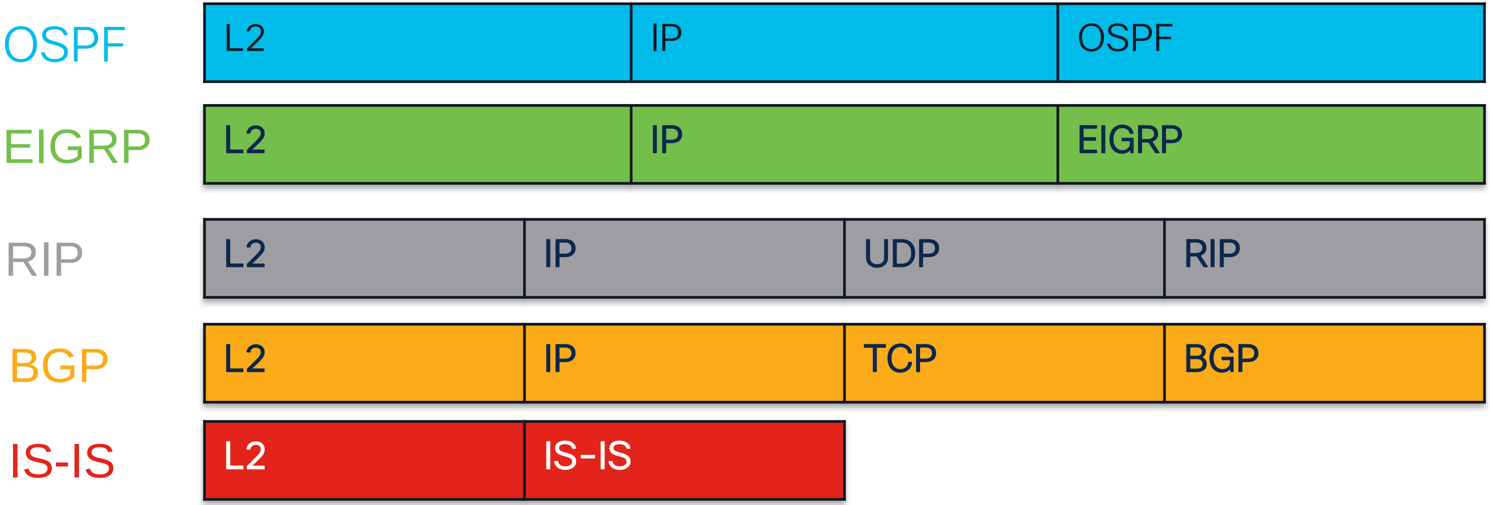
What role is it? (2)



OSPF Packet Types

Packet Encapsulation

- OSPF is encapsulated directly into L3 as an IP packet



OSPF	L2	IP	OSPF	
EIGRP	L2	IP	EIGRP	
RIP	L2	IP	UDP	RIP
BGP	L2	IP	TCP	BGP
IS-IS	L2	IS-IS		

Packet Communication

- OSPF Packets are sent over media using one of the following MAC addresses:

Name	Destination MAC	Destination IP
AllSPFRouters	0100.5e00.0005	224.0.0.5
AllDRouters	0100.5e00.0006	224.0.0.6

Packet Types (1)

- All packet types have a **common** 24-bit header that includes fields:
  - Version
  - Type
  - Packet Length
  - Router-ID
  - Area-ID
  - Checksum
  - Authentication Type
  - Authentication

Packet Types (2)

- Type 1 – Hello
- Type 2 - Database descriptors (DBD)
- Type 3 - Link-state request (LSR)
- Type 4 - Link-state update (LSU)
- Type 5 - Link-state acknowledgement (LSA)

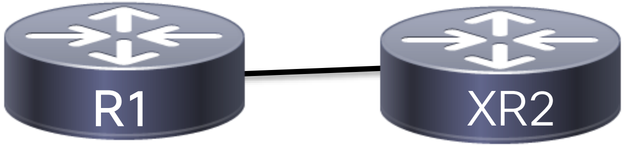
Type 1 - Hello Packets

- Contains parameters that allow for discovery of OSPF-capable routers to form adjacencies (or neighbor relationships!) in the segment
- Neighbor detection and maintenance
- Used to perform DR/BDR election in multi-access networks (Broadcast / NBMA)
- Hello periodicity may vary depending the network type configured

Note: Timers must match for adjacency to be established

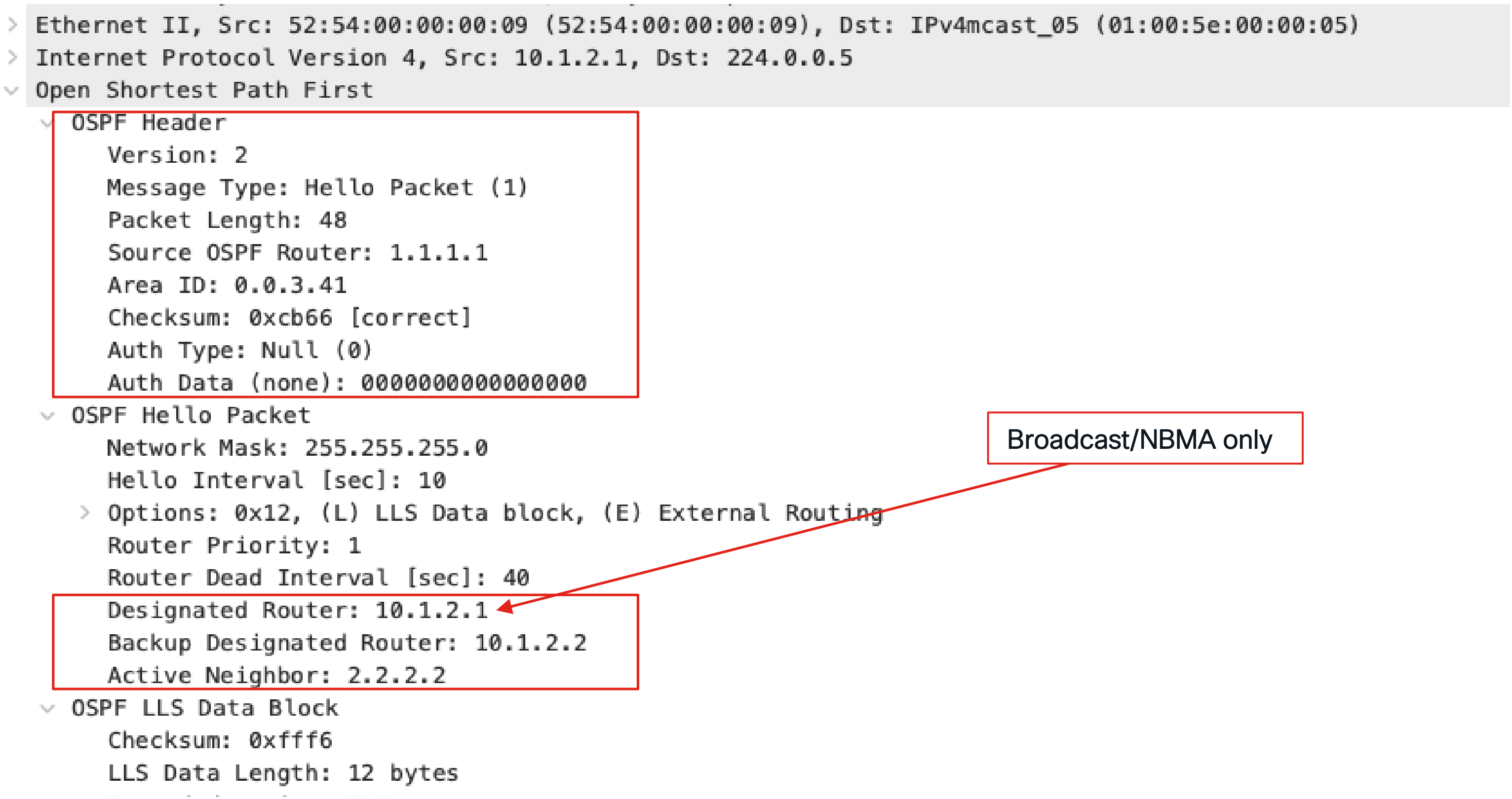
Hello Configuration: IOS-XR

```
router ospf 1
 area 809
  interface GigabitEthernet0/0/0/0
   dead-interval 33
   hello-interval 11
```



Must match between R1 and XR2!

Type 1 – Hello

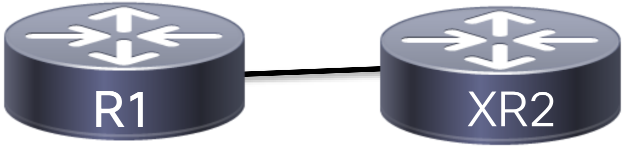


```
> Ethernet II, Src: 52:54:00:00:00:09 (52:54:00:00:00:09), Dst: IPv4mcast_05 (01:00:5e:00:00:05)
> Internet Protocol Version 4, Src: 10.1.2.1, Dst: 224.0.0.5
v Open Shortest Path First
   v OSPF Header
        Version: 2
        Message Type: Hello Packet (1)
        Packet Length: 48
        Source OSPF Router: 1.1.1.1
        Area ID: 0.0.3.41
        Checksum: 0xcb66 [correct]
        Auth Type: Null (0)
        Auth Data (none): 0000000000000000
   v OSPF Hello Packet
        Network Mask: 255.255.255.0
        Hello Interval [sec]: 10
      > Options: 0x12, (L) LLS Data block, (E) External Routing
        Router Priority: 1
        Router Dead Interval [sec]: 40
        Designated Router: 10.1.2.1
        Backup Designated Router: 10.1.2.2
        Active Neighbor: 2.2.2.2
   v OSPF LLS Data Block
        Checksum: 0xfff6
        LLS Data Length: 12 bytes
```

Broadcast/NBMA only

Type 2 - Database descriptor

- Exchanged when the adjacency is initialized (**ExStart**) to describe the link-state database
- Contains a brief description of the router's advertisements to allow for database synchronization by the election of Master/Slave relationships in the **OSPF** adjacency process
- Holds the Maximum Transmission Unit (**MTU**) of the OSPF-enabled interface



Type 3 - Link-state request (LSR)

- After reviewing database descriptors (DD), the OSPF router may proceed with launching link-state requests
- Link-state requests allow for querying link-state advertisements (LSA’s) to keep the most up-to-date version of the database
- The Link State Request packet is used to request the pieces of the neighbour's database that are more up-to-date. Multiple Link State Request packets may need to be used.

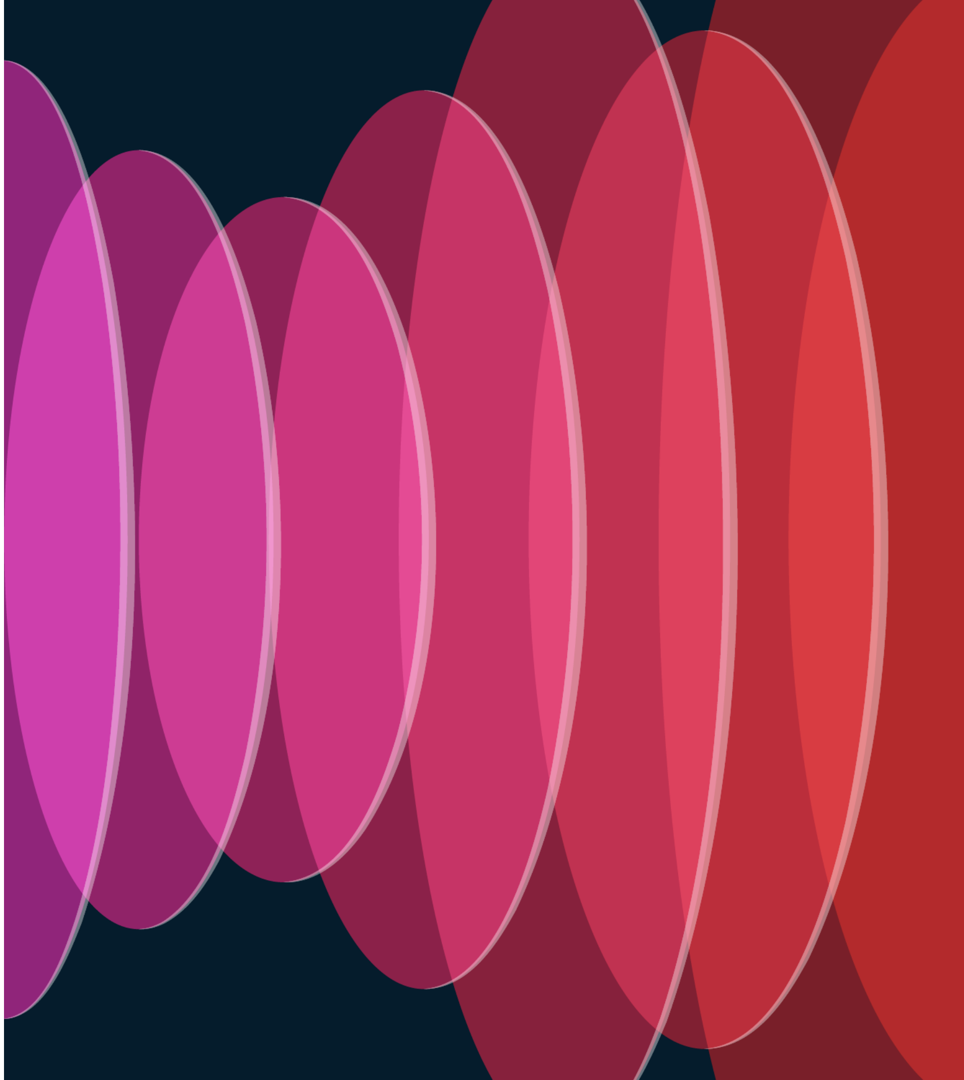
Type 4 - Link-state update (LSU)

- Allow for flooding link-state advertisements (LSA) in OSPF
- LSUs contain one or multiple link-state advertisements (LSA’s) and are sent as a response to link-state requests or due to network events that grant

Type 5 - Link-state acknowledgement

- To confirm the receipt of link-state updates (LSU) packets
- Allows to have reliable exchange of LSU during initial link-state database synchronization and network events
- Link-state acknowledgements contain the LSA headers that have been received

Visualizing Packet Types



LSA Types

Link-state advertisements (LSA)

- LSAs convey network-layer reachability information alongside with topological information about the routing domain - Different LSA types exist:
  - Type 1- Router-LSA
  - Type 2 – Network-LSA
  - Type 3 – NetSummary-LSA
  - Type 4 - ASBR Summary-LSA
  - Type 5 – External-LSA
  - Type 7 – NSSA-External-LSA
  - Type 9/10/11 – Opaque-LSAs (Link, Area, Domain)

Type 1 – Router LSA

- As OSPF is enabled in a router will always flood a Router LSA, network layer reachability information (NLRI) and topological information (adjacency descriptions)

```
r1#show ip ospf database router self-originate
            OSPF Router with ID (1.1.1.1) (Process ID 1)
                  Router Link States (Area 809)
  LS age: 1150
  Options: (No TOS-capability, DC)
  LS Type: Router Links
  Link State ID: 1.1.1.1
  Advertising Router: 1.1.1.1
  LS Seq Number: 80000031
  Checksum: 0xF2A9
  Length: 72
  Number of Links: 4
    Link connected to: a Stub Network
     (Link ID) Network/subnet number: 1.1.1.1
     (Link Data) Network Mask: 255.255.255.255
      TOS 0 Metrics: 1
<snip>
```

Type 2 – Network-LSA

- Network-LSA is originated by the DR and lists the collection of nodes in the multi-access segment

```
r2#show ip ospf database network self-originate

            OSPF Router with ID (2.2.2.2) (Process ID 1)

                  Net Link States (Area 809)
  LS age: 284
  Options: (No TOS-capability, DC)
  LS Type: Network Links
  Link State ID: 10.1.2.2 (address of Designated Router)
  Advertising Router: 2.2.2.2
  LS Seq Number: 80000001
  Checksum: 0x21F5
  Length: 32
  Network Mask: /24
         Attached Router: 2.2.2.2 < r2
         Attached Router: 1.1.1.1 < r1
```

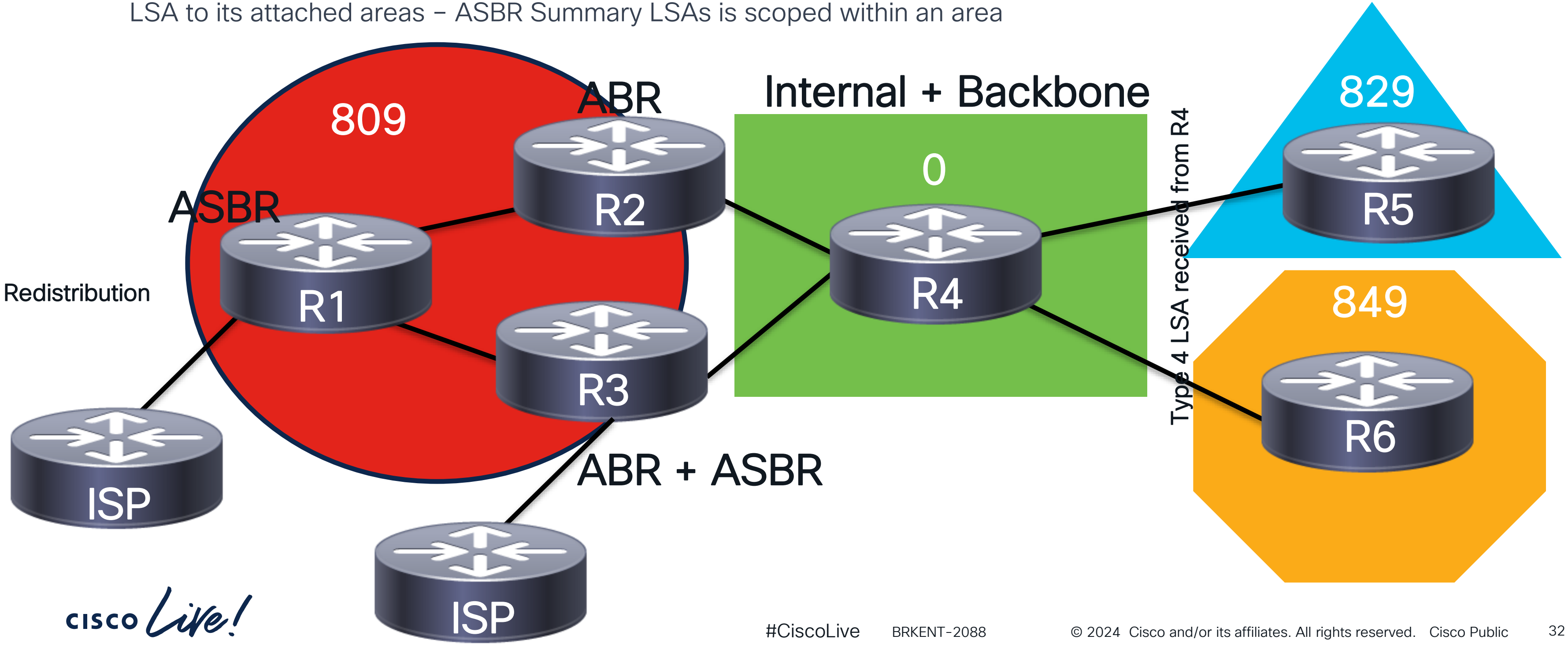
Type 3 – NetSummary LSA

- Conveys prefix information as information is sent between areas through area border routers (ABRs)
- Most confusing LSA name is **NetSummary LSA** ☺ as it does not summarize prefixes, but embodies the simplification of reachability information between areas

```
r2#show ip ospf database summary 2.2.2.222 self-originate

            OSPF Router with ID (2.2.2.2) (Process ID 1)

                  Summary Net Link States (Area 809)

  LS age: 912
  Options: (No TOS-capability, DC, Upward)
  LS Type: Summary Links(Network)
  Link State ID: 2.2.2.222 (summary Network Number)
  Advertising Router: 2.2.2.2
  LS Seq Number: 80000001
  Checksum: 0x5AF4
  Length: 28
  Network Mask: /32
         MTID: 0   Metric: 1
```

