# Spanning-Tree Protocol

Spanning-Tree Protocol and Algorithm

# Introduction

Le **Spanning-Tree** offre une solution de détection de boucles dans des LAN commutés, et offre également la possibilité de maintenir des redondance alternatives de lien pour prévenir d'éventuelle pannes.

La première version du **Spanning-Tree** (STP, 802.1d) demandait un certain temps lors de la reconfiguration de la topologie, il a été remplacé par le **Rapid Spanning-Tree Protocol** (RSTP 802.1w) et le **Multiple Spanning-Tree Protocol** (MSTP, 802.1s).

**RSTP** est donc une extension de **STP**, réduisant les temps de (re)configuration de la topologie active à 2sec contre 30 pour STP.

# Bridge

Par Bridge (= pont), on désigne tout appareil permettant de relier 2 segments d'un réseau, cela peut être un routeur, un concentrateur, un commutateur, etc. Étant donné que le STP s'utilise plus généralement sur un commutateur, j'utilise ici le terme commutateur pour représenter le terme Bridge.

# Fonctionnement d'un commutateur

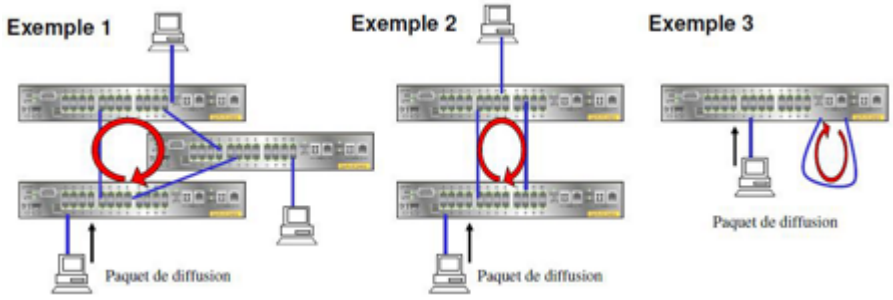
Un commutateur est un appareil possédant au minimum 2 ports et qui permet de diriger les trames Ethernet.

Chaque port du commutateur transmet et reçoit des trames depuis et vers le LAN sur lequel il est attaché. Un commutateur utilise une base de donnée appelée **Filtering Database** dans lequel sont enregistrées les adresses MAC associées à chaque port, c'est ce qui permet au commutateur de savoir ou diriger les trames dans un réseau.

Lorsqu'une trame est reçue avec une adresse MAC de destination qui n'est pas incluse dans sa **Filtering Database**, le commutateur envoie la trame sur tous ses ports actifs. De même les trames émises en diffusion sont émise sur tous les ports actifs.

Lorsqu'une boucle est présente, les trames sont re-émises et l'on obtient rapidement une tempête de diffusion. Afin de détecter ces boucles et d'éviter les tempêtes de diffusion, les commutateurs utilisent le protocole et l'algorithme Spanning-Tree.

# Les boucles sur un réseau Ethernet




# État et rôle des ports

Chaque port d'un commutateur possède à un instant donné, un état, ainsi qu'un rôle.

**RSTP définit 3 états de port :**

**Discarding** indique que le port est exclu de la topologie active et n'enregistre pas d'adresse Mac. En fonction de son rôle il peut cependant rester en écoute pour intercepter les information STP.

**Learning** est en mode écoute et enregistre les adresses MAC dans la Filtering Database.

**Forwarding** correspond à un mode pleinement fonctionnel d'un port, il reçoit et émet des trames.

**RSTP définit également 4 rôle de port :**

**Root** un commutateur ne peut en avoir qu'un, il s'agit du chemin le plus court vers le commutateur Root. A noter qu'un commutateur Root n'en a pas.

**Designated** Par défaut tout port qui n'est ni **Root**, **Alternate** ou **Backup** est un port **Designated**. Pour faire plus simple, un port Designated est un port qui envoie les trames en direction du port Root d'un commutateur.

**Alternate** un port qui possède un chemin vers le port root, mais qui n'a pas été retenu, il s'agit d'une route alternative.

