STP – Spanning Tree Protocol

Introduzione

- Sviluppato da Perlman nel 1985 per la Digital Equipment Corporation (DEC); in seguito standardizzato dal IEEE come 802.1d
- Una LAN di grandi dimensioni deve essere necessariamente divisa in SEGMENTI.
  - Ethernet : ridurre il numero di hosts per segmento e quindi il numero di collisioni
  - Token Ring : ridurre il numero di hosts su segmento e quindi il tempo di attesa del token

Il Bridge (1)

- Il BRIDGE permette di :
  - Isolare i domini di collisioni dei segmenti Ethernet basati su Hub
  - Consentire la comunicazione tra nodi appartenenti a segemnti diversi, condividendo quindi un dominio di broadcast.

Il Bridge (2)

- Il Bridge “apprendendo” (*learning mode*) la corrispondenza tra proprie porte ed indirizzi fisici (MAC) dei singoli hosts (MAC-Address-Table*):
  - Inoltra il trafico da un segmento all’altro quando:
    - il segmento del ricevente è differente da quello del trasmittente (***forwarding mode***)
    - ignora quale sia il segmento del ricevente (***flooding mode***)
    - Il traffico è di tipo broadcast
  - Non inoltra il traffico da un segmento all’altro quando:
    - Il trasmittente ed il ricevente appartengono allo stesso segmento (***filtering mode***)

* Le voci che compongono la MAC-AddressTable si formano mano a mano che il traffico raggiunge i bridges. A questo punto vengono poste in una “cache” da dove venfono rimosse allo scadere di un timeout

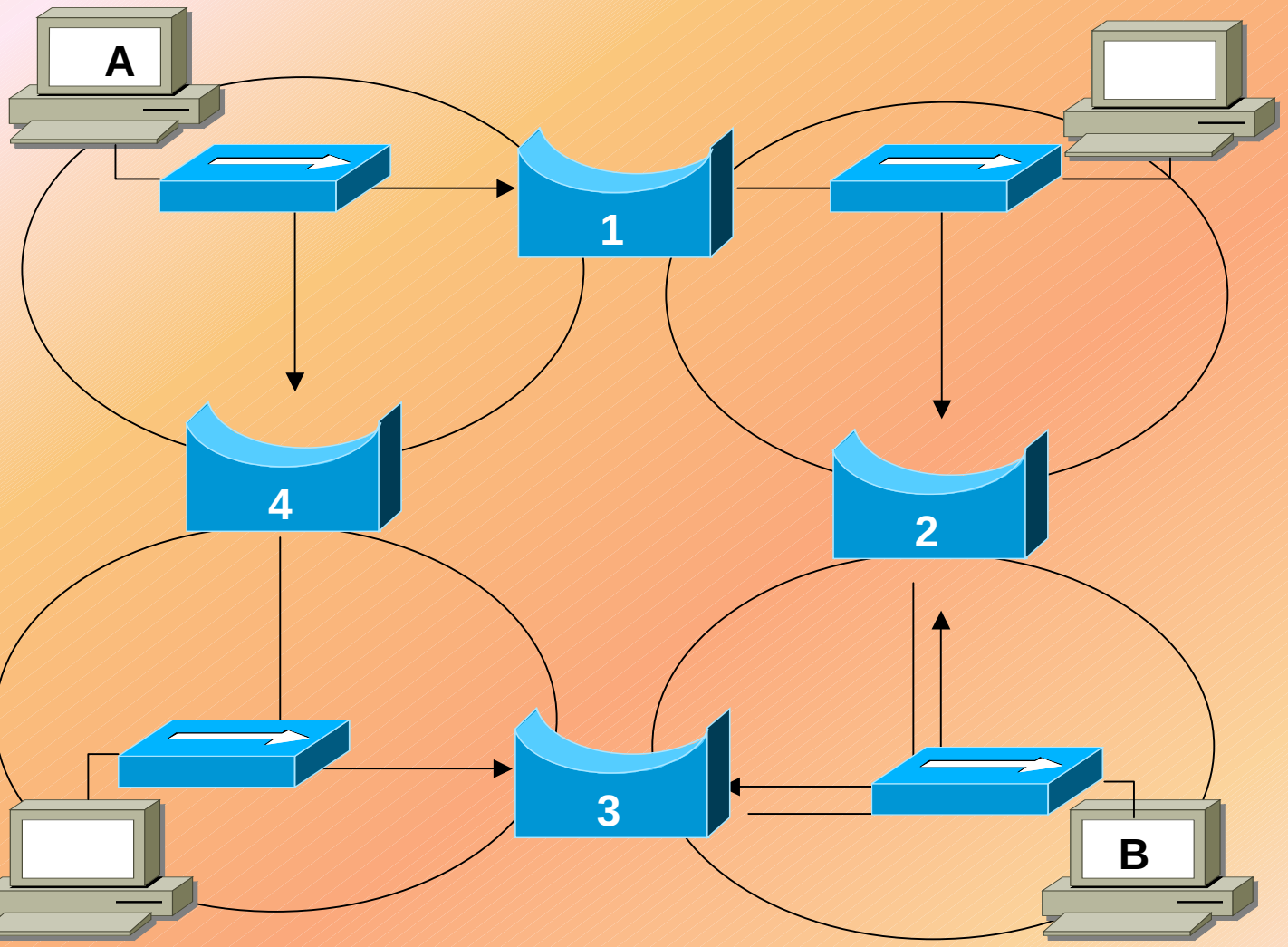
I Bridge (3)

