

# Spanning Tree Protocol

Alessio Carta

[Networking](#)

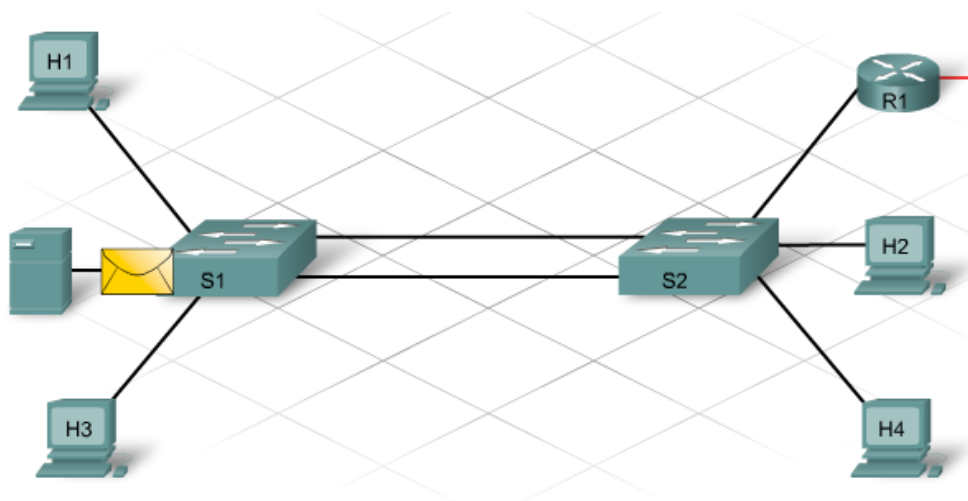
Read Time: 12 mins

Ultima modifica: 18 Giugno 2019

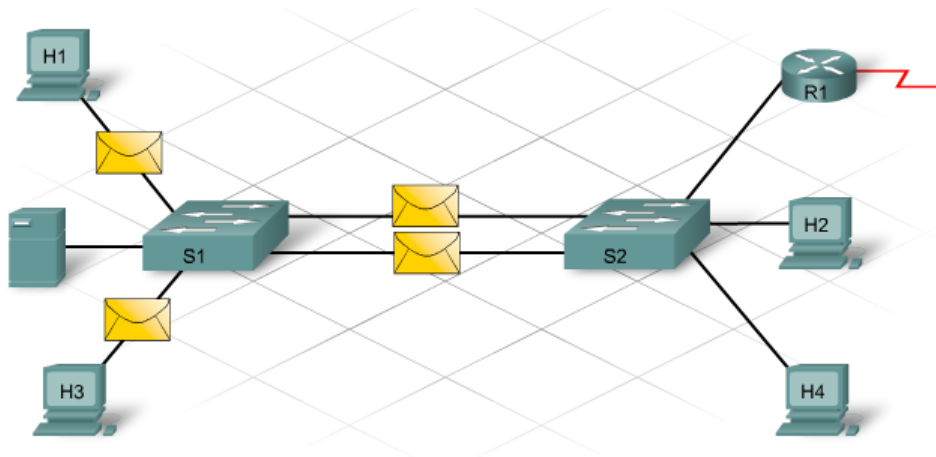
In molte situazioni è utile collegare tra loro più switch, per garantire l'espansione della rete; è utile avere anche dei collegamenti ridondati, poiché se un cavo che collega tra loro due switch si rompe oppure le porte dove è connesso hanno dei problemi, la comunicazione tra host sui differenti switch viene a mancare.

I collegamenti ridondati vanno però gestiti con attenzione, perché possono creare dei cicli infiniti (loop) all'interno della rete, generando frame duplicati o peggio dei **broadcast storm** (letteralmente tempeste di broadcast). Per capire meglio, vediamo cosa succede quando viene inviato un frame in modalità broadcast, osservando le sequenze qui sotto.

