

Spanning-Tree Protocol

STP är ett nätverksprotokoll som används för att eliminera loopar i L2 nät. Artikeln beskriver först klassisk spanning-tree och sedan nyare varianter. I STP så beskrivs switchar som bryggor.

Beskrivning av STP

STP gör detta genom att se till att vilka interface ej tar emot eller skickar trafik. Dessa portar hamnar i ett läge som kallas **Blocking**. Portar som fortfarande skickar och tar emot trafik är i liget **Forwarding**.

STP protokollmeddelanden kallas för Bridge Protocol Data Units (BPDU). Det finns två typer av BPDUs, Configuration BPDU (typ 0x00) och Topology Change Notification BPDU (typ 0x80). En Configuration BPDU kan även kallas för **Hello BPDU** eller helt enkelt **Hellos**.

Se bild på vad som ingår i en BPDU:

image.png

Switchar och portar identifieras baserat på sitt ID i BPDUn. ID:t har två delar, en del som kan programmeras som är *priority* och en som är statisk som ej kan konfigureras. Både switchar och portars prioritet kan konfigureras.

STP baseras på möjligheten att jämföra två eller flera Configuration BPDUs och välja den som är bättre - den som är **superior**. För att räkna ut vilken som är superior letar STP efter det först förekommande lägsta värdet av följande parametrar från top till botten:

- Root Bridge ID (RBID)
- Root Path Cost (RPC)
- Sender Bridge ID (SBID)
- Sender Port ID (SPID)
- Receiver Port ID (RPID, ingår ej i en BPDU utan räknas ut lokalt)

När det först förekommande lägsta värdet hittas så avslutas processen och den bästa Configuration BPDUn har hittats.

STP-processen

För att välja vilka portar som trafik ska forwardas på och vilka trafik ska blockas på så följer STP en trestegsprocess:

1. Välj root switch. Den switch med lägst bridge ID blir root.
2. Uträkning av root-port på övriga switchar. Det är portar på switchar som ej är root som leder till root. Den bästa BPDUn har tagits emot på denna port.
3. Välj Designated Port för varje nätsegment. Vid flera inkopplingar till samma nätsegment så är det denna port som BPDUerna skickas ut på.

Bridge ID

Formatet av Bridge ID har förändrats från den första versionen av 802.1D. Idag så definieras vanliga STP i 802.1D-2004.

image.png

Formatet ändrades för att få stöd för flera spanning-tree instanser per switch utan att behöva reservera en MAC-adress per VLAN, nu återanvänds samma MAC-adress för att bygga Bridge IDn. Extension konfigureras med globala kommandot `spanning-tree extend system-id`. Äldre switchar tillåter att man tar bort kommandot och går tillbaka till gamla formateringen. Nya switchar tillåter inte att kommandot tas bort från konfigurationen.

Prioritet konfigureras per VLAN med det globala konfigurationskommandot `spanning-tree vlan 1337 priority 4096`. Det går att specificera flera VLAN i samma rad, ex. `spanning-tree vlan 50-51,1337 priority 4096`. Prioritetsvärdet sätts i steg på 4096. Lägst prioritet är bäst. Default är 32768.

```
catalyst9300-lab(config)#spanning-tree vlan 1337 priority 500
% Bridge Priority must be in increments of 4096.
% Allowed values are:
0    4096 8192 12288 16384 20480 24576 28672
32768 36864 40960 45056 49152 53248 57344 61440
```

Best practice är att inte använda sig av prioritet 0 utan något högre än så, ex 8192. Väljer man 8192 kan man alltid sätta in en till brygga med lägre prioritet.

Val av root switch

Enbart en switch kan vara root i en STP domän. När en switch startar och initierar STP så antar den att den själv är root, den skapar Configuration BPDUn därefter och skickar ut dessa. Om switchen tar emot en bättre Configuration BPDUn så slutar den anse sig själv som root och slutar skicka Configuration BPDUs, istället så forwardar den de bättre BPDUerna från root-switchen.

