

Nätverksteknik B - Redundans på lager 2

Lennart Franked

Informationssystem och -teknologier (IST)
Mittuniversitetet

28 januari 2019

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 that must be sure to span
so packet 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.

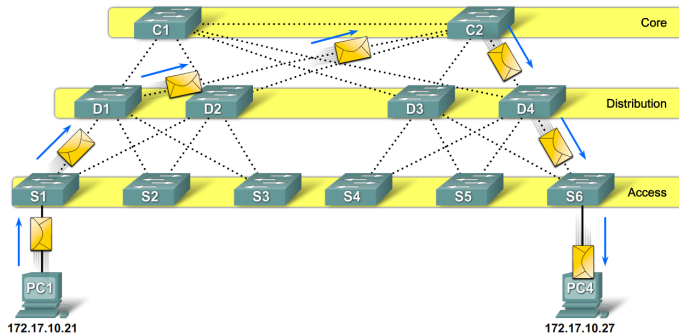
—

Radia Perlman

- Ger fysisk redundans i ett switchat nätverk.
- Ökar tillförlitligheten samt tillgängligheten i nätverket.

Problem

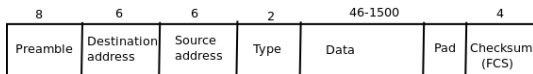
Redundans



Figur: Redundant nätverk[2]

Problem

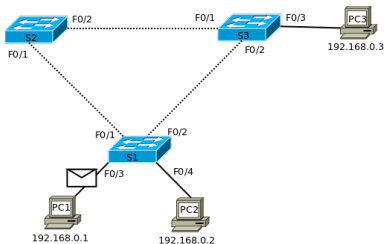
- Redundans orsakar loopar på lager 2
- Ethernetramar *saknar* ett TTL field.



Figur: DIX Ethernet header

Broadcast loop

Redundans



Figur: Broadcast loop[2]

Tabell: S1 MAC table

Destination	Port
PC1	F0/3

Tabell: S2 MAC table

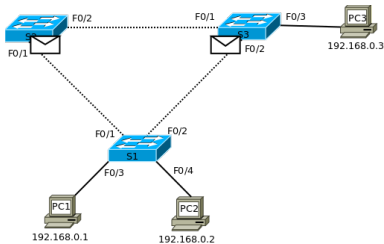
Destination	Port

Tabell: S3 MAC table

Destination	Port

Broadcast loop

Redundans



Figur: Broadcast loop[2]

Tabell: S1 MAC table

Destination	Port
PC1	F0/3

Tabell: S2 MAC table

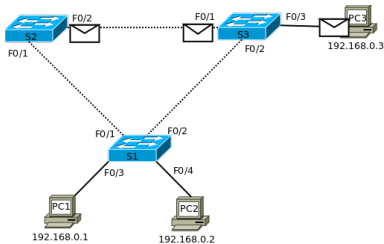
Destination	Port
PC1	F0/1

Tabell: S3 MAC table

Destination	Port
PC1	F0/2

Broadcast loop

Redundans



Figur: Broadcast loop[2]

Tabell: S1 MAC table

Destination	Port
PC1	F0/3

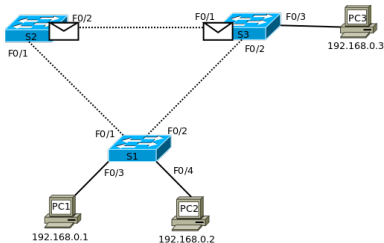
Tabell: S2 MAC table

Destination	Port
PC1	F0/1

Tabell: S3 MAC table

Destination	Port
PC1	F0/2

Broadcast loop



Figur: Broadcast loop[2]