**Backup** un port connecté au port Alternate d'un autre commutateur.

# Principe du STP

Pour qu'une topologie active soit construite, un commutateur va être élu Root, chaque commutateur utilisera le chemin le plus court vers ce commutateur pour propager les trames et éviter ainsi les boucles. Une fois la topologie active construite, les commutateurs vont s'envoyer des informations Spanning-Tree à intervalle régulier afin de garantir que la topologie active n'est pas modifiée. L'absence de réception d'informations Spanning-Tree de la part d'un commutateur peut signifier un lien coupé ou une panne, il sera alors nécessaire de redéfinir la topologie active.

Les informations Spanning-Tree sont diffusées dans des RST BPDU (Bridge Protocol Data Unit). Ces trames sont émises en multicast à une adresse 01-80-C2-00-00-0[0 à F] suivant le type de protocole.

Ces trames contiennent plusieurs informations :

| Nb Octets | Champ |
|---|---|
| **2** | Protocol Identifier |
| **1** | Protocol Version Identifier |
| **1** | BPDU Type |
| **1** | Flags |
| **8** | Root Identifier |

**4** Root Path Cost

**8** Bridge Identifier

**2** Port Identifier

**2** Message Age

**2** Max Age

**2** Hello Time

**2** Forward Delay

**1** Version 1 Length

**Protocol Identifier** Il prend la valeur 0000 0000 0000 0000, qui identifie le RSTP et les versions antérieures de Spanning-Tree.

**Protocol Version Identifier** Prend la valeur 0000 0010 (0000 0000 pour les versions précédentes à RSTP)

**BPDU Type** Ce champ prend la valeur 0000 0010 qui dénote une configuration RST BPDU. (0000 0000 pour un configuration STP)

**Flags** Ces flags indiquent l'état de la topologie active ainsi que l'état et le rôle du port par lequel est envoyé le BPDU.

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Topology Change | Proposal | Port rôle | | Learning | Forwarding | Agreement | Topology Change (ACK) |

**Note** En jaune, ces flags ne sont disponible qu'à partir de RSTP, il sont inutilisés dans des BPDU compatible STP.

**Version 1 Length** Prend la valeur 0000 0000, il a été ajouté pour permettre de distinguer des versions future de STP qui pourraient inclure des informations supplémentaire.

# Commutateur root

Pour construire une topologie active sans boucle, un commutateur Racine doit être élu. Par défaut tous les commutateurs se considèrent root, il émettent alors leurs BPDU sur tout leur ports. Tous les commutateurs recevrons donc le BPDU de tous les commutateurs. Ainsi, chaque commutateur saura qui a été élu root. Pour déterminer le commutateur Root, chaque commutateur possède un identifiant. Celui ci est composé de l'adresse MAC du commutateur, plus un Identifiant de priorité, sur 2 octets (32768 par défaut, modifiable par pas de 4096). Cet identifiant est paramétrable et permet d'influencer l'élection du commutateur Root si on le souhaite.

Cet identifiant est sous la forme **ID : MAC**

Le commutateur ayant le plus petit identifiant est alors élu comme commutateur Root. Lorsqu'un commutateur reçoit un BPDU avec un Root Identifier plus petit que celui qu'il a enregistré, il le remplace et le mémorise. Tous les autres commutateurs vont alors converger leur trafic en direction de ce commutateur Root.

# Root Path Cost

Afin de déterminer le plus court chemin vers le commutateur Root, une valeur est calculée par chaque commutateur et transmises aux autres commutateurs directement reliés. Ainsi, chaque commutateur pourra déterminer son chemin le plus court.

Le commutateur Root aura donc un Root Path Cost de 0. Les autres commutateurs vont donc calculer leur valeur en fonction du Root Path Cost reçu, plus le coût du port, permettant de déterminer quel sera le meilleur chemin vers le commutateur Root.

Les valeurs correspondantes au niveau de priorité de port sont :

| Débit du SubLan | Valeurs Recommandées | Plages recommandées |
|---|---|---|
| 1 Mbps | 20 000 000 | 2 000 000-200 000 000 |
| 10 Mbps | 2 000 000 | 200 000-20 000 000 |
| 100 Mbps | 200 000 | 20 000-200 000 |
| 1 Gbps | 20 000 | 2 000-200 000 |
| 10 Gbps | 2 000 | 200-20 000 |

# Bridge Identifier

L'identifiant du commutateur, comme nous l'avons vu, se compose d'une valeur de priorité suivie de l'adresse MAC du commutateur.