Uträkning av root-port

När root-switchen är vald måste samtliga andra switchar räkna ut sin väg till root-switchen, alltså bestämma vilken port som är root-port. Processen är enligt följande:

1. Root-switchen skickar ut Configuration BPDUs varannan sekund per default (Hello Time-värdet). BPDUn innehåller root-switchen ID i RBID och SBID. RPC är satt till 0 och SPID är egress-interfacet.
2. Varje icke-root switch som tar emot en configuration BPDUn på en viss port lägger till portens **cost** till RPC-värdet i den mottagna BPDUn, vilken resulterar i en **resulting BPDUn**. Den port som har tagit emot den bästa Configuration BPDUn blir Root Port.
3. Configuration BPDUs som mottages på Root portar skickas vidare på samtliga designated portar efter uppdatering av RPC, SBID, SPID och MessageAge värdena. Configuration BPDUs som mottages på andra portar analyseras men skickas ej vidare.
4. Configuration BPDUs skickas ej vidare ut på Root portar eller ut på portar som stabiliseras till Blocking state. Configuration BPDUs som eventuellt skulle skickas ut på dessa interface skulle vara ointressanta för mottagaren.

Uträkning av Designated-port

Ut på ett Ethernet-segment ska det bara finnas en Designated-port. I andra änden av en Designated-port ska läget vara Root-port. Enbart Designated-ports skickar vidare Configuration BPDUs. Portar som inte är Root-port eller Designated-port flyttas till Blocking.

Costs

Ett interface tilldelas en spanning-tree kostnad som används för uträkningen av RPC (Root Path Cost). De äldre standarderna av spanning-tree ser inte skillnad på cost på ett 1 gbit/s interface mot ett 10 gbit/s interface, det har uppdaterats i senare versioner. För att använda sig av 802.1D-2004 costs globalt i switchen kan man aktivera det med kommandot `spanning-tree pathcost method long`. Per default är `spanning-tree pathcost method short` aktiverat.

Länkhastighet	Pre-802.1D-1998 Cost	802.1D-1998 Cost	802.1D-2004 Cost
10 Mbit/s	100	100	2000000
100 Mbit/s	10	19	200000
1 Gbit/s	1	4	20000
10 Gbit/s	1	2	2000
100 Gbit/s	x	x	200
1 Tbit/s	x	x	20

Det går även att tilldela ett interface en cost manuellt med kommandot `spanning-tree cost x` i interface-konfigläge. Det går även att sätta per VLAN med `spanning-tree vlan 10 cost x`.

Konvergering

När en STP domän har stabiliserat sig efter uppdatering så har den konvergeras. Beräkningar sker dock varje gång en Configuration BPDUn mottages, vilket alltså sker per default varannan sekund då en ny Configuration BPDUn genereras av root-switchen enligt Hello Time-värdet.

Topologin kan, och kommer, förstås att förändras. I STP-termer är det en **topology change** som inträffar när:

- En Topology Change Notification BPDU tas emot på en Designated port i en switch
- En port förflyttas till Forwarding state och switchen har minst en Designated Port
- En port förflyttas från Learning eller Forwarding till Blocking
- En switch blir root

Switchar som märker en förändring i topologin börjar skapa BPDUs med uppdaterad information. Grannar som tar emot dessa behandlar de som vanligt, processar och ev. skickar ut på designated ports.

När STP konvergeras enligt den nya topologin så kan det finnas poster i Content Address Memory (CAM, mac address table) som inte längre stämmer. STP kan då berätta för switchen att oanvända poster ska rensas i förtid. Det måste ske på samtliga switchar och en kort timer måste användas som är likk Forward Delay timer, default 15 sekunder. En Topology Change Notification (TCN) BPDU skickas från switchen som märkte topologiförändringen till root. TCN BPDU tas emot på ev. designated ports och vidarebefodras ut på root-portar. Acknowledgements skickas tillbaka för varje hopp genom att märka Topology Change Acknowledgment (TCA) bit i Configuration BPDUs. Root-bryggan skapar vid mottagande av TCN BPDU en Configuration BPDU med Topology Change (CN) bit satt till 1 vilket instruerar alla switchar att minska aging time av CAM poster till sekunderna i ForwardDelay timer.

Interface-lägen

I STP finns flera interface states. Några av dem har redan nämnts här innan. Det finns fler states i klassisk STP än i Rapid STP.

