23 Febbraio 2024

La Galassia del Routing IP

Il cuore dell'Internet

II^ puntata – Routing *Link State*

Tiziano Tofoni

Note di *Copyright*

- Questo insieme di diapositive è protetto dalle leggi sul *copyright* e dalle disposizioni dei trattati internazionali. Il titolo ed i *copyright* relativi alle diapositive (ivi inclusi, ma non limitatamente, ogni immagine, fotografia, animazione, video, audio, musica e testo), in accordo con gli artt. 12 e seguenti della Legge 633/1941, sono di proprietà dell'autore Tiziano Tofoni (di seguito 'l'autore').

- Le diapositive possono essere utilizzate esclusivamente per scopi di studio nell'ambito dei corsi tenuti dall'autore.

- Ogni altra utilizzazione o riproduzione (ivi incluse, ma non limitatamente, le riproduzioni su supporti ottici/magnetici, su reti di calcolatori o stampate) in toto o in parte è vietata, se non esplicitamente autorizzata per iscritto, a priori, da parte dell'autore.

- L'informazione contenuta in queste diapositive è ritenuta essere accurata alla data della pubblicazione. Essa è fornita per scopi meramente didattici e non per essere utilizzata in progetti di impianti, prodotti, reti, ecc. In ogni caso essa è soggetta a cambiamenti senza preavviso. L'autore non si assume alcuna responsabilità per il contenuto di queste diapositive (ivi incluse, ma non limitatamente, la correttezza, completezza, applicabilità, aggiornamento dell'informazione).

- In ogni caso non può essere dichiarata conformità all'informazione contenuta in queste diapositive.

- In ogni caso questa nota di *copyright* non deve mai essere rimossa e deve essere riportata anche in utilizzi parziali.

Di cosa parlerò …

#1 Gli "ingredienti" fondamentali

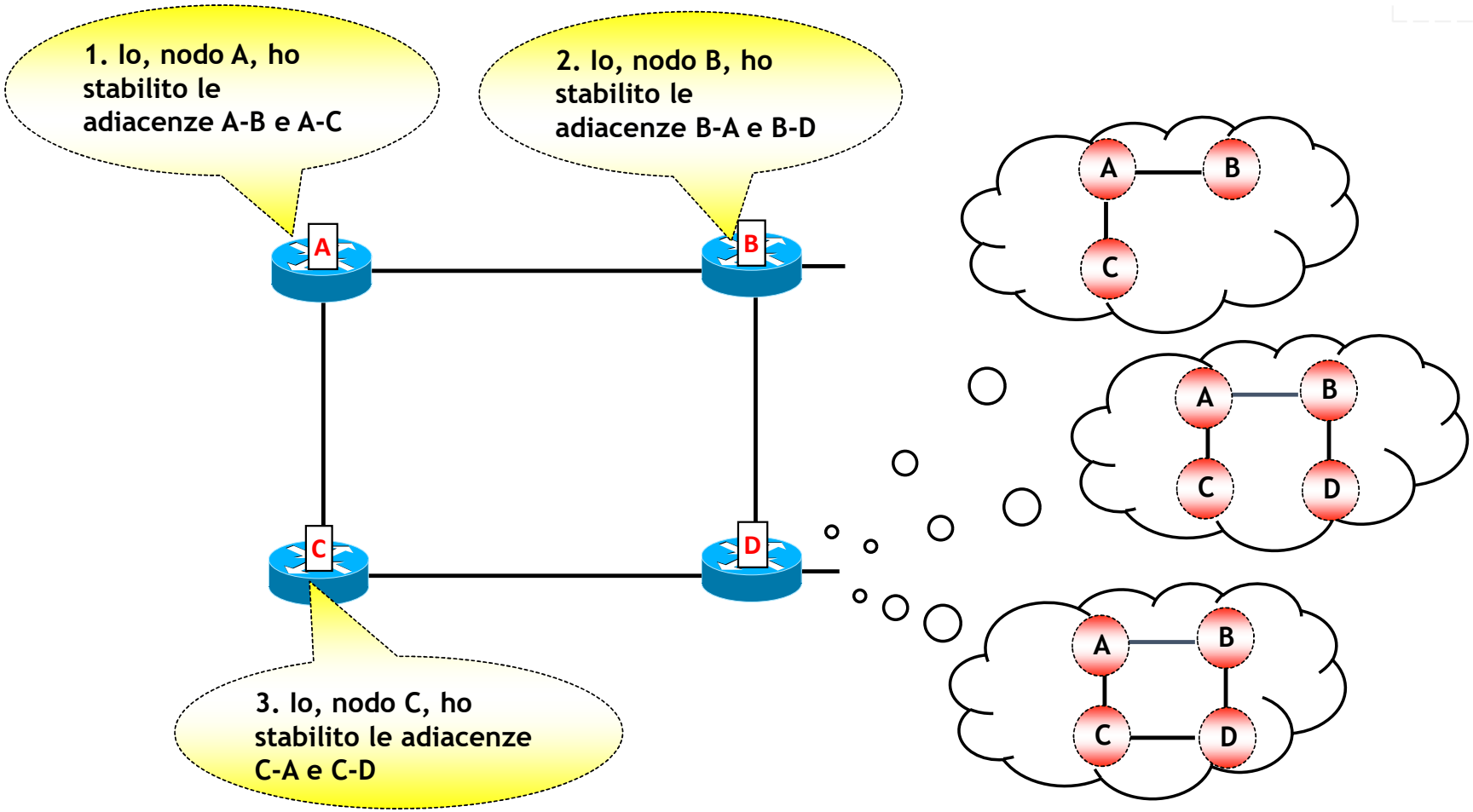
#2 Il *Link State DataBase* (LSDB)

