NETWORKERS 2004

ADVANCE OSPF DEPLOYMENT

SESSION RST-4301

Agenda

Cisco.com

- OSPF Deployment Tips
- OSPF Dialup Design Tips
- Fast Convergence
- Non-Stop Forwarding
- Stub Router Advertisement
- Flood Reduction

Market Segments

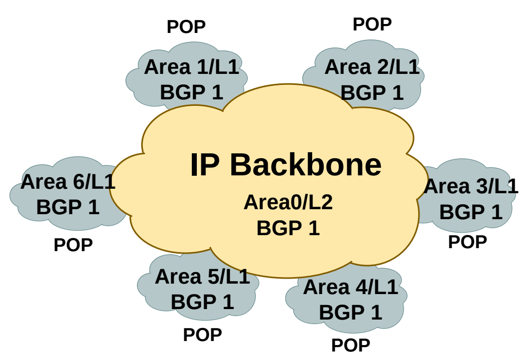
Cisco.com

- **Market segments**
  - **a) Service providers**
  - **b) Enterprise**
    - **Manufacturing**
    - **Retail**

SP Deployment Characteristics

Cisco.com

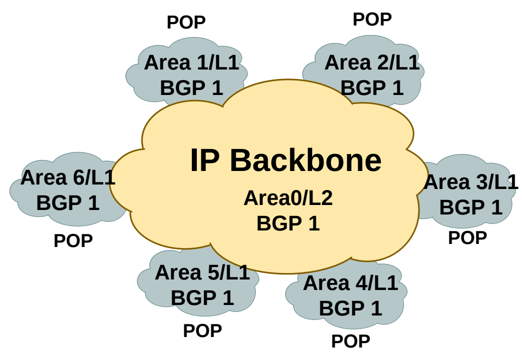
- **SPs should have only one instance of IGP running throughout network (exceptions are there)**
- **BGP carries external reachability**
- **IGP carries only next-hop (loopbacks are better for e.g., next-hop-self)**



SP Architecture

Cisco.com

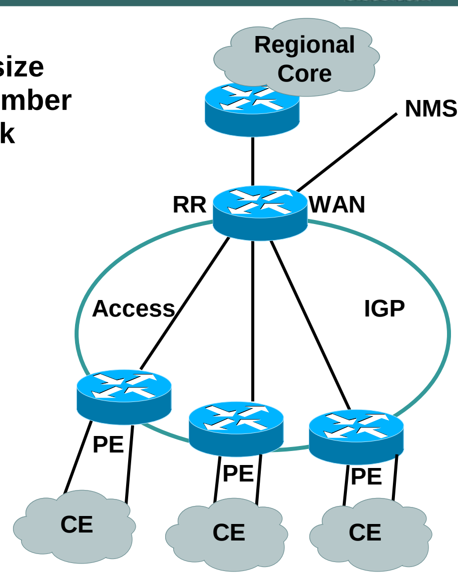
- Major routing information is 110K via BGP
- Largest known IGP routing table is ~6–7K
- Total of 117K
- 6K/117K~5% of IGP routes in an ISP network
- A very small factor but has a huge impact on network convergence!



SP Architecture

Cisco.com

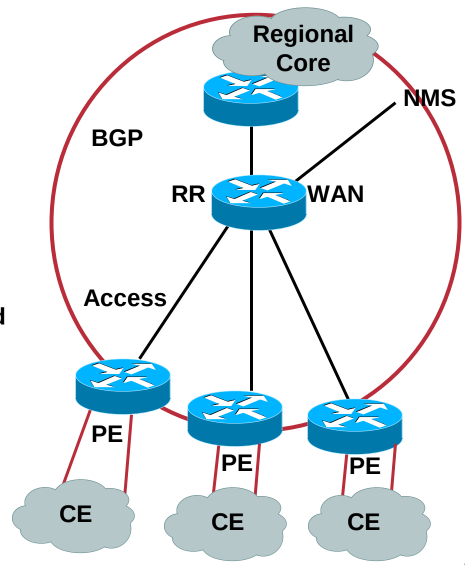
- You can reduce the IGP size from 6K to approx the number of routers in your network
- This will bring really fast convergence
- Optimized where you must and summarize where you can
- Stops unnecessary flapping



SP Architecture

Cisco.com

- The link between PE-CE needs to be known for management purpose
- BGP next-hop-self should be done on all access routers—unless PE-CE are on shared media (rare case)
- This will cut down the size of the IGP
- For PE-CE link do redistributed connected in BGP
- These connected subnets should ONLY be sent through RR to NMS for management purpose; this can be done through BGP communities

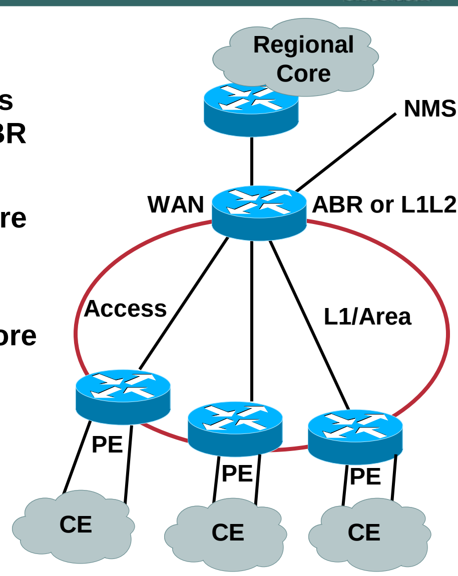


RST-4301
9722_05_2004_c2

SP Architecture

Cisco.com

- Where do we define area boundaries? WAN routers can be L1L2 in ISIS or ABR in case of OSPF
- Hide the pop infrastructure from your core
- Traffic engineering if needed can be done in core from WAN routers

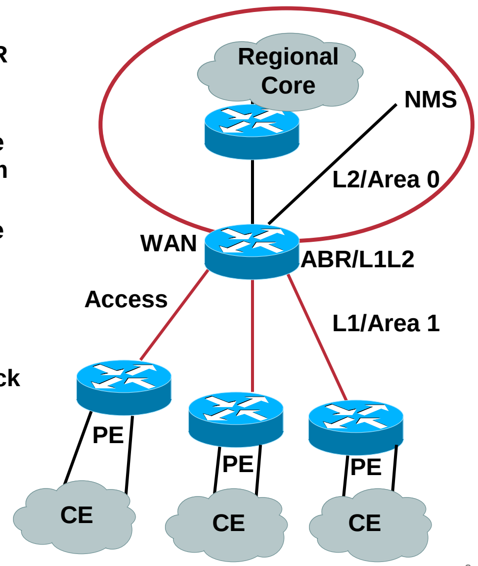


RST-4301
9722_05_2004_c2

SP Architecture

Cisco.com

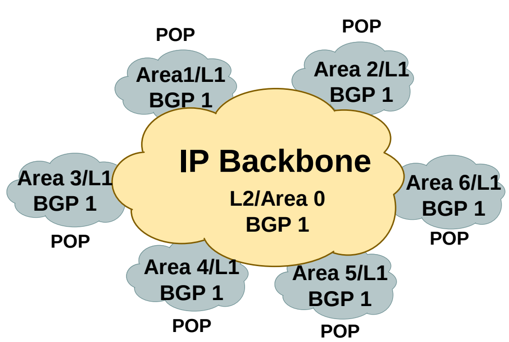
- Physical address between ABR and PE should be in a contiguous blocks
- These physical links should be filtered via Type 3 filtering from area 0 into other areas
- Why? To reduce the size of the routing table within each pop
- Every area will carry only loopback addresses for all routers
- Only NMS station will keep track of those physical links
- PE device will not carry other Pop's PE's physical address in the routing table



