Camada de Ligação (Ethernet)

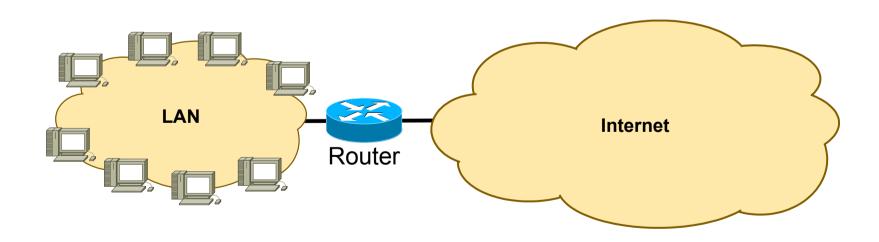
Redes e Serviços

Licenciatura em Engenharia Informática DETI-UA

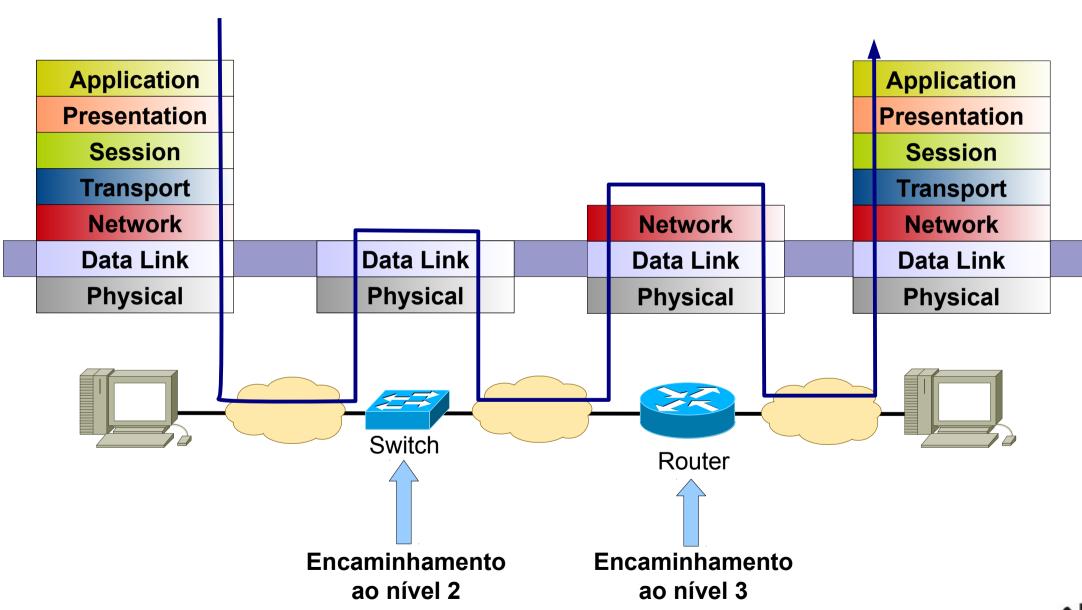


Redes Locais (LAN)

- Interligam estações relativamente "próximas" através de ligações partilhadas
- Tecnologias
 - Ethernet, Token Ring, 802.11, FDDI, ...



Camada de Ligação



Equipamentos de Interligação (resumo)

• Hub/Repetidor:

- Opera no nível físico (camada OSI 1)
- Regenera os sinais
- Hub = repetidor com múltiplas portas
- Não é usado nas redes locais actuais

Switch/Bridge:

- Tipo store-and-forward
- Opera ao nível da camada da ligação (camada OSI 2)
- Interliga dois ou mais domínios de colisões
- Comuta com base nos endereços MAC
- Switch = bridge com múltiplas portas

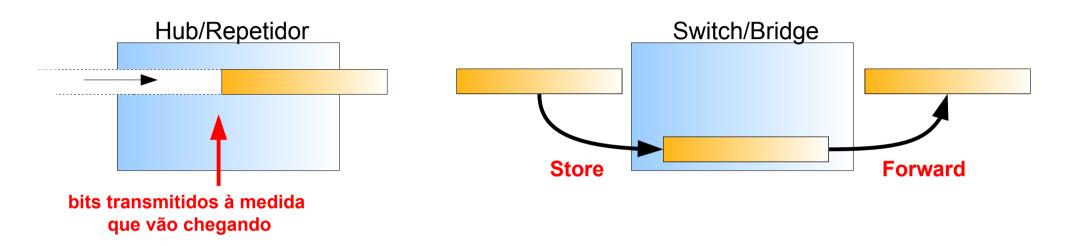
Router:

- Tipo store-and-forward
- Opera no nível de rede (camada OSI 3)
- Comuta com base nos endereços de nível 3 (ex. IP, IPX, AppleTalk)

Ethernet (802.3)

- Tecnologia LAN mais bem sucedida
- Inventada pela Xerox Palo Alto Research Center (PARC)
- A Xerox, DEC e Intel definiram em 1978 a norma para a Ethernet 10Mbps
- Usa "Carrier Sense/Multiple Access" with "Collision Detect" (CSMA/CD)
 - Carrier Sense: os nós conseguem perceber se o canal de comunicação está ocupado.
 - Multiple Access: múltiplos nós podem aceder simultaneamente.
 - Collision Detect: os nós "escutam" a rede enquanto transmitem, podem detetar colisões.

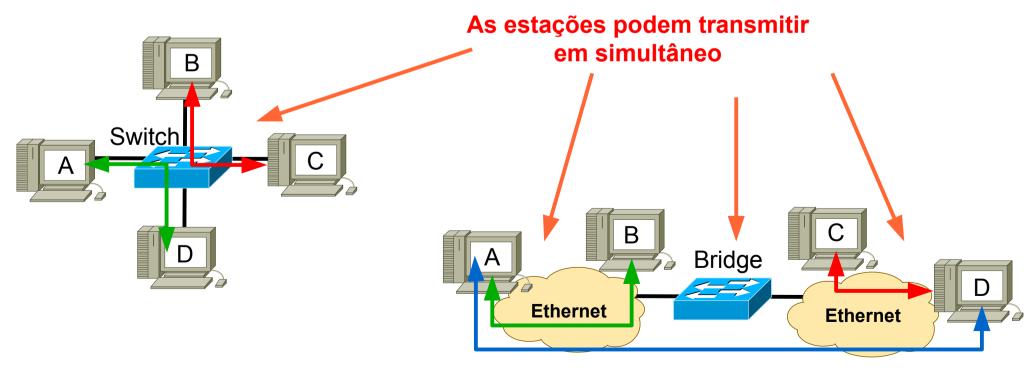
Switches/Bridges versus Hubs/Repetidores (I)



- Os repetidores/hubs interligam segmentos de LANs
- As bridges/switches interligam diferentes LANs
- Funções adicionais das bridges/switches:
 - Store & Forward + Filtragem
 - Em vez de expedir os pacotes para todas as portas pode expedir apenas para a porta da estação destino
 - As portas podem operar a diferentes taxas de transmissão

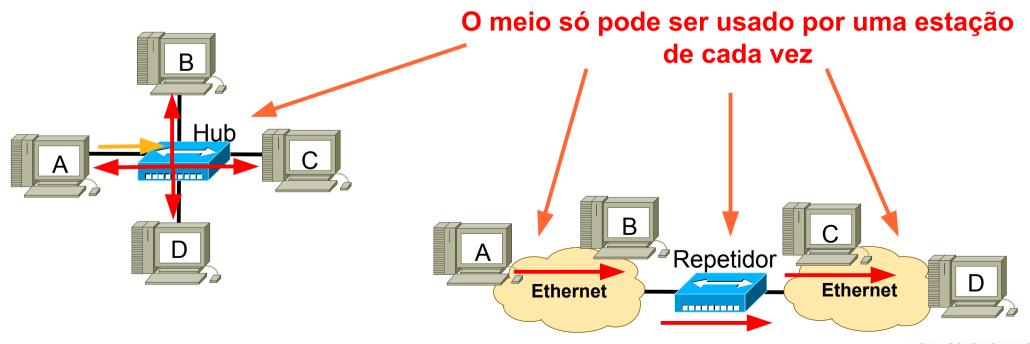
Switches/Bridges versus Hubs/Repetidores (II)