Tabell: S1 MAC table

Destination	Port
PC1	F0/3

Tabell: S2 MAC table

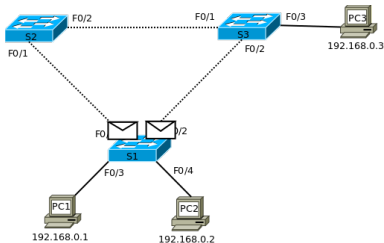
Destination	Port
PC1	F0/2

Tabell: S3 MAC table

Destination	Port
PC1	F0/1

Broadcast loop

Redundans



Figur: Broadcast loop[2]

Tabell: S1 MAC table

Destination	Port
PC1	F0/1

Tabell: S2 MAC table

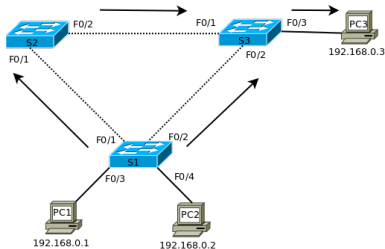
Destination	Port
PC1	F0/2

Tabell: S3 MAC table

Destination	Port
PC1	F0/1

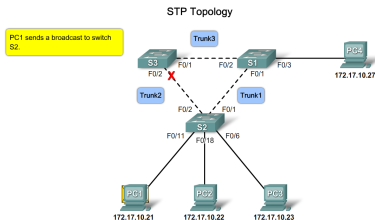
Resultat av lager 2 loopar:

- Broadcast stormar
- Dubbla unicasts
- Congestions
 - ▶ Packetförluster
 - ▶ Ökad fördröjning



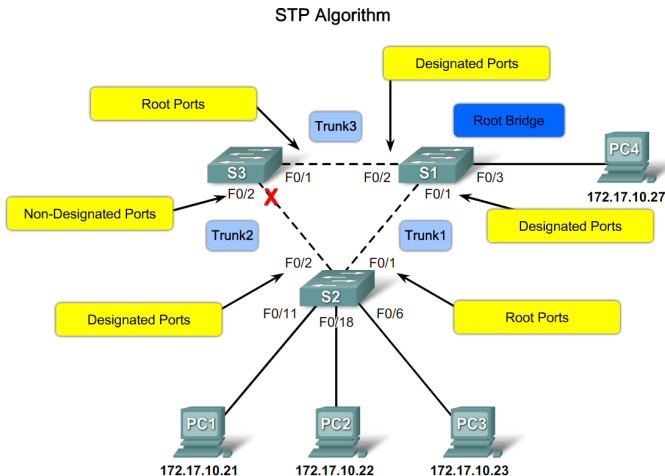
Figur: Dubbla unicasts

- Utvecklad av Radia Perlman i början av 80-talet
- Publicerad som IEEE 802.1D 1990
- Försäkrar om att enbart en logisk väg finns mellan samtliga noder på nätverket.
- Blockerar redundanta vägar.



Figur: Spanning Tree Protocol[2]

Spanning-tree Algorithm



Figur: Översikt – Spanning-tree Algorithm[2]

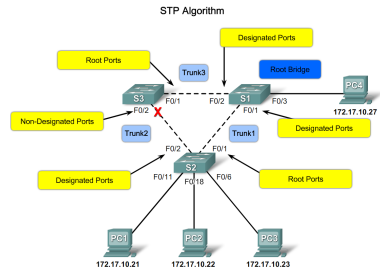
Huvudkomponenterna i STP

- Root bridge
- Spanning-tree Algorithm (STA)
- Port Roles
- Bridge Protocol Data Unit (BPDU)

Rot-brygga

- Agerar referenspunkt för uträkningen av spanning-tree.
- Utses genom en valprocess baserad på ett Bridge-ID (BID).
- Lägsta ID blir rot-brygga.
- BID utbyts med hjälp utav BPDU-ramar.