Vid konvergering så kan det finnas Designated Ports eller Root Ports som ska gå från Blocking till Forwarding. Men för att inte skapa några loopar så lyssnar de, Listening, först ett tag och lär sig, Learning, innan de blir forwarding. I klassisk spanning-tree så varar varje övergående tillstånd i 15 sekunder.

State	Växlar data paket?	Lär sig source MAC-adresser från mottagna paket?	Övergående eller stabilt tillstånd?
Blocking	Nej	Nej	Stabilt
Listening	Nej	Nej	Övergående
Learning	Nej	Ja	Övergående
Forwarding	Ja	Ja	Stabilt
Disabled	Nej	Nej	Stabilt

I klassisk 802.1D spanning-tree gäller interface tillståndet en hel länk.

Exempelkonfiguration 802.1D

```
spanning-tree vlan 1 priority 28672
```

```
interface FastEthernet0/1
```

```
spanning-tree vlan 1 cost 100
```

Show kommandon 802.1D

Kommando	Förklaring
show spanning tree vlan 1337 [detail]	Se status för ett visst VLAN i en STP instans
show spanning tree root	Se vägar till rootbrygga
show spanning-tree [detail]	Information om samtliga STP instanser

STP Standarder

802.1D, Spanning-Tree Protocol

802.1D är den första standarden för L2-bridging vilket inkluderade den första versionen av STP. 802.1D STP är idag föråldrad och är ersatt med 802.1D-2004 STP. 802.1D och 802.1D-2004 har beskrivits ovan.

Per-Vlan Spanning Tree

PVST är en Cisco proprietär implementation av STP. Numera är PVST+ som används. Interface-lägen kan i PVST+ förhandlas per VLAN över en trunk. Det gör att det går att lastbalansera VLAN över olika länkar. VLAN 1 STP instans används för kommunikation med switchar som inte klarar av PVST+. Formatteringen av 802.1D och PVTST+ STP paket skiljer sig och därmed kan switchen avgöra om den ska prata PVST+ full ut eller använda sig av VLAN 1.

802.1w, Rapid STP

Det som en gång var 802.1w har nu införlivas i 802.1D, då den första standarden för STP är föråldrad och används inte längre. Målet med RSTP är att förbättra STPs konvergeringstid till under en sekund. Portar har i RSTP porttillstånd (state), portroller (roles) och porttyper (types). Antalet

porttillstånd har förändrats från 5 i 802.1D STP till 3 i RSTP

Porttillstånd

Administrativt tillstånd	Porttillstånd STP 802.1D	Porttillstånd RSTP
Disabled	Disabled	Discarding
Enabled	Blocking	Discarding
Enabled	Listening	Discarding
Enabled	Learning	Learning
Enabled	Forwarding	Forwarding

I RSTP innebär Discarding att ingen trafik switchas, tas emot eller att source MAC-adresser installeras i CAM. En port i Discarding processar dock fortfarande BPDUs, beroende på roll skickar ut BPDUs och tar emot inter-switch signalprotokoll såsom DTP, VTP, CDL, LLDP, PAgP, LACP eller LOOP. Discarding är standardläge för nya portar som tas upp med undantag för Edge-portar som direkt är i Forwarding.

I RSTP är porttillstånd och portroller två olika saker. De fyra portroller som finns är:

Root port	Samma som i 802.1D
Designated port	Samma som i 802.1D
Alternate port	En möjlig ersättare för switchens Root Port. Tar emot BPDUs som är sämre än de som kommer via Root-porten. Vid händelse att Root-porten går ned blir Alternate befodrad till Root direk
Backup port	En möjlig ersättare för switchens Designated Port. Är redo att bli Designated port men blir inte det direkt utan väntar på timers.

Porttyper och länktyper

Porttyper i 802.1w är antingen *Edge port* eller *Non-Edge port*. En edge port konfigureras med `spanning-tree portfast` i interface-konfigurationsläge. En Edge Port är direkt i Forwarding läge. BPDUs skickas ut på en Edge port men tas BPDUs emot så stängs den ned och går igenom RSTP-processen, Discarding, Learning och till slut åter i Forwarding. Portar är per default Non-Edge.

