universidade de aveiro
theoria poiesis praxis

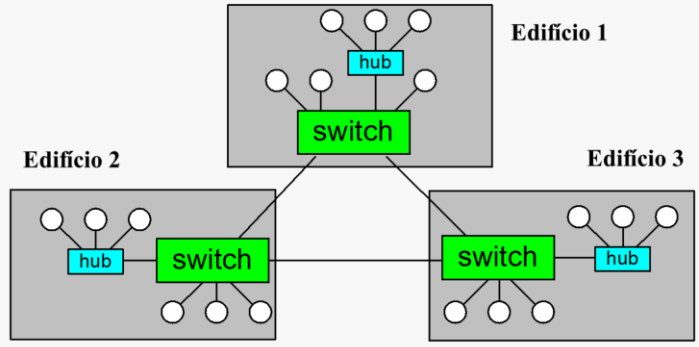
Bridges/Switches: Spanning Tree (IEEE 802.1D)

Fundamentos de Redes

Mestrado Integrado em Engenharia de Computadores e Telemática

DETI-UA, 2018/2019

Interligação de Bridges/Switches com redundância

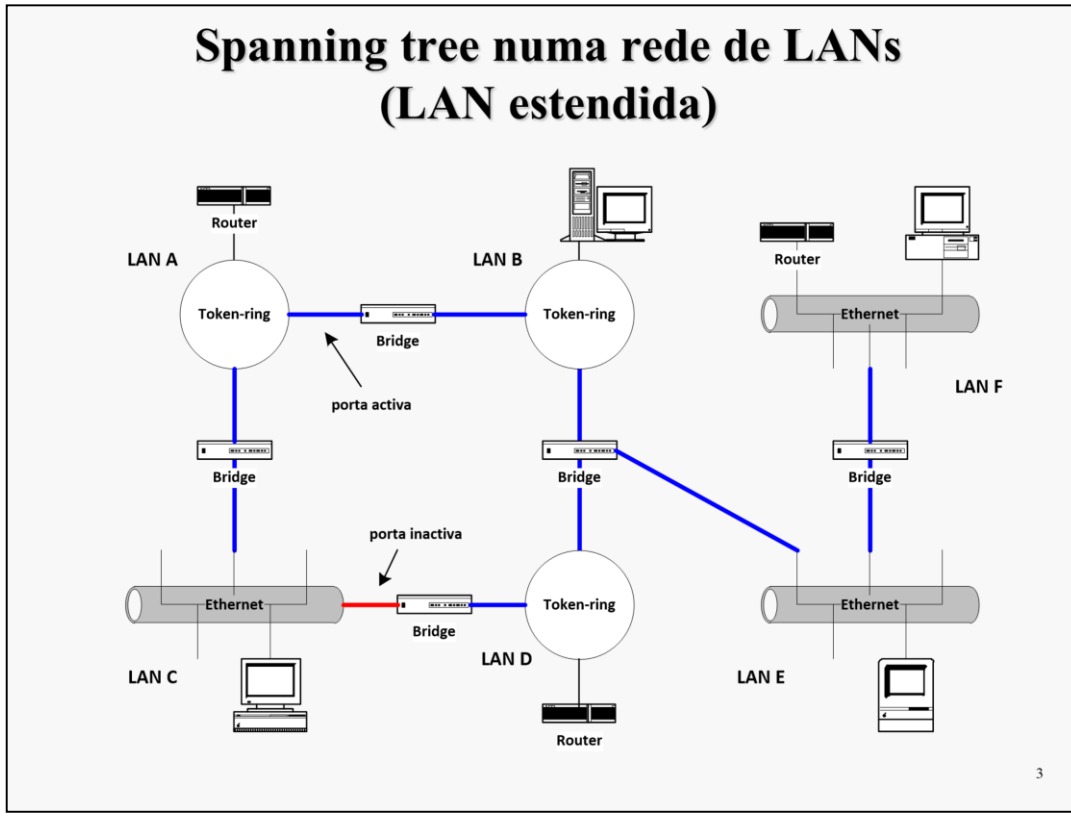


- **Objectivo**: proporcionar capacidade de recuperação a falhas na rede
- **Problema**: redundância lógica provoca colapso de comunicações devido, por exemplo, às tramas MAC para o endereço de broadcast

Redundancy of Bridge/Switch networks

When connecting different bridges/switches, it is desirable to have a redundant topology, i.e., a set of connections that can maintain the whole network connected when some connections fail. In the above example, a network composed by 3 switches, each one in a different building, has 3 connections in such a way that if one connection fails, the remaining two connections still connect all switches.

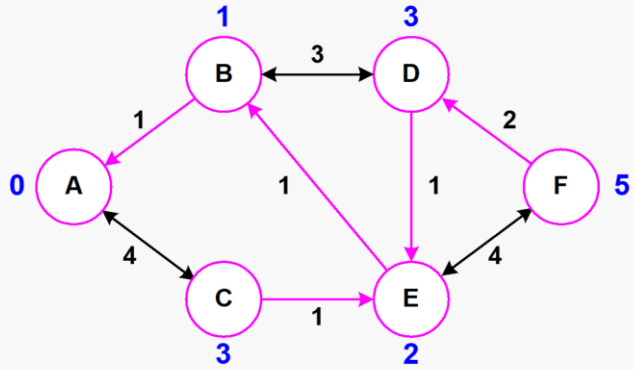
Nevertheless, at logical level, routing cycles cause communications collapse because, for example, broadcast frames will circulate in the network for ever (why?). To solve this issue, a protocol is used between switches that enables them to decide which connections should be forwarding frames and which connections should be discarding frames. This protocol is the IEEE 802.1D STP (Spanning Tree Protocol).



Spanning Tree on an extended LAN

In the above example, an extended LAN is implemented using bridges to interconnect LAN segments of different technologies. Bridges change information (through the IEEE 802.1D STP) in order to decide which ports become active and which ports become inactive in such a way that all LANs are interconnected through a single routing path. In the above example, a single port (the red port) became inactive.

Equações de Bellman



Quando os custos das ligações são não negativos, então:

Custo do percurso mínimo de um nó para A

=

Custo do arco que une esse nó ao nó que se lhe segue no percurso mínimo

+

Custo do percurso mínimo desse nó para o nó A

Bellman equations

Consider a graph where each edge is assigned with a positive cost value and one node is selected as the root node. When all other nodes aim to compute the minimum cost path from them to the root node, the Bellman equations can be used. For a given node, the Bellman equations state that:

The cost of the minimum cost path to the root node

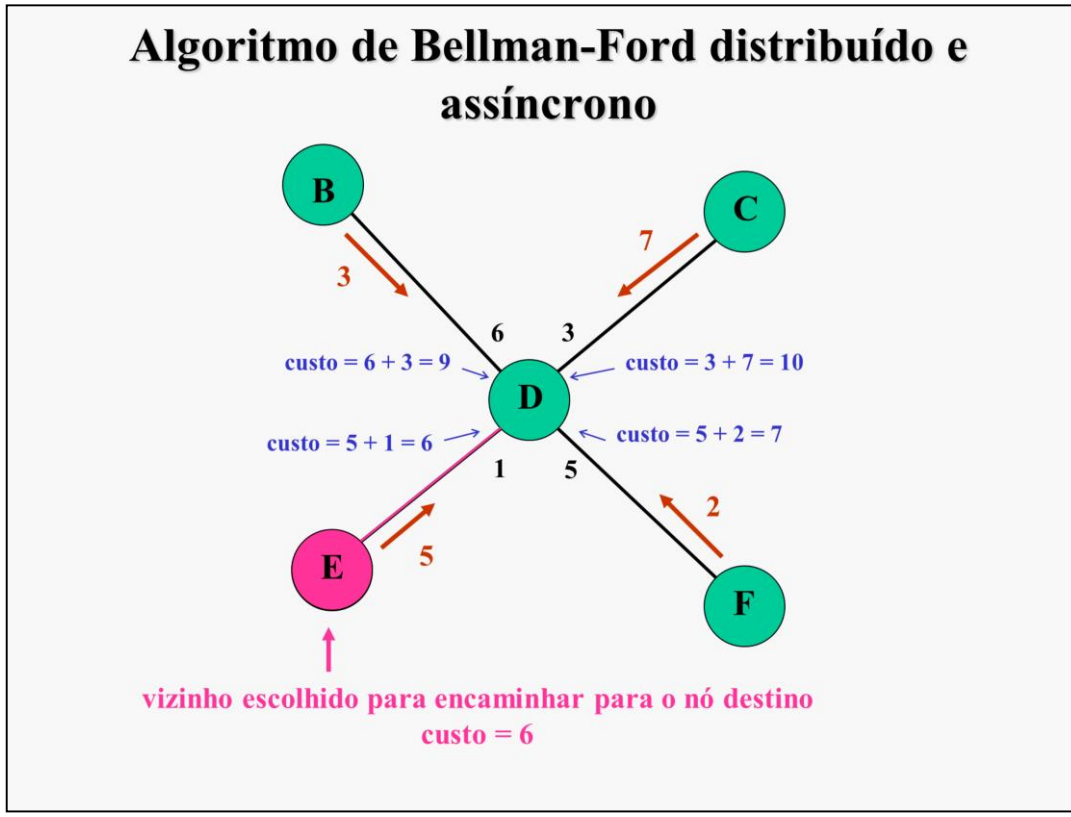
=

The cost of the link to the next node in the minimum cost path to the root node

+

The cost of the minimum cost path of the next node to the root node

In the above example, the pink arcs highlight the minimum cost paths from every node to root node A. If we consider, for example, node E, we can see that the cost of the minimum cost path from E to A (which is 2) = the cost of link {E,B} (which is 1) + the cost of the minimum cost path from B to A (which is 1).



Bellman-Ford algorithm

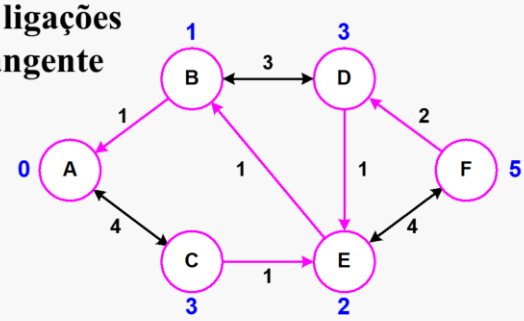
One algorithm to compute the minimum cost path from one origin node to a root node is the Bellman-Ford algorithm. This algorithm can run in a distributed way among all nodes where each node solves the Bellman equations based on information from its neighbor nodes, and the information exchange between nodes can be asynchronous.

At each node *i*, the algorithm runs as follows. Node *i* sends periodically to all its neighbor nodes its own estimation D_i of the cost of its minimum cost path to the root node. When it receives the estimations from its neighbors, it updates its cost with $D_i = \min(d_{ij} + D_j)$ among all neighbor nodes *j*. The neighbor node *j* such that $d_{ij} + D_j$ is minimum becomes the next node towards the root node.

