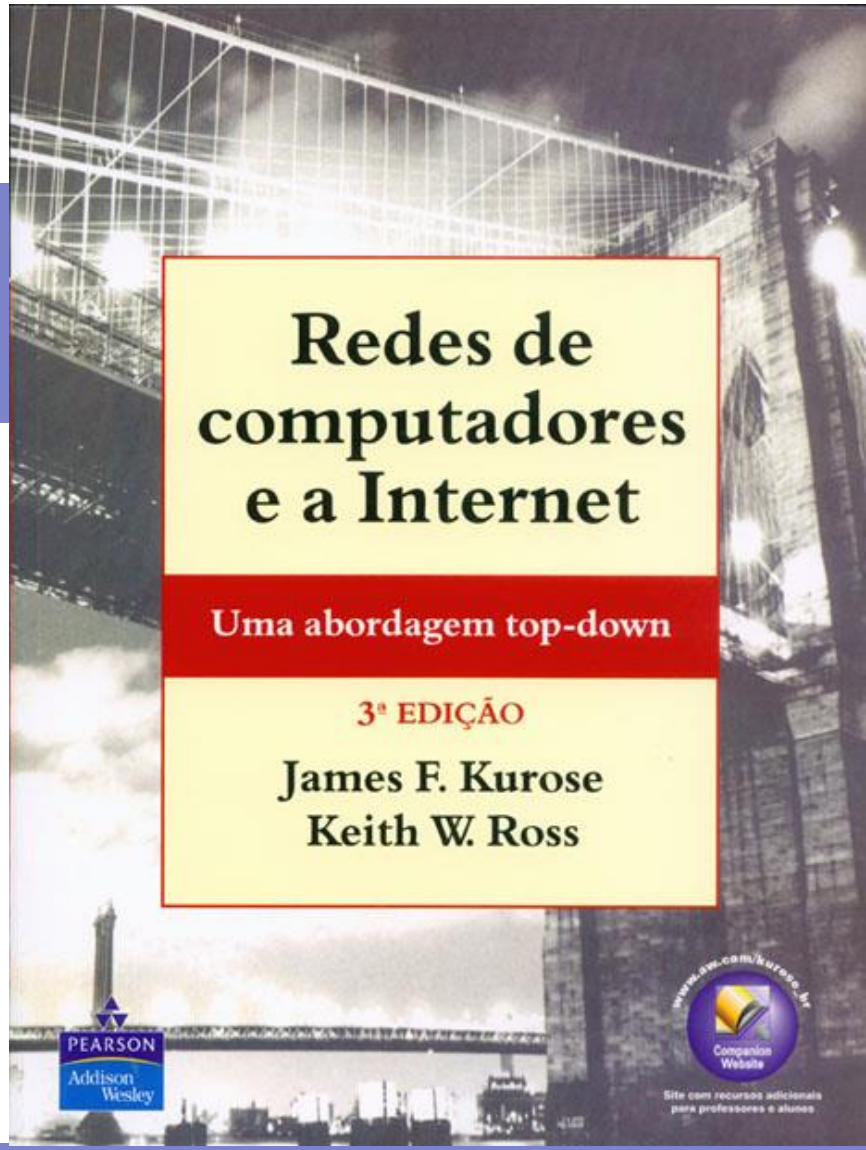
Redes de computadores e a Internet

Capítulo 4

A camada de rede



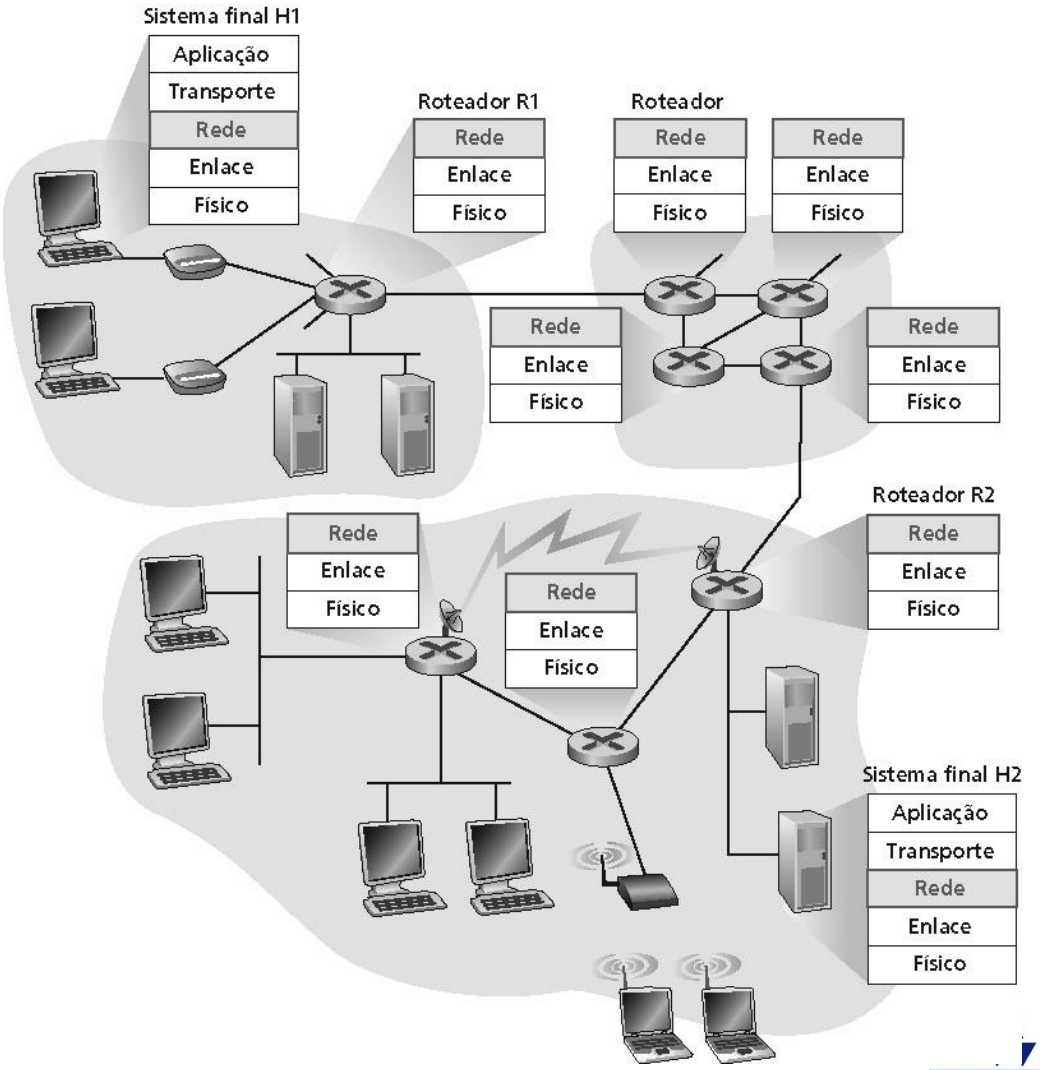
(Alterado por Atslands Rocha)

4 A camada de rede

- 4. 1 Introdução
- 4.2 O que há dentro de um roteador
- 4.3 Redes de datagrama
- 4.4 IP: Protocolo da Internet
  - Formato do datagrama
  - Endereçamento IPv4
  - ICMP
  - IPv6
- 4.5 Algoritmos de roteamento
  - Link state
  - Distance vector
  - Roteamento hierárquico
- 4.6 Roteamento na Internet
  - RIP
  - OSPF
  - BGP

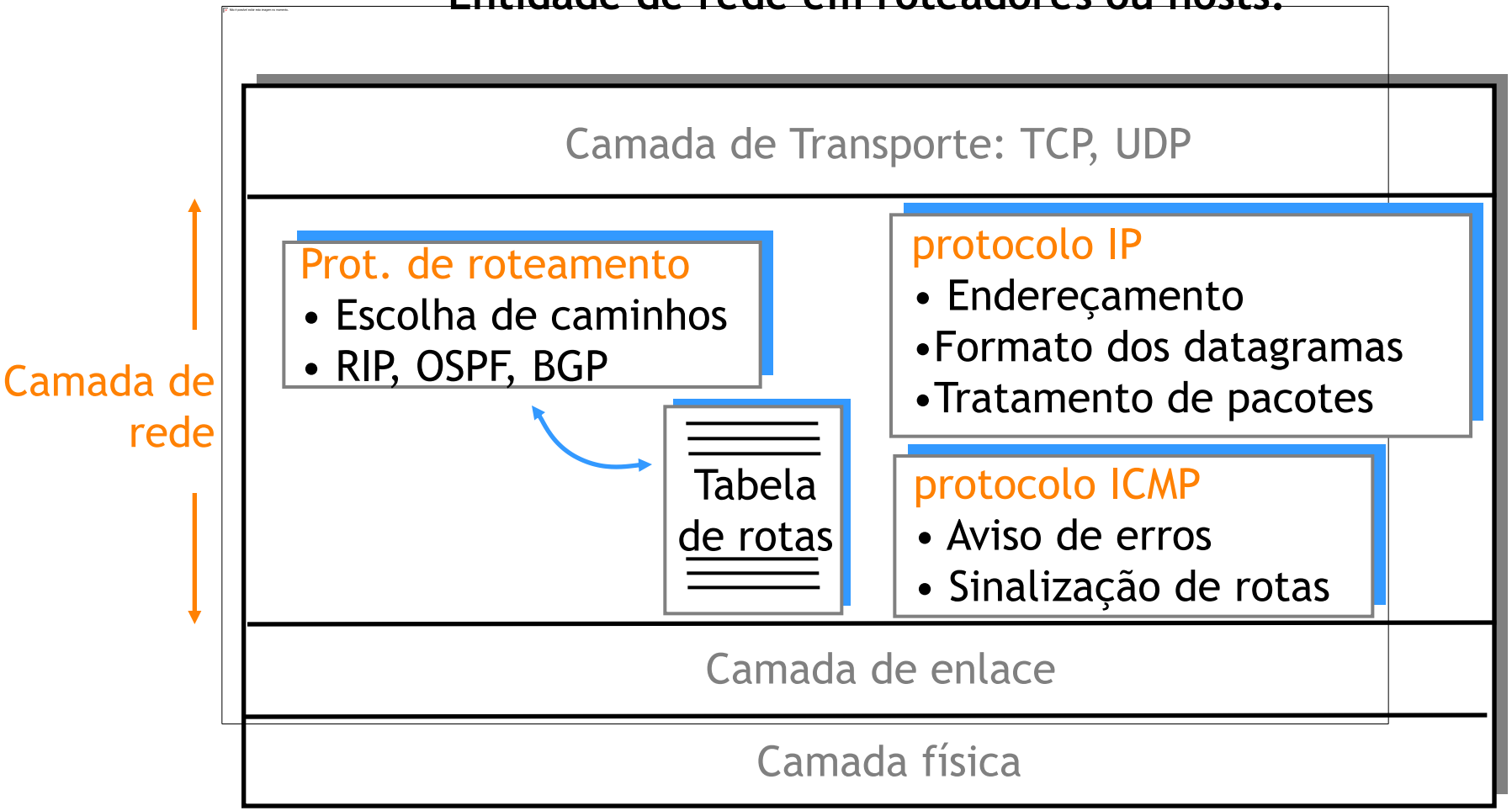
4 A camada de rede

- Protocolos da camada de rede em *cada* hospedeiro e roteadores.
- Roteador examina campos de cabeçalho em todos os datagramas IP que passam por ele.



4 A camada de rede

Entidade de rede em roteadores ou hosts:

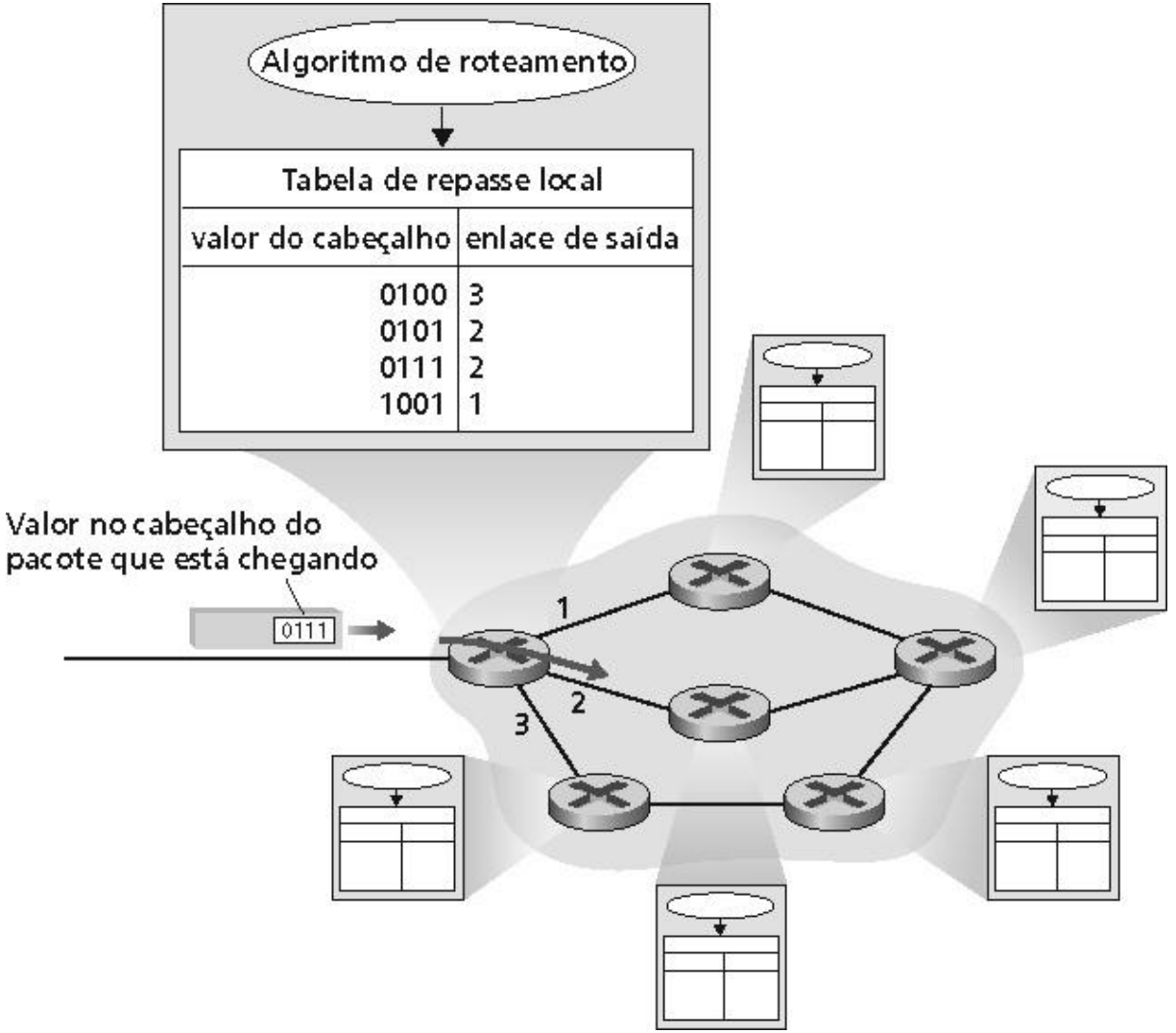


4 Serviços da camada de rede

- **Comutação:** mover pacotes da entrada do roteador para a saída apropriada do roteador.