Best practice är att samtliga portar som inte ska ansluta till andra switchar konfigureras som Edge-portar. Det gör att användaren slipper vänta på RSTP-uträkning när de kopplar in något i nätverksuttaget och det undviker eventuell nedtid orsakad av Proposal/Agreement processen som beskrivs längre ned.

Det finns även två länktyper, *Point-to-point link* och *Shared link*. PTP-länkar är en länk till en annan switch och Shared är till ett segment med två eller flera switchar. I moderna nätverk är länkar PTP.

I äldre nätverk där interfacet går upp i halv-duplex så blir länktypen Shared. Det sker om en hub sitter inkopplad mellan switcharna. Länktypen kan konfigureras på interfacet med `spanning-tree link-type { point-to-point | shared }`.

BPDUer

I 802.1w finns bara en BPDU-typ. TCN används ej av RSTP. I STP så skickas Configuration BPDUs av rootswitchen. I RSTP så skickas Configuration BPDUs av samtliga RSTP-switchar, det går alltså att jämföra med Hello-paket i ett routingprotokoll. Om en switch slutar tar emot Configuration BPDUs på ett interface kan switchen med säkerhet anta att felet på är på den länken och snabbt förklara de emottagna sparade BPDUerna som irrelevanta och ta bort dem efter 3 gånger så lång tid som Configuration BPDU intervallen.

Hanteringen av sämre BPDUer har förbättrats gentemot STP. I RSTP behöver switchen inte vänta på att de bättre BPDUerna time:ar ut enligt MessageAge-MaxAge värdena. Skulle det mottagna BPDU-värdet på ett interface förändras, ex. bli bättre eller sämre, så accepteras det direkt och uträkning för porttillstånd sker.

Proposal/Agreement Process

Förbättringar har även skett för tillägg av nya länkar som har en bättre kostnad till root. För att undvika loopar vid inkoppling av nya länkar finns Proposal/Agreement processen. För att undvika loopar måste portrollsförändringen ske på den switch som förändrar vilket interface som är mot root. Switchen kan placera alla Non-Edge Designated ports i porttillståndet Discarding innan nya Root-porten blir Forwarding. Det skulle undvika loopar men är trafikstörande. För att undvika att vänta på ForwardDelay-timern två gånger (uträkning för root på switch 1 och designated på switch 2) så skickas signaler mellan switcharna när det är säkert att placera Designated-ports i Forwarding med *Proposal/Agreement*.

När en ny PTP länk tas upp mellan switchar är interfacen per default i Designated Discarding. BPDUs skickas ut med Proposal-bit:en flaggad. Om en av portarna märker att BPDUn är bättre ska portrollen bli Root. En RSTP-switch som tar emot en bättre BPDU på root-interfacet sätter per automatik samtliga Non-Edge Designated ports i läget Discarding, proceduren kallas för *Sync*. En switch i sync är i princip isolerad från nätverket och orsakar inga loopar. Nu är det säkert att flytta den nya Root-porten från Discarding till Forwarding och signalera till överliggande switch att den även kan byta sitt tillstånd på Designated porten till Forwarding. Det genomförs genom att skicka en Configuration BPDU med Agreement-biten flaggad efter sync, vid mottagande av BPDUn så flyttar överliggande switch porttillståndet till Forwarding direkt.

Nu kommer Proposal/Agreement processen att upprepas på eventuella Non-Edge portar nedströms.

Värt att komma ihåg från detta är att nedtid i L2-nät kan bero på direkta länkfel, indirekta länkfel eller tillägg av nya bättre länkar som kräver omräkning för att undvika loopar.

Topologiförändringar

I RSTP räknas endast förändringen från Non-edge port från Non-forwarding till Forwarding som en topologiförändring. En non-edge port som har blivit forwarding skulle kunna ha en bättre väg till MAC-adresser än vad som tidigare fanns och CAM behöver uppdateras. Förlust av en Non-edge port från forwardingläget räknas ej som en topologiförändring.