In the above example, node D has received from its neighbor nodes the estimation costs 3 (from node B), 7 (from node C), 5 (from node E) and 2 (from node F). For each neighbor node, D sums the neighbor estimation with the cost of the link connected to it. Node D obtains a cost of 6 + 3 = 9 for node B, 3 + 7 = 10 for node C, 5 + 1 = 6 for node E and 5 + 2 = 7 for node F. The minimum value is 6 through neighbor node E. Therefore, the current estimation of node D is that the minimum cost path to the root node is 6 and it is through node E.

Encaminhamento baseado em Spanning Trees

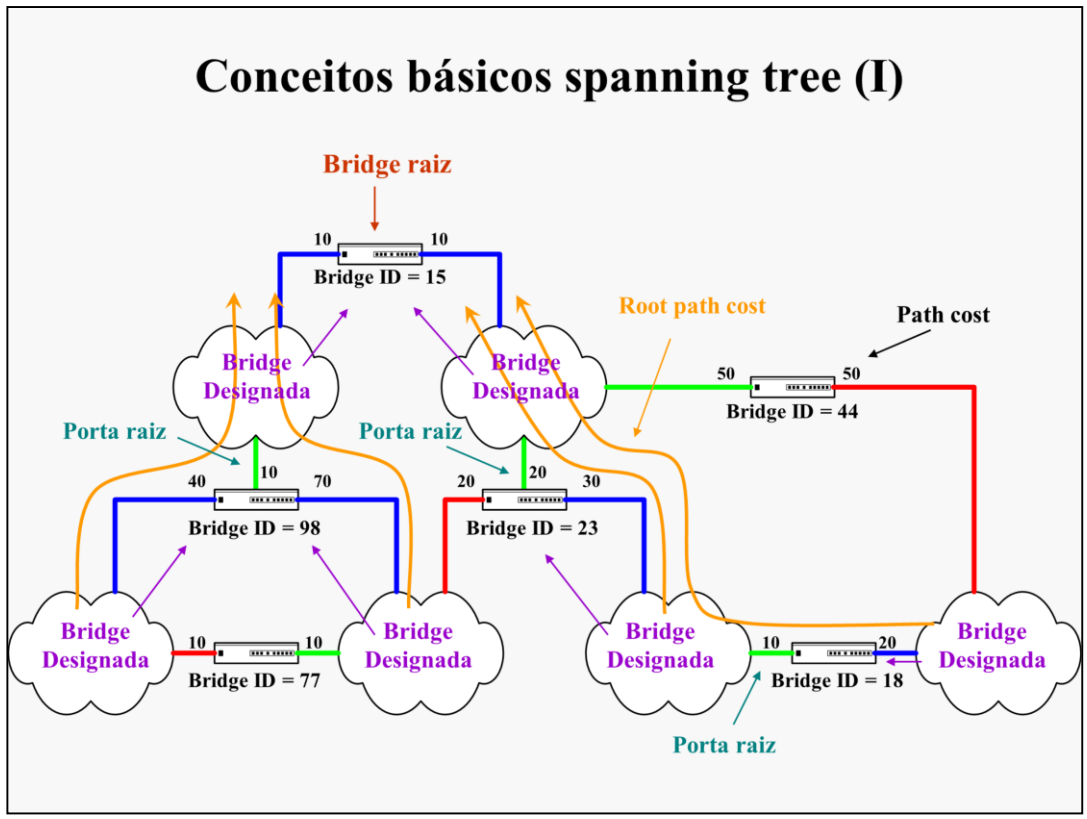
- **É escolhida uma bridge como nó origem**
- **As outras bridges usam o algoritmo de Bellman-Ford assíncrono e distribuído para calcular o vizinho no percurso de custo mínimo para o nó origem**
- **<u>As ligações compostas pelos percursos de custo mínimo</u> (de todas as outras bridges para a origem) <u>definem uma árvore abrangente</u> (spanning tree)**
- **As portas activas são as das ligações que compõem a árvore abrangente**
- **É necessário um critério para desempatar quando há múltiplos percursos de custo mínimo**



Routing based on spanning trees

Routing between Layer 2 switches/bridges is based on spanning trees. In the IEEE 302.1D STP:

- One switch is selected as the Root Bridge
- The other switches use the distributed and asynchronous Bellman-Ford algorithm to compute the next switch in the minimum cost path to the Root Bridge
- The connections of the minimum cost paths from all switches to the Root Bridge define a spanning tree
- The switches activate their ports belonging to the spanning tree
- An additional criterion is required when multiple minimum cost paths exist (coming later)



IEEE 802.1D basic concepts (I)

Each switch/bridge has a Bridge ID value. The bridge with lowest Bridge ID value is selected as the Root Bridge.

Each switch/bridge port has a Path Cost value (which must be a positive integer). The Root Path Cost of each bridge is the cost of the minimum cost path from it to the Root Bridge and the Root Port of a bridge is the port that provides the Root Path Cost (the Root Path Cost is the sum of the Path Cost values of the output ports of the bridges towards the Root Bridge).

In the above example, consider the bridge with Bridge ID = 18. For this bridge, there are the following paths to the Root Bridge (with the following costs):

BridgeID=18 → BridgeID=23 → Root Bridge (with total cost 10 + 20 = 30)

BridgeID=18 → BridgeID=44 → Root Bridge (with total cost 20 + 50 = 70)

BridgeID=18 → BridgeID=23 → BridgeID=98 → Root Bridge (with total cost 10 + 20 + 10 = 40)

BridgeID=18 → BridgeID=23 → BridgeID=77 → BridgeID=98 → Root Bridge (with total cost 10 + 20 + 10 + 10 = 50)

The path with the minimum cost value is the first one and, therefore, the Root Path Cost of the bridge with Bridge ID = 18 is 30 and its Root Port is the one with cost 10. Note that the Root Bridge has a Root Path Cost of 0 and has no Root Port (Root Ports are highlighted in green).

Each LAN has a Designated Bridge and a Designated Port (which is the port of the Designated Bridge connected to the LAN). For each LAN, its Designated Bridge is the bridge providing the minimum Root Path Cost from the LAN to the Root Bridge.

In the example, the rightmost LAN has two bridges connected to it. The bridge with Bridge ID = 18 provides a Root Path Cost of 30 while the bridge with Bridge ID = 44 provides a Root Path Cost of 50. Therefore, the bridge with Bridge ID = 18 is the Designated Bridge of this LAN and its port (the one with cost 20) is its Designated Port (Designated Ports are highlighted in blue).

Conceitos básicos spanning tree (II)

- **Bridge ID – cada bridge é identificada por um endereço que contém:**
  - **2 octetos de prioridade, configurável pelo gestor da rede +**
  - **6 octetos fixos (um dos endereços MAC das portas da bridge, ou qualquer outro endereço único de 48 bits)**
  - **A prioridade tem precedência sobre o campo de octetos fixos**
- **Bridge raiz (*Root Bridge*) – bridge que está na raiz da spanning tree; bridge com menor Bridge ID**
- **Path cost – custo associado a cada porta da bridge (pode ser configurado pelo gestor da rede)**

8

IEEE 802.1D basic concepts (II)

Each switch/bridge has a Bridge ID value. The Bridge ID is 8 bytes long:

- the 2 most significant bytes are the priority of the switch; they are 0x8000 by default but can be configured by the network manager;

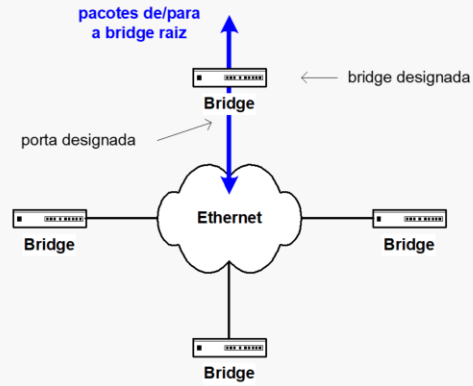
- the 6 least significant bytes are the switch MAC address: an IEEE address, which is set by the manufacturer and is guaranteed to be unique.

The Root Bridge is the Bridge in the root of the spanning tree and it is the switch with the lowest Bridge ID value.

The Path Cost is a positive value associated to each switch port (it can be configured by the network manager).

Conceitos básicos spanning tree (III)

- **Bridge designada (*Designated Bridge*) – bridge que, numa LAN, é responsável pelo envio de pacotes BPDU da LAN para a bridge raiz e vice-versa; a bridge raiz é a bridge designada em todas as LANs a que está ligada**
- **Porta designada (*Designated Port*) – porta que, numa LAN, é responsável pelo envio de pacotes BPDU da LAN para a bridge raiz e vice-versa (uma das portas da bridge designada)**



9

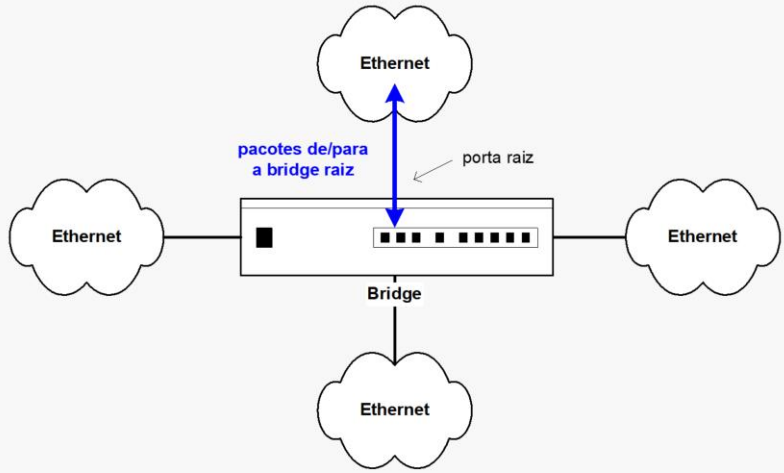
IEEE 802.1D basic concepts (III)

The Designated Bridge of a LAN is the bridge responsible to forward BPDU packets from the LAN towards the Root Bridge and vice-versa (the Root Bridge is the Designated Bridge of all LANs that is attached to).

The Designated Port of a LAN is the port of the bridge responsible to forward BPDU packets from the LAN towards the Root Bridge and vice-versa.

Conceitos básicos spanning tree (IV)

- **Porta raiz (*Root Port*) – porta que, numa bridge, é responsável pela recepção/transmissão de pacotes BPDU de/para a bridge raiz**



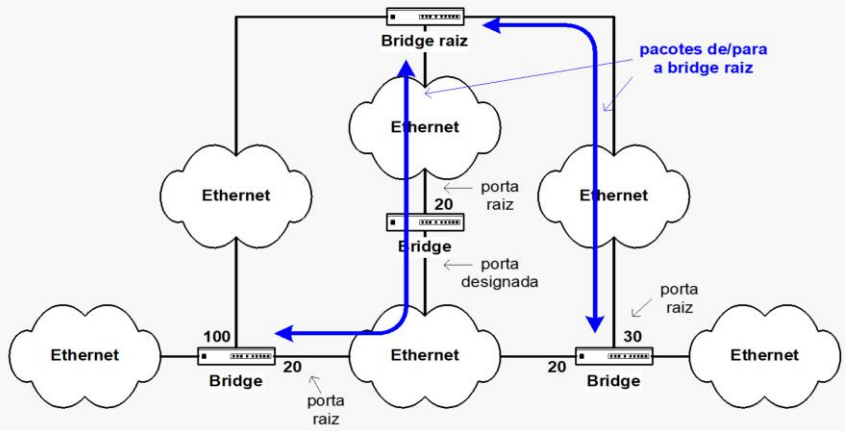
10

IEEE 802.1D basic concepts (IV)

The Root Port of a bridge is the port responsible to send/receive BPDU packets to/from the Root Bridge.