Type 4 – ASBR Summary LSA

- Generated by an ABR to signal the areas that it knows how to reach ASBR
- When the **E-bit** is set in the Router LSA to signal that the local router is an ASBR, the ABR will generate a Type 4 LSA to its attached areas – ASBR Summary LSAs is scoped within an area



Type 5 – External LSA

- External LSAs are originated from an ASBR, they describe stub/IP prefixes as originated from external domains through redistribution
- The scope of External LSAs are from within the Autonomous System (AS)

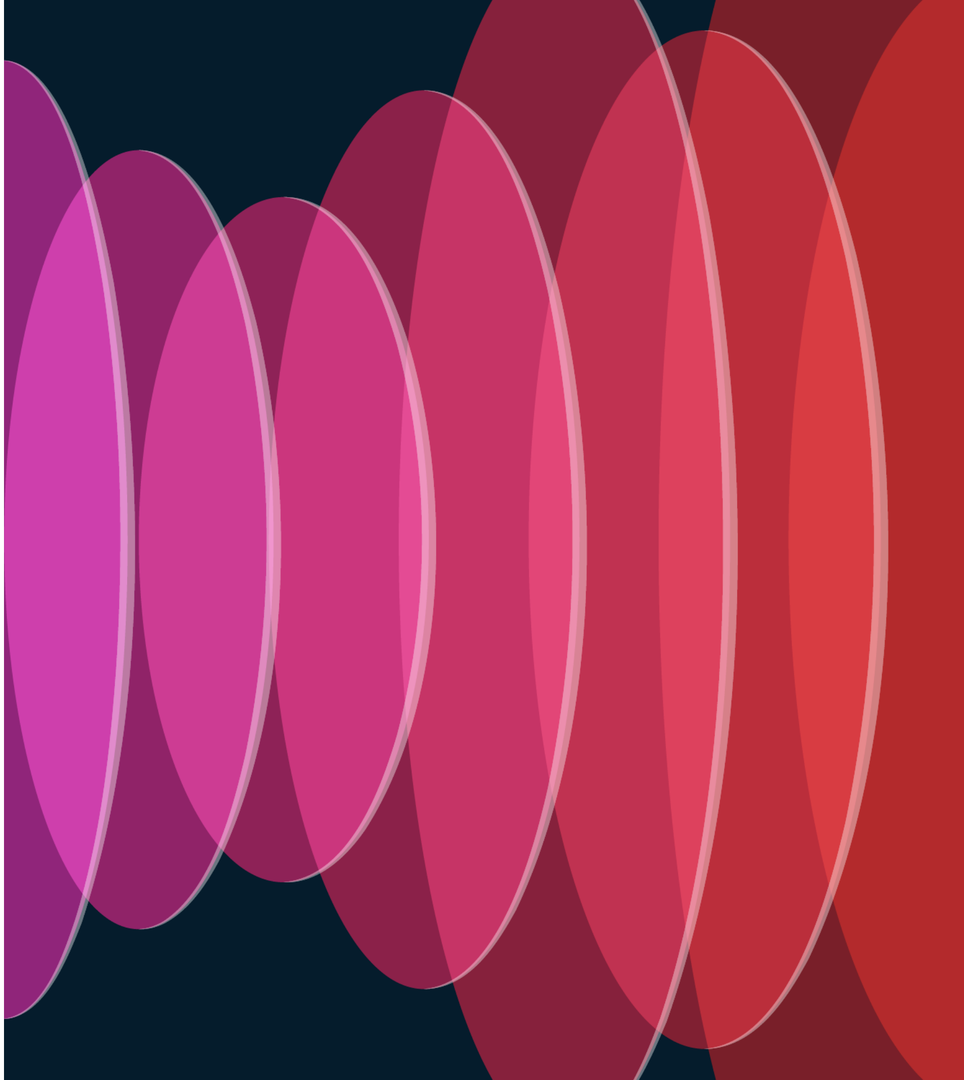
```
r1#show ip ospf database external
            OSPF Router with ID (1.1.1.1) (Process ID 1)
                  Type-5 AS External Link States
  LS age: 5
  Options: (No TOS-capability, DC, Upward)
  LS Type: AS External Link
  Link State ID: 1.1.1.11 (External Network Number )
  Advertising Router: 1.1.1.1
  LS Seq Number: 80000002
  Checksum: 0xE2B3
  Length: 36
  Network Mask: /32
         Metric Type: 2 (Larger than any link state path)
         MTID: 0
         Metric: 10
         Forward Address: 0.0.0.0
         External Route Tag: 1
```

Type 7 – NSSA External LSA

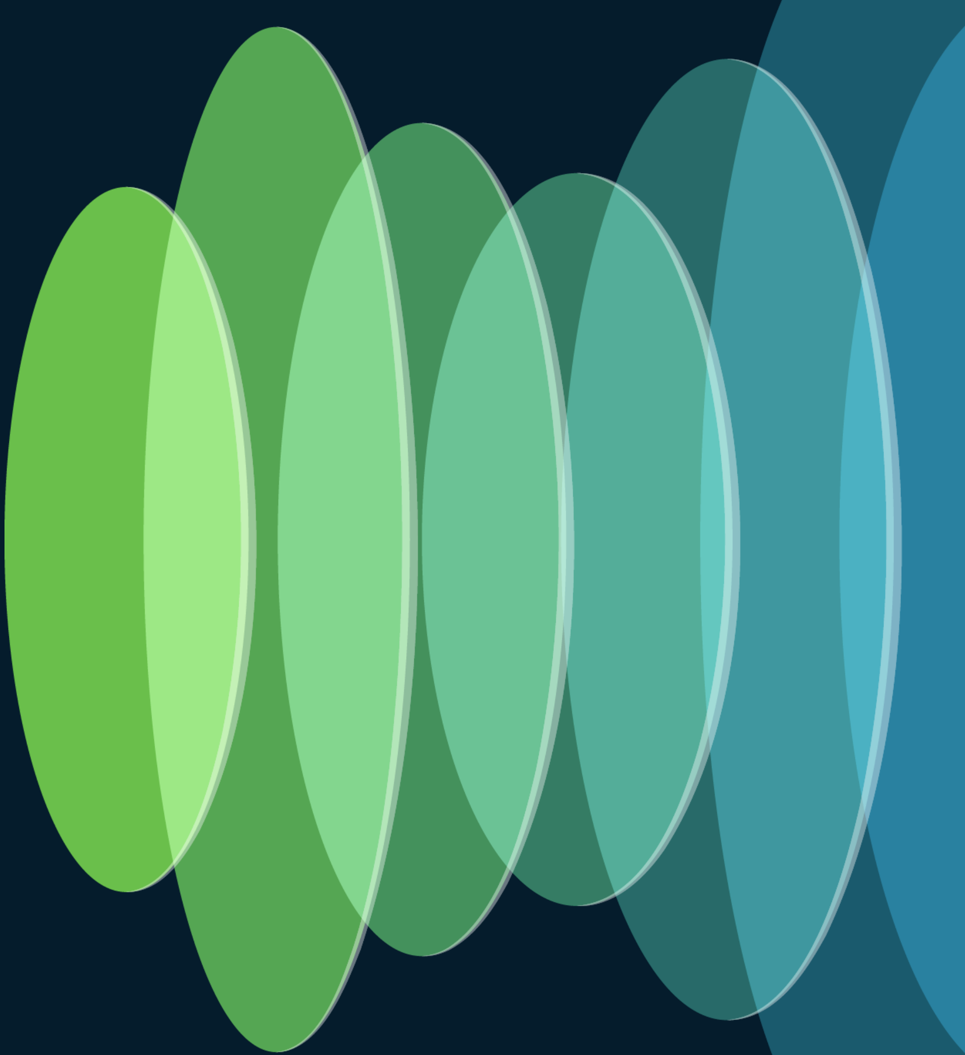
- Special use-case LSA for Not-So-Stubby Areas (NSSA) scenarios
- Used to allow redistribution of external routing sources within an NSSA
- The scope of a Type 7 LSA is within the area it was originated

```
r1#show ip ospf database nssa-external
            OSPF Router with ID (1.1.1.1) (Process ID 1)
                  Type-7 AS External Link States (Area 809)
  LS age: 19
  Options: (No TOS-capability, Type 7/5 translation, DC, Upward)
  LS Type: AS External Link
  Link State ID: 1.1.1.11 (External Network Number )
  Advertising Router: 1.1.1.1
  LS Seq Number: 80000001
  Checksum: 0x8207
  Length: 36
  Network Mask: /32
         Metric Type: 2 (Larger than any link state path)
         MTID: 0
         Metric: 10
         Forward Address: 1.1.1.1
         External Route Tag: 1
```

LSA Verification



Network Types, Adjacencies, Designated Router



Network Types (1)

- Several network types are supported:
  - Broadcast
  - Point-to-point
  - Non-broadcast multiaccess (NBMA)
  - Point-to-multipoint
  - Point-to-multipoint non-broadcast
  - Loopback

Network Types (2)

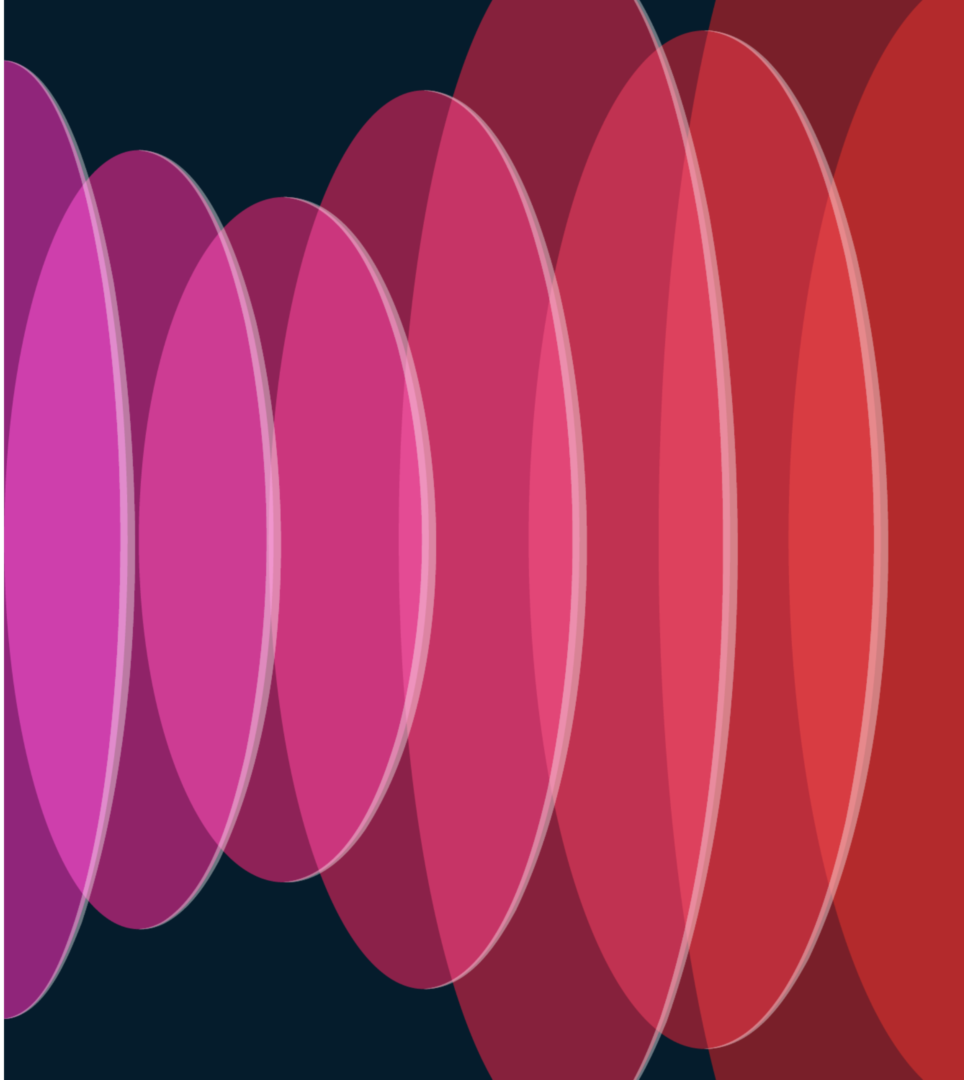
IOS-XE

```
r1(config)#interface gigabitEthernet 0/0
r1(config-if)#ip ospf network ?
  broadcast            Specify OSPF broadcast multi-access network
  non-broadcast        Specify OSPF NBMA network
  point-to-multipoint  Specify OSPF point-to-multipoint network
  point-to-point       Specify OSPF point-to-point network
```

IOS-XR

```
RP/0/0/CPU0:r1(config)#router ospf 1
RP/0/0/CPU0:r1(config-ospf)#area 0
RP/0/0/CPU0:r1(config-ospf-ar)#interface gigabitEthernet 0/0/0/0
RP/0/0/CPU0:r1(config-ospf-ar-if)#network ?
  broadcast            Specify OSPF broadcast multi-access network
  non-broadcast        Specify OSPF NBMA network
  point-to-multipoint  Specify OSPF point-to-multipoint network
  point-to-point       Specify OSPF point-to-point network
```

Configuring Network Types

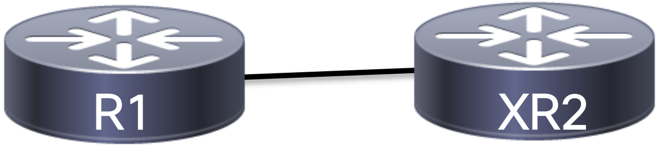


Adjacency Requirements (1)

Parameters	Must Match	Must be Unique
Router-ID	-	Yes
Area ID	Yes	-
Subnet Mask	Yes – Only in Broadcast	-
Stub area flag	Yes	-
Hello/Dead intervals	Yes	-
MTU	[Yes]	-
Authentication	Yes	-

Adjacency Requirements (2)

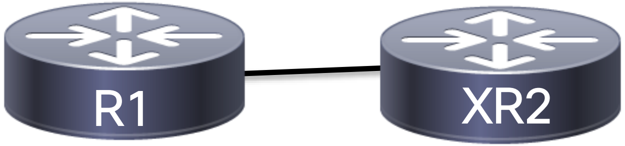
Mismatch hello / dead intervals



```
r2(config-if)#ip ospf dead-interval 11
r2(config-if)#end
<snip>
*Apr 20 18:30:32.384: OSPF-1 HELLO Gi0/0: Send hello to 224.0.0.5 area 809
from 10.1.2.2
*Apr 20 18:30:33.743: %SYS-5-CONFIG_I: Configured from console by console
*Apr 20 18:30:36.824: OSPF-1 HELLO Gi0/0: Rcv hello from 1.1.1.1 area 809
10.1.2.1
*Apr 20 18:30:36.824: OSPF-1 HELLO Gi0/0: Mismatched hello parameters from
10.1.2.1
*Apr 20 18:30:36.824: OSPF-1 HELLO Gi0/0: Dead R 40 C 11, Hello R 10 C 10
```

Adjacency Requirements (3)

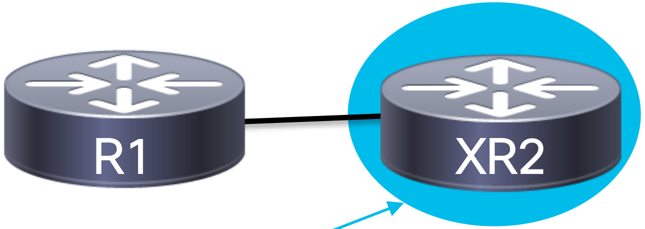
Router-ID



```
r2(config)#router ospf 1
r2(config-router)#router
r2(config-router)#router-id 1.1.1.1 << Same RID as R1
% OSPF: Reload or use "clear ip ospf process" command, for this to take
effect
r2(config-router)#end
Reset ALL OSPF processes? [no]: yes
<snip>
*Apr 20 18:35:40.180: %OSPF-5-ADJCHG: Process 1, Nbr 1.1.1.1 on
GigabitEthernet0/0 from FULL to DOWN, Neighbor Down: Interface down or d
*Apr 20 18:35:42.306: %OSPF-4-DUP_RTRID_NBR: OSPF detected duplicate router-
id 1.1.1.1 from 10.1.2.1 on interface GigabitEthernet0/0
```

Adjacency Requirements (4)

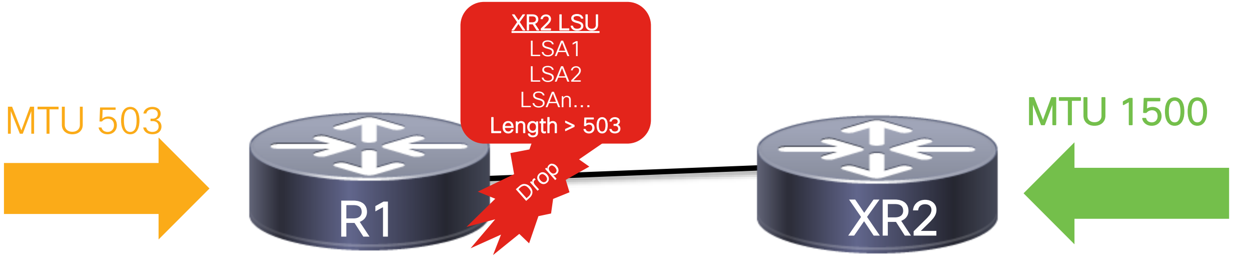
Stub area flag



```
router ospf 1
 router-id 2.2.2.2
 area 809 nssa
<snip>

*Apr 20 18:56:30.544: OSPF-1 HELLO Gi0/0: Send hello to 224.0.0.5 area 809
from 10.1.2.2
*Apr 20 18:56:36.148: OSPF-1 HELLO Gi0/0: Rcv hello from 1.1.1.1 area 809
10.1.2.1
*Apr 20 18:56:36.149: OSPF-1 HELLO Gi0/0: Hello from 10.1.2.1 with mismatched
NSSA option bit
```

Adjacency Requirements (5)

MTU Mismatch

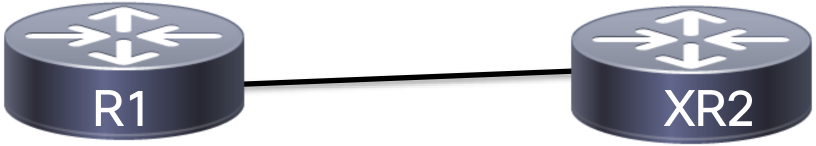


```
r1#show ip interface gigabitEthernet 0/0 | include MTU
  MTU is 513 bytes
<snip>

*Apr 21 12:39:56.140: OSPF-1 ADJ   Gi0/0: Rcv DBD from 2.2.2.2 seq 0x71E6 opt 0x52 flag 0x7 len
32  mtu 1500 state EXSTART
*Apr 21 12:39:56.140: OSPF-1 ADJ   Gi0/0: Nbr 2.2.2.2 has larger interface MTU
```

Note: Fix the MTU issue instead of ignoring the MTU with **ip ospf mtu-ignore!** ☺

Adjacency Requirements (6)

Subnet mask (Broadcast-only)



```
*Apr 20 19:00:13.619: OSPF-1 HELLO Gi0/0: Mismatched hello parameters from
10.1.2.2
*Apr 20 19:00:13.620: OSPF-1 HELLO Gi0/0: Dead R 40 C 40, Hello R 10 C 10
Mask R 255.255.255.0 C 255.255.255.252
*Apr 20 19:00:16.571: OSPF-1 HELLO Gi0/0: Send hello to 224.0.0.5 area 809
from 10.1.2.1
```

Adjacency States (1)

- **Down** – Initial state of a neighbor conversation, it indicates that there is no recent information **received** from the neighbor
- **Attempt** – Applicable to NBMA networks only. Indicates that no recent information has been received from the neighbor
- **Init** – Hello packet has been received from the neighbor, but no bidirectional communication is established
- **2-Way** – Communication is bidirectional, all parameters match, and a **neighbor relationship** is established