- Problematiche :
  - Se la rete si basa sulla presenza di singoli bridge per la connessione dei segmenti, avremo un “Single Point Of Failure” in corrispondenza del Bridge, con conseguente fragilità del sistema (Scarsa Ridondanza)
- Soluzione:
  - Aggiungere altri Bridges per avere la possibilità di cammini multipli per il traffico in caso di “failure”.
  - ATTENZIONE !!! Questa soluzione rischia di creare problemi ancora maggiori di quanti non ne risolva. Vediamo quali…

I Loop (1)

- Un Loop si verifica quando su una rete i pacchetti vengono inoltrati su percorsi che chiudendosi su loro stessi, danno vita ad un traffico senza fine che consuma inutilmente la banda a disposizione.
- Se poi il traffico è broadcast, si verifica una situzione chiamata di “Broadcast Storm” che oltre ad avere effetti deleterei sulla banda trasmissiva, oblliga tutti gli host a rallentare per analizzare il traffico indirizzato a tutti

I Loop (2)



I Loop (3) – Broadcast Storm

- Host A invia un messaggio broadcast
- I bridges 1 e 4 lo inoltrano rispettivamente ai bridges 2 e 3
- Il bridge 3 inoltra al 4 il broadcast ricevuto dal bridge 2
- Il bridge 2 inoltra al 1 il broadcast ricevuto dal bridge 3
- E così via senza fine.

- MA CI SONO ALTRI PROBLEMI…

I Loop (4)

- **Copie multiple**:
  - Quando l'host A spedisce un pacchetto (unicast) all'host B, quest'ultimo ne riceve due copie. Una attraverso il percorso 4-3 ed una attraverso il percorso 1-2.
- **Corruzione della "MAC-Address-Table"**:
  - L'integrità di Questa tabella costruita in modo dinamico dal bridge, può risentire del traffico in loop, in quanto il bridge riceve su entrambe le proprie porte il traffico inviato da un singolo host.
  - Quando B risponde al messaggio di A, i Bridges 2 e 3 filtreranno il pacchetto ritenendo erroneamente sulla base delle propie tabelle corrotte che A si trovi sullo stesso segmento di B. Quseto avviene perché i due bridge hanno visto arrivare il pacchetto di A per B anche sulle porte che li connettono direttamente.