#3 L'algoritmo per la ricerca dei percorsi ottimi

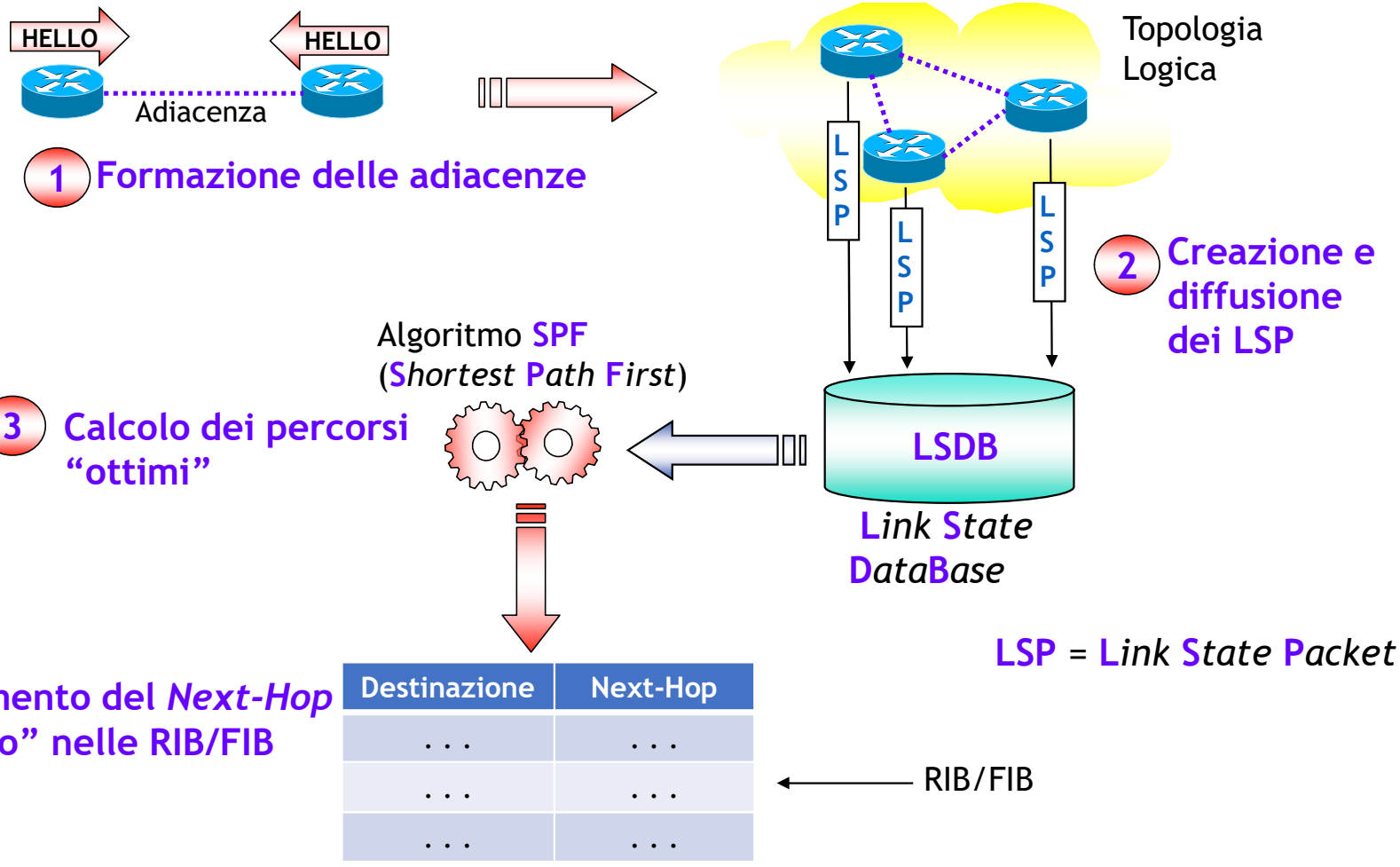
#4 Aspetti generali di convergenza

#5 I protocolli *Link State*

Principio del *Link State*

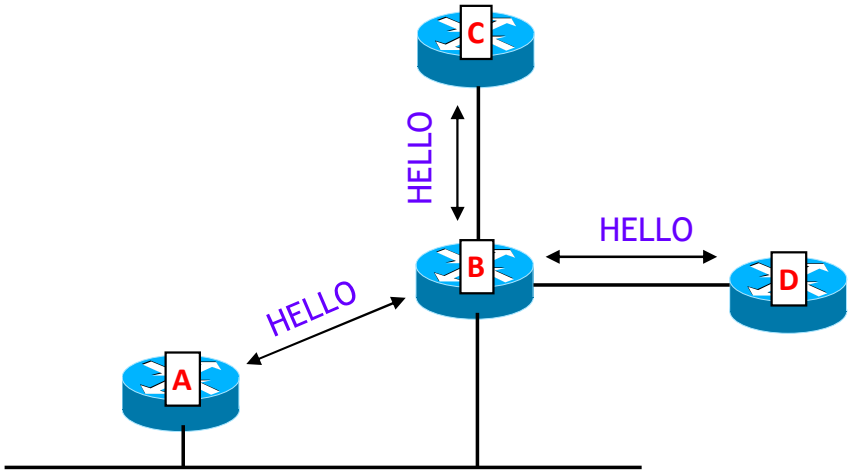


Gli "ingredienti" fondamentali



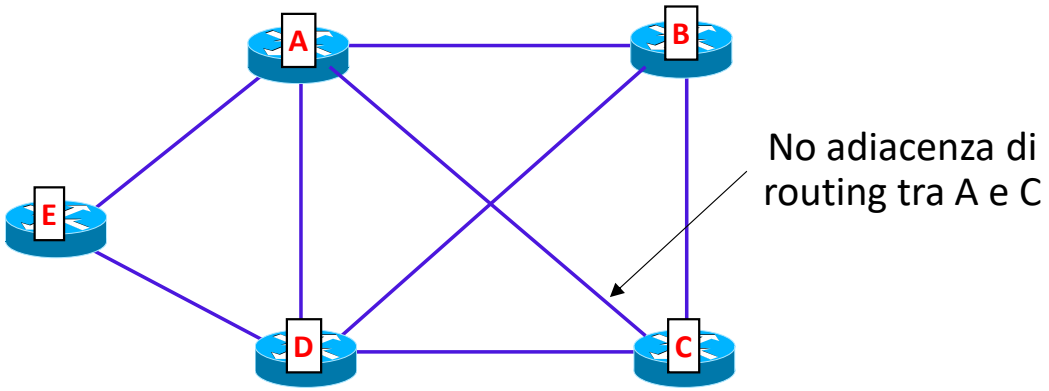
Neighbor Greetings (Hello)

- Ogni router impara il suo ambito locale (*link* e nodi adiacenti) per mezzo di un meccanismo di *Neighbor Greetings*
  - Basato sull'invio periodico di messaggi HELLO
  - Periodicità elevata per il riconoscimento delle variazioni sulle adiacenze in tempi ragionevoli
- Due router che si scambiano regolarmente messaggi HELLO formano una *neighborship*
  - Nota: in funzione del protocollo adottato potrebbero formare una adiacenza di routing

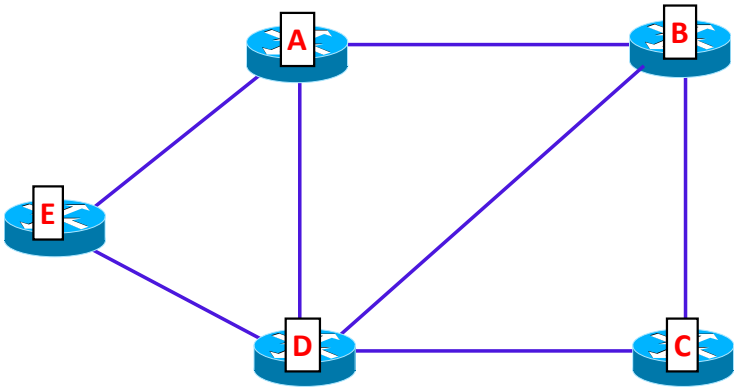


Topologia fisica e topologia logica

- Topologia fisica: è l'insieme dei router e dei segmenti di rete (tipicamente *broadcast* o punto-punto) che li interconnettono
- Topologia logica: è l'insieme dei router e delle adiacenze di routing che li interconnettono
- NOTA: i protocolli di routing *Link State* determinano i percorsi ottimi sulla base della topologia logica



Topologia fisica

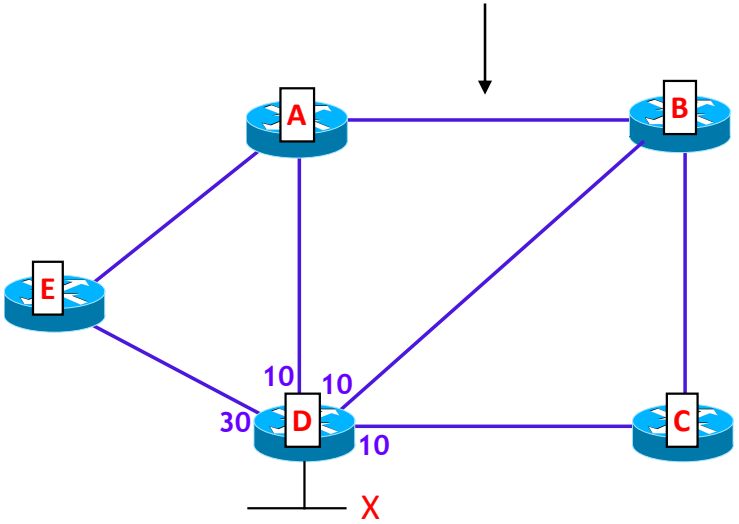


Topologia logica

I *Link State Packet*

- Messaggi utilizzati per annunciare la topologia logica locale
  - Generati indipendentemente da ogni router
  - Ogni router inserisce l'elenco dei nodi adiacenti, la metrica per raggiungere il nodo adiacente e le reti IP direttamente connesse

Topologia logica

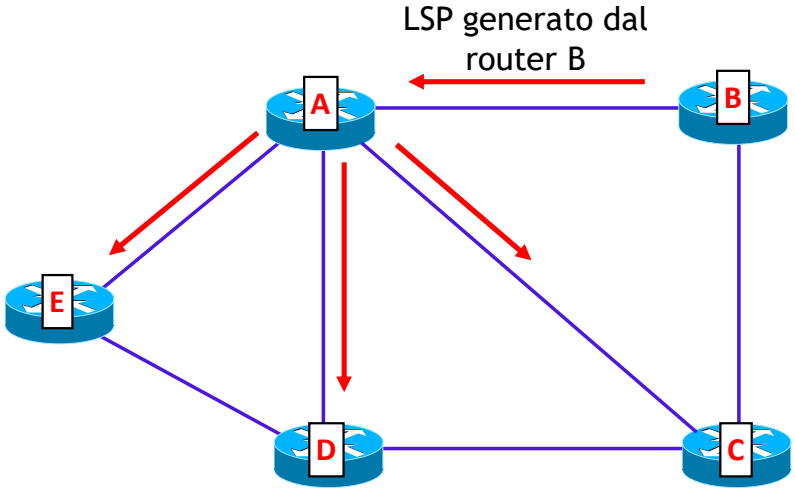


LSP del router D

Prefissi D.C.	X
Adiacenze	
Neighbor	Metrica
A	10
B	10
C	10
E	30

Diffusione dei *Link State Packet*

- I LSP sono diffusi al resto della rete tramite un meccanismo di *Selective Flooding*
  - Ogni router che riceve un LSP lo invia a tutti i router adiacenti ad esclusione di quello dal quale lo ha ricevuto
- Meccanismo semplice, efficace, ma poco efficiente
  - Ogni router può ricevere più copie dello stesso LSP
  - In una rete di N router completamente magliata, ogni router riceve N-1 copie dello stesso LSP

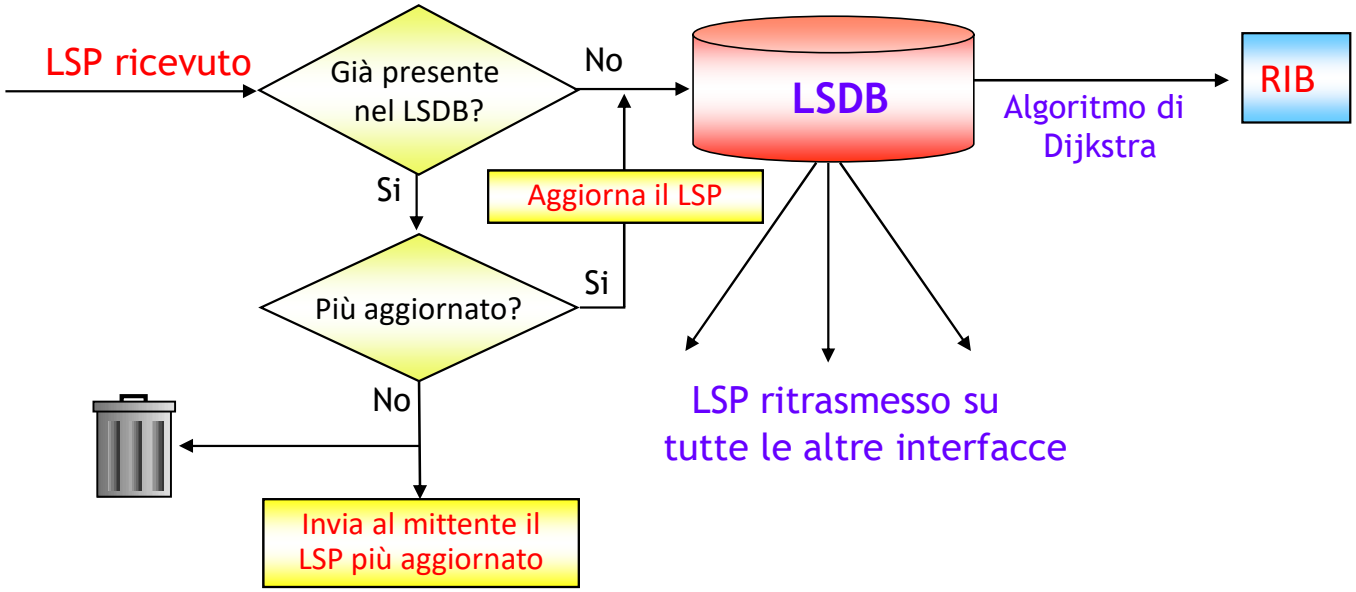


Di cosa parlerò …

- #1 Gli "ingredienti" fondamentali
- #2 Il *Link State DataBase* (LSDB)
- #3 L'algoritmo per la ricerca dei percorsi ottimi
- #4 Aspetti generali di convergenza
- #5 I protocolli *Link State*