- Spanning-Tree Algorithm
bestämmer vilka portar som
måste blockeras för att förhindra
uppkomsten av en loop.
- Tre typer av portar.
 - ▶ Root port
 - ★ Närmast rot-bryggan.
 - ▶ Designated port
 - ★ Icke-rot port. Används för
att vidarebefordra ramar.
 - ▶ Nondesignated port
 - ★ Blockerad port för att
förhindra routing loopar.
 - ▶ Disabled port
 - ★ Administrativt shut down.



Figur: Spanning Tree Protocol[2]

- Beräknar bästa vägen till rot-bryggan.
- Varje länk mellan två switchar ges en kostnad.
- Vägen till rotbryggan är summan av alla utgående portars kostnad.
- Bästa vägen till roten är den väg med lägst kostnad.

Tabell: Path cost

Link	Cost (revised IEEE)	cost (old IEEE)
10Gbit/s	2	1
1Gbit/s	4	1
100Mbit/s	19	10
10Mbit/s	100	100

- Första 4 fälten identifierar ramtyp.
- Nästa 4 fält identifierar rotbryggan och kostnaden att nå denna.
- Sista 4 fält innehåller uppdateringsinformation.
- BPDU skickas varannan sekund på alla trunkportar.

The BPDU Fields

Field #	Bytes	Field
4	2	Protocol ID
	1	Version
	1	Message type
	1	Flags
8	8	Root ID
	4	Cost of path
	8	Bridge ID
	2	Port ID
12	2	Message age
	2	Max age
	2	Hello time
	2	Forward delay

Figur: En BPDU-ram[2]



- BPDU Timers.

- ▶ Hello time - BPDU sändningsintervall
- ▶ Forward delay - Tid spenderad in lyssnings och inlärningsläget.
- ▶ Maximum age - Hur länge en switch ska spara BPDU-information

BPDU Timers

Hello time	The hello time is the time between each BPDU frame that is sent on a port. This is equal to 2 seconds by default, but can be tuned to be between 1 and 10 seconds.
Forward delay	The forward delay is the time spent in the listening and learning state. This is by default equal to 15 seconds for each state, but can be tuned to be between 4 and 30 seconds.
Maximum age	The max age timer controls the maximum length of time a switch port saves configuration BPDU information. This is 20 seconds by default, but can be tuned to be between 6 and 40 seconds.

Figur: BPDU Timers[2]

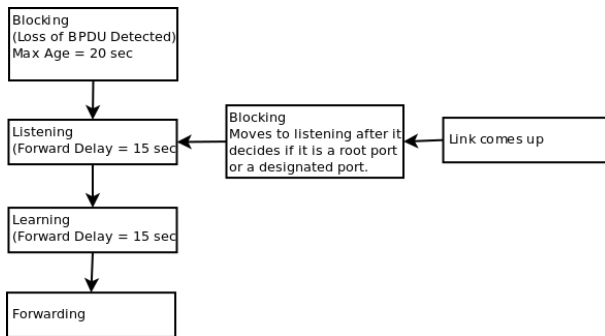
- Varje port som är involverad i STP, går igenom fem tillstånd.
 - Blocking - Deltar inte i vidarebefordring av ramar.
 - Listening - Utser en rotbrygga, rotport, designerad port, icke-designerad port.
 - Learning - Förbereder sig att delta i vidarebefordring av ramar.
 - Forwarding - Deltar i den aktiva STP-topologin.
 - Disabled - Administrativt down.

Port States

Processes	Blocking	Listening	Learning	Forwarding	Disable
Receives and process BPDUs	✓	✓ ¹	✓	✓	✗
Forward data frames received on interface	✗	✗	✗	✓	✗
Forward data frames switched from another interface	✗	✗	✗	✓	✗
Learn MAC addresses	✗	✗	✓	✓	✗

¹Return to blocking if not lowest cost path to root bridge

Figur: Porttillstånd[2]



Figur: STP port tillståndsovergångar [2]

Cisco PortFast

- Övergång från 'blocked' till 'forwarding' sker direkt.
- Minskar tiden som en accessport måste vänta.
- Används på accessportar anslutna till en slutnod.

