

ISIS Fast Convergence

IS-IS Tutorial

Agenda

- **Definition**
- **Objective**
- **Technology**
- **Conclusion**

IS-IS Tutorial

Loss of Connectivity / Convergence

- Incentive to reduce the loss of connectivity (LoC)
- Availability
 - 99.999% per day \Leftrightarrow 0.9sec of downtime
- VoIP
 - 40msec LoC: glitch
 - 1, 2 sec LoC: call drop

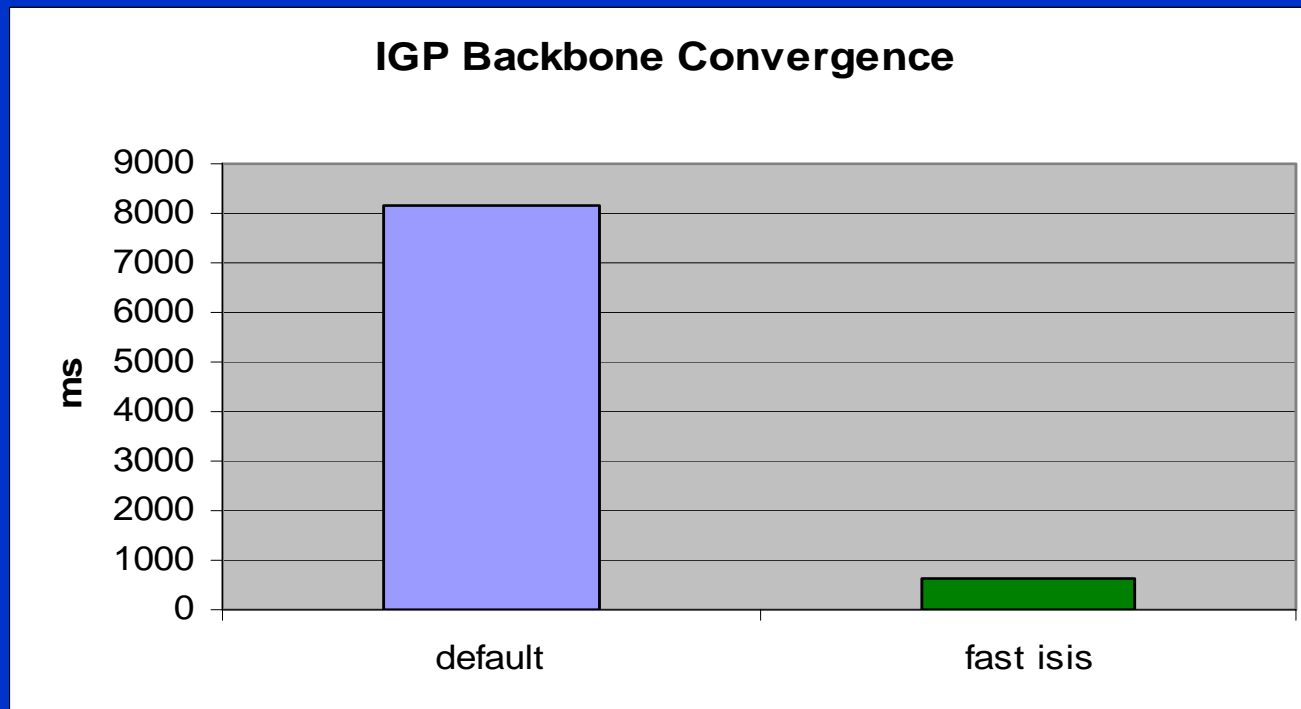
Loss of Connectivity

- **IGP Backbone Convergence:**
 - the time it takes for connectivity to be restored upon link/node failure/addition for an IP flow starting on an edge access router and ending on another edge access router, excluding any variation of BGP routes.
- **For this session, IGP = ISIS**

Historical ISIS Convergence

- 10 to 30 seconds
- Not excellent
- In the past, focus has been more on stability than on fast convergence
 - typical trade-off