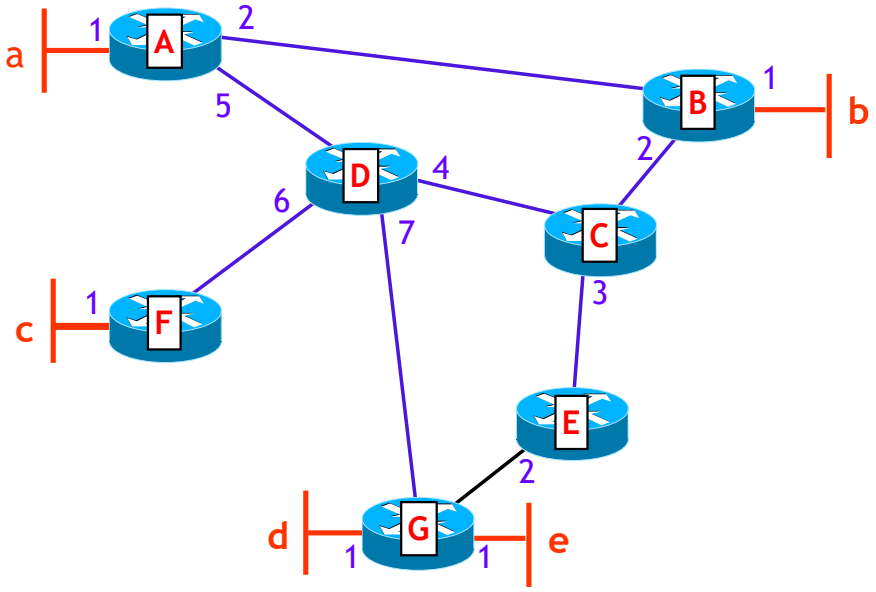
Il *Link State DataBase*

- I router possiedono un archivio di LSP (*LSDB*, *Link State DataBase*)
- Un router che riceve un LSP controlla l'archivio
  - Se il LSP ricevuto non è presente nel LSDB viene aggiunto
  - Se presente nel LSDB e il LSP ricevuto è più recente di quello già presente, lo sostituisce, altrimenti lo scarta e invia al mittente la copia aggiornata



Il *Link State DataBase*

- I LSP memorizzati formano una mappa completa della rete
- Principio fondamentale dei protocolli *Link State*: tutti i router della topologia logica devono avere i LSDB sincronizzati
  - Nel routing gerarchico in realtà solo i router di un'area devono avere i LSDB sincronizzati



LSDB

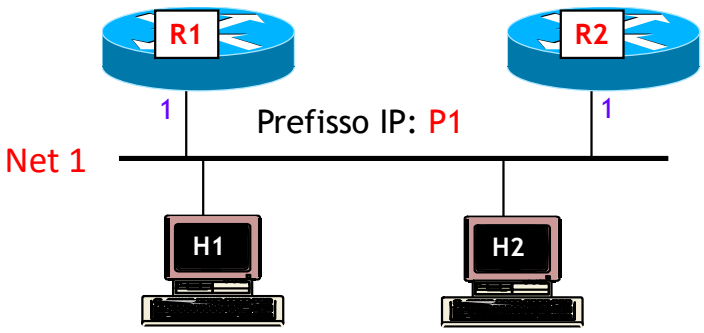
A	B/2 D/5 a/1
B	A/2 C/2 b/1
C	B/2 D/4 E/3
D	A/5 C/4 F/6 G/7
E	C/3 G/2
F	D/6 c/1
G	D/7 E/2 d/1 e/1

(identico su ogni router)

X/n = Nodo/Metrica

LSDB e reti periferiche

- Reti periferiche (*stub network*)
  - Ospitano *End System* (*Host*)
    - Normalmente non hanno capacità di routing
    - Non generano LSP
- Nel LSDB possono essere sostituite con un elemento unico valido per tutta la rete



LSDB ideale

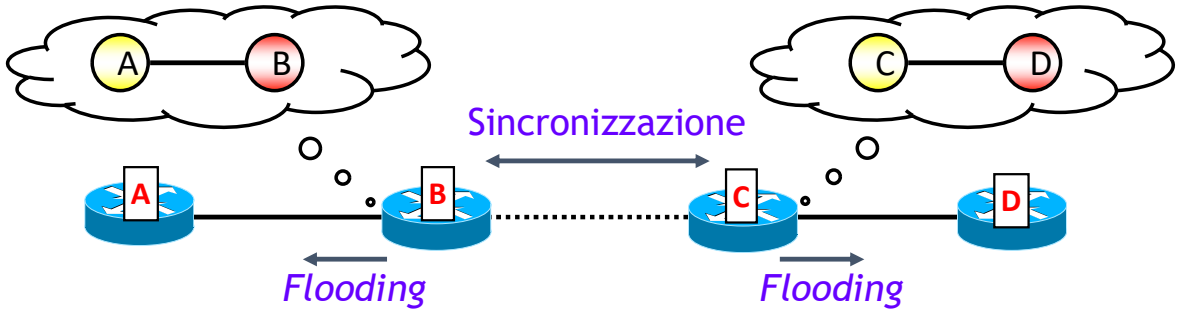
R1	R2/1	H1/1	H2/1
R2	R1/1	H1/1	H2/1
H1	R1/1	R2/1	H2/1
H2	R1/1	R2/1	H1/1

LSDB reale

R1	R2/1	P1/1
R2	R1/1	P1/1

Sincronizzazione dei LSDB

- Utilizzata per
  - Popolare immediatamente il LSDB di un router appena acceso
  - Allineare i LSDB dei router in caso di partizionamento della rete
    - Ambedue i router impareranno qualcosa l'uno dall'altro
- Procedura
  - Attivazione della procedura di *Neighbor Greetings*
  - Se viene rilevata una nuova adiacenza, inizia una fase di sincronizzazione dei rispettivi LSDB
    - La sincronizzazione viene ripetuta per ogni adiacenza
    - I LSP non conosciuti verranno anche inviati in *flooding* agli altri router della rete

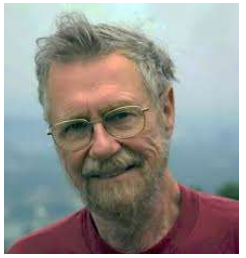


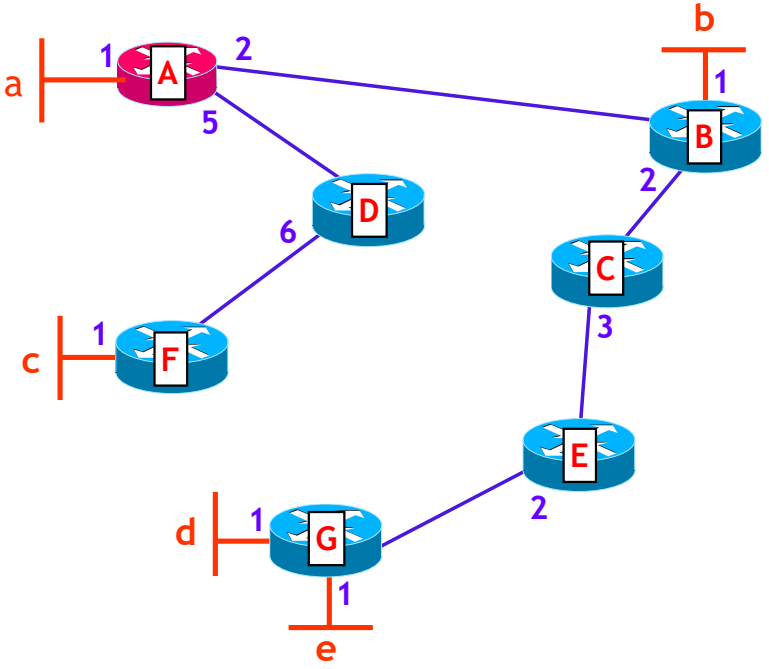
Di cosa parlerò …

- #1 Gli "ingredienti" fondamentali
- #2 Il *Link State DataBase* (LSDB)
- #3 L’algoritmo per la ricerca dei percorsi ottimi
- #4 Aspetti generali di convergenza
- #5 I protocolli *Link State*

L’algoritmo di Dijkstra

- Ogni router calcola indipendentemente i suoi percorsi ottimali verso tutte le destinazioni conosciute, applicando alla topologia logica della rete l'algoritmo di Dijkstra o SPF (*Shortest Path First*)





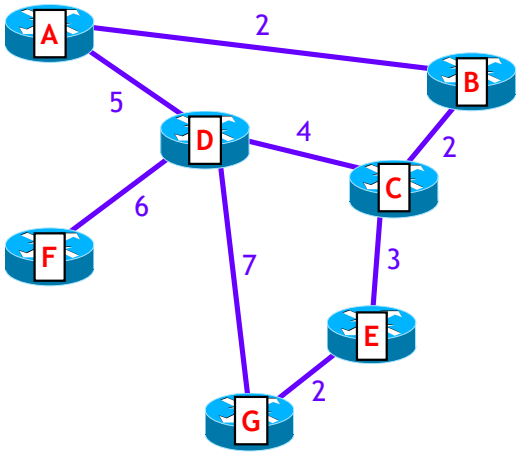
Dest.	Next-Hop	Costo
a	diretto	1
b	B	3
c	D	12
d	B	10
e	B	10
B	diretto	2
C	B	4
D	diretto	5
E	B	7
F	D	11
G	B	9

RIB di A

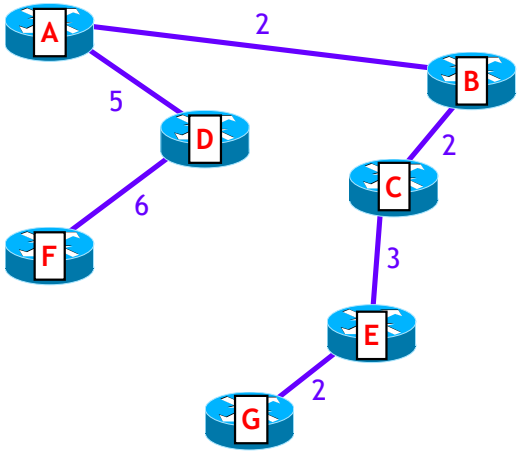
Ottimizzazioni dell'algoritmo di Dijkstra