- Com Switches/Bridges
 - As colisões deixam de ser um problema
 - A interligação é feita ao nível da ligação (camada OSI 2)



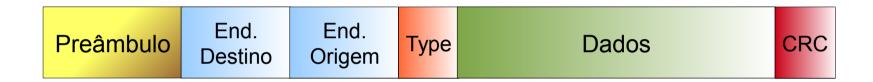
Repetidores/Hubs definem Domínios de Colisão

- Os hubs/repetidores interligam segmentos do mesmo tipo de LANs
 - Interligação apenas ao nível físico (camada OSI 1)
- Os hubs já (praticamente) não existem nas redes locais actuais



Trama Ethernet (1)

O adaptador de rede do emissor encapsula o datagrama IP (ou outro protocolo da camada de rede) na trama Ethernet

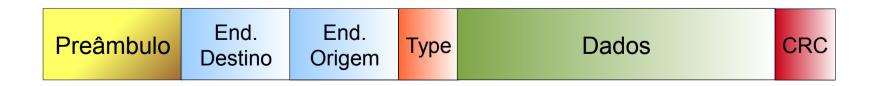


• Preâmbulo:

- 7 bytes com o padrão 10101010 seguidos de um byte com o padrão 10101011
- É usado para sincronizar os relógios de emissão (e recepção) do emissor e do receptor.

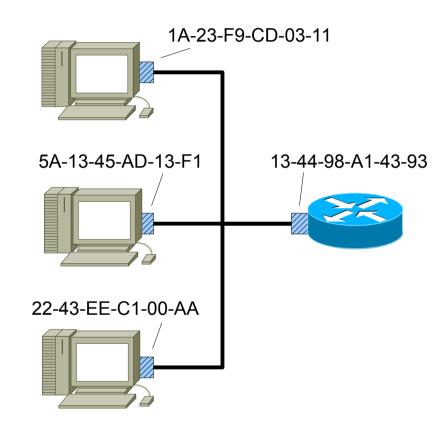
Trama Ethernet (2)

- Endereços: 6 bytes
 - Se o adaptador de rede receber uma trama com um endereço de destino igual ao seu, ou se for o endereço de broadcast, passa os dados da trama ao protocolo da camada de rede
 - Senão, o adaptador descarta a trama
- Type: Indica qual o protocolo da camada de rede (tipicamente IP, mas pode ser IPX, AppleTalk, etc...)
- CRC: Usado para detecção de erros de transmissão
 - Se for detectado um erro no receptor a trama é descartada



Endereços MAC

- Endereço MAC (LAN ou físico ou Ethernet):
 - Função: permitir a transmissão de tramas entre interfaces ligados fisicamente (que estão na mesma rede)
 - Tem 48 bits (para a maioria das LANs)
 - Embebido de origem na ROM da placa de rede
 - Algumas placas permitem a sua alteração por software
 - Notação hexadecimal
 - Cada placa/adaptador de rede tem um endereço MAC
 - Endereço de broadcast: FF-FF-FF-FF-FF



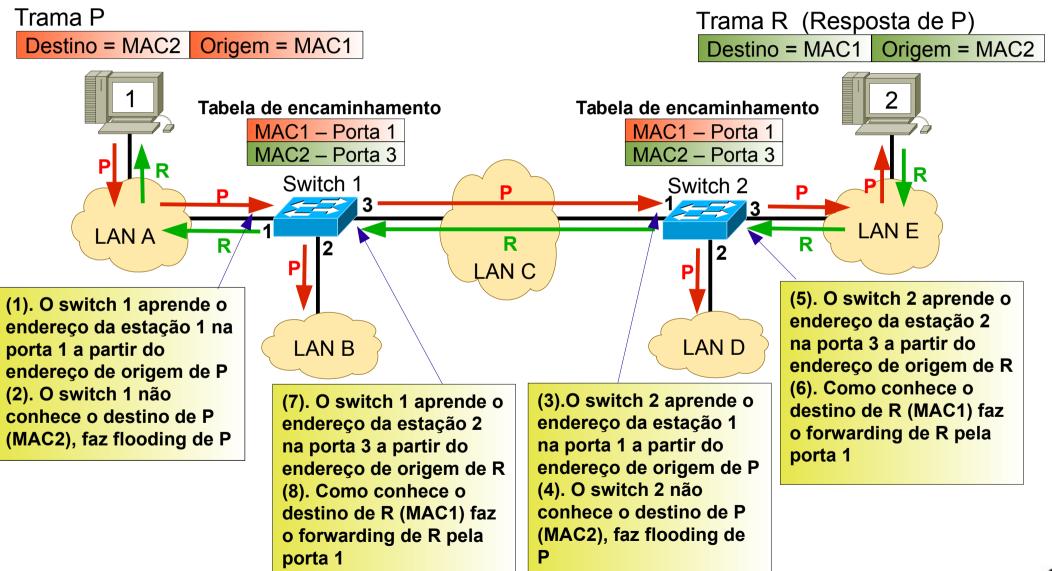
Operações básicas de um switch

- Os switches possuem uma tabela de encaminhamento (camada 2)
- Quando um switch recebe uma trama MAC numa porta:

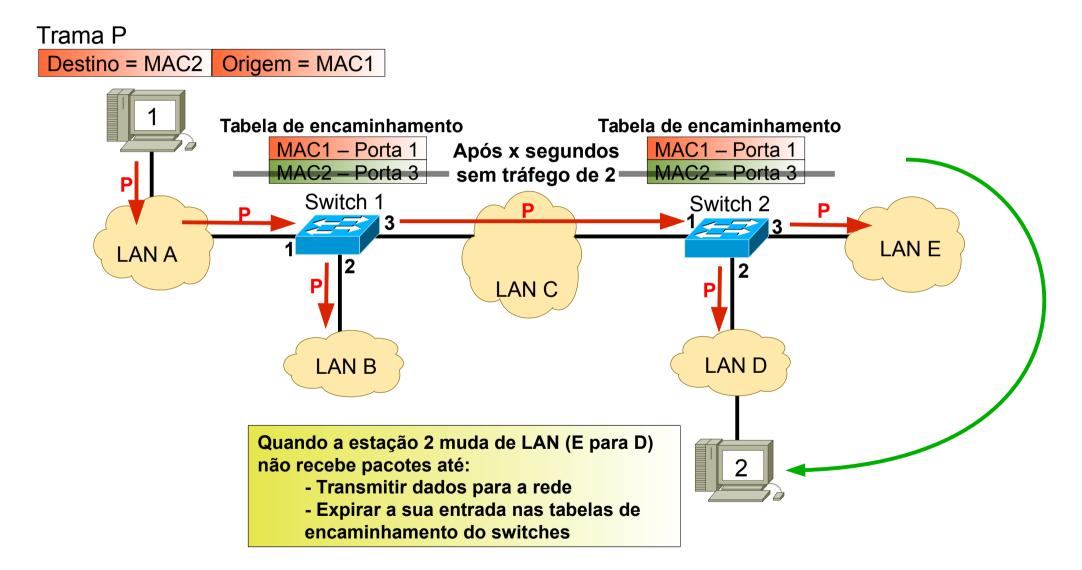
MAC	Porta
00:11:11:11:11:11	1
00:22:22:22:22	1
A1:33:33:33:33	2
44:44:44:44:44	3
55:55:55:00:00:55	3

- Regista na sua tabela de encaminhamento a porta em que recebeu a trama e o endereço MAC de origem da trama
- Procura o endereço MAC de destino da trama na tabela de encaminhamento para reencaminhar a trama
- Mecanismo de forwarding:
 - Quando o endereço MAC destino da trama existe na tabela de encaminhamento, o switch envia a trama apenas pela porta registada na sua tabela
- Mecanismo de flooding:
 - Quando o endereço MAC destino da trama não existe na tabela de encaminhamento, o switch envia a trama por todas as portas exceto pela porta em que recebeu a trama