- **Roteamento:** determinar a rota a ser seguida pelos pacotes desde a origem até o destino.
  - **Algoritmos de roteamento**

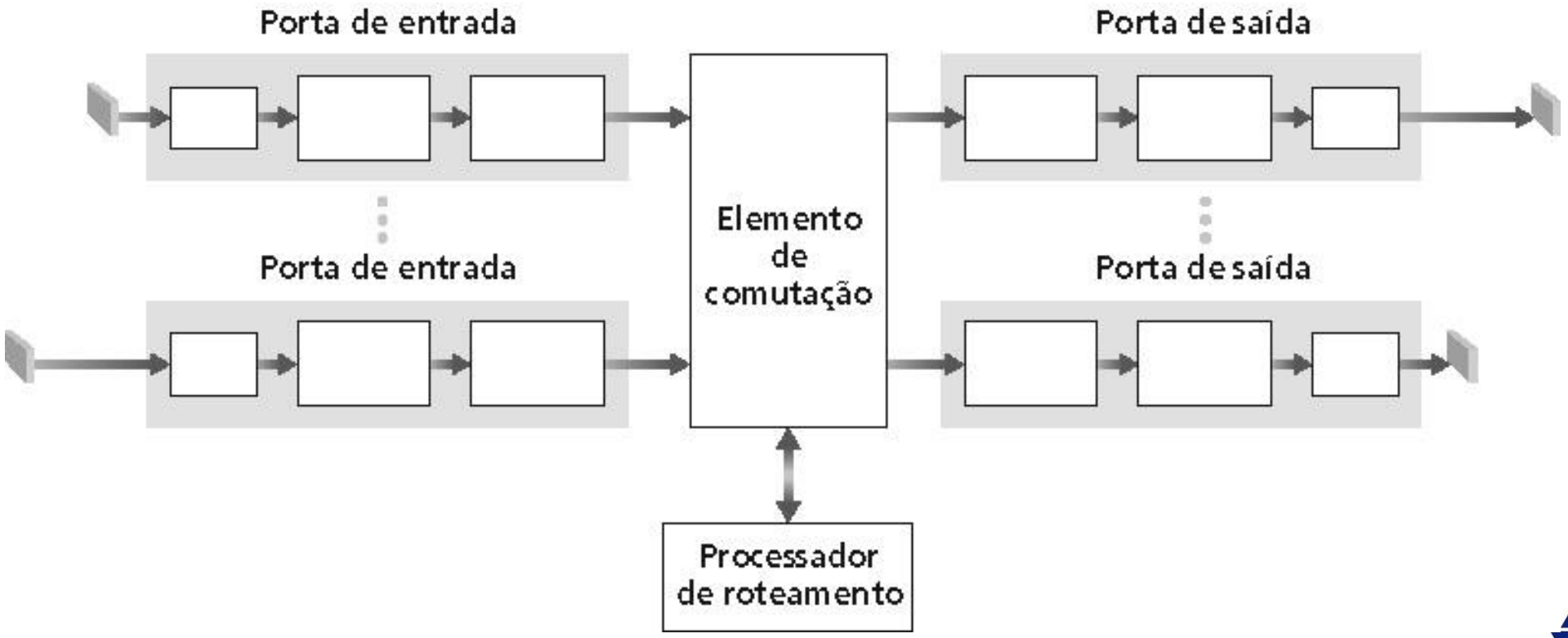
4 Interação entre roteamento e comutação



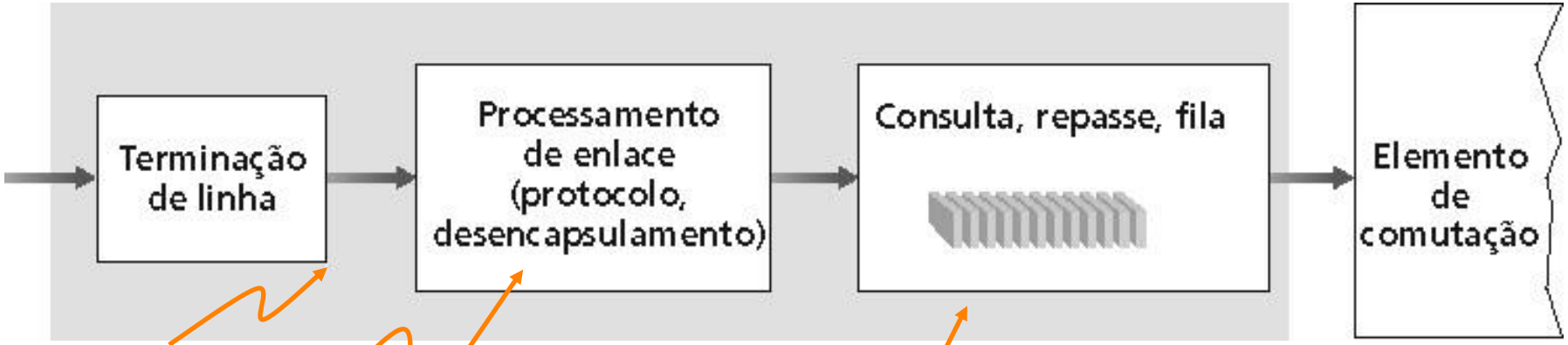
4 Visão geral da arquitetura do roteador

Duas funções-chave do roteador:

- Executar algoritmos/protocolos (RIP, OSPF, BGP).
- **Comutar** os datagramas do link de entrada para o link de saída.



4 Funções da porta de entrada



Camada física:
recepção de bits

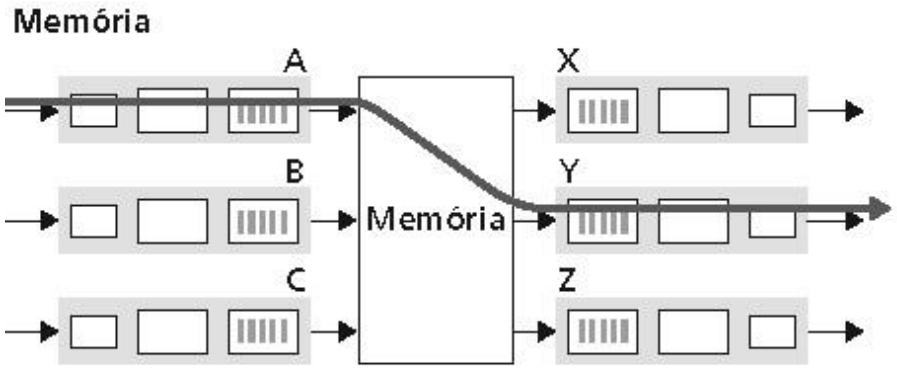
Camada de enlace:
ex.: Ethernet

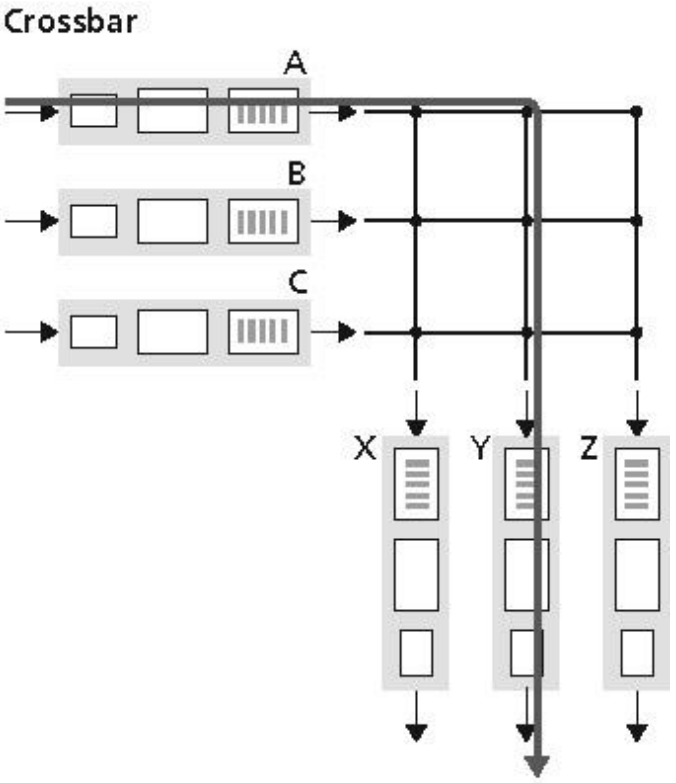
Comutação descentralizada:

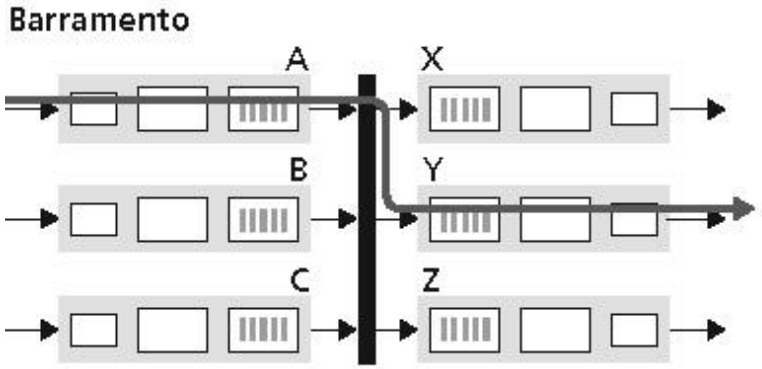
- Dado o destino do datagrama, procura a porta de saída usando a tabela de comutação na memória da porta de entrada.

- Fila: Quando os datagramas chegam mais rápido do que a taxa de comutação.

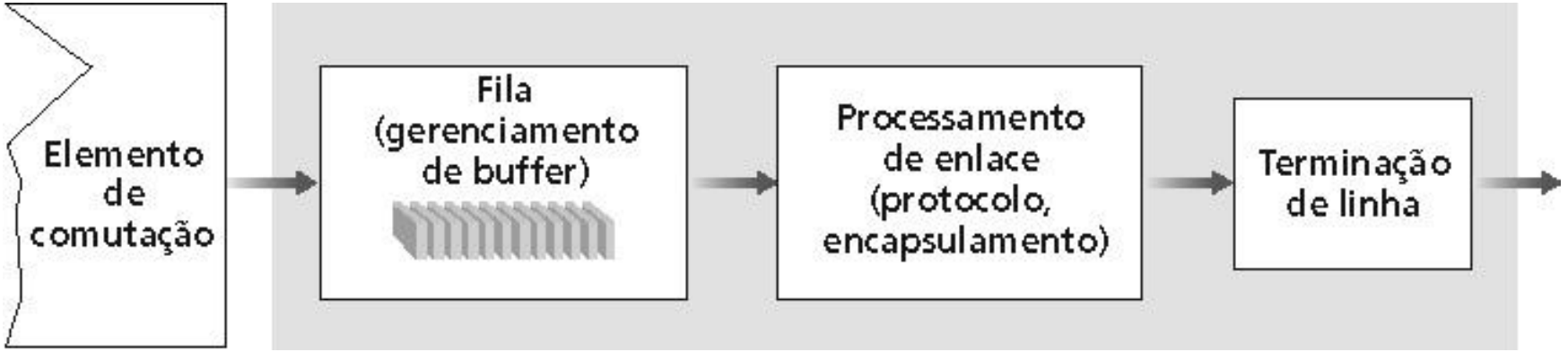
4 Três tipos de estrutura de comutação







4 Portas de saída



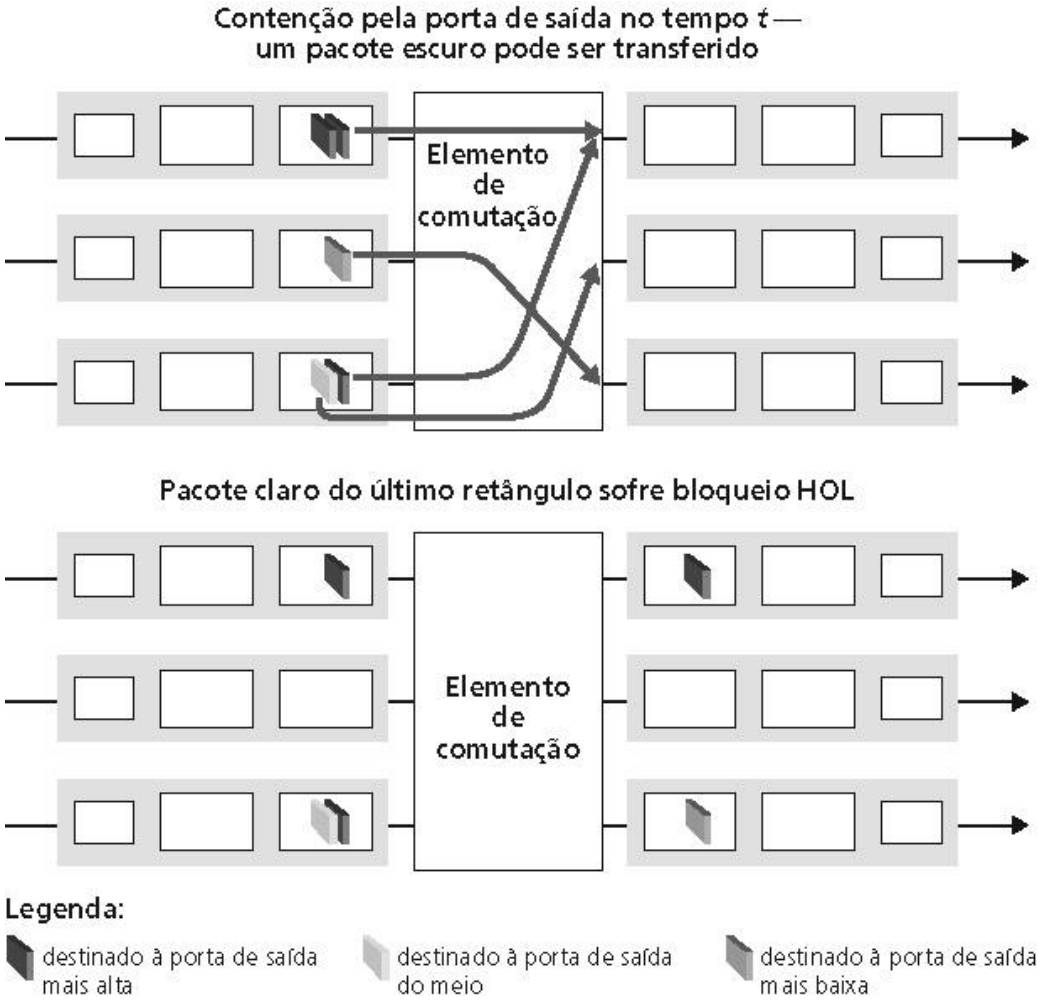
- **Buffering** necessário quando datagramas chegam mais rápido do que a taxa de transmissão.

- **Disciplina de agendamento** escolhe entre os datagramas na fila para transmissão.

4 Enfileiramento na porta de saída

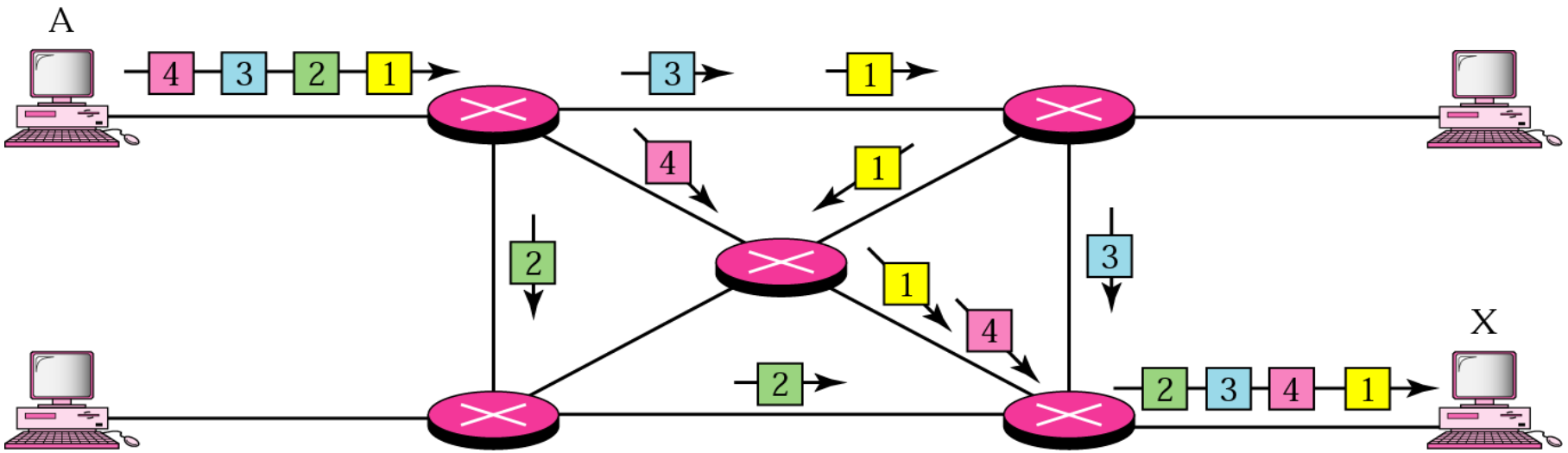
- **Enfileiramento (atraso) e perda devido ao buffer overflow da porta de saída!**

- **Bloqueio Head-of-the-Line (HOL):** datagrama na frente da fila impede os outros na fila de se moverem para adiante.



4 Redes de datagrama

- Não existe estabelecimento de conexão na camada de rede.
- Roteadores: não existe estado sobre conexões fim-a-fim.
- Pacotes são encaminhados pelo endereço do destino:
  - Pacotes para o mesmo destino podem seguir diferentes rotas.