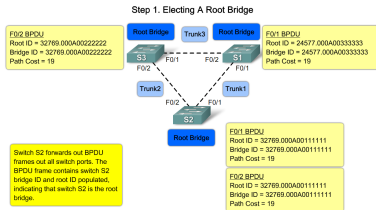
Example

```
Switch(config-if)#spanning-tree portfast
```

STP konvergeringsprocess kan ses som tre etapper.

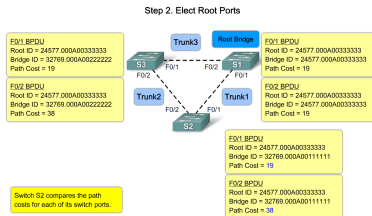
- Utse rotbryggan.
- Utse the rotportar.
- Utse designerad och icke-designerade portar.

- Baserad på den lägsta BID.
- Om två switchar har samma prioritet. Lägst MAC-adress kommer avgöra.



Figur: Utse rot[2]

- Jämför kostnaden mot rotbryggan.
- Den port som har lägst kostnad till rotbryggan utses till rotport.
- Om kostnaden är identisk, används portprioriteten.
- Om portprioriteten är identisk, används lägst port-nummer.



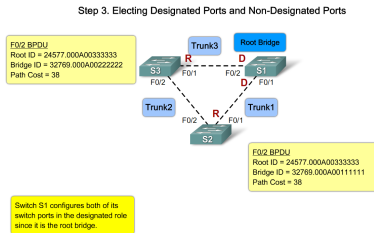
Figur: Elect root port[2]

Designerad och Icke-designerad port

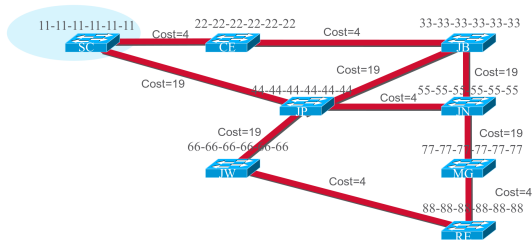
STP

Länken mellan två
icke rotbryggor
jämnförs...

- Port med lägst kostnad till rot blir designerad port.
- När flera portar har samma kostnad för att nå rotbryggan:
 1. Lägst sändande BID
 2. Lägst sändande PID
- Port med den högsta kostnaden till rot blir

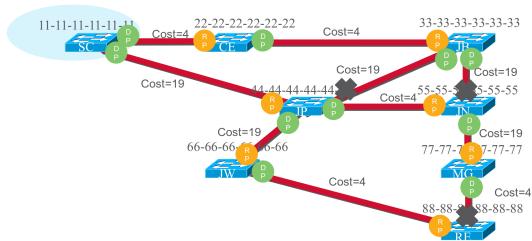


Figur: Utse designerad port[2]



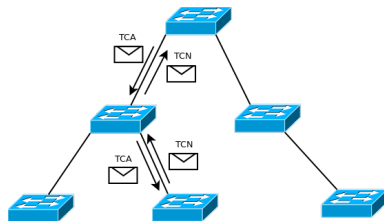
Figur: Topologi STP [1]

- Beräkna Rotport.
- Beräkna Designerad port.
- Sätt ett kryss över ej använda länkar.



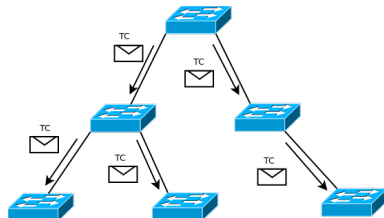
Figur: Topologi STP [1]

- Om en förändring i topologin sker. Den switch som först märker detta informerar rotbryggan, med hjälp utav TCN (Topology Change Notification)
- Varje Switch som tar emot en TCN, svarar med en TCA (Topology Change Acknowledgment)
- Dessa ramar skickas tills dess att roten svarar.