Conceitos básicos spanning tree (V)

- **Cada bridge tem associado um custo do percurso para a raiz (*Root Path Cost*), igual à soma dos custos das portas portas raiz no percurso de menor custo para a bridge**



11

IEEE 802.1D basic concepts (V)

Each bridge has an associated Root Path Cost which is the cost of the minimum cost path from it to the Root Bridge. The Root Path Cost is the sum of the Path Cost values of the Root Ports of the bridges towards the Root Bridge.

Conceitos básicos spanning tree (VI)

- A **porta raiz** é, em cada bridge, a porta que fornece o melhor percurso (de menor custo) para a raiz
- A **porta designada** é, em cada LAN, a porta que fornece o melhor percurso para a raiz

As portas activas em cada bridge são a porta raiz + as portas designadas

As restantes portas ficam inactivas (blocking)

12

IEEE 802.1D basic concepts (VI)

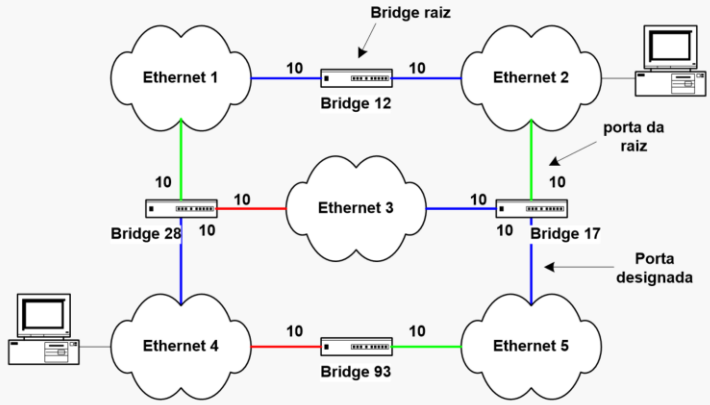
The Root Port is, on each bridge, the port providing the minimum cost path to the Root Bridge.

The Designated Port is, on each LAN, the port providing the minimum cost path to the Root Bridge.

On each bridge, the Root Port and the Designated Ports become active while the others become inactive (i.e., blocking).

Conceitos básicos spanning tree (VII)

- **Numa bridge, se houver múltiplas portas com o menor custo para a raiz, a porta raiz é a que liga à próxima bridge com menor BridgeID (exemplo: Bridge 93 na figura)**
- **Numa LAN, se houver múltiplas bridges com o menor custo para a raiz, a bridge designada é a que tiver menor BridgeID (exemplo, Ethernet 3 na figura).**



13

IEEE 802.1D basic concepts (VII)

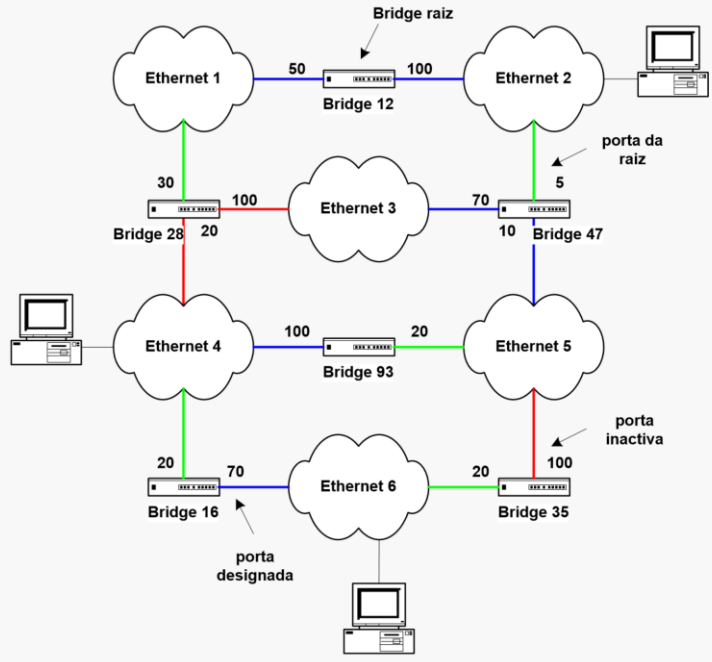
In a bridge, when multiple ports provide the same minimum Root Path Cost, the Root Port is the port connected to the bridge with the lowest Bridge ID value. In the above example, the bridge with Bridge ID = 93 has a Root Path Cost of 20 on both ports. It selects the port connected to Ethernet 5 because it is connected to a bridge with Bridge ID = 17 while the other port is connected to a bridge with Bridge ID = 28.

In a LAN, if multiple bridges provide the same minimum Root Path Cost, its Designated Bridge is the one with the lowest Bridge ID value. In the above example, both bridges connected to Ethernet 3 provide a Root Path Cost of 10. The Designated Bridge of Ethernet 3 is the bridge with Bridge ID = 17 (which is lower than 28).

Exemplo – spanning tree (I)

Bridges designadas

Eth1	12
Eth 2	12
Eth 3	47
Eth 4	93
Eth 5	47
Eth 6	16



14

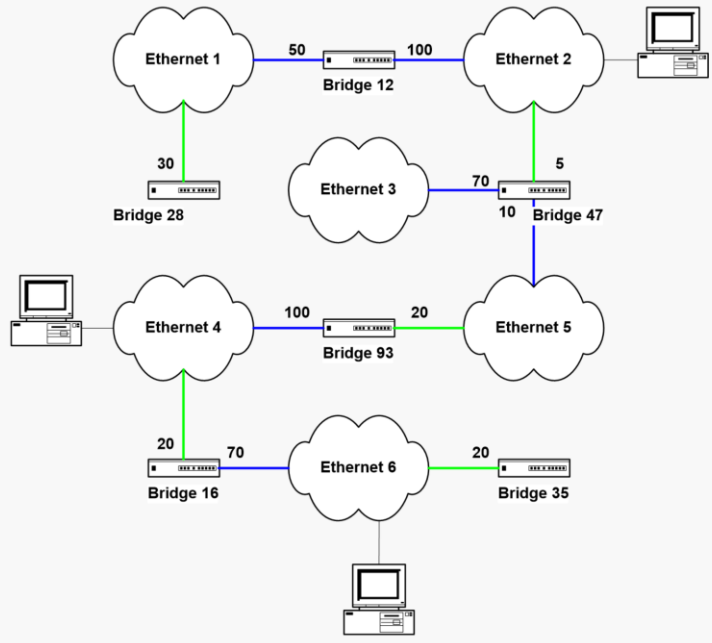
Example – spanning tree (I)

In any network of bridges (or switches) running IEEE 802.1D STP, provided that the Bridge ID and Path Cost values are known, we can always determine the Root Ports and the Designated Ports and, consequently, we can predict the ports which are active (the blue and green ports in the above example) and the ones which are inactive (the red ports in the above example). The table indicates the Designated Bridges of each LAN in the example.

Exemplo – spanning tree (II)

Bridges designadas

Eth1	12
Eth 2	12
Eth 3	47
Eth 4	93
Eth 5	47
Eth 6	16



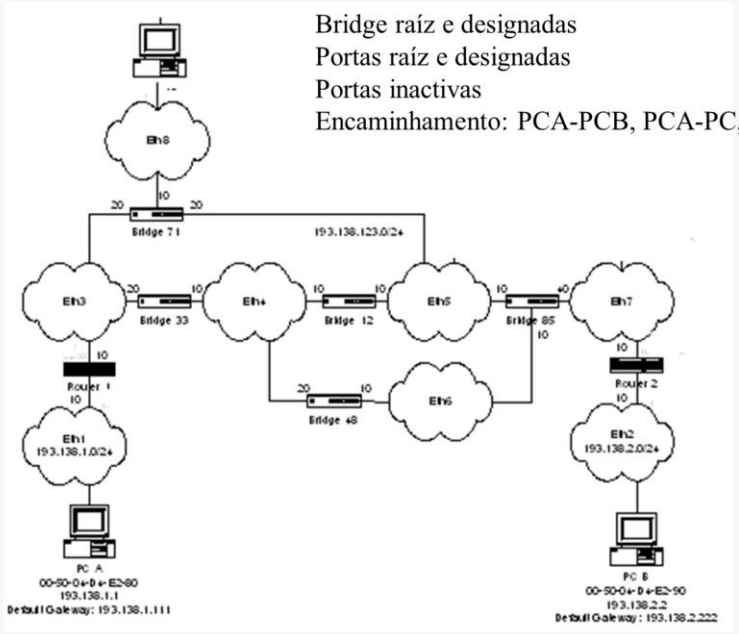
15

Example – spanning tree (II)

When bridges decide to make inactive some ports, we obtain a logical network (where data frames are forwarded) which is a spanning tree. Note that every bridge is always logically connected through at least one port (its Root Port). A bridge with a single active port behaves as a single attached host.

The above figure represents the logical network obtained by the spanning tree of the previous slide.

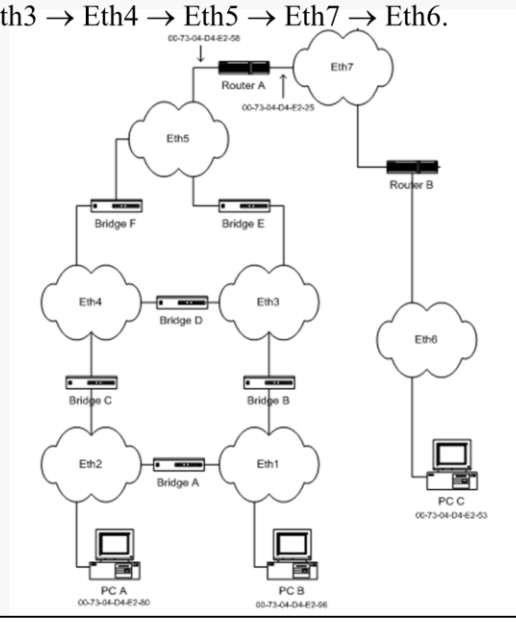
Exemplo I



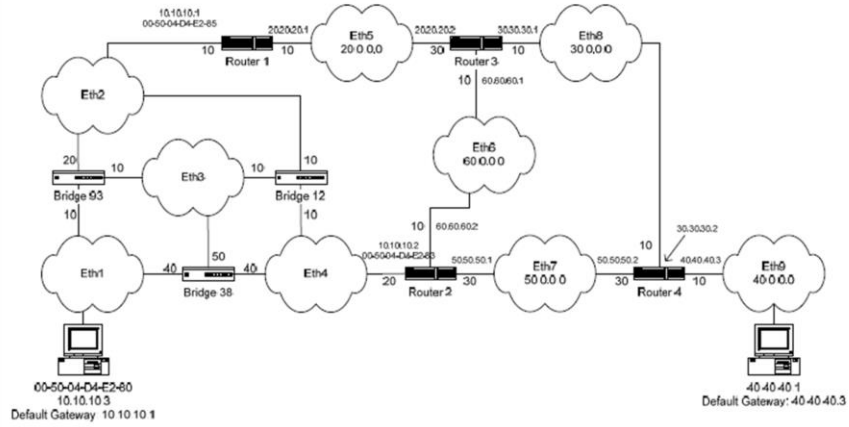
Bridge raíz e designadas
Portas raíz e designadas
Portas inactivas
Encaminhamento: PCA-PCB, PCA-PC, PCB-PC

Exemplo II

Indique as configurações a efectuar nos elementos de rede por forma a que quando é efectuado um ping do PC A para o PC C o ICMP Echo Request siga o percurso Eth2 → Eth1 → Eth3 → Eth4 → Eth5 → Eth7 → Eth6.