Aprendizagem de endereços em switches

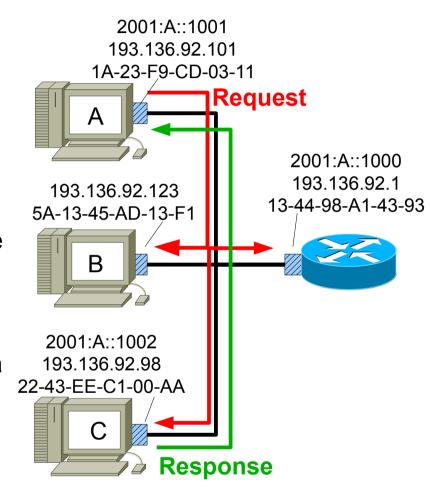


Tempo de vida das entradas das tabelas de encaminhamento



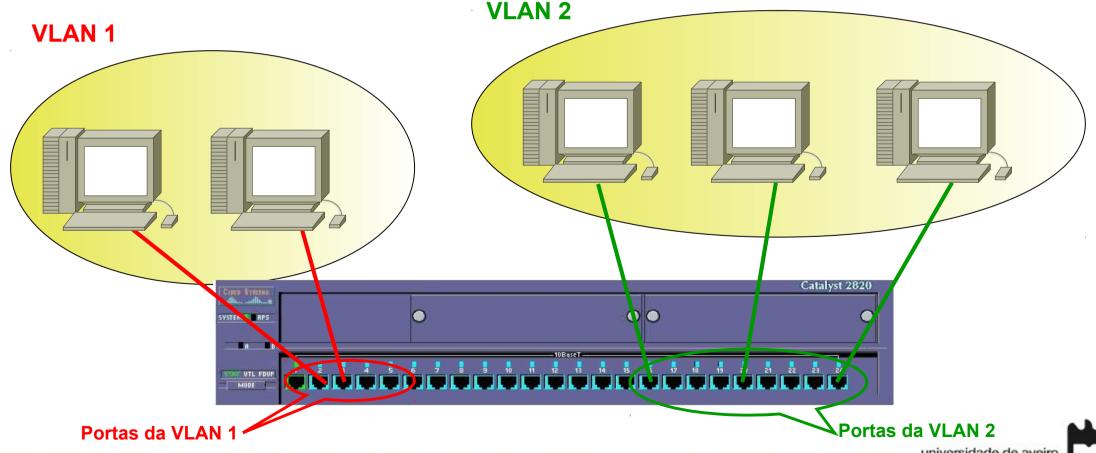
Resolução de Endereços Físicos

- Quando "A" precisa de contactar "C" por IP
 - "A" precisa do endereço MAC de "C"
 - Só sabe o endereço IPv4 ou IPv6
 - Se o endereço de "C" não estiver na tabela de endereços físicos conhecidos
 - IPv4: Tabela ARP
 - → IPv6: Tabela de vizinhos (NDP)
 - "A" envia um pedido para a rede com o endereço IP de "C" e a pedir o endereço MAC de C
 - IPv4: pacote ARP, enviado em broadcast (para todos)
 - → IPv6: pacote ICMPv6, enviado em multicast (para um grupo restrito)
 - "C" verifica que o seu IP vem no pedido e responde diretamente para "A" (MAC destino=MAC de A) com a indicação do seu MAC.
 - "A" atualiza a sua tabela ARP/NDP
- A resolução de endereços físicos apenas existe dentro da mesma rede local (mesmo domínio de broadcast)



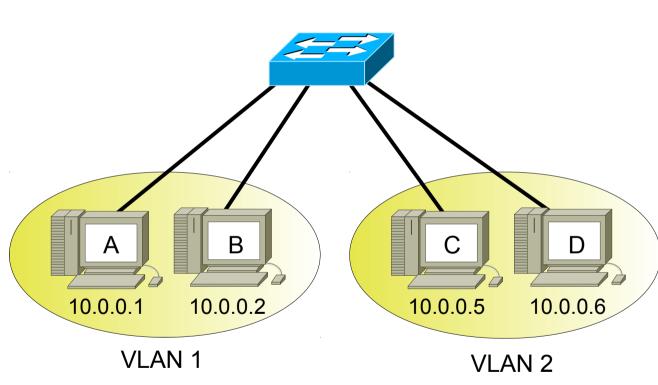
LANs virtuais (VLAN)

- Uma LAN virtual (VLAN) é um grupo de máquinas/utilizadores com um conjunto de requisitos/características comuns que comunicam no mesmo domínio de broadcast.
 - Independente da sua localização física (switch a que estão ligados).
- Máquinas em VLAN diferentes não comunicam ao nível da camada 2.
 - Para comunicar têm de o fazer ao nível da camada 3.



Exemplo – VLANs

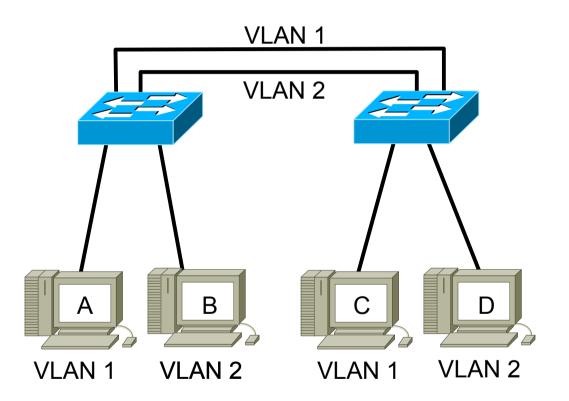
Ping enviado por 10.0.0.1

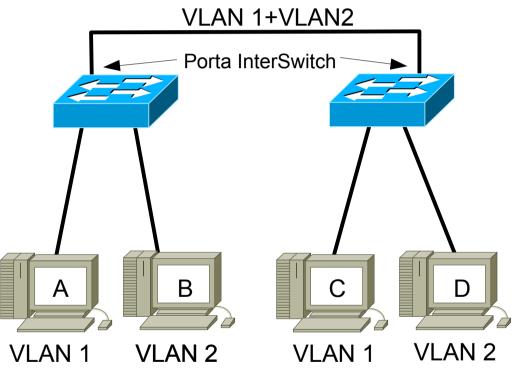


```
# ping 10.0.0.2
Pinging 10.0.0.2 with 32 bytes of data:
Reply from 10.0.0.2: bytes=32 time<10ms TTL=128
Ping statistics for 10.0.0.2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 0ms, Average = 0ms
# ping 10.0.0.5
Pinging 10.0.0.5 with 32 bytes of data:
Request timed out.
Request timed out.
Request timed out.
Request timed out.
Ping statistics for 10.0.0.5:
    Packets: Sent = 4, Received = 0, Lost = 4 (100% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 0ms, Average = 0ms
# ping 10.0.0.6
Pinging 10.0.0.6 with 32 bytes of data:
Request timed out.
Request timed out.
Request timed out.
Request timed out.
```

Interligação de switches

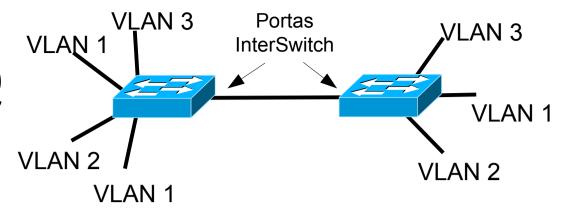
- Usando uma ligação física por VLAN
- Usando porta(s) InterSwitch
 - Também designadas por trunk(s)