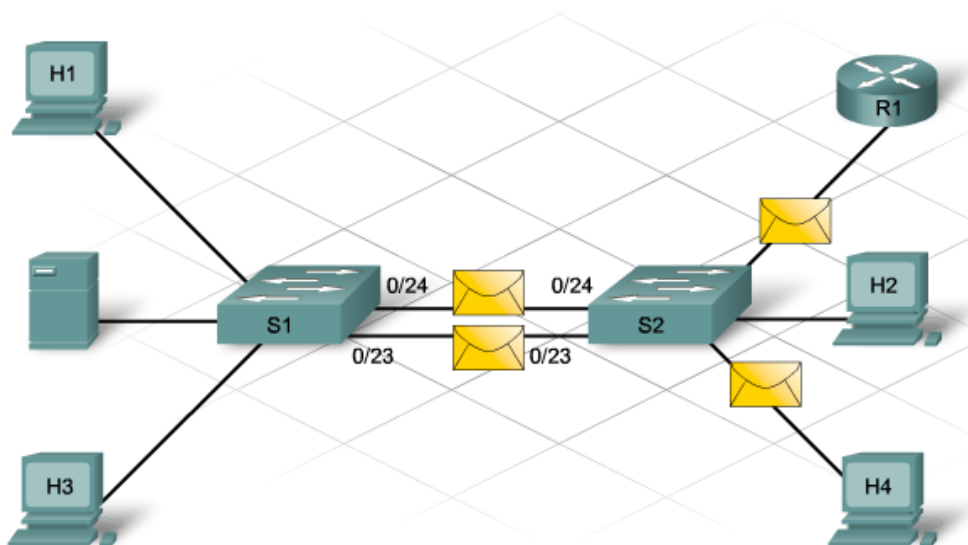
Il server (sul lato sinistro) invia un messaggio in broadcast:



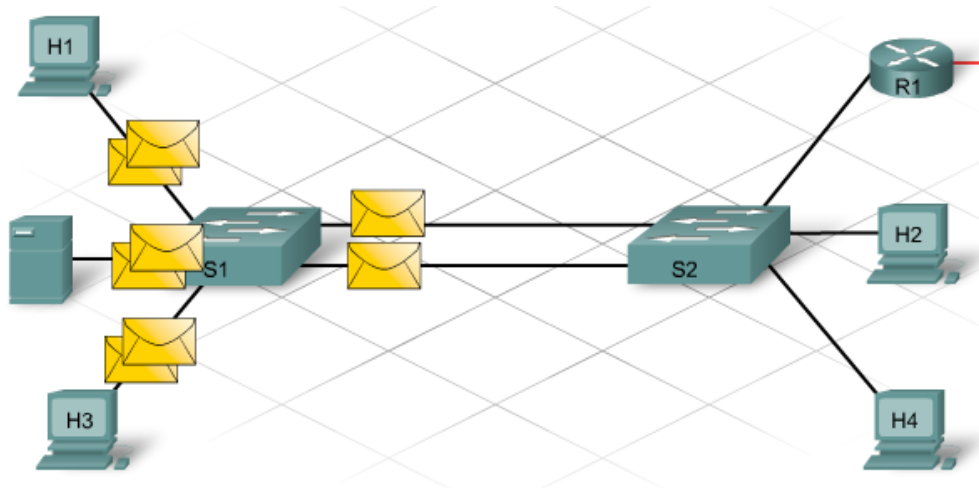
Lo switch S1 inoltrerà il frame su tutte le porte, tranne quella di arrivo:



Lo switch S2 riceve 2 frame e li elabora. Quello che arriva sull'interfaccia 0/24, essendo di tipo broadcast, verrà replicato su tutte le porte tranne la 0/24 (quella di arrivo). Lo stesso avviene per il frame arrivato sull'interfaccia 0/23, replicato su tutte le porte tranne la 0/23 (quella di arrivo). A questo punto 2 frame identici ripartono verso le interfacce dello switch S1:



Sullo switch 1 i 2 frame in arrivo saranno duplicati a causa del meccanismo appena descritto, ossia ognuno di essi sarà inoltrato verso tutte le porte dello switch tranne quella di arrivo. Il server riceverà messaggi broadcast duplicati, in quantità crescente, in ciclo continuo e infinito:



L'algoritmo STP (Spanning Tree Protocol) si occupa proprio di risolvere questo problema, fornendo un meccanismo per la disattivazione dei collegamenti ridondanti in una rete commutata. STP è un protocollo relativamente autosufficiente e richiede configurazioni minime. Alla loro accensione, gli switch con STP abilitato controllano la rete per verificare l'esistenza di percorsi ciclici (loop): se vengono rilevati, alcune porte di collegamento vengono bloccate, spezzando in questo modo l'anello generato dai percorsi ridondanti. Le altre porte vengono lasciate attive per la trasmissione delle trame.