- *Incremental SPF* (*iSPF*)
  - Metodo che consente di evitare il ricalcolo dell'albero SPF quando la variazione della topologia non interessa l'albero SPF originale

- *Partial Route Calculation* (*PRC*)
  - Metodo che consente di evitare il ricalcolo dell'albero SPF quando la topologia logica rimane invariata

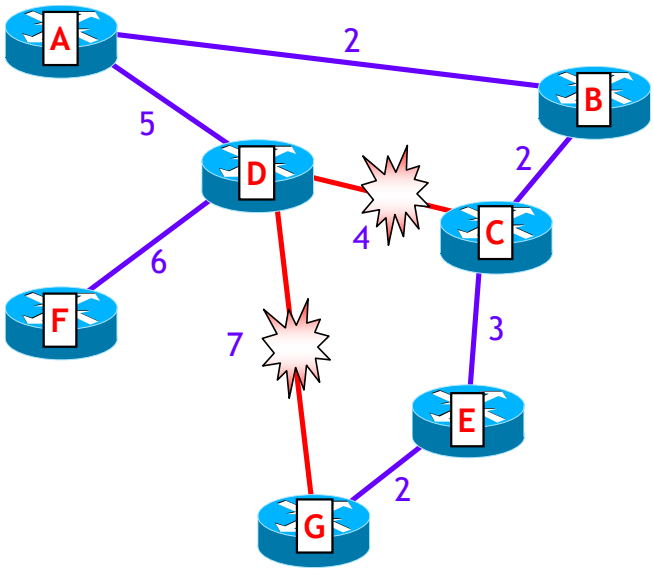


Topologia logica originale

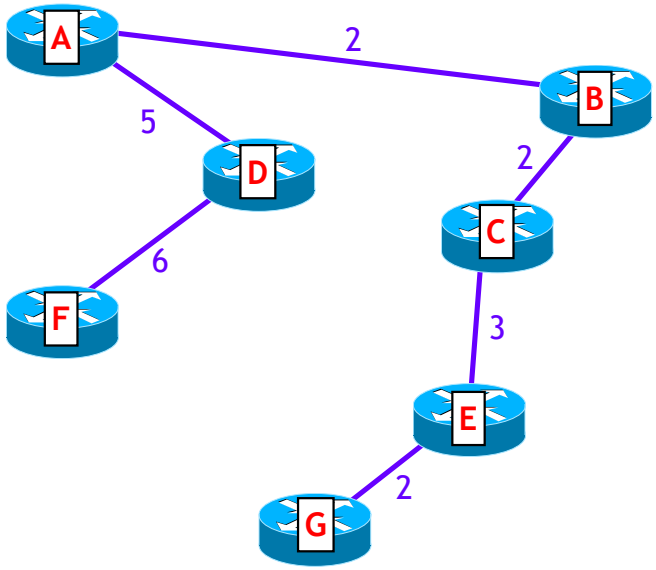


Albero SPF determinato dal router A

Incremental SPF (1/3)



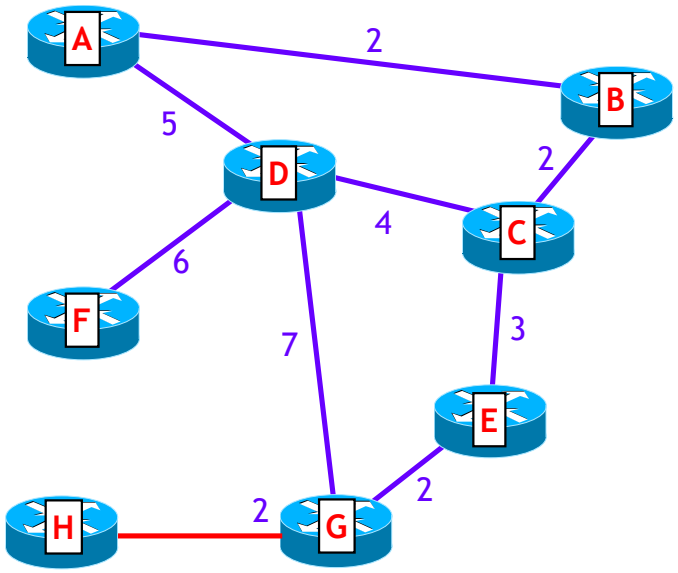
Topologia logica originale



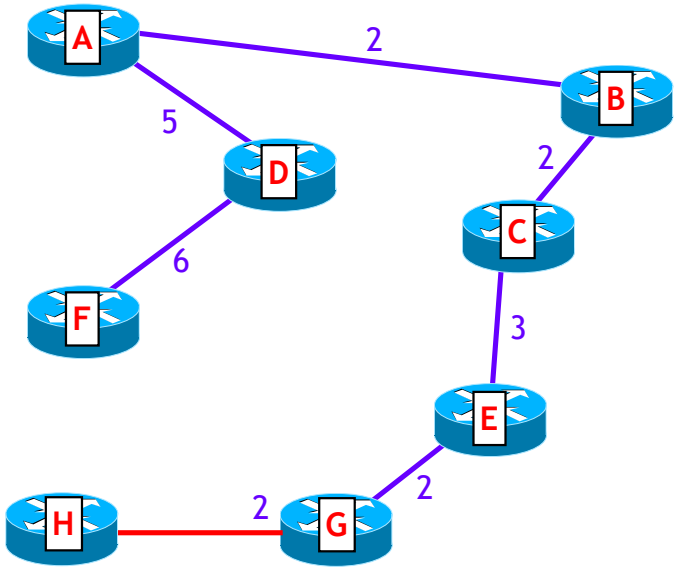
Nuovo albero SPF
determinato dal nodo A

- Osservazione fondamentale: la perdita delle adiacenze DC e/o DG non altera l'albero SPF!
  - Conseguenza: nel caso di perdita delle adiacenze D-C e/o D-G il ricalcolo dell'albero SPF non è necessario

Incremental SPF (2/3)



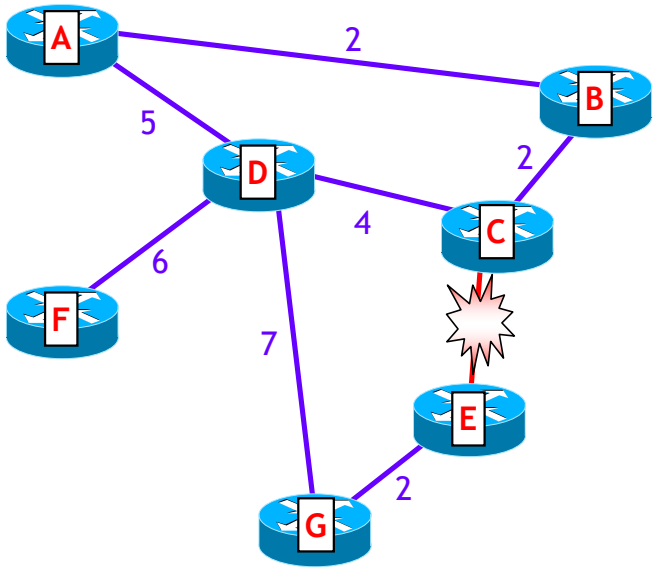
Topologia logica originale
dopo l'aggiunta del nodo H



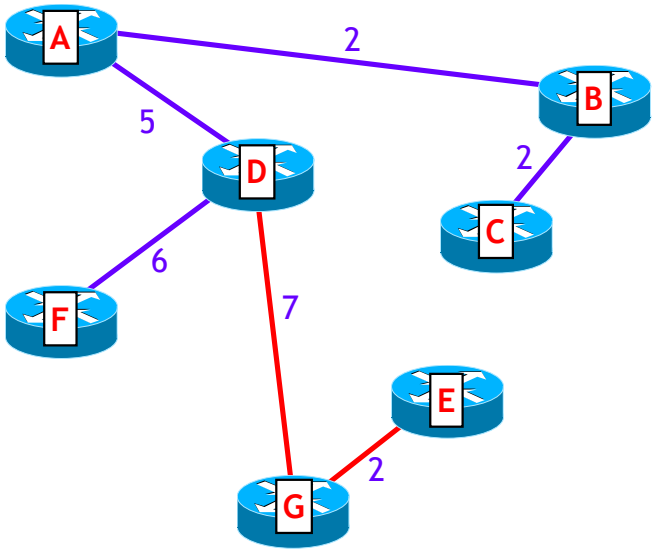
Albero SPF dopo
l'aggiunta del nodo H

- Nel caso in cui alla topologia originale venga aggiunto un nodo "foglia", l'albero SPF è semplicemente "esteso" con l'aggiunta del nuovo nodo foglia

Incremental SPF (3/3)