# Port Identifier

Chaque port possède un numéro unique sur un commutateur. Chaque port possède également un niveau de priorité. Ce niveau de priorité est paramétrable est permet donc de modifier la priorité d'un port administrativement.

L'identifiant d'un port est donc la somme du niveau de priorité + le numéro du port.

Le niveau de priorité d'un port va de 0 à 240 (128 par défaut), modifiable par pas de 16.

# Message Age

Ce champ est incrémenté à chaque commutateur traversé. Si ce champ devient supérieur au champ Max Age, le PBDU est détruit.Les 3 paramètres suivant servent à détecter les pannes :

**Max Age** Fixe le délai à partir duquel un commutateur n'ayant pas reçu de BPDU sur son port racine considère qu'un problème est posé.

**Hello Time** Lorsque la topologie active a atteint un état stable, le commutateur Root se met alors à émettre des BPDU à intervalle régulier. Ces BPDU sont ensuite retransmis de proche en proche vers tous les autres commutateurs. Ce mécanisme permet de garantir que la topologie active est toujours la même, et permet à un commutateur de détecter une panne lorsqu'il ne reçoit plus de BPDU sur un port au bout d'un temps donné.Par défaut le commutateur Root émet ces BPDU toutes les 2 secondes, il est possible de modifier cette valeur administrativement.

**Forward Delay** Fixe le délai à respecter pour que le port d'un commutateur passe d'un état bloqué à un état actif.

# Cheminement de l'élection Root et de l'arbre

# Vecteur de priorité

Le vecteur de priorité est le vecteur de priorité Spanning-Tree maintenu pour un port donné. Il se présente sous la forme :

**Port priority vector** = *RootBrigeID : RootPathCost : DesignatedBridgeID : DesignatedPortID : BridgePortID*

Ce vecteur permet de déterminer le rôle de chaque port dont notamment le port root.

Un commutateur A recevant un message de configuration sur un port $P_A$ d'un port désigné $P_B$ d'un commutateur B proclamant un Identifiant Root et un RootPathCost de $RPC_B$ :

*message priority vector* = $R_B : RPC_B : B : P_B :_A$

# Diffusion de son BPDU

Le commutateur A va donc diffuser sur ses autres ports un BPDU en fonction des informations qu'il a reçu. Ainsi, si B est bien Root, il conservera son identifiant, va ajouter son Root Path Cost, son identifiant, et l'identifiant de son port d'émission. Ainsi un commutateur C qui reçoit son BPDU sur un port PC d'un port $P_A$ du commutateur A sera :

**message priority vector = $R_B$ : $RPC_B$+$PPC_A$ : A : $P_A$ : $P_C$**

Lorsqu'un commutateur B se proclame Root le vecteur de priorité root sera : **B : 0 : B : 0 : 0**

donc son BPDU sera :

**message priority vector = B : 0 : B : $P_B$ : $P_B$**

# Assignation du rôle des ports

Grâce au vecteur de priorité root, un commutateur est en mesure de déterminer le port le plus proche du commutateur Root. Ce port sera donc son **Port Root**. Si ce commutateur a d'autres lien vers le commutateur root, mais qui n'ont pas été élus root, ces ports sont des **Port Alternate**. Ils seront à l'état **Disabled**.

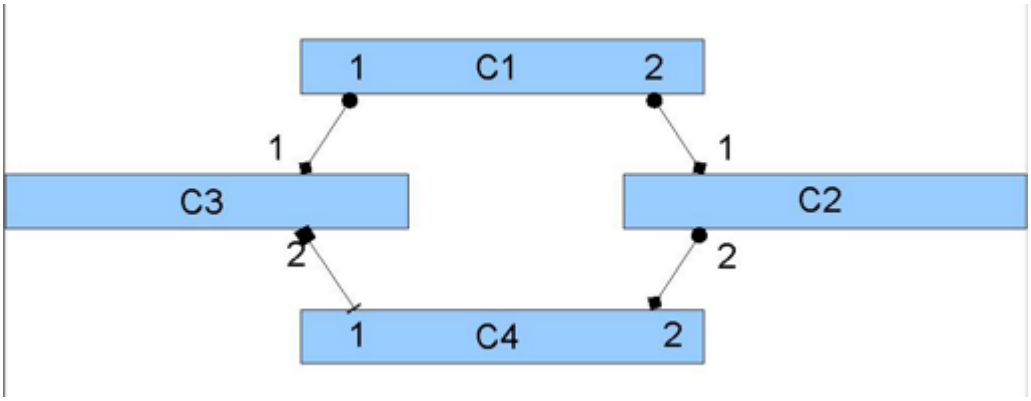
Un commutateur dont un port est relié à un **port Root** est un **Port Designated**, il est à l'état **Forwarding**.