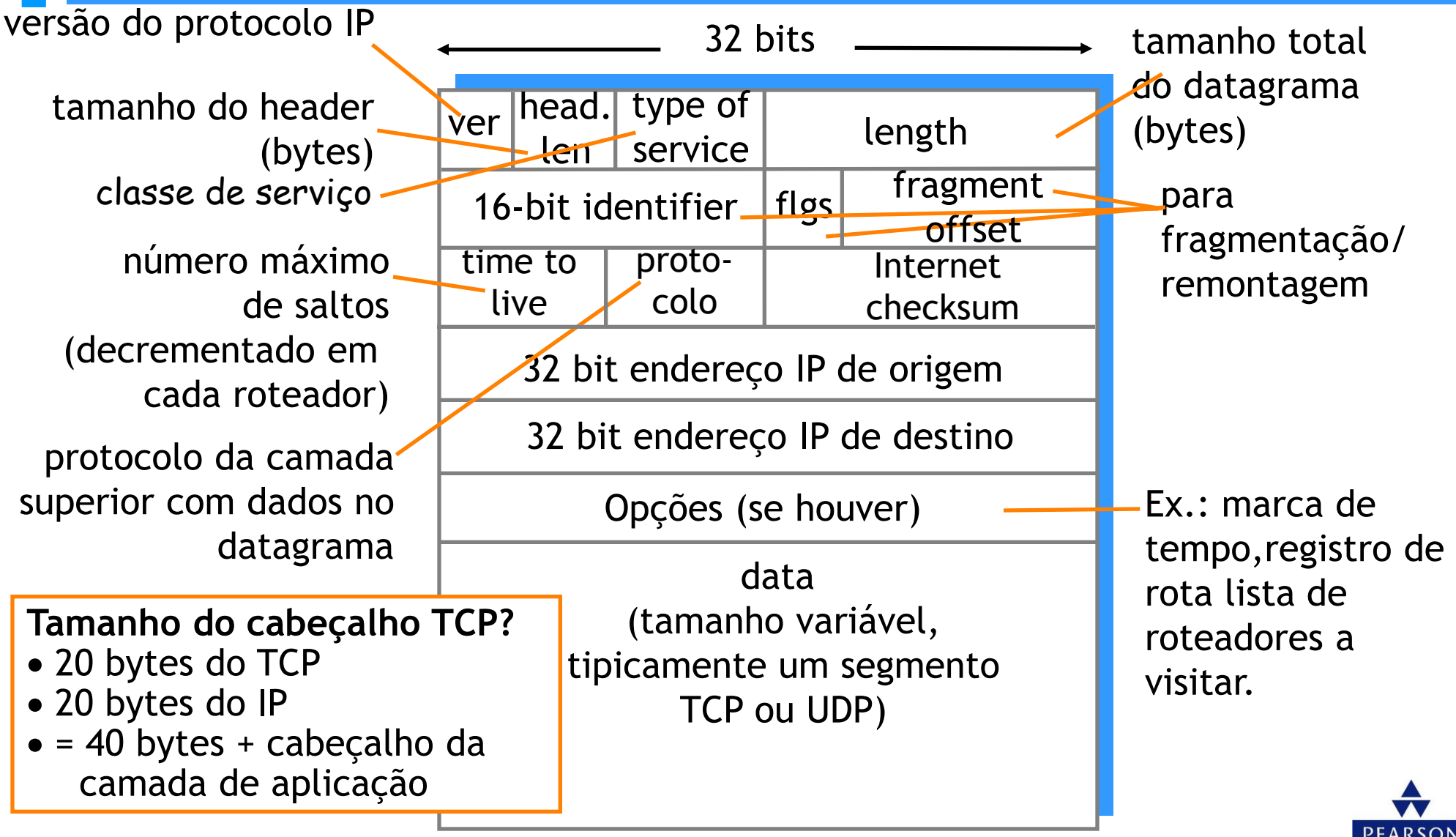
4 Datagrama: Resumo

- Internet
- Dados trocados entre computadores
  - Serviço elástico, requisitos de atraso não críticos.

- Sistemas finais inteligentes
  - A rede é simples; a complexidade fica nas pontas.

- Muitos tipos de enlaces
  - Características diferentes.
  - Difícil obter um serviço uniforme.

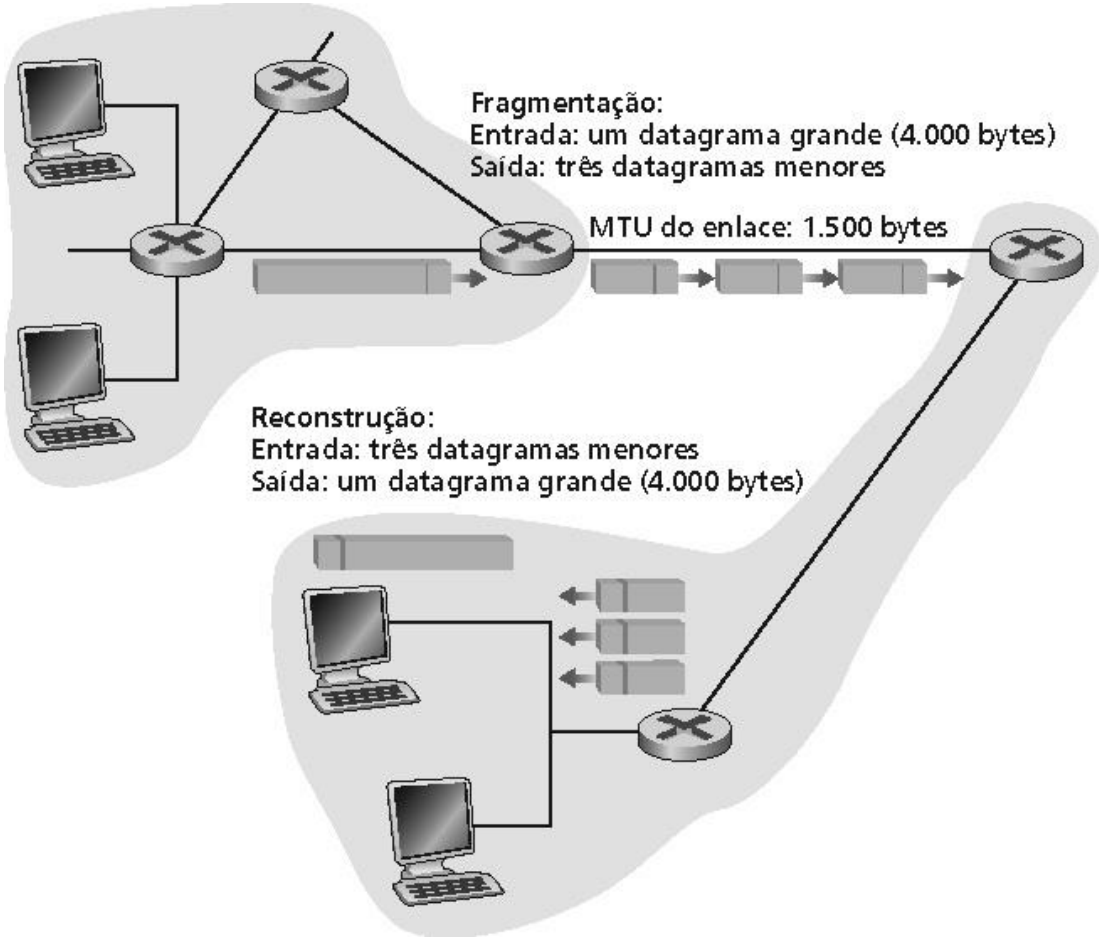
4 Formato do datagrama IPv4



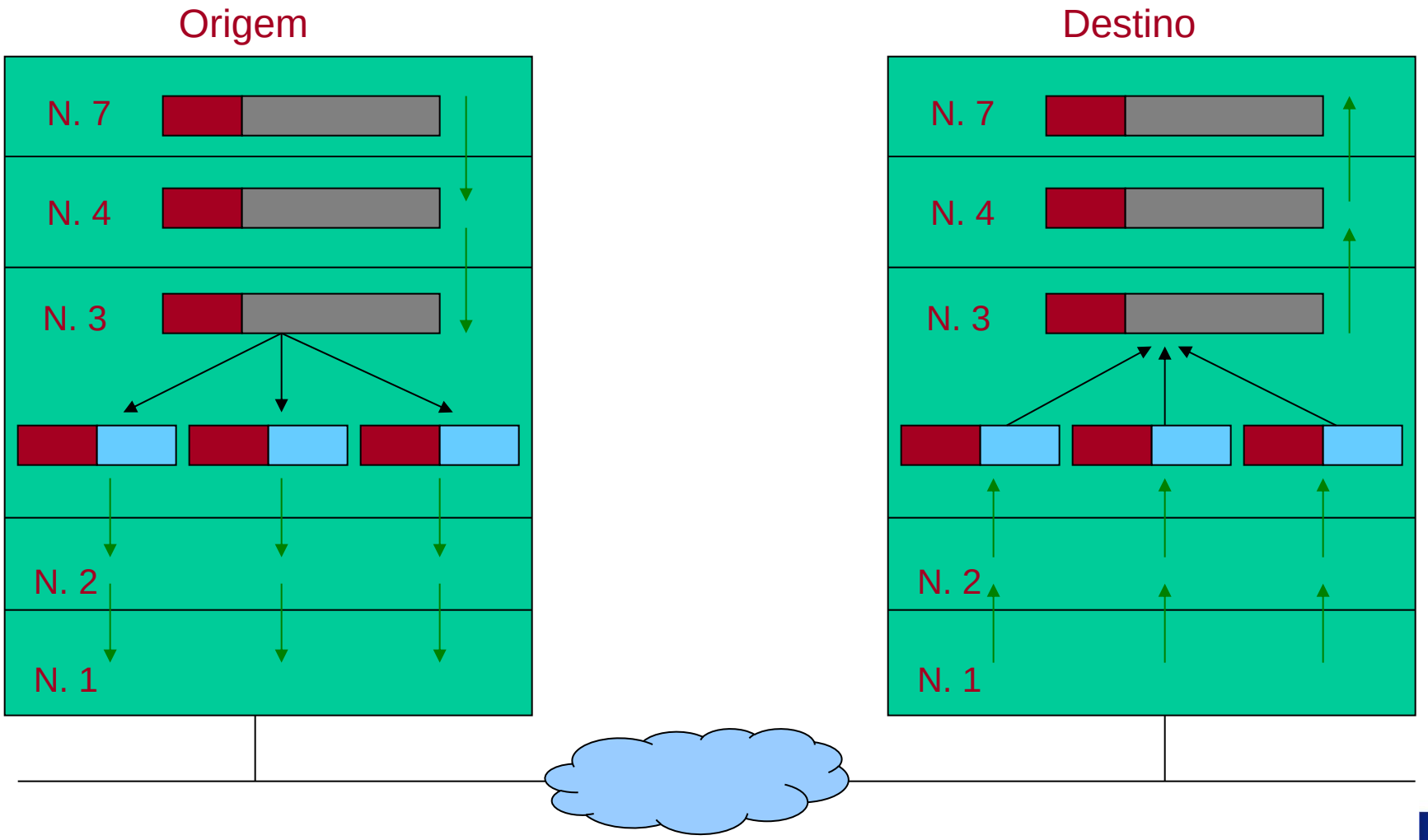
Tamanho do cabeçalho TCP?
- 20 bytes do TCP
- 20 bytes do IP
- = 40 bytes + cabeçalho da camada de aplicação

4 IP fragmentação e remontagem

- Tipos de enlaces diferentes possuem MTU diferentes (Ethernet: 1518 bytes)
- Datagramas IP grandes devem ser fragmentados dentro da rede
- Remontagem ocorre apenas no destino final (não ocorre em roteadores!).
- O cabeçalho IP é usado para identificar e ordenar datagramas relacionados

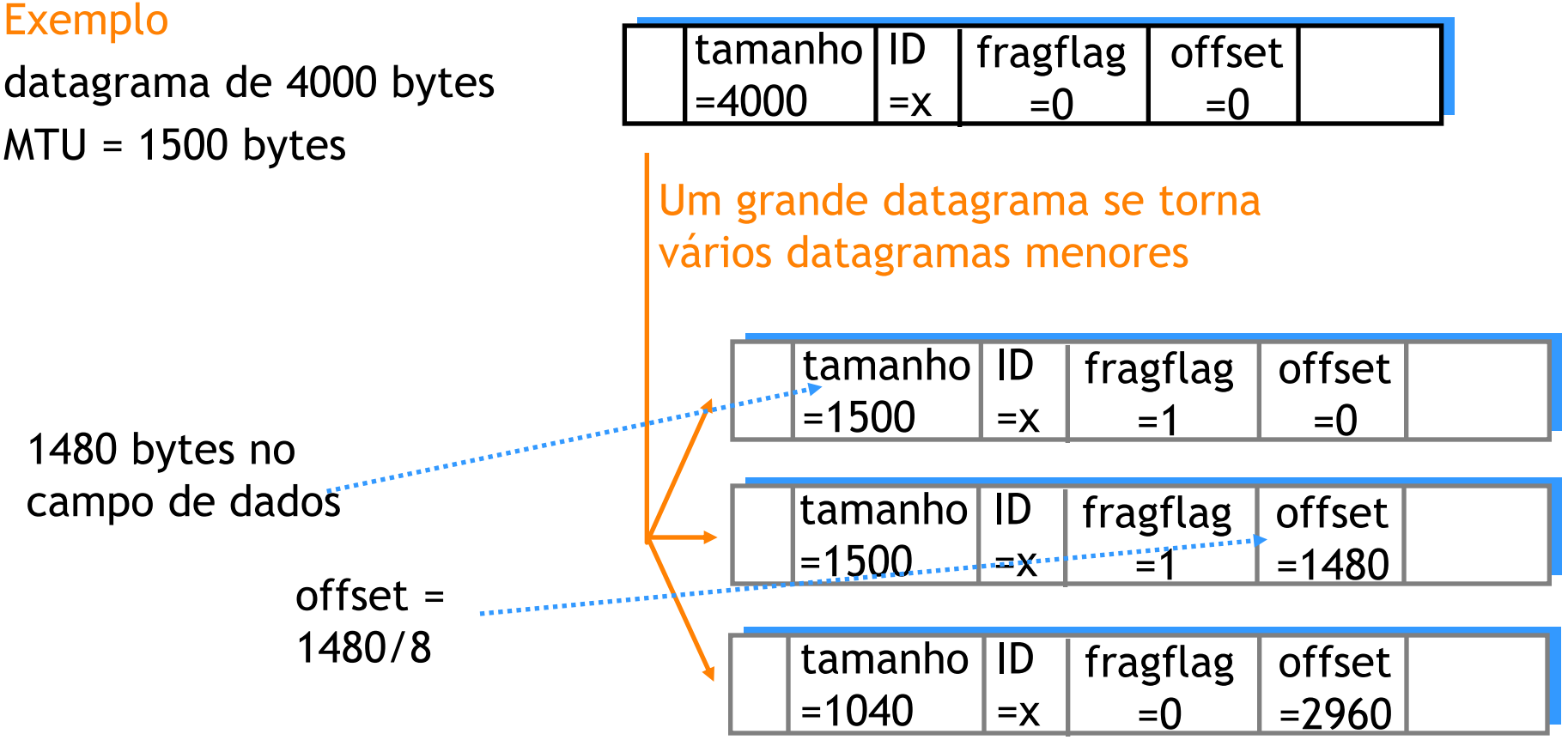


4 IP fragmentação e remontagem



4 IP fragmentação e remontagem

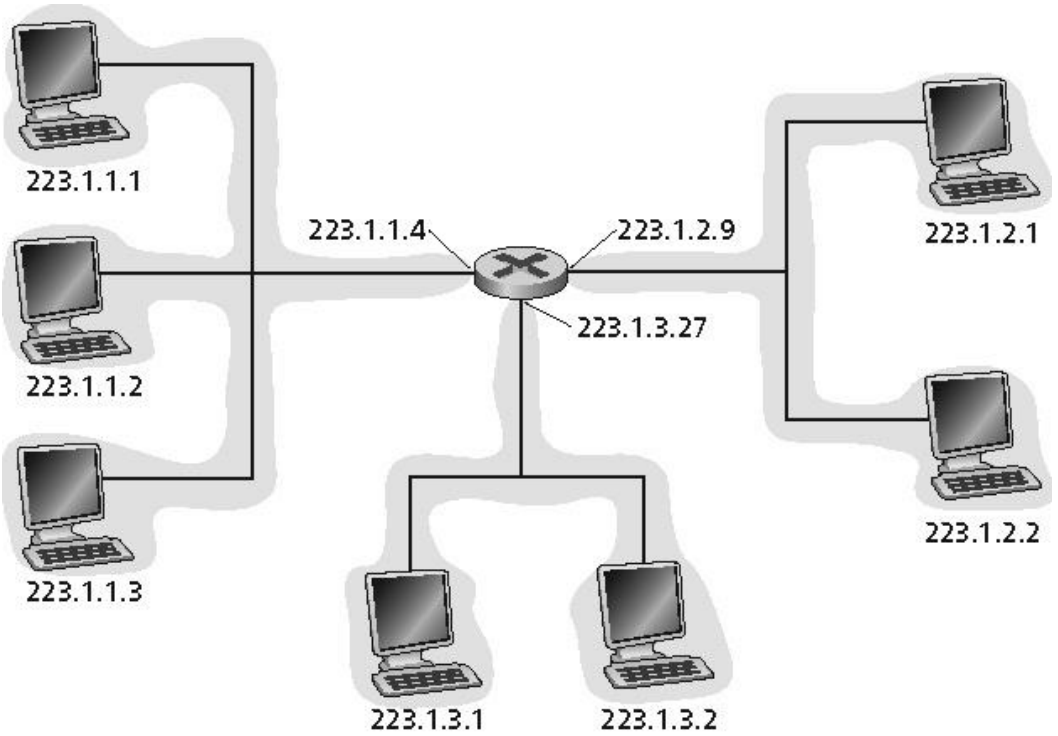
- Exemplo
- datagrama de 4000 bytes
- MTU = 1500 bytes



4 Endereçamento IP: Introdução

- Endereço IP: identificador de 32 bits para **interfaces** de roteadores e hosts.
- Cada campo pode ter valor

  0 - 255 ($2^8 = 256$)
- Endereço IP inadequado = 200.280.23.47
- Interface: conexão entre roteador ou host e enlace físico
  - Roteador tem tipicamente múltiplas interfaces.
  - Hosts podem ter múltiplas interfaces.
  - Endereços IP são associados com interfaces, não com o host ou com o roteador.