Som tidigare nämnt används ej TCN BPDUer i RSTP. Istället skickas en configuration BPDU med TC biten flaggad. När en topologiförändring märks så sker följande:

- tcWhile-timern sätts till samma värde som Hello timern plus en sekund på Root-porten och Designated-portar.
- Tömmer direkt CAM från MAC-adresser som switchen har lärt sig på dessa portar
- Skickar Configuration BPDUer ut på dessa portar enligt Hello-intervall tills att tcWhile timern går ut

Information om topologiförändringen propageras väldigt fort enligt ovan. Edge-portar skapar aldrig en topologiförändring.

Rapid Per-VLAN Spanning-Tree (RPVST+)

RPVST+ är en version av RSTP som utförs per VLAN basis. Konfigureras globalt med `spanning-tree mode rapid-pvst`. Edge-portar kan konfigureras per interface eller default globalt med `spanning-tree portfast default`.

802.1s, Multiple STP

MSTP är en IEEE industristandard implementation av STP. MSTP tillåter att man grupperar olika VLAN i olika instanser och utför STP-uträkningar inom den instansen, till skillnad från PVST där uträkning sker per VLAN. MSTP använder sig av timers från 802.1w RSTP och är därmed ett snabbt protokoll. MSTP är ett bra val om man har flera switchleverantörer i nätet eller får slut på antalet STP instanser som kan köras i ett nät.

MSTP principer

MSTP organiserar nätverket i en eller flera *regioner*. En region är en samling switchar som tillsammans använder MSTP på ett konsekvent sätt. De har samma antal MSTP instanser och kopplar samma VLAN till dessa instanser. En av de största fördelarna med MSTP kontra PVST är att enbart en BPDU behövs skickas för samtliga MSTP instanser. I PVST så skickas BPDUer per VLAN. MSTP tillåter som högst 65 instanser.

Instans 0 är en speciell instans. Den kallas för *Internal Spanning Tree* (IST). Den finns per default även om ingen annan instans är skapad, samtliga VLAN mappas till IST om inget annat konfigureras. IST är den enda instans som integrerar med STP utanför MSTP-regionen.

Interoperabilitet mellan MSTP och andra STP-versioner

För att MSTP ska fungera med andra STP-versioner måste MSTP se ut för sin granne som om den enbart kör en STP-instans. MSTP gör detta genom att använda IST/instans 0 på gränsportar och genom att skicka STP/RSTP BPDUs ut på dessa gränsportar. IST pratar alltså på alla regioner vägnar mot STP/RSTP.

Mot PVST blir det lite mer komplext, men grundtanken är densamma som ovan. En enda representant från MST och PVST väljs, MSTP utför något som kallas PVST Simulation för konsekvent interoperabilitet med PVST. I riktning från MSTP till PVST så replicerar MSTP-switchen IST BPDUs till PVST BPDUs för alla aktiva VLAN. I riktning PVST till MSTP tar MSTP VLAN 1 som representant för hela PVST regionen och processar BPDUs informationen i IST.

Konfiguration

```
spanning-tree mst configuration
name MSTP
revision 1
instance 1 vlan 1-500
instance 2 vlan 501-1000

! Går att kontrollera mapping i det här läget med show pending och show current
exit

spanning-tree mst 0 priority 0
spanning tree mst 1 priority 40962
spanning-tree mst 2 priority 8192

! Går även att synka med VTPv3
vtp version 3
vtp mode server mst
do vtp primary mst
```

Skydda och optimera STP

PortFast

En Edge-port eller PortFast-port är en port som skippar spanning-tree uträkningen och hamnar i Forwarding direkt. Skulle BPDUs tas emot på porten så stängs den. Konfigurera globalt med `spanning-tree portfast default` eller direkt på interface med `spanning-tree portfast`. Är default-läge satt stängs det av på port med `spanning-tree portfast disable`. Det går även att göra på trunkportar med `spanning-tree portfast trunk`, gör det aldrig på portar mot andra switchar. Vanligt användningsområde

för portfast trunk är mot ex. Cisco UCS eller lagringsmiljöer.

BPDU Guard

BPDU Guard är en säkerhetsfunktion som slås igång per port eller globalt per PortFast port som stänger porten när en BPDU mottages via errdisable. På interface nivå aktiveras det med `spanning-tree bpduguard enable` och globalt för samtliga portar aktiveras det med `spanning-tree bpduguard default`. Det globala kommandot aktiverar det på portar som verkar som PortFast portar. BPDU Guard kan avaktiveras perport med `spanning-tree bpduguard disable`. Automagisk errdisable recovery går att konfigurera med `errdisable recovery cause bpduguard`.