Un commutateur dont le port est relié à un **port Alternate** est un **port Backup**, il est à l'état **Forwarding**.

# Détection et changement de la topologie

Lorsqu'un changement dans la topologie est détectée, de nouveaux messages de type TCN (Topologie Change Notification) sont transmis au travers du réseau pour indiquer qu'il est nécessaire d'opérer les modifications de topologie.

# Exemple




C1 étant le root, il émet toutes les 2sec des BPDU sur ses ports désignés. Ces paquets sont retransmis sur tous les ports désignés des autres commutateurs. Les PBDU sont également transmis sur les ports alternatifs pour indiquer que la connexion existe toujours.

Si le lien entre C2 et C4 est rompu, les BPDU ne seront plus transmis. Ainsi après le délai MaxAge passé sans avoir reçu de BPDU, C4 va :

- Considérer que sont port 2 n'est plus son port root.
- Étant donné qu'il possède un port alternatif, il va le nommer port root.
- Il vide sont cache MAC associé au port 2.
- Ce port 2 devient un port désigné.

puis C3 :

- Après un délai MaxAge dépassé sans avoir reçu de BPDU, il va décider que sont port 2 ne remplis plus son rôle de port Backup.
- Il va le passer en port désigné.
- Il le passe d'abord à l'état d'écoute.
- Il supprime les entrées MAC de son port Root.
- Après un premier Forward Delay son port 2 passe à l'état d'apprentissage
- Puis après un nouveau Forward Delay, le port devient actif.

C3 ayant vidé son cache associé à son port Root, il transmet l'information à C1, qui à son tour va vider son port 2, et enfin C2 fera de même.

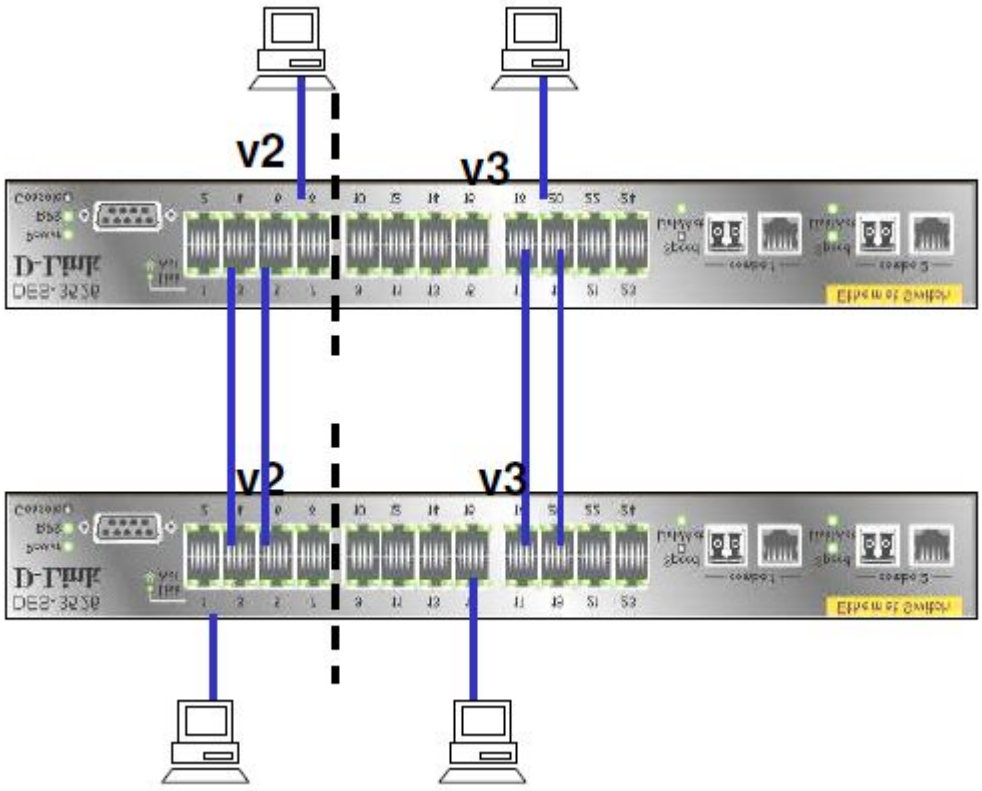
Une fois fois la topologie active redéfinie, le commutateur Root recommence à émettre des BPDU à intervalle régulier.