Figur: Svar på topologiförändringar[2]

- Rotbryggan svarar med en 'Topology Change' ram.
- Varje switch som mottar detta, ändrar age-timer till samma värde som forward delay (standard 15 sekunder).
- Detta gör att switchen kan lära om sig vilka mac-adresser som finns tillgängliga på nätverket.



Figur: Topology Change[2]

Cisco and STP Variants

Cisco Proprietary	PVST • Uses the Cisco proprietary ISL trunking protocol • Each VLAN has an instance of spanning tree • Ability to load balance traffic at layer-2 • Includes extensions BackboneFast, UplinkFast, and PortFast
	PVST+ • Supports ISL and IEEE 802.1Q trunking • Supports Cisco proprietary STP extensions • Adds BPDU guard and Root guard enhancements
	rapid-PVST+ • Based on IEEE802.1w standard • Has faster convergence than 802.1D
IEEE Standard	RSTP • Introduced in 1982 provides faster convergence than 802.1D • Implements generic versions of the Cisco proprietary STP extensions • IEEE has incorporated RSTP into 802.1D, identifying the specification as IEEE 802.1D-2004
	MSTP • Multiple VLANs can be mapped to the same spanning-tree instance • Inspired by the Cisco Multiple Instances Spanning Tree Protocol (MISTP), • IEEE 802.1Q-2003 now includes MSTP

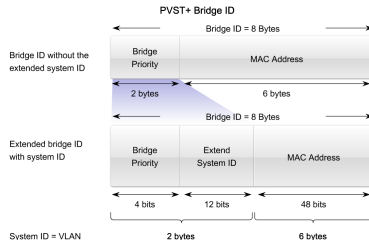
Figur: Varianter av STP[2]

Per VLAN Spanning Tree +

PVST+

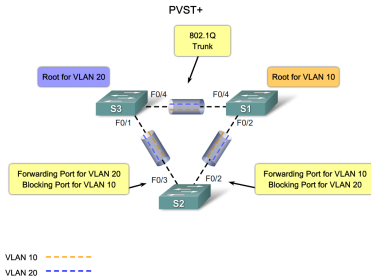
PVST+

- Stödjer multipla VLAN.
- Endast ett träd per VLAN.
- Tillåter oss att använda olika rotbryggor för varje VLAN.
- Standard STP-protokoll på en Cisco Catalyst switch.



Figur: BPDU Bridge ID[2]

Per VLAN Spanning Tree + PVST+



Figur: PVST+[2]

Example (Macro)

```
Switch1(config)#spanning-tree  
vlan 10 root primary  
Switch1(config)#spanning-tree  
vlan 20 root secondary
```

Example (Static)

```
Switch1(config)#spanning-tree  
vlan 10 priority 0  
Switch1(config)#spanning-tree  
vlan 20 priority 4096
```

IEEE 802.1w

- Slutgiltig version publicerades 2002-02-07
- Det protokoll som föredras att använda för att förhindra L2 loopar.
- Baserad på många förbättringar Cisco gjorde på STP.
- Bakåtkompatibelt med STP.
- Samma process för att utse rotbrygga som i 802.1D.
- Samma BPDU format.

Tabell: Porttillstånd

STP (802.1D) Port State	RSTP (802.1W) Port State	Port included in active topology	Is Port learning MAC Addresses?
Disabled	Discarding	No	No
Blocking	Discarding	No	No
Listening	Discarding	Yes	No
Learning	Learning	Yes	Yes
Forwarding	Forwarding	Yes	Yes