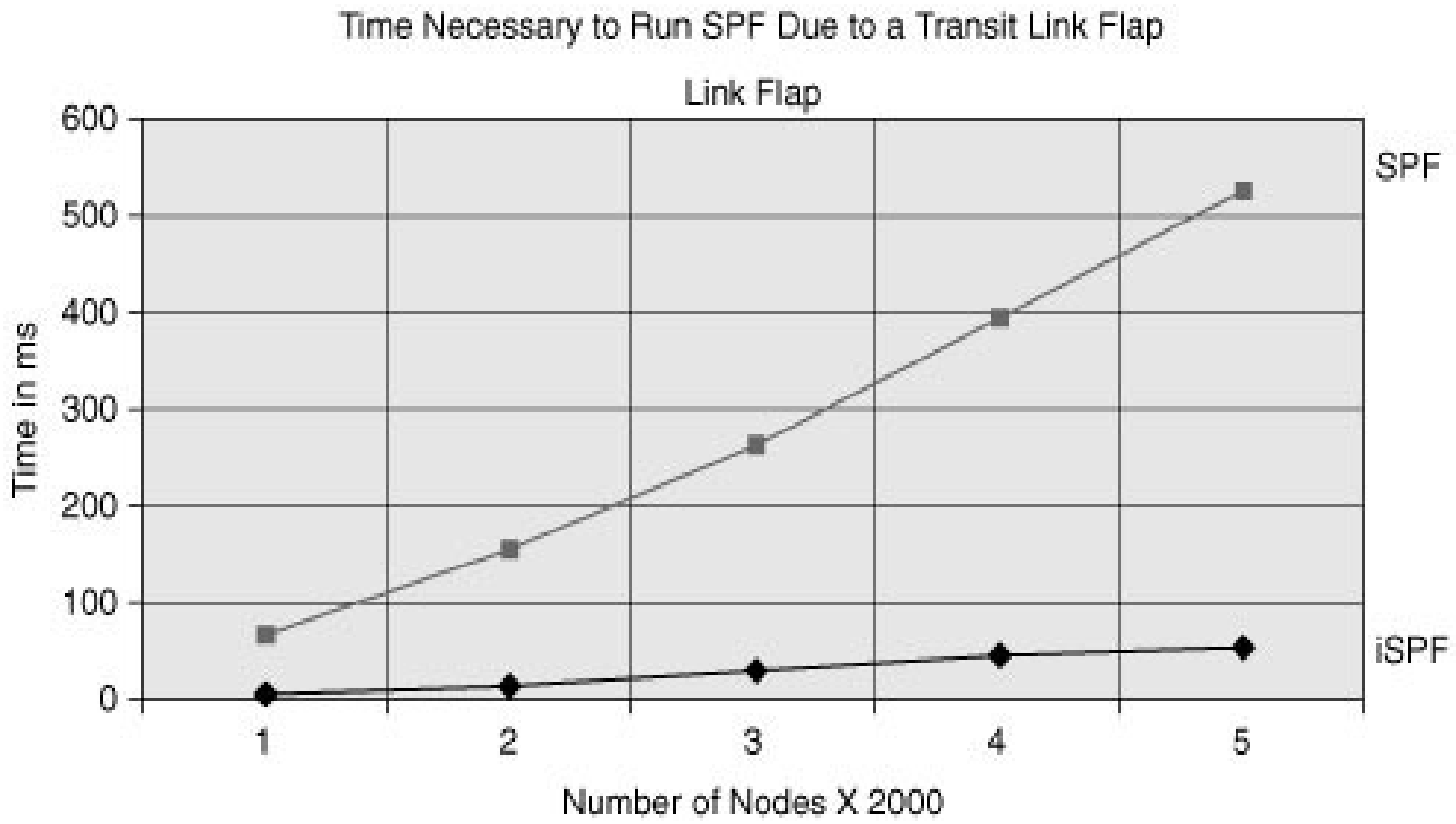
Topologia logica originale



Nuovo albero SPF
determinato dal nodo A

- Nel caso di perdita di una adiacenza che è parte dell'albero SPF originale, è sufficiente il ricalcolo dell'albero SPF verso i soli router a valle dell'adiacenza persa

SPF vs Incremental SPF



Fonte: http://ciscodocuments.blogspot.it/2011/06/chapter-3-developing-optimum-design-for_405.html

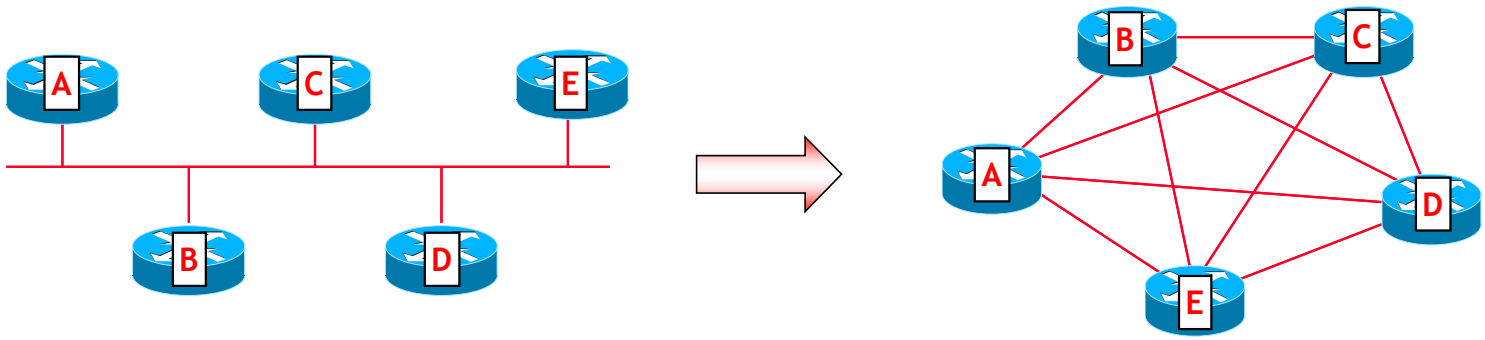
Partial Route Calculation

- Per determinare un percorso verso una *stub network* può essere utilizzata una *Partial Route Calculation* (*PRC*)

- Caratteristica fondamentale
  - Non è richiesto un ricalcolo dell'intero albero SPF, ma solo l'individuazione del router dove è attestata la *stub network*
  - L'aggiunta di nuove *stub network* o il fuori servizio di qualcuna di queste non richiede il ricalcolo dell'albero SPF

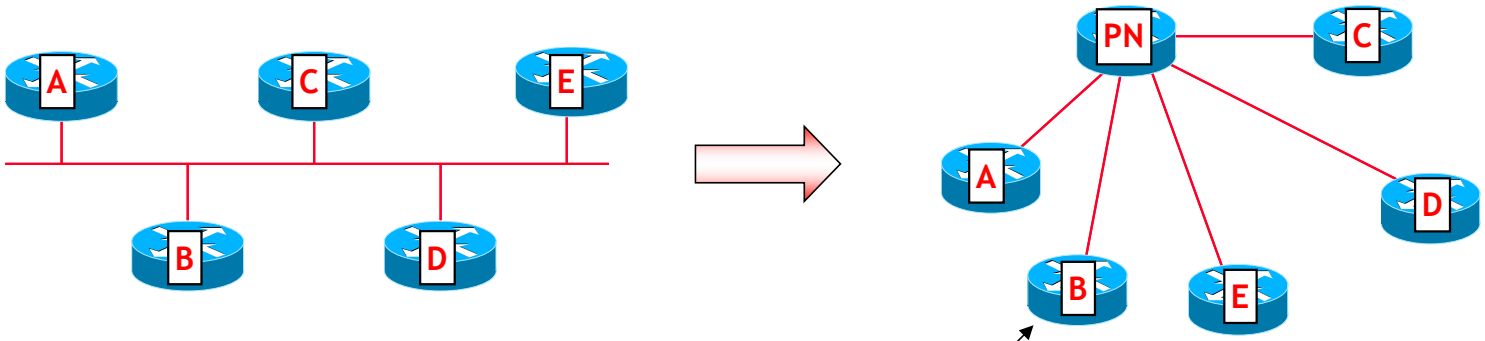
Rappresentazione delle reti *broadcast* (1/3)

- *N* nodi su rete *broadcast* = *N*(*N*-1)/2 *link* logici (adiacenze)
  - I LSP diventano molto grandi: ogni router avrebbe decine di adiacenze
  - Complessità dell'algoritmo di Dijkstra esplode (proporzionale al numero di *adiacenze*)



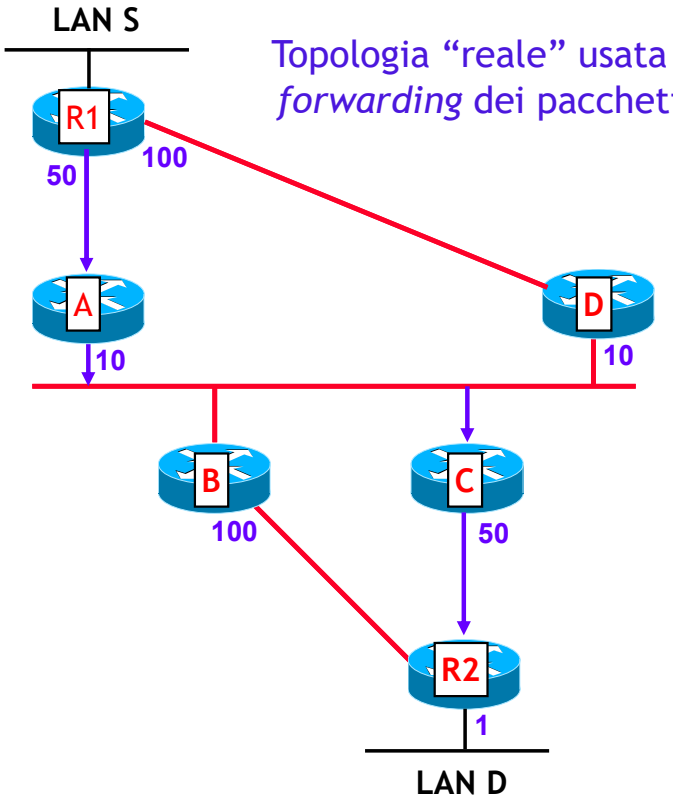
Rappresentazione delle reti *broadcast* (2/3)

- Soluzione: passare da una topologia logica completamente magliata a una a stella
- Centro Stella = Pseudo-Nodo (PN)
  - Nodo fittizio che trasforma la topologia equivalente da maglia completa a stella

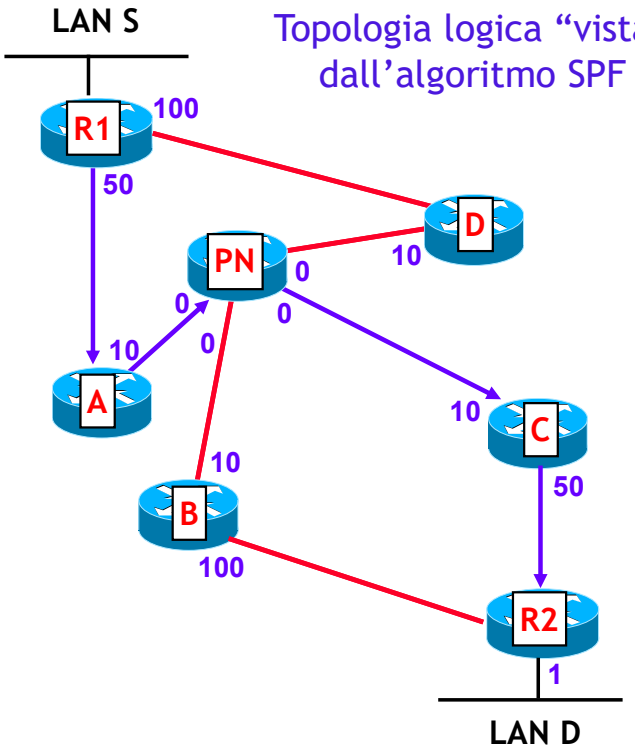


Topologia "virtuale" usata solo per l'algoritmo *Link State*; il routing dei pacchetti avviene in modalità classica