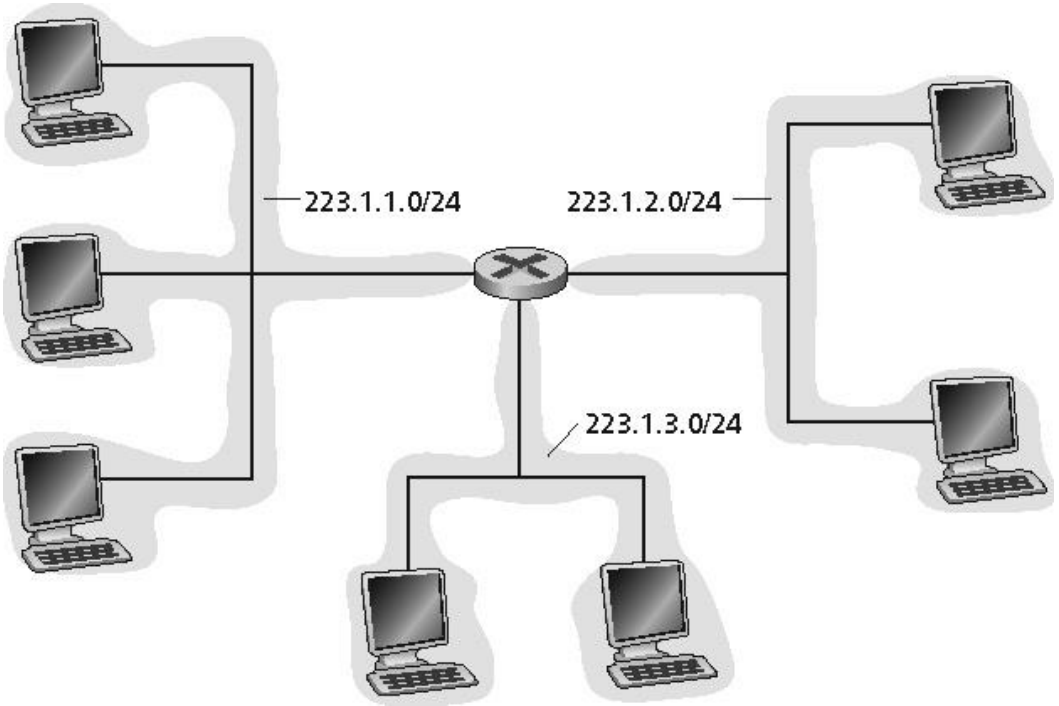


223.1.1.1 = 11011111 00000001 00000001 00000001

223 1 1 1

4 Sub-redes

- Endereço IP:
  - Parte da sub-rede.
  - Parte do host.
- O que é uma sub-rede?
  - Interfaces de dispositivo com a mesma parte de sub-rede do endereço IP.
  - Podem alcançar fisicamente uns aos outros sem intervenção de roteador.

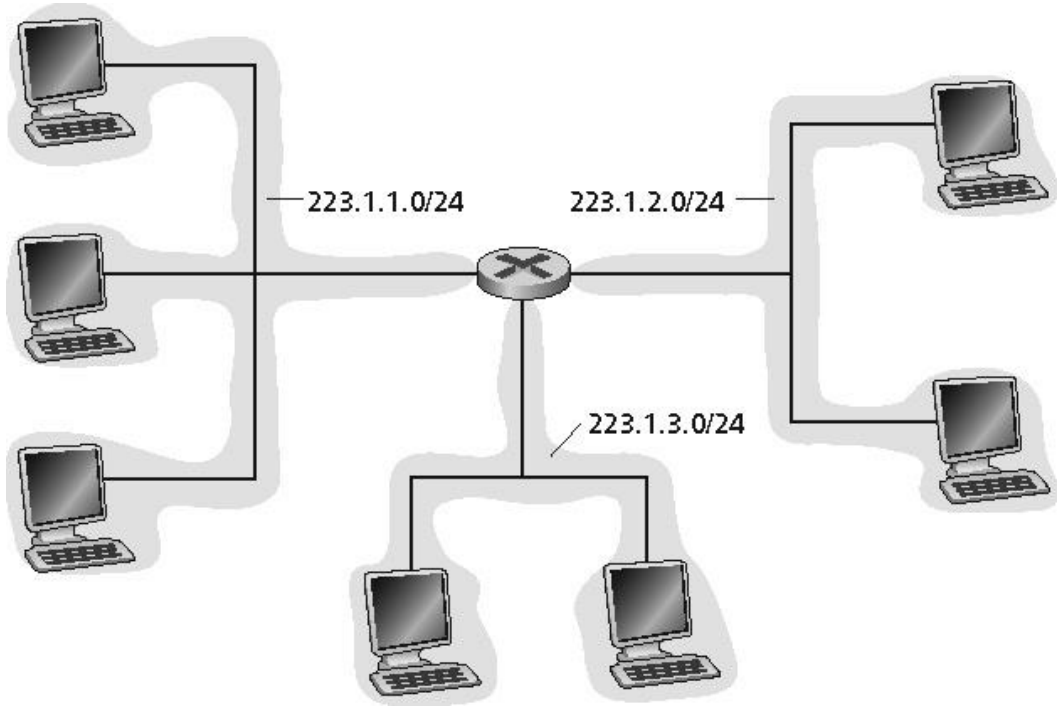


rede consistindo de 3 sub-redes

4 Sub-redes

Receita

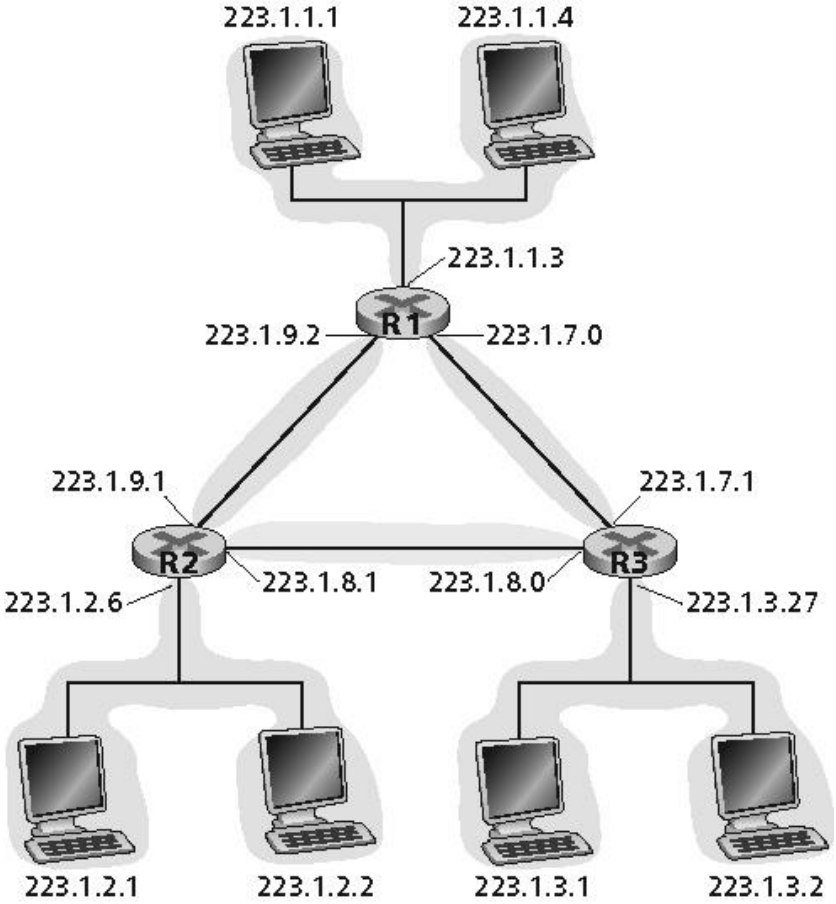
- Para determinar as sub-redes, destaque cada interface de seu host ou roteador, criando ilhas de redes isoladas. Cada rede isolada é considerada uma sub-rede.



máscara de sub-rede: /24 ou 255.255.255.0
(11111111.11111111.11111111.00000000)

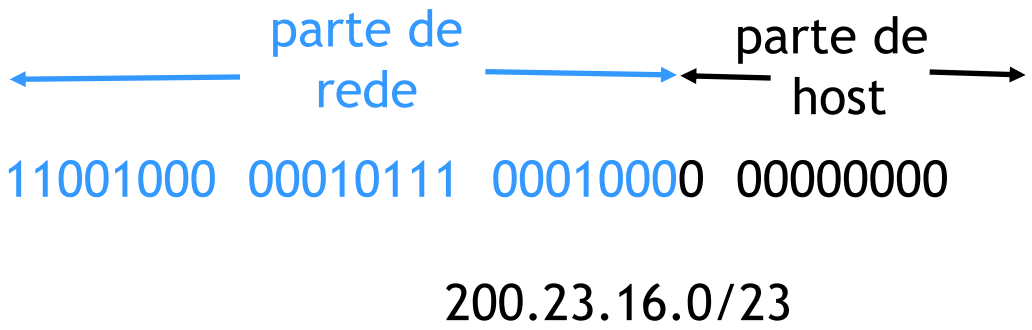
Sub-redes

Quantas?



4 Endereçamento IP: CIDR

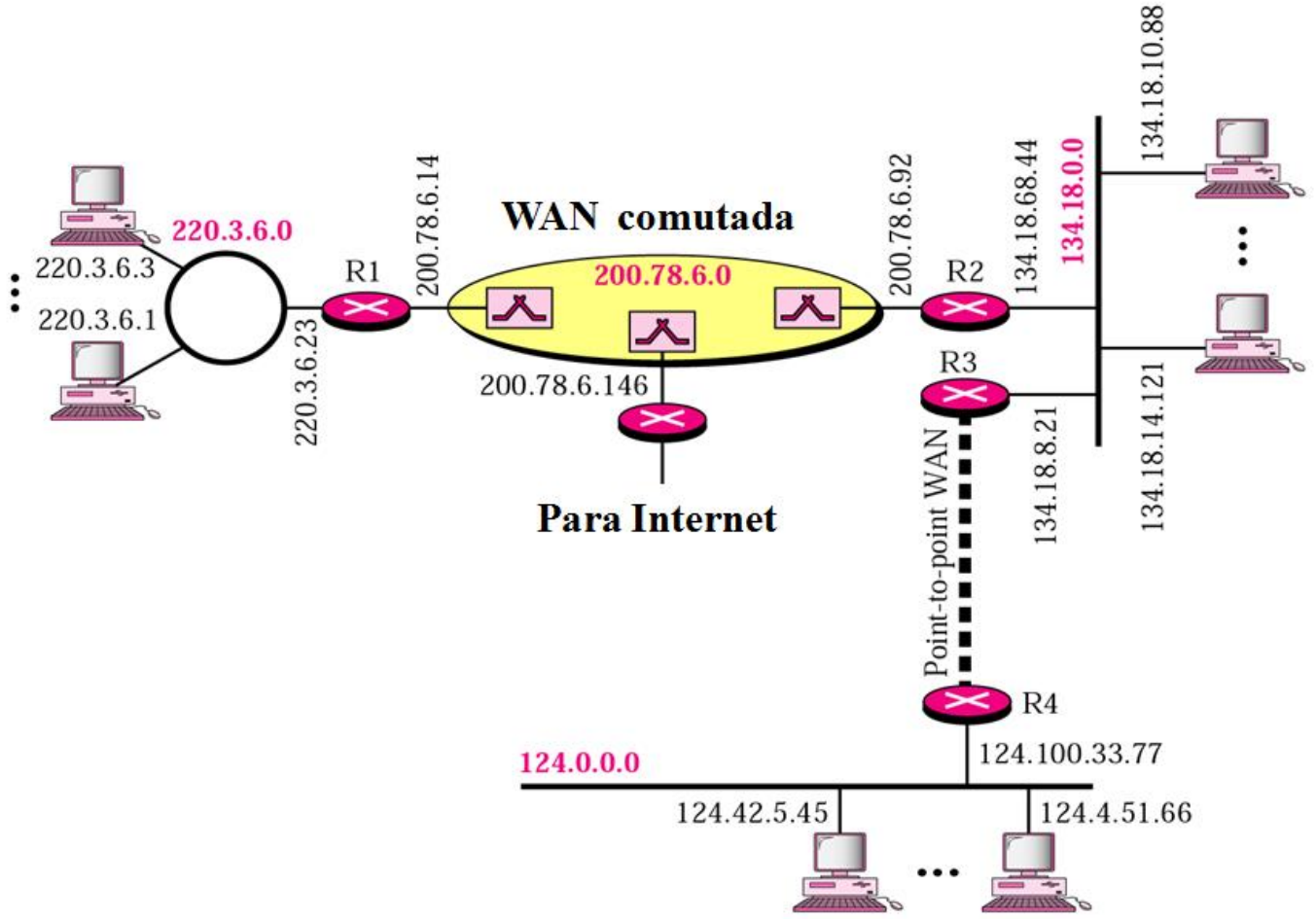
- CIDR: Classless InterDomain Routing
  - A porção de endereço de rede tem tamanho arbitrário
  - Formato do endereço: a.B.C.D/x, em que x é o número de bits na parte de rede do endereço



4 Endereços IP reservados...

- Endereço local
  - Identifica a própria máquina (127.0.0.1);
  - Permite a comunicação entre aplicações situadas na mesma máquina;
  - Permite realização de testes locais.
- Endereço de Rede;
  - Identifica a própria rede.
  - A porção de host possui valor zero (ex: 129.47.0.0).
- Endereço de difusão (*broadcast);*
  - identifica todas as máquinas na rede.
- Em qualquer rede, o primeiro e o último endereços não podem ser usados!
- Três faixas para endereços privados: 10.0.0.0/8, 192.168.0.0/24 e 172.16.0.0/16 – 172.31.0.0/16.

Endereços de rede



4 Como obter um endereço IP

P.: Como o ISP obtém seu bloco de endereço?

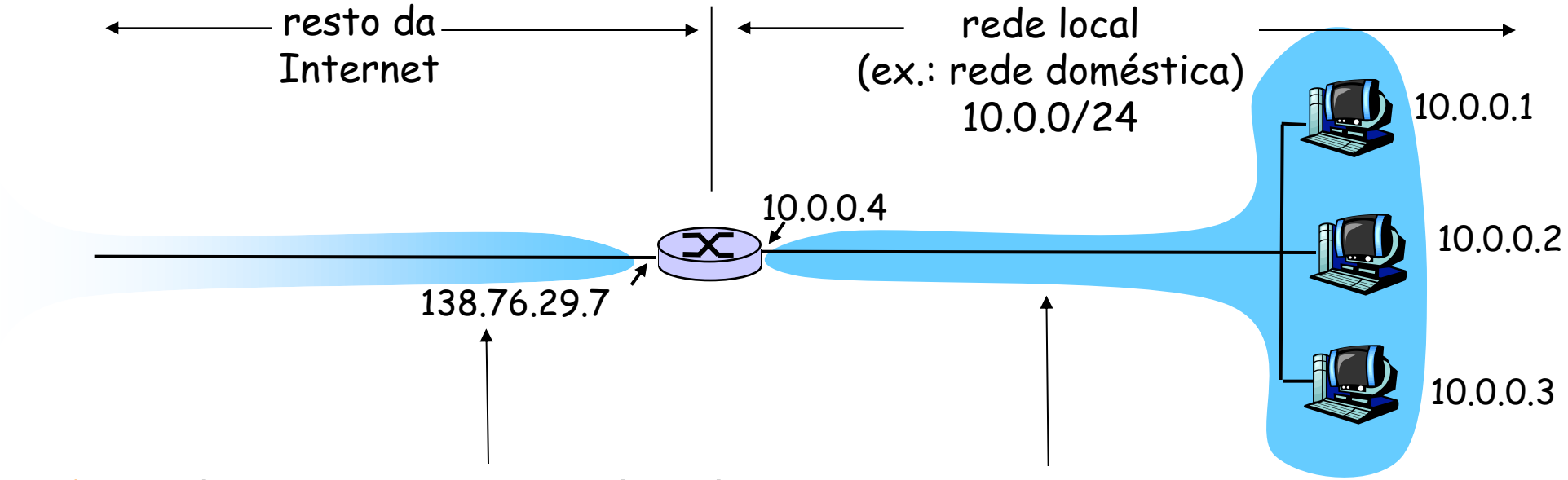
R.: ICANN: internet corporation for assigned names and numbers

- Aloca endereços
- Gerencia DNS
- Atribui nomes de domínios e resolve disputas

P.: Como um host obtém endereço IP ?