Adjacency States (2)

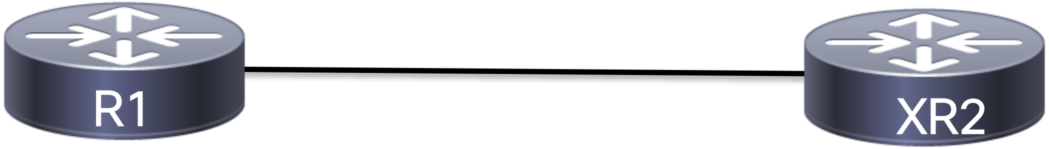
- **ExStart** – Adjacency formation starts in this state, it allows to perform the Master election for the exchange of the initial Database Descriptor sequence number
- **Exchange** – The neighbor is exchanging its link-state database by sending the DBD packets.
- **Loading** – Link state requests (LSR) packets are sent to neighbors asking for up-to-date LSAs
- **Full** – The routers are fully adjacent and database synchronization has finished

Point-to-Point Adjacency (1)

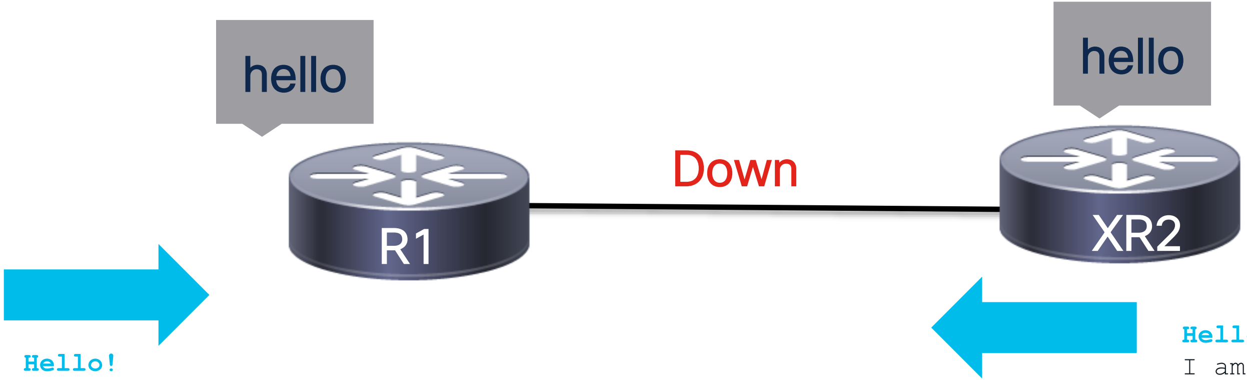
- Adjacencies are formed in a P2P environment without electing DR/BDRs
- A single communication channel is used in P2P networks, 224.0.0.5/AllSPFRouters

```
interface gigabitethernet0/0
 ip ospf network point-to-point
```

```
router ospf 1
   area 809
     interface gi0/0/0/0
       network point-to-point
```



Point-to-Point Adjacency (2) - Down

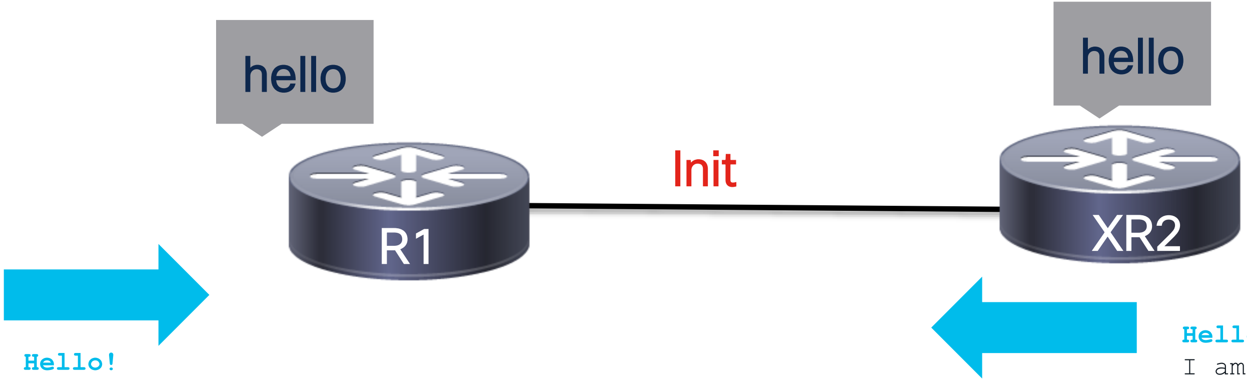


```
Hello!
I am 1.1.1.1
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

```
Hello!
I am 2.2.2.2
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

Initially routers in P2P networks are in Down state

Point-to-Point Adjacency (3) - Init

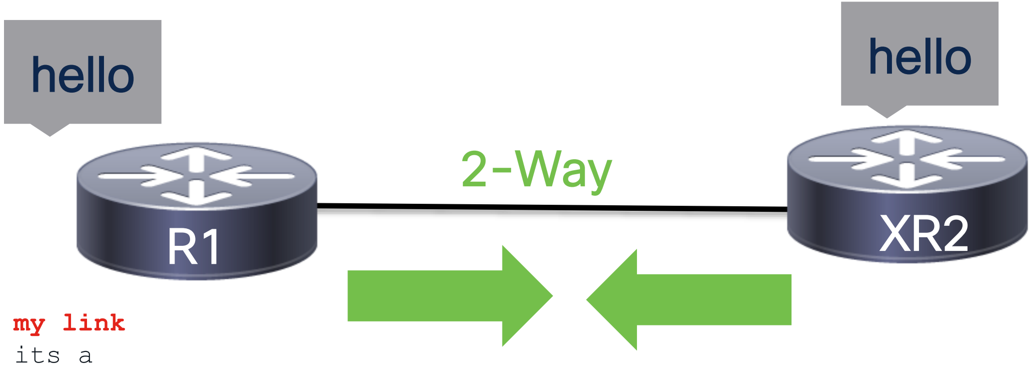


```
Hello!
I am 1.1.1.1
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

```
Hello!
I am 2.2.2.2
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

As soon as hellos are received from a neighbor, these are processed and routers will put the adjacency state as Init

Point-to-Point Adjacency (4) - 2-Way

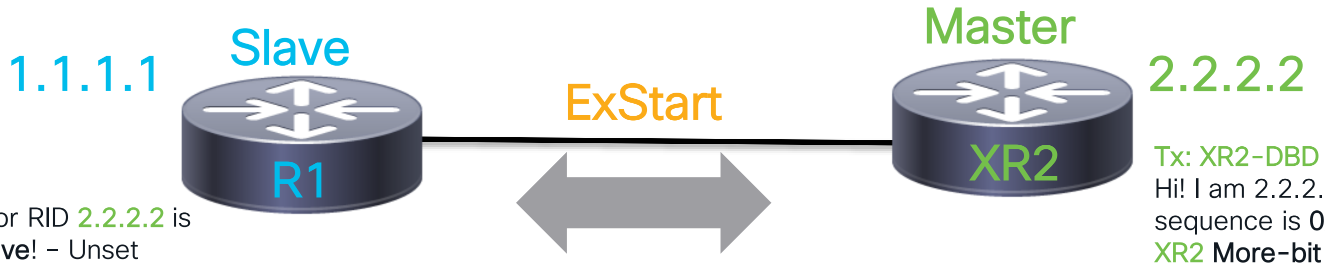


```
Hello!

I am 1.1.1.1
I see: 2.2.2.2 in my link
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

```
Hello!

I am 2.2.2.2
I see: 1.1.1.1 in my link
I am in area 809, its a
regular area
I have authentication
type: null
Hello interval is: 10s
Dead interval is: 40s
<...>
```

To acknowledge the existence of each-other, and once the adjacency parameters are validated, routers will add each other in the subsequent **Hello** packets as a sign of bidirectional communication establishment, this is 2-Way

Point-to-Point Adjacency (5) – ExStart



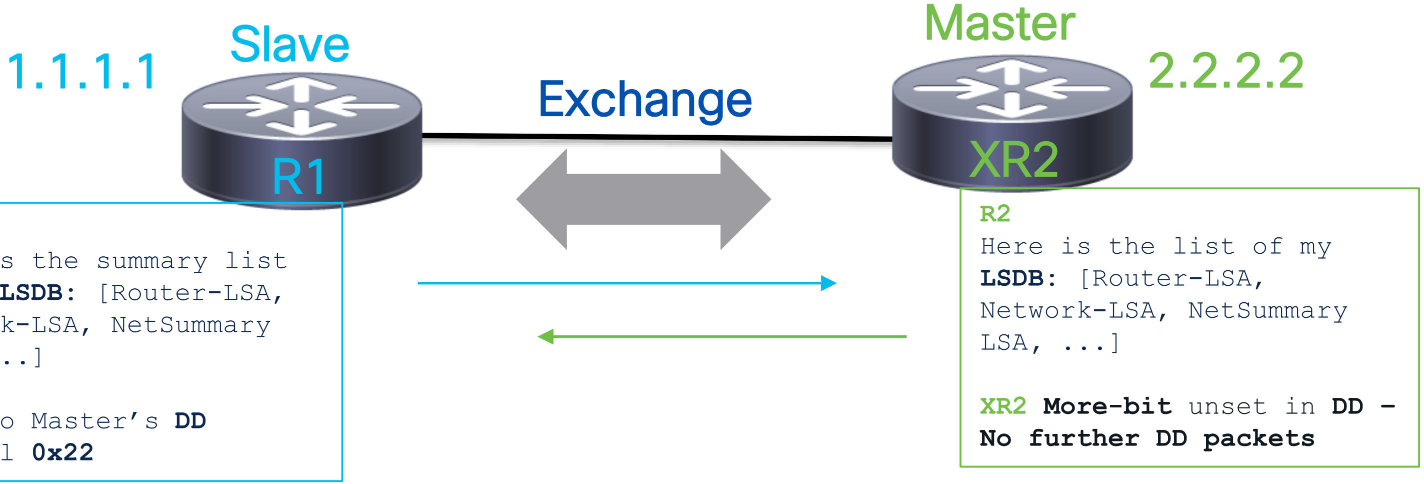
R1 knows Neighbor RID 2.2.2.2 is higher, **we are Slave**! – Unset **Master-bit** from **DD**

Tx: XR2-DBD
Hi! I am 2.2.2.2 and my initial DD sequence is **0x22** – I am **Master**!
XR2 **More-bit** set in **DD**

Master/Slave election in the ExStart state begins, since **XR2** has the **highest router-ID**, it wins the **Master/Slave** process.

The initial **DD** sequence to be used is the one sent by the Master/XR2.

Point-to-Point Adjacency (6) – Exchange



R1
Here is the summary list of my **LSDB**: [Router-LSA, Network-LSA, NetSummary LSA, ...]

R1 echo Master’s **DD** initial **0x22**

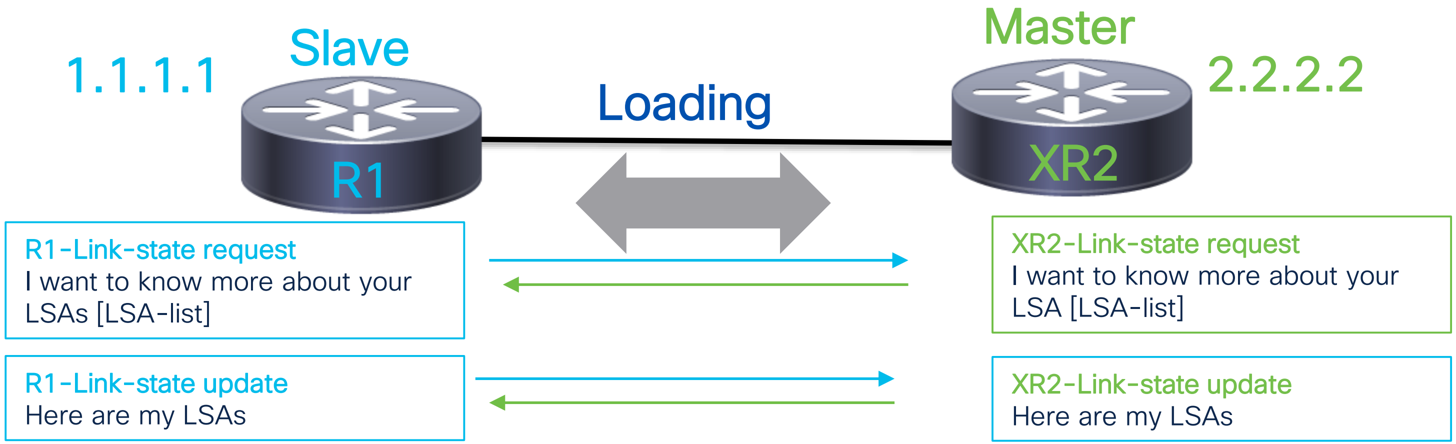
R2
Here is the list of my **LSDB**: [Router-LSA, Network-LSA, NetSummary LSA, ...]

XR2 More-bit unset in **DD** – **No further DD packets**

The slave sends summarized list of **DBD** packets containing the link-state advertisements headers. This will be used to further request the most recent LSAs.

No explicit LS Ack, Master/Slave echo the DD sequence as ACK

Point-to-Point Adjacency (7) – Loading



1.1.1.1

Slave

R1

Loading

Master

2.2.2.2

XR2

R1-Link-state request
I want to know more about your LSAs [LSA-list]

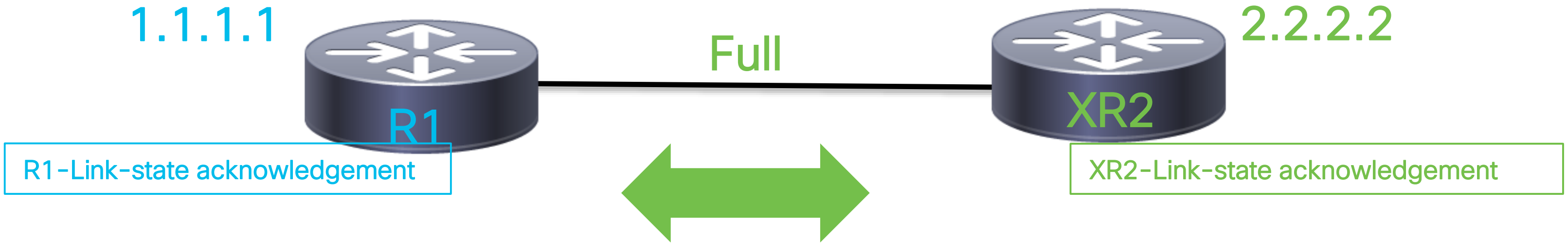
XR2-Link-state request
I want to know more about your LSA [LSA-list]

R1-Link-state update
Here are my LSAs

XR2-Link-state update
Here are my LSAs

In the **Loading** state the routers will perform LS Requests, LS Updates, and LS Acknowledgements to reliably exchange the link-state advertisements.

Point-to-Point Adjacency (8) – Full

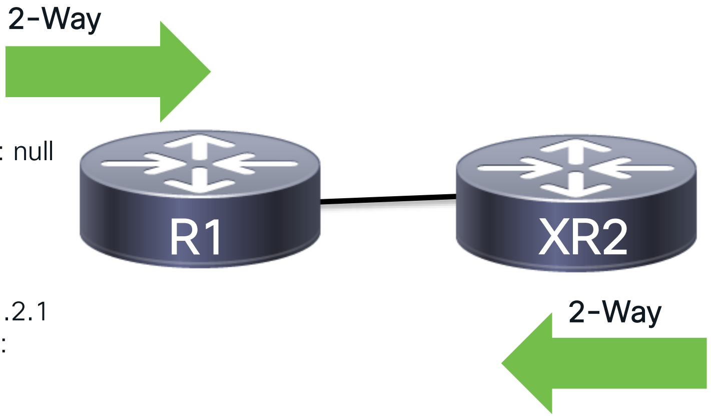


```
OSPF-1 ADJ   Gi0/0: Interface going Up
OSPF-1 ADJ   Gi0/0: Interface state change to UP, new ospf state P2P
OSPF-1 ADJ   Gi0/0: 2 Way Communication to 2.2.2.2, state 2WAY

OSPF-1 ADJ   Gi0/0: Nbr 2.2.2.2: Prepare dbase exchange
OSPF-1 ADJ   Gi0/0: Send DBD to 2.2.2.2 seq 0x1B76 opt 0x52 flag 0x7 len 32
OSPF-1 ADJ   Gi0/0: Rcv DBD from 2.2.2.2 seq 0x7B3D opt 0x52 flag 0x7 len 32  mtu 1500 state EXSTART

OSPF-1 ADJ   Gi0/0: NBR Negotiation Done. We are the SLAVE
OSPF-1 ADJ   Gi0/0: Nbr 2.2.2.2: Summary list built, size 7
OSPF-1 ADJ   Gi0/0: Send DBD to 2.2.2.2 seq 0x7B3D opt 0x52 flag 0x2 len 172
OSPF-1 ADJ   Gi0/0: Rcv DBD from 2.2.2.2 seq 0x7B3E opt 0x52 flag 0x1 len 92  mtu 1500 state EXCHANGE

OSPF-1 ADJ   Gi0/0: Exchange Done with 2.2.2.2
OSPF-1 ADJ   Gi0/0: Synchronized with 2.2.2.2, state FULL