SP Architecture

Cisco.com

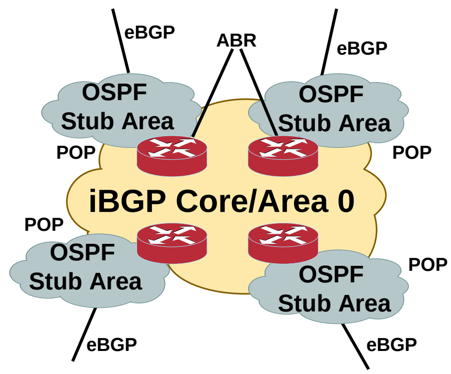
- Area 0 will contain all the routes
- This is the most intelligent form of routing and also there will not be too many routes in IGP
- If there are 500 pops and every pop contains 4 routers; then instead of having 6K routes you will only have 2K
- This is scalable and hack proof network!



Accidental Redistribution Prevention (OSPF)

Cisco.com

- Areas should be defined as stub to prevent accidental redistribution of eBGP into OSPF
- Type 3 LSA filtering should be used at ABR's and only routers' loopbacks should be allowed to leak into other areas
- Loopback should be in private address space to make LSA type 3 filtering easier; for e.g., 10.0.0.0/8
- iBGP routes can not be redistributed into IGP by default
- NMS resides in area 0 here



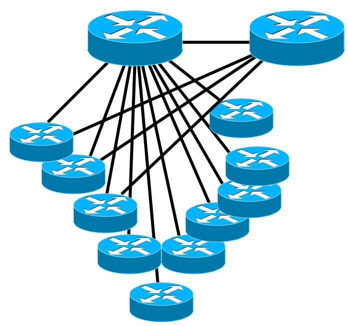
Market Segments

Cisco.com

- Market segments
  - a) Service providers
  - b) Enterprise
    - Manufacturing
    - Retail

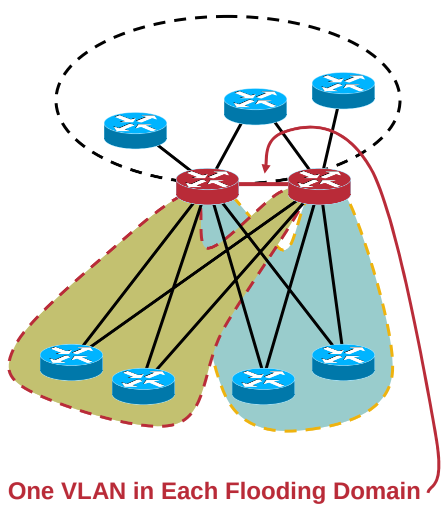
Enterprise Retail

- OSPF is not very good choice for hub and spokes
- EIGRP, ODR, RIPv2 and BGP are better choice here
- Enterprise BGP is not complicated
- You do not need to play with lot of attributes



Enterprise Retail

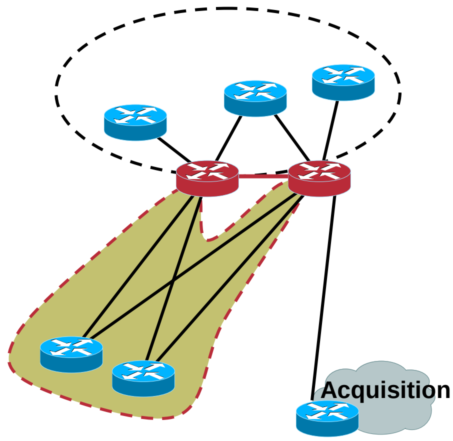
- The link between 2 hub routes should be equal to the number of areas
- Summarization of areas will require specific routing information between the ABR's
- This is to avoid suboptimal routing
- As you grow the number of areas, you will grow the number of VLAN/PVC's
- This is protocol limitation



Enterprise Retail

Cisco.com

- Spoke router in one area will contains route for the other spoke router since they are in the same area
- Acquisitions and merger will create another sets of problem
- Rearrangement of topology required if the area or the router limitation has been reached
- Very difficult to preserve the protocol's hierarchical design

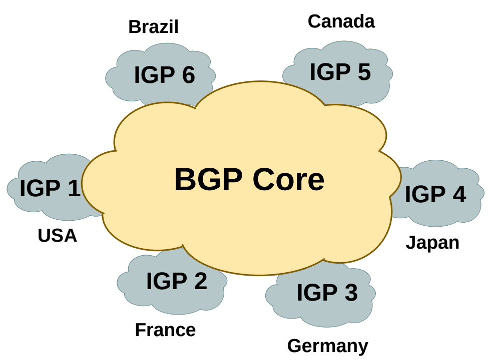


RST-4301
9722_05_2004_c2

Enterprise Manufacturing

Cisco.com

- Can have multiple 'islands' of IGPs
- Islands tied together by a BGP core
- May be a requirement for redistribution



RST-4301
9722_05_2004_c2

Agenda

Cisco.com

- OSPF Deployment Tips
- OSPF Dialup Design Tips
- Fast Convergence
- Non-Stop Forwarding
- Stub Router Advertisement
- Flood Reduction

Dial Backup and OSPF

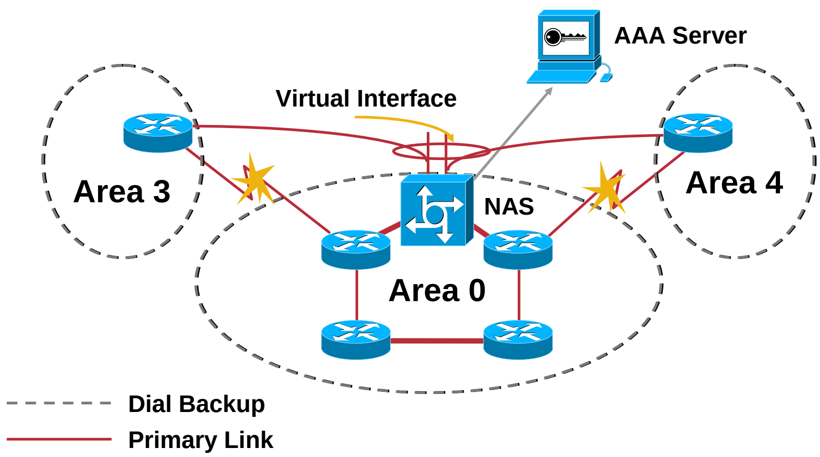
Cisco.com

- Dial-on-demand techniques:
  - Use virtual profiles
  - Virtual interface gets assigned to the area associated with the calling router
  - Area configuration and IP address for virtual interface are dynamic