É necessária uma forma de distinguir os pacotes de diferentes VLANs

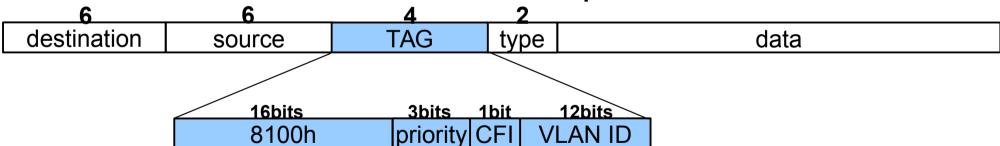
Norma IEEE802.1Q



Pacote Ethernet sem etiqueta VLAN

6	6	2	
destination	source	type	data

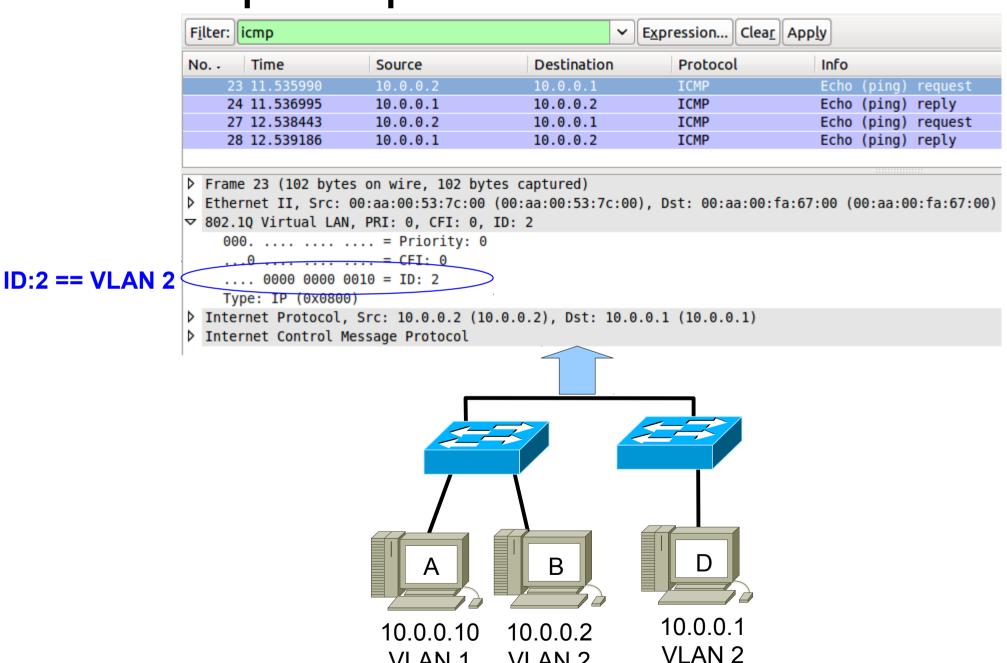
Pacote Ethernet com etiqueta VLAN



- Priority: Prioridade do tráfego de acordo com a norma 802.1q (valores 0 a 7)
- CFI: usado para compatibilidade com tecnologias mais antigas (sempre zero na Ethernet)
- VLAN ID: identificados da VLAN



Exemplo – portas InterSwitch/Trunk

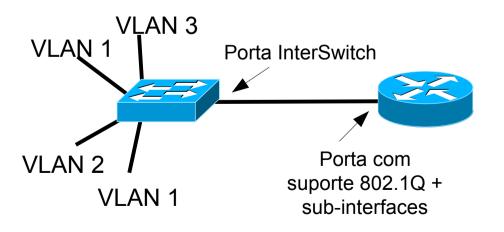


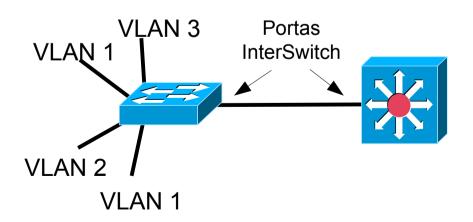
VLAN 2

VLAN 1

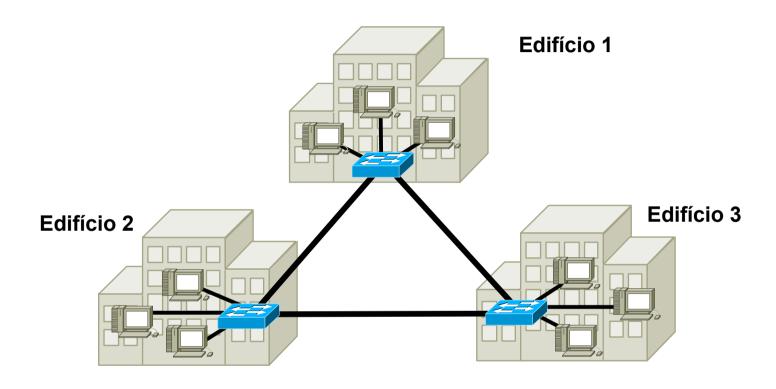
Encaminhamento entre VLANs

- Para comunicar entre VLAN diferentes é necessário um router.
 - É equivalente à comunicação entre duas LAN distintas.
- As soluções mais comuns são:
 - Usar um router com suporte do protocolo 802.1Q,
 - Ligar um porta de inter-switch ao interface físico de um router,
 - O interface físico é sub-dividido em sub-interfaces, um para cada VLAN.
 - Usar um switch L3
 - Faz encaminhamento entre as diferentes VLAN.



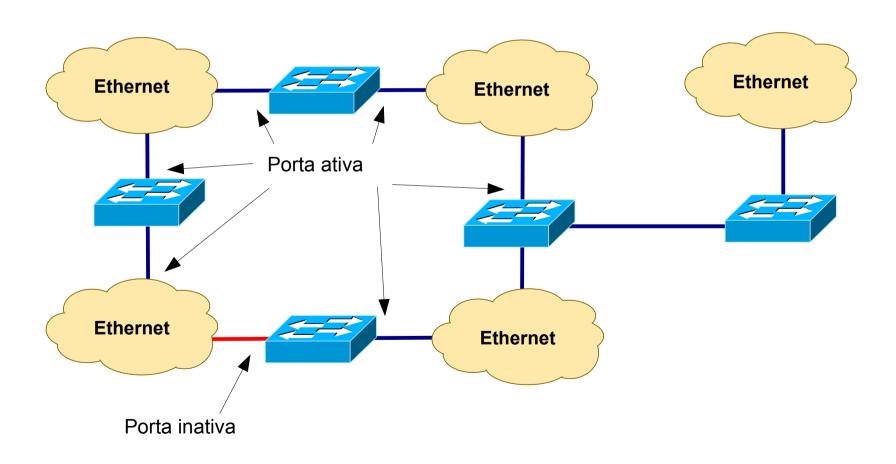


Interligação de Switches com redundância

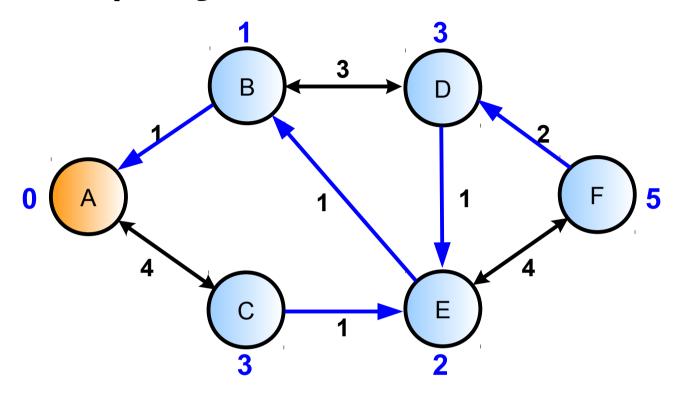


- Objectivo: proporcionar capacidade de recuperação a falhas na rede
- Problema: redundância lógica provoca colapso de comunicações devido ao envio de tramas MAC para o endereço de broadcast

Spanning tree numa rede de LANs (LAN estendida)



Equações de Bellman



Quando os custos das ligações são não negativos, então:

Comprimento do percurso mínimo de um nó para A



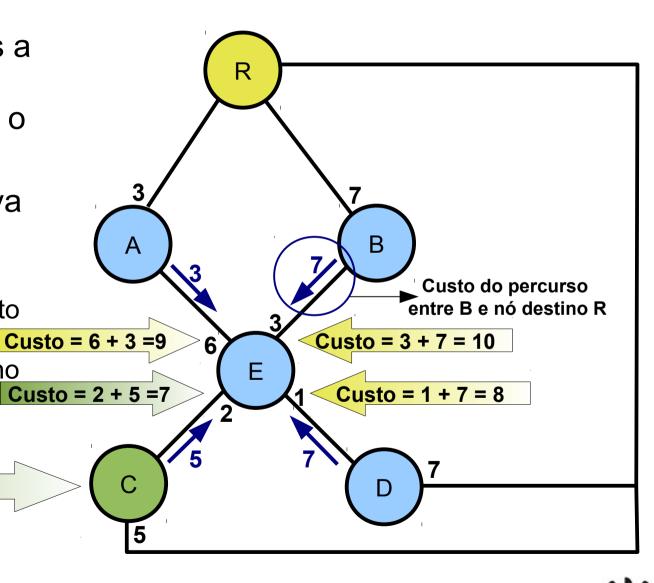
Comprimento do arco que une esse nó ao nó que se lhe segue no percurso mínimo