10. Considere a rede da figura, com 9 redes locais Ethernet (Eth1 a Eth9) interligadas por bridges e routers. Na figura estão indicados os custos das portas dos routers e das bridges. Todos os endereços IP atribuídos têm a máscara 255.0.0.0 associada.



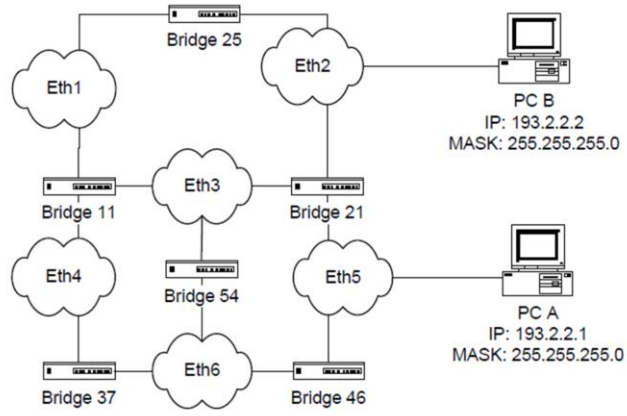
a) Relativamente à rede de bridges, indique qual a bridge raiz, qual a bridge designada em cada rede local, qual a porta da raiz em cada bridge e quais as portas activas em cada bridge.

4. Considere a rede da figura seguinte constituída por 6 segmentos Ethernet interligados por bridges com o protocolo Spanning Tree activo (os *BridgeIDs* são indicados na figura e os *PortID* das portas de cada bridge são dados pelo valor da rede Eth a que estão ligadas). Considere um custo de 50 para as portas da bridge 25 e um custo de 100 para as portas de todas as outras bridges.

4.1. Indique justificadamente qual a porta raiz de cada bridge e qual a bridge designada de cada rede física. (*2.0 valores*)

4.2. Indique justificadamente por que redes Eth passam os pacotes ICMP gerados por um *ping* entre o PC A e o PC B. (*2.0 valores*)

4.3. Para haver comunicação IP entre os PCs, é preciso configurar endereços IP nas bridges? Justifique. (*2.0 valores*)

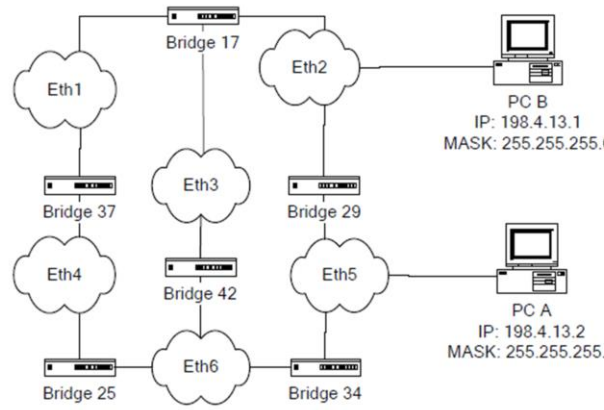


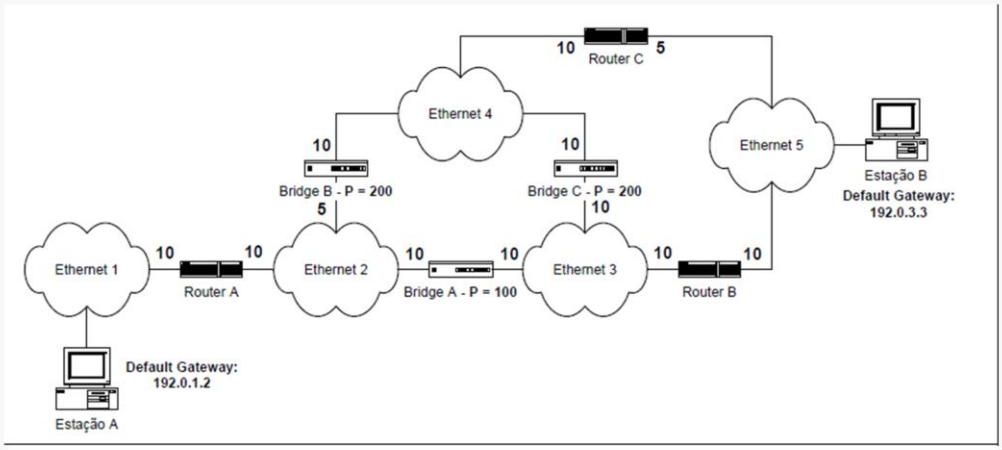
Considere a rede da figura constituída por 6 segmentos Ethernet interligados por bridges com o protocolo Spanning Tree activo (os *BridgeIDs* são indicados na figura e os *PortID* das portas de cada bridge são dados pelo valor da rede Eth a que estão ligadas). Considere um custo de 20 para as portas das bridges 25 e 37 e um custo de 40 para as portas de todas as outras bridges.

3.1. Indique justificadamente qual a porta raíz de cada bridge e qual a bridge designada de cada rede Eth. (*1.5 valores*)

3.2. Indique justificadamente por que redes Eth passam os pacotes ICMP gerados por um *ping* entre o PC A e o PC B. (*1.5 valores*)

3.3. Se for executado um *ping* no PC A para o endereço 198.4.13.1 numa situação em que a tabela ARP do PC A está vazia, diga justificando qual a tabela de encaminhamento da bridge 25 após a execução deste *ping*. (*1.5 valores*)

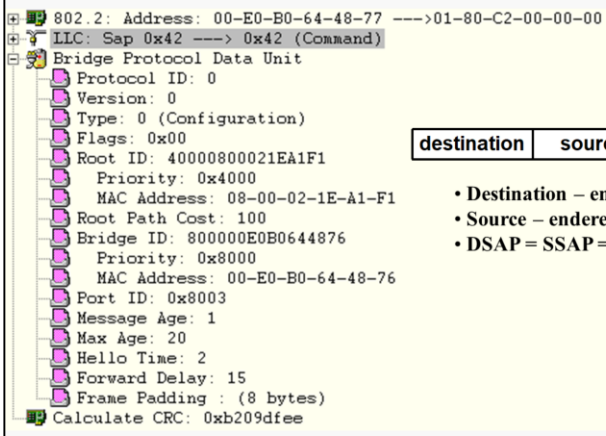




2.3. Na solução apresentada em 2.2., indique justificadamente qual a porta raiz de cada bridge e qual a bridge designada de cada segmento Ethernet. (*1.5 valores*)

Protocolo IEEE 802.1D
BPDUs (Bridge Protocol Data Units)

- **Para construir e manter a spanning tree as bridges trocam mensagens especiais entre si, designadas por Bridge Protocol Data Units (BPDUs)**
- **Existem dois tipos:** ***Configuration*** **e** ***Topology Change Notification***



destination	source	DSAP	SSAP	BPDU

- **Destination – endereço multicast atribuído a todas as bridges**
- **Source – endereço MAC da porta que enviou a BPDU**
- **DSAP = SSAP = 01000010 = 42 hex**

IEEE 802.1D BPDUs (Bridge Protocol Data Units)

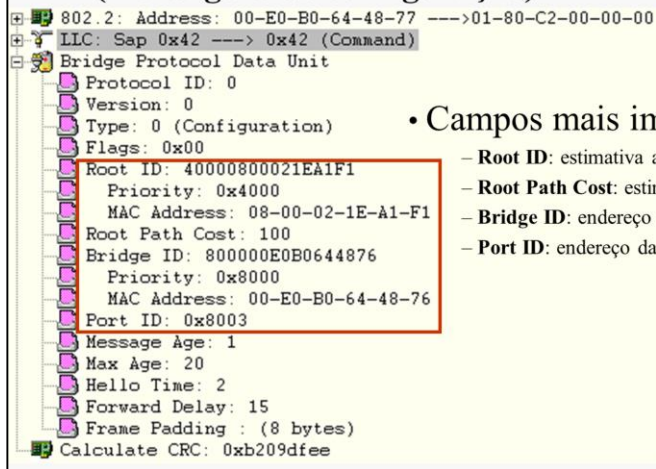
To built, maintain and reconfigure the spanning tree, bridges exchange between them special protocol messages, named Bridge Protocol Data Units (BPDUs).

There are two types of BPDU messages: Configuration BPDUs (Conf-BPDUs) and Topology Change Notification BPDUs (TCN-BPDUs).

BPDU messages are encapsulated in 802.3 format Layer 2 frames where the source address is the bridge MAC address and the destination address is the multicast address 01:80:C2:00:00:00. In the LLC layer, both Source Service Access Point (SSAP) and Destination Service Access Point (DSAP) have the code 0x42.