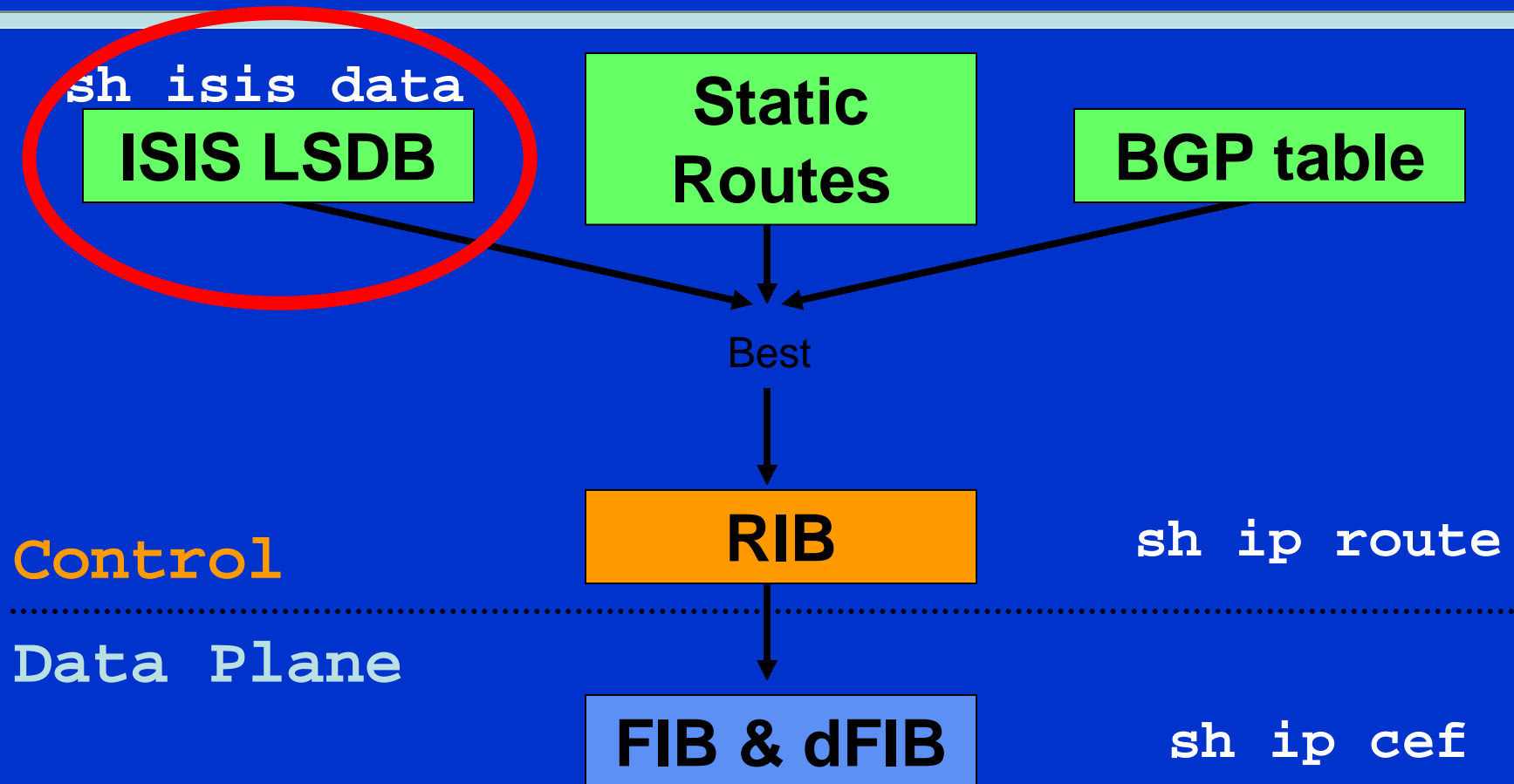
What this presentation will explain



- **ISIS Convergence in 1 or 2 second is realistic**

IS-IS Tutorial

LSDB, RIB and FIB



IS-IS Tutorial

SPF optimisations

IS-IS Tutorial

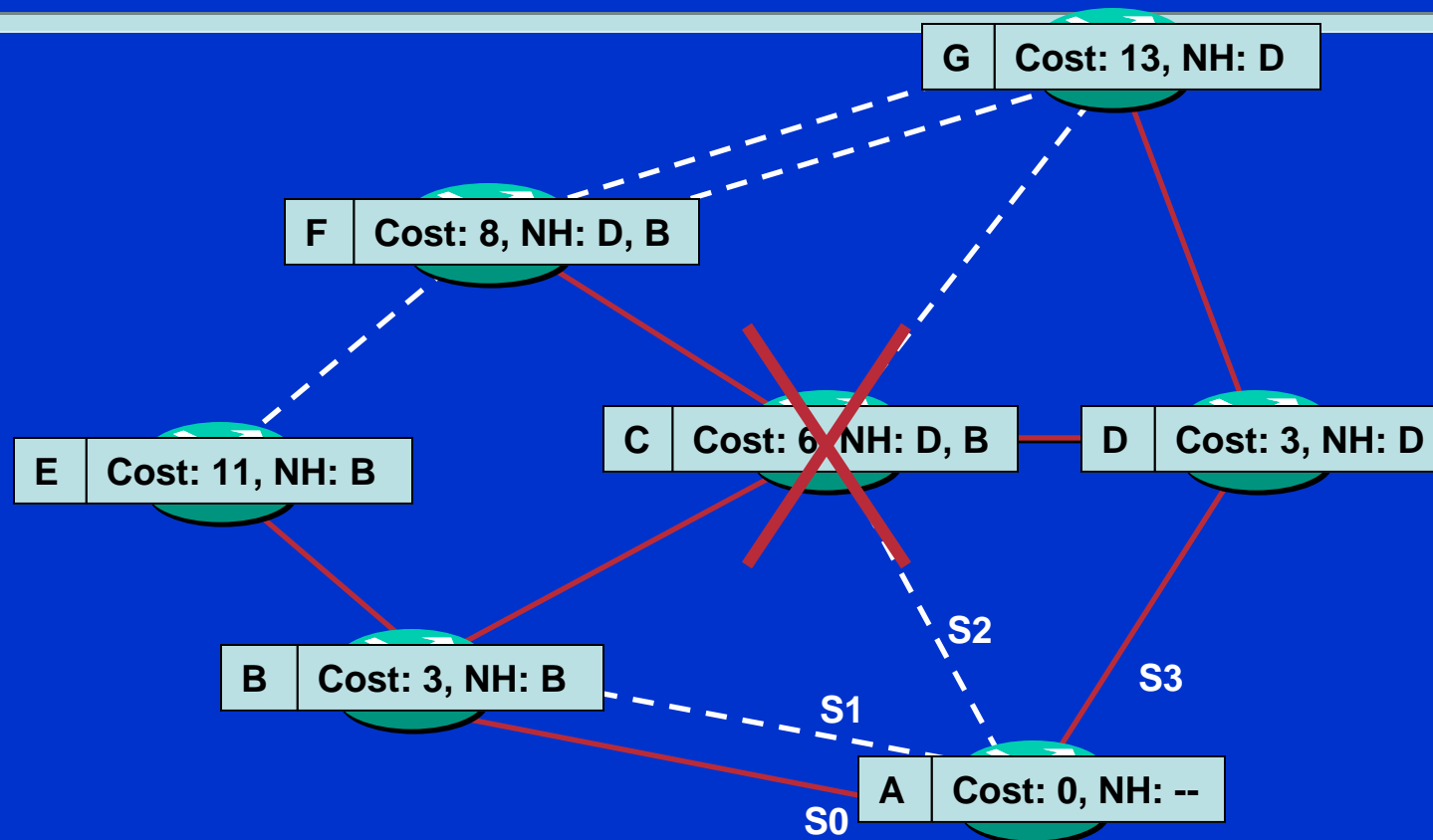
Shortest Path First Algorithm:

SPF and PRC

- Basic Implementation
 - **Any change (link, node, leave)**
 - → Recompute the whole SPT and the whole RIB
- Decouple SPT and RIB
 - **If any topology change (node, link)** Called "SPF"
 - → Recompute SPT and the RIB
 - **If only an IP prefix changes**
 - → Keep the SPT, just update the RIB for the nodes whose prefixes have changed Called "PRC"

