

# Spanning tree protocol

**Mario Baldi**

Politecnico di Torino

<http://staff.polito.it/mario.baldi>

**Pietro Nicoletti**

Studio Reti

<http://www.studioreti.it>

Basato sul capitolo 4 di:

M. Baldi, P. Nicoletti, "Switched LAN", McGraw-Hill, 2002, ISBN 88-386-3426-2

# Nota di Copyright

Questo insieme di trasparenze (detto nel seguito slide) è protetto dalle leggi sul copyright e dalle disposizioni dei trattati internazionali. Il titolo ed i copyright relativi alle slide (ivi inclusi, ma non limitatamente, ogni immagine, fotografia, animazione, video, audio, musica e testo) sono di proprietà degli autori indicati a pag. 1.

Le slide possono essere riprodotte ed utilizzate liberamente dagli istituti di ricerca, scolastici ed universitari afferenti al Ministero dell'Istruzione, dell'Università e della Ricerca, per scopi istituzionali, non a fine di lucro. In tal caso non è richiesta alcuna autorizzazione.

Ogni altro utilizzo o riproduzione (ivi incluse, ma non limitatamente, le riproduzioni su supporti magnetici, su reti di calcolatori e stampate) in toto o in parte è vietata, se non esplicitamente autorizzata per iscritto, a priori, da parte degli autori.

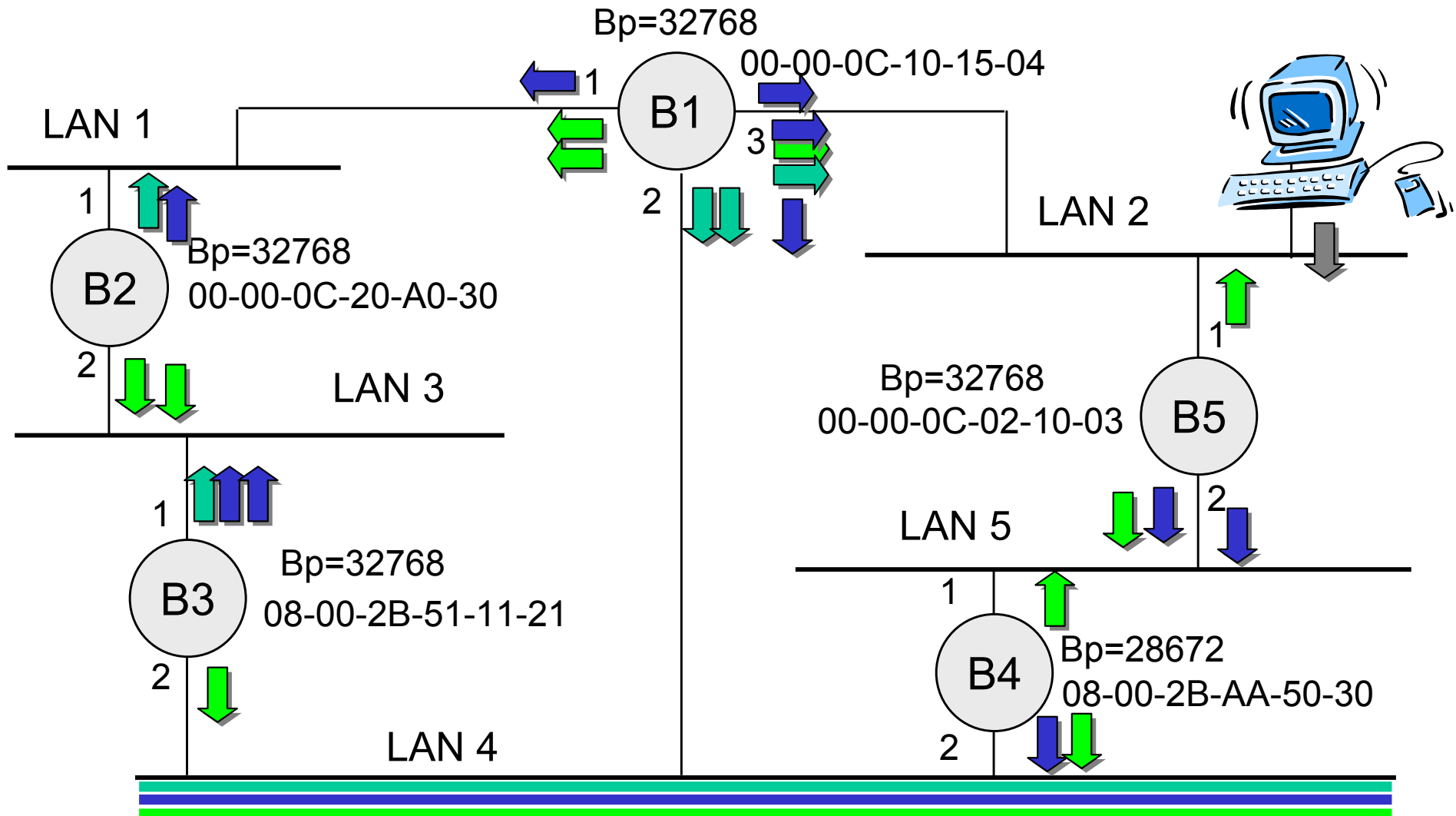
L'informazione contenuta in queste slide è ritenuta essere accurata alla data dell'edizione. Essa è fornita per scopi meramente didattici e non per essere utilizzata in progetti di impianti, prodotti, reti, ecc. In ogni caso essa è soggetta a cambiamenti senza preavviso. Gli autori non assumono alcuna responsabilità per il contenuto di queste slide (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 slide.