RSTP Portroller

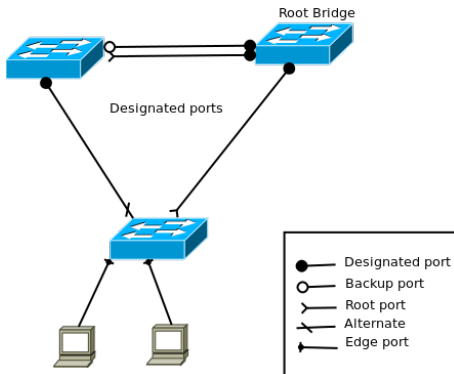
- Rotport
 - ▶ Den port som mottar bästa kostnaden mot rotbryggan.
 - ▶ Närmast roten sett utifrån ett kostnadsperspektiv.
- 'Designated Port'
 - ▶ Den port som skickar den bästa BPDU på det segment som den är kopplad till.
- 'Alternate'
 - ▶ Tar emot en BPDU som är bättre än sin egen, men sämre än den BPDU som tas emot från rotporten.
 - ▶ BPDU som kommer från en annan brygga.
 - ▶ Alternativ väg till roten.
 - ▶ Blockerad port.

RSTP Portroller

- Backup
 - ▶ Tar emot en bättre BPDU än sin egen, men sämre än den som tas emot från rotporten.
 - ▶ Båda BPDUerna kommer från samma brygga.
 - ▶ Blockerad port.
 - ▶ Redundant länk till roten.

Accessport

- Edge Port
 - ▶ Samma funktion som Cisco's PortFast.
 - ▶ Slutar agera 'edge port' om en BPDU tas emot på denna port.



Figur: RSTP Länktyper

BPDU in RSTP

- 802.1D bryggor vidarebefordrade enbart BPDU.
- Gjorde det omöjligt att veta ifall en rotbrygga eller designerad brygga var nere.
- Varje brygga skickar sin egen BPDU i RSTP
- Fungerar som en 'keep-alive' mekanism..
- Tillåter att fel upptäcks snabbt.

Rapid Transition

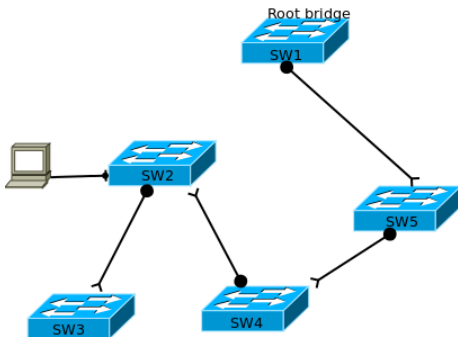
- När 802.1D används, måste en designerad port vänta $2 \times \text{forward delay}$ innan den kan övergå till 'forwarding state'
- RSTP håller koll på vilken typ av länk som används, och ifall det är möjligt att göra en snabb övergång.
- Två variabler används för att bestämma ifall en snabb övergång kan ske eller inte:
 - ▶ Edge port
 - ▶ Link type
 - ★ point-to-point (Full-duplex)
 - ★ shared (Half-duplex)

Länktyper

- Snabb övergång endast möjligt på 'edge-portar' och Point-to-Point länkar.
- På delade länkar används timers.

Proposal and agreement process i RSTP

RSTP

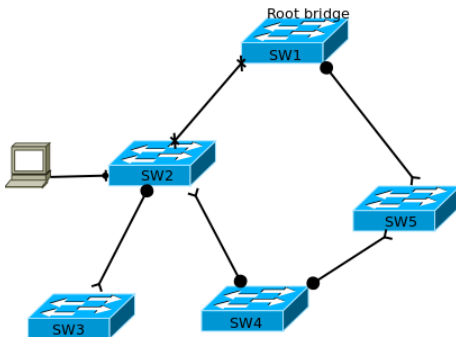


- En länk skapas mellan SW1 och SW2.