Rappresentazione delle reti *broadcast* (3/3)



Topologia "reale" usata il *forwarding* dei pacchetti



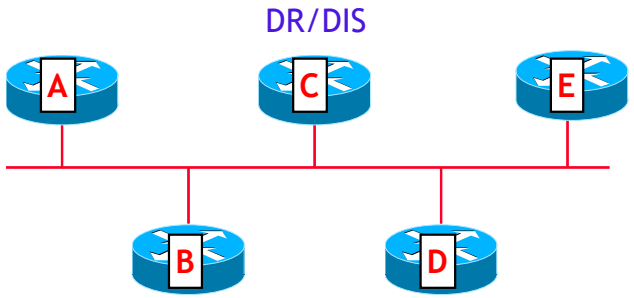
Topologia logica "vista" dall'algoritmo SPF

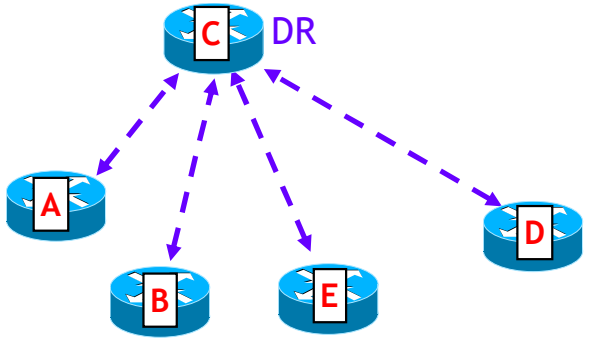
Costo ottimo LAN S → LAN D =
50 + 10 + 50 + 1 = 111

Costo ottimo LAN S → LAN D =
50 + 10 + 0 + 50 + 1 = 111

Il *Designated Router* (DR)

- Nei segmenti *broadcast* è troppo oneroso che ogni router sincronizzi il suo LSDB con quello di tutti gli altri router del segmento
  - Soluzione: definire un *Designated Router* (DR)
- Il DR è un router del segmento *broadcast* che ha due compiti principali
  - Annunciare agli altri router del dominio di routing la presenza del segmento *broadcast*
  - Mantenere sincronizzati tutti i LSDB dei router del segmento *broadcast*
- L'elezione del DR avviene con una procedura che dipende dal protocollo utilizzato
  - NOTA: nel protocollo IS-IS il DR viene indicato come DIS (*Designate Intermediate System*)

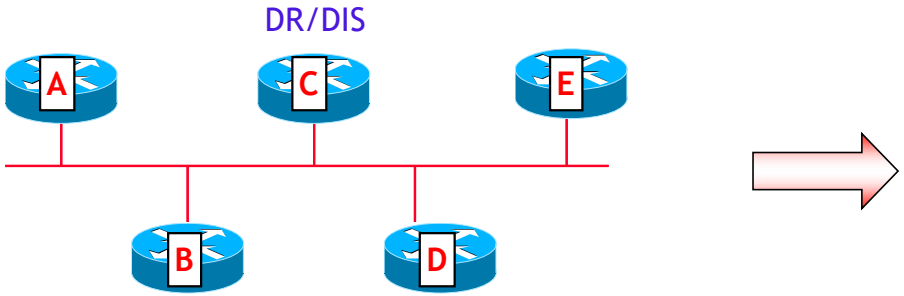


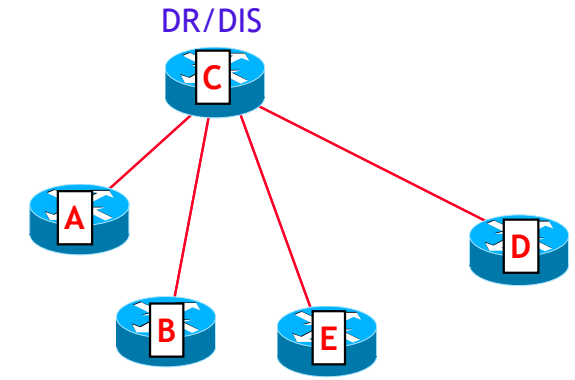


NOTA: DR/DIS ≠ PN

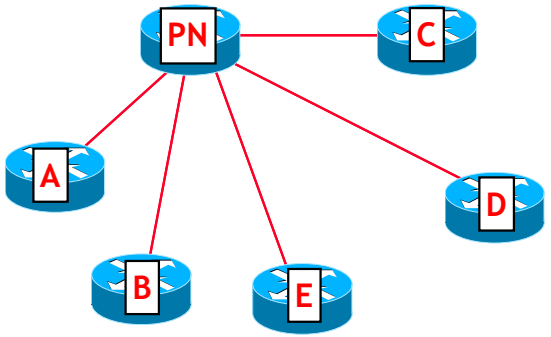
Tipi di adiacenze

- Due tipi di adiacenze
  - Adiacenze topologiche: sono stabilite tra ciascun router del segmento *broadcast* e lo Pseudo-Nodo (PN)
  - Adiacenze di sincronizzazione: sono stabilite tra ciascun router del segmento *broadcast* e il DR/DIS





Adiacenze di sincronizzazione

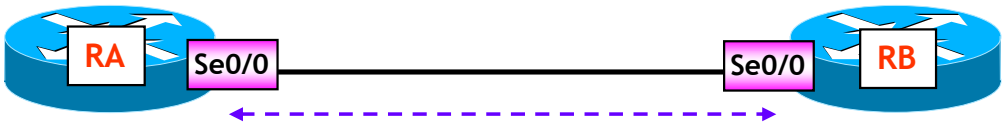


Adiacenze topologiche

NOTA: DR/DIS ≠ PN

Rappresentazione delle reti punto-punto

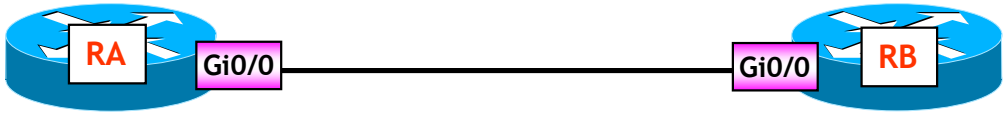
- Tipicamente si ha tra router collegati a Livello 2 attraverso protocolli punto-punto (es. PPP, HDLC Cisco, PVC ATM o FR)
- Nei segmenti di rete punto-punto non vi è bisogno di un DR
- Le adiacenze di sincronizzazione e topologiche coincidono



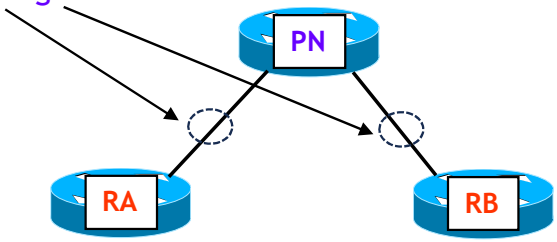
Adiacenza
(di sincronizzazione e topologica)

Rappresentazione delle reti punto-punto Ethernet

- Quando il collegamento punto-punto è realizzato con un collegamento Ethernet *back-to-back*, è buona regola disabilitare l'elezione del *DR*
  - L'elezione del *DR* comporta un inutile spreco di CPU e aumenta il numero di *link* visti dall'algoritmo SPF da 1 a 2
  - Idea descritta nella RFC 5309 - *Point-to-Point Operation over LAN in Link State Routing Protocols*