%OSPF-5-ADJCHG: Process 1, Nbr 2.2.2.2 on GigabitEthernet0/0 from LOADING to FULL, Loading Done
```

Neighbor Relationships vs. Adjacencies

- **Non-DR/BDR** routers (**DROHTERS**) stay in **2-Way** state between them, this is known as a neighbor relationship as DROTHERs cannot exchange the link-state database directly between them.
- All routers become adjacent (**FULL**) with DR and BDR

Hello!
I am 1.1.1.1
I am in area 809
I have authentication type: null
My network mask is 255.255.255.0
Hello interval is: 10s
Dead interval is: 40s
Designated router is: 10.1.2.1
Backup designated router: 10.1.2.2



Hello!
I am 2.2.2.2
I am in area 809
I have authentication type: null
My network mask is 255.255.255.0
Hello interval is: 10s
Dead interval is: 40s
Designated router is: 10.1.2.1
Backup designated router: 10.1.2.2

Adjacency over Broadcast networks

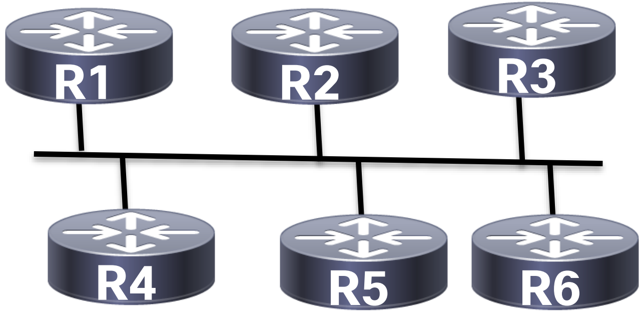
- A **DR** should be elected to fully exchange the link-state database within the multi-access network, **election is performed in a per-link-basis**
- The DR has two (2) main functions:

  1) DR originates the **Network-LSA** listing all routers attached in the segment including the **DR** itself

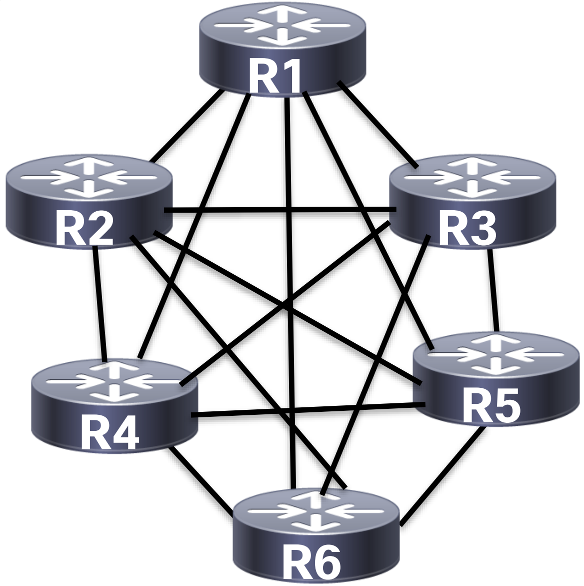
  2) The **DR** is the only router that can become fully adjacent with all routers in the segment, making the DR the central point of reference for **LSDB** synchronization

Designated Router (DR) (1)

- Without the **DR/Pseudonode**, the graph within the multi-access segment is more complex



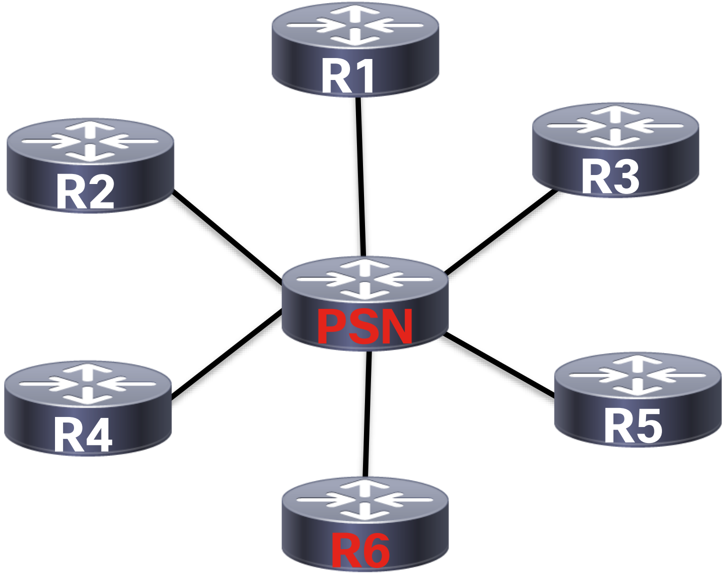
Multiaccess segment



No DR

Designated Router (DR) (2)

- With the Designated Router (**DR**) the graph is simplified to a collection of point-to-point links towards the **DR/Pseudonode**



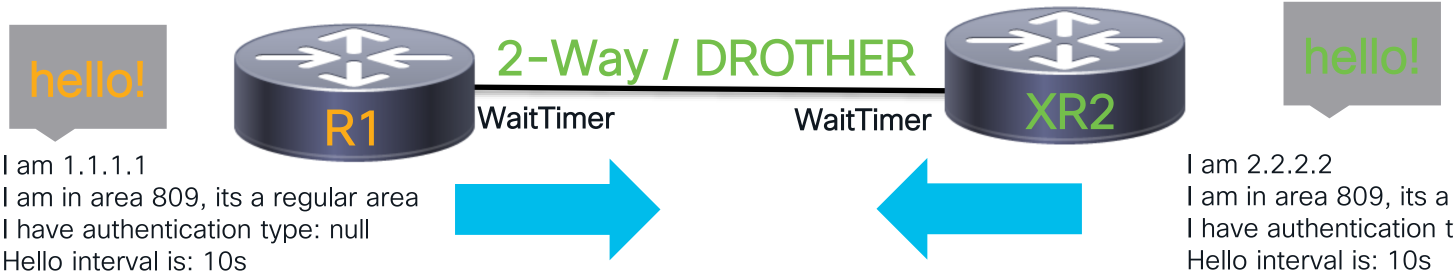
Designated Router (DR) (3)

- Designated Router (**DR**) election is **non-preemptive**, once DR is selected it cannot be overthrown (hmm... really?)
- During initialization, the router waits for the **WaitTimer** set in the interface (defaults to the configured HoldTime)
- The criteria of **DR** selection is:
  - a. Highest interface priority (default 1, range 0 - 65535)
  - b. Highest router-ID

- Note: Priority zero (0) has special meaning of non-eligibility, therefore it sets the router to always be a **DROHTER**.

Designated Router (DR) Election (1)

- When the interface is OSPF-enabled, the router waits for a period known as the **WaitTimer** to validate the existence of a **DR** in the segment

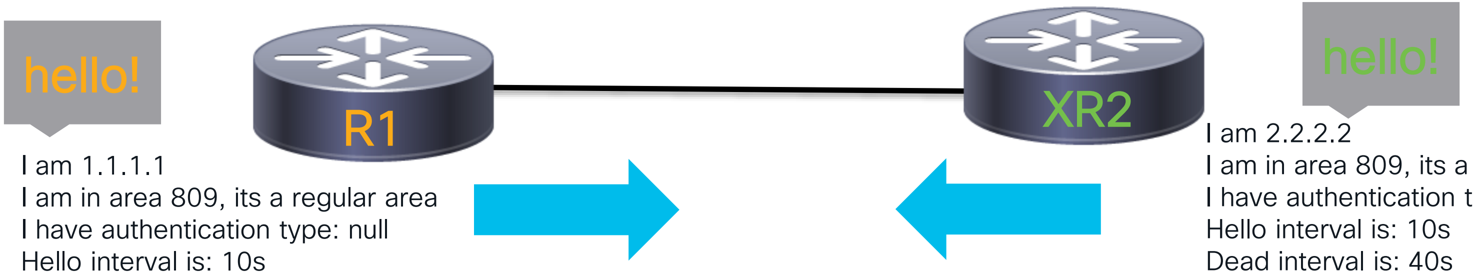


I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 2.2.2.2
DR: 0.0.0.0
BDR: 0.0.0.0

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 1.1.1.1
DR: 0.0.0.0
BDR: 0.0.0.0

Designated Router (DR) Election (2)

- Begin the election process by filling out the **DR/BDR** fields in the **Hello** packet based on criteria (**priority, highest RID**)
- If there is no **DR** in the **segment**, the router elects itself as the **BDR** to promote itself as the **DR** - **Yes, algorithmically BDR is elected first!** ☺

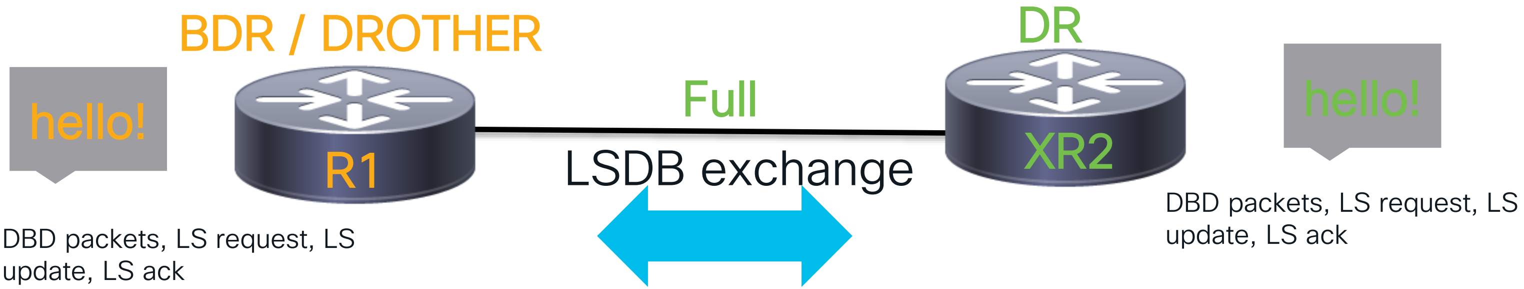


I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 2.2.2.2
DR: 2.2.2.2
BDR: 1.1.1.1

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 1.1.1.1
DR: 2.2.2.2
BDR: 1.1.1.1

Designated Router (DR) Election (3)

- After the DR is selected, the database descriptor (DD) and LSDB synchronization will happen as usual, during which all routers will form adjacencies with the DR/BDR and exchange their LSDB contents using the DR.

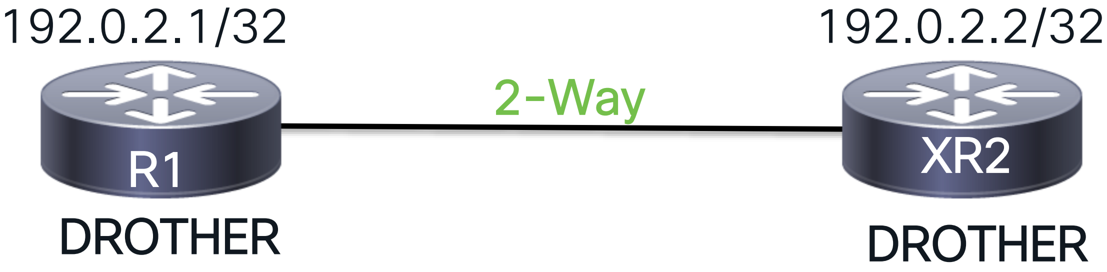


DBD packets, LS request, LS update, LS ack

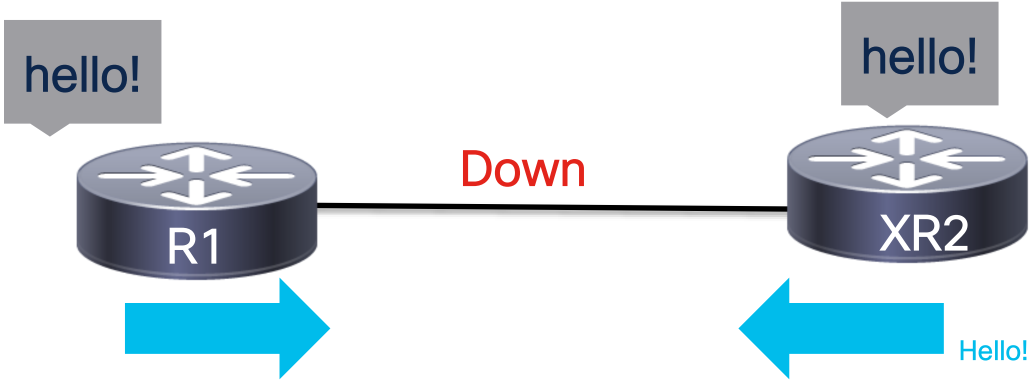
DBD packets, LS request, LS update, LS ack

Designated Router (DR) (4)

- What happens if routers are DROTHER?



Broadcast Adjacency (1) - Down



Hello!

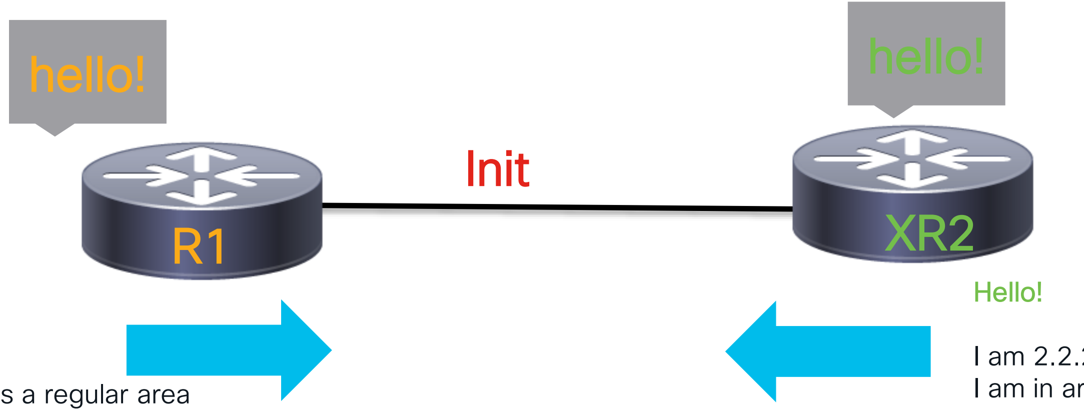
I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 0.0.0.0
DR: 0.0.0.0
BDR: 0.0.0.0

Hello!

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 0.0.0.0
DR: 0.0.0.0
BDR: 0.0.0.0

Initially routers in Broadcast networks are in Down state

Broadcast Adjacency (3) - Init



Hello!

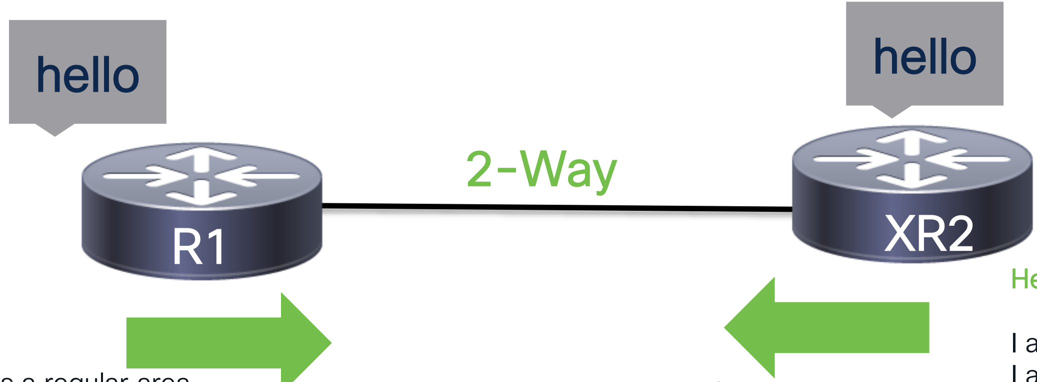
I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 0.0.0.0
DR: 0.0.0.0
BDR: 0.0.0.0

Hello!

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 0.0.0.0
DR: 0.0.0.0
BDR: 0.0.0.0

As soon as hellos are received from a neighbor, these are processed and routers will put the adjacency state as Init

Broadcast Adjacency (4) – 2-Way (1)



Hello!

I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 2.2.2.2
DR: 0.0.0.0
BDR: 0.0.0.0

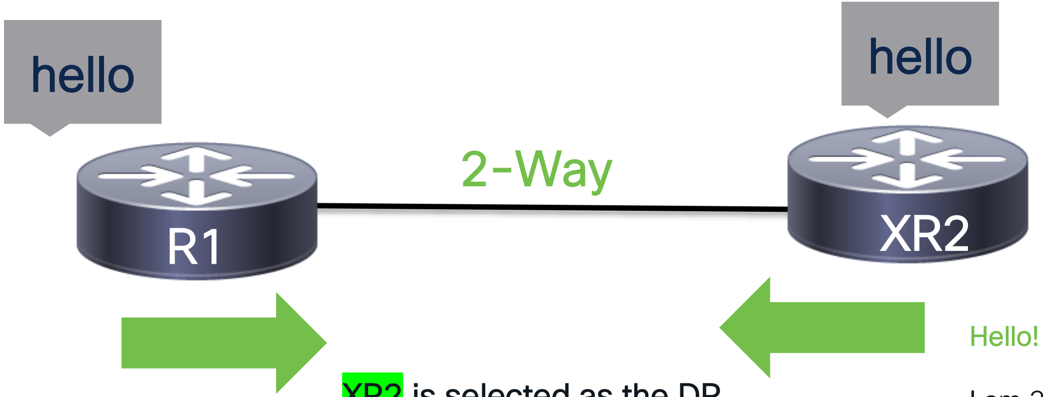
The DR election begins! ☺

Hello!

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 1.1.1.1
DR: 0.0.0.0
BDR: 0.0.0.0

Routers will acknowledge each other’s presence and move to 2-way, additionally the **DR election procedure begins here!**

Broadcast Adjacency (4) – 2-Way (2)



XR2 is selected as the DR

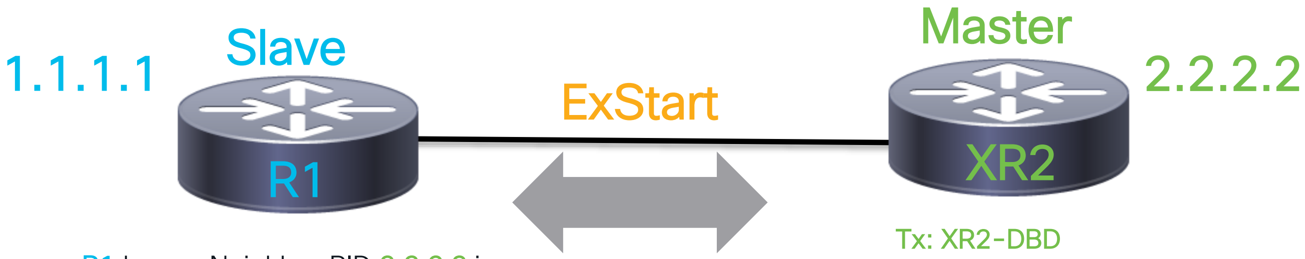
Hello!

I am 1.1.1.1
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 2.2.2.2
DR: 2.2.2.2
BDR: 1.1.1.1

Hello!

I am 2.2.2.2
I am in area 809, its a regular area
I have authentication type: null
Hello interval is: 10s
Dead interval is: 40s
Priority: 1
Active Neighbor: 1.1.1.1
DR: 2.2.2.2
BDR: 1.1.1.1

Broadcast Adjacency (5) – ExStart



R1 knows Neighbor RID 2.2.2.2 is higher, **we are Slave**! – Unset **Master-bit** from **DD**

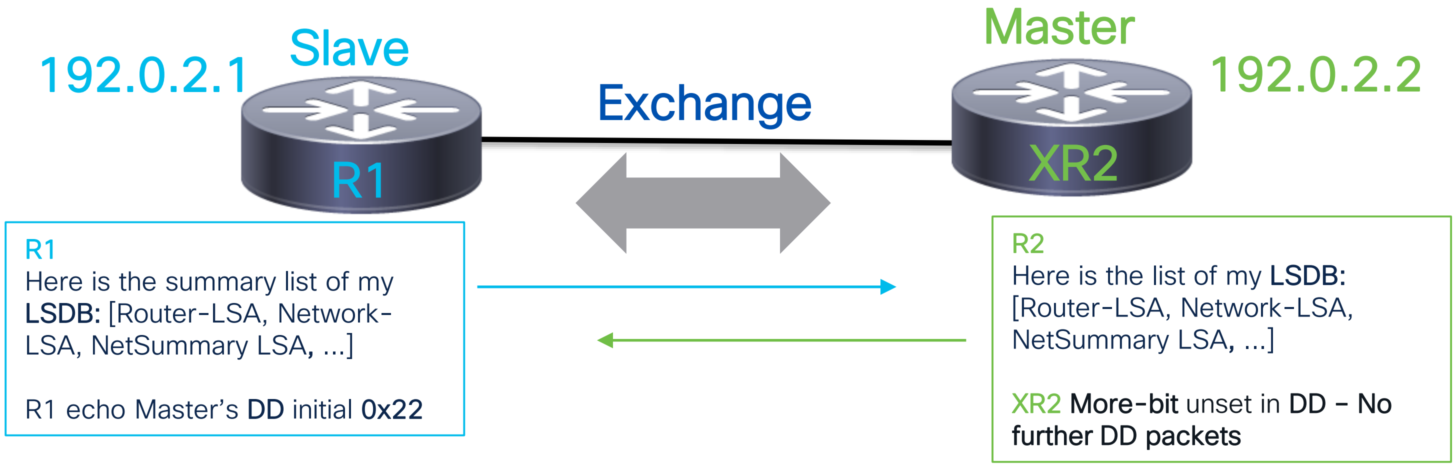
Tx: XR2-DBD
Hi! I am 2.2.2.2 and my initial DD sequence is **0x22** – I am **Master**!

XR2 **More-bit** set in **DD**

Master/Slave election in the ExStart state begins,
since **XR2** has the **highest router-ID**, it wins the **Master/Slave** process.

The initial **DD** sequence to be used is the one sent by the Master/XR2.

Broadcast Adjacency (6) – Exchange



R1
Here is the summary list of my **LSDB:** [Router-LSA, Network-LSA, NetSummary LSA**,** ...]

R1 echo Master’s **DD** initial **0x22**

R2
Here is the list of my **LSDB:** [Router-LSA, Network-LSA, NetSummary LSA**,** ...]

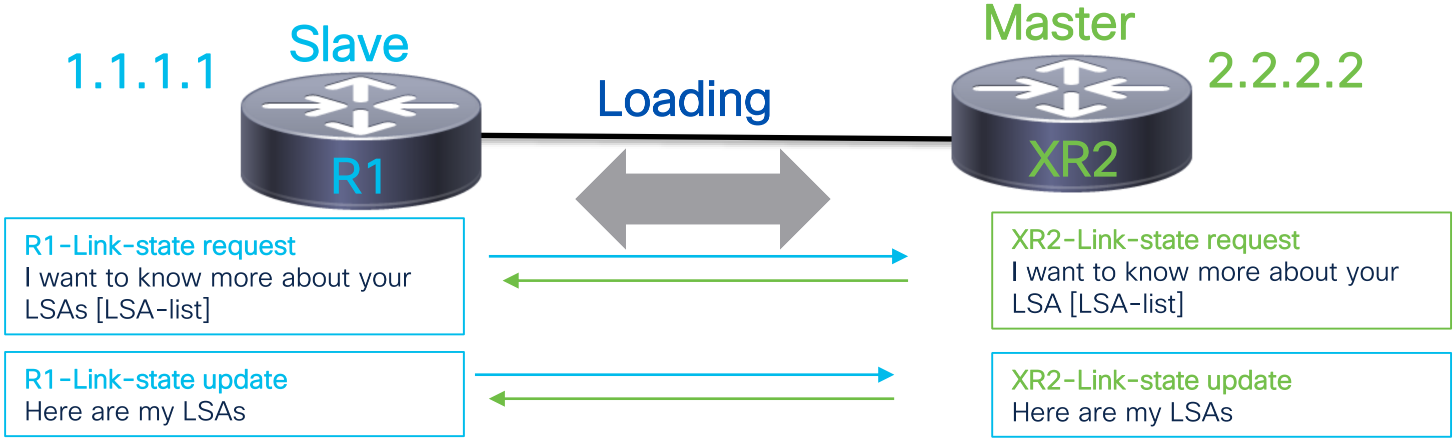
XR2 **More-bit** unset in **DD – No further DD packets**

The slave sends summarized list of **DBD** packets containing the link-state advertisements headers. This will be used to further request the most recent LSAs.

Note: **DD** with packets are not explicitly acknowledged using link-state acknowledgement packets, rather, they use an “echo” mechanism starting from the initial DD sequence during the exchange phase. As the Master and Slave exchange DD packets, the routers will send each other’s sequence back as acknowledgement.

Note: DR/BDR and Master/Slaves functions are decoupled.

Broadcast Adjacency (7) – Loading



1.1.1.1

Slave

R1

Loading

Master

2.2.2.2

XR2

R1-Link-state request
I want to know more about your LSAs [LSA-list]

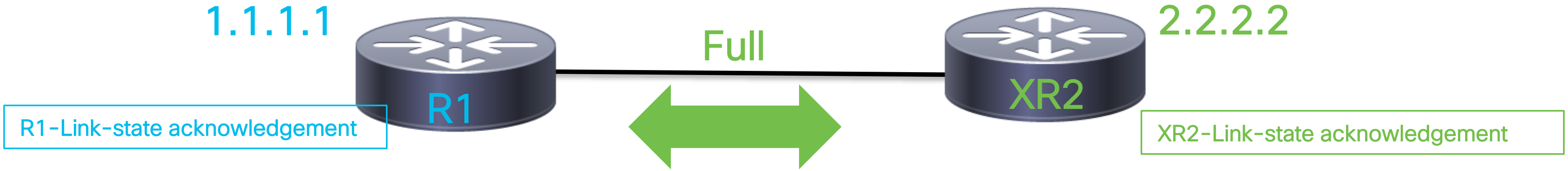
XR2-Link-state request
I want to know more about your LSA [LSA-list]

R1-Link-state update
Here are my LSAs

XR2-Link-state update
Here are my LSAs

In the **Loading** state the routers will perform **LS Requests**, **LS Updates**, and **LS Ack** to reliably exchange the **LSAs**

Broadcast Adjacency (8) – Full

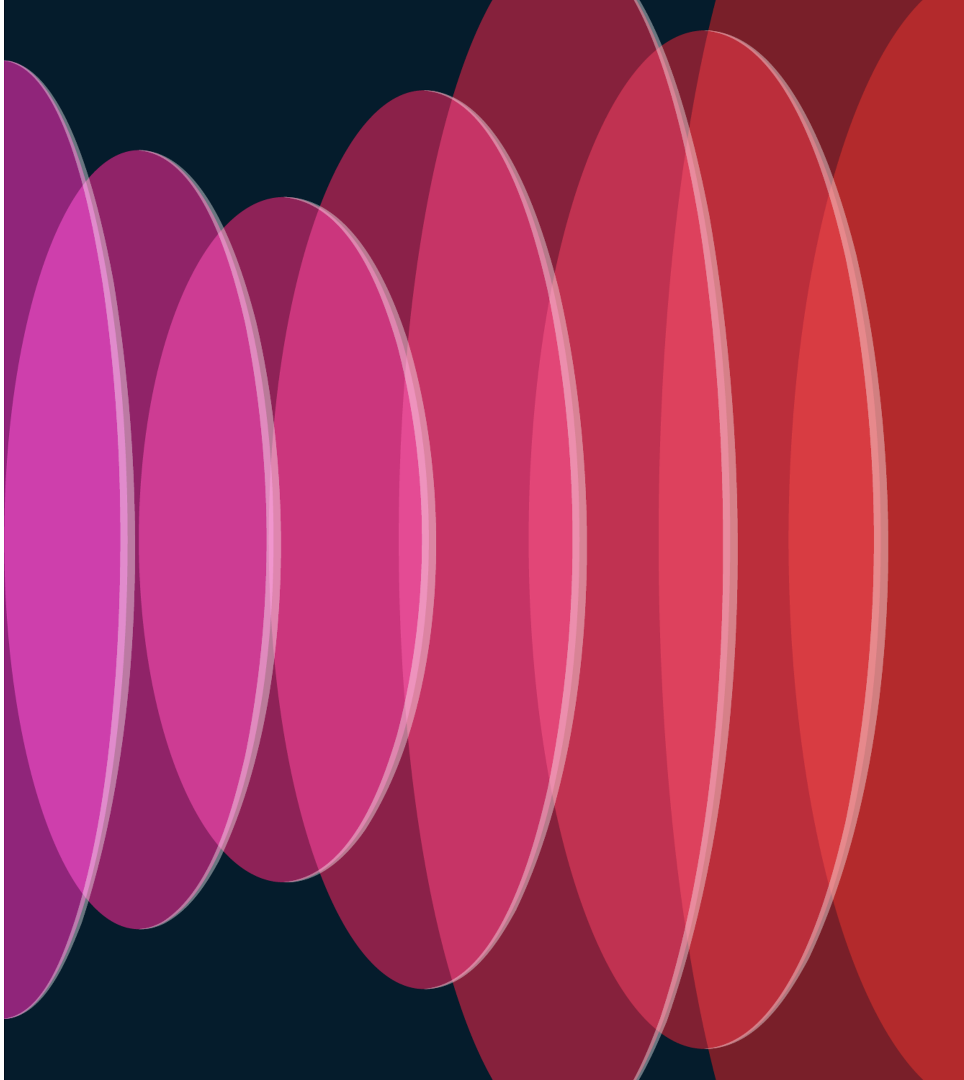


R1-Link-state acknowledgement

XR2-Link-state acknowledgement