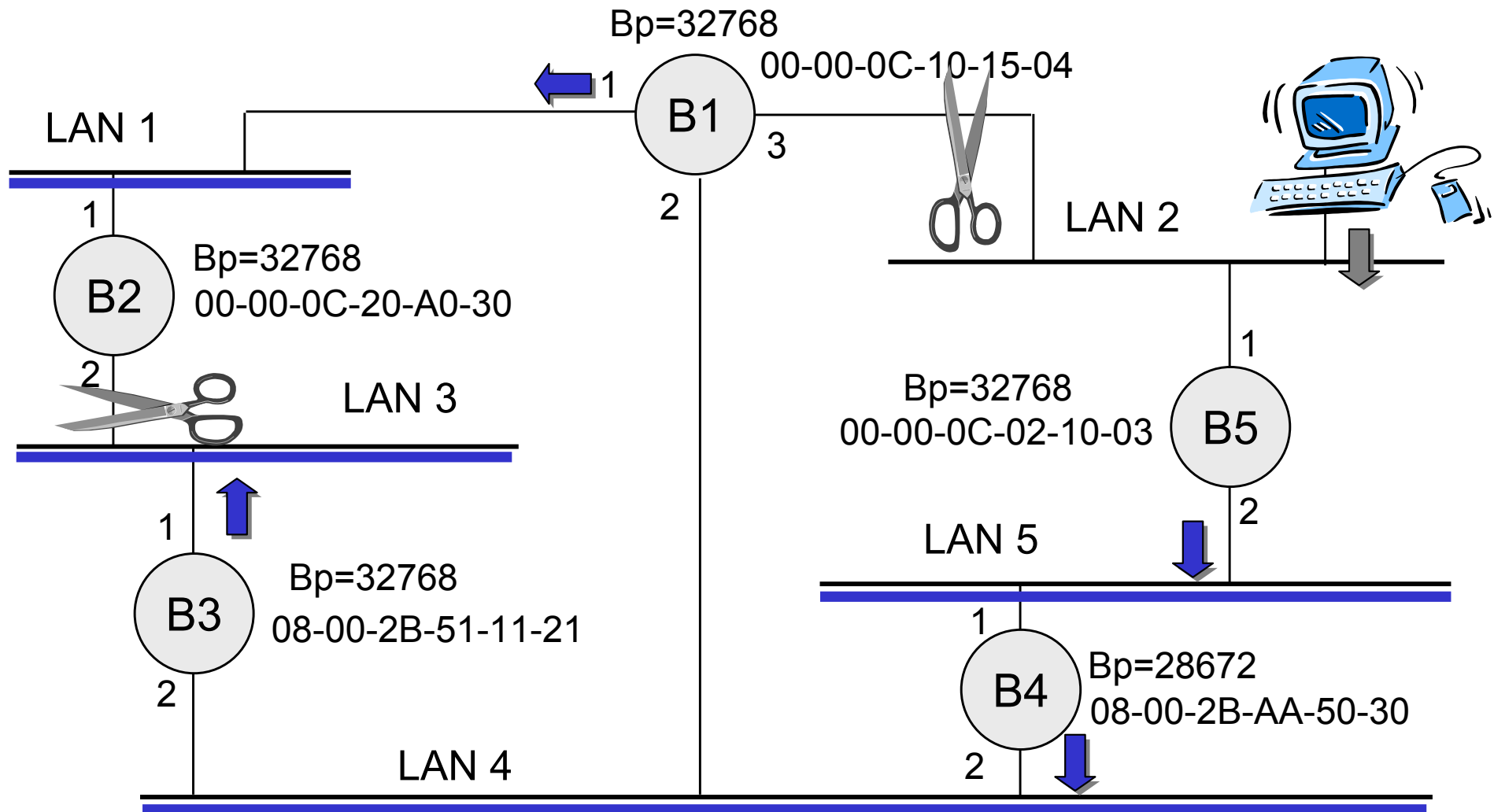
In ogni caso questa nota di copyright non deve mai essere rimossa e deve essere riportata anche in utilizzi parziali.