Shortest Path First Algorithm

Full SPF: Node Down



Loss of this node breaks the SPT, therefore SPF is run to rebuild the SPT

IS-IS Tutorial

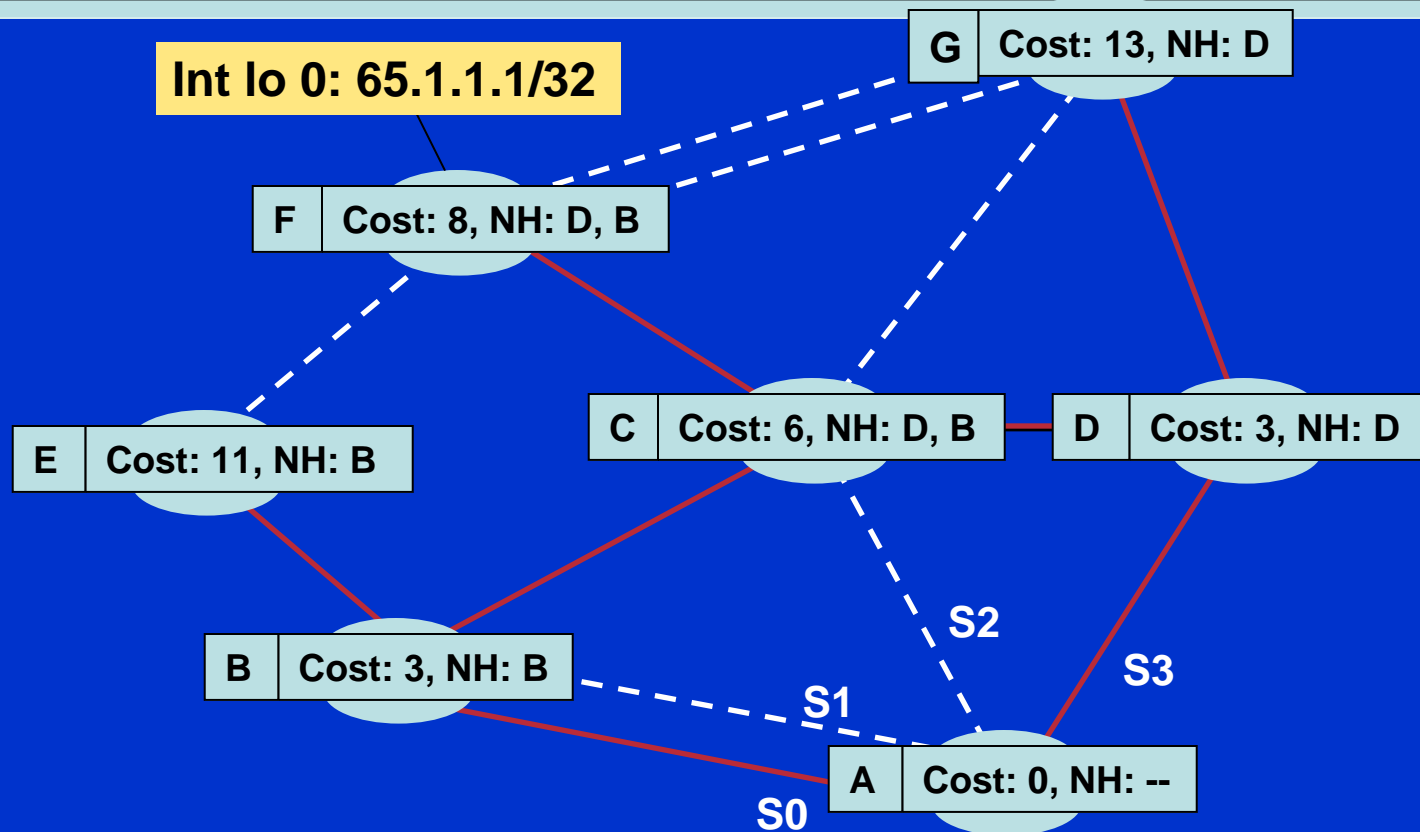
Shortest Path First Algorithm

Partial Route Calculation (PRC)

- When new LSPs are received, each router will check what has changed in the LSP
- **If only leaf routes have changed, the SPT does not need to be rebuilt**
 - As long as the new LSP still advertises the same neighbors with the same metric
- In this case we only reinstall IP prefixes of the newly received LSP into the routing table
- **PRC consists on taking the new LSP and reinstalling its prefixes into the routing table**

Shortest Path First Algorithm

Partial Route Calculation (PRC)



PRC here consists in just adding 65.1.1.1/32 in the RIB; the SPT is not affected

IS-IS Tutorial

Shortest Path First Algorithm

SPF, PRC & LSP Generation

- **Initial wait before SPF, PRC and LSP generation**
 - 5.5 seconds for SPF & PRC
 - 50 msec for LSP generation
- **Configurable minimum interval between consecutive events**
 - By default is
 - 5 seconds for LSP generation
 - 10 seconds for SPF & PRC calculation

Shortest Path First Algorithm

Exponential Backoff Timer

- IS-IS throttles its main events
 - **SPF computation**
 - **PRC computation**
 - **LSP generation**
- Throttling slows down convergence
- Not throttling can cause melt-downs
- Find a compromise...
- The goal is to **react fast to changes** but then to **backoff under constant churn** to avoid network meltdown

Shortest Path First Algorithm

Exponential Backoff Timer

- **Extended syntax**
spf-interval <a> [<c>]
<a> seconds between SPF runs (seconds)
 milliseconds between first trigger and SPF
<c> milliseconds between first and second SPF
- **Same syntax for**
prc-interval
lsp-gen-interval

Shortest Path First Algorithm

Exponential Backoff Timer

- Example: **spf-interval 10 100 1000**
(a) (b) (c)

- We decide to run an SPF
 - Wait 100 msecs, then run SPF ($b=100$)
 - Wait at least 1 second before running a second SPF if needed ($c = 1000$)
 - If we need to run a 3rd SPF, right after, wait at least 2 seconds ($c = 2c$)
 - Wait at least 4 sec before next SPF, then 8 sec, then 10 sec, 10 sec, ... ($c = \text{MIN}(2c, a)$)

Shortest Path First Algorithm

Exponential Backoff Timer

- When the network calms down, and there were no triggers for **at least twice the minimum interval** (20 secs in this example), then go back to fast behavior (after 100 msec initial wait)