```
003252: OSPF-1 ADJ   Gi0/0: Interface state change to UP, new ospf state WAIT
003253: OSPF-1 ADJ   Gi0/0: 2 Way Communication to 2.2.2.2, state 2WAY
...
003257: OSPF-1 ADJ   Gi0/0: Nbr state is 2WAY
003258: OSPF-1 ADJ   Gi0/0: end of Wait on interface
...
003259: OSPF-1 ADJ   Gi0/0: DR/BDR election
003260: OSPF-1 ADJ   Gi0/0: Elect BDR 2.2.2.2
003261: OSPF-1 ADJ   Gi0/0: Elect DR 2.2.2.2
003280: OSPF-1 ADJ   Gi0/0: Rcv DBD from 2.2.2.2 seq 0x5A2A opt 0x52 flag 0x7 len 32  mtu 1500 state EXSTART
003281: OSPF-1 ADJ   Gi0/0: NBR Negotiation Done. We are the SLAVE
003282: OSPF-1 ADJ   Gi0/0: Nbr 2.2.2.2: Summary list built, size 4
003283: OSPF-1 ADJ   Gi0/0: Send DBD to 2.2.2.2 seq 0x5A2A opt 0x52 flag 0x2 len 112
003284: OSPF-1 ADJ   Gi0/0: Rcv DBD from 2.2.2.2 seq 0x5A2B opt 0x52 flag 0x1 len 92  mtu 1500 state EXCHANGE
003285: OSPF-1 ADJ   Gi0/0: Exchange Done with 2.2.2.2
003286: OSPF-1 ADJ   Gi0/0: Send LS REQ to 2.2.2.2 length 60
003287: OSPF-1 ADJ   Gi0/0: Send DBD to 2.2.2.2 seq 0x5A2B opt 0x52 flag 0x0 len 32
003289: OSPF-1 ADJ   Gi0/0: Synchronized with 2.2.2.2, state FULL
```

Verifying Adjacencies



LSDB Synchronization

Flooding Pre-Checks

- Validate LSA checksum
- Check if LSA type is valid
- Check if External-LSA are received over Stub areas
- LSA received with **MaxAge** set to maximum, then discard it

Flooding Events

- Event changes that cause flooding of new information is OSPF include:
  - Adjacency state
  - Router ID
  - Area ID
  - DR re-election
  - Transit metric cost

  - **Note**: If changes are triggered, affected LSA must be reflooded

Link State Database

- Link-State Database (LSDB) contents draw a detailed map of the network topology within a particular scope
- OSPF maintains independent LSDBs for each level
- LSDB stores all Link State Advertisements (LSAs) of a particular area

Link State Database Synchronization

- All routers operating at the same scope (in the same area) must have identical LSDB contents
  - LSDB contents must be always synchronized between routers
- Synchronizing LSDB contents requires
  - Exchanging LSAs during initial synchronization when a new adjacency comes up, and anytime an LSA is updated
  - Acknowledging exchanged LSAs using **LS Ack** packets
  - On broadcast network types, using DR as a synchronization reference using 224.0.0.6/AllDRRouters

LSDB Synchronization on point-to-point links (1)

- When a new adjacency comes up between two routers on a point-to-point link, they synchronize their LSDBs in a simple way
  - Each router schedules database descriptors (**DD**) packets to be sent to the neighbor and elect the **Master/Slave** relationship, the highest RID wins the Master election
  - Master will send the initial **DD** sequence with the **Initial** and **More** bits set in the **DD** packet indicating that more packets are to follow
  - As DD packets are exchanged containing the aggregate view of the participating router's LSDBs, LS Request packets are sent to

LSDB Synchronization on point-to-point links (2)

Cont.

- As DD packets are exchange containing the aggregate view of the participating router's LSDBs, **LS Request** packets are sent if the received LSAs are:
  - New(er): Store it and schedule it for acknowledgment in a **LS Ack**
  - Identical: Schedule an acknowledgment in a **LS Ack**
  - Older: Schedule our own LSA to be flooded to the neighbor
- LSA stays scheduled for sending to the neighbor only if it is newer

LSDB Synchronization on broadcast networks (1)

- On broadcast networks, pairwise synchronization of a new router with every existing neighbor would be both complex and useless
- Instead, DR becomes the reference point for database synchronization among all routers on the network
  - Relying on transitivity: *If I know the same as DR, and if you know the same as DR, then I and you know the same, too*
  - Every router’s goal: Make the DR LSDB and own LSDB identical
- As opposed to IS-IS, all OSPF routers on a broadcast network are not fully adjacent and will only accept LS Updates from the DR directly – DR is the relay for LS Updates

LSDB Synchronization on broadcast networks (1)

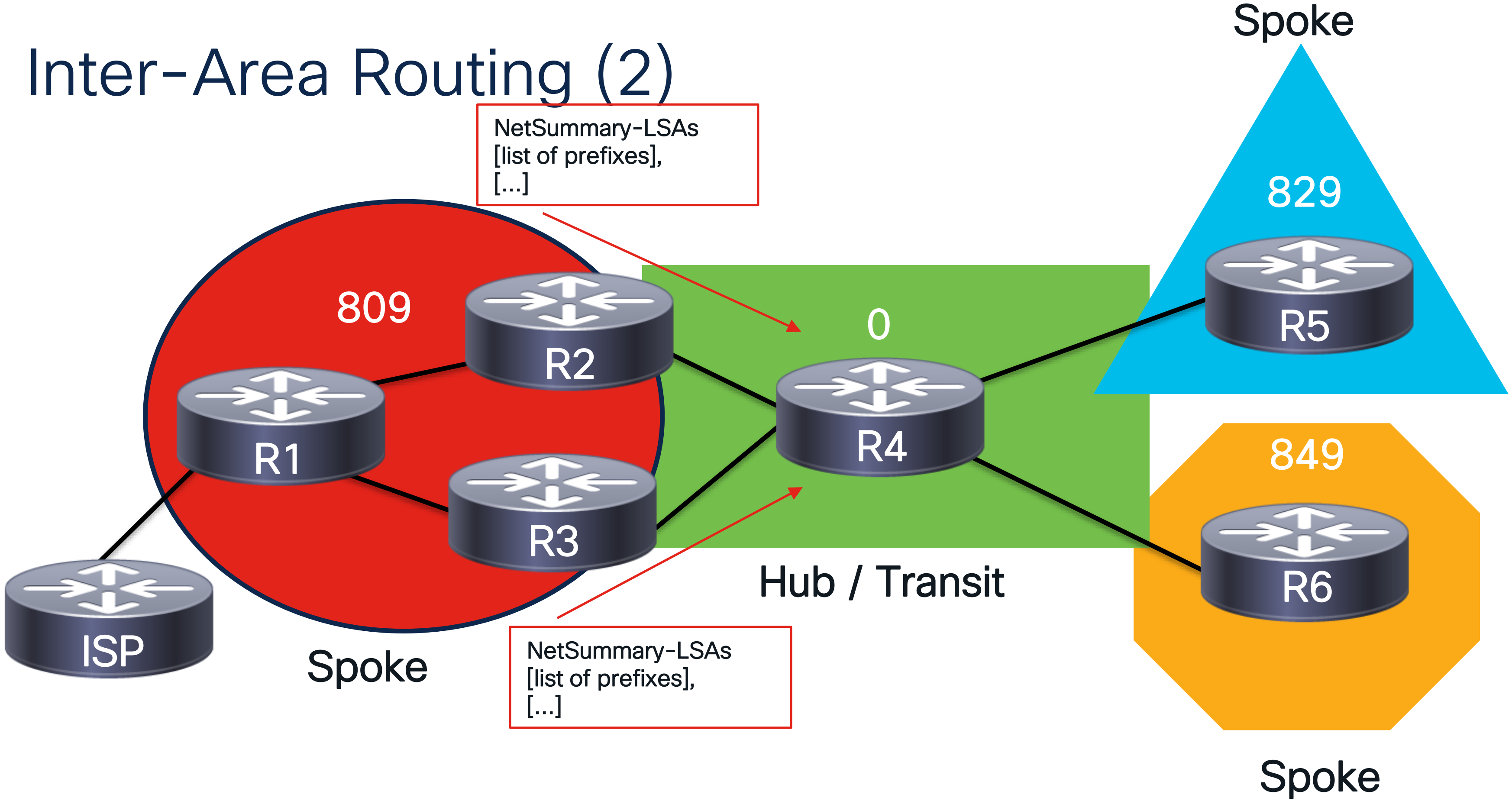
- Each router on the broadcast network compares uses the DR to synchronize, If the router knows about a(n)…
  - Newer LSA: Just flood it onto the DR/BDR. Other routers will learn the newest information through the DR relay downstream to the adjacent routers
  - Identical LSA: Acknowledge the received LS Update and no further processing is performed
  - Older LSA: Ask for an updated LSA using a LS Request directed to the DR

Inter-Area Routing

Inter-Area Routing (1)

- The inter-area routing in **OSPF** works as a **distance vector** protocol, where all network complexity within an area is hidden from the backbone and other areas
- As advertisements pass through the area border routers (**ABRs**), **NetSummary-LSAs** are generated with the ABR as the attachment point to summarize the topological information – **Routing by rumor!**
- A start topology is enforced with the backbone area at the center (hub) that other areas (spokes) must transit

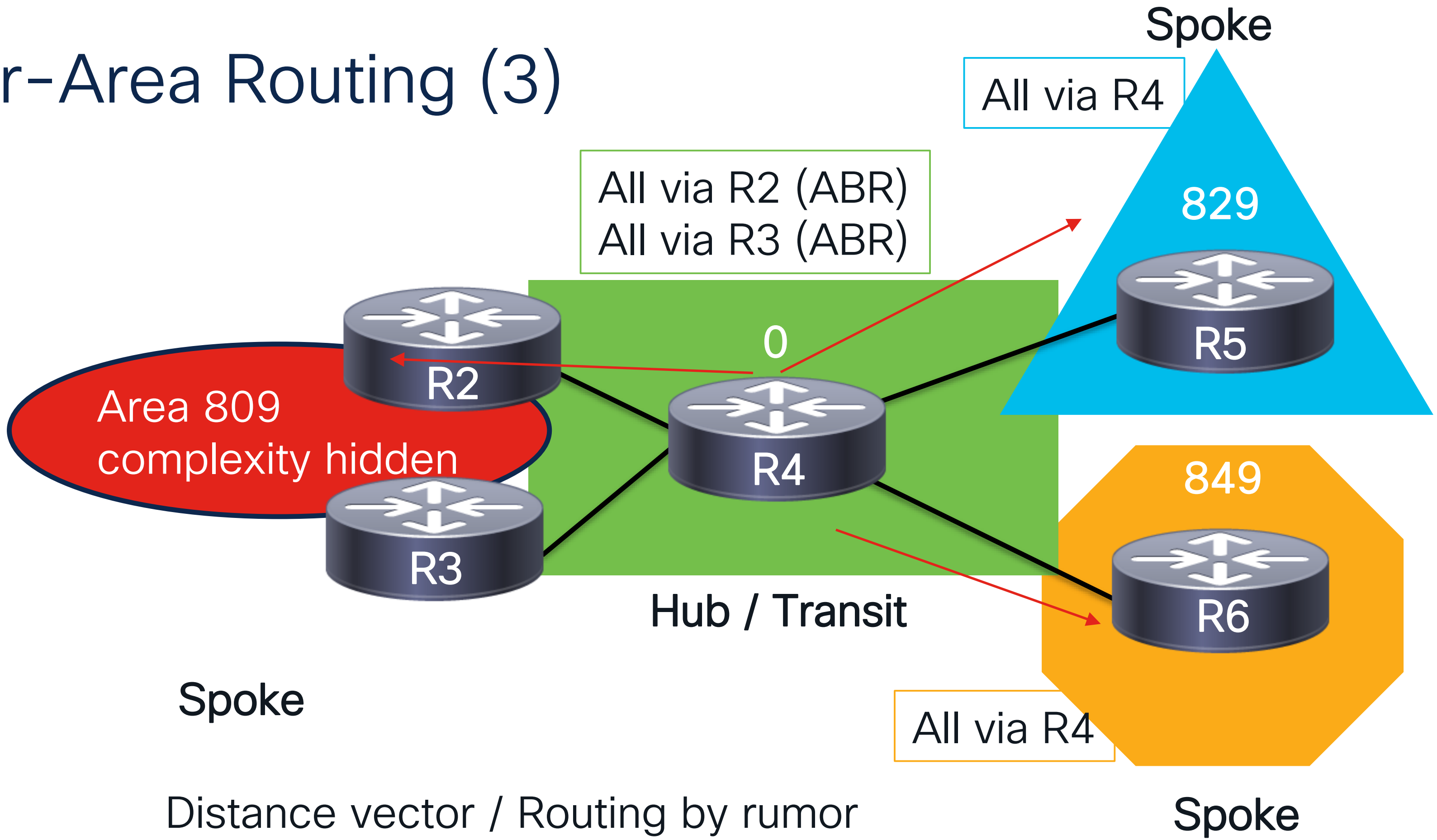
Inter-Area Routing (2)



NetSummary-LSAs
[list of prefixes],
[...]

NetSummary-LSAs
[list of prefixes],
[...]

Inter-Area Routing (3)

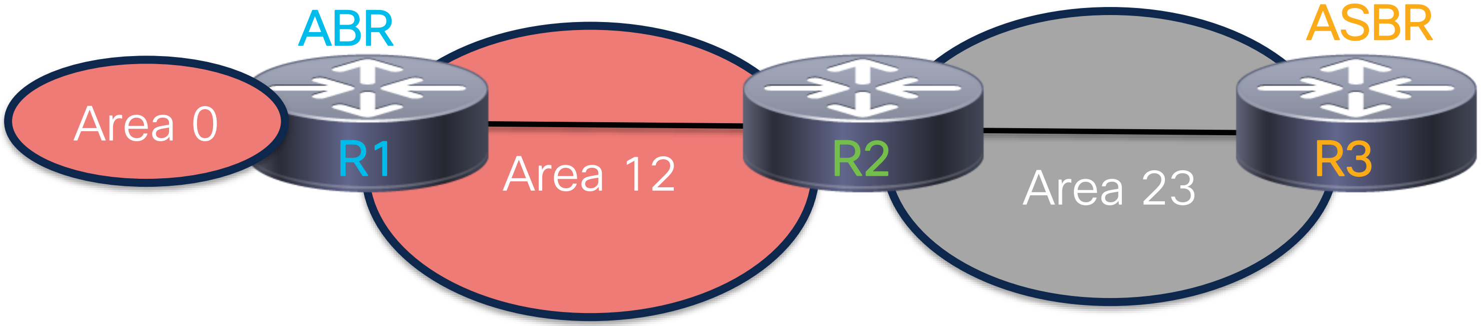


Distance vector / Routing by rumor

Inter-Area Loop Prevention (1)

- Area border routers (ABRs) are routers that have an interface attached to the backbone and is not in the DOWN state, only the ABR is allowed to generate **NetSummary-LSAs**
- ABR will never use **NetSummary-LSAs** coming from non-backbone areas
- As ABRs generate **NetSummary-LSAs**, they insert their router ID in the advertisement to prevent LSA feedback

Inter-Area Loop Prevention (2)



R1