- Definido pelo administrador do sistema
- DHCP: dynamic host configuration protocol: obtém dinamicamente endereços IP de um servidor

4 NAT: Network Address Translation



todos os datagramas que **saem** da rede local possuem o **mesmo** e único endereço IP do NAT de origem: 138.76.29.7, números diferentes de portas de origem

datagramas com origem ou destino nesta rede possuem endereço 10.0.0/24 para origem, destino (usualmente)

4 NAT: Network Address Translation

- Motivação: redes locais podem utilizar apenas um endereço IP:
  - Não é preciso alocar uma faixa de endereços do ISP: apenas um endereço IP é usado para todos os dispositivos.
  - Podem-se alterar os endereços dos dispositivos na rede local sem precisar notificar o mundo exterior.
  - Pode-se mudar de ISP sem alterar os endereços dos dispositivos na rede local.
  - Dispositivos da rede local não são explicitamente endereçáveis ou visíveis pelo mundo exterior (um adicional de segurança).

4 NAT: Network Address Translation

Implementação: o roteador NAT deve:

Datagramas que saem: substituir (endereço IP de origem, porta #) de cada datagrama para (endereço IP do NAT, nova porta #)

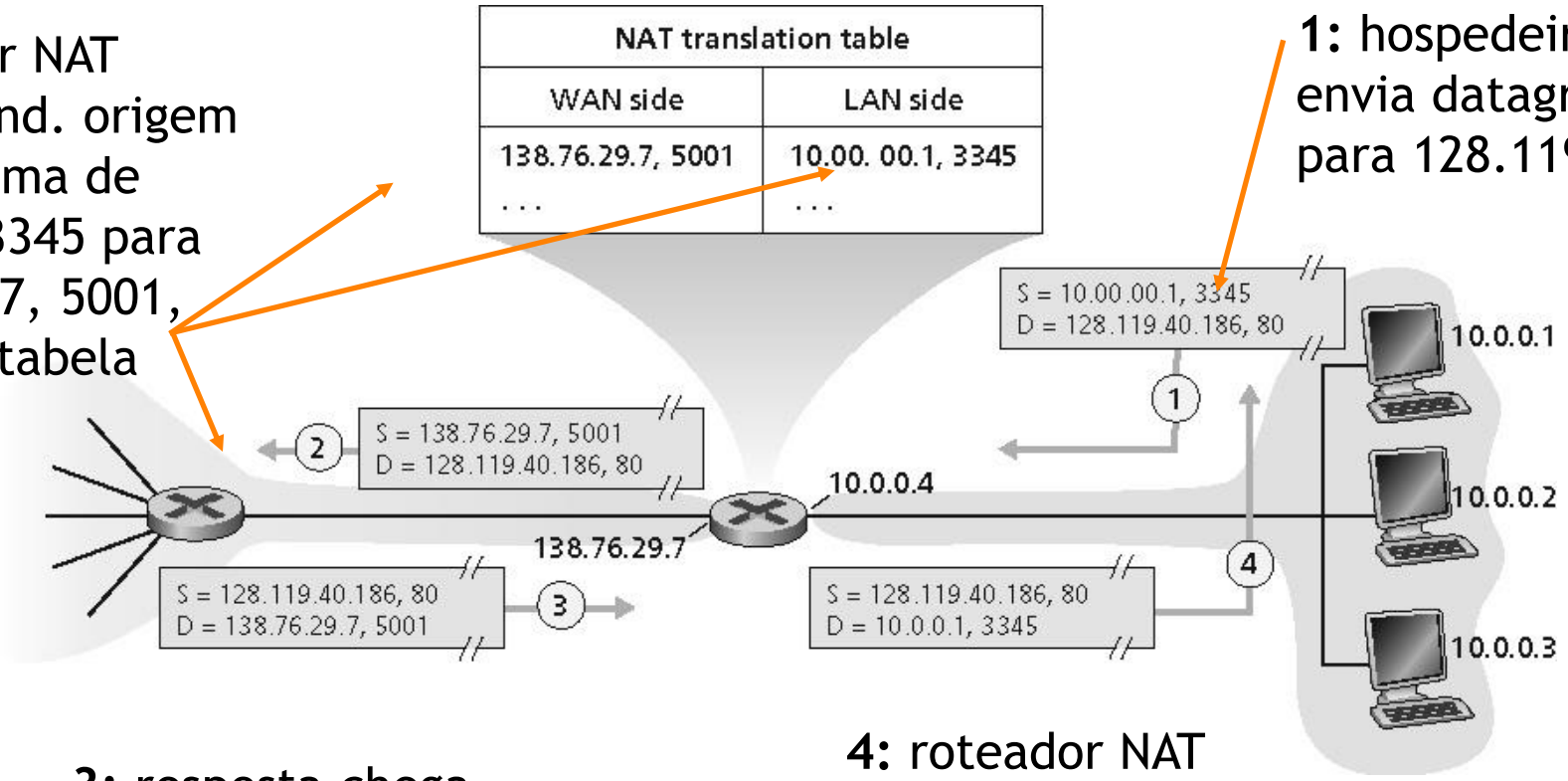
. . . clientes/servidores remotos responderão usando (endereço IP do NAT, nova porta #) como endereço de destino.

- **Lembrar (na tabela de tradução do NAT)** cada (endereço IP de origem, porta #) para o par de tradução (endereço IP do NAT, nova porta #).
  - Campos da tabela do NAT: IP de origem, Porta de Origem, IP do NAT, Porta do NAT, IP de destino, Porta de destino, Protocolo usado dentro do pacote, Hora do último pacote enviado ou recebido

- **Datagramas que chegam: substituir** (endereço IP do NAT, nova porta #) nos campos de destino de cada datagrama pelos correspondentes (endereço IP de origem, porta #) armazenados da tabela NAT

4 NAT: Network Address Translation

2: roteador NAT substitui end. origem do datagrama de 10.0.0.1, 3345 para 138.76.29.7, 5001, atualiza a tabela



1: hospedeiro 10.0.0.1 envia datagrama para 128.119.40, 80

3: resposta chega endereço de destino: 138.76.29.7, 5001

4: roteador NAT substitui o endereço de destino do datagrama de 138.76.29.7, 5001 para 10.0.0.1, 3345

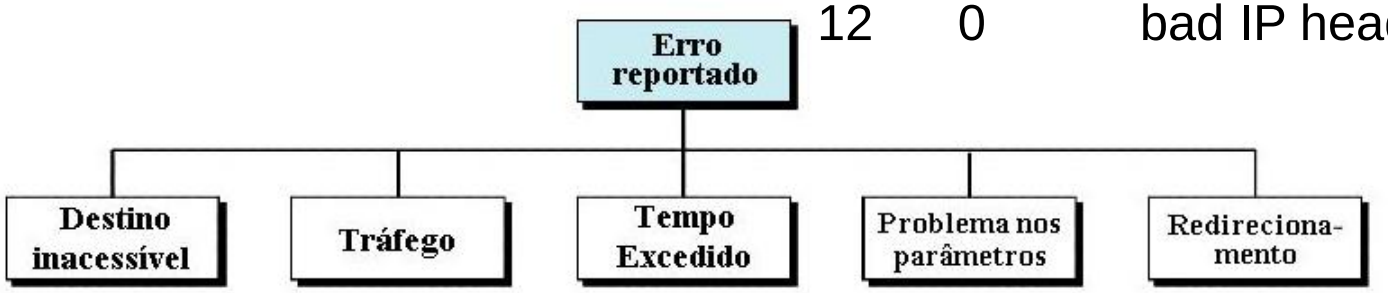
4 NAT: Network Address Translation

- Campo número de porta com 16 bits:
  - 65.535 conexões simultâneas por protocolo com um único endereço de LAN!
- NAT é controverso:
  - Número de porta deveria identificar processo e não hospedeiro.
  - Roteadores deveriam processar somente até a camada 3.
  - Violação do argumento fim-a-fim.
  - A possilidade de NAT deve ser levada em conta pelos desenvolvedores de aplicações, ex., Interfere em aplicações P2P.
  - A escassez de endereços deveria ser resolvida pelo IPv6.

4 ICMP: Internet Control Message Protocol

- Usado por hosts e roteadores para troca de informação de controle da camada de rede.
  - Error reporting.
  - Echo request/reply (usado pela aplicação ping).
- Transporte de mensagens:
  - Mensagens ICMP transportadas em datagramas IP.

Tipo	Código	descrição
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

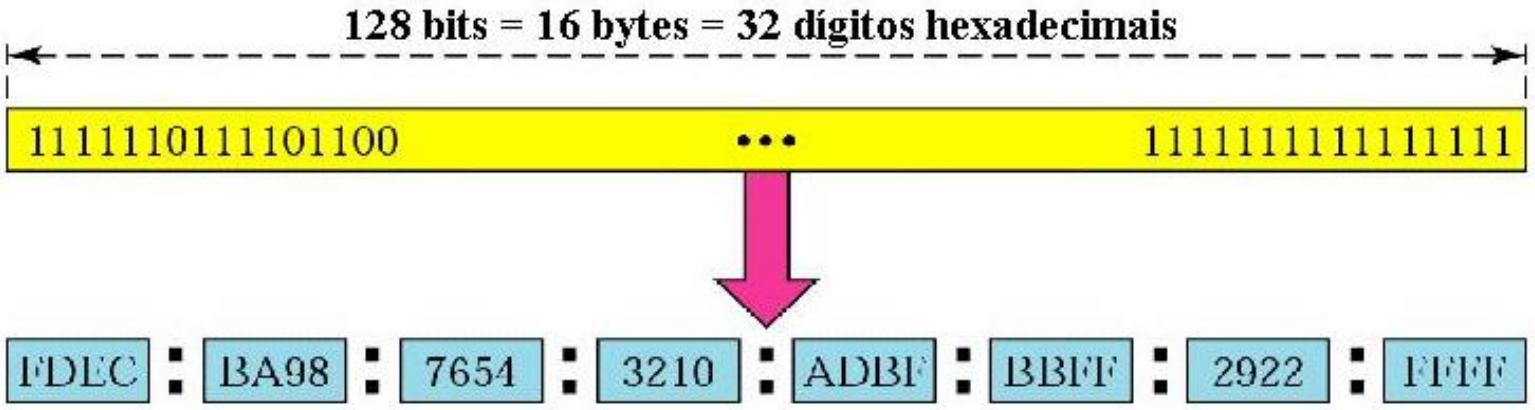


4 Traceroute e ICMP

- O transmissor envia uma série de segmentos UDP para o destino
  - O 1º possui TTL = 1
  - O 2º possui TTL = 2 etc.
  - nº de porta improvável.
- Quando o enésimo datagrama chega ao enésimo roteador:
  - O roteador descarta o datagrama.
  - E envia à origem uma mensagem ICMP (type 11, code 0).
  - A mensagem inclui o nome do roteador e o endereço IP.
- Quando a mensagem ICMP chega, a origem calcula o RTT.
- O traceroute faz isso três vezes.
- Critério de interrupção
- O segmento UDP finalmente chega ao hospedeiro de destino.
- O destino retorna o pacote ICMP “hospedeiro unreachable” (type 3, code 3).
- Quando a origem obtém esse ICMP, ela pára.

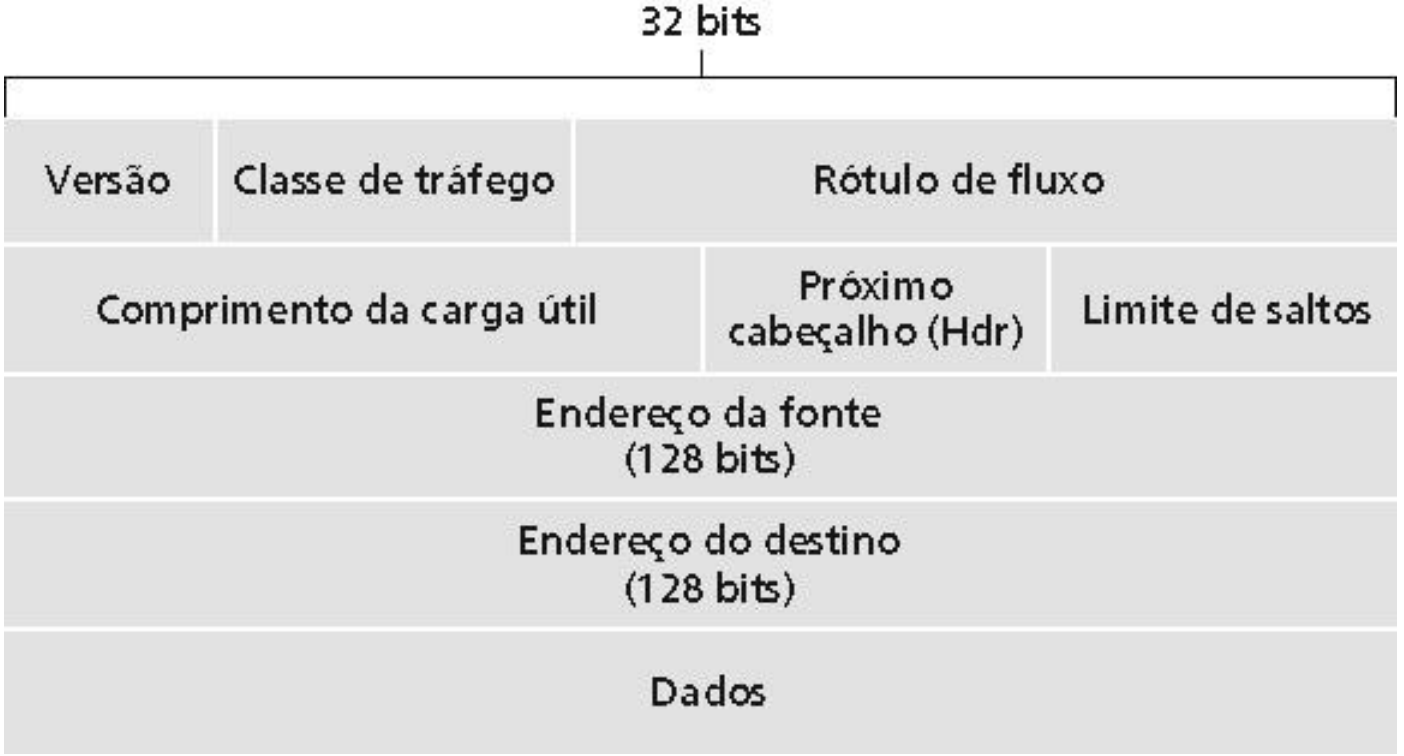
4 Cabeçalho IPv6

- Motivação inicial: o espaço de endereços de 32 bits está próximo de ser completamente alocado! Previsão 2008 e 2018!
- Motivação adicional:
  - Melhorar o formato do cabeçalho para permitir maior velocidade de processamento e de transmissão.
  - Mudanças no cabeçalho para incorporar mecanismos de controle de QoS.
  - Formato do datagrama IPv6:
  - Cabeçalho fixo de 40 bytes. Não é permitida fragmentação nos roteadores (apenas fonte e destino). Se roteador recebe um pacote muito grande, ele descarta!



4 Cabeçalho IPv6

Priority: permitir definir prioridades diferenciadas para vários fluxos de informação.
Fluxo label: identifica datagramas do mesmo "fluxo." (conceito de "fluxo" não é bem definido).
Next header: identifica o protocolo da camada superior ou um cabeçalho auxiliar.