Dial Backup and OPSF

Cisco.com

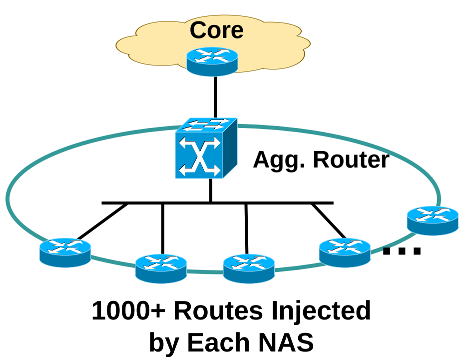
Virtual Profiles



Dialup Design Practices

Cisco.com

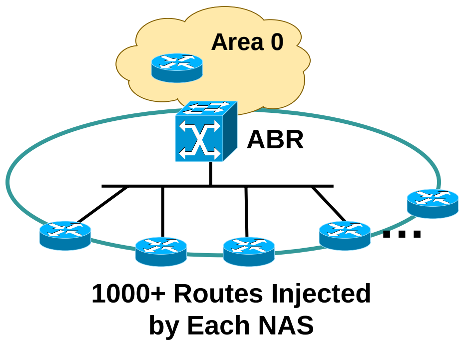
- Two kinds of Pools can be defined on NAS: Static Pools and Distributed Pools
- Static Pool: address range remain within a single NAS—easier to manage from routing perspective
- Dynamic Pool: address range may be distributed into multiple NAS's—hard to from a routing perspective



Dialup Design with Static Pool Addresses

Cisco.com

- Three ways to propagate dialup routes from NAS:
  - Either Static route to pool address to null 0 with redistribute static on NAS or
  - Assign the pool add on a loopback on NAS with OSPF p2p network-type including loopback in an OSPF area or
  - Static route on ABR for the pool address pointing towards NAS (ASBRs)—this is a preferred method because summarization can be done at ABR
- Static pool do not require redistribute connected subnets on NAS



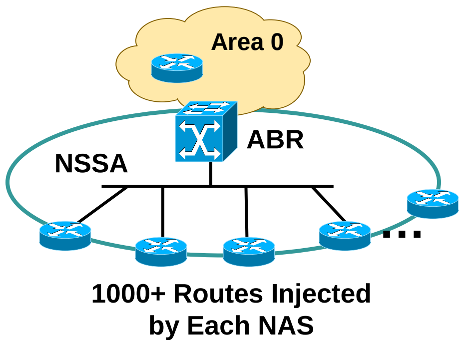
RST-4301
9722_05_2004_c2

21

Dialup Design with Dynamic Pool Addresses

Cisco.com

- Distributed pool REQUIRES REDISTRIBUTE CONNECT SUBNETS
- If pool is distributed, you can't summarize the pools at ABR because of REDISTRIBUTE CONNECTED SUBNETS on NASs' unless it's an NSSA, why?
- NSSA can summarize routes at ABR or ASBR



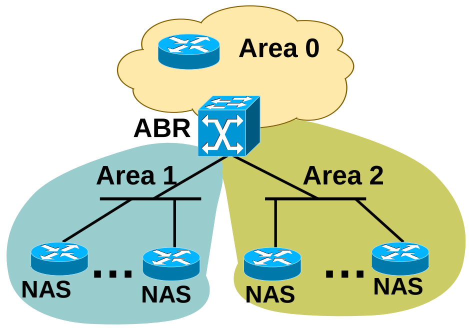
RST-4301
9722_05_2004_c2

22

Dialup Design Practices Scalability Issues

Cisco.com

- If an area has too many routes injected by NAS then break it up in more than one area
- Area should be configured as NSSA for controlling type 5 at ABR level
- NSSA ABR can filter type 5 originated by NAS servers
- Configure totally NSSA so one area type 5 will not go into other areas



1000+ Routes Injected by Each NAS

RST-4301
9722_05_2004_c2

Agenda

Cisco.com

- OSPF Deployment Tips
- OSPF Dialup Design Tips
- Fast Convergence
- Non-Stop Forwarding
- Stub Router Advertisement
- Flood Reduction

RST-4301
9722_05_2004_c2

Network Convergence

Cisco.com

- **Network convergence is the time needed for traffic to be rerouted to the alternative or more optimal path after the network event**
- **Network convergence requires all affected routers to process the event and update the appropriate data structures used for forwarding**

Network Convergence

Cisco.com

- **Network convergence is the time required to:**

  Detect the event

  Propagate the event

  Process the event

  Update the routing table/FIB

Network Convergence: Event Detection

Cisco.com

- **When physical interface changes state, driver must notify the routing process**
  - This should happen in a ms range
  - Carrier-delay—default is 2s
- **Some events are detected by IGP**
  - LAN is a typical example
    - Neighbor is lost, but interface is UP/UP
  - Hello mechanism has to detect the neighbor lost
    - Min Hello interval is 1 second currently

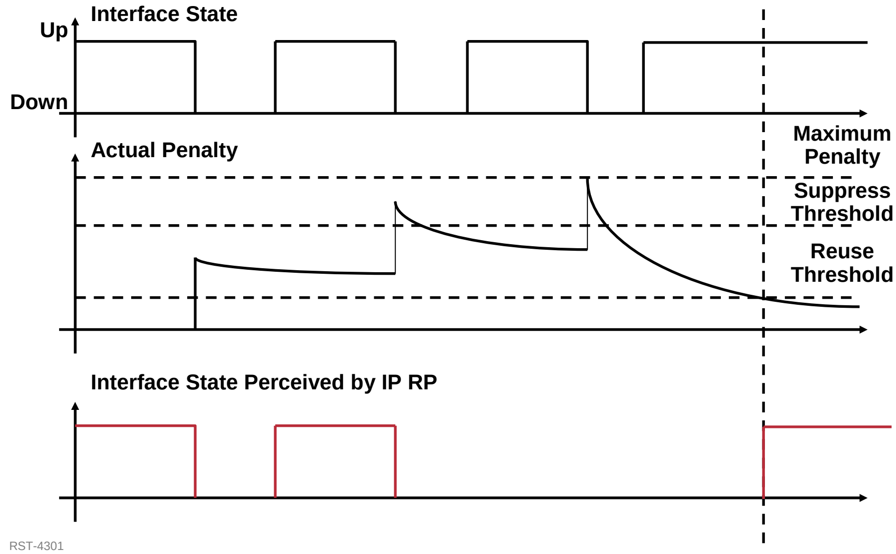
Interface Event Dampening

Cisco.com

- **When interface state changes both RT and RPs are notified**
- **Unstable interface cause excessive RT processing and RP updates**
- **Purpose is to dampen the interface from the RT and RPs perspective (only works for IP)**
- **CSCdt88027**
- **CLI:**

interface x/y
dampening [half-life] [reuse suppress max-time] [restart [penalty]]

Interface Event Dampening



Network Convergence: Subsecond Hellos

- **Problem:**

  At what frequency should hellos be sent?