```
interface Loopback0
 ip address 192.0.2.1 255.255.255.255
 ip ospf 1 area 0 << (ABR!)
!
interface GigabitEthernet0/0.12
 encapsulation dot1q 23
 ip address 10.1.2.1 255.255.255.0
 ip ospf network point-to-point
 ip ospf 1 area 12
```

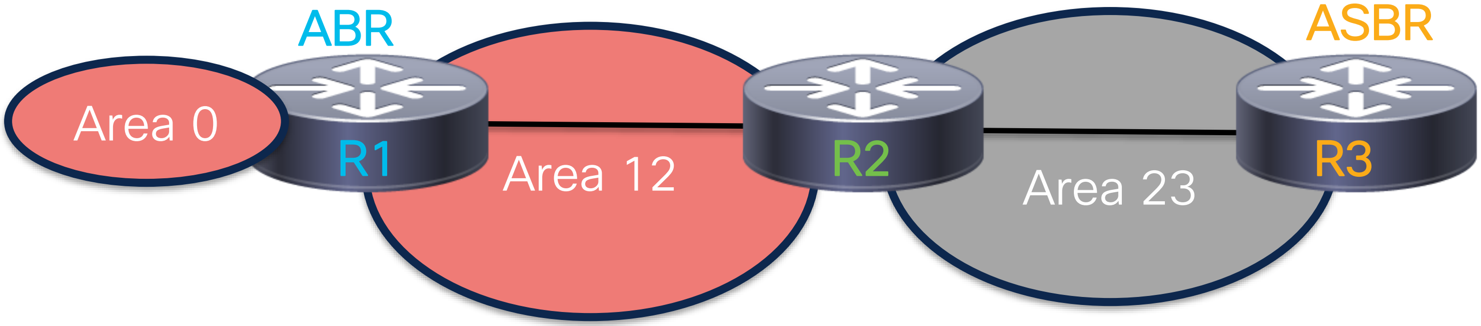
R2

```
interface Loopback0
 ip address 192.0.2.2 255.255.255.255
!
interface GigabitEthernet0/0.12
 encapsulation dot1q 12
 ip address 10.1.2.2 255.255.255.0
 ip ospf network point-to-point
 ip ospf 1 area 12
!
interface GigabitEthernet0/0.23
 encapsulation dot1q 23
 ip address 10.2.3.2 255.255.255.0
 ip ospf network point-to-point
 ip ospf 1 area 23
```

R3