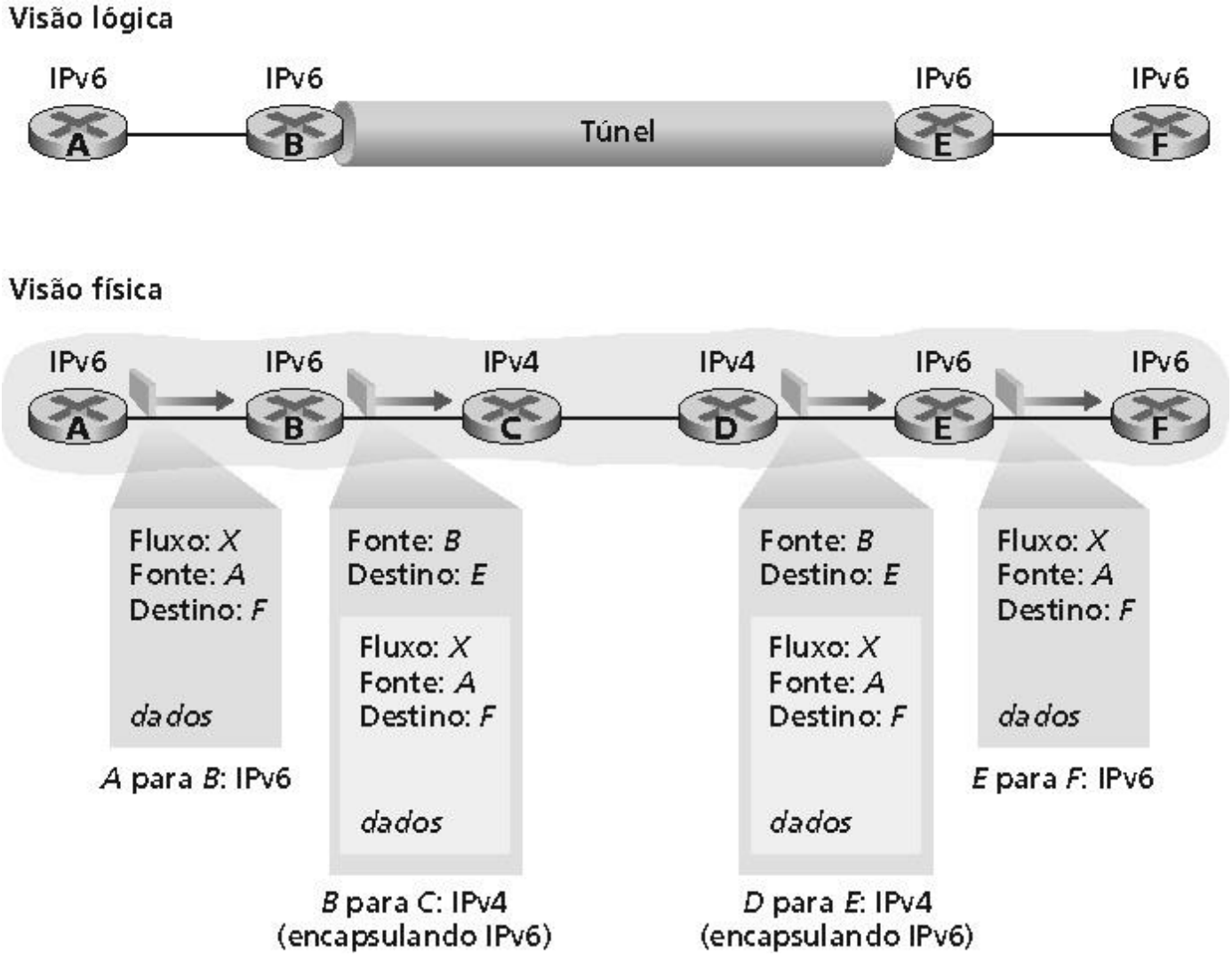
4 Outras mudanças do IPv4

- Checksum: removido inteiramente para reduzir o tempo de processamento em cada salto (redundante com a camada de transporte).
- Options: são permitidas, mas são alocadas em cabeçalhos suplementares, indicados pelo campo "Next header".
- ICMPv6: nova versão do ICMP
  - Tipos de mensagens adicionais , ex.: "Packet Too Big".
  - Funções de gerenciamento de grupos multicast.

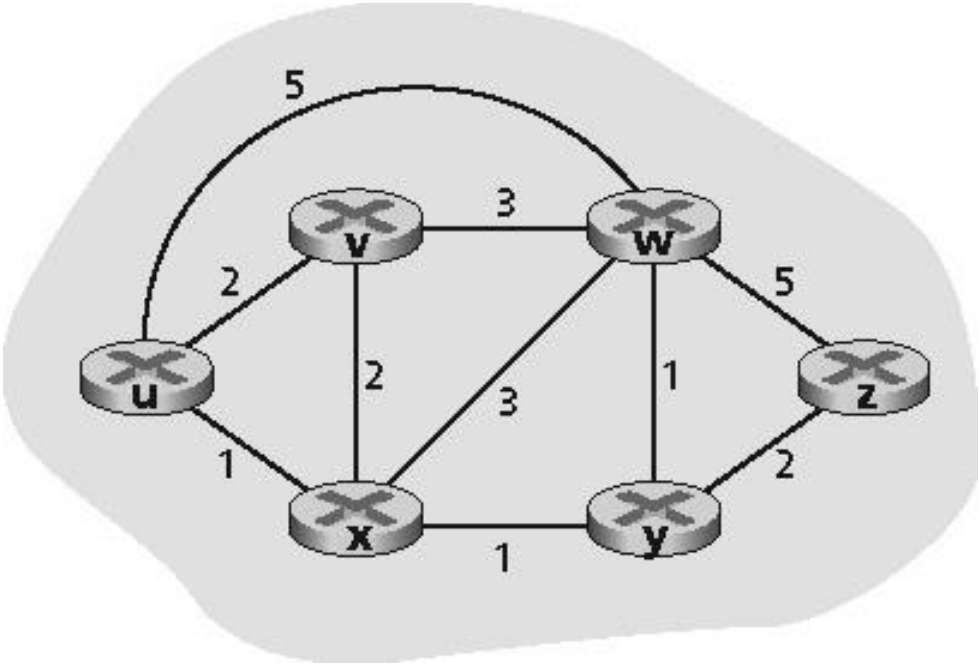
4 Transição do IPv4 para IPv6

- Nem todos os roteadores poderão ser atualizados simultaneamente
  - Não haverá um dia da vacinação!
  - Como a rede irá operar com roteadores mistos de IPV4 e IPV6?
- **Tunelamento**: IPv6 transportado dentro de pacotes IPv4 entre roteadores IPv4.

4 Tunelamento



4 Abstração do grafo

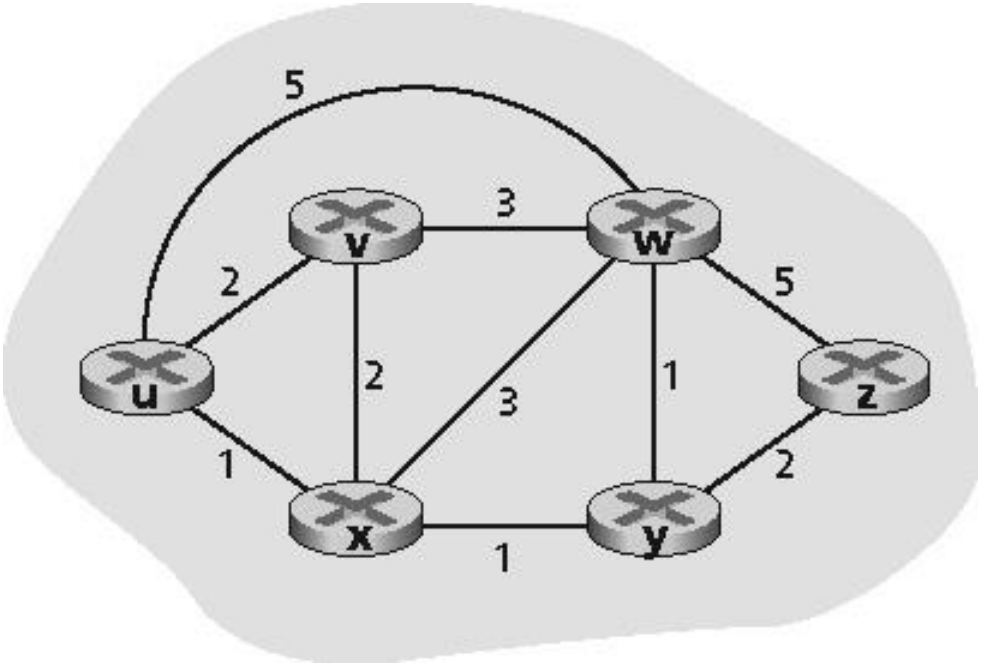


Grafo: G = (N,E)

N = conjunto de roteadores = { u, v, w, x, y, z }

E = conjunto de links ={ (u,v), (u,x), (v,x), (v,w), (x,w), (x,y), (w,y), (w,z), (y,z) }

4 Abstração do grafo: custo



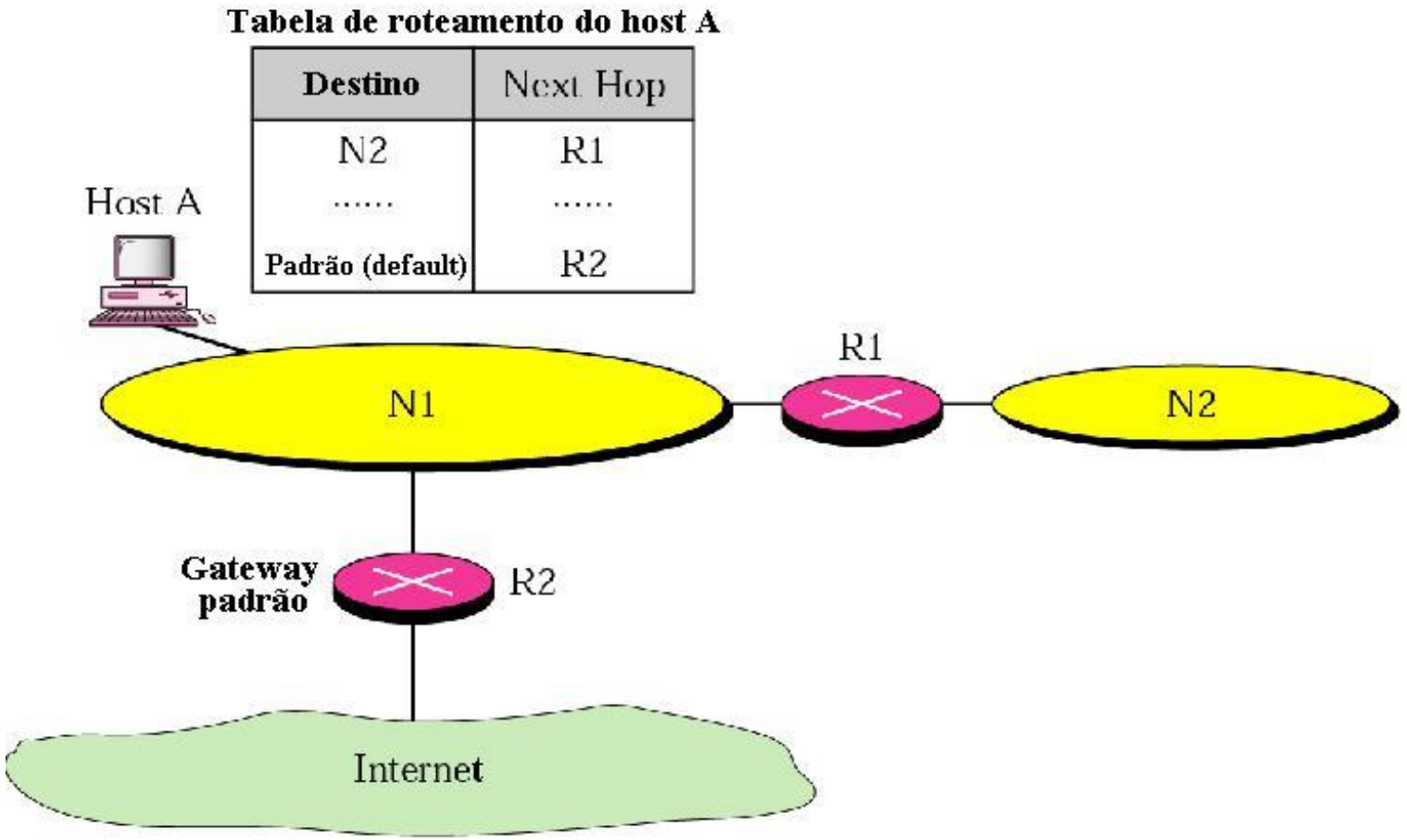
- c(x,x') = custo do link (x,x')
  - ex., c(w, z) = 5
- Custo poderia ser sempre 1, ou inversamente relacionado à largura de banda ou ao congestionamento

Custo do caminho $(x_1, x_2, x_3, \ldots, x_p) = c(x_1,x_2) + c(x_2,x_3) + \ldots + c(x_{p-1},x_p)$

Questão: Qual é o caminho de menor custo entre *u* e *z* ?

Algoritmo de roteamento: algoritmo que encontra o caminho de menor custo

4 Conceito: Rota Padrão



4 Classificação dos algoritmos de roteamento

Informação global ou descentralizada

Global:

- Todos os roteadores têm informações completas da topologia e do custos dos enlaces
- Algoritmos “link state” (estado dos enlaces)

Descentralizada:

- Roteadores só conhecem informações sobre seus vizinhos e os enlaces para eles
- Processo de computação interativo, troca de informações com os vizinhos
- Algoritmos “distance vector” (vetor de distância)

Estático ou dinâmico?

Estático:

- As rotas mudam lentamente ao longo do tempo.

Dinâmico:

- As rotas mudam mais rapidamente.
  - Podem responder a mudanças no custo dos enlaces;
  - Atualizações periódicas.

4 Algoritmo de roteamento link-state

- Algoritmo de Dijkstra
- Topologia de rede e custo dos enlaces são conhecidos por todos os nós.
  - Implementado via “link state broadcast”.
  - Todos os nós têm a mesma informação.
- Computa caminhos de menor custo de um nó (fonte) para todos os outros nós
  - Fornece uma tabela de roteamento para aquele nó.
- Convergência: após k iterações, conhece o caminho de menor custo para k destinos.

Notação:

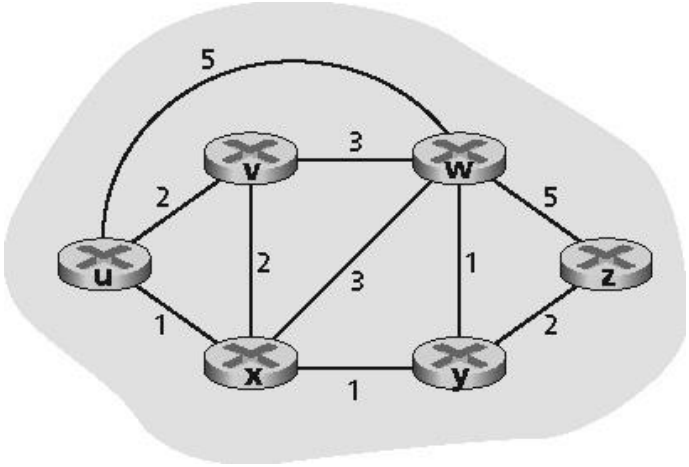
- C(i,j): custo do enlace do nó i ao nó j. Custo é infinito se não houver ligação entre i e j.
- D(v): valor atual do custo do caminho da fonte ao destino V.
- P(v): nó predecessor ao longo do caminho da fonte ao nó v, isto é, antes do v
- N’: conjunto de nós cujo caminho de menor custo é definitivamente conhecido .

4 Algoritmo de Dijsktra

```
1  Inicialização:
2    N' = {u}
3    para todos os nós v
4      se v é adjacente a u
5        então D(v) = c(u,v)
6        senão D(v) = ∞
7
8   Loop
9     ache w não em N' tal que D(w) é um mínimo
10    acrescente w  a N'
11    atualize D(v) para todo v adjacente a w e não em N':
12       D(v) = min( D(v), D(w) + c(w,v) )
13    /* novo custo para v é ou o custo anterior para v ou o menor
14     custo de caminho conhecido para w mais o custo de w a v */
15  até que todos os nós estejam em N'
```

4 Exemplo: Algoritmo de Dijsktra

Passo	início N	D(v),P(v)	D(w),P(w)	D(x),P(x)	D(y),P(y)	D(z),P(z)
0	u	2,u	5,u	1,u	∞	∞
→ 1	ux	2,u	4,x		2,x	∞
→ 2	uxy	2,u	3,y			4,y
→ 3	uxyv		3,y			4,y
→ 4	uxyvw					4,y
→ 5	uxyvwz					



4 Algoritmo vetor de distância

Idéia básica:

- Cada nó envia periodicamente sua própria estimativa de vetor de distância aos vizinhos.
- Quando o nó x recebe nova estimativa de DV do vizinho, ele atualiza seu próprio DV usando a equação B-F:

$$D_x(y) = \min_v\{c(x,v) + D_v(y)\} \quad \text{para cada nó } y \in N$$

- Ao menos em condições naturais, a estimativa $D_x(y)$ *converge para o menor custo atual* $d_x(y)$

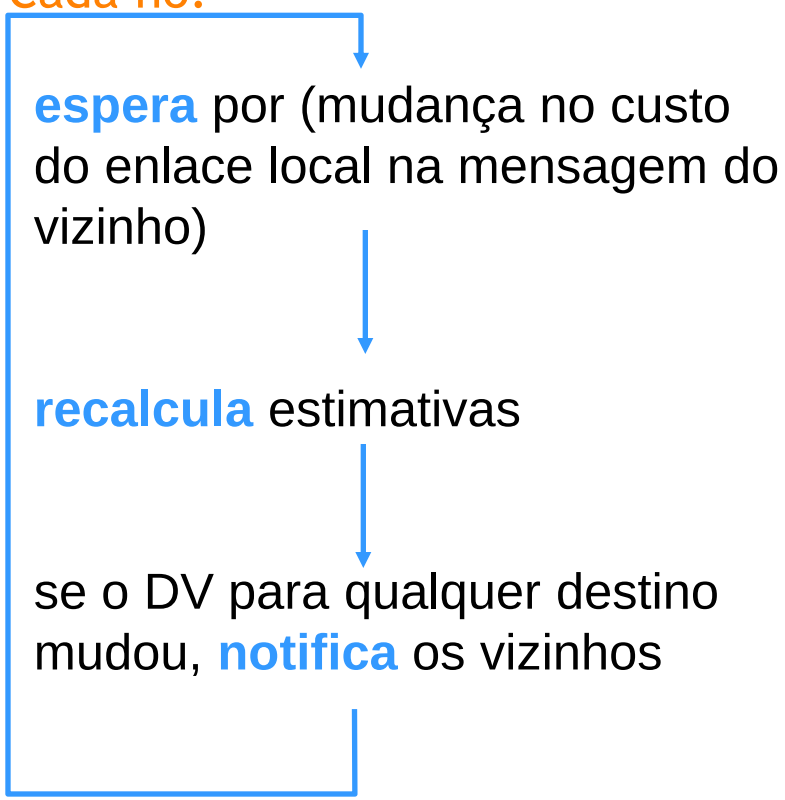
4 Algoritmo vetor de distância

Iterativo, assíncrono: cada iteração local é causada por:

- Mudança no custo do enlace local.
- Mensagem de atualização DV do vizinho.