Root Guard

Root guard är en funktion som ignorerar bättre BPDUs som kommer in på en port så att den inte blir root-port. Aktiveras på portbasis. Vid mottagande av en för bra BPDU så placeras porten i Blocking tills att mottagande av bättre BPDUs på porten avstannar. Konfigureras med `spanning-tree guard root` i interface-konfigläge.

Verifiering:

```
c9500(config-if)#spanning-tree guard root
```

```
Aug 16 07:31:30: %SPANTREE-2-ROOTGUARD_CONFIG_CHANGE: Root guard enabled on port Port-channel31.
```

```
c9500#show spanning-tree interface port-channel 31 detail
```

```
Port 2311 (Port-channel31) of VLAN0007 is designated forwarding
```

```
Port path cost 20000, Port priority 128, Port Identifier 128.2311.
```

```
Designated root has priority 7, address cc70.ed70.2c00
```

```
Designated bridge has priority 7, address cc70.ed70.2c00
```

```
Designated port id is 128.2311, designated path cost 0
```

```
Timers: message age 0, forward delay 0, hold 0
```

```
Number of transitions to forwarding state: 1
```

```
Link type is point-to-point by default
```

```
Root guard is enabled on the port
```

```
BPDU: sent 2579422, received 4
```

BPDU Filter

BPDU Filter är en funktion för att begränsa sändning och mottagande av BPDUs på en port. Det går att konfigureras globalt för Edge-portar med `spanning-tree portfast bpdupfilter default`, dock skickas 11 BPDUs vid upplänk och den processar fortfarande mottagna BPDUs. Kan stängas av per Edge-

port med `spanning-tree bpdupfilter disable`. Konfigureras `spanning-tree bpdupfilter enable` i Interface-konfigläge så avstannar sändning och mottagande av BPDUer totalt.

UDLD

Unidirectional Link Detection (UDLD) är en Cisco-propertiär teknik som används för att upptäcka och eventuellt stänga länkar som har trafikfel i ena riktningen. UDLD verkar över L2. UDLD verkar i två lägen, normal och aggressiv. Om en switch slutar ta emot UDLD-paket i normal mode så försöker den att återansluta mot sin granne 8 gånger men gör ingenting med interfacet efter försöken. I aggressiv mode så försöker switchen återansluta mot sin granne 8 gånger och tar ned interface i err-disable efter återanslutningsförsöken. UDLD kan konfigureras globalt med `udld { enable | aggressive }`, där enable är normal mode, eller per interface med `udld port [ aggressive ]`. Verifiering sker med `show udld` och `show udld neighbors`. En port som har blivit err-disabled kan tas up med `shut/no shut` eller genom `udld reset`.

Loop Guard

STP Loop Guard är en funktion som skyddar mot loopar vid trafik i ena riktningen. Loop Guard ser till att Root och Alternate portar ej kan bli Designated om de slutar ta emot BPDUer. Porten placeras då i loop-inconsistent blocking state. De tas tillbaka till ordinarie läge när BPDUs börjar tas emot igen. Loop Guard konfigureras globalt med `spanning-tree loopguard` default och per interface med `spanning-tree guard loop`.

Bridge Assurance

Bridge Assurance är en teknik för RPVST+ och MSTP som används på P2P-länkar. Bridge Assurance är en förlängning av idén med loop guard. Bridge Assurance ändrar reglerna för sändandet av BPDUer, när Bridge Assurance är aktiverat på en port så skickas alltid BPDUer vid varje Hello-intervall oavsett vilken STP roll switchen har. En Bridge Assurance-skyddad port måste alltid ta emot BPDUer, gör den inte det så placeras porten i BA-inconsisten blocking state tills att den tar emot BPDUer igen. Konfigureras globalt med `spanning-tree bridge assurance` och per interface med `spanning-tree protfast network`. Båda sidor av en länk måste vara konfigurerade för Bridge Assurance.

Backbone Fast

Backbone fast är en Cisco propertiär teknik som ska minska konvergeringstid vid återhämtning från indirekta länkfel.