# Problemi dei bridge con i percorsi chiusi

Completa saturazione della rete in pochi secondi: **broadcast storm**



# Soluzione: eliminare i percorsi chiusi



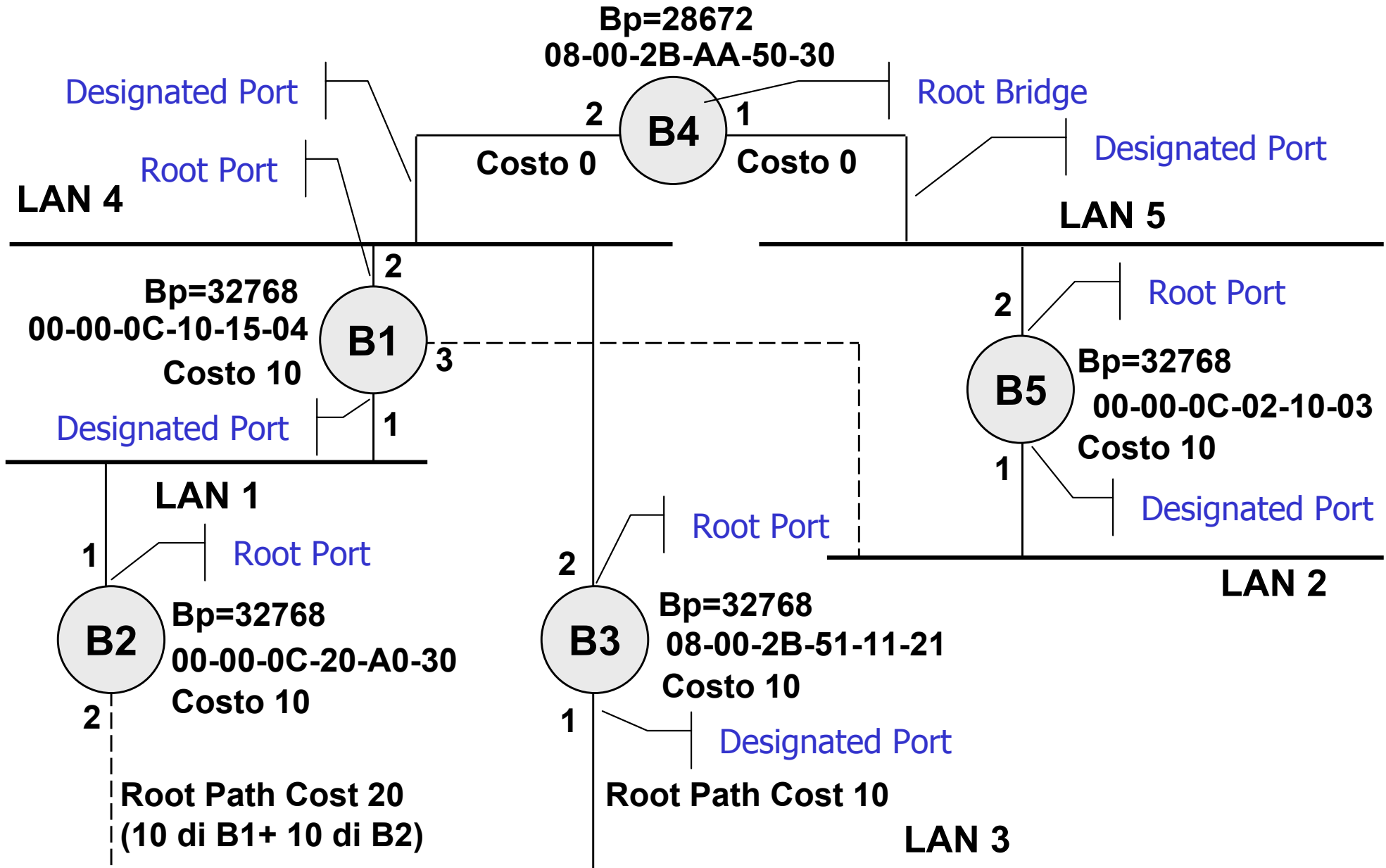
# Spanning Tree Protocol

- Eseguito dal processo Spanning Tree
- Definito da IEEE 802.1D
- Trasforma una rete con maglie in un albero
  - Eliminazione di percorsi circolari chiusi (loop)

Tre fasi:

- Elezione del ***root bridge***
  - Radice dell'albero (spanning tree) che si vuole costruire
- Selezione della ***root port***
  - Una tra le porte di un bridge è usata per raggiungere il root bridge
- Selezione delle ***designated port***
  - Una tra le porte collegate ad una LAN è usata per ricevere e inoltrare pacchetti

