Evolutions of the Spanning Tree Protocol

Fulvio Risso
Politecnico di Torino

Copyright notice

- This set of transparencies, hereinafter referred to as slides, is protected by copyright laws and provisions of International Treaties. The title and copyright regarding the slides (including, but not limited to, each and every image, photography, animation, video, audio, music and text) are property of the authors specified on page 1.
- The slides may be reproduced and used freely by research institutes, schools and Universities for non-profit, institutional purposes. In such cases, no authorization is requested.
- Any total or partial use or reproduction (including, but not limited to, reproduction on magnetic media, computer networks, and printed reproduction) is forbidden, unless explicitly authorized by the authors by means of written license.
- Information included in these slides is deemed as accurate at the date of publication. Such information is supplied for merely educational purposes and may not be used in designing systems, products, networks, etc. In any case, these slides are subject to changes without any previous notice. The authors do not assume any responsibility for the contents of these slides (including, but not limited to, accuracy, completeness, enforceability, updated-ness of information hereinafter provided).
- In any case, accordance with information hereinafter included must not be declared.
- In any case, this copyright notice must never be removed and must be reported even in partial uses.
Evolutions of the STP

- Rapid Spanning Tree protocol
  - Fast convergence
  - Appropriate for current network topologies (no hubs, only point-to-point links)

- Multiple Spanning Tree
  - Metropolitan area networks
  - Coexistence of STP and RSTP within the same domain
Rapid Spanning Tree (RSTP, 802.1w) (1)

- Fast convergence (less than 1 second)
- Interoperable with IEEE 802.1D
  - Without fast convergence
- Modern solution for mission critical BLAN
- Operates only on point-to-point links
  - Direct connections (no hubs)
- Replaces other proprietary solutions with fast convergence introduced by many vendors
- Defines a set of symbols to be used in network design
- Major improvements
  - Fast topology convergence
  - Fast update of the filtering database
Fast convergence

Requirements

- All the switches must be 802.1w
- Links that may create a mesh between different switches
  - Must be point-to-point (twisted pair or fiber), full duplex
    - STP supported also shared medium (e.g. coax)
  - No links terminated on hubs (although still possible at the edge; often marked as “shared links”)
  - These links guarantee that devices at both sides of the link detect the fault at the same time and that they both start the convergence process at almost the same time

Convergence is usually on the order of ~ 10ms

- Faults are immediately detected in the physical layer
- New transition rules for ports helps to improve the convergence
Improvements introduced for a fast service recovery over STP:

- Fast detection of failures over links
- Failure propagation to establish a new topology and port rules
- Fast transition of the ports to the forwarding state
- Unicast entries update
- Multicast entries update
- Removal of entries related to local changes
- Notifications and appropriate entries removal
- Establish alternative paths before a failure occurs

Detection based on the physical layer:
- Topology map to restore services
- Propagation and resolution of the new topology

Detection based on the protocol:
- Root Port transition
- Design. Port transition

Fast update of local entries
- Notification of the new requirements
Port States and Roles

- Port States
  - Possible operational states with respect to data frames
    - Discarding, Learning, Forwarding
  - Port Roles
    - Role of the port within the STP topology
      - Root, Designated, Alternate, Backup, Edge
Port states in RSTP

- Only 3 states left (from 5 defined in STP)
- Disabled, Blocking and Listening merged in the same state

<table>
<thead>
<tr>
<th>STP (802.1D) Port State</th>
<th>RSTP (802.1w) Port State</th>
<th>Is Port Included in Active Topology?</th>
<th>Is Port Learning MAC Addresses?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disabled</td>
<td>Discarding</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Blocking</td>
<td>Discarding</td>
<td>No (Alternate or Backup)</td>
<td>No</td>
</tr>
<tr>
<td>Listening</td>
<td>Discarding</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Learning</td>
<td>Learning</td>
<td>Yes (on the way to become Root or Designated)</td>
<td>Yes</td>
</tr>
<tr>
<td>Forwarding</td>
<td>Forwarding</td>
<td>Yes (Root or Designated)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
New Port Roles (1)