La Soluzione : STP

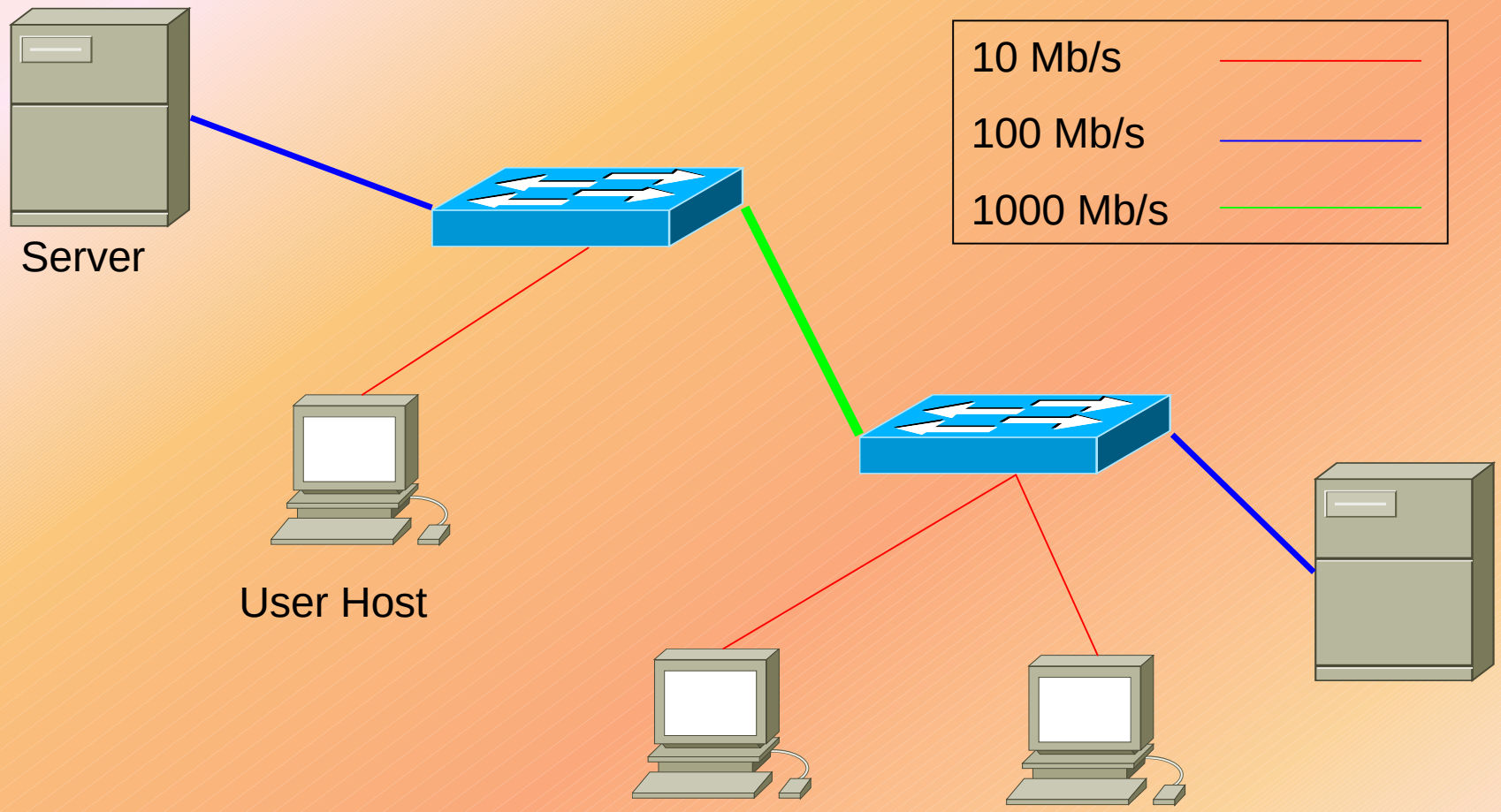
- Lo Spanning Tree Protocol è un protocollo che consente di realizzare topologie “logicamente” prive di loop, grazie alla gestione dinamica dello stato delle porte

  (forwardin o blocking).
- Il concetto fondante del STP è la presenza di un solo Root Bridge e di una serie di Non-Root Bridges.

Bridge vs Switch

- Un Switch è un bridge “micro-segmentato”
  ovvero ogni host ha una connessione punto-punto con una propria porta sul switch (riduzione della probabilità di collisione)
- Relaizzazione in Hardware (APIC) degli algoritmi per l’inoltro ed il controllo del traffico (maggiore velocità rispetto al bridge)
- 2 tipi di modalità di inoltro:
  - Store and forward (pro:controllo sui frames prima dell’inoltro; contro: lentezza)
  - Cut through (pro:velocità; contro:nessun controllo sul frame)