  Neighbor can be declare down due to the Hello(s) being dropped/lost over a very short congestion period

  Router may not be able to send Hellos fast enough if most of its resources are used for other tasks

Network Convergence: Event Detection

Cisco.com

- OSPF Hello packet contains values of Hello and Dead interval
  - Must match between neighbors
  - Values are in seconds

…		
HelloInterval	Options	Rtr Pri
RouterDeadInterval		
…		

Network Convergence: OSPF Subsecond Hellos

Cisco.com

- CSCdu67116:
  - DeadInterval—minimum 1 second
  - Hello multiplier is used to specify how many Hellos to send within 1 second
  - Hello interval will be advertised as 0 second
  - CLI:
    - *interface x/y*
    - *ip ospf dead-interval minimal hello-multiplier <3-20>*

Network Convergence: OSPF Subsecond Hellos


```
Topvar(config)#int eth 1/3
Topvar(config-if)#ip ospf dead-interval minimal hello-multiplier 3
Topvar#sh ip ospf int eth 1/3
Ethernet1/3 is up, line protocol is up
 Internet Address 100.1.1.1/24, Area 0
 Process ID 1, Router ID 100.0.0.1, Network Type BROADCAST, Cost: 10
 Transmit Delay is 1 sec, State BDR, Priority 1
 Designated Router (ID) 100.0.0.2, Interface address 100.1.1.2
 Backup Designated router (ID) 100.0.0.1, Interface address 100.1.1.1
 Timer intervals configured, Hello 333 msec, Dead 1, Wait 1, Retransmit 5
   Hello due in 149 msec
…
Topvar#sh ip ospf neigh det
 Neighbor 100.0.0.2, interface address 100.1.1.2
   In the area 0 via interface Ethernet1/3
   Neighbor priority is 1, State is FULL, 6 state changes
   DR is 100.1.1.2 BDR is 100.1.1.1
   Options is 0x52
   LLS Options is 0x1 (LR)
   Dead timer due in 896 msec
…
```


Fast Hello’s: Scaling Issues

Scaling Is a Major Issue

300 Interfaces x 10 Neighbors/Interface = 3000 Neighbors

---

3 Hello Packets per Second on Each Interface

Router Has to Generate 900 Hellos per Second

3000 Neighbors Each Send 3 Hellos per Second to This Router

Router Has to Accept and Process 9000 Hellos per Second

Router Has to Deal with 9900 Hellos per Second

One Hello Every 10,000th of a Second

Network Convergence

Cisco.com

- Network convergence is the time required to:

  Detect the event

  Propagate the event

  Process the event

  Update the routing table/FIB

RST-4301
9722_05_2004_c2

Network Convergence: OSPF Event Propagation

Cisco.com

- After an event has been detected, LSA is generated to reflect the change
- LSA is not generated immediately

  OSPF_LSA_DELAY_INTERVAL—500ms delay

  Only used when generating router and network LSA

  The reason for this delay is to collect any changes that happens during the delay interval and include them all in the new LSA

RST-4301
9722_05_2004_c2

Network Convergence: OSPF Event Propagation

Cisco.com

- Origination of the new instance of the LSA is limited at the originator

  MinLSInterval

  The minimum time between distinct originations of any particular LSA; the value of MinLSInterval is set to 5 seconds

- Old CLI:

  timers lsa-interval <sec>"

  5s by default (as in spec)

Network Convergence: OSPF Event Propagation

Cisco.com

- Receiving of the LSAs is limited by the receiver

  MinLSArrival

  "For any particular LSA, the minimum time that must elapse between reception of new LSA instances during flooding. LSA instances received at higher frequencies are discarded. The value of MinLSArrival is set to 1 second."

- Old CLI:

  timers lsa-arrival <sec>"

  1 second by default (as in spec)

OSPF LSA Generation Exp. Backoff

Cisco.com

- LSAs are generated quickly after initial triggers
- Prolonged instability slows down the LSA generation process
- CSCdt21415
- CLI

timers throttle lsa all <lsa-start> <lsa-hold> <lsa-max>

timers lsa arrival <timer>

all values are in ms

OSPF LSA Generation Exp. Backoff

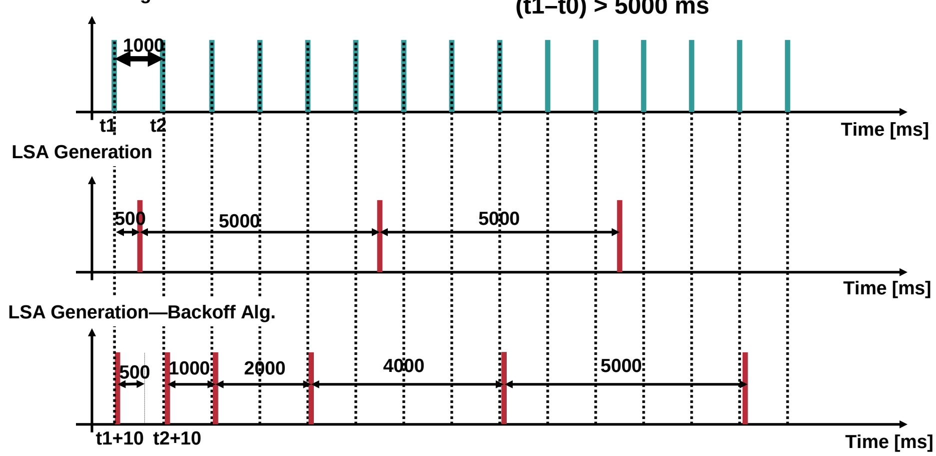
Cisco.com

timers throttle lsa all 10 500 5000

previous LSA generation at t0

(t1–t0) > 5000 ms

Events Causing LSA Generation



OSPF LSA Generation Backoff

Cisco.com

- Current RFC:
  - MinLSArrival is set to 1 second
  - LSA instances received at higher frequencies are discarded
- Need to set the MinLSArrival to *'lsa-hold'*
- *timers lsa arrival <'lsa-hold'>*

Network Convergence: Event Propagation

Cisco.com

- LSA has to be processed on each node
  - Detect if the LSA is newer/older
  - If the LSA is newer, detect if it carries any change
    - Number of links in the LSA (link comparison)
    - Size of the database (search)
  - If change detected schedule SPF (full/partial)
  - Install in the database
- Above processing time is rather small
  - For Router-LSA with 10 links: 0.5ms
  - For Router-LSA with 100 links: 1ms

Network Convergence: Event Propagation

Cisco.com

- **If LSA is declared newer, it's flooded over a certain set of interfaces**
  - Excluding the neighbor from which the LSA has been received
- **LSA is not flooded immediately**
  - Link state update packets are paced
  - Pacing timer is 33ms by default (jittered by 10%)

Network Convergence: OSPF Event Propagation

Cisco.com

- **With default values and no retransmission each node can add 33ms delay to the event propagation**
- **CSCds86112 make packet pacing configurable**
  - Default values are 33 msec/66 msec
    - *timers pacing flood <timer>*
    - *timers pacing retransmission <timer>*

Network Convergence

Cisco.com

- Network convergence is the time required to:
  - Detect the event
  - Propagate the event
  - Process the event
  - Update the routing table/FIB