- Blocking state replaced by Alternate/Backup
- Other roles (Root, Designated) remain the same
- Alternate
  - Alternate port provides an alternate path to the Root Bridge
  - Therefore, it can replace the root port if it fails
New Port Roles (2)

- Backup
  - If we have a LAN directly connected to two ports of the same bridge, the port that is not Designated will be Backup
  - It provides a backup path for reaching that LAN from the bridge
  - Cannot guarantee an alternate connectivity to the root bridge
  - Backup Ports exist only where there are two or more connections from a given Bridge to a given LAN
    - It does no longer exist on modern networks
New Port Roles (3)

- Edge ports
  - Port that connects to a single end-station
    - No other switches are served from that point
  - Need an explicit configuration from the network admin
  - Can lose their status in case a BPDU is received
  - Edge ports do not work like other ports:
    - As soon detect a Link Integrity Test go immediately to Forwarding state without Listening e Learning state
      - No longer have to wait 30s (2 * Max Forward Delay) before having the port fully operational
    - Edge ports become immediately “Designated”
    - A port changing state do not cause a Topology Change Notification BPDU transmission through root port
Modifications in the BPDU

- Modified
  - BPDU Type (now 2)
  - BPDU Version (now 2)
  - Flags
- Old bridges can discard new BPDUs
Principles of the new algorithm

- The tree is created in the same way as STP
  - Root bridge election
  - Root port definition
  - Designated port definition

- Costs and other parameters are handled the same way

- The port not selected as root or designated will become:
  - Alternate if connected to a port on different bridge
  - Backup if connected to a different port of the same bridge
Improvements against STP

- New BPDU Handling
  - BPDU are Sent Every Hello-Time
  - Faster Aging of Information
  - Accepts Inferior BPDUs
- Rapid Transition to Forwarding State
  - Edge Ports
  - Link Type
(1) BPDU sent every Hello Time

- BPDU are sent every Hello Time
  - In STP, non-root bridges simply relay BPDUs when receive them from the root port
    - If the root bridge dies, nobody generates BPDU till Max_Age expires
  - Hello Time: default 2 sec
(2) Faster Aging of Information

- If Hello not received for 3 consecutive times, the current BPDU of the root bridge is declared obsolete
  - No need to wait for Max_Age, although still valid parameter
  - Quick failure detection
- BPDU are then used as “keep-alive” between bridges
- Corollary
  - If a bridge fails to receive BPDUs from a neighbor, it is certain that the connection to that neighbor is lost
  - In STP, the problem might have been anywhere from that bridge to the root
- In presence of Full-Duplex links, failures are detected much faster (no need for 3*Hello Time)
(3) Accepts Inferior BPDUs (1)

- If a bridge receives inferior BPU from its Designated or Root Bridge, it immediately accepts it and replaces the one previously stored.

Example

- B1 detects loss of connectivity from RB.
- B1 starts promoting itself as a root bridge.
- B2 and B3 accepts this info because it comes from the root port.
  - This info is maintained only if B1 has better BridgeID (otherwise B2/B3 will reply with a better BPDU).
(3) Accepts Inferior BPDUs (2)

- Example
  - B1 detects loss of connectivity from RB
  - B1 starts promoting itself as a root bridge
  - B2 does not accept this info because it does not come from the root port
  - B2 sends immediately back its own BPDU to B1 with the right RB
  - B1 changes its active port to Root
(4) Rapid Transition to Forwarding State: Edge Ports

- Edge ports connect end users
- Edge ports transition immediately in Forwarding State
  - No listening/learning stage
  - No delayed connectivity for the end user
- No Topology Change Notification when the link changes its status
- If it receives a BPDU it loses immediately its edge port status and it becomes a normal port
(5) Rapid Transition to Forwarding State: Link Type

- Rapid transition to Forwarding State
  - Achieved only on Edge Ports and Point-to-Point links
- Point-to-point link: if the port is in Full Duplex
  - A port in Half Duplex is considered as a “shared port”
Convergence with STP: example

- The Root Bridge adds a new link toward A
  - Link goes into Listening (no data)
    - No loop in the network
  - BPDU generated by root propagated down to B, C, D
  - B, C, D update their STP Topology (e.g. Root Port)
    - A, B, C were previously reachable through D
    - All links involved still in blocking
    - D will move port P1 in Blocking state
  - A, B, C and their leaves unreachable for 2 * ForwardDelay
    - Link Root-A in Forwarding state after 2* Forward Delay
Convergence with RSTP: example