LAN basata su Switch : La Banda

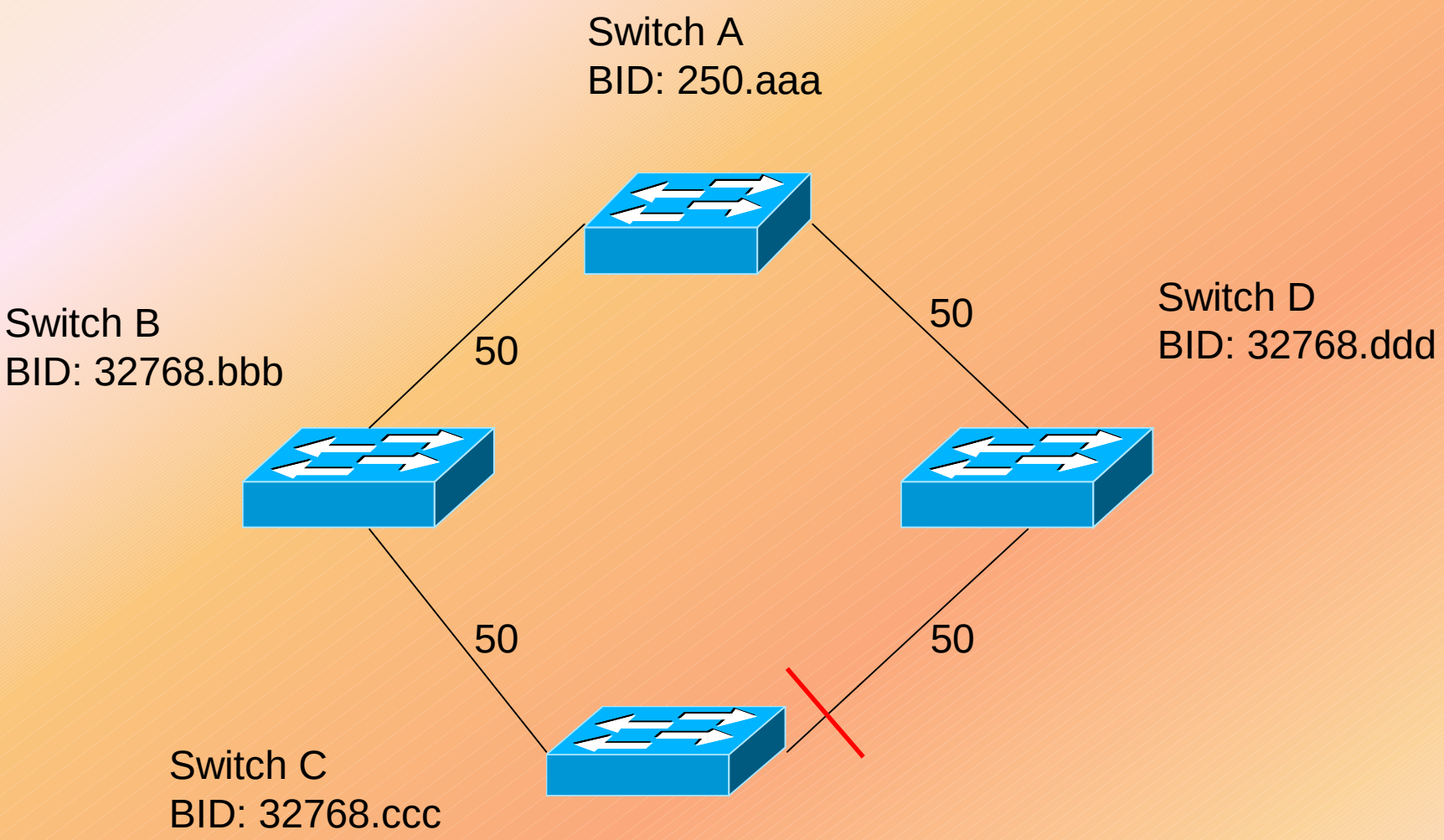


Root Bridge: L'Elezione

- Ogni bridge viene impostato con un "Bridge ID" = Bridge Priority (2 byte) + MAC (6 byte)
- La Bridge Priority di default per Cisco è settata a 32768 (0x8000).
- Il bridge che viene eletto come Root è:
  - Quello con Bridge Priority più bassa
  - Quello con MAC più basso, a parità di Bridge Priority
- Consigliabile posizionare il Root Bridge al centro della topologia, in modo da ottimizare i tempi di convergenza per tutti i non-root bridge
- Consigliabile configurare un Backup Root Bridge con Bridge Priority di poco maggiore a quella del root bridge, in modo che sia scelto dal processo di elezione nel caso di guasto del Root Bridge.
- I messaggi BPDU (Bridge Protocol Data Unit) sono inviati ogni 2 secondi in modo che i bridges apprendano i reciproci Bridge ID e quindi possano eleggere il Root Bridge.
- Massimo tempo di elezione = 2sec * [(Numero Bridge)-1]
  Es. 5 Bridge ➔ 2sec*(5-1) ➔ 8sec