# Risultato dello spanning tree protocol



# Per fissare il tutto ... un po' di poesia

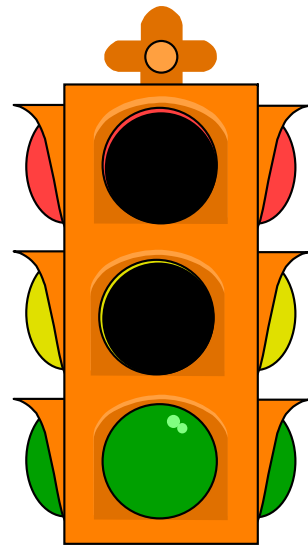
I think that I shall never see  
A graph more lovely than a tree.  
A tree whose crucial property  
Is loop-free connectivity.  
A tree which must be sure to span  
So packets can reach every LAN.  
First the Root must be selected  
By ID it is elected.  
Least cost paths from Root are traced  
In the tree these paths are placed.  
A mesh is made by folks like me.  
Then bridges find a spanning tree.

[Radia Perlman]

# Stato delle porte

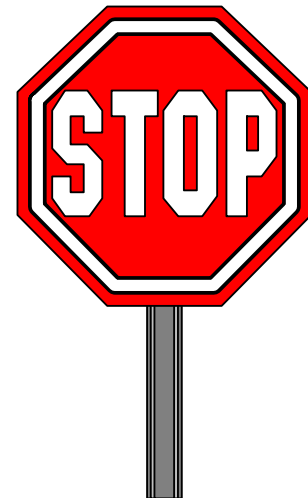
## ■ *Forwarding*

- Root port e designated port
- La porta è utilizzata per inoltrare pacchetti
- I pacchetti ricevuti dalla porta vengono elaborati, ed eventualmente inoltrati dal bridge



## ■ *Blocking*

- Tutte le altre porte
- Non vengono inoltrati pacchetti sulla porta
- I pacchetti ricevuti sono ignorati



# Bridge Protocol Data Unit (BPDU)

- Trasmesse ad indirizzo multicast predefinito
- **Configuration** BPDU
- **Topology Change Notification** BPDU

Dest. Addr.	Source Addr.	Length	DSAP	SSAP	Control	BPDU	
Multicast 01-80-C2 00-00-00	Singlecast Indirizzo Bridge	XY	042H	042H	XID	Configuration BPDU oppure Topology Change Notification BPDU	FCS

**BPDU: Bridge Protocol Data Unit**  
**DSAP: Destination Service Access Point**  
**SSAP: Source Service Access Point**

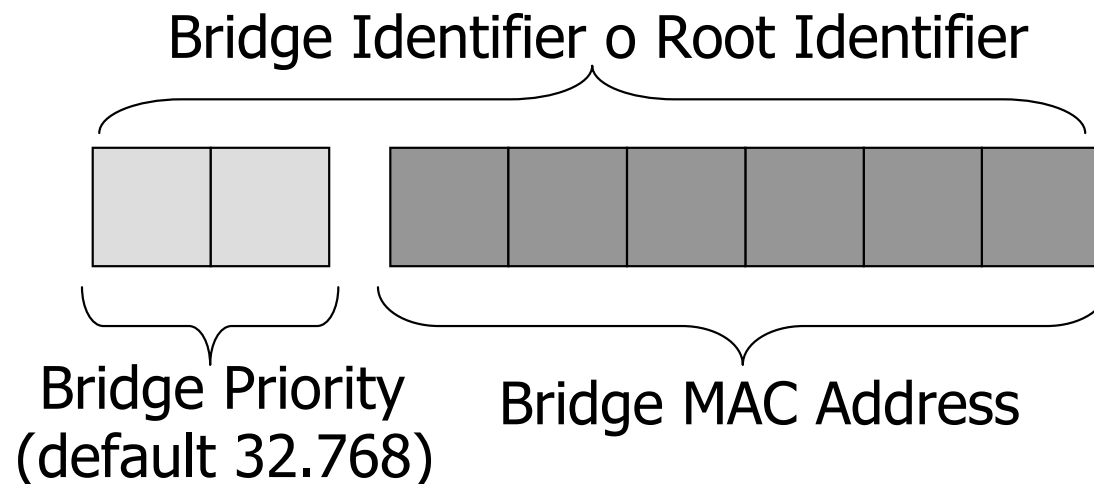
Byte 1÷2	<b>Protocol Identifier:</b>	<b>00-00</b>
3	<b>Protocol Version Identifier:</b>	<b>00</b>
4	<b>BPDU Type:</b>	<b>00</b>
5	TC	Flags
		TCA
	<b>Root Identifier</b>	
6÷13	primi 2 byte = Bridge Priority	successivi 6 byte = Indirizzo MAC del Root Bridge
14÷17	<b>Root Path Cost</b>	
	<b>Bridge Identifier</b>	
18÷25	primi 2 byte = Bridge Priority	successivi 6 byte = Indirizzo MAC del Bridge che trasmette la BPDU
	<b>Port Identifier</b>	
26÷27	primo byte = Port Priority	secondo byte = numero di porta
28÷29	<b>Message Age</b>	
30÷31	<b>Max Age</b>	
32÷33	<b>Hello Time</b>	
34÷35	<b>Forward Delay</b>	