Network Convergence: OSPF Event Processing

Cisco.com

- If there was a change detected in the LSA during the flooding, an SPF (full/partial) is scheduled
- Full SPF computation is delayed to protect router CPU resources
  - spf-delay interval—5 seconds by default
  - spf-holdtime—10 seconds by default
  - Old CLI (before CSCdt21362)
  - *timers spf <spf-delay> <spf-holdtime>*
  - Both timers in seconds
- Partial SPF is not delayed

Network Convergence: OSPF SPF Backoff

Cisco.com

- **spf-delay and spf-holdtime** protect the router resources, but **significantly slows down the convergence**
- **Solution:**
  - Exponential SPF backoff (CSCdt21362)
  - spf-delay/spf-holdtime changed to dynamic timers
  - Values configured in ms
    - *timers throttle spf <spf-start> <spf-hold> <spf-max>*
  - <spf-start>—Delay between receiving a change to SPF calculation
  - <spf-hold>—Delay between first and second SPF calculation
  - <spf-max>—Maximum wait time in milliseconds for SPF calculations

Network Convergence: SPF

Cisco.com

- **Two types of SPF**
  - Full SPF
    - Triggered by the change in router or network LSA
    - SPT tree is recomputed
    - All LSA types (Type-1/2/3/4/5/7) are processed
  - Partial SPF
    - Triggered by the change in Type-3/4/5/7 LSA
    - If triggered by Type-3/ all Type-3 LSAs that contribute to the certain destination are processed
    - If triggered by Type-5/7 all Type-5/7 LSAs that contribute to the certain destination are processed
    - If triggered by Type-4 all Type-4 LSAs that announce a certain ASBR and all Type-5/7 LSAs are processed

Network Convergence: SPF Time

Cisco.com

- SPF calculation time
  - Full spf:
    - Depends on:
      - Number of nodes/links in the area
      - Number of Type-3/4/5/7 LSAs
    - Some experimental numbers (GSR/7500)
      - 50 nodes fully connected topology~10ms
      - 100 node fully connected topology~25ms
      - 500 nodes~50 ms
      - 1000 nodes~100 ms
  - Partial SPF:
    - Fast—less then 0.5 ms

RST-4301
9722_05_2004_c2

49

Network Convergence: SPF

Cisco.com

- Any change in the router/network LSA triggers full SPF
  - Some changes does not represent the topology change:
    - Stub network UP/DOWN
    - IP address change on link
  - During the full SPF the whole SPT is rebuilt
    - Change in the topology may not require the whole SPT rebuild
    - Major part of the tree may stay the same in many cases

RST-4301
9722_05_2004_c2

50

Network Convergence: Incremental SPF

Cisco.com

- **Incremental SPF**
  - Modified Dijkstra algorithm
  - We keep the unchanged part of the tree
  - We rebuild only the affected parts of the tree
  - Re-attach the affected parts of the tree to the unchanged part of the tree

Network Convergence: Incremental SPF

Cisco.com

- **Gain of incremental SPF depends on how far (topologically) the change happens from the calculating node**
- **If the change affects only a small part of the topology, gain is significant**
  - We were able to run SPF and update the RT for the 1000 node network in less then 10 ms
- **If the change is close to the calculating node and affect almost the whole topology, there will be no gain in i-spf**

Network Convergence: Incremental SPF

Cisco.com

- There are always nodes close (closer) to the topological change and nodes that are remote (more remote) to it
- Flooding takes some time—nodes that are most remote from the change are usually notified last
- If full SPF runs on all nodes regardless of the change, then routers notified as last about it will converge last (giving that it takes same amount of time to run SPF on each node)
- With i-SPF, more remote the node is from the change, less work it needs to do during i-SPF, resulting in faster network wide convergence

RST-4301
9722_05_2004_c2

Network Convergence: Incremental SPF

Cisco.com

- Stub link down event (IP prefix lost):

Full SPF:

```
Sep 25 14:07:37.795:   OSPF: Begin SPF at 187751.852ms, process time 149100ms
Sep 25 14:07:37.795:   spf_time 2d04h, wait_interval 10s
Sep 25 14:07:37.839:   OSPF: End SPF at 187751.896ms, Total elapsed time 44ms
Sep 25 14:07:37.839:   Intra: 44ms, Inter: 0ms, External: 0ms
Sep 25 14:07:37.839:   R: 506, N: 786, Stubs: 620
Sep 25 14:07:37.839:   SN: 0, SA: 0, X5: 0, X7: 0
Sep 25 14:07:37.839:   SPF suspends: 0 intra, 0 total
```

Incremental SPF:

```
Sep 25 14:06:27.715:   OSPF: Begin SPF at 187681.772ms, process time 149016ms
Sep 25 14:06:27.715:   spf_time 2d04h, wait_interval 10s
Sep 25 14:06:27.719:   OSPF: End SPF at 187681.776ms,Total elapsed time 4ms
Sep 25 14:06:27.719:   Incremental-SPF: 0ms
Sep 25 14:06:27.719:   Intra: 0ms, Inter: 0ms, External: 0ms
Sep 25 14:06:27.719:   R: 0, N: 0, Stubs: 1
Sep 25 14:06:27.719:   SN: 0, SA: 0, X5: 0, X7: 0
Sep 25 14:06:27.723:   SPF suspends: 0 intra, 0 total
```

RST-4301
9722_05_2004_c2

Network Convergence: Incremental SPF

Cisco.com

- **Link up Event—part of the network becomes reachable:**

Full SPF:

```
Sep 25 14:27:13.463:    OSPF: Begin SPF at 188927.520ms, process time 149760ms
Sep 25 14:27:13.463:    spf_time 2d04h, wait_interval 5s
Sep 25 14:27:13.515:    OSPF: End SPF at 188927.572ms, Total elapsed time 52ms
Sep 25 14:27:13.515:    Intra: 48ms, Inter: 0ms, External: 0ms
Sep 25 14:27:13.515:    R: 488, N: 758, Stubs: 598
Sep 25 14:27:13.515:    SN: 0, SA: 0, X5: 0, X7: 0
Sep 25 14:27:13.515:    SPF suspends: 0 intra, 0 total
```

Incremental SPF:

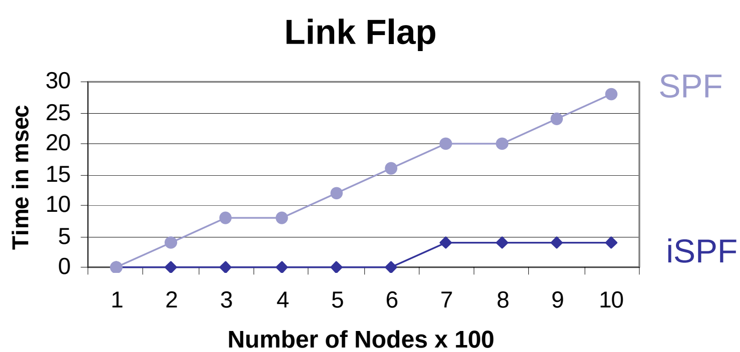
```
Sep 25 14:23:13.467:    OSPF: Begin SPF at 188687.524ms, process time 149612ms
Sep 25 14:23:13.467:    spf_time 2d04h, wait_interval 5s
Sep 25 14:23:13.479:    OSPF: End SPF at 188687.536ms, Total elapsed time 12ms
Sep 25 14:23:13.479:    Incremental-SPF: 0ms
Sep 25 14:23:13.479:    Intra: 8ms, Inter: 0ms, External: 0ms
Sep 25 14:23:13.479:    R: 18, N: 29, Stubs: 22
Sep 25 14:23:13.479:    SN: 0, SA: 0, X5: 0, X7: 0
Sep 25 14:23:13.479:    SPF suspends: 0 intra, 0 total
```