Configuration BPDUs

- **A configuração da spanning tree é feita pelas Conf - BPDUs (mensagens de configuração)**

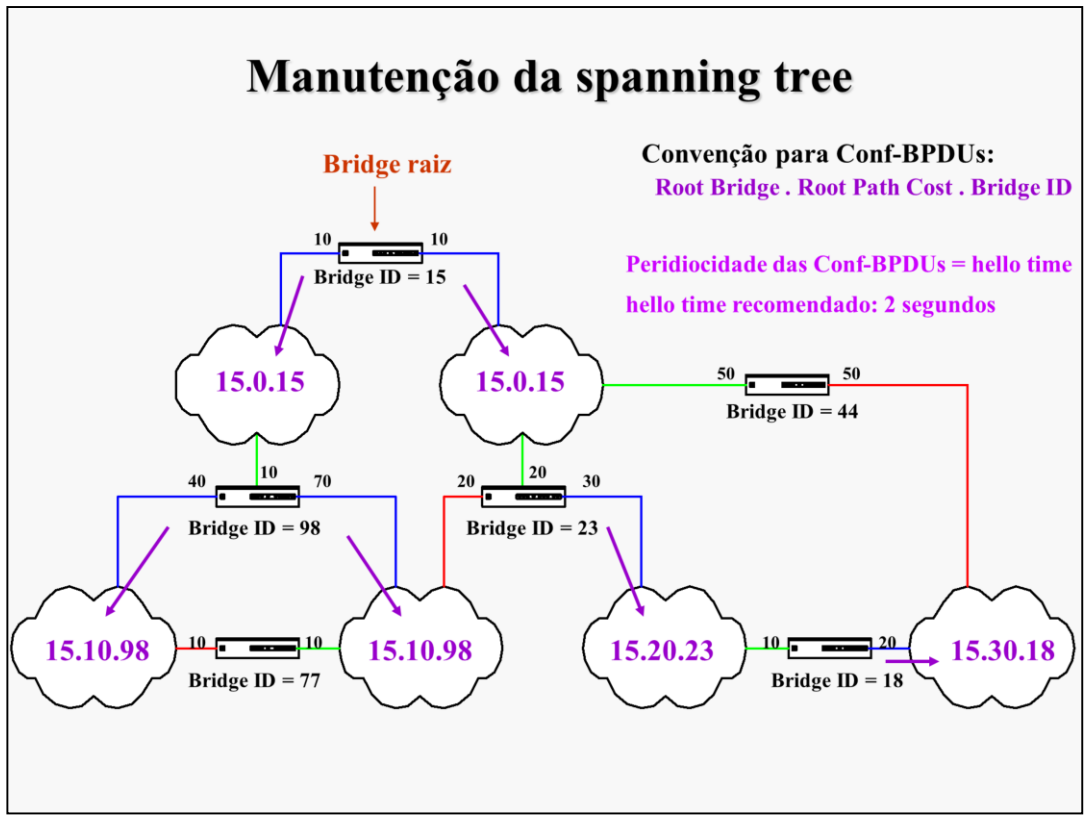


- Campos mais importantes:
  - **Root ID**: estimativa actual do endereço da bridge raiz
  - **Root Path Cost**: estimativa actual do custo para a bridge raiz
  - **Bridge ID**: endereço da bridge que envia a mensagem de configuração
  - **Port ID**: endereço da porta que envia a mensagem de configuração

Configuration BPDUs

The spanning tree configuration on each bridge is done based on the exchange of Conf-BPDU messages. To achieve this goal, the most important Conf-BPDU fields are:

- The Root ID field: the current Root ID estimation of the sender bridge
- The Root Path Cost field: the current Root Path Cost estimation of the sender bridge
- The Bridge ID field: the Bridge ID of the sender bridge
- The Port ID field: the Port ID of the port through which the message is sent.



Maintaining the spanning tree

To maintain the spanning tree: (i) the Root Bridge sends periodically Conf-BPDU messages through its Designated Ports (the periodicity is Hello Time and its recommended value is 2 seconds), (ii) for each Conf-BPDU message received on a Root Port, each bridge calculates its own Conf-BPDU message and sends it through its own Designated Ports.

Consider the above example, the Root Bridge (with Bridge ID = 15) sends periodically on all its Designated Ports a Conf-BPDU with: (i) Root Bridge = 15, (ii) Root Path Cost = 0 and (iii) Bridge ID = 15 (in short, 15.0.15).

For each Conf-BPDU received by the bridge with Bridge ID = 98 on its Root Port, the bridge sends on all its Designated Ports a Conf-BPDU with: (i) the Root Bridge = 15 (equal to the received Root Bridge), (ii) Root Path Cost = 10 (which is the sum of the received Root Path Cost 0 with the Path Cost 10 of its Root Port) and (iii) Bridge ID = 98 (which is its own Bridge ID).

All other bridges behave in the same way as described to the bridge with Bridge ID = 98.

Ordenação das mensagens de configuração

- **Uma mensagem de configuração C_1 diz-se melhor que outra C_2 se**
  - **o Root ID de C_1 for inferior ao de C_2**
  - **sendo os Root ID idênticos, o Root Path Cost de C_1 for inferior ao de C_2**
  - **sendo idênticos o Root ID e o Root Path Cost, o Bridge ID de C_1 for inferior ao de C_2**
  - **sendo idênticos o Root ID, o Root Path Cost e o Bridge ID, o Port ID de C_1 for inferior ao de C_2**

Root ID	Root Path Cost	Bridge ID	Port ID
18	27	32	2
18	27	32	4
18	27	43	1
18	35	23	3
23	31	45	2

Order of configuration messages

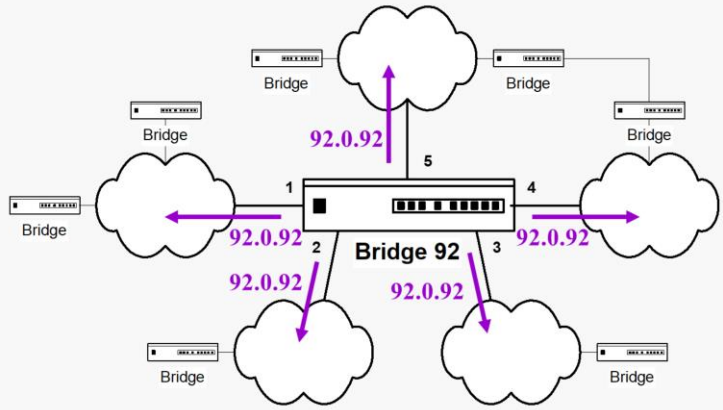
When comparing two Conf-BPDU messages C_1 and C_2, message C_1 is better than message C_2 if:

- the Root ID field is lower in C_1 than in C_2
- the Rood ID field is equal and the Root Path Cost field is lower in C_1 than in C_2
- the Root ID and Root Path Cost fields are equal and the Bridge ID field is lower in C_1 than in C_2
- the Root ID, Root Path Cost and Bridge ID fields are equal and the Port ID field is lower in C_1 than in C_2

The above table shows the content of 5 Conf-BPDU messages ordered from the lowest order message to the highest order message.

Construção da spanning tree (I)

- **Cada bridge assume inicialmente que é a bridge raiz (faz Root Path Cost = 0); envia mensagens de configuração em todas as suas portas**

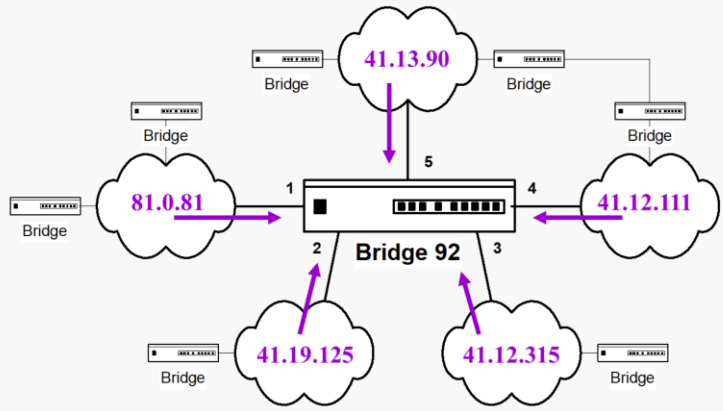


Building the spanning tree (I)

When a bridge is switched on, it starts considering that it is the Root Bridge, since it did not receive yet any Conf-BPDUs from its neighbor switches.

In the above example, the switch with Bridge ID = 92 starts sending the appropriate Conf-BPDU message 92.0.92 through all its ports.

Construção da spanning tree (II)



melhores mensagens recebidas na Bridge 92 até um dado instante

Estimativas da Bridge 92 (assumindo que os custos das portas são unitários)

- **Bridge raiz = 41**
- **Porta raiz = 4**
- **Custo para a raiz = 12 + 1**

Building the spanning tree (II)

In any time instant, the switch assumes that:

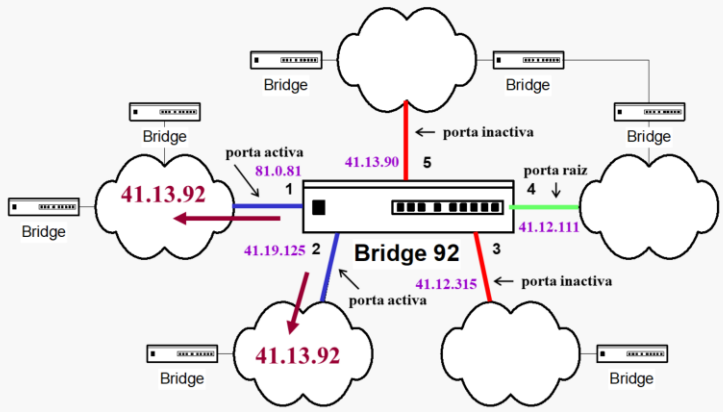
- the Root ID is the lowest Root ID value of the Conf-BPDU messages received from its neighbors;
- its Root Path Cost is given by the Bellman equations;
- its Root Port is the port connecting it to the neighbor switch providing the Root Path Cost.

Consider the above example that shows the content of the Conf-BPDU messages received by a switch from its neighbor switches (consider the Path Cost = 1 in all switch ports).

In the above example, the switch assumes that the Root ID is 41 because this is the lowest value among all received Conf-BPDU messages. Then, it assumes that its Root Port is port 4 since it provides a Root Path Cost of 12 + 1 = 13 which is the lowest value among all its ports.

(continue in the next slide)

Construção da spanning tree (III)



mensagens enviadas pela Bridge 92 - 41.13.92

Building the spanning tree (III)

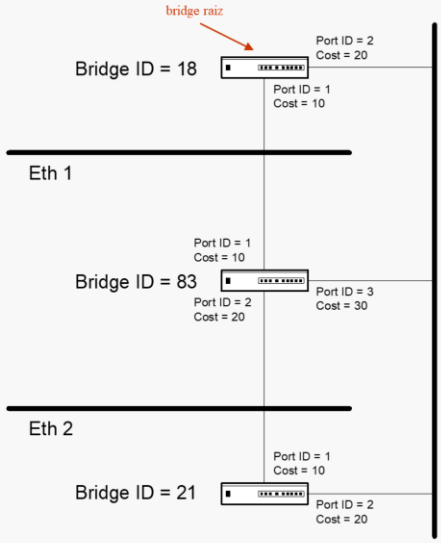
(continue from the last slide)

With the previous information, the Conf-BPDU message to be send to the neighbors is 41.13.92. Besides the Root Port (which becomes active for data frame forwarding), the switch compares its own Conf-BPDU message with the ones received on each port. When its message is better, it sends its message through the port and sets it as a Designated Port (activating the port for data frame forwarding). When its message is worst, it does not send its message and blocks the port for data frames forwarding.

In the above example, the message 41.13.92 is better than message 81.0.81 received on port 1 and is better than message 41.19.125 received on port 2. Therefore, these ports become Designated Ports and are activated for data frame forwarding.

On the other end, the message 41.13.92 is worse than message 41.12.315 received on port 3 and is worse than message 41.13.90 received on port 5. Therefore, these ports are set in blocking state for data frames forwarding.

Construção da spanning tree (IV)



- **A bridge 83 envia 83.0.83 em Eth1, Eth2 e Eth3**
- **A bridge 21 envia 21.0.21 em Eth2 e Eth3**
- **A bridge 83 envia 21.20.83 em Eth1 (porta raiz = 2)**
- **A bridge 18 envia 18.0.18 em Eth1 e Eth3**
- **A bridge 21 envia 18.20.21 em Eth2**
- **A bridge 83 envia 18.10.83 em Eth2 (porta raiz = 1)**

O algoritmo convergiu!