Distribuído:

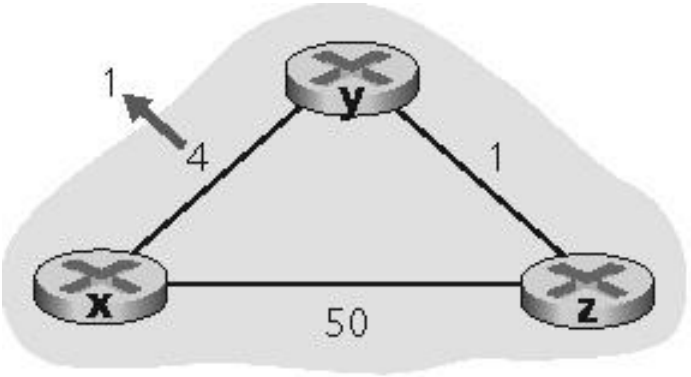
- Cada nó notifica os vizinhos **apenas** quando seu DV mudar.
  - Os vizinhos então notificam seus vizinhos, se necessário.



4 Vetor de distância: mudanças no custo do enlace

Mudanças no custo do enlace:

- Nó detecta mudança no custo do enlace local.
- Atualiza informações de roteamento, recalcula o vetor de distância.
- Se o DV muda, notifica vizinhos.



a.

No tempo t_0, *y* detecta a mudança no custo do enlace, atualiza seu DV e informa seus vizinhos.
No tempo t_1, *z* recebe a atualização de *y* e atualiza sua tabela.
Ele calcula o menor custo novo para *x* e envia seu DV para os vizinhos.
No tempo t_2, *y* recebe a atualização de que *z* atualizou sua tabela de distância.
O vetor de distância de *y* não muda e então *y* *não* envia nenhuma mensagem.

"boas notícias viajam depressa"

4 Comparação dos algoritmos LS e VD

Complexidade

- LS: com n nós, E links, O(NE) mensagens enviadas.
- DV: trocas somente entre vizinhos.
  - Tempo de convergência varia

Tempo de convergência

- LS: algoritmo $O(N^2)$ exige mensagens O(NE)
  - Pode ter oscilações
- DV: tempo de convergência varia
  - Pode haver loops de roteamento

Robustez: o que acontece se um roteador funciona mal?

Ls:

- Nós podem informar custos de **link** incorretos.
- Cada nó calcula sua própria tabela de roteamento

Dv:

- Nó DV pode informar custo de **caminho** incorreto
- Tabela de cada nó é usada por outros
  - Propagação de erros pela rede

4 Roteamento hierárquico

Autonomia administrativa

- Internet = rede de redes.
- Cada administração de rede pode querer controlar o roteamento na sua própria rede .

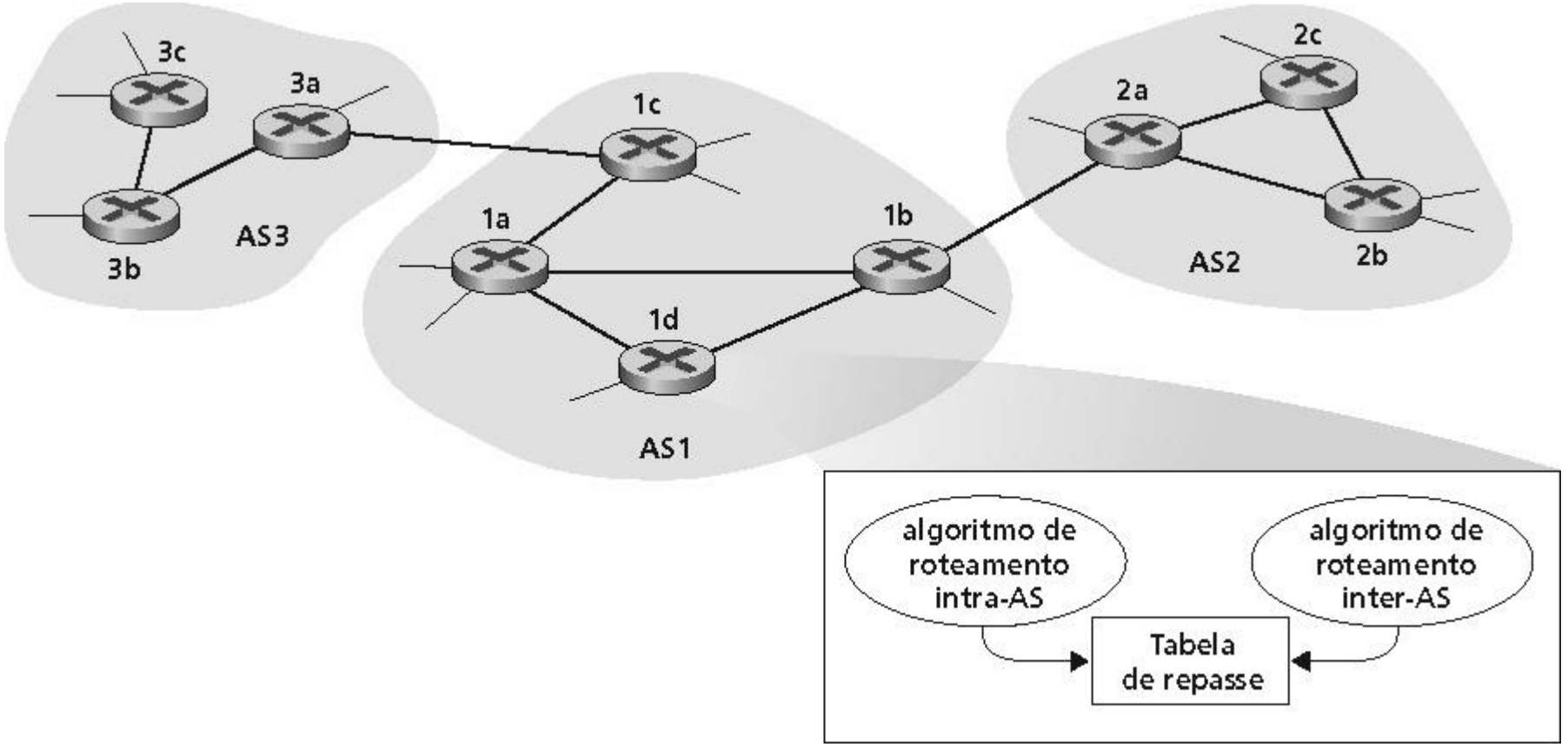
Agrega roteadores em regiões, “sistemas autônomos ” (AS).

- Roteadores no mesmo AS rodam o mesmo protocolo de roteamento.
  - Protocolo de roteamento “intra-AS”.
  - Roteadores em diferentes AS podem rodar diferentes protocolos de roteamento.

Roteador Gateway

- Link direto para um roteador em outro AS.

4 ASs interconectadas



- Tabela de roteamento é configurada por ambos algoritmos, intra- e inter-AS
  - Intra-AS estabelece entradas para destinos internos.
  - Inter-AS e intra-As estabelecem entradas para destinos externos.

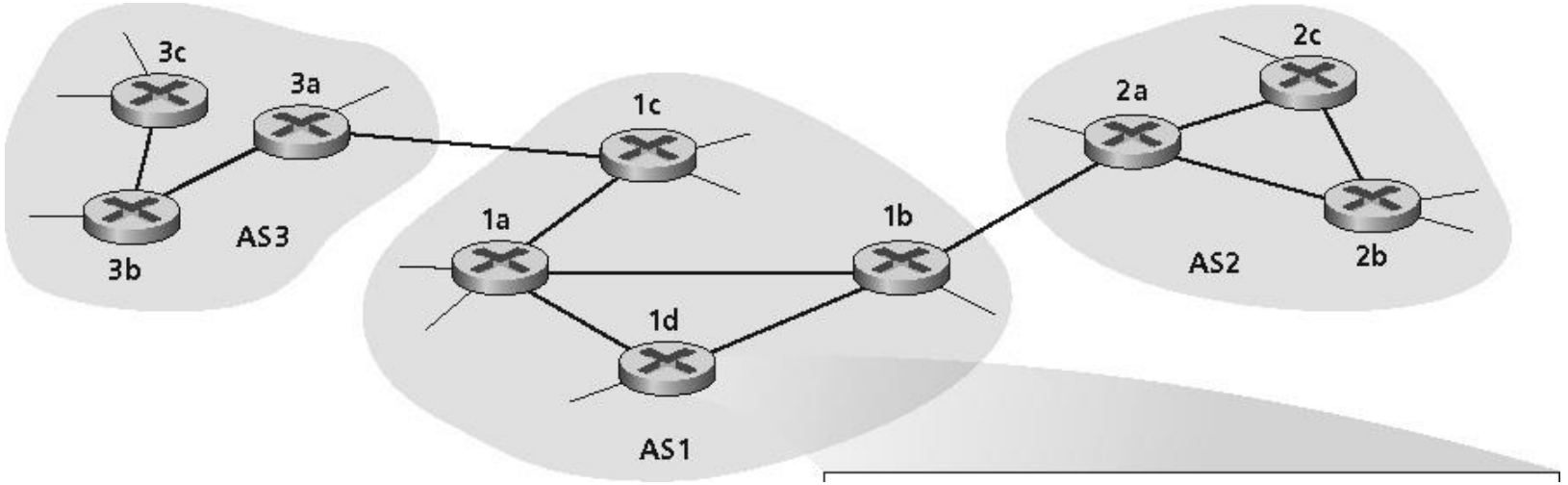
4 Tarefas Inter-AS

- Suponha que um roteador no AS1 receba um datagrama cujo destino seja fora do AS1.
  - O roteador deveria encaminhar o pacote para os roteadores gateway, mas qual deles?

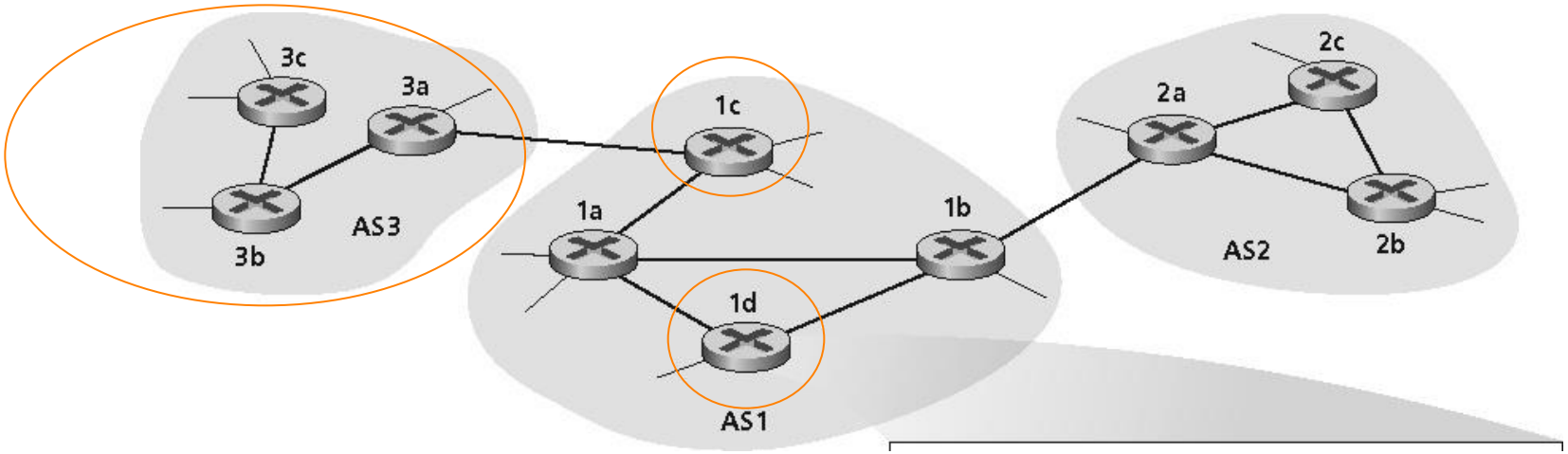
AS1 precisa:

1. Aprender quais destinos são alcancáveis através de AS2 e através de AS3.
2. Propagar suas informações de alcance para todos os roteadores em AS1.

Tarefa para o roteamento inter-AS routing!



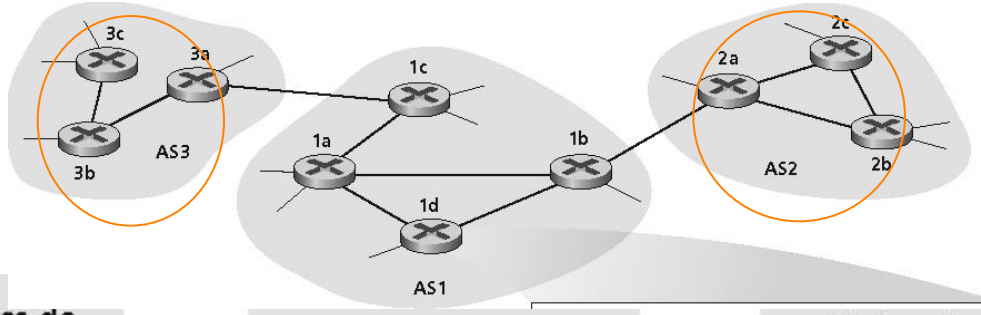
4 Exemplo: Ajustando a tabela de roteamento no roteador 1d



- Suponha que AS1 aprende pelo protocolo inter-AS protocol que a sub-rede **x** é alcancável através de AS3 (gateway 1c) mas não através de AS2.
- O protocolo intra-AS propaga informações de alcance para todos os roteadores internos.
- Baseado nas informações de roteamento intra-AS, o roteador 1d determina que sua interface **I** está no caminho de menor custo para 1c.
- Coloca na tabela de roteamento a entrada **(x,I).**

4 Exemplo: Escolhendo entre múltiplos ASs

- Agora suponha que AS1 aprende pelo protocolo inter-AS que a sub-rede **x** é alcançavel através de AS3 *e* através de AS2.
- Para configurar a tabela de roteamento, o roteador 1d deve determinar por qual gateway ele deve encaminhar os pacotes para o destino x.
- Isso também é tarefa para o protocolo de roteamento inter-AS.
- Roteamento de "batata-quente": envia o pacote para o mais próximo de dois roteadores.



Aprende por um protocolo inter-AS que a sub-rede x pode ser alcançada via vários roteadores de borda → Usa informações de roteamento do protocolo intra-AS para determinar custos de caminhos de menor custo para cada um dos roteadores de borda. → Roteamento da batata quente: escolhe o roteador de borda que tenha o menor custo. → Determina, pela tabela de repasse, a interface I que leva ao roteador de borda de menor custo. Adiciona (x,I) à tabela de repasse.