# Configuration BPDU

- **Root Identifier**
  - identificatore del bridge assunto come root
- **Root Path Cost**
  - costo per raggiungere il bridge che ha originato la Configuration BPDU lungo il percorso lungo cui il messaggio ha transitato
- **Bridge Identifier**
  - identificatore del bridge che ha originato la Configuration BPDU
- **Port Identifier**
  - identificativo della porta del bridge attraverso cui la Configuration BPDU è stata generata

# Bridge Identifier

- Ogni bridge ha un indirizzo MAC per ogni interfaccia
  - Uno viene scelto per costruire l'identificatore del bridge
- Ad ogni bridge è associata una priorità: **bridge priority**
  - Valore di default per garantire funzionamento "plug&play"
- Si elegge root bridge quello con bridge identifier minore
  - La bridge priority minore determina il root bridge
  - A parità di bridge priority viene eletto il bridge con indirizzo minore



# Root Bridge Election

- Inizialmente ogni bridge assume di essere root bridge
  - Inserisce il proprio bridge identifier nel campo root identifier
  - Genera Configuration BPDU con periodo **hello time**
    - Default: 2 secondi
- Ogni bridge confronta il proprio bridge identifier con il campo root identifier delle Configuration BPDU ricevute
  - Se è minore, continua a generare Configuration BPDU
    - Il campo root identifier contiene il suo bridge identifier
  - Altrimenti, inoltra le Configuration BPDU (flooding) ricevute
    - Il campo root identifier contiene il bridge identifier del candidato a divenire root bridge
- Alla fine solo il root bridge origina Configuration BPDU

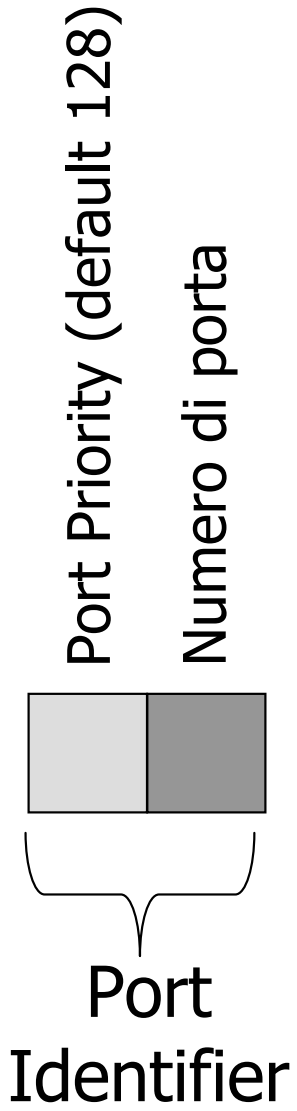
# Root Port Selection

- Porta sul cammino di costo minimo verso il root bridge
  - Riceve Configuration BPDUs generate dal root bridge
  - Inoltra trame verso il root bridge
- Ad ogni porta è associato un costo
- Le Configuration BPDUs contengono il costo del cammino attraversato
  - Campo `root path cost`
  - Il costo associato alla porta di ricezione viene sommato al campo `root path cost` da ogni bridge attraversato
- Viene scelta root port quella che riceve Configuration BPDUs con priorità maggiore
- Le Configuration BPDUs ricevute dalla root port sono inoltrate sulle altre porte

# Priorità delle Configuration BPDUs

Una Configuration BPDUs ha priorità maggiore di un'altra se soddisfa, nell'ordine, una delle seguenti condizioni

- Il valore del campo `root path cost` è minore
  - Aggiornato sommando il costo (path cost) associato alla porta di ricezione il valore nella Configuration BPDUs ricevuta
- Il valore del campo `bridge identifier` è minore
- Il valore del campo `port identifier` è minore
- Il parametro ***port identifier*** associato alla porta di ricezione è minore



## Designated port selection

- Su una LAN su cui è collegata più di una porta non root vengono inoltrate più copie di Configuration BPDU
  - Hanno fatto percorsi diversi dal root bridge alla LAN
- Un bridge collegato alla LAN tramite porta non root riceve Configuration BPDU inoltrate da altri bridge
- La porta è scelta come designated se le Configuration BPDU inoltrate hanno priorità più elevata di quelle ricevute
  - Ogni altra porta è portata in stato blocking
  - Solo la designated port inoltra le BPDU sulla LAN