- ✓**A bridge raiz envia periodicamente 18.0.18**
- ✓**A bridge 83 retransmite com 18.10.83 em Eth2**

Building the spanning tree (IV)

Consider the network of bridges in the figure above. Consider the initial state where the bridges are switched on in the following order: Bridge 83, Bridge 21 and Bridge 18.

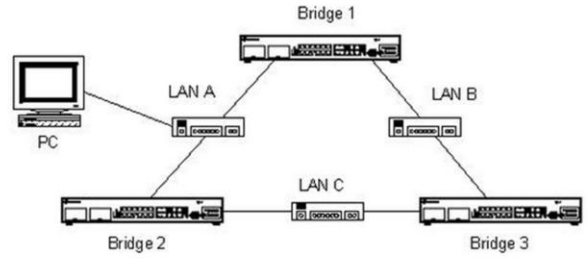
- First, Bridge 83 is switched on, assumes it is the Root Bridge and sends the message 83.0.83 to Eth1, Eth2 and Eth3 (all ports become Designated Ports).
- Then, Bridge 21 is switched on, assumes it is the Root Bridge and sends the message 21.0.21 to Eth 2 and Eth3 (all ports become Designated Ports).
- The previous message is received by Bridge 83 in Eth2 and Eth3. Bridge 83 now knows it is not the Root Bridge. It assumes its port 2 as its Root Port (with Root Path Cost = 20), sends the message 21.20.83 to Eth1 (port 1 becomes the Designated Port of Eth1) and blocks its port 3.
- Then, Bridge 18 is switched on, assumes it is the Root Bridge and sends the message 18.0.18 to Eth1 and Eth 3 (all ports become Designated Ports).
- The previous message is received by Bridge 21 in Eth3. Bridge 21 now knows it is not the Root Bridge. It assumes its port 2 as its Root Port (with Root Path Cost = 20) and sends the message 18.20.21 to Eth2 (port 1 becomes the Designated Port of Eth2).
- Meanwhile, Bridge 83 has assumed its port 1 as its Root Port (with Root Path Cost = 10) and since its message 18.10.83 is better than the message 18.20.21 received in Eth2, it sends its message 18.10.83 to Eth2 (port 2 becomes the Designated Port of Eth2) and activates it.
- The previous message makes Bridge 21 to block port 1 for data frames.

The spanning tree algorithm has converged. From now on, Bridge 18 (which is the Root Bridge) sends periodically the message 18.0.18 to Eth1 and Eth3. Every time this message is received by Bridge 83 on port 1 (its Root Port), it sends the message 18.10.83 to Eth2.

Exercício

Considere a rede seguinte com 3 bridges e o protocolo Spanning Tree activo.
No PC são capturados pacotes BPDU enviados pela Bridge 2 com o seguinte conteúdo:

Root ID: 80000011D823EA30
Root Path Cost: 50
Bridge ID: 800000E07B51A3B0
PortID: 8003

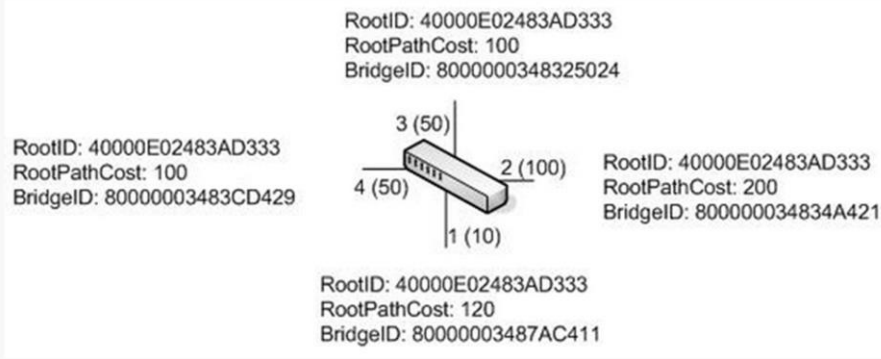


Das afirmações que se seguem, assinale como verdadeiras as afirmações correctas e assinale como falsas as afirmações incorrectas:
a) A bridge 3 é a bridge raíz
b) O BridgeID da bridge 2 é 80000011D823EA30
c) A porta que liga a bridge 2 à LAN C tem custo 50
d) A bridge 2 está ligada à LAN C pela porta número 3
e) A bridge 1 tem uma prioridade menor que 0x8000

30

Exercício

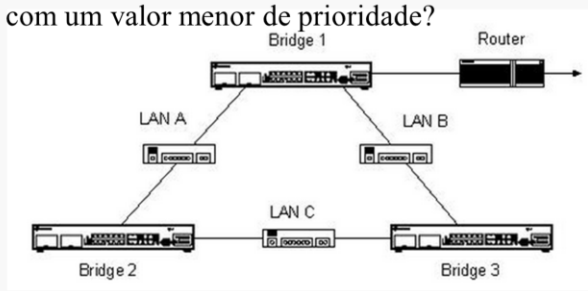
- **Estado de cada porta?**



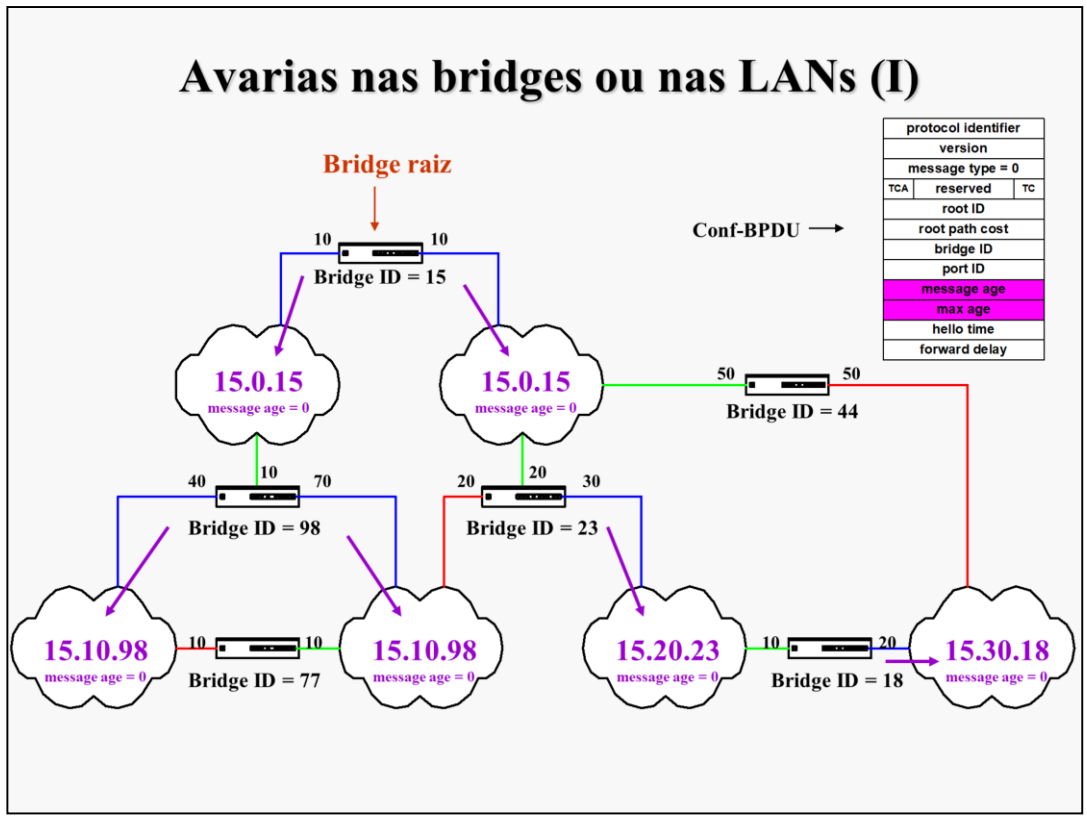
31

Exercício

- Colocar uma das portas da Bridge 2 bloqueada
  - bridge 2 com o maior valor de prioridade; as portas da bridge 2 com um custo 100; as portas das bridges 1 e 3 com um custo 10?
  - bridge 2 com o maior valor de prioridade?
  - portas da bridge 2 com um custo 100; as portas das bridges 1 e 3 com um custo 10?
  - bridges 1 e 3 com um valor de prioridade menor que a bridge 2; as portas de todas as bridges configuradas com o mesmo custo?
  - bridge 1 com um valor menor de prioridade?



32

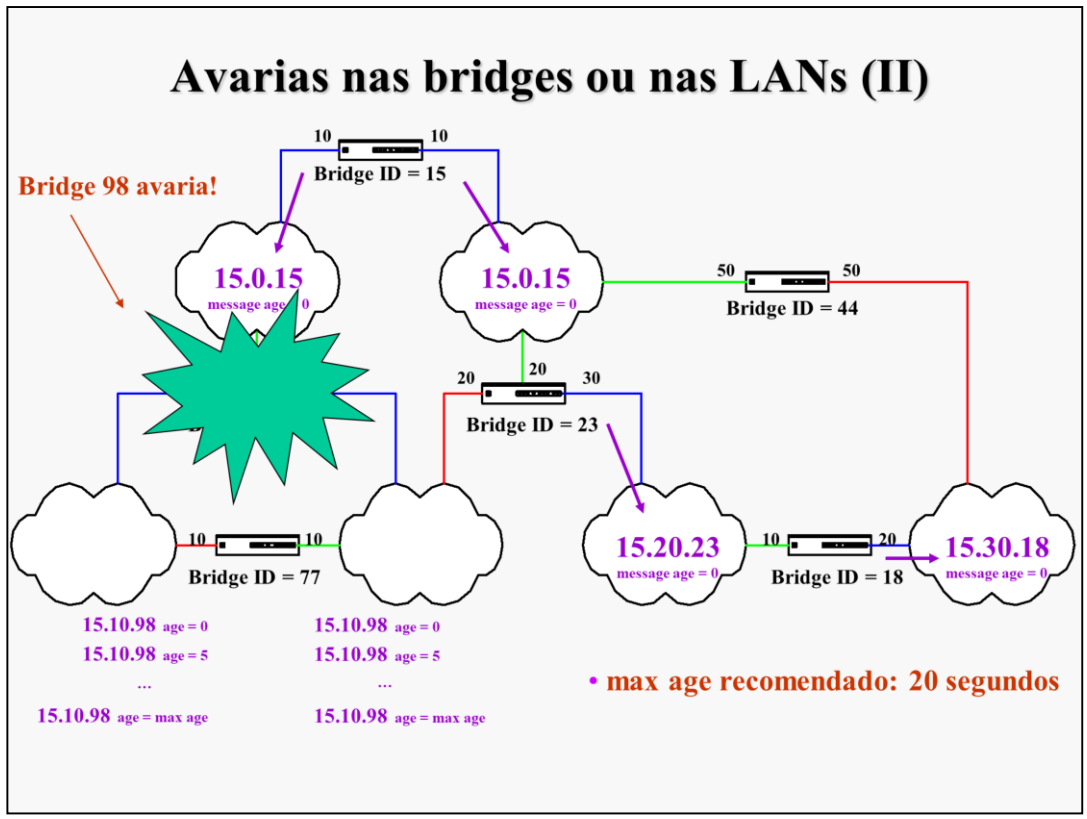


Failures in bridges or LANs (I)

The Root Bridge sends periodically the Conf-BPDU messages with the **message age** field set to zero and the **max age** field set to a preconfigured value (the recommended **max age** value is 20 seconds).