```
interface Loopback0
 ip address 192.0.2.3 255.255.255.255
!
interface GigabitEthernet0/0.23
 encapsulation dot1q 23
 ip address 10.2.3.3 255.255.255.0
 ip ospf network point-to-point
 ip ospf 1 area 23
!
router ospf 1
 redistribute connected subnets
```

Inter-Area Loop Prevention (3)



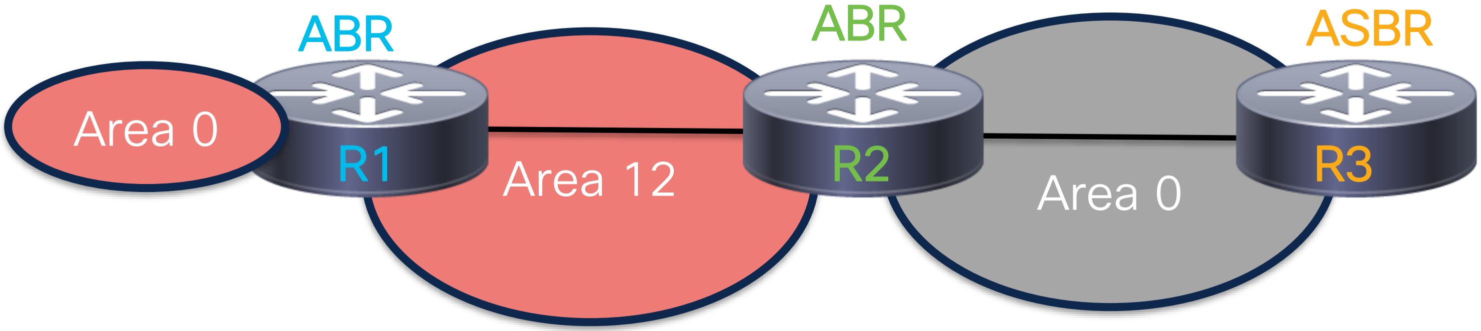
Inter-Area

1. Will R1 receive and install R1’s summary LSA for 192.0.2.1/32?
2. Will R3 receive and install R1’s prefix for 192.0.2.1/32?

External

1. Will R2 receive and install R3’s external prefix for 192.0.2.3/32?
2. Will R1 receive and install R3’s external prefix for 192.0.2.3/32?

Inter-Area Loop Prevention (4)



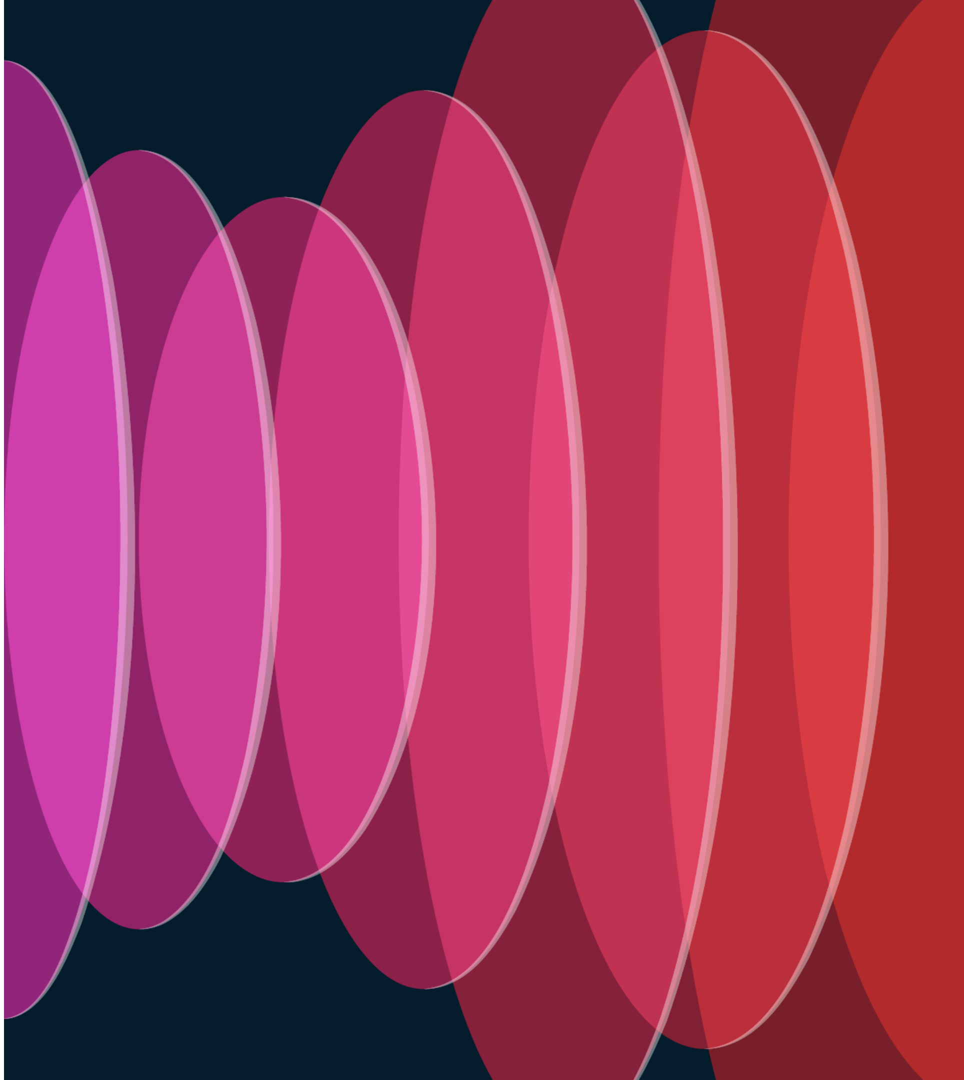
Inter-Area

1. Will R2 receive and install R1's summary LSA for 192.0.2.1/32?
2. Will R3 receive and install R1's prefix for 192.0.2.1/32?

External

1. Will R2 receive and install R3's external prefix for 192.0.2.3/32?
2. Will R1 receive and install R3's external prefix for 192.0.2.3/32?

Inter-Area Loop Prevention



Special Areas

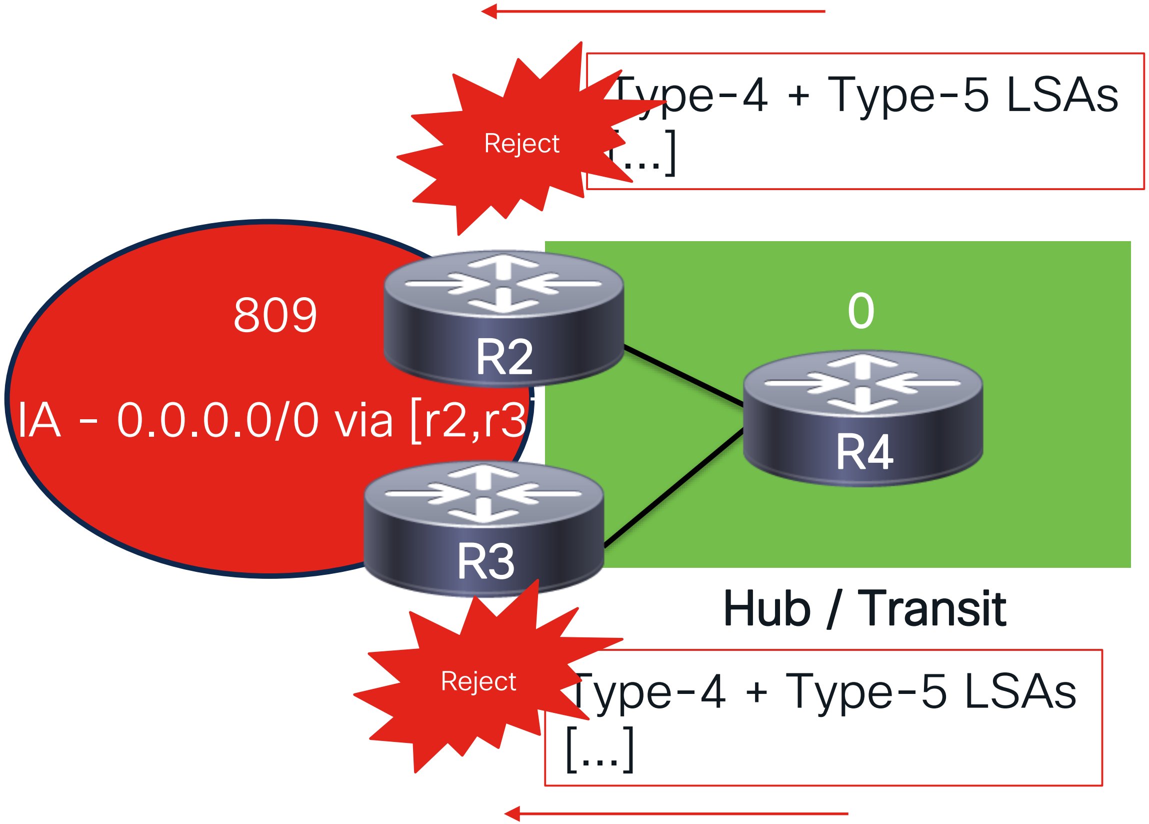
Special Area Types (1)

- Special areas have characteristics to allow/disallow certain link-state advertisements (LSA)
- Special Areas are:
  - Stub
  - Totally-Stubby
  - Not-So-Stubby Areas (NSSA)
  - Totally Not-So-Stubby Areas (Totally NSSAs)

Stub (1)

- Only NetSummary-LSAs are allowed through the Stub area
- Any external data structures (i.e., Type-4/5 LSAs) will be blocked
- Relies on default route generated from ABR for external routing

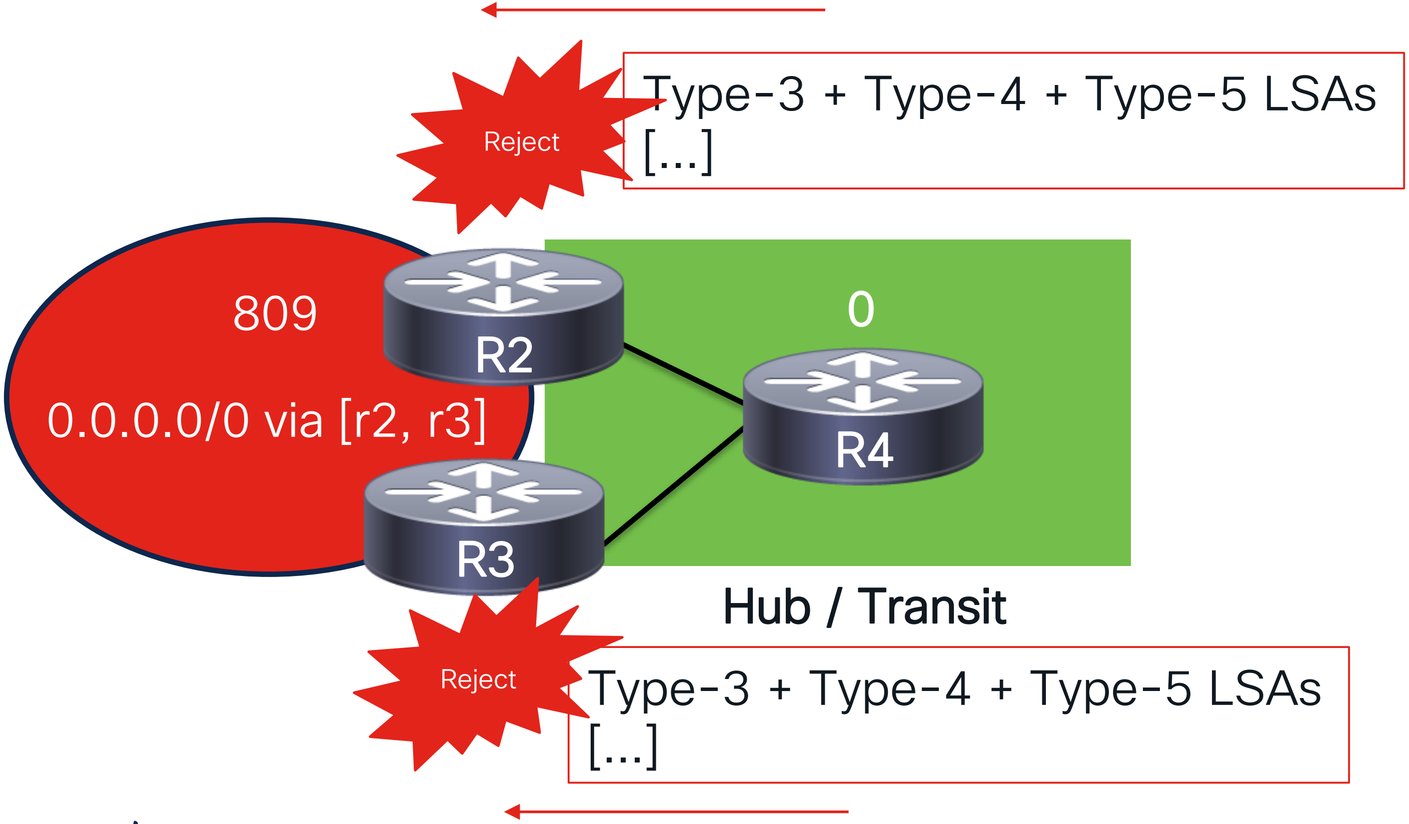
Stub (2)


Reject
Type-4 + Type-5 LSAs
[...]
809
IA – 0.0.0.0/0 via [r2,r3]
R2
0
R4
R3
Hub / Transit
Reject
Type-4 + Type-5 LSAs
[...]

Totally Stub (1)

- Any Inter-Area and External data structures (i.e., Type-3, Type-5 LSAs) and will be blocked
- Relies on default route generated from ABR to route towards inter-area and external sources

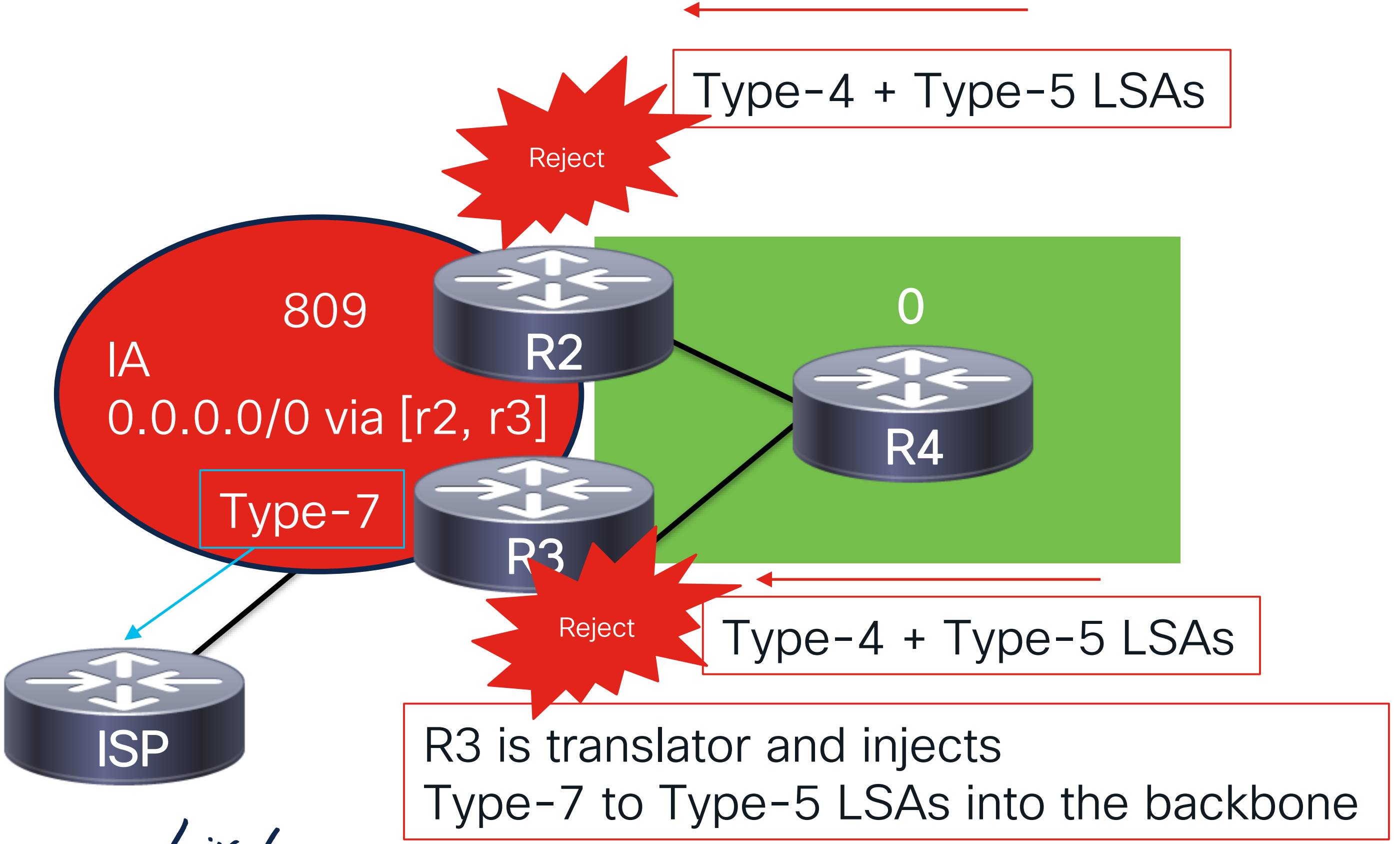
Totally Stub (2)



Not-So-Stubby-Area (NSSA) (1)

- Any External data structures (i.e., Type-5 LSAs) and will be blocked from coming from the backbone into the NSSA
- Relies on default route generated from ABR to route towards external sources outside the NSSA
- Allows for External routing using Type-7/NSSA-External-LSA
- ABR within the NSSA will perform Type-7 to Type-5 translation, highest RID wins the translator role

Not-So-Stubby-Area (NSSA) (2)



Type-4 + Type-5 LSAs

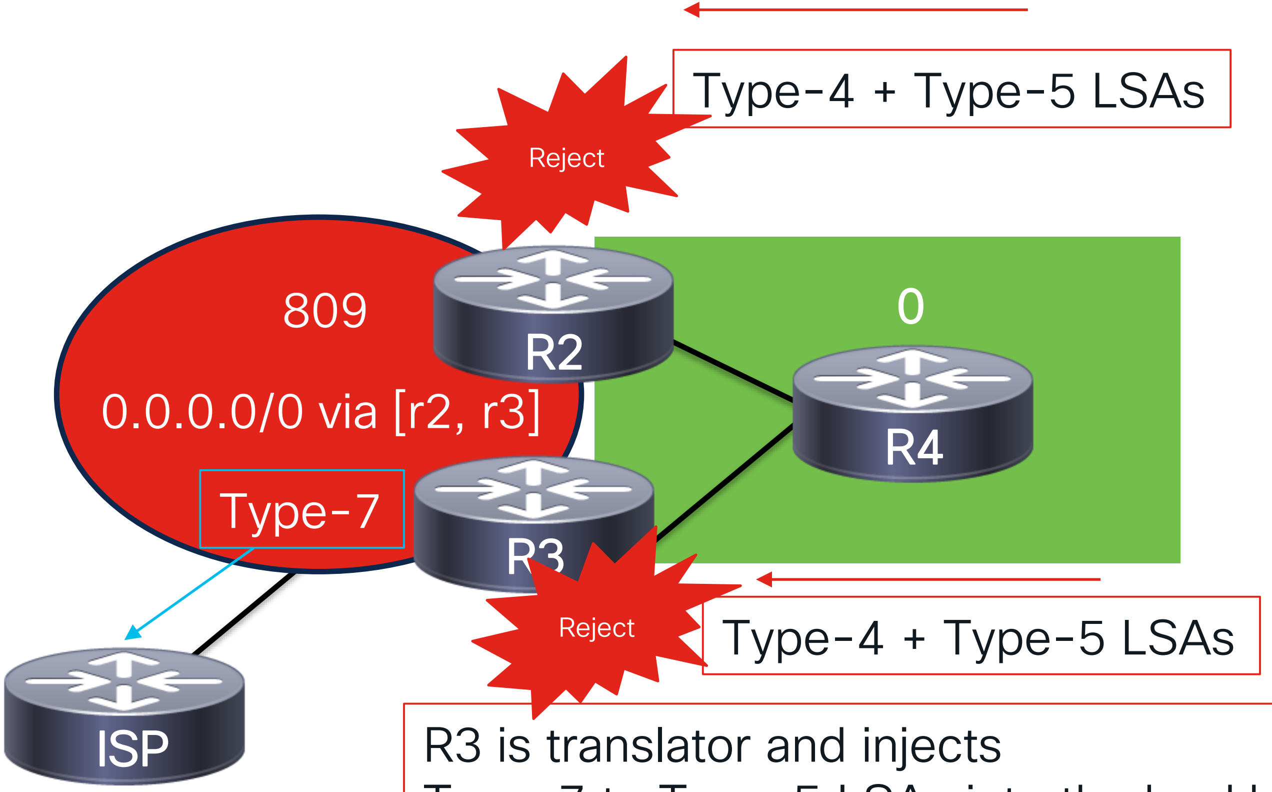
Type-4 + Type-5 LSAs

R3 is translator and injects
Type-7 to Type-5 LSAs into the backbone

Totally Not-So-Stubby-Area (Totally NSSA) (1)

- Any Inter-Area and External data structures (i.e., Type-3, Type-5 LSAs) and will be blocked from coming from the backbone into the NSSA
- Relies on default route generated from ABR to route towards inter-area and external sources outside the NSSA
- Allows for inter-area and external routing using Type-7/NSSA-External-LSA
- ABR within the NSSA will perform Type-7 to Type-5 translation, highest RID wins the translator role

Totally Not-So-Stubby-Area (Totally NSSA) (2)

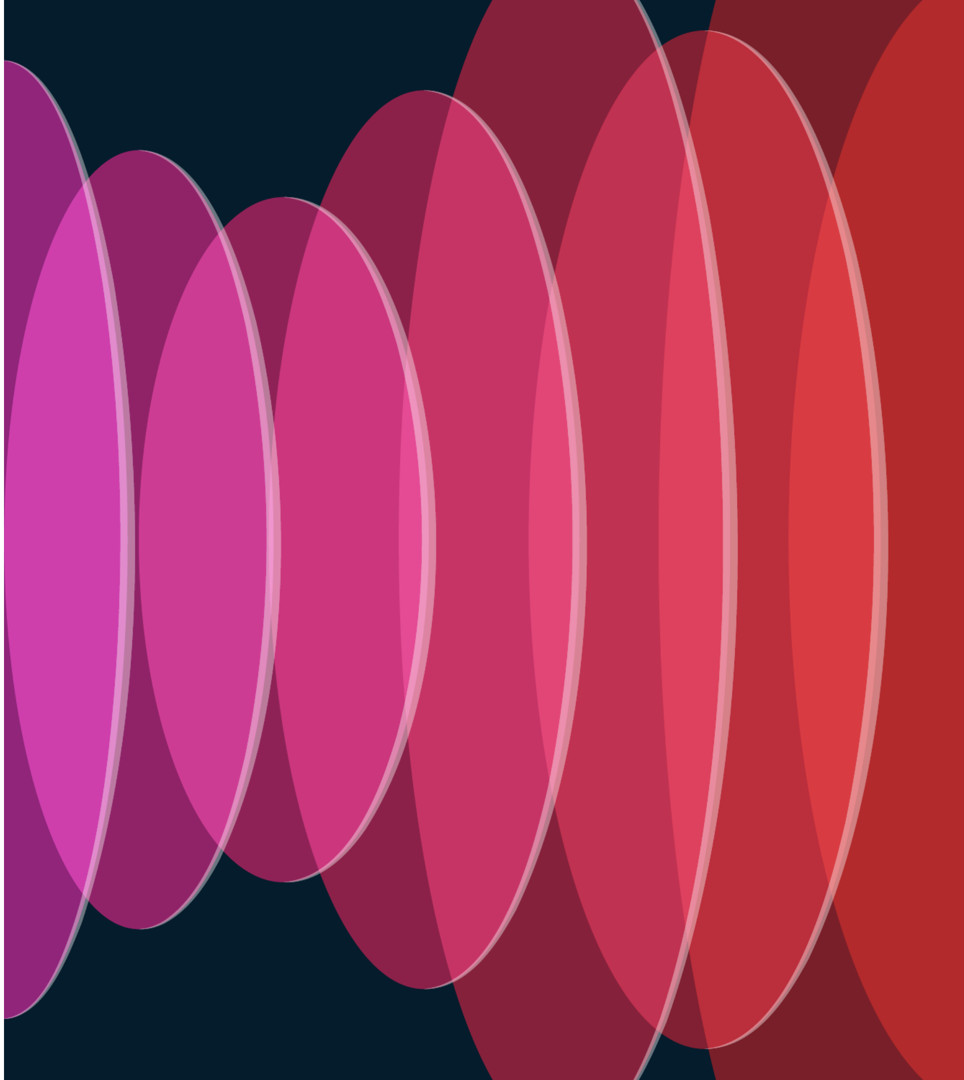


Type-4 + Type-5 LSAs

Type-4 + Type-5 LSAs

R3 is translator and injects
Type-7 to Type-5 LSAs into the backbone

Configuring Special Areas



Path Selection

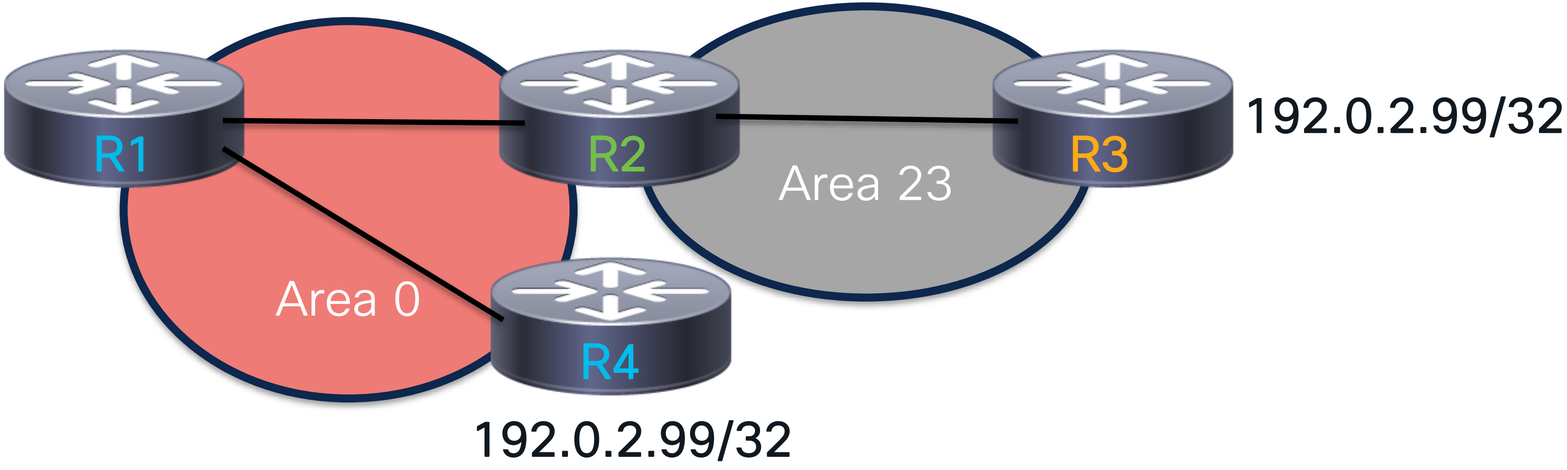
Path Selection (1)

- OSPF employs a **strict** path selection rule, where the order is applicable as follows:
  - Intra-Area (O)
  - Inter-Area (O IA)
  - External Type 1 (E1)
  - External Type 2 (E2)
  - NSSA Type 1 (N1)
  - NSSA Type 2 (N2)

  - Note: There are nuances! ☺

Scenarios: E1 vs. E2 Path Selection

- R1 and R2 are redistributing the same prefix of 192.0.2.99/32, R2, verify the different scenarios of External path selection ☺

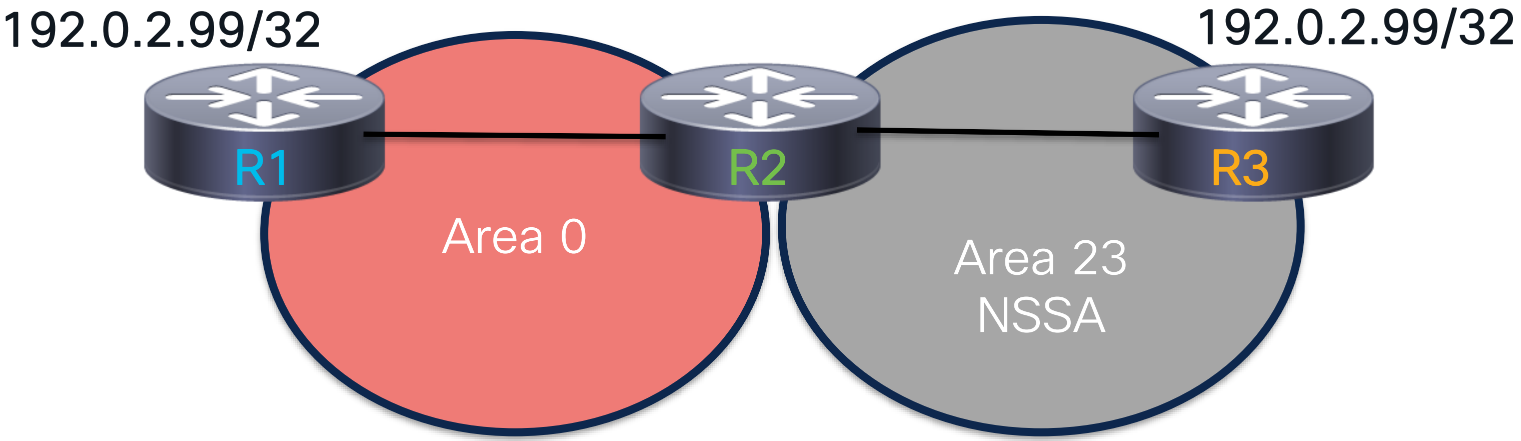


RFC 1587 vs. RFC 3101

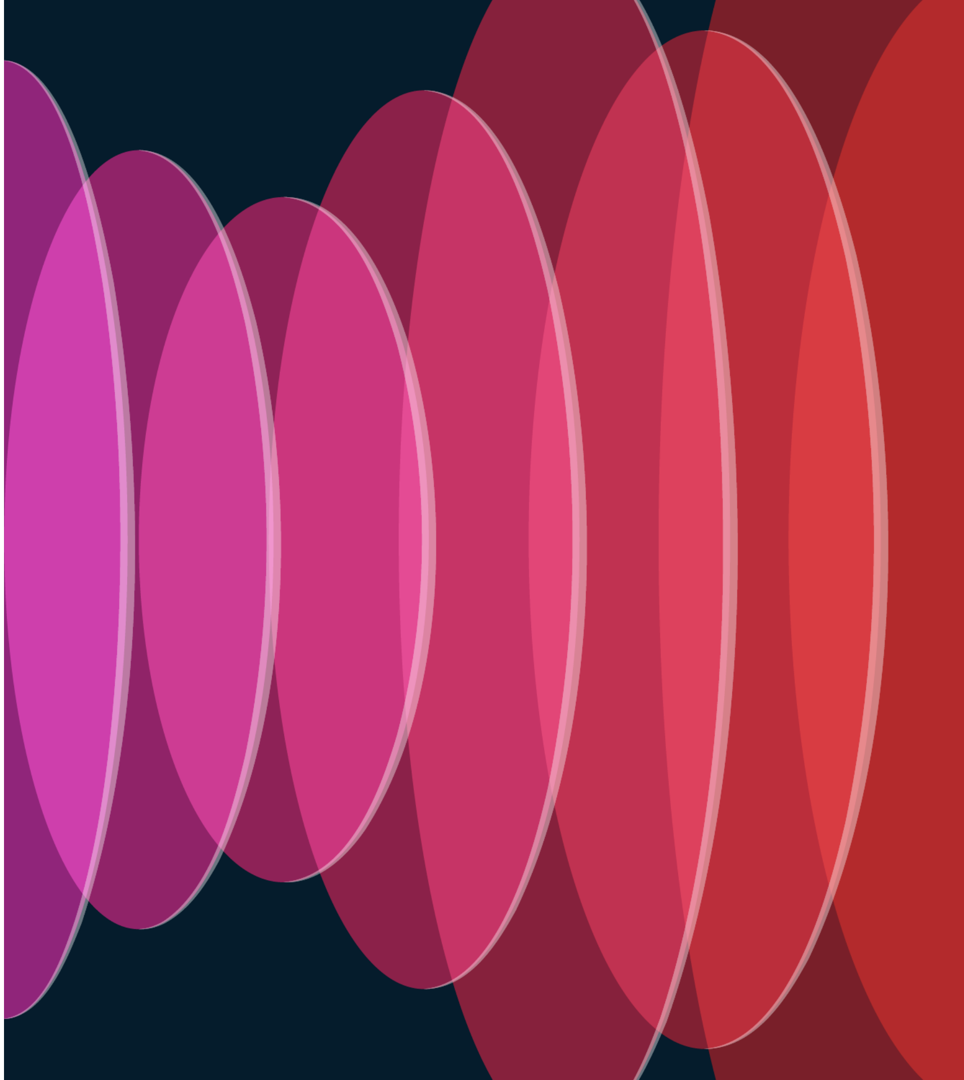
- By default RFC 3101 is enabled in IOS-XR and IOS-XE for path selection criteria (can be tweaked in the CLI)
- If the cost of the path is same, then the selection is as follows:
  - 1. A Type-7 LSA with the P-bit set
  - 2. A Type-5 LSA
  - 3. The LSA with the higher router ID

Scenarios: N1 vs. N2

- R1 and R2 are redistributing the same prefix of 192.0.2.99/32, R2, verify the different scenarios of NSSA path selection ☺



Path Selection



Security Hardening

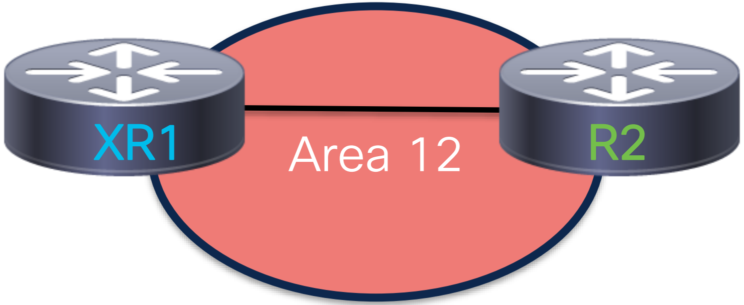
Authentication (1)

- Different authentication types exist:
  - Type 0: Null authentication (default)
  - Type 1: Simple-text authentication
  - Type 2: Cryptographic Authentication (RFC 5709)
    - SHA-1
    - SHA-256
    - SHA-384
    - SHA-512

Type 0: Null Authentication

- Default – no extra configuration is required

Type 1: Simple-text Authentication



XR1

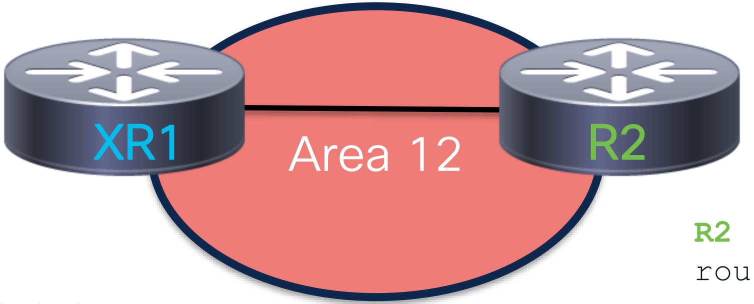
```
router ospf 1
 router-id 1.1.1.1
 area 0
  interface GigabitEthernet0/0/0/0
   authentication-key encrypted 060506324F41
   authentication
   network point-to-point
```

R2