Root Path Cost

- I Non-Root Bridges decidono quale percorso seguire (e quindi quale porte porre in stato di *forwarding*) sulla base del **costo** dei vari links.
- Il Costo dipende dalla “BandWidth” del link:
  - 10 Mbps ➔ Costo = 100
  - 45 Mbps ➔ Costo = 39
  - 100 Mbps ➔ Costo = 19
  - 622 Mbps ➔ Costo = 6
  - 1000 Mbps ➔ Costo = 4
- Tramite il traffico BPDU, i bridges determinano la rotta migliore sommando i costi dei vari tratti e scegliendo quella con costo totale inferiore (cioè quella dotata di maggiore larghezza di banda)
- In caso di parità di costo, viene scelto il bridge che abbia il Bridge ID più basso.
- La porta scelta sulla base del costo diviene la *Designated Port* e viene posta in stato di forwarding
- La porta scartata viene posta in stato di *blocking*, ma lasciata comunque attiva nel caso si verificasse un guasto sul lato che connette l’atra porta al Root Bridge



Nell’esempio, il Switch C sceglie il percorso verso il Root Bridge A che passa per il Switch B, perché a parità di costo (50 + 50 in entrambi i lati) sceglie il BID più basso (che a parità di Priority è dato dal MAC inferiore).

Stati di una Porta

- **Disabled**: nessuna attività sulla porta
- **Listening**: stato presente nelle fasi di elezione del Root Bridge e nella scelta di root port, designated port e deignated bridge
- **Learning**: dopo le suddette scelte, il bridge passa a costruire la Bridge Table
- **Forwarding**: scaduto il Forward Delay Timer (default = 15 sec), la porta passa nello stato di inoltro
- **Blocking**: la porta non inoltra il traffico, ma rimane attiva per eventuali riconfigurazioni topologiche

BPDU: Dettagli (1)

- Il Root Bridge invia i BPDU ogni 2 sec (Hello Timer)
- Se un non-root bridge non riceve messaggi BPDU per 20 sec (Maximum Age Time), ricalcola sulla base delle propri tabelle il Root Cost ed il Root Path.

Prima di cambiare lo stato della porta da “blocking” in “forwarding” attende che sia scaduto il Forward Delay, per evitare di formare loops dovuti a cambiamenti topologici troppo veloci.

BPDU: Dettagli (2)

- Ci sono 2 tipi di BPDU:
  - Configuration BPDU: messaggi utilizzati per eleggere il root-bridge, scegliere il non-root bridges, il root-cost, la root-port
  - Topology BPDU: quando un bridge riceve un “Topology Configuration Change (TCN-BPDU)”, questi messaggi vengono inviati a ritroso in modo che risalendo i nodi dell’albero giungano al Root-Bridge, il quale invia un Ack per avvisare i bridges a valle di non inviare altri TCN-BPDU ed inoltre li avverte di impostare il timeout delle loro MAC-Address-Table passando dal valore default di 300 sec (5 minuti) a quello di 15 sec (uguale al Forward Delay). Questo è necessario in quanto altrimenti accadrebbe che i bridges, pur configurando le porte con la nuova topologia, non potrebbero comunicare per la presenza di dati ormai non più aggiornati nelle MAC-Address-Tables

VLAN & STP (1)

- Le VLANs (Virtual Area Networks) sono dei domini logici di broadcast che risiedono su un unico dominio di broadcast fisico (un switch)
- La comunicazione tra i Switches che connettono fisicamente le VLANs avviene tramite una connessione chiamata “TRUNK” che incanala il traffico appartenente a differenti VLANs rendendo però incomunicanti le varie VLANs tra loro.
- I protocolli di Trunkink sono 2:
  - Inter-Switch Link (ISL) di Cisco
  - IEEE 802.1q
- Perché diverse VLAN possano comunicare è necessaria la presenza di almeno un router connesso agli switch (**router on a stick**)

VLAN & STP (2)

- Cisco consente che ogni VLAN abbia una propria implementazione del STP, ovvero ogni VLAN può avere il proprio Root-Bridge ed inoltre una porta può essere in stato di blocking per una VLAN e di forwarding per un’altra.

  Questo permette di ottimizzare i root-path, ma inevitabilmente richiede potenze di calcolo maggiori a carico delle CPU dei bridges.

  Infatti 10 VLANs gestite da 2 switches, richiederebbero 20 BPDU al secondo.

VLAN & STP (3)

- PROBLEMATICHE
  - SCALABILITA': STP è un protocollo abbastanza scalabile ma ovviamente il crescere del numero di bridges comporta un maggiore traffico di rete, un maggiore carico sulle CPU dei bridges e soprattutto una maggiore probabilità di incorrere in problemi dovuti a malfunzionamenti dei bridges.