# Parametri di configurazione

- I bridge sono plug&play
  - Sono in grado di funzionare con i valori di default
- Bridge priority
  - Range: 0 - 61440
  - Default/raccomandato: 32768
  - Incremento consigliato (IEEE 802.1t): 4096
- Port priority
  - Range: 0 - 240
  - Default/raccomandato: 128
  - Incremento consigliato (IEEE 802.1t): 16
- Path cost
  - Range: 0 - 65535
  - Raccomandato (IEEE 802.1D):  $1000/(\text{velocità in Mb/s})$

# Path cost raccomandato da IEEE 802.1D rev1998

<b>Velocità porta</b>	<b>Valore raccomandato</b>	<b>Intervallo di valori raccomandati</b>	<b>Intervallo di valori accettabili</b>
4Mb/s	250	100 - 1000	1 - 65535
10 Mb/s	100	50 - 600	1 - 65535
16 Mb/s	62	40-400	1 - 65535
100 Mb/s	19	10 - 60	1 - 65535
1 Gb/s	4	3 - 10	1 - 65535
10 Gb/s	2	1 - 5	1 - 65535

# Trattamento delle configuration BPDU

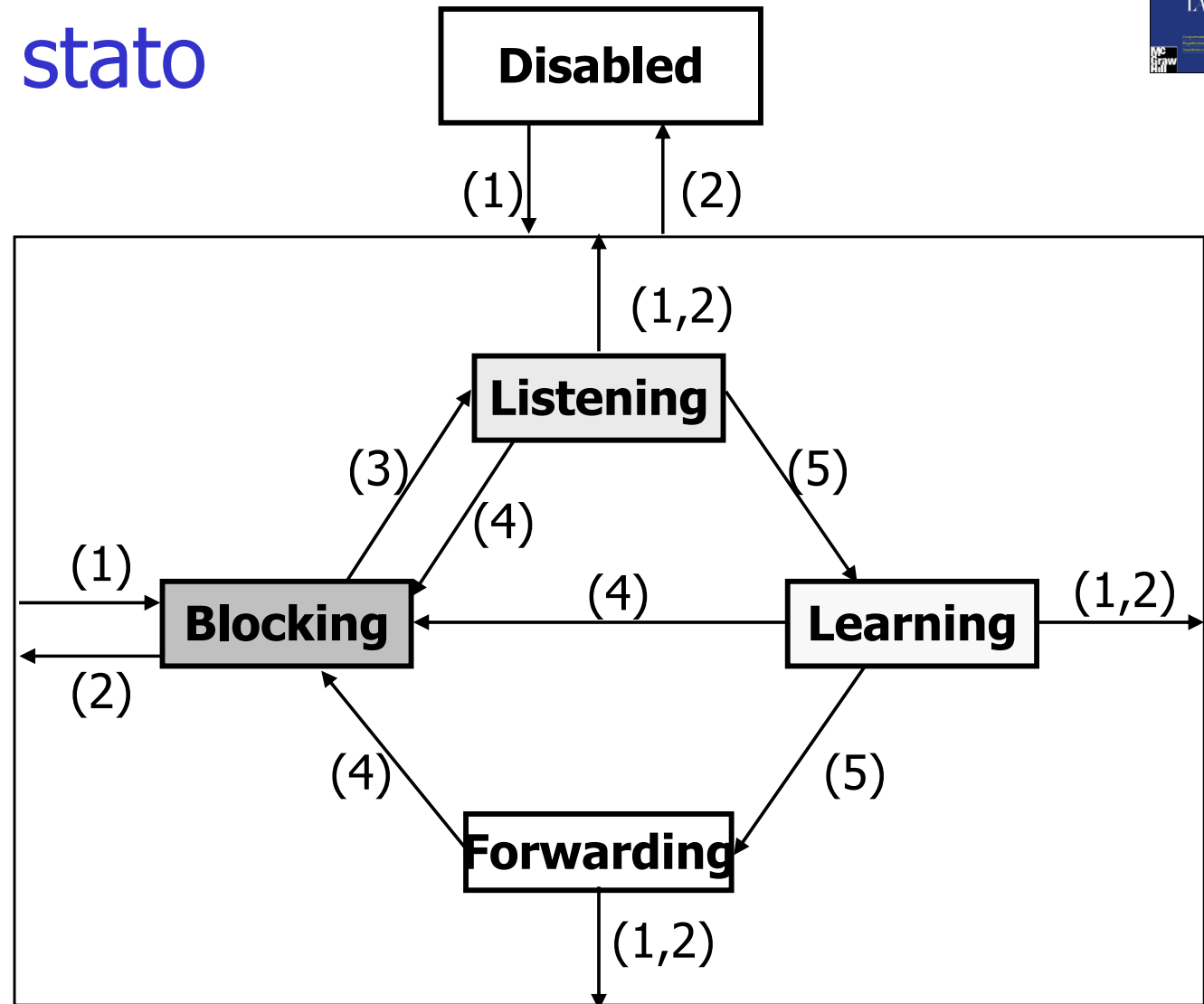
- Una configuration BPDU ricevuta dalla root port è inoltrata su tutte le altre porte (flooding)
- Il bridge aggiorna ogni copia:
  - Somma al `root path cost` il parametro `path cost` associato alla porta di ricezione
  - Inserisce il proprio `bridge identifier` nel campo omonimo
  - Inserisce il `port identifier` della porta su cui la BPDU è inoltrata nel campo omonimo