When a bridge receives a Conf-BPDU, it sends its own Conf-BPDUs also with the **message age** field set to zero and the **max age** field received from the Root Bridge.

Therefore, the age information is propagated over the whole network.

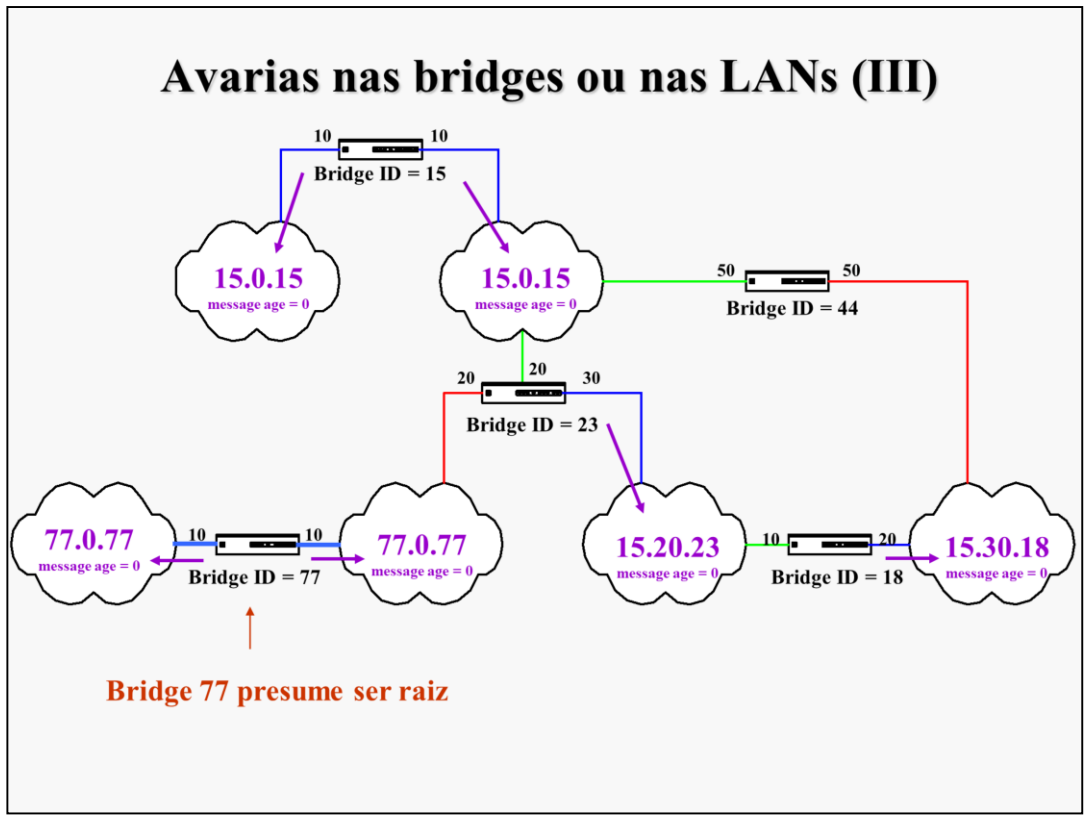


Failures in bridges or LANs (II)

When a bridge (or a LAN) fails in the path from the Root Bridge to a particular bridge, this bridge stops receiving Conf-BPDU messages from the Root Bridge. Therefore, the age of information received in the last Conf-BPDU message will reach max age seconds and the information will be discarded.

In the above example, one bridge has failed and the bridge with Bridge IB = 77 stopped receiving Conf-BPDU messages in both its ports.

(continue in the next slide)

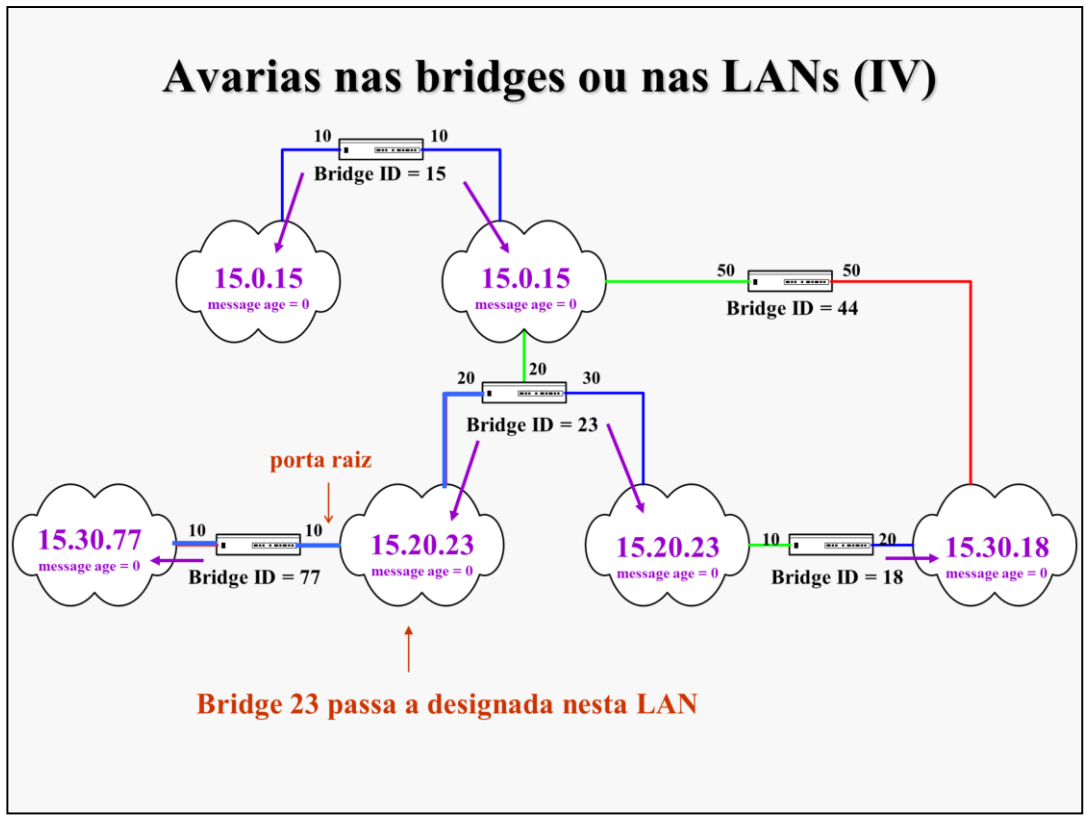


Failures in bridges or LANs (III)

(continue from the last slide)

In this case, since the information discarded is related to both bridge ports, the bridge assumes it is the Root Bridge and sends the appropriate Conf-BPDU message 77.0.77.

(continue in the next slide)



Failures in bridges or LANs (IV)

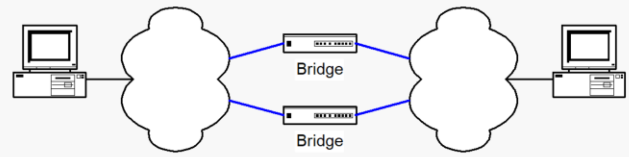
(continue from the last slide)

Since the bridge with Bridge ID = 23 has a better message (with content 15.20.23) than the message received (with content 77.0.77), in the next time it receives a Conf-BPDU message in its Root Port, it will send its Conf-BPDU message and activate the port connecting again the whole network.

The Conf-BPDU message with content 15.20.23 enables the bridge with Bridge ID = 77 to select appropriately its own Root Port and to set the other port as a Designated Port.

Existência de ciclos temporários

- **Após alteração da topologia da rede:**
  - **pode existir perda temporária de conectividade se uma porta que estava inactiva na topologia antiga ainda não se apercebeu que deverá estar activo na nova topologia**
  - **podem existir ciclos temporários se uma porta que estava activa na topologia antiga ainda não se apercebeu que deverá estar inactiva na nova topologia**
- **Para minimizar a probabilidade de se formarem ciclos temporários as bridges são obrigadas a esperar algum tempo antes de permitirem que uma das suas portas passe do estado inactivo para o estado activo; o tempo de espera é função do parâmetro forward delay**



Existence of Routing cycles

When the topology of the network changes, the configured spanning tree might also change. In this case:

- some pairs of terminals might experience connectivity lost for some time, if a previous inactive port takes too long to become active;
- temporarily routing cycles might exist, if a previous active port takes too long to become inactive.

Routing cycles impose instability in the network and must be prevented (at the cost of penalizing connectivity lost).

To minimize the probability of routing cycles establishment, bridges/switches set some delay when a port is to change from an inactive state to an active state. This delay is controlled by the **forward delay** parameter, which is recommended by the standard to be 15 seconds.

Estados das portas da bridge

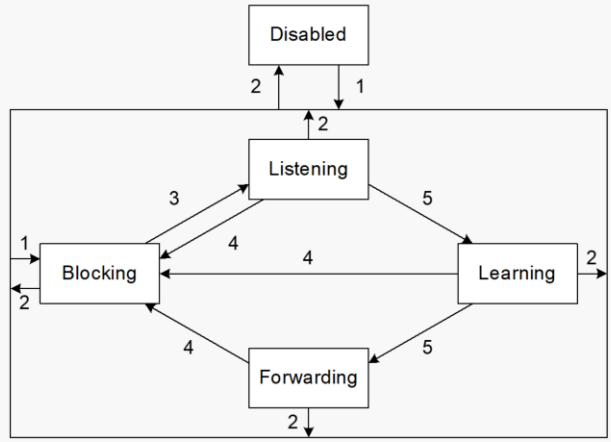
- **Estado** blocking**: os processo de aprendizagem e de expedição de pacotes estão inibidos; recebe e processa mensagens de configuração**
- **Estado** listening**: os processo de aprendizagem e de expedição de pacotes estão inibidos; transita para o estado** learning **após um tempo de permanência neste estado igual a** forward delay**; recebe e processa mensagens de configuração**
- **Estado** learning**: o processo de aprendizagem está activo mas o processo de expedição de pacotes está inibido; transita para o estado** forwarding **após um tempo de permanência neste estado igual a** forward delay**; recebe e processa mensagens de configuração**
- **Estado** forwarding**: é o estado activo; tanto o processo de aprendizagem com o processo de expedição de pacotes estão activos; recebe e processa mensagens de configuração**
- **Estado** disabled**: os processo de aprendizagem e de expedição de pacotes estão inibidos; não participa no algoritmo de spanning tree**

Port states on bridges/switches

Each port can be in one of the following states:

- Blocking state: this is the inactive state; both MAC address learning and data frame forwarding/flooding processes are disabled; the reception and processing of BPDU messages is enabled.
- Listening state: both MAC address learning and data frame forwarding/flooding processes are disabled; the reception and processing of BPDU messages is enabled; the port moves to the learning state after forward delay seconds.
- Learning state: the MAC address learning process is enabled but the data frame forwarding/flooding process is disabled; the reception and processing of BPDU messages is enabled; the port moves to the forwarding state after forward delay seconds.
- Forwarding state: this is the active state; both MAC address learning and data frame forwarding/flooding processes are enabled; the reception and processing of BPDU messages is enabled.
- Disabled state: the port is disabled by management both for data and BPDU frames.