STP definisce un albero che attraversa tutti gli switch della rete. Per evitare i loop nello switching, lo Spanning Tree Protocol:

- forza alcune interfacce ad uno stato di blocco o standby
- lascia le altre interfacce nello stato di inoltro
- riconfigura la rete attivando il percorso lasciato in standby, nel caso in cui il percorso principale non fosse più disponibile

STP prevede la presenza di uno switch designato come root, detto **root bridge**, ovvero radice dello spanning tree. Il root bridge comunica con altri switch utilizzando particolari trame dette **BPDU**, **Bridge Protocol Data Units**, inviate in multicast ogni 2 secondi a tutti gli altri switch. Le trame BPDU contengono varie informazioni, di seguito descritte:

Byte 1÷2	Protocol Identifier:	00-00
3	Protocol Version Identifier:	00
4	BPDU Type:	00
5	Flags	TC TCA
6÷13	Root Identifier	first 2 bytes = Bridge Priority last 6 bytes = MAC address of the Root Bridge
14÷17	Root Path Cost	
18÷25	Bridge Identifier	first 2 bytes = Bridge Priority last 6 bytes = MAC address of the Bridge that sends BPDU
26÷27	Port Identifier	first byte = Port Priority second byte = port number
28÷29	Message Age	
30÷31	Max Age	
32÷33	Hello Time	
34÷35	Forward Delay	

- Protocol ID: indica il tipo di protocollo (STP); è sempre 0.
- Version ID: indica la versione del protocollo STP; è 0 per STP 802.1D.
- Type: indica il tipo di messaggio BPDU; è sempre 0.
- Flags: include i bit TC=Topology Change, TCN e TCA.
- Root ID: identifica il root bridge; è composto da 2 byte, che sono la priorità, più altri 6, che costituiscono l'indirizzo MAC dello switch riconosciuto come root bridge.
- Patch Cost: costo totale del percorso per raggiungere il root bridge; questo campo viene aggiornato dopo ogni passaggio sugli switch;
- Bridge ID (BID): è l'identificatore del bridge che ha ritrasmesso la BPDU; è composto da 2 byte, che sono la priorità, più altri 6, che sono l'indirizzo MAC dello switch.
- Port ID: è l'identificatore della porta da cui è stata trasmessa la BPDU. E' composto da due byte: il primo indica la priorità, mentre il secondo è l'identificativo della porta.
- Message Age: è il tempo trascorso dalla generazione della BPDU, quindi indica da quanto tempo si è eletto il Root Bridge.
- Max Age: se Message Age supera questo tempo limite, la BPDU viene scartata, e viene rilanciata l'elezione.
- Hello Time: tempo che intercorre fra la generazione di due BPDU.
- Forward Delay: indica il tempo di permanenza nello stato di listening prima di passare nello stato di learning e nello stato di learning prima di passare nello stato di forwarding.

Il valore del Path Cost indicato nella BPDU rappresenta la velocità di un link. In base allo standard 802.1d del 1998, questo valore aveva le seguenti corrispondenze:

<b>Banda</b>	<b>Costo STP</b>
--------------	------------------

4 Mbit/s	250
10 Mbit/s	100
16 Mbit/s	62
100 Mbit/s	19
1 Gbit/s	4
2 Gbit/s	3
10 Gbit/s	2

Con lo standard 802.1d del 2004, per il path cost sono stati introdotti valori a 32 bit, consentendo maggior granularità nell'assegnazione dei costi. Le corrispondenze sono state aggiornate come indicato nella tabella sotto:

Banda	Costo STP
4 Mbit/s	5,000,000
10 Mbit/s	2,000,000
16 Mbit/s	1,250,000
100 Mbit/s	200,000
1 Gbit/s	20,000
2 Gbit/s	10,000
10 Gbit/s	2,000

## Gli stati delle porte

Appena uno switch con STP abilitato si avvia, ogni porta può attraversare una serie di 5 stati:

1. **disabled**
2. **blocking (blocco)**
3. **listening (ascolto)**
4. **learning (apprendimento)**
5. **forwarding (trasmissione)**

Lo stato di **disabled** si ha quando l'amministratore disabilita la porta dello switch con il comando shutdown. In base allo stato in cui si trova la porta, i corrispondenti LED sullo switch cambiano da una luce arancione lampeggiante ad una luce verde fissa. Possono passare fino a 50 secondi prima che una porta sia pronta per trasmettere trame. L'esistenza di questi stati è motivata dal fatto che se un collegamento si interrompe, uno switch potrebbe reagire ed inoltrare il traffico prima che la convergenza sia completa in tutta la rete. Questo può creare percorsi doppi e physical loops. Per evitare ciò lo switch mette le porte in due stati transizionali, prima di passare dal blocking al forwarding. Questi stati sono appunto **listening e learning**; il primo mette il bridge in ascolto di BPDU contenenti soluzioni migliori, l'altro impara la nuova posizione dei MAC address per scrivere la nuova MAC address table. Nel caso di caduta di qualche link, sarà necessario far ripartire l'algoritmo STP, quindi è necessario che anche le porte blocked possano ricevere BPDU. A questo punto si è praticamente generato un albero, raggiungendo lo scopo di non avere problemi di loop.