RST-4301
9722_05_2004_c2

Network Convergence: Incremental SPF

Cisco.com

- **Time it takes to run the SPF with the transit link flap**



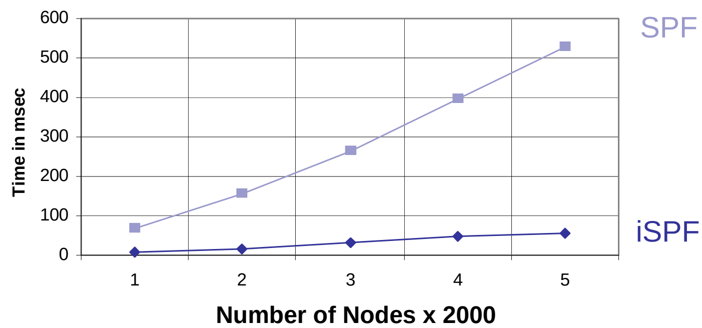
RST-4301
9722_05_2004_c2

Network Convergence: Incremental SPF

Cisco.com

- **Time it takes to run the SPF with the transit link flap**

Link Flap



Network Convergence: OSPF Incremental SPF

Cisco.com

- **CLI**

  Router ospf <process number>
  ispf

- **'debug ip ospf'**

  Includes i-spf related output

- **'sh ip ospf'**

```
Routing Process "ospf 1" with ID 170.99.99.99 and Domain ID 0.0.0.1
    Supports only single TOS(TOS0) routes
    Supports opaque LSA
    It is an area border and autonomous system boundary router
    Redistributing External Routes from,
    SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
    Incremental-SPF enabled
    Minimum LSA interval 5 secs; minimum LSA arrival 1 secs
```

Network Convergence

Cisco.com

- **Network convergence is the time required to:**
  - Detect the event
  - Propagate the event
  - Process the event
  - Update the routing table/FIB

Network Convergence: Update the Routing Table/FIB

Cisco.com

- **Based on some experiments, this represents a significant time in the convergence**
- **Projects are under way to make the RIB faster, more scalable and to improve the FIB info download to the linecards**

Network Convergence: Summary

Cisco.com

- With the mentioned features deployed sub-second network convergence is possible with OSPF
- Fast convergence must not affect the stability and scalability of the network

Agenda

Cisco.com

- OSPF Deployment Tips
- OSPF Dialup Design Tips
- Fast Convergence
- Non-Stop Forwarding
- Stub Router Advertisement
- Flood Reduction

NSF

Cisco.com

- 2 Issues (with the existing OSPF deployments)
  - When a RP/LC gets reset, we tear down the whole adjacency
  - If LSDB has to be resynched, we have to put the neighbor FSM state into ExStart State, which causes the adjacencies to be removed from the router-LSAs

NSF

Cisco.com

- Issues addressed (via): [Hellos]
  - From a restarting-router point-of-view
    - OSPF Hello Packets will carry additional information which includes
      - The LR Bit
      - The RS Bit
    - In the EO-TLV which is a part of LLS data block
    - By doing this, the "restarting router" is saying to its neighbors not to reset/tear down their adjacency to this router

NSF

Cisco.com

- **Issues addressed (via): [Hellos]**
  - **From the neighboring-router(s) point-of-view**
    - **Should skip the two-way connectivity check with the announcing neighbor**
      - **Not generate a 1-WayReceived event for the neighbor if it does not find its own router ID in the list of neighbors**

NSF

Cisco.com

- **Issues addressed (via): [LSDB Sync]**
  - **LSDBD packet includes a new**
    - **R Bit**
  - **Indicating the OOB Resynchronization**
  - **OPSF Nei. Data structure includes a new field OOBResync indicating the router is currently performing OOB LSDB resynchronization**

NSF

Cisco.com

The Options Field

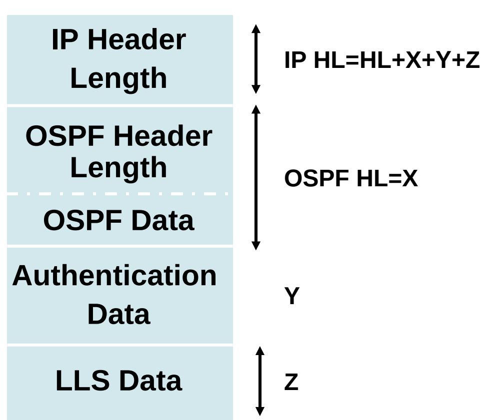
*	O	DC	L	N/P	MC	E	*

E	Describes the Way AS External-LSAs Are flooded
MC	Describes Whether IP MultiCast Datagrams Are Flooded
N/P	Handling of Type-7 LSAs
DC	Describes the Router's Handling of Demand Circuits
O	Describes the Router's Willingness to Receive and Forward Opaque-LSAs
L	Link Local Signaling (LLS)

NSF

Cisco.com

LLS Data Block



NSF

- **Two new TLVs have been defined as part of LLS data block**
  - Extended Options TLV
  - Cryptographic Authentication TLV

	# of Octets
Checksum	2
LLS Data Length	2
LLS TLVs	Variable

NSF

- **Extended Options TLV**
  - Type 1
  - 4-Bytes long
  - Should appear once in the LLS data block

	# of Octets
EO-TLV #: 1	2
Length	2
Extended Options	4

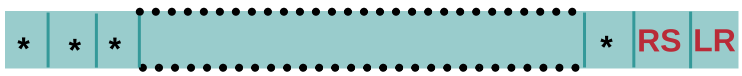
NSF

- EO-TLV has the following information carried in the Data Portion of the TLV

    (in the Extended Options)

    LR Bit (LSDB Resynchronization)

    RS Bit (Restart Signal)

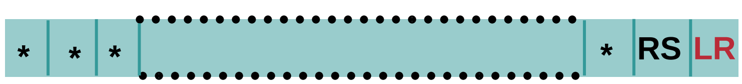


NSF

- LR Bit (LSDB Resynchronization)

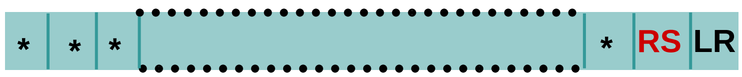
    To announce OOB Resynchronization capability to the neighbors as part of both Hello and DBD packets

    Uses 0x00000001 in the EO-TLV [part of LLS]