```
router ospf 1
 router-id 2.2.2.2
!
interface GigabitEthernet0/0
 ip address 10.1.2.2 255.255.255.0
 ip ospf authentication
 ip ospf authentication-key cisco
 ip ospf network point-to-point
 ip ospf 1 area 0
 duplex auto
 speed auto
```

Note: Use '**service password-encryption**' command to encrypt the passwords in plain-text at config

Type 2: Cryptographic Authentication



XR1

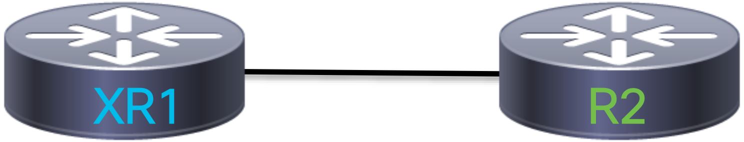
```
router ospf 1
 router-id 1.1.1.1
 area 0
  interface GigabitEthernet0/0/0/0
   authentication-key encrypted 060506324F41
   authentication
   network point-to-point
```

R2

```
router ospf 1
 router-id 2.2.2.2
!
interface GigabitEthernet0/0
 ip address 10.1.2.2 255.255.255.0
 ip ospf authentication
 ip ospf authentication-key cisco
 ip ospf network point-to-point
 ip ospf 1 area 0
 duplex auto
 speed auto
```

Note: Use '**service password-encryption**' command to encrypt the passwords in plain-text at config

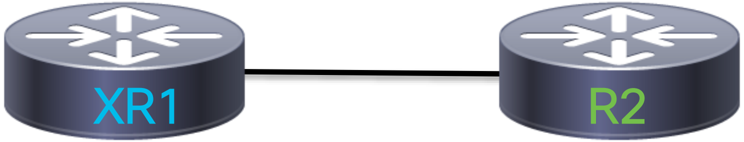
Database Protection: Maximum LSA



XR/XE

```
router ospf 1
 max-metric router-lsa
 max-lsa 23
```

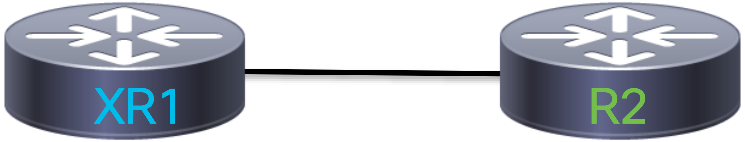
Redistribution Limit



```
XR-only
router ospf 1
 maximum redistributed-prefixes [1-4294967295]
```

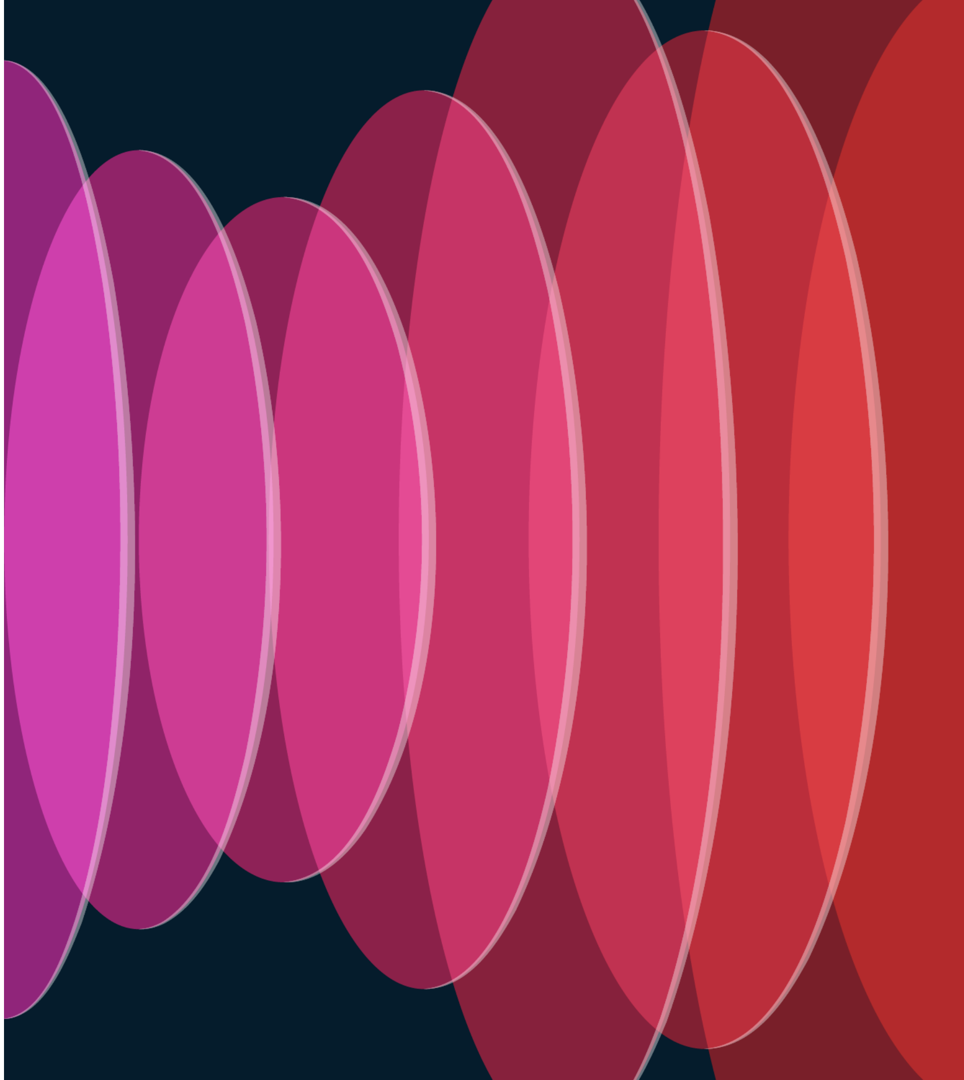
Generalized TTL Security

- Mitigates targeted attacks against OSPF that rely on the TTL
- A receive threshold is configured with the max number of hops that a packet may have travelled. The value for this hop-count argument is a number from 1 to 254, with a default of 1.



- **Note**: Be careful and avoid causing an outage! ☺

Security Hardening



Optimization Features

Prefix Suppression (1): Point-to-Point

- For each numbered point-to-point network, a router has two link descriptions in its router-LSA: one Type 1 link (point-to-point) describing the neighboring router, and one Type 3 link (stub) describing the assigned IPv4 subnet

XE

```
r2(config)#int gi 0/0
r2(config-if)#ip ospf prefix-suppression
```

XR

```
router ospf 1
 area 0
  interface GigabitEthernet0/0/0/0
   prefix-suppression
```

Prefix Suppression (2): Broadcast Networks

- A broadcast network joins many (more than two) routers and supports the capability to address a single physical message to all of the attached routers
- A special subnet mask value of 255.255.255.255 MUST be used in the network-LSA to hide a transit-only broadcast network.
- **Food for thought**: What if a router not-capable of RFC 6860 receives a Network-LSA with a subnet mask of 255.255.255.255?

Stub Router

- Used to advertise a system is out-of-service and cannot be used as transit
- Announces **max-metric** in the Router-LSA

```
XR/XE

router ospf 1
 max-metric router-lsa
```

Flood Reduction

- The OSPF Flooding Reduction feature works by reducing unnecessary refreshing and flooding of already known and unchanged information
- To achieve this reduction, the LSAs are now flooded with the higher bit set, thus making them **DoNotAge (DNA)** LSAs.

XR

```
router ospf 1
 interface <INT>
  flood-reduction
```

XE

```
Interface <INT>
 ip ospf flood-reduction
```

Loopback as Stub Network

- If a loopback is required to be announced with a subnet mask other than /32, the **loopback-as-stub** feature is required

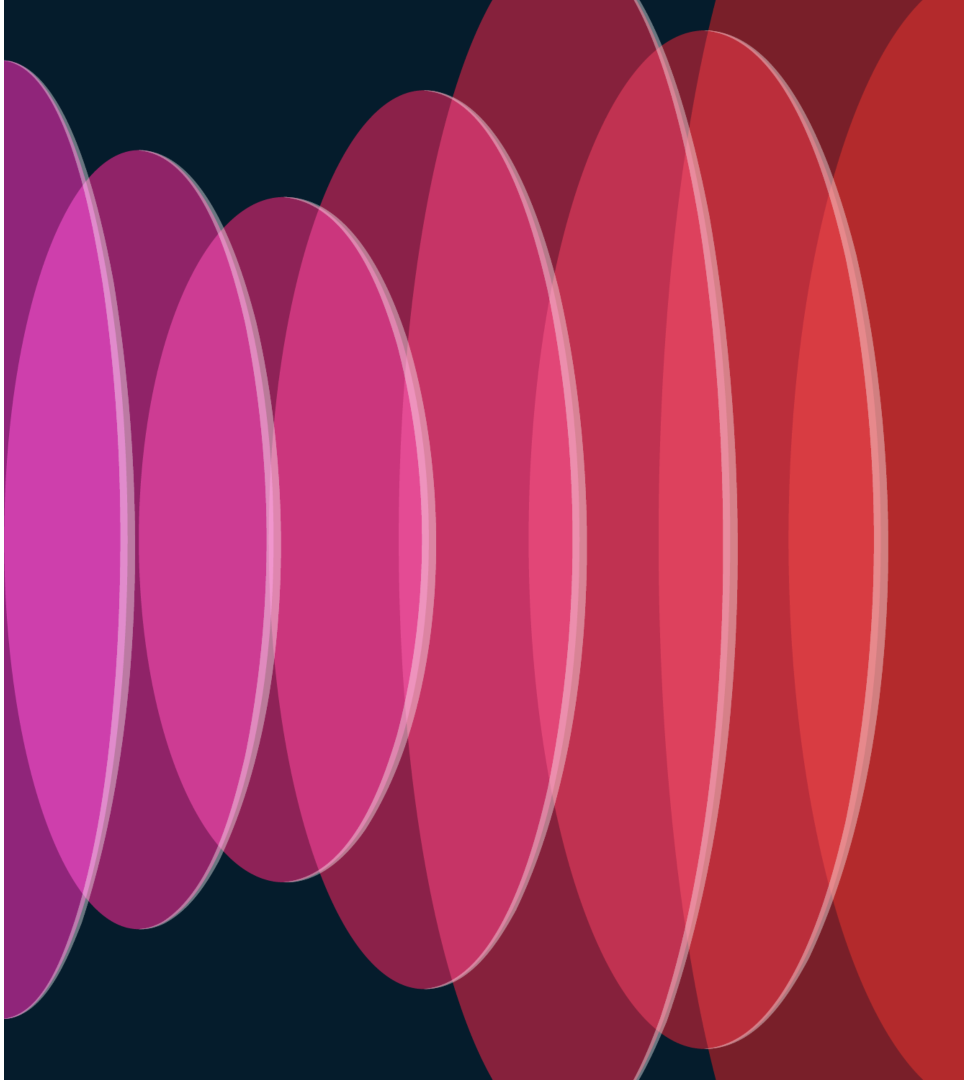
XR

```
router ospf 1
area 0
 interface Loopback0
  loopback stub-network enable
```

XE

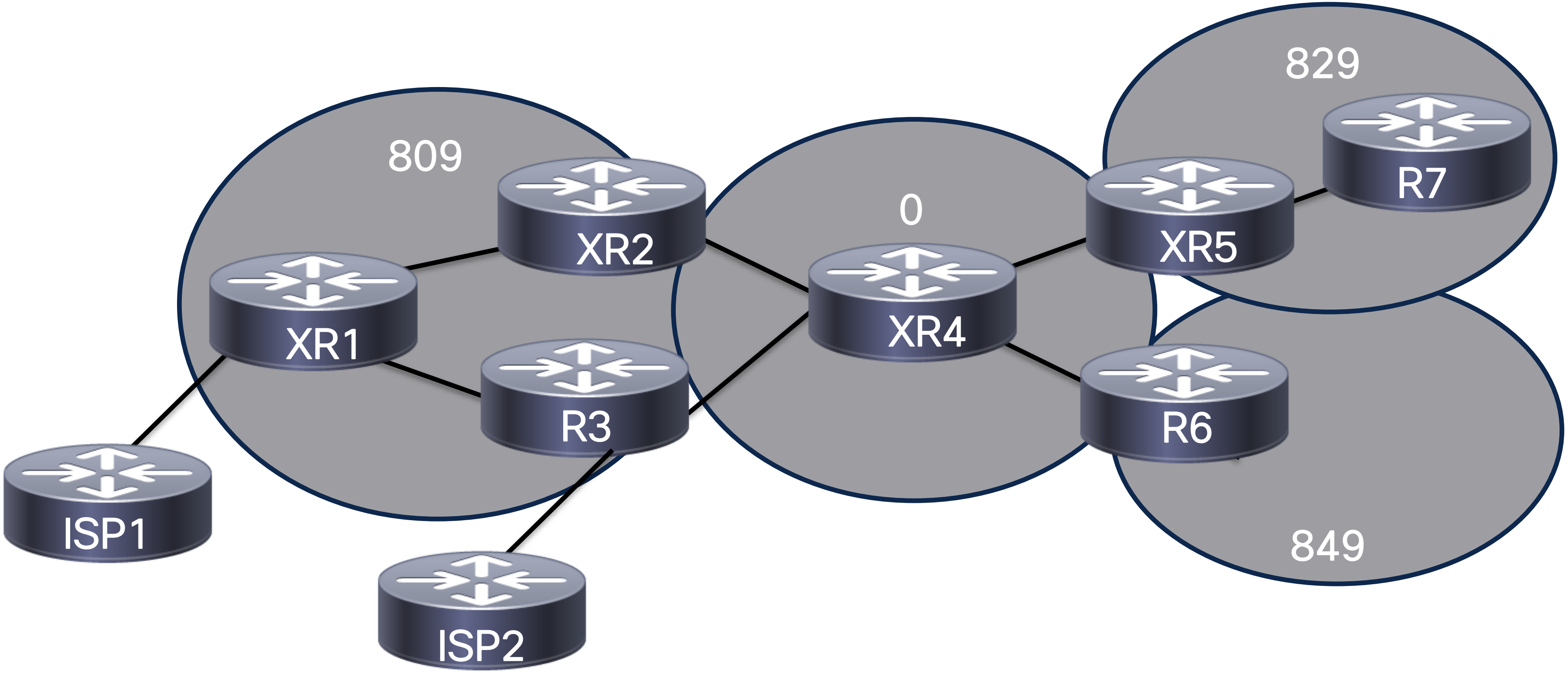
```


interface Loopback0
 ip ospf network point-to-point
```

Optimization Features



Stupid Routing Tricks!

Final Topology



Scenario 1: Aggregate Metrics

- **What?**: Aggregate the loopback11 and loopback111 on XR1
- **Question**: What path will XR4 use to route traffic to the aggregate **172.16.1.0/24**?

- **RFC 1583**: Uses the lowest metric of the components for the aggregated prefix
- **RFC 2328**: Uses the largest metric of the components for the aggregated prefix

Scenario 2: NSSA Translator

- **What?**: Make area 809 an NSSA, and force both ABRs to translate prefixes as they are advertised to the backbone areas

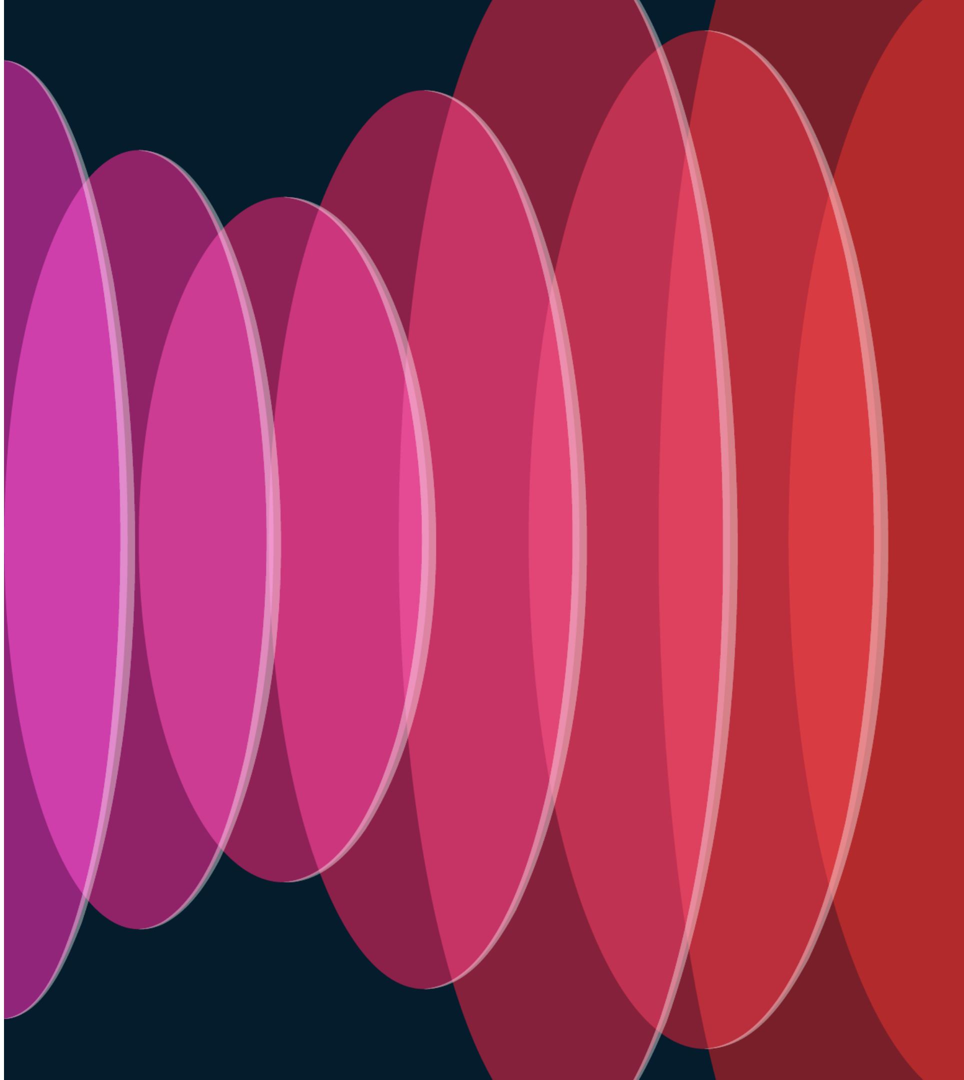
- Can you … Is it possible? ☺

Scenario 3: P-bit trick

- **What?**: Contain the advertisement of 192.0.2.1/32 within the NSSA only, do not use the **nssa-only** option in the **redistribute** statement.

- Can you … Is it possible? ☺

Stupid Routing Tricks!



Complete Your Session Evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to **win 1 of 5 full conference passes** to Cisco Live 2025.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn **exclusive prizes!**

Complete your surveys in the **Cisco Live mobile app.**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: eariasso@cisco.com

The bridge to possible

Thank you

#CiscoLive