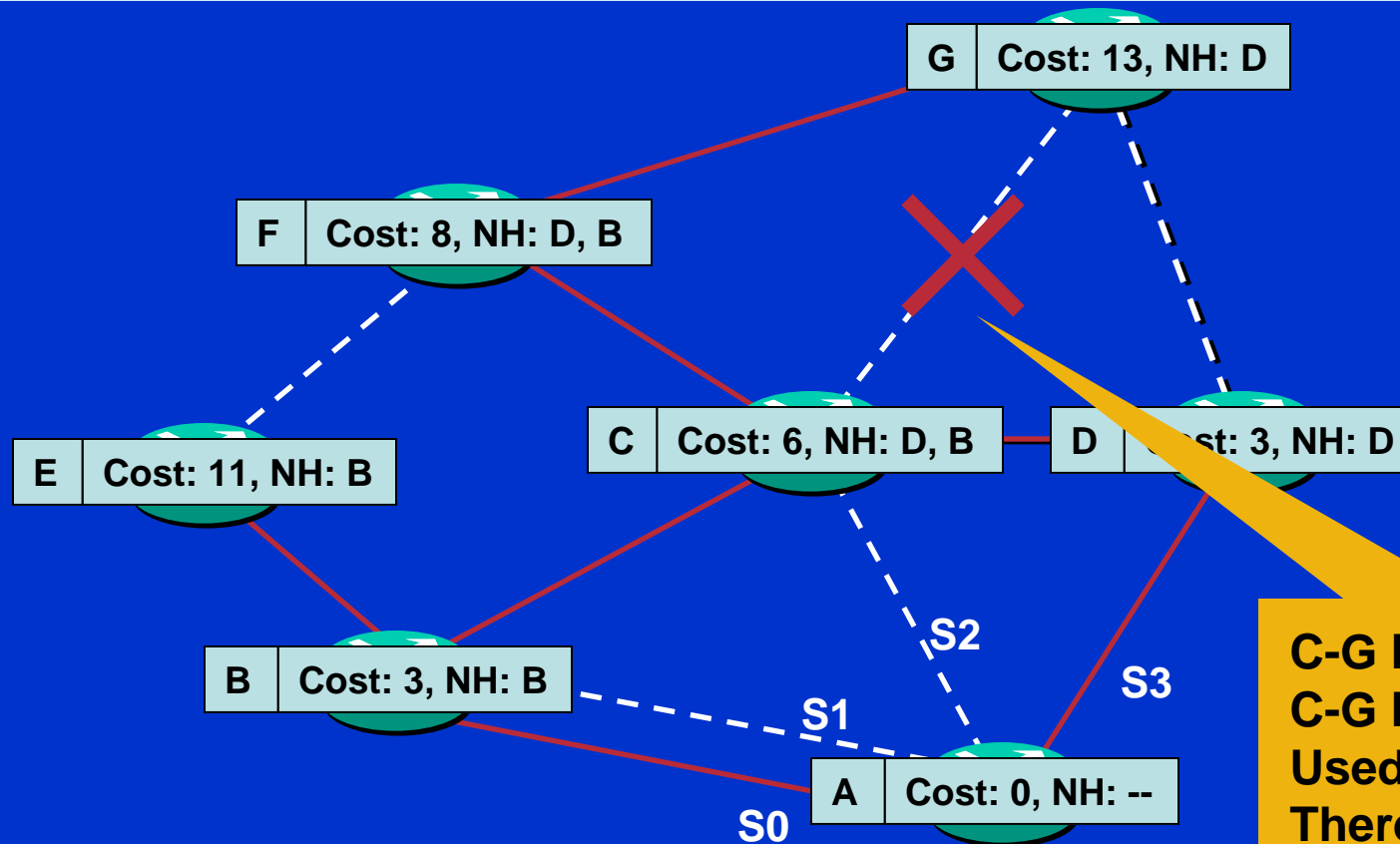
Shortest Path First Algorithm

Incremental SPF (i-SPF)

- Basic Operation
 - We **keep the unchanged part** of the tree
 - We **rebuild only the affected parts** of the tree
 - **Re-attach affected parts of the tree to the unchanged part of the tree**
- More information is kept in the SPT
 - Have to store and compare current and new information
 - The primary lists we maintain are:
 - Parents List
 - Neighbour List
 - Memory impact for using i-SPF (but not significant)
- Based on the changed information, the SPT is “modified” in order to reflect the changes

Shortest Path First Algorithm

Incremental SPF (i-SPF)



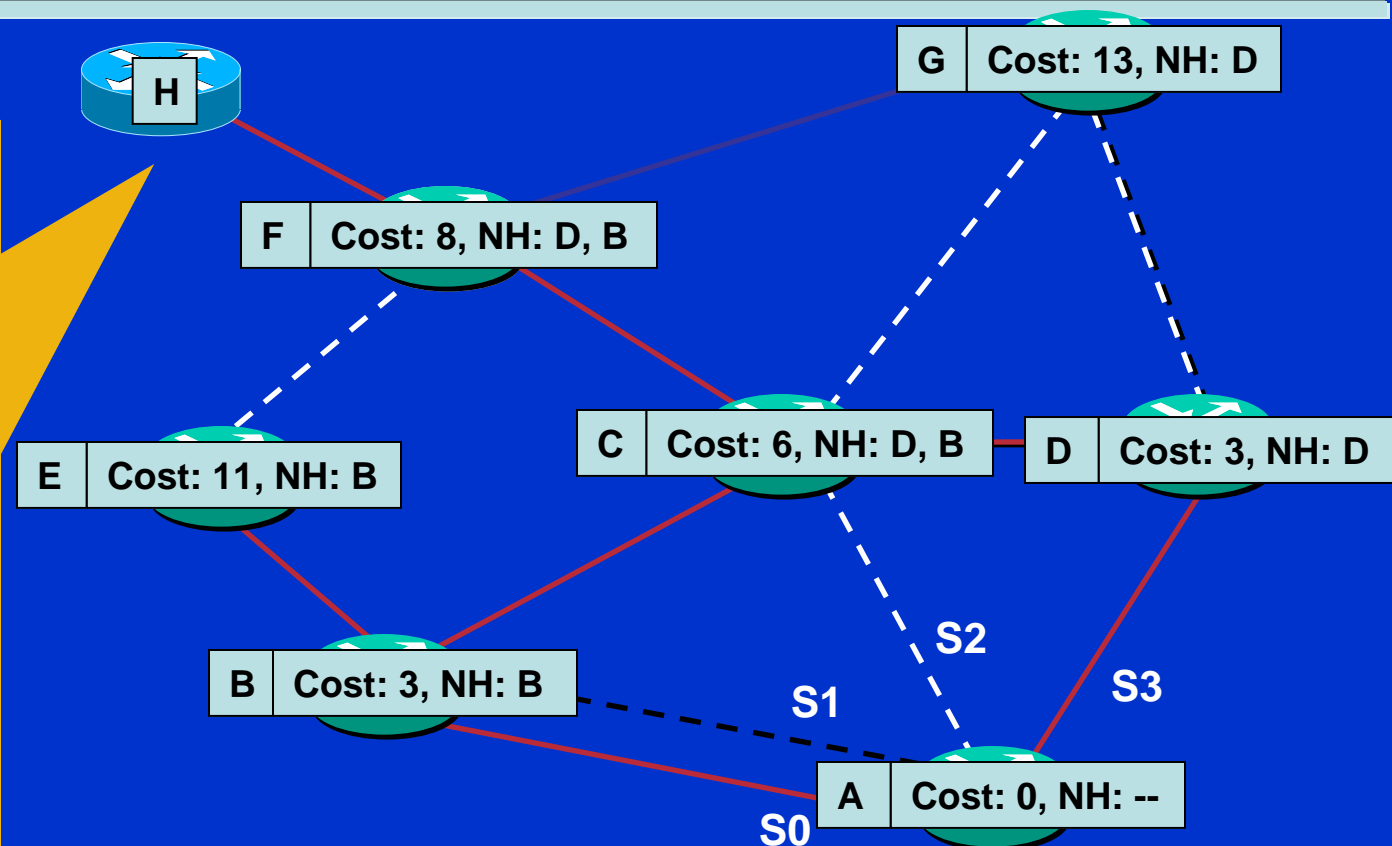
**C-G Link Is Down.
C-G Link Was Not
Used in SPT.
Therefore - No
Need to Run SPF.**