# Timer

- Hello time
  - Periodicità per la generazione di Configuration BPDU
  - Range: 1 - 10 secondi - Raccomandato: 2 secondi
- Forward delay timer
  - Ritardo in alcuni cambiamenti di stato delle porte
  - Tempo di rimozione veloce di entry del filtering database
  - Range: 4 - 30 secondi - Raccomandato: 15 secondi
- Max age
  - Intervallo tra la ricezione di una Configuration BPDU e "sblocco" di una porta
  - Range: 6 - 40 secondi - Raccomandato: 20 secondi

I bridge adottano i valori annunciati dal root bridge all'interno delle configuration BPDU

# Diagramma di stato delle porte



- (1) Management o inizializzazione
- (2) Management o guasto (non connessa o no Link Integrity Test)
- (3) Seleziona come Designated o Root Port
- (4) Seleziona come "*non Designated Port*"
- (5) Scadenza Forward Delay timer

# Stati delle porte

## ■ Listening

- Riceve trame
- Non inoltra trame
- Non aggiorna forwarding data base
- Elabora BPDU ricevute
- Ritrasmette BPDU

## ■ Forwarding

- Riceve trame
- Inoltra trame
- Aggiorna forwarding data base
- Elabora BPDU ricevute
- Ritrasmette BPDU

## ■ Learning

- Riceve trame
- Non inoltra trame
- Aggiorna forwarding data base
- Elabora BPDU ricevute
- Ritrasmette BPDU

## ■ Blocking

- Riceve trame
- Non inoltra trame
- Non aggiorna forwarding data base
- Elabora BPDU ricevute
- Non ritrasmette BPDU

# Cambiamento topologico

Byte	
1÷2	Protocol Identifier: 00-00
3	Protocol Version Identifier: 00
4	BPDU Type: 80

- Filtering database può non essere aggiornato
  - Eliminazione delle entry per garantire raggiungibilità grazie al flooding
  - Apprendimento di nuove entry
- Il bridge che rileva un cambiamento invia una Topology Change Notification BPDU attraverso la root port
  - I bridge che la ricevono la inoltrano sulla propria root port
- Il root bridge risponde con una Configuration BPDU con il bit **topology change** impostato
  - I bridge che la ricevono dalla root port la inoltrano impostando il bit **topology change acknowledgment**
- I bridge, che rilevano il cambiamento di topologia, eliminano le entry dopo un tempo forward delay

# Rilevamento di un cambiamento topologico

- Rilevamento di un malfunzionamento di livello fisico
  - Fallimento del Link Integrity Test
- Mancata ricezione periodica di Configuration BPDU
  - Una porta in stato di blocking inizializza un timer al valore del parametro max age
  - Se il timer scade, la porta passa in stato listening
    - Il collegamento dal root bridge alla LAN ha problemi
  - Allo scadere del forward delay passa in stato learning
  - Allo scadere del forward delay passa in stato forwarding
    - Transizioni ritardate per evitare oscillazioni
  - Recupero del guasto richiede 50 secondi