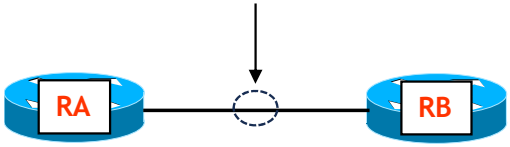


Adiacenze topologiche



Topologia vista dall'algoritmo SPF
senza disabilitare l'elezione del *DR*

Adiacenza topologica e di sincronizzazione

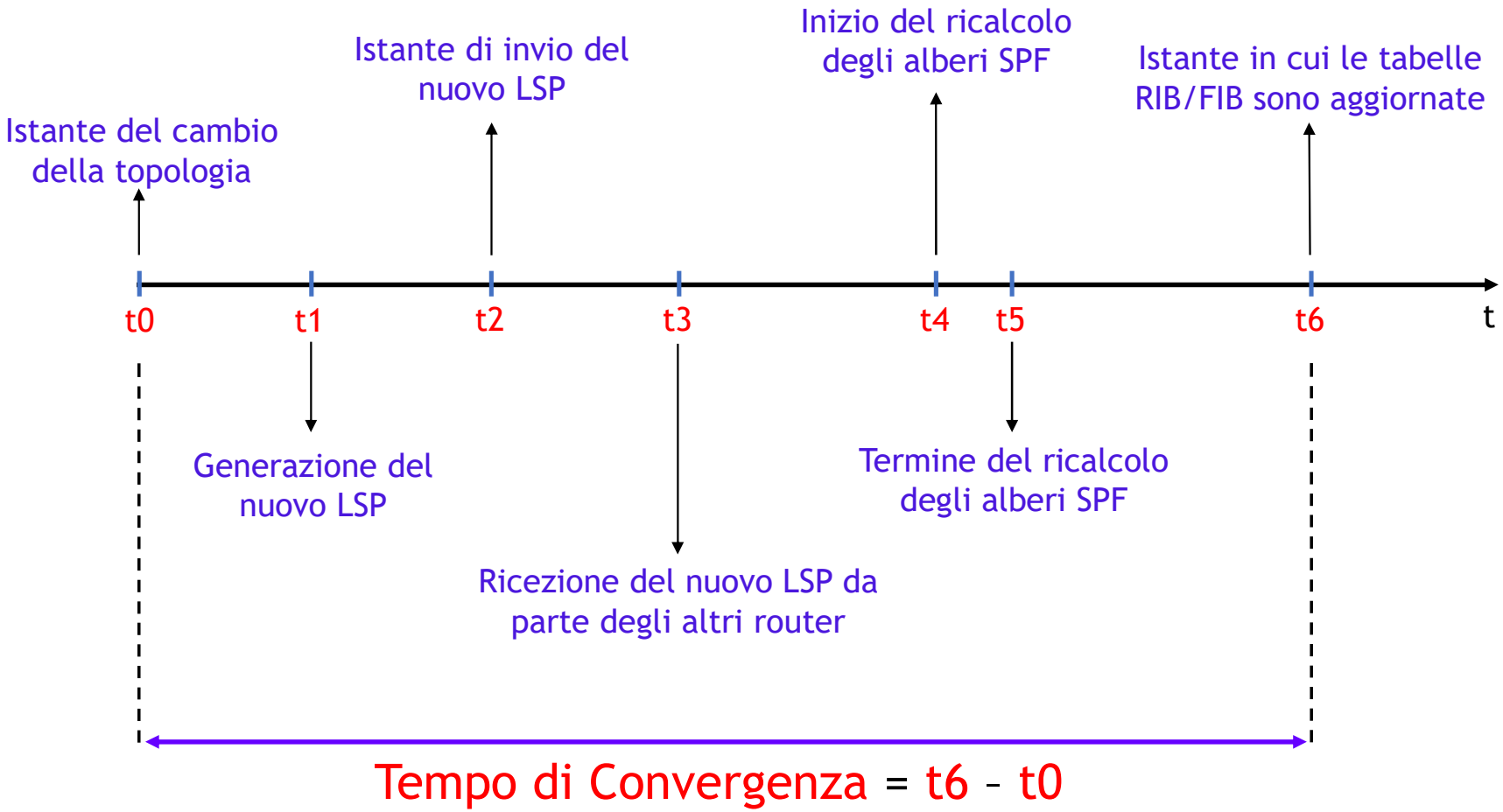


Topologia vista dall'algoritmo SPF
disabilitando l'elezione del *DR*

Di cosa parlerò …

- #1 Gli "ingredienti" fondamentali
- #2 Il *Link State DataBase* (LSDB)
- #3 L'algoritmo per la ricerca dei percorsi ottimi
- #4 Aspetti generali di convergenza
- #5 I protocolli *Link State*

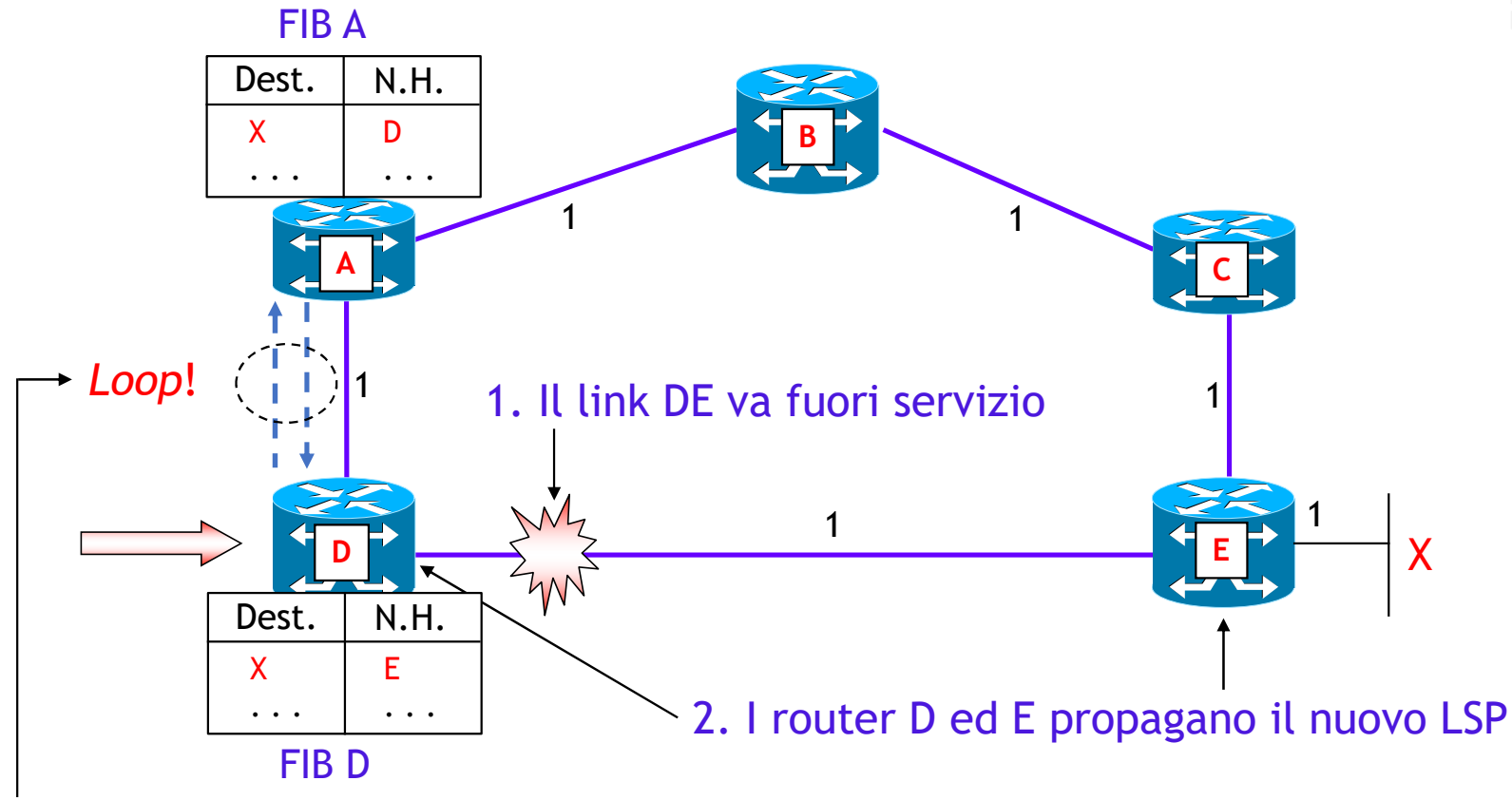
Il tempo di convergenza ...



In campo ...

- Convergenza "classica" (es. OSPF)
  - Tempi di convergenza > 5 sec (*worst case* ≈ 50 sec, con *timer* di default e rilevazione della perdita di una adiacenza via messaggi HELLO)
  - Pro: funzionalità ampiamente collaudata e disponibile per tutti i protocolli
  - Contro: tempi di convergenza elevati, ricalcolo dei percorsi svolto dalla CPU
- Convergenza "veloce"
  - Tempi di convergenza < 1 sec (*best case* ≈ 150 msec, con SPF *tuning* + BFD)
  - Pro: funzionalità ampiamente supportata (ma non disponibile per tutti i protocolli)
  - Contro: ricalcolo dei percorsi svolto dalla CPU
- *Loop-Free Alternate* (LFA)
  - Tempi di convergenza < 50 msec
  - Pro: *Next-Hop* alternativo direttamente disponibile nella FIB; evita la formazione di *microloop*
  - Contro: non applicabile a qualsiasi topologia di rete

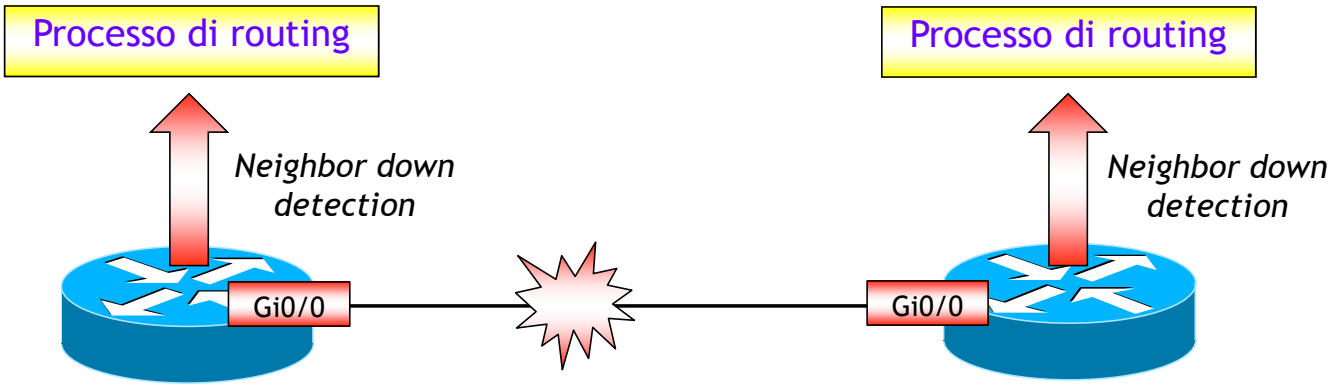
Il problema dei *microloop*



Rilevazione della perdita di una adiacenza

- La rilevazione veloce della perdita di una adiacenza è la prima priorità per ottenere una convergenza veloce di qualsiasi protocollo di routing

- Metodi disponibili
  - *Tuning* dei timer associati ai messaggi HELLO (periodo e *Holdtime*)
  - Rilevamento a livello fisico
  - *Bidirectional Forwarding Detection* (BFD)



Rilevamento a livello fisico

- Nei collegamenti *back-to-back* la perdita di una adiacenza può essere rilevata a livello fisico

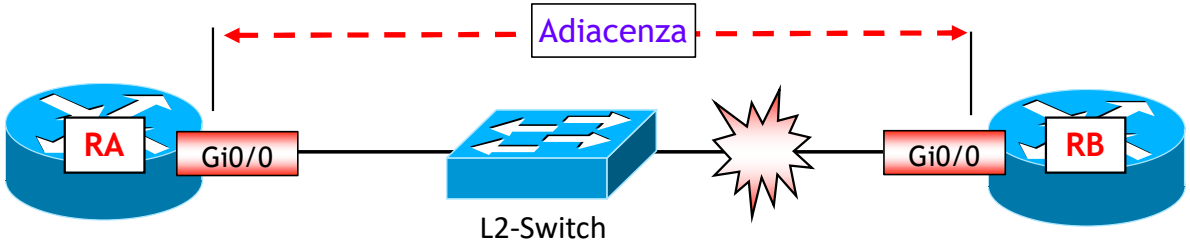
- Il tempo di comunicazione al processo di routing è regolato da *timer* che sono *hardware-dependent*
  - Interfacce Ethernet: *debounce timer* (negli switch) e *carrier-delay* (nei router)
  - Interfacce POS: *delay triggers line*