IS-IS Tutorial

Shortest Path First Algorithm

Incremental SPF (i-SPF)

F Reports a New Neighbor. The SPT Need Only be Extended Behind F. There Is No Need for Router A to Recompute the Whole SPT. Router A Will Compute SPF from Node F.



Shortest Path First Algorithm

Incremental SPF (i-SPF)

- The greater the distance from root that the change happens, the less time i-SPF will take to update the SPT
- If the change is close to the calculating node and affects the whole topology, there will be little or no gain in using i-SPF
- Time needed by i-SPF is not easily predictable
 - Tests demonstrated an average gain of more than 80%

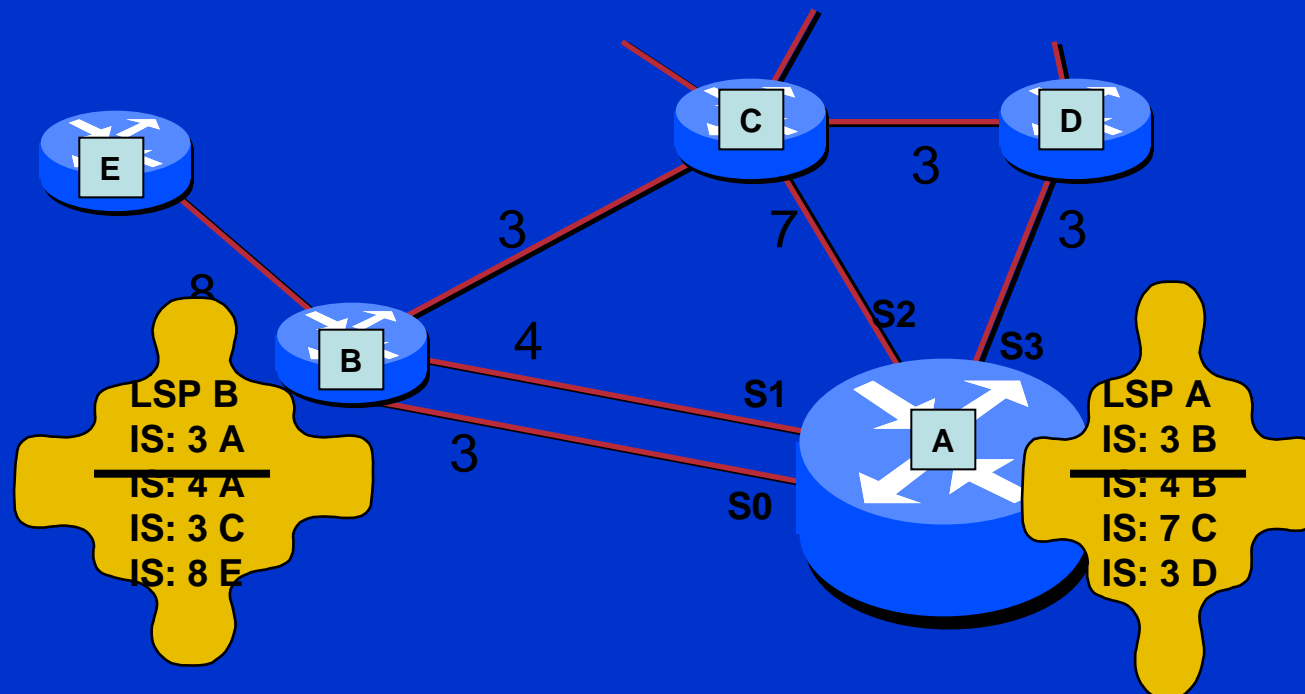
SPF, PRC, I-SPF: summary

- **Only a leaf change**
 - **PRC**
- **Graph impacted**
 - **normal-SPF: recompute the full SPT and hence reinserts all the ISIS routes in the RIB**
 - **I-SPF: only recomputes the part of the SPT that is affected. Only the leaves from that part are affected.**