Diagrama de estados das portas



1 - Porta activada por gestão ou por inicialização
2 - Porta desactivada, por gestão ou falha
3 - Algoritmo selecciona como sendo porta designada ou porta raiz
4 - Algoritmo selecciona como não sendo porta designada ou porta raiz
5 - Forwarding timer expira

Port state transition diagram

The above diagram specifies the possible transitions between the states of each port.

Note that the port can move from any state directly to the blocking state but the opposite is not possible: to move from the blocking state to the forwarding state, the port must first move to the listening state (staying there for forward delay seconds), then, to the learning state (staying there for an additional forward delay seconds) and, finally, to the forwarding state.

Tempo de vida das entradas das tabelas de encaminhamento

- **Tempo de vida demasiado longo – pode haver um número exagerado de pacotes perdidos quando a estação muda de localização.**
- **Tempo de vida demasiado curto – o tráfego na rede pode ser exagerado devido ao processo de *flooding***

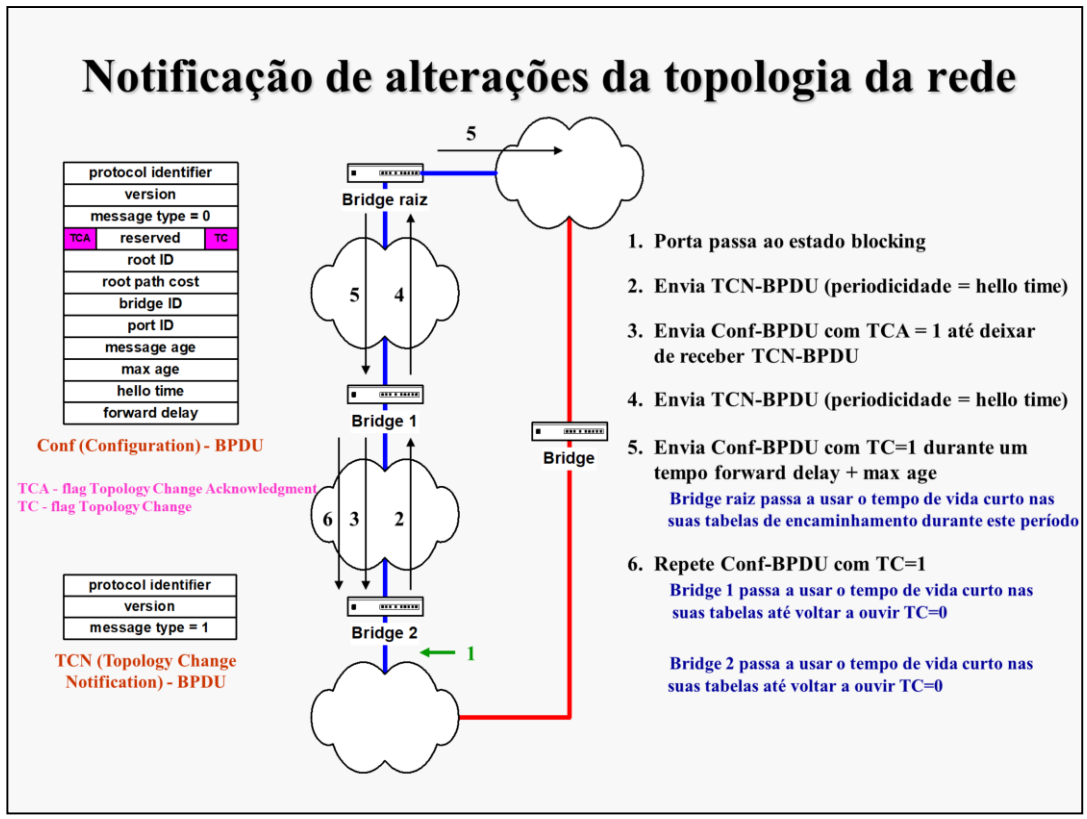
- **Existem dois tempos de vida:**
  - **Longo: usado por defeito (valor recomendado = 5 minutos)**
  - **Curto: usado quando a spanning tree está em reconfiguração (valor recomendado = 15 segundos) - exige processo de notificação de alterações da topologia da rede**

Aging time of MAC Address Table entries

The aging time of each entry of the MAC Address Table should be neither too long (which penalizes connectivity lost when the entry becomes invalid, due to terminal location change, for example) nor too short (which penalizes the network load due to more flooded frames).

The standard proposes two values for the aging time:

- the long aging time value (recommended value: 5 minutes) – used by default
- the short aging time value (recommended value: 15 seconds) – used when the spanning tree is being reconfigured, which requires a Topology Change Notification process (next slide).



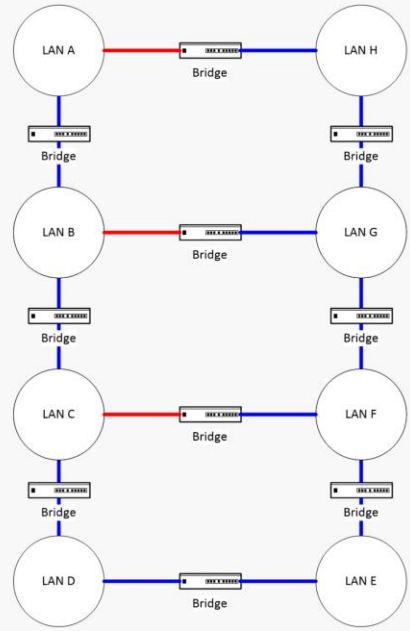
Topology Change Notification process

The Topology Change Notification process involves the flags TC (Topology Change) and TCA (Topology Change Acknowledgement) of the Conf-BPDU messages and the second type of BPDU messages: the TCN-BPDU messages. This process is triggered by any bridge/switch that changes (for some reason) the state of one of its ports. The idea is that the bridge detecting a change of the spanning tree notifies this fact to the Root Bridge and the Root Bridge notifies all bridges to use the short aging time on their MAC Address Tables. Consider the above example where bridge 2 has a port that has changed from the forwarding state to the blocking state:

- bridge 2 starts sending TCN-BPDU messages (with a periodicity of Hello Time) through its Root Port until the next received Conf-BPDU message has the TCA flag set to 1;
- bridge 1, when receiving a TCN-BPDU message in a Designated Port, sets the next Conf-BPDU message to be sent to the same Designated Port with the flag TCA set to 1 (to notify the other bridge that it has received the TCN-BPDU message); then, bridge 1 starts sending TCN-BPDU messages (with a periodicity of Hello Time) through its Root Port until the next Conf-BPDU message has the TCA flag set to 1;
- the Root Bridge, when receiving a TCN-BPDU message in a Designated Port, sets the next Conf-BPDU message to be sent to the same Designated Port with the flag TCA set to 1 (to notify the other bridge that it has received the TCN-BPDU message); at the same time, the Root Bridge starts using the short aging time and starts sending the periodic Conf-BPDU messages with the flag TC set to 1 for all its Designated Ports;
- all other bridges, when receiving a Conf-BPDU message on their Root Port with the flag TC = 1, start using the short aging time and send their Conf-BDPU messages also with TC = 1;
- the Root Bridge sends the Conf-BPDU messages with TC = 1 during forward delay + max age seconds (a time duration which is considered enough for the spanning tree to reconfigure);
- all other bridges, when receiving a Conf-BPDU message on their Root Port with the flag TC = 0, start using again the long aging time and send their Conf-BDPU messages also with TC = 0.

Desempenho do encaminhamento via spanning tree

Exemplo de encaminhamento via spanning tree ineficiente



Performance of spanning tree routing

Note that the spanning tree routing might be inefficient. In the above example, the frames exchange between terminal hosts attached to LAN A and terminal hosts attached to LAN H will have a poor performance since these frames are routed over 7 different bridges and occupy transmission resource on all other LANs. In fact, other spanning tree solutions can have a better overall performance by, for example, minimizing the average number of routing hops between all pairs of LANs.

Nonetheless, the network manager can always set the desired spanning tree by setting appropriate Bridge ID values and/or Path Cost values on the bridges/switches.

Rapid Spanning Tree Protocol (RSTP)

- **Norma IEEE 802.1W**
- **RSTP providencia uma convergência muito mais rápida que o STP**
  - **STP pode demorar 30 a 50 segundos a convergir após uma alteração de topologia**
  - **O RSTP redefine os estados possíveis das ligações e o algoritmo de convergência dos switches**
  - **RSTP consegue tempos de convergência de 6 segundos (no caso de falha de switches) ou de poucos milisegundos (no caso de falhas de ligação).**
- **Standard IEEE 802.1D-2004 inclui o RSTP e torna obsoleto o IEEE 802.1D original.**

Rapid Spanning Tree Protocol (RSTP)

The Rapid Spanning Tree Protocol (RSTP) was first defined in the standard IEEE 802.1W.

RSTP is similar to STP in the way the spanning tree is set based on the Bridge ID and Path Cost values. Nevertheless, it provides a much faster spanning tree convergence time:

- STP can take some 30 to 50 seconds to converge to the new spanning tree when there is a network topology change.
- RSTP redefines the possible states of the bridge/switch ports and the way the switches exchange BPDU messages between them.
- RSTP obtains convergence times of up to 6 seconds (in case of switch failures) or just a few milliseconds (in the case of connection failures).

Meanwhile, a new standard (the IEEE 802.1D-2004) was proposed which includes RSTP and makes obsolete the original IEEE 802.1D standard.

Multiple Spanning Tree Protocol (MSTP)

- **O MSTP, originalmente definido na norma IEEE 802.1S, foi mais tarde integrado na norma IEEE 802.1Q-2005**
- **O MSTP permite agrupar as VLANs e configurar uma Spanning Tree por cada grupo de VLANs.**
  - **Quando existem múltiplas VLANs, uma única Spanning Tree funciona mas não permite balancear o tráfego pelas diferentes ligações.**
  - **Com o MSTP, é possível usar todas as ligações em que cada ligação está ativa para umas spanning trees e inactiva para outras spanning trees.**

Multiple Spanning Tree Protocol (MSTP)

The Multiple Spanning Tree Protocol (MSTP), originally defined in the standard IEEE 802.1S, was later included in the VLAN protocol standard IEEE 802.1Q-2005.

MSTP enables to group the different VLANs and to configure one Spanning Tree for each group of VLANs:

- When the network supports multiple VLANs, one single Spanning Tree supporting all VLANs works but this solution does not enable the balancing of the data traffic through the different network connections.
- With MSTP, it is possible to balance the data traffic through all connections where some connections are active for some Spanning Trees and inactive for other Spanning Trees.