# Vidage des adresses MAC

Lorsqu'un changement de topologie opère, certains commutateurs doivent vider les adresses MAC apprise sur certains ports. Ainsi, le port Root de C4, relié au port désigné C2, ne reçoit plus de BPDU, changera son port Root, il lui sera nécessaire de vider les adresse MAC apprise par son ancien port Root. Le port désigné du C2 devra également vider ses adresses MAC associés, puis remonter l'information en direction du commutateur Root, afin que tous les commutateurs intermédiaire vident leur adresses MAC associé à leur port désigné ou root qui menait vers le C4.

# MSTP

MSTP est une variante de RSTP. RSTP est unique sur un commutateur, ainsi lorsque plusieurs vlan existent, il n'est pas possible de mettre en place RSTP sur chaque vlan. MSTP a été conçu pour pouvoir disposer d'un Spanning-Tree par vlan.

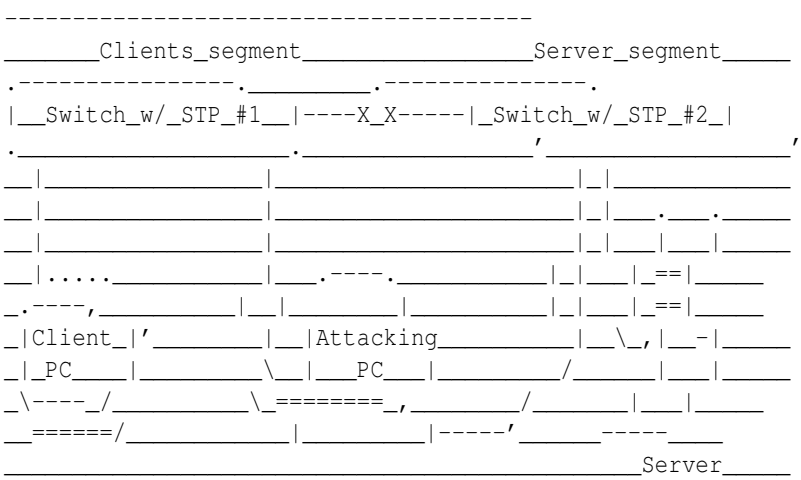



# Problèmes de sécurité lié à STP

Plusieurs problèmes de sécurités ont été découvert avec STP. Il est ainsi très facile d'émettre des BPDU avec un identifiant de commutateur inférieur au Root, et de décrémenter l'identifiant de priorité toutes les secondes, une fois à 0, on repars à sa valeur initial et on recommence. On crée ainsi une reconfiguration permanente, certains port même, n'arriverons jamais à atteindre un état Forwading, Paralysant une partie du réseau.

Il existe de nombreuses façon de détourner STP, une attaque des plus redoutable est de créer un MITM. Une machine attaquante est connecté d'un coté sur un commutateur A et de l'autre un commutateur B. Il émet un STP proclamant le Root, de nouvelles élections sont faite, il devient le root, tous les commutateurs convergents vers l'attaquant, il peut ainsi espionner le trafic qui passe entre 2 commutateurs.

Illustration



Il est difficile de détecter des attaques de ce genre. On pourrait par exemple utiliser un IDS qui compare la topologie active avec une topologie de référence, mais cela impliquerai qu'on ne peut pas changer la topologie sans devoir modifier la topologie de référence, de plus une attaque pourrait être menée au même moment que la topologie change, ainsi la topologie de référence serait alors faussée.

Spanning-Tree devrait donc être désactivé partout où il n'est pas nécessaire. De plus pour améliorer la sécurité, il est préférable d'utiliser des Routeurs, ou des switch/routeurs, de segmenter son réseau et d'utiliser des protocoles comme OSPF. Ce protocole gère bien mieux la topologie du réseau, et la segmentation adéquat du réseau en sous-réseau permet d'éviter de nombreuses attaques, notamment de type MITM.