- The Root Bridge adds a new link toward A
  - Both ports put in Blocking
  - Now, a new negotiation takes place
  - A receives a Root BPDU from its upper port and it blocks all non-edge designated ports
    - Operation called “sync”
  - Then, A explicitly authorizes the RB to put its port in Forwarding State
    - A puts that link in forwarding state too
    - No loops can occur, since downstream ports of bridge A are now blocked
- Process is repeated on B and C: they both negotiate to activate their upstream ports through a SYNC operation
- The explicit authorization sent in the SYNC process replaces the 2*Forward Delay of STP
Port negotiation (Proposal / Agreement) (1)

- A new link is created between RB and B1
  - Both P1 and P2 transition in Designated Blocking state (Discarding frames) until they receive a BPDU from the other party
  - They both send BPDU with the “Proposal” bit
  - B1 learns that the best path is no longer through port P6, but through port P2
Port negotiation (Proposal / Agreement) (2)

- B1 starts a SYNC operation
  - SYNC = I need to be sure that no loop can occur
  - It puts in Discarding state all the ports that are active in the current STP topology (i.e., Designated and Root ports)
  - Now, B1 can unblock port P1 by sending a BPDU with the “Acknowledgement” bit set
    - The BPDU is an exact copy of the one received by RB plus the ACK flag
    - This is needed by RB to understand exactly which BPDU was acknowledged
  - Link P1-P2 goes in Forwarding
- The process may restart on all ports in Discarding State below B1 (P3, P4, P6)
Fault detected on the root port

- If an Alternate port exist, it is promoted to Root
  - In case more than one Alternate exist, the best one is promoted
  - This path will become the new path toward the root bridge

- If no Alternate ports exist, the Bridge starts promoting itself as a new Root Bridge
Topology Change Detection

- Only non-edge ports that move to forwarding state cause a Topology Change
  - A loss of connectivity (a port moving to Blocking) is no longer considered a Topology Change
    - It was in 802.1D
- An RTSP bridge that detects a topology change
  - Starts the TC timer (2*Hello Time) for all entries related to its Designated Ports and its Root Port
  - It flushes all MAC addresses associated to these ports
  - It propagates BPDU with the TC bit on all these ports till the TC timer expires
Topology Change Propagation (1)

- When a bridge receives a BPDU with the TC bit set from a neighbor:
  - It clears the MAC addresses learned on all its ports, except the one that receives the topology change
  - It starts the TC timer and sends BPDUs with TC set on all its designated ports and root port
    - RSTP no longer uses the specific TCN BPDU, unless a legacy bridge needs to be notified
- TCN floods very quickly across the entire network
  - No longer involved the Root Bridge (such as in the STP) in order to generate BPDUs with the TC set
Topology Change Propagation (2)

The originator of the TC directly floods this information through the network.
Compatibility with 802.1D (1)

- Interoperability possible
  - Migration delay timer (3 sec) starts when a port comes up
    - During this time, the current STP or RSTP mode associated to the port is locked
    - When the migration delay expires, the port adapts to the mode that corresponds to the next BPDU it receives
    - If the port changes its mode of operation as a result of a BPDU received, the migration delay restarts
  - 802.1D timers (Forward Delay, Max_age) used only as a backup

- Fast convergence is lost

- Problem: an 802.1w bridge that starts operating in 802.1D compatibility mode does not turn back in 802.1w unless an explicit configuration is taken
Compatibility with 802.1D (2)

- Example
  - A has the Designated Port over the LAN
  - A receives a 802.1D BPDU from C
    - Please note that C advertises itself as Designated Port, since it cannot understand the 802.1w BPDU generated by Bridge A
  - A changes in compatibility mode and starts generating 802.1D BPDU
  - If A has the best BridgeID, C will put its port in Blocking State
Conclusions

- Efficient
- Fast convergence (often < 1s)
- Can replace many proprietary protocols
- Possibility to mix devices from different vendors
- Interoperability problems with STP
  - Mostly used with Multiple Spanning Tree (MST)