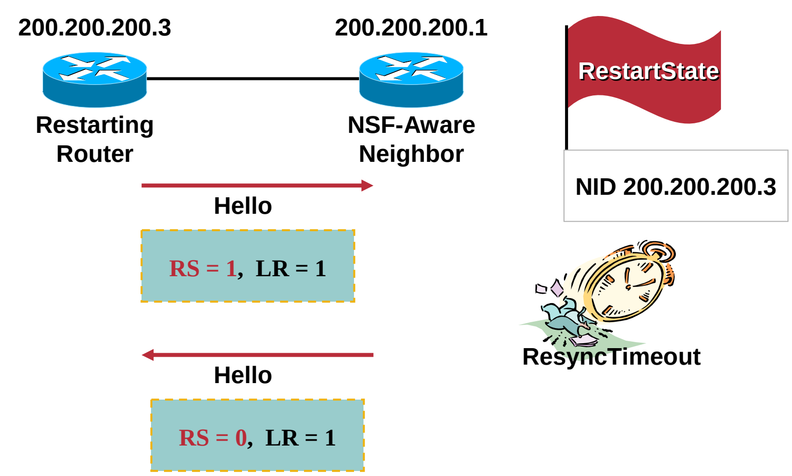


NSF

- RS Bit (Restart Signal)
  - "Restarting Router" should set this Bit in the EO-TLV attached in the Hello Packet to tell the neighbors to preserve their adjacencies
  - "Receiving Router" with RS Bit should not generate a 1-WayReceived event for the neighbor
  - "Receiving Router" should send a Unicast Hello back to "Restarting Router" in reply to a Hello Packet with the RS Bit
  - The RS Bit may not be set in Hello Packets longer than RouterDeadInterval seconds
  - Uses 0x00000002 in the EO-TLV [part of LLS]



OSPF: NSF Hello Exchange



NSF

Cisco.com

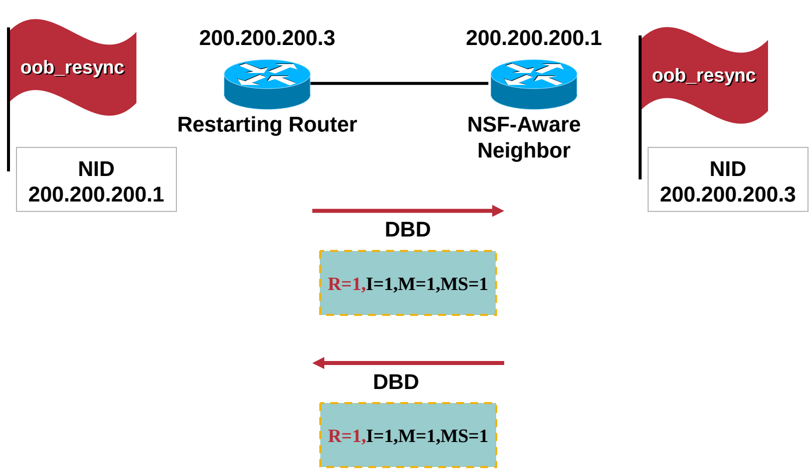
- The Database Description packet will include a new flag indicating OOB LSDB Resynchronization

0	0	0	0	R	I	M	MS

R—LSDB Resync Active Bit

OSPF: OOB Resync

Cisco.com



OSPF NSF CLI

Cisco.com

- The following commands are configured under “router ospf <proc>”

  nsf [enforce global]

  The “enforce global” option causes NSF processing to terminate for entire OSPF process, whereas the default behavior is that NSF will abort per-interface where non-NSF-aware neighbors are discovered

  timers nsf wait <interval>

  Hidden command—For setting the interface wait interval during NSF restart, as well as the first NSF Restart Timer period

  timers nsf flush <interval>

  Hidden command—For setting how much time to wait after RIB convergence, before flushing stale LSAs

RST-4301
9722_05_2004_c2

OSPF Configuration Example(s)

Cisco.com

- Restarting-Router

  Show IP OSPF [Abbreviated Output]

```
HA-Router#show ip ospf
 Routing Process "ospf 1" with ID 200.200.200.3
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
 Supports Link-local Signaling (LLS)
 Non-Stop Forwarding enabled, last NSF restart 00:25:00 ago
          (took 32 secs)
```

RST-4301
9722_05_2004_c2

OSPF Configuration Example(s)

Cisco.com

- **NSF-Aware Neighbor**

Show IP OSPF [Abbreviated Output]

```
NSF-Aware-Router#show ip ospf
 Routing Process "ospf 1" with ID 200.200.200.1
 Number of areas in this router is 1. 1 normal 0 stub 0 nssa
 External flood list length 0
 Non-Stop Forwarding enabled
    Router is not operating in SSO mode
```


OSPF Configuration Example(s)

Cisco.com

- **Restarting-Router**

Show IP OSPF Neighbor Detail [Abbreviated Output]

```
HA-Router#show ip ospf neighbor detail
 Neighbor 200.200.200.1, interface address 172.1.1.21
    In the area 1 via interface POS3/0
    Neighbor priority is 0, State is FULL, 7 state changes
    DR is 0.0.0.0 BDR is 0.0.0.0
    Options is 0x58
    LLS Options is 0x1 (LR), last OOB-Resync 00:26:13 ago
    Dead timer due in 00:00:31
    Neighbor is up for 00:26:34
```

OSPF Configuration Example(s)

Cisco.com

- **NSF-Aware Neighbor**

Show IP OSPF Neighbor Detail [Abbreviated Output]

```
NSF-Aware-Router#show ip ospf neighbor detail
 Neighbor 200.200.200.3, interface address 172.1.1.22
    In the area 1 via interface POS4/0
    Neighbor priority is 0, State is FULL, 42 state changes
    DR is 0.0.0.0 BDR is 0.0.0.0
    Options is 0x58
    LLS Options is 0x1 (LR), last OOB-Resync 00:46:35 ago
    Dead timer due in 00:00:34
```

RST-4301
9722_05_2004_c2

OSPF Configuration Example(s)

Cisco.com

Show IP OSPF Database

```
HA-Router#show ip ospf database
        OSPF Router with ID (200.200.200.3) (Process ID 1)
                Router Link States (Area 1)
Link ID         ADV Router     Age    Seq#       Checksum Link count
200.200.200.1   200.200.200.1 1332   0x80000068 0x7EA5    2
200.200.200.3 S200.200.200.3 1302   0x8000006C 0xE7B3    3
```

RST-4301
9722_05_2004_c2

NSF: Non-Stop Forwarding More Details

Cisco.com

- **draft-nguyen-ospf-oob-resync-xx.txt**

  Describes OSPF out-of-band resynchronization (oob-resync)

- **draft-nguyen-ospf-restart-xx.txt**

  Describes protocol restart signaling

- **draft-nguyen-ospf-lls-xx.txt**

  Describes extending the OSPF options field

Agenda

Cisco.com

- OSPF Deployment Tips
- OSPF Dialup Design Tips
- Fast Convergence
- Non-Stop Forwarding
- Stub Router Advertisement
- Flood Reduction

Stub Router Advertisement

Cisco.com

- There are situations, when we may want to avoid using the router that is connected in a network
  - Some time before/after upgrade, scheduled downtime
  - When the router is converging a large RT—typically BGP with full Net routes; OSPF has converged, but we do not want to send the transit traffic through the box until BGP has converged
  - While the router is reconfigured