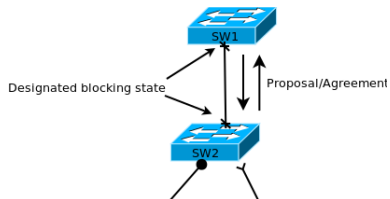
Figur: RSTP topology

Proposal and agreement process i RSTP – SW1 <-> SW2

RSTP



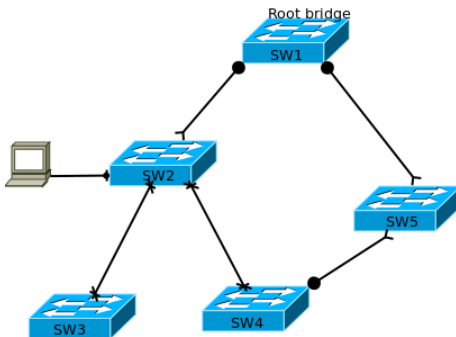
Figur: RSTP topology



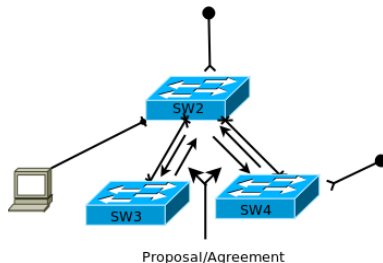
Figur: Förhandling mellan SW1 och SW2

Proposal and agreement process i RSTP – SW2 <-> SW3,SW4

RSTP



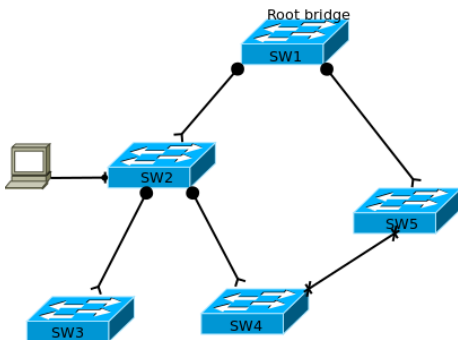
Figur: RSTP topology



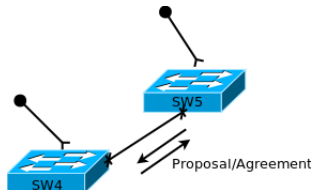
Figur: Förhandling mellan SW2 - SW3 och SW2 - SW4

Proposal and agreement process i RSTP – SW4 <-> SW5

RSTP



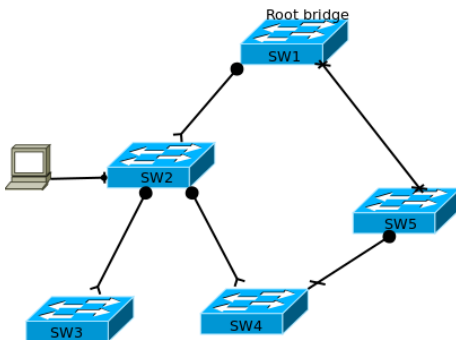
Figur: RSTP topology



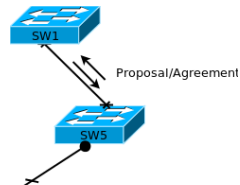
Figur: Förhandling mellan SW4 och SW5

Proposal and agreement process i RSTP – SW5 <-> SW1

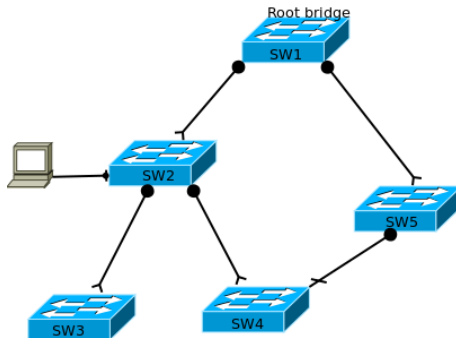
RSTP



Figur: RSTP topology



Figur: Förhandling mellan SW5 och SW1



Figur: RSTP topology



Cisco Networking Cisco Networking Academy. *Scaling Networks V6 Companion Guide*. Companion Guide. Cisco Press, 2017. ISBN: 9781587134340.



Scaling networks : companion guide. 2014.