- *Best-practice*
  - *Link debounce timer* = 10 msec
  - *Carrier-delay* = 0 msec, o nel caso di protezione ottica, a un valore leggermente superiore al tempo di protezione (50-60 msec)

Bidirectional Forwarding Detection (BFD)

- BFD è un protocollo molto leggero che consente un test continuo su una comunicazione bidirezionale
  - Basato su un meccanismo di HELLO molto veloci (valore consigliato del periodo degli HELLO 300 ms e *Holdtime* 900 ms)
  - Utilizza UDP

- Utilizzato tra router adiacenti
  - Tipicamente non necessario in collegamenti POS, ATM e Frame Relay
  - Utile in collegamenti Ethernet

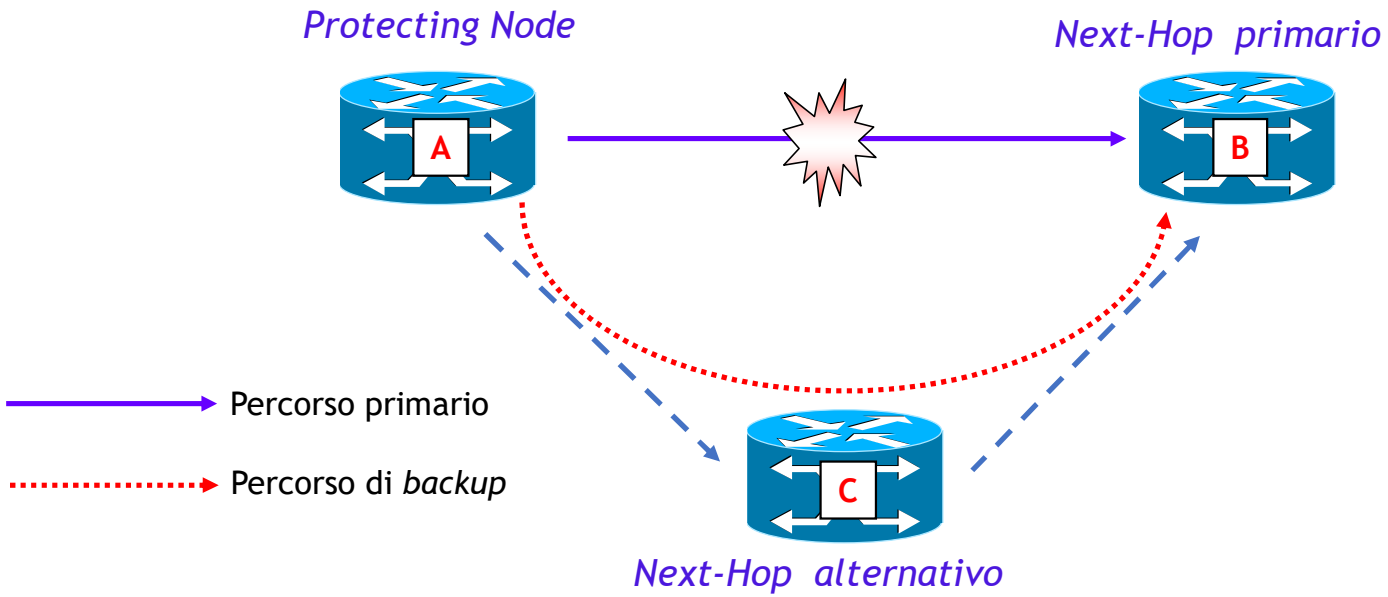


SPF per-prefix prioritization

- Non tutti i prefissi hanno la stessa importanza ...
  - *Subnet* che contengono gli indirizzi IP delle sorgenti di un servizio multicast
  - Indirizzi di *loopback* utilizzati per le sessioni iBGP
  - . . .

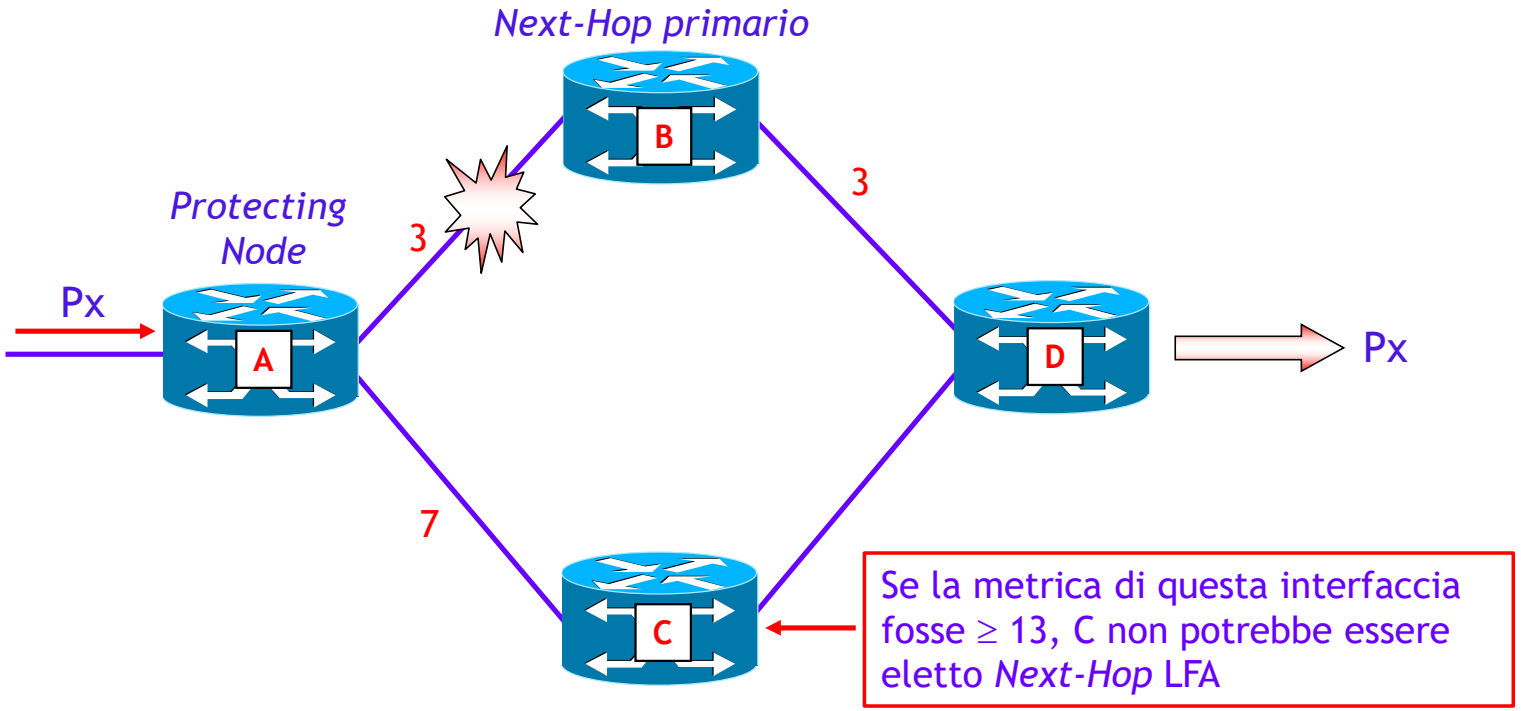
- Nel ricalcolo dell'albero SPF è più opportuno prima considerare i prefissi più importanti (*SPF per-prefix prioritization*)
  - I prefissi più importanti finiscono nella RIB e FIB prima dei prefissi meno importanti
  - Conseguenza: maggiore velocità di convergenza per i prefissi più importanti

Loop Free Alternate (LFA)



- Ogni nodo, oltre al *Next-Hop* primario, determina, se possibile, un *Next-Hop* alternativo
  - Il *Next-Hop* alternativo viene preinstallato nella FIB e utilizzato in caso di perdita di una adiacenza
  - Idea simile ai concetti di *successor*/*feasible successor* di EIGRP e di FRR MPLS-TE

Il problema: l'esistenza del *Next-Hop* LFA



- NOTA IMPORTANTE: l'esistenza di un *Next-Hop* LFA non è sempre garantita ma dipende dalla topologia della rete e dalla natura del fuori servizio (nodo e/o *link*)
  - Nell'esempio, in caso di fuori servizio del link A-B, A ottiene un nuovo percorso verso *Px* dopo la normale convergenza del protocollo IGP

Di cosa parlerò …

- #1 Gli "ingredienti" fondamentali
- #2 Il *Link State DataBase* (LSDB)
- #3 L'algoritmo per la ricerca dei percorsi ottimi
- #4 Aspetti generali di convergenza
- #5 I protocolli *Link State*

I due protocolli di routing *Link State* ...

- OSPF (*Open Shortest Path First*): è un protocollo *Link State* standard sviluppato dall'IETF
  - RFC 1131 (OSPFv1, Ottobre 1989) - *The OSPF Specification*
  - RFC 2328 (OSPFv2, Aprile 1998) – *OSPF Version 2*
  - RFC 5340 (OSPFv3, Luglio 2008) – *OSPF for IPv6*
  - RFC 5838 (OSPFv3-AF, Aprile 2010) - *Support of Address Families in OSPFv3*

- IS-IS (*Intermediate System - Intermediate System*) è un protocollo *Link State* standard sviluppato per il routing in ambiente OSI e successivamente esteso per il routing in ambiente IPv4/v6
  - ISO/IEC 10589 (Febbraio 1990): *IS-IS intra-domain routing exchange protocol*
  - *Integrated IS-IS* è una estensione del protocollo IS-IS per il routing in ambiente misto IP/OSI
    - RFC 1195 (Dicembre 1990): *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*

Ultima Diapositiva (finalmente …)