Topology and Leaf Optimizations

IS-IS Tutorial

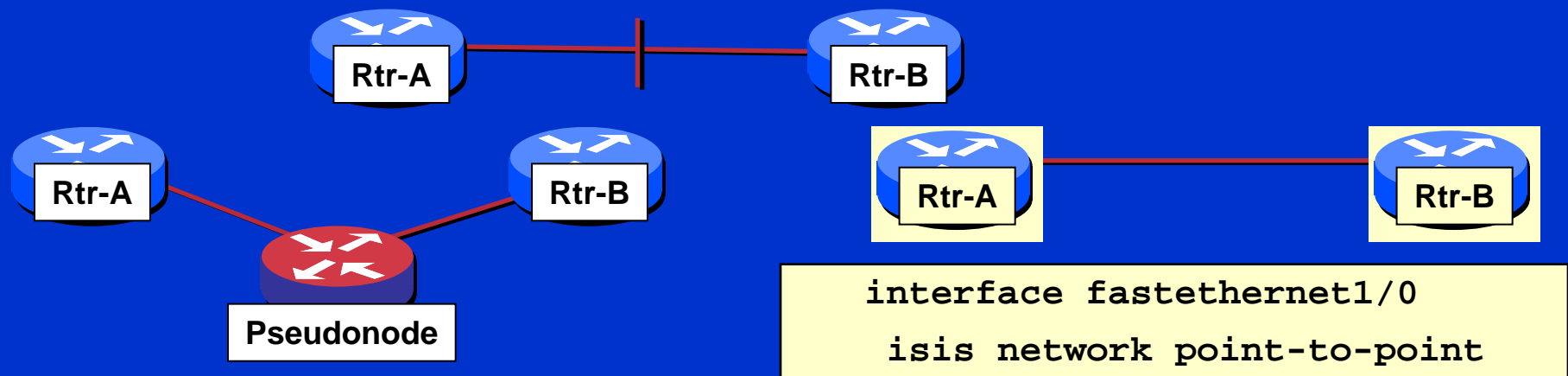
Parallel point-to-point adjacencies



- Only best parallel adjacency is reported

IS-IS Tutorial

P2P mode for back-to-back GE



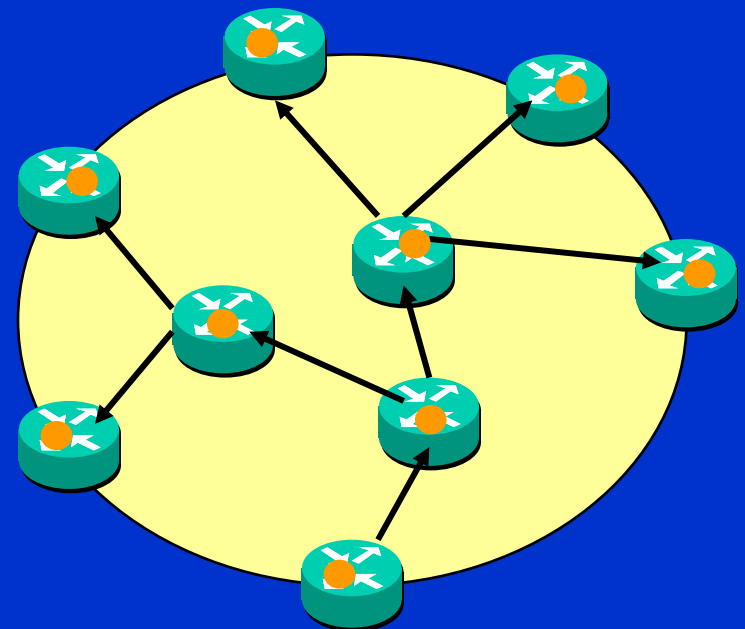
- No DIS election
- No CSNP transmission
- No Pseudo-node and extra link
- 12.0(22)S

IS-IS Tutorial

Speeding up route installation

- Limit the # of leaves in the IGP
 - only the BGP speakers are needed ()
 - rest: I-BGP
 - 12.0(22)S

```
router isis  
advertise passive-only
```



IS-IS Tutorial

Local RIB – Smart Ordering

- **Local_RIB updates the RIB according to this order:**
 1. **prefixes with high tag**
 2. **/32 prefixes without high tag**
 3. **Any other prefixes**

Tagging the “important” prefixes

On the interface with
an important ISIS prefix:

```
interface ../..  
  ip router isis  
  isis tag <XX>
```

On all ISIS routers:

```
router isis  
  ip fast high <XX>
```

- A usual target is ~ 400 important prefixes
- If the important prefixes are not /32 and/or if there are lots of /32's that are not important, then one should tag the important prefixes
- 12.0(22)S

IS-IS Tutorial

SPF, PRC and LSP-gen Exponential BackOff Timers

IS-IS Tutorial

Backoff timer algorithm

- IS-IS throttles its main events
 - SPF computation
 - PRC computation
 - LSP generation
- Throttling slows down convergence
- Not throttling can cause melt-downs
- The scope is to react fast to the first events but, under constant churn, slow down to avoid to collapse

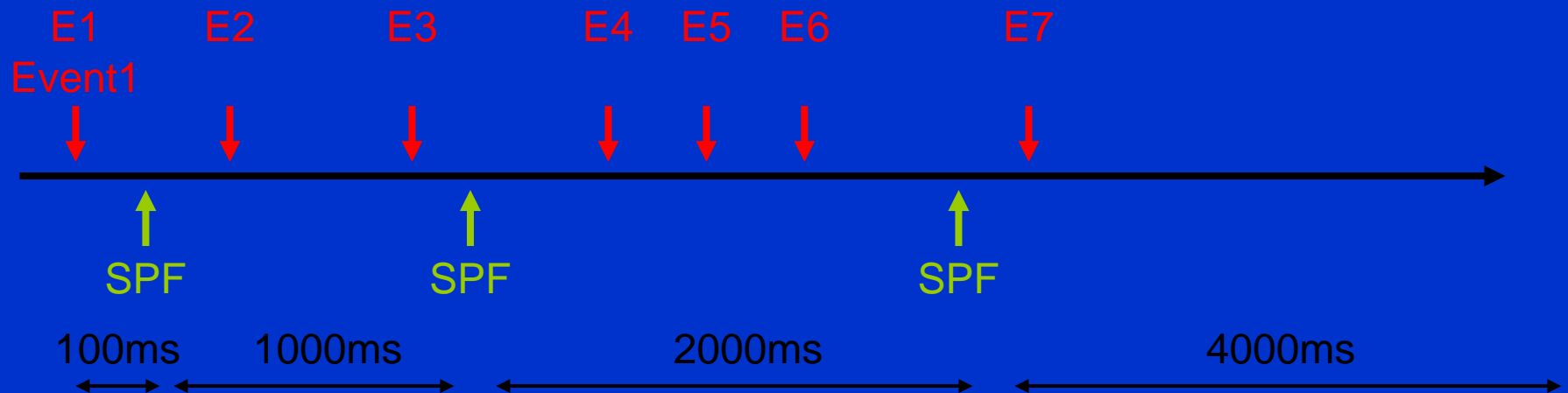
Backoff timer algorithm

```
spf-interval <Max> [<Init> <Inc>]
```