Algoritmo de Bellman-Ford distribuído e assíncrono

- Cada nó transmite de tempos a tempos a sua estimativa do custo do percurso entre ele e o nó destino
- Cada nó recalcula o estimativa do custo até ao nó destino:
 - Somando as estimativas recebidas dos vizinhos ao custo do ligação/porta por onde receberam o anúncio do vizinho
 - Escolhem o menor valor

Vizinho escolhido por E para encaminhar para o nó destino R



Encaminhamento baseado em Spanning Trees

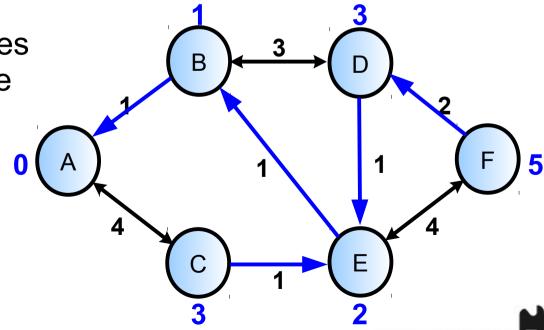
- É escolhido um switch como nó origem/raiz
- Os outros switches usam o algoritmo de Bellman-Ford assíncrono e distribuído para calcular o vizinho no percurso de custo mínimo para o nó origem/raiz

 As ligações compostas pelos percursos de custo mínimo (de todos os outros switches para a origem/raiz) definem uma árvore abrangente

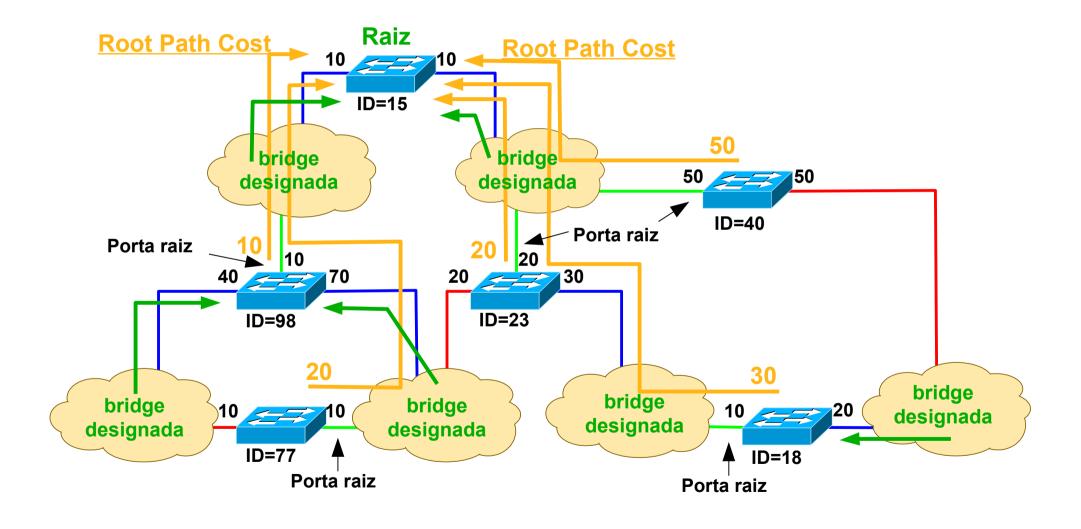
(spanning tree)

 As portas ativas são as das ligações que compõem a árvore abrangente

 É necessário um critério para desempatar quando há múltiplos percursos de custo mínimo

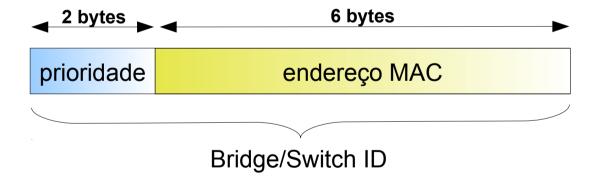


Conceitos básicos spanning tree (I)



Conceitos básicos spanning tree (II)

- Bridge/Switch ID cada switch é identificado por um endereço que contém:
 - 2 octetos de prioridade, configurável pelo gestor da rede
 - 6 octetos fixos (um dos endereços MAC das portas do switch, ou qualquer outro endereço único de 48 bits)
 - A prioridade tem precedência sobre o campo de octetos fixos

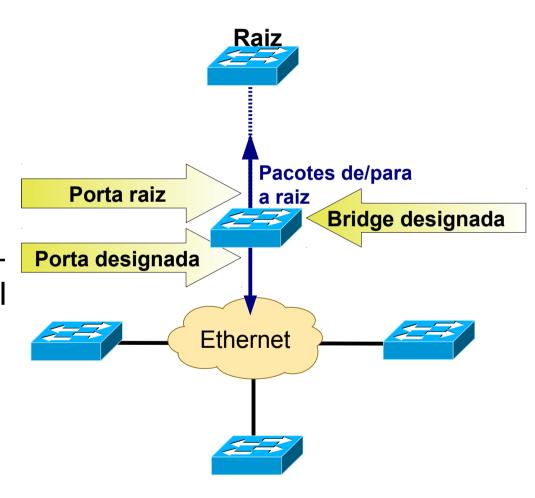


- Switch/bridge raiz (Root) switch que está na raiz da spanning tree
 - Switch com menor ID
- Path cost custo associado a cada porta do switch
 - Pode ser configurado pelo gestor da rede



Conceitos básicos spanning tree (III)

- Bridge designada (Designated Bridge) – bridge que, numa LAN, é responsável pelo envio de pacotes da LAN para a 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 da LAN para a raiz e vice-versa (uma das portas da bridge designada)
- Porta raiz (Root Port) porta responsável pela recepção/transmissão de pacotes de/para a bridge raiz



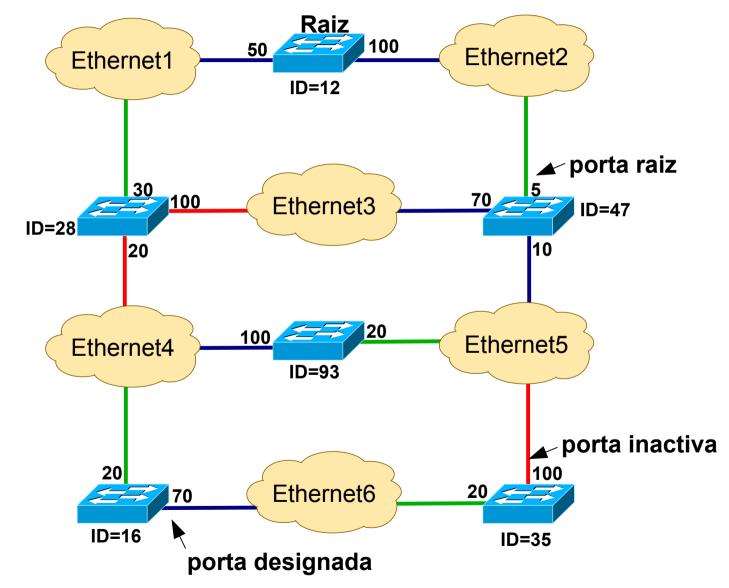
Conceitos básicos spanning tree (IV)

- Cada bridge tem associado um custo do <u>percurso para a raiz</u> (<u>Root Path Cost</u>), igual à soma dos custos das portas que recebem pacotes enviados pela raiz (portas raiz) no percurso de menor custo para a bridge
- 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)

Exemplo – spanning tree (I)

Bridges designadas

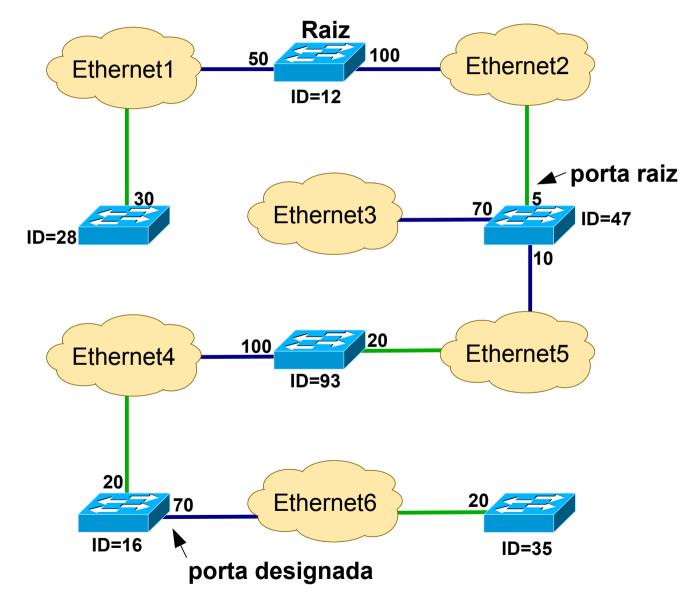
Eth1	12	
Eth 2	12	
Eth 3	47	
Eth 4	93	
Eth 5	47	
Eth 6	16	