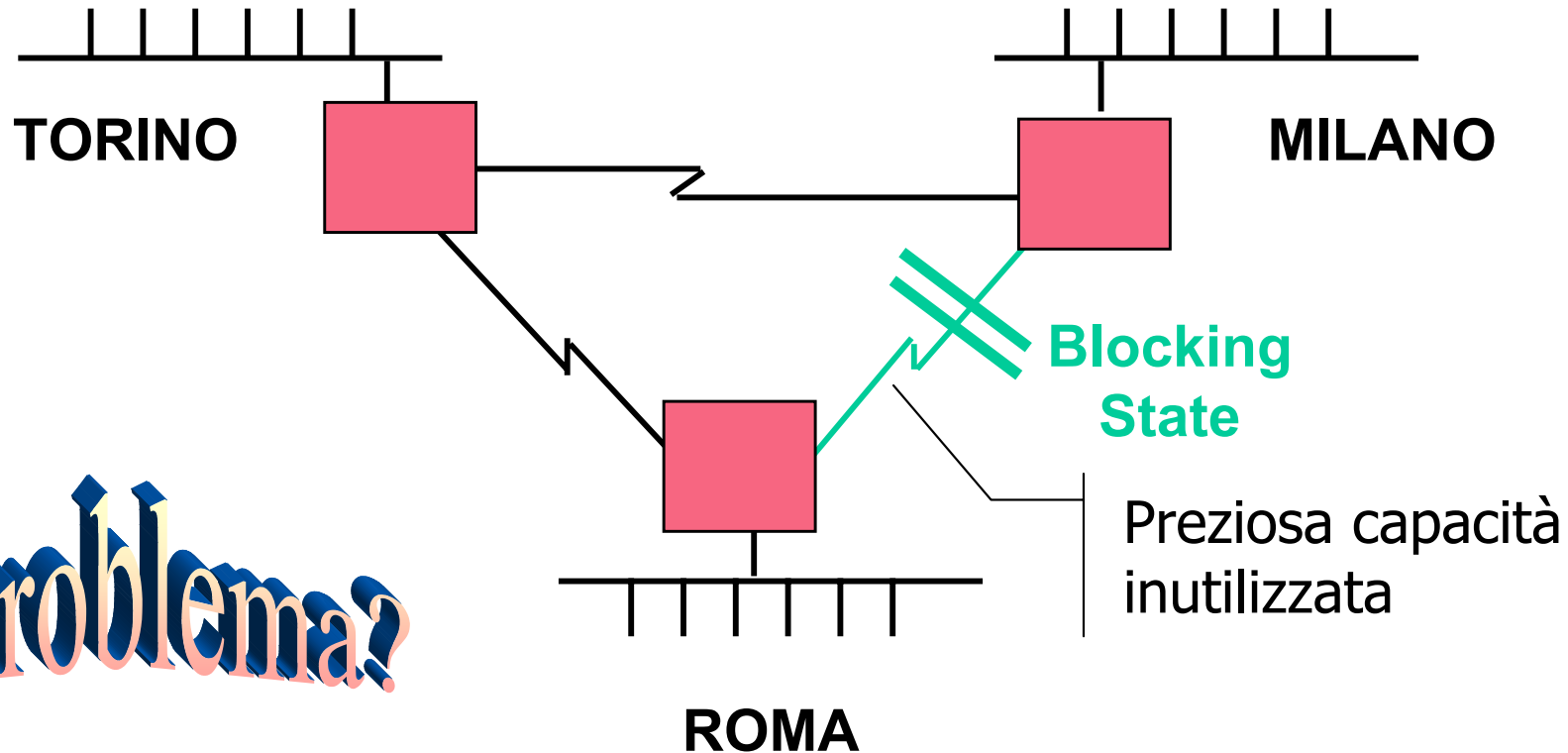
# Taratura dei timer

- Valori raccomandati assicurano corretto funzionamento con 7 bridge in cascata
  - In IEEE 802.1D si parla di max bridge diameter
- Modificando i timer si possono
  - Ridurre i tempi di convergenza
  - Aumentare il diametro massimo della rete
- Non facile da realizzare in modo efficace
  - Valori non ottimali possono
    - Peggiorare la reattività della rete ai cambiamenti topologici
    - Compromettere il buon funzionamento della rete (Loop!!!)
- Difficile da gestire
  - Se si cambia l'apparato root bridge è importante cambiare il valore dei timer

# Taratura dei timer

- IEEE 802.1D definisce i calcoli tramite cui ottenere un valore ottimale per i parametri a partire da
  - `max bridge diameter` → numero massimo di bridge in cascata
  - `maximum bridge transit delay` → massimo tempo che una BPDU impiega ad attraversare un bridge
    - Dalla ricezione alla ritrasmissione, includendo l'elaborazione
- Si deve diminuire il `maximum bridge transit delay`
- Calcolare il valore risultante per i vari timer
  - Normalmente `hello time` è il doppio del `maximum bridge transit delay`
  - In molti casi di ottimizzazione `hello time` viene posto a 1 s

# Spanning Tree su LAN estese



il problema?

Unico albero per tutto il traffico

La soluzione?

Albero per mittente

# Un grosso problema: collegamenti unidirezionali

- Le BPDU sono propagate da root a foglie
  - Non bidirezionalmente
- Se si rompe la direzione di propagazione, l'altro estremo considera la porta designated
  - Bridge rotto non se ne accorge
  - Si crea un loop
- Può essere temporaneo, ma genera broadcast storm

# Broadcast storm a seguito di guasto