- **Maximum interval:** Maximum amount of time the router will wait between consecutive executions
- **Initial delay:** Time the router will wait before starting execution
- **Incremental interval:** Time the router will wait between consecutive execution. This timer is variable and will increase until it reaches **Maximum-interval**

IS-IS Tutorial

spf-interval 10 100 1000

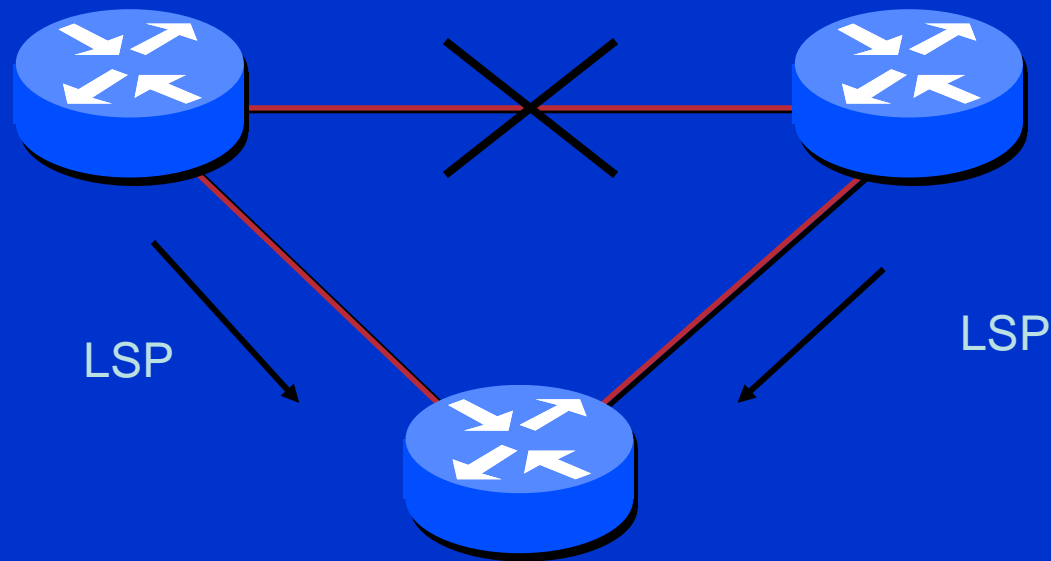


- Then 8000ms
- Then maxed at 10sec
- 20s without Trigger is required before resetting the SPF timer to 100ms

Default Values

- **Maximum-interval:**
 - SPF: 10 seconds
 - PRC: 5 seconds
 - LSP-Generation: 5 seconds
- **Initial-wait:**
 - SPF: 5.5 seconds
 - PRC: 2 seconds
 - LSP-Generation: 50 milliseconds
- **Incremental-interval:**
 - SPF: 5.5 seconds
 - PRC: 5 seconds
 - LSP-Generation: 5 seconds

Two-Way Connectivity Check



- For propagating Bad News, 1! LSP is enough

IS-IS Tutorial

Timers for Fast Convergence

```
router isis  
  spf-interval M I E  
  prc-interval M I E
```

- **I: link failure suggestion: 1ms**
 - 5.5 sec faster than default reaction!
 - Optimized for the going down mode
- **E: [10 msec to average SPF time]**
- **M: default should be fine in most cases**

LSP Pacing and Flooding

IS-IS Tutorial

LSP Pacing and Flooding

```
int pos x/x  
isis lsp-interval <>
```

- **Inter-LSP Pacing:**
 - Suggestion: keep default: 33msecs inter-LSP gap
 - backoff protection
 - full database download

```
router isis  
ip fast-convergence
```

- **Flooding**
 - flooding is done before SPF
- **Pacing**
 - the first 5 LSP's are no longer paced
- **Requires dv39355 (12.0(22)S)**

Link Protocol Properties

IS-IS Tutorial

Link Protocol Properties

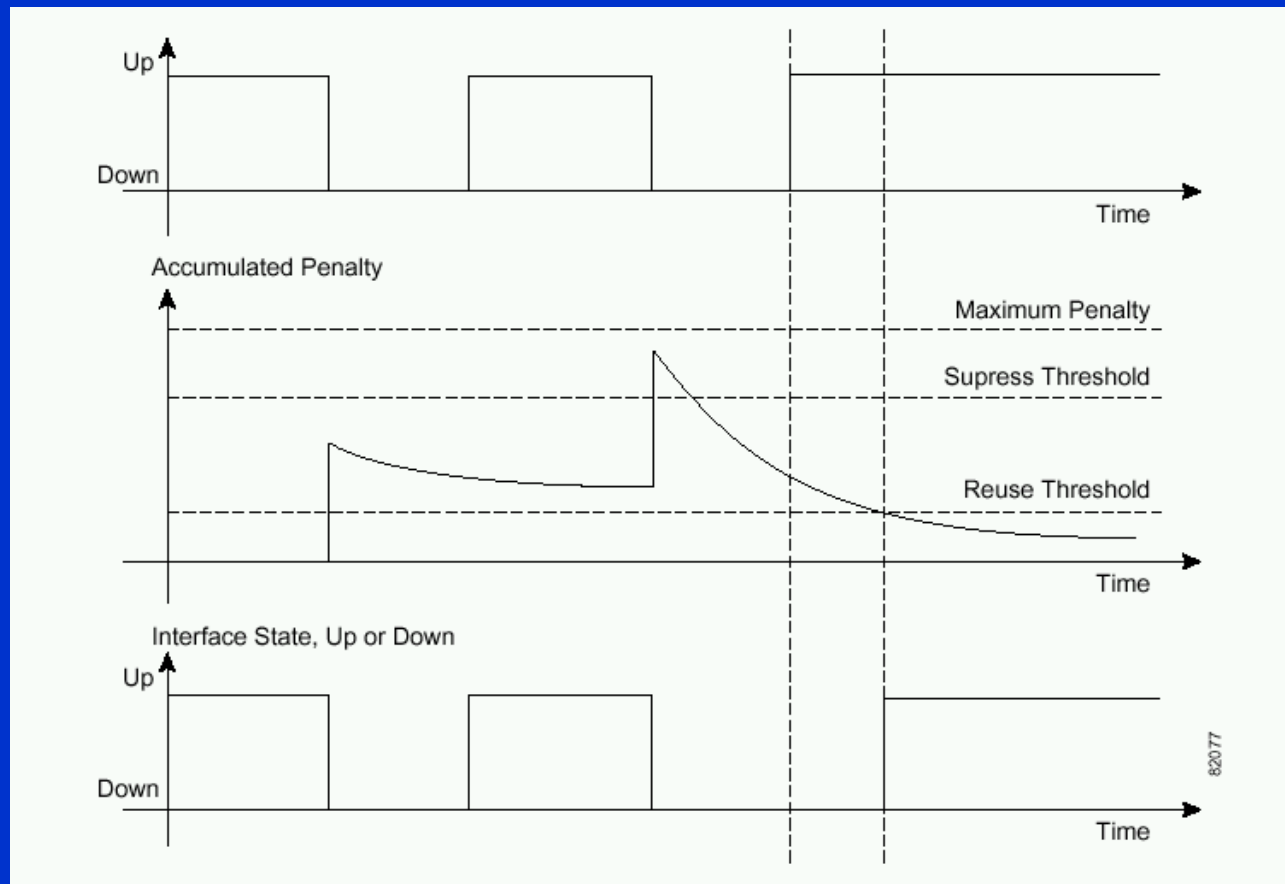
- **Link Failure Detection**
 - the faster and more reliable, the better
- **Dampening flapping links**
 - Fast signalling of a Down information
 - Stable signalling of an UP information
 - Freeze a flapping link in Down status