Exemplo – spanning tree (II)

Bridges designadas

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



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

IEEE 802.3 Ethernet

Destination: 01:80:c2:00:00:00 (01:80:c2:00:00:00)

Source: 00:16:e0:9a:c3:92 (00:16:e0:9a:c3:92)

Length: 39

Logical-Link Control

DSAP: Spanning Tree BPDU (0x42) SSAP: Spanning Tree BPDU (0x42)

Control field: U, func=UI (0x03)

Spanning Tree Protocol

Protocol Identifier: Spanning Tree Protocol (0x0000)

Protocol Version Identifier: Spanning Tree (0)

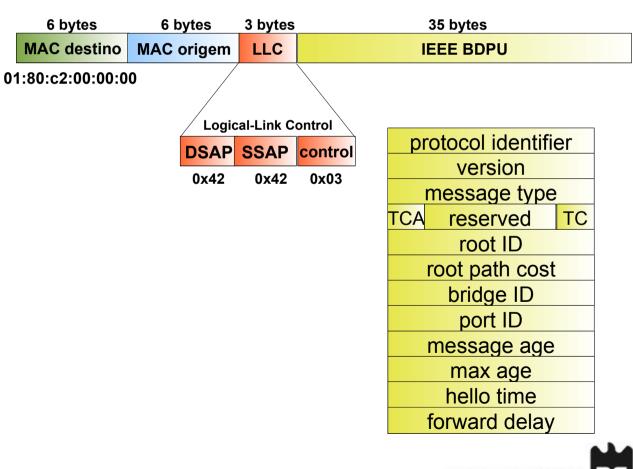
BPDU Type: Configuration (0x00) Root ID: 32768 / 00:05:1a:4e:fd:58

Root Path Cost: 200004

Bridge ID: 32768 / 00:16:e0:9a:c3:80

Port ID: 0x8012 Message Age: 1 Max Age: 20 Hello Time: 2

Forward Delay: 15



Configuration BPDUs

 A configuração da Spanning Tree é feita pelas Conf - BPDUs (mensagens de configuração)

IEEE 802.3 Ethernet

Destination: 01:80:c2:00:00:00 (01:80:c2:00:00:00)

Source: 00:16:e0:9a:c3:92 (00:16:e0:9a:c3:92)

Length: 39

Logical-Link Control

DSAP: Spanning Tree BPDU (0x42) SSAP: Spanning Tree BPDU (0x42) Control field: U, func=UI (0x03)

Spanning Tree Protocol

Protocol Identifier: Spanning Tree Protocol (0x0000)

Protocol Version Identifier: Spanning Tree (0)

BPDU Type: Configuration (0x00)

Root ID: 32768 / 00:05:1a:4e:fd:58

Root Path Cost: 200004

Bridge ID: 32768 / 00:16:e0:9a:c3:80

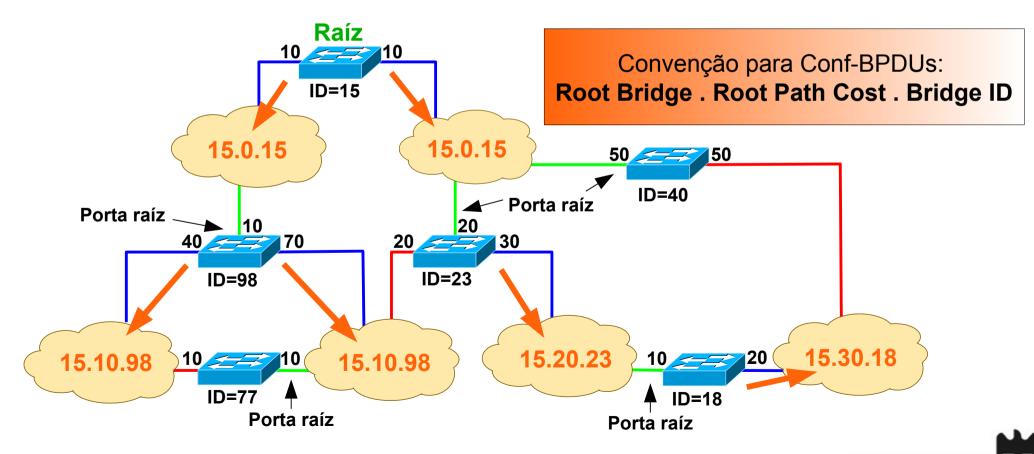
Port ID: 0x8012
Message Age: 1
Max Age: 20
Hello Time: 2
Forward Delay: 15

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: identificação da porta que envia a mensagem de configuração
 - Prioridade (1 byte) + Número da porta

Manutenção da spanning tree

- Periodicamente as bridges enviam pelas portas designadas Conf-BPDUs
 - Periodicidade das mensagens Conf-BPDUs = hello time
 - Hello time recomendado: 2 segundos



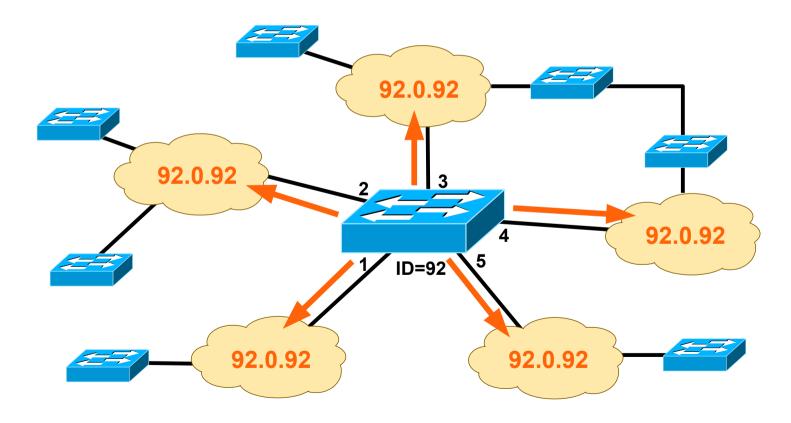
Ordenação das mensagens de configuração

- Uma mensagem de configuração C1 diz-se melhor que outra C2 se:
 - o Root ID de C1 for inferior ao de C2
 - sendo os Root ID idênticos, o Root Path Cost de C1 for inferior ao de C2
 - sendo idênticos o Root ID e o Root Path Cost, o Bridge ID de C1 for inferior ao de C2
 - sendo idênticos o Root ID, o Root Path Cost e o Bridge ID, o Port ID de C1 for inferior ao de C2

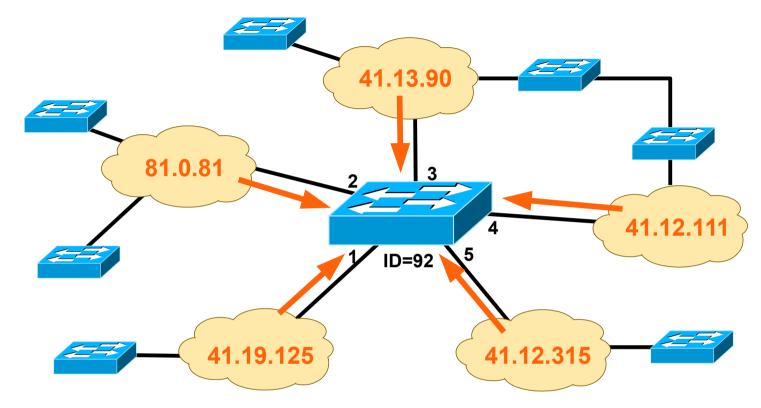
Root ID	Root Path Cost	Bridge ID	Port ID	
18	27	32	2	
18	27	32	4	
18	27	43	1	
18	35	23	3	
23	31	45	2	