Stub Router Advertisement

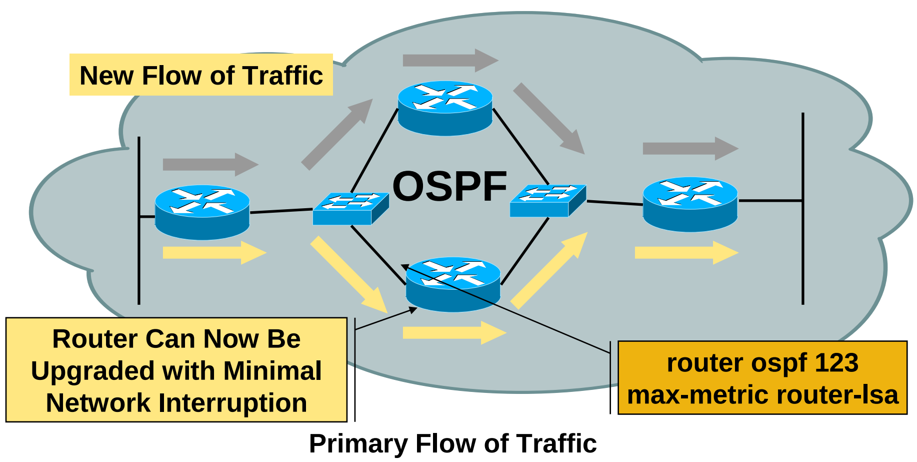
Cisco.com

- Router-LSA advertises:
  - (1) p2p links to other router
  - (2) Transit network
  - (3) Stub network
  - (4) VL
- 1, 2, 4 are transit links
  - Transit links are used for traffic that passes through the router (destination is not directly connected to the router)

Stub Router Advertisement

- If we want to avoid transit traffic through the box we set the metric of all transit links advertised by the router in its router LSA to 0xFFFF (max metric)
- All other routers will avoid the router and use alternative paths if such paths exist
- If no other paths are available:
  - Routers that implements RFC1247—links with max-metric will be simply ignored
  - Routers that implements RFC2328—links with max-metric will be used if no alternative paths exists

OSPF Stub Router Advertisement



Stub Router Advertisement

- max-metric router-lsa [on-startup {wait-for-bgp | <announce-time>}]

  Syntax description

• router-lsa	Always originate router-LSAs with maximum metric
• on-startup	Set max-metric temporarily after reboot
• announce-time	Time, in seconds, router-LSAs are originated with max-metric (default is 600s)
• wait-for-bgp	Let BGP decide when to originate router-LSA with normal metric (i.e., stop sending router-LSA with max-metric)

Stub Router Advertisement

Output of "show ip ospf" command:

```
 Routing Process "ospf 1998" with ID 14.18.134.155
   Supports only single TOS(TOS0) routes
   Supports opaque LSA
   It is an area border and autonomous system boundary router
   Redistributing External Routes from,
      static, includes subnets in redistribution
 Originating router-LSAs with maximum metric, Time remaining: 00:01:18
      Condition: on startup while BGP is converging, State: active
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
   Minimum LSA interval 5 secs. Minimum LSA arrival 1 secs
   Number of external LSA 7. Checksum Sum 0x47261
   Number of opaque AS LSA 0. Checksum Sum 0x0
 .....
```

Stub Router Advertisement

Cisco.com

Output of "show ip ospf database router" command:

```
Exception Flag: Announcing maximum link  costs
LS age: 68
     Options: (No TOS-capability, DC)
     LS Type: Router Links
     Link State ID: 14.18.134.155
     Advertising Router: 14.18.134.155
     LS Seq Number: 80000002
     Checksum: 0x175D
     Length: 60
     Area Border Router
     AS Boundary Router
     Number of Links: 3

Link connected to: a Transit Network
        (Link ID) Designated Router address: 192.1.1.11
        (Link Data) Router Interface address: 192.1.1.14
         Number of TOS metrics: 0
          TOS 0 Metrics: 65535 (metric used for local
calculation: 10)

       Link connected to: a Transit Network
        (Link ID) Designated Router address: 10.1.145.11
        (Link Data) Router Interface address:
10.1.145.14
         Number of TOS metrics: 0
          TOS 0 Metrics: 65535 (metric used for local
calculation: 10)

       Link connected to: a Stub Network
        (Link ID) Network/subnet number: 9.11.12.0
        (Link Data) Network Mask: 255.255.255.0
         Number of TOS metrics: 0
          TOS 0 Metrics: 1
```


Agenda

Cisco.com

- **OSPF Deployment Tips**
- **OSPF Dialup Design Tips**
- **Fast Convergence**
- **Non-Stop Forwarding**
- **Stub Router Advertisement**
- **Flood Reduction**

Flood Reduction

Cisco.com

- Each LSA has its age
- When LSA is originated its age is set to 0
- LSA is flushed from the area/domain when its age reaches MAXAGE (3600s)
- Each router must periodically refresh all self-generated LSAs
  - Refresh period is +/- 1800s

Flood Reduction

Cisco.com

- Why do we need to refresh periodically?
  - At the time the spec was written it looked to be a good to refresh to keep the databases in sync
  - Flooding is reliable and all changes are propagated reliably to all routers in area

Flood Reduction

Cisco.com

- OSPF has the concept of preventing the refresh for DC (Demand Circuit), like dialup links, ISDN,…
  - Do Not Age LSAs
  - DN bit
- DC has some drawbacks…
  - Hellos are not sent over DC, which can prevent the neighbor loss detection

Flood Reduction

Cisco.com

- We can generalize the Do Not Age concept
  - FR (Flood Reduction) link—same as DC, but Hellos are sent over it as over any p2p link
  - All LSAs sent over FR link will have the DN bit set
  - Those LSA will never timeout in the databases of routers behind the FR link

Flood Reduction

Cisco.com

- LSAs does not need to be refreshed over the FR links at all
- Changed LSA are flooded over FR link
- 'ip ospf flood-reduction' on the interface
- CSCdp80470
- draft-pillay-esnault-ospf-flooding-xx.txt
- Possible extension is to refresh at configured interval

Summary

Cisco.com

What We Learned?

- Effective OSPF Deployment Tips. How to use those techniques in your OSPF network
- Understanding of how OSPF Fast Convergence works
- Understanding of how Non Stop Forwarding works
- We learned about the Stub Router advertisement and where to use this feature
- We learned about Flood reduction feature and where to use it

Recommended Reading

Cisco.com

- RFC 2328(OSPF)
- 1587(NSSA) New RFC 3101
- 1793(Demand Circuit)
- RFC 2370 (for opaque support)
- Large-scale IP network solutions: CCIE Professional Development by Khalid Raza and Mark Turner ISBN: 1578700841
- Troubleshooting IP Routing Protocols by Faraz Shamim, Zaheer Aziz, Johnson Liu and Abe Martey ISBN: 1587050196

Available on-site at the Cisco Company Store

Complete Your Online Session Evaluation!

Cisco.com

WHAT: Complete an online session evaluation and your name will be entered into a daily drawing

WHY: Win fabulous prizes! Give us your feedback!

WHERE: Go to the Internet stations located throughout the Convention Center

HOW: Winners will be posted on the onsite Networkers Website; four winners per day