Dampening

```
int foo x/x  
dampening
```

- IP Event Dampening automatically identifies and selectively dampens a local interface that is flapping.
- Dampening an interface removes the interface from the network until the interface stops flapping and becomes stable.
- Improved convergence times and stability throughout net
 - isolation of failures so that disturbances are not propagated
 - reduction of CPU resource utilization
- 12.0(22)S

Dampening



- **configurable exponential decay mechanism**
- IS-IS Tutorial

```
int foo x/x  
  dampening[half reuse] [supp max [restart]]
```

- **half-life-period**
 - Time [1 to 30sec, def=15sec] after which a penalty is decreased by half
- **reuse-threshold**
 - When penalty < this value, the interface is unsuppressed. The range of the reuse value is from 1 to 2000; the default is 1000
- **suppress-threshold**
 - When penalty >= value: the intf is suppressed. The range is from 1 to 2000; the default is 2000
- **max-suppress-time**
 - Maximum time [1sec to 20000sec, def = 4*half-life] an intf can be suppressed.
- **restart-penalty**
 - penalty of the interface upon router reload [1 to 20000, def=2000]

POS – Detection of a link failure

- **Pos delay trigger line:**
 - hold time before reacting to a line failure
 - default is: immediate reaction
- **Pos delay trigger path:**
 - hold time before reacting to a path failure
 - default is: no reaction
- **Carrier-delay**
 - hold time between the end of the pos delay holdtime and the bring down of the IOS interface
 - default: 2000 msec

POS – Detection of a link failure

```
int pos 1/0  
  carrier-delay msec 0
```

- Redundant for POS interfaces
 - if PPP, you need dz45716 integrated in 12.0(24.1)S. Without this fix, configure carrier-delay to 20 msec

POS – Detection of a link failure

```
int pos 1/0  
  carrier-delay msec 0  
  pos delay triggers line 150  
  pos delay triggers path 150
```

R1-ADM--PROTECTED_SONET_net--ADM-R2

- Should delay a little to allow for SONET protection.
- Conservative Suggestion: 150msec

POS – Detection of a link failure

```
int pos 1/0
  carrier-delay msec 0
  pos delay triggers line 0
  pos delay triggers path 0
```

R1-ADM--UNprotected_SONET_net--ADM-R2

- **Should react as fast possible**
 - line default ok
 - path default not ok

POS – Detection of a link failure

```
int pos 1/0  
  carrier-delay msec 0  
  pos delay triggers line 150
```

R1-DWDM--PROTECTED_DWDM_net--DWDM-R2

- Should delay for DWDM protection
 - Conservative suggestion: 150msec
 - failure will be section or line

POS – Detection of a link failure

```
int pos 1/0  
  carrier-delay msec 0  
  pos delay triggers line 0
```

R1-DWDM--UNPROTECTED_DWDM_net--DWDM-R2

- Should react asap
 - line: default ok
 - path: not needed: default ok

POS – Bringing a down link back up

- Upon failure clearance, POS Driver will wait 10seconds + <Carrier-Delay> before turning the interface back up, hence before triggering ISIS convergence

POS – Best for Convergence

- **Very fast Link failure detection**
 - no need to tune the ISIS hello/holdtime
- **Native anti-flap property of POS**
 - down info is signalled very fast
 - up info is confirmed for 10s before relaying to interface
 - full dampening by enabling interface dampening

Other types of Links

- **Link Failure Detection**
 - If the native mode is too slow or if the link has no failure detection capability
 - ISIS Hello/Holdtime tuning

Fast Hello's

```
int serial0  
isis hello-interval minimal  
isis hello-multiplier 4
```

- Fast hello's allow a dead timer of 1 second
- POS much faster/reliable
- Only useful when layer1/2 can't help!

Operating this Design

IS-IS Tutorial

Monitoring link/adjacency status

```
int pos x/x
  pos ais-shut
  pos report lrldi
  pos report lais
  pos report prdi
  pos report pais

router isis
  log-adjacency-change all
```

- **ais-shut:** ensures that when you shut the interface, the router sends an AIS alarm
- **Reports:** enable sys-logging
- **Log-adjacency-change:** enable sys-logging

Before reloading a router

Config saved in NVRAM

```
router isis
```

```
    set-overload-bit on-startup <5-86400>
```

Just before reload:

```
router isis
```

```
    set-overload-bit
```

Reload and Do not save nvram

- Set overload bit makes sure no other router uses the local router for transit traffic

ISIS Monitor

- Router not used for normal traffic, but part of ISIS
 - debug isis spf-statistics
 - debug isis spf-triggers
 - sh proc cpu | include CPU utilization
 - sh proc cpu | include IS-IS
 - sh isis spf-log
- Allow to analyze statistics on
 - link/node failure
 - SPF duration and CPU consumption
 - extrapolate impact of link/node failure statistics on availability

IS-IS Tutorial

Conclusion

- **IGP convergence needs to be optimized for Tight-SLA Services**
- **New development speed up convergence without stability compromise**
- **Test results indicate that convergence in 1-to-2 second is realistic**
- **For sub-100ms Convergence, local action based on precomputed tables is required**