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



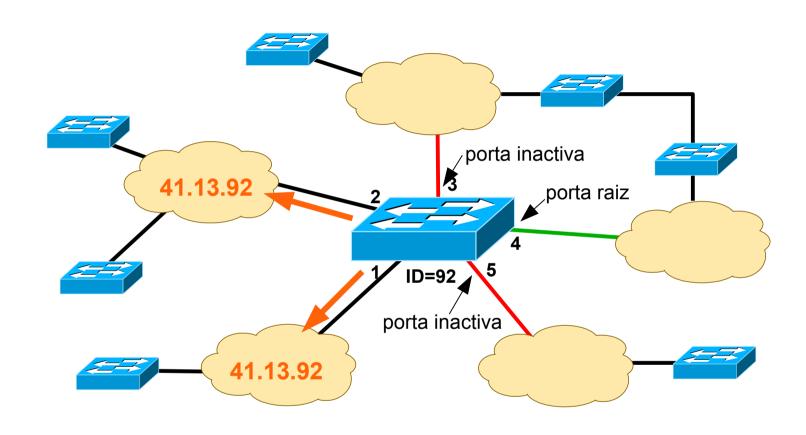
Construção da spanning tree (II)



Melhores mensagens recebidas na Bridge 92 até um dado instante

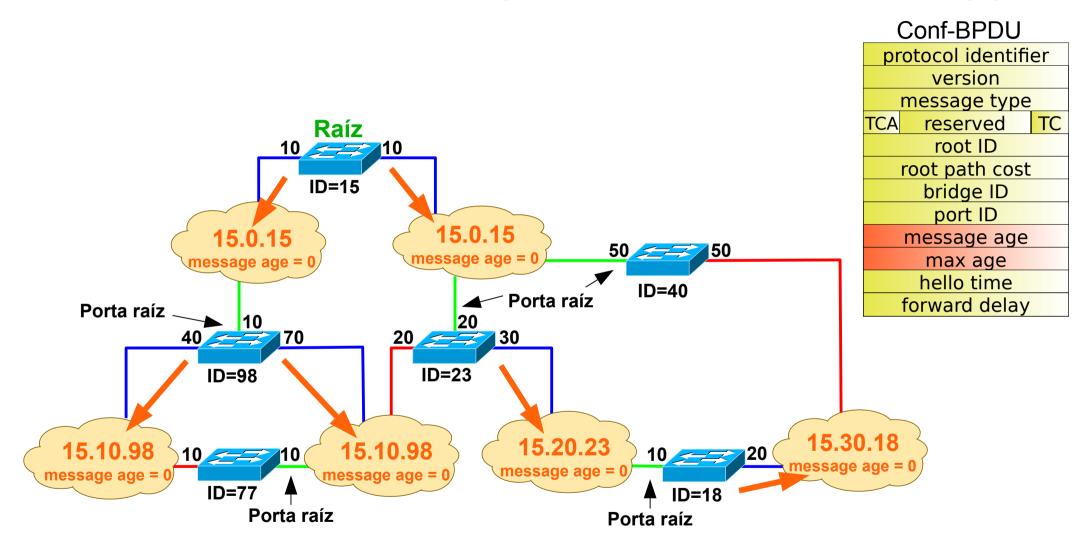
Estimativas da Bridge 92 (assumindo que os custos das portas são unitários)

Construção da spanning tree (III)

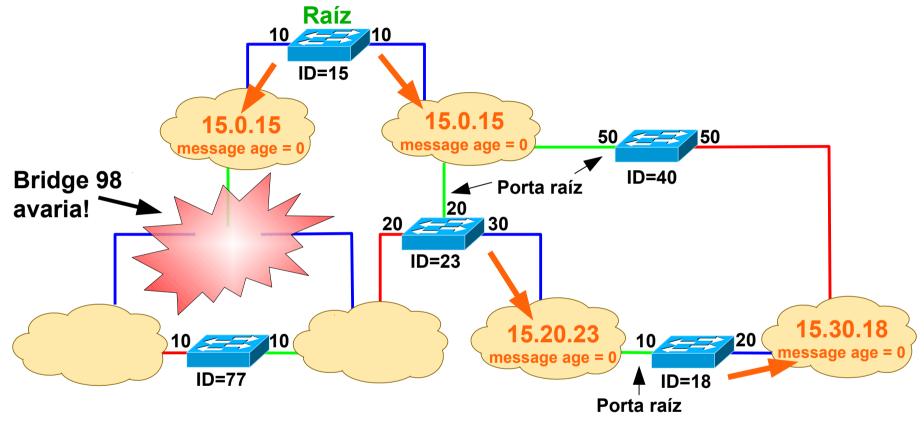


Mensagens enviadas pela Bridge 92 - 41.13.92

Avarias nas bridges ou nas LANs (I)



Avarias nas bridges ou nas LANs (II)



15.10.98 age = 0

15.10.98 age = 5

15.10.98 age = 10

15.10.98 age = max age

15.10.98 age = 0

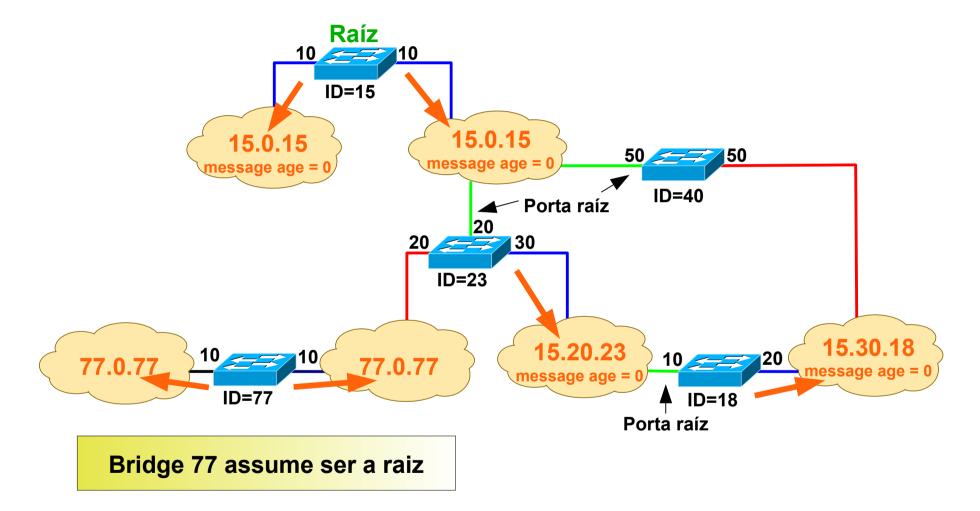
15.10.98 age = 5

15.10.98 age = 10

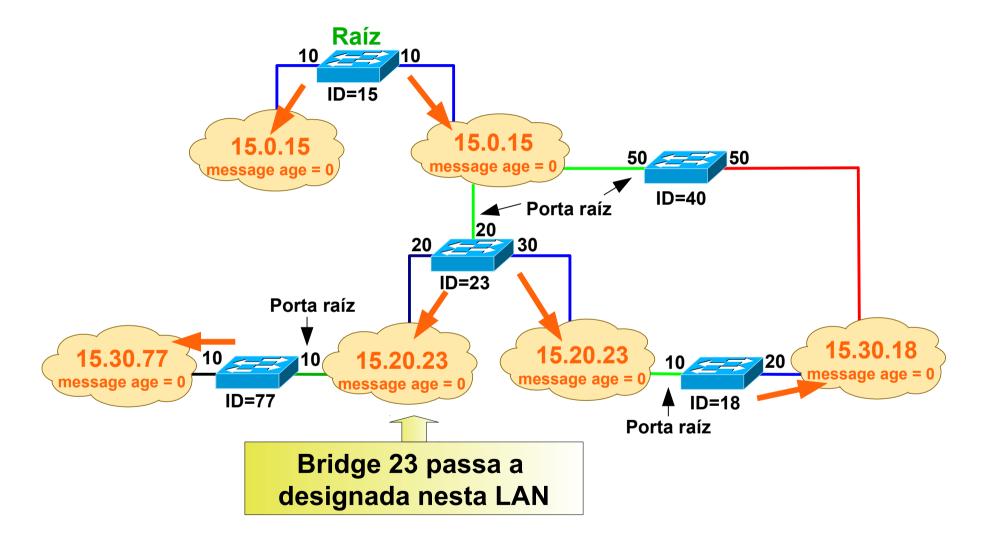
15.10.98 age = max age

max age recomendado = 20 segundos

Avarias nas bridges ou nas LANs (III)