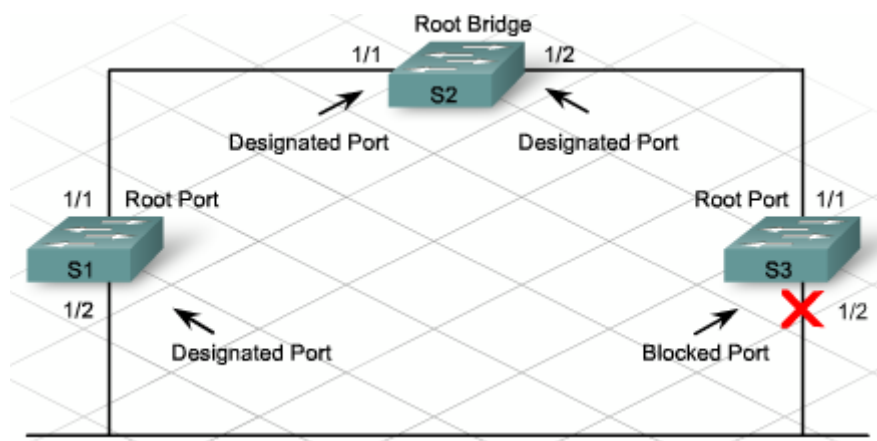
## I ruoli delle porte

Per la definizione dei ruoli delle porte sono previsti 3 passaggi fondamentali:

1. elezione del **root bridge**
2. elezione delle **root port**
3. elezione delle **designated port**

**1.** Il root bridge viene eletto in base al valore presente nel campo **Bridge Identifier (BID)** della BPDU: **“vince” lo switch con il valore minore**. Come già indicato, il BID è composto da 2 byte di Bridge Priority (di default 32768, modificabile dall'amministratore entro un range 0-65535) più 6 byte costituiti dal MAC-address. Ad esempio, se uno switch ha come MAC address AA-11-BB-22-CC-33, il BID sarà 32768:AA-11-BB-22-CC-33. Se si vuol forzare l'elezione del root bridge in favore di uno switch specifico, è sufficiente variare al ribasso la sua priorità. **All'inizio, ogni switch ritiene di essere il root bridge**: genera quindi una BPDU, impostando nel campo Root ID il proprio identificatore. In questa fase si ha che Root ID = Bridge ID. Il frame viene quindi inoltrato a tutte le porte. Quando un altro switch riceve una BPDU, confronta il Root ID ricevuto con il proprio Bridge ID; se il proprio Bridge ID è inferiore, lo switch continuerà a ritenersi root, proseguendo nell'invio delle BPDU e propagando il proprio identificativo come root; se invece il valore ricevuto è minore, lo switch riconoscerà il nuovo root e invierà le BPDU inserendo il Root ID corretto. In base a questo meccanismo, dopo un po' di tempo tutti gli switch sapranno chi è il root bridge.

**2.** Tutti gli apparati che non sono diventati root bridge, avranno una porta configurata come **root port**, ossia una porta che connette lo switch al root bridge tramite il link a minor costo. Ogni switch calcola il percorso più breve verso il root bridge stesso. Le BPDU inviate dal root avranno un **costo pari a 0**. Uno switch che riceve una BPDU dal root vi aggiunge il costo della porta tramite cui ha ricevuto la BPDU stessa (vedere tabella sul valore del Patch Cost nei paragrafi precedenti). A questo punto verificherà su quale delle sue interfacce riceve la BPDU con costo minore: l'interfaccia con questa caratteristica diventerà **root port**. Di seguito tutti gli switch, tramite il meccanismo appena descritto, identificheranno la propria root port. In caso porte di pari costo, viene scelta quella col minor **Port ID**, costituito da una priorità (di default 128) e dal numero della porta. Ad esempio, una Fa0/3 avrà un Port ID di 128.3.