Avarias nas bridges ou nas LANs (IV)



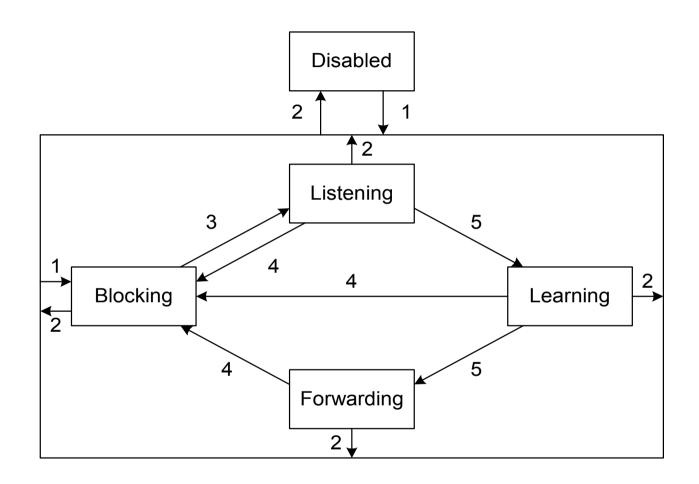
Existência de ciclos temporários

- Após alteração da topologia da rede:
 - Pode existir perda temporária de conectividade se uma porta que estava inativa na topologia antiga ainda não se apercebeu que deverá estar ativo na nova topologia
 - Podem existir ciclos temporários se uma porta que estava ativa na topologia antiga ainda não se apercebeu que deverá estar inativa 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 inativo para o estado ativo; o tempo de espera é função do parâmetro forward delay

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

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

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

Notificação de alterações da topologia

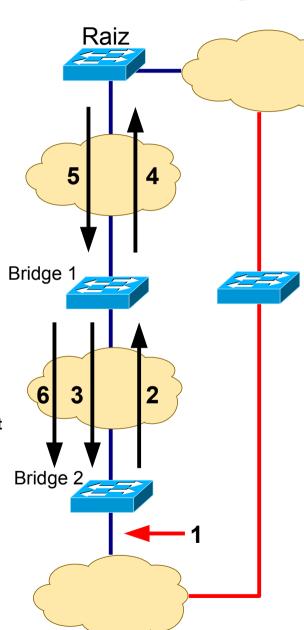
Conf (Configuration) BPDU

	· · · · · · · · · · · · · · · · · · ·		
protocol identifier			
version			
message type = 0			
TCA	reserved	TC	
root ID			
root path cost			
bridge ID			
port ID			
message age			
max age			
hello time			
	forward delay		

TCA - flag Topology Change Acknowledgment TC - flag Topology Change

TCN (Topology Change Notification) **BPDU**

protocol identifier version message type = 1



- 1. Porta passa ao estado blocking
- 2. Envia TCN-BPDU (periodicidade = hello time)
- 3. Envia Conf-BPDU com TCA = 1 até deixar de receber TCN-BPDU
- 4. Envia TCN-BPDU (periodicidade = hello time)
- 5. Envia Conf-BPDU com TCA = 1 até deixar de receber TCN-BPDU e com TC=1 durante um tempo forward delay + max age

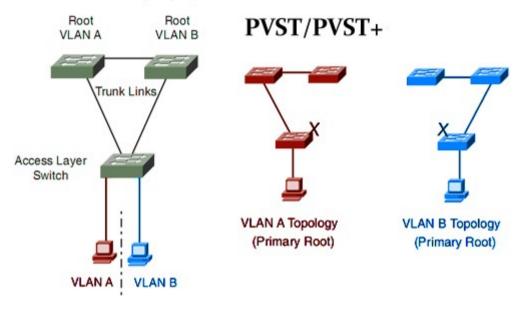
Bridge raiz passa a usar o tempo de vida curto nas suas tabelas de encaminhamento durante este período

6. Repete Conf-BPDU com TC=1

Bridge 1 passa a usar o tempo de vida curto nas suas tabelas até voltar a ouvir TC=0

Bridge 2 passa a usar o tempo de vida curto nas suas tabelas até voltar a ouvir TC=0

- As versões proprietárias do STP da Cisco designam-se por:
 - Per-VLAN Spanning Tree (PVST).
 - Per-VLAN Spanning Tree Plus (PVST+).
 - Criam uma spanning tree distinta para cada VLAN.
 - Diferentes raízes, custos, portas bloqueadas, etc...
- As pacotes BPDU de diferentes VLAN são distinguidos (nos trunks):
 - PVST: usando o ISL (protocolo proprietário da Cisco equivalente ao 802.1Q).
 - PVST+: usando o 802.1Q.



```
Ethernet II, Src: c2:00:05:7f:f1:01 (c2:00:05:7f:f1:01), Dst: PVST+ (01:00:0c:cc:cc)
802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 1
  000. .... = Priority: 0
  ...0 .... = CFI: 0
  Length: 50
Logical-Link Control
Spanning Tree Protocol
  Protocol Identifier: Spanning Tree Protocol (0x0000)
  Protocol Version Identifier: Spanning Tree
  BPDU Type: Configuration (0x00)
D BPDU flags: 0x00
                                                    ldentificador da VLAN
▶ Root Identifier: 32768 / 0 / c2:00:05:7f:00:00
  Root Path Cost: 0
Bridge Identifier: 32768 / 0 / c2:00:05:7f:00:00
  Port identifier: 0x802a
  Message Age: 0
  Max Age: 20
  Hello Time: 2
```

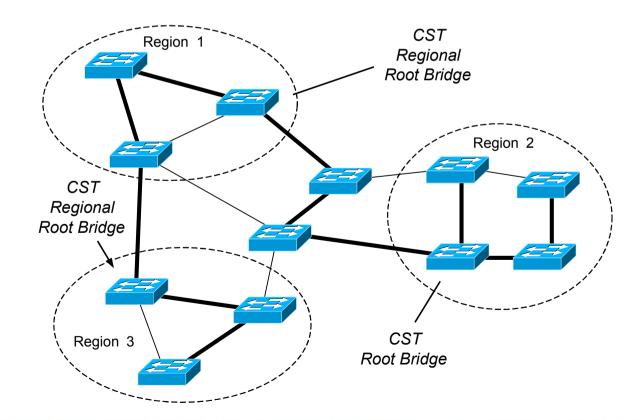
- IEEE 802.1p
 - É uma extensão do IEEE 802.1Q
 - Providencia qualidade de serviço com base em prioridades
 - Define o campo User Priority (3 bits) que permite 8 níveis de prioridade.
 - A norma recomenda
 - Prioridade 7 : tráfego crítico da gestão de rede
 - Prioridades 5–6 : tráfego sensível ao atraso (voz e vídeo)
 - Prioridades 1–4 : tráfego sensível à variação do atraso (streaming)
 - Prioridade 0 : restante tráfego (transferência de ficheiros)

- IEEE 802.1w Rapid Spanning Tree Protocol
 - É uma extensão do IEEE 802.1D
 - Acelera os tempos de convergência da Spanning Tree em caso de alteração da topologia de rede
 - Altera os estados e as funções de cada porta
 - Introduz um mecanismo de negociação entre bridges
 - Usa os bits entre as flags TCA e TC

Conf (Configuration) BPDU

protocol identifier			
version			
message type = 0			
TCA	reserved	TC	
root ID			
root path cost			
bridge ID			
port ID			
message age			
max age			
hello time			
forward delay			

- IEEE 802.1s Multiple Spanning Tree Protocol
 - Permite criar múltiplas Spanning Trees e atribuir cada VLAN a uma das Spanning Trees criadas
 - Usa os protocolos IEEE 802.1W (VLANs) e IEEE 802.1w (RSTP)
 - Permite criar múltiplas regiões
 - Fora das regiões funciona como o protocolo IEEE 802.1D



Referências

- IEEE 802.1w Rapid Spanning Tree Protocol
 - "Understanding Rapid Spanning Tree Protocol (802.1w)"
 - http://www.cisco.com/warp/public/473/146.html
- IEEE 802.1s Multiple Spanning Tree Protocol
 - "Understanding Multiple Spanning Tree Protocol (802.1s)"
 - http://www.cisco.com/warp/public/473/147.html