*Nell'esempio proposto, dopo l'elezione del root bridge vinta dallo switch S2, prendendo in considerazione lo switch S1 vediamo che esso può ricevere BPDU su tutte e due le interfacce 1/1 e 1/2. Le BPDU ricevute sulla 1/1 avranno un costo pari a 0, a cui va sommato il costo della porta stessa; quelle ricevute sulla 1/2 avranno un costo pari a quello della porta 1/1 di S3, a cui va sommato il costo della stessa 1/2 di S1. La root port di S1 sarà pertanto la 1/1.*

**3.** Secondo il protocollo STP, ogni segmento di rete può avere una sola porta designata, tecnicamente detta **designated port**. Questa porta permette l'inoltro del traffico verso il root bridge. La designated port è quella con root path cost inferiore e, in caso di parità, quella appartenente allo switch con BID inferiore. Le altre porte che si affacciano sullo stesso segmento di rete divengono delle **blocked ports**, in grado di ricevere BPDU ma non di inoltrare traffico.

## Topology change

Con lo Spanning Tree a regime, le BPDU vengono emesse solo dai Root Bridge. Se nel periodo di tempo indicato dal campo Max Age non si ricevono BPDU, viene rilanciata l'elezione tramite un processo di topology change. Si verifica un cambio nella topologia quando una porta di trunk nello stato di forwarding va in down, oppure quando uno switch ha una nuova porta che passa nello stato di forwarding. Lo switch che rileva l'evento invia verso il root bridge una trama di **TCN, Topology Change Notification**; il root Bridge informa tutti gli switch della rete del cambiamento in corso, dando via ad una rinegoziazione dell'STP.

## I tempi dello spanning tree

Gli switch con STP attivo inviano le BPDU nella rete ogni 2 secondi. STP rimane in continuo ascolto per verificare tramite queste BPDU che non vi siano collegamenti interrotti o loop in rete. Se si verifica l'interruzione di un collegamento, STP interviene per cambiare le porte dallo stato di blocked a quello di forwarding o viceversa. Il cambiamento però non è istantaneo. Il ricalcolo dei percorsi migliori nella rete varia dai 30 ai 50 secondi per ogni switch; in questo intervallo di tempo, nelle porte interessate da STP, i dati non possono transitare.

Schematizzando:

1. Le porte iniziano ad operare in uno stato di blocking (arancione fisso), per impedire la formazione di un loop; trascorrono sino a **20 secondi** prima di passare allo stato di listening.
2. Nello stato di listening (arancione lampeggiante), si possono ricevere BPDU dagli switch vicini; tuttavia, non sono concessi l'inoltro di trame e l'apprendimento dei mac-address. Trascorrono **15 secondi** per il passaggio allo stato di learning: in questo tempo viene verificata la presenza di altre porte in trunk con le quali si potrebbero generare dei loop. Se c'è un loop si ritorna nello stato di blocking, altrimenti si passa allo stato di learning.

3. Quando la porta si trova nello stato di learning (arancione lampeggiante), si possono ricevere BPDU dagli switch vicini; non è ancora concesso l'inoltro di trame, ma si possono apprendere i mac-address. Trascorrono **15 secondi** per il passaggio allo stato di forwarding.
4. Lo stato di forwarding (verde lampeggiante) è l'ultimo della serie; la porta inoltra le trame, riceve BPDU, apprende mac-address.

Per alcune applicazioni questo tempo di attesa potrebbe essere eccessivo; inoltre l'intervento frequente di STP per modificare l'albero è sintomo di problemi, ed ha un impatto negativo sulle operazioni in rete. Esistono comunque diverse soluzioni per ridurre al minimo i tempi di ricalcolo da parte di STP. Le principali sono PortFast, UplinkFast e BackboneFast, di proprietà Cisco; non essendo standard aperti, non possono essere utilizzati in reti con switch di marche diverse. Vediamole nel dettaglio:

### **PortFast**

La funzione PortFast di STP permette ad una porta di passare immediatamente allo stato di forwarding, saltando gli stati intermedi di listening e learning. L'uso di PortFast è utile nel collegamento diretto tra una porta ed una workstation o un server, permettendo a tali macchine un accesso immediato alla rete, senza le attese dovute ai tempi di convergenza di STP.

### **UplinkFast**

Accelera la scelta di una nuova root port quando un collegamento salta.

### **BackboneFast**

Fornisce una veloce convergenza nel caso in cui si presenti un cambiamento nella topologia di STP. E' una funzione usata spesso ai livelli Distribution e Core Layers, dove più switch sono interconnessi tra loro.