4 Roteamento intra-AS

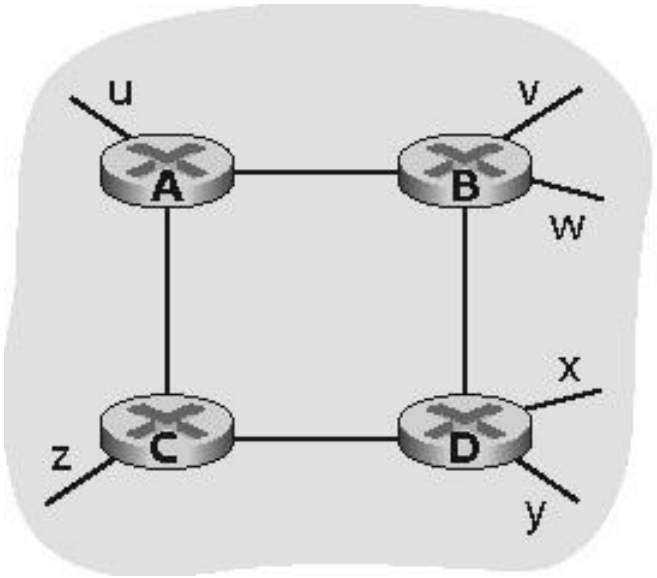
- Também conhecido como Interior Gateway Protocols (IGP)
- Protocolos de roteamento intra-AS mais comuns:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (proprietário da Cisco)

4 RIP (Routing Information Protocol)

- Algoritmo do tipo vetor distância.
- Métrica de distância: # de saltos (máx. = 15 saltos)

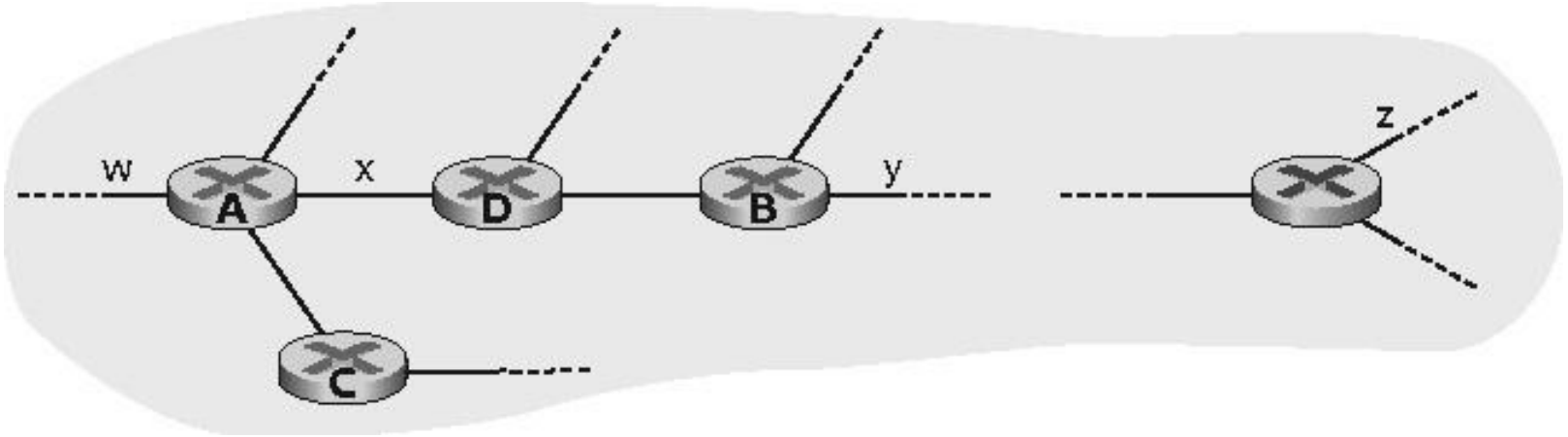
Vetores de distância: trocados a cada 30 s via Response Message (também chamado advertisement, ou anúncio).

- Cada anúncio indica rotas para até 25 redes de destino.



Destino	Saltos
u	1
v	2
w	2
x	3
y	3
z	2

4 RIP: Exemplo



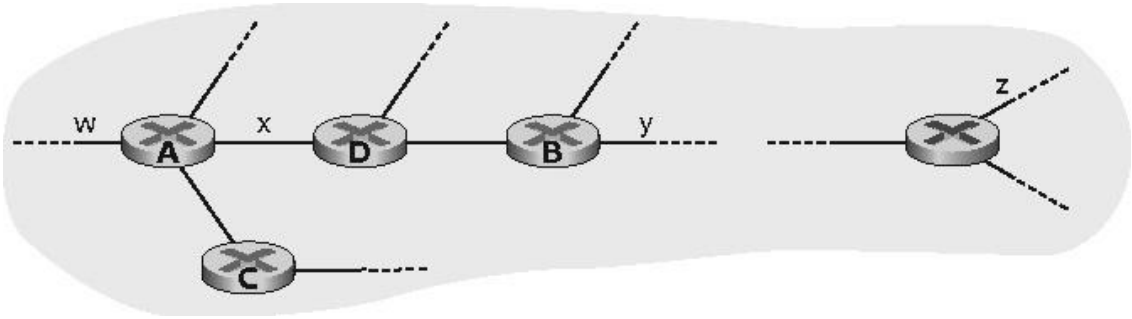
rede de destino	roteador seguinte	núm. de saltos para dest.
w	A	2
y	B	2
z	B	7
x	--	1
....	....	....

Tabela de roteamento em D

4 RIP: Exemplo

dest.	próximos saltos	
w	-	-
x	-	-
z	C	4
....	...	...

Anúncio de A para D



rede de destino	roteador seguinte	núm. de saltos até dest.
w	A	2
y	B	2
z	~~B~~ A	~~7~~ 5
x	--	1
....	....	....

Tabela de Roteamento em D

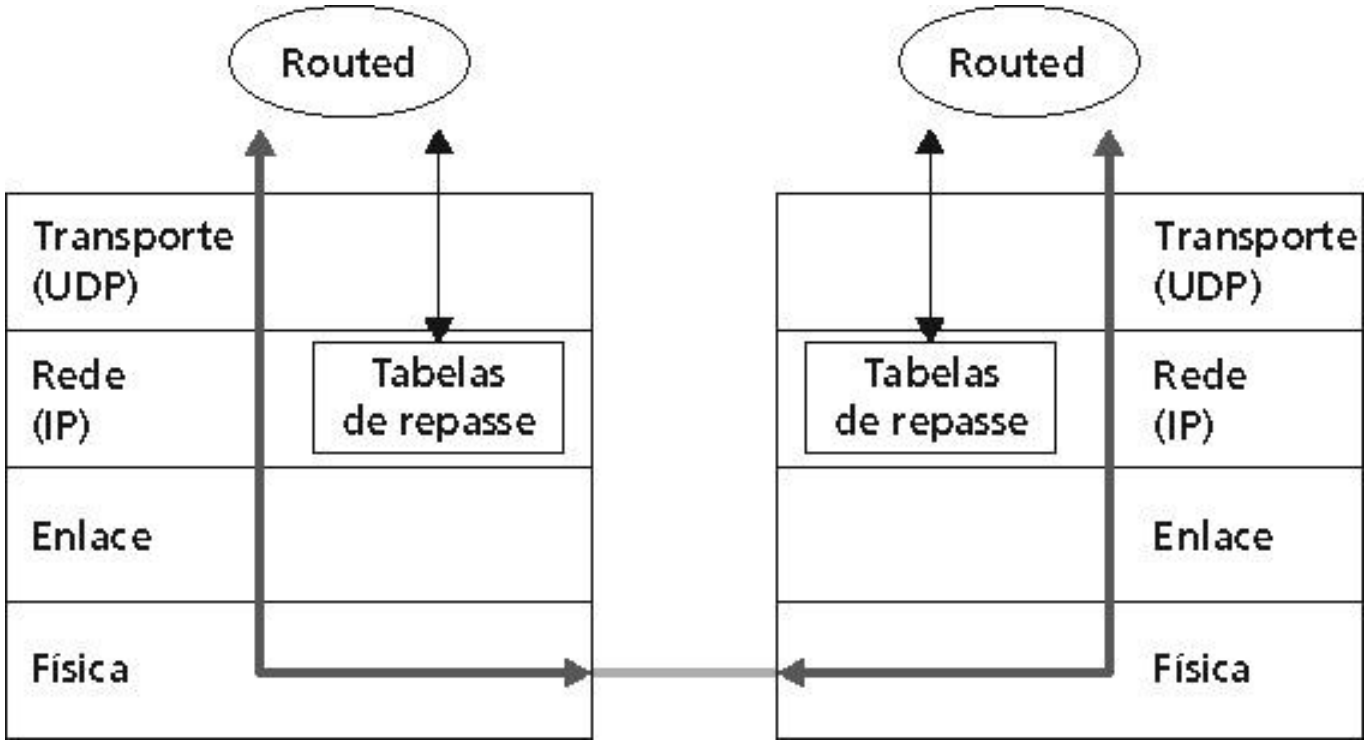
4 RIP: falha de enlances e recuperação

Se não há um aviso depois de 180 s --> o vizinho e o enlace são declarados mortos.

- Rotas através do vizinho são anuladas.
- Novos anúncios são enviados aos vizinhos.
- Os vizinhos por sua vez devem enviar novos anúncios (se suas tabelas de rotas foram alteradas).
- A falha de um enlace se propaga rapidamente para a rede inteira.

4 RIP Processamento da Tabela de Rotas

- As tabelas de roteamento do RIP são manipuladas por um processo de aplicação chamado route-d (daemon)
- Anúncios são enviados em pacotes UDP com repetição periódica



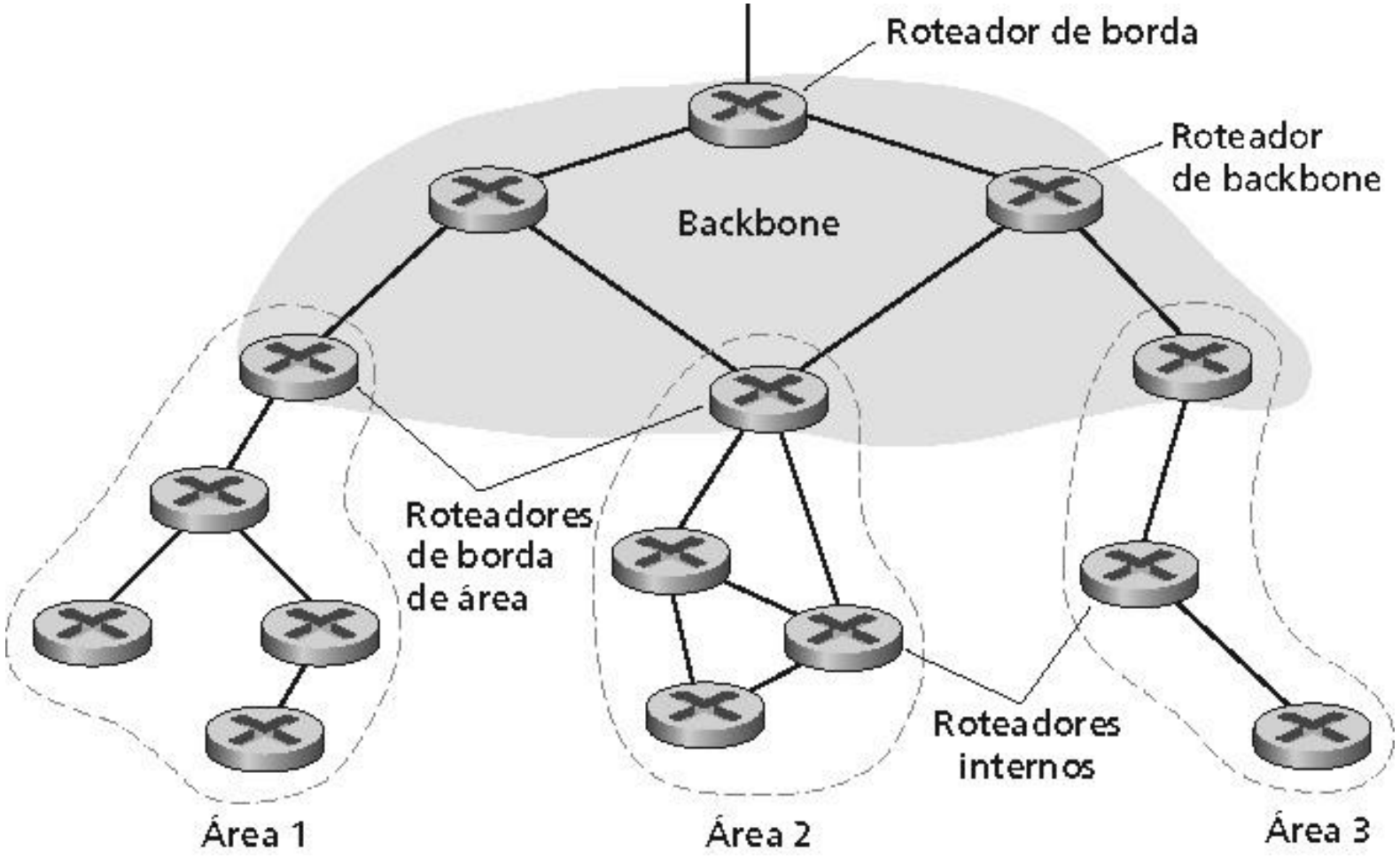
4 OSPF (Open Shortest Path First)

- "open": publicamente disponível.
- Usa algoritmo do tipo link state .
  - Disseminação de pacotes LS.
  - Mapa topológico em cada nó.
  - Usa algoritmo de Dijkstra para cálculo de rotas.
- Anúncios do OSPF transportam um registro para cada roteador vizinho.
- Anúncios são distribuídos para todo o AS (via flooding).

4 OSPF características “avançadas”

- Segurança: todas as mensagens do OSPF são autenticadas (para previnir intrusões maliciosas).
- Múltiplos caminhos de mesmo custo são permitidos (o RIP só permite um caminho).
  - Neste caso, há balanceamento de carga!
- Para cada link, múltiplas métricas de custo para TOS diferentes (ex., custo de enlace por satélite definido baixo para tráfego de “melhor esforço” e alto para serviços de tempo real).
- Integra tráfego uni- e multicast :
  - Multicast OSPF (MOSPF) usa a mesma base de dados de topologia do OSPF.
- OSPF hierárquico: OSPF para grandes domínios.

4 OSPF hierárquico



4 OSPF hierárquico

- Hierarquia de dois níveis: área local e backbone.
  - Anúncios de link state apenas nas áreas.
  - Cada nó tem a topologia detalhada da área, mas somente direções conhecidas (caminhos mais curtos) para redes em outra áreas.
- Roteadores de borda de área: “resumem” distâncias para redes na própria área e enviam para outros roteadores de borda de área.
- Roteadores de backbone: executam o roteamento OSPF de forma limitada ao backbone.
- Roteadores de borda: conectam-se a outros ASs.

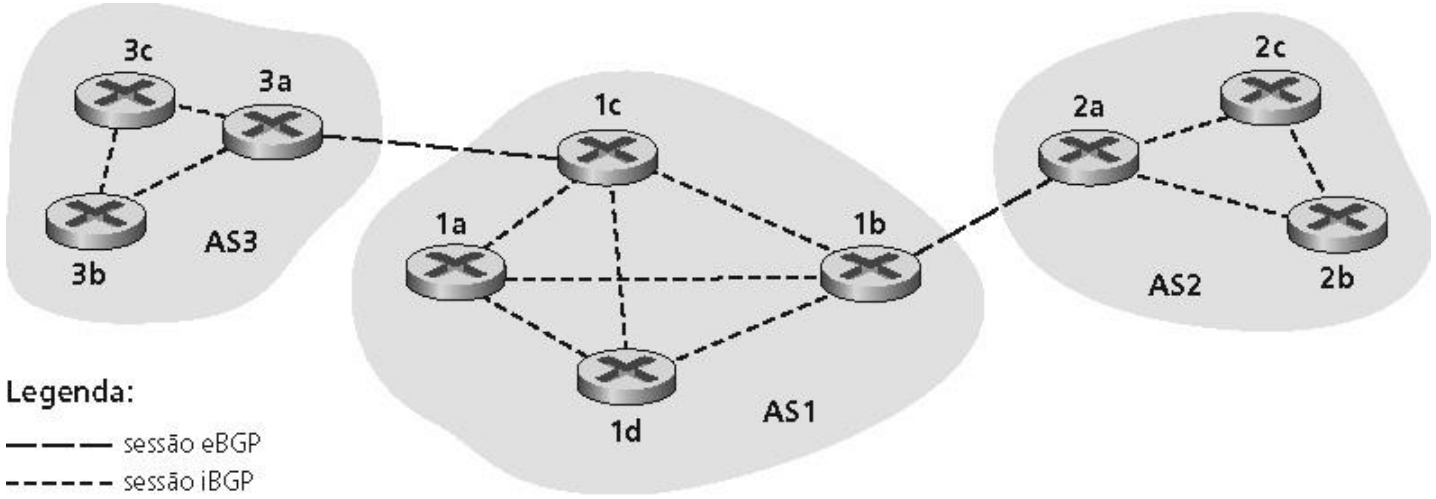
4 Roteamento inter-AS da Internet: BGP

BGP (Border Gateway Protocol): é o padrão de fato para uso na Internet.

- BGP provê cada AS dos meios para:
  1. Obter informações de alcance de sub-rede dos ASs Vizinhos
  2. Propagar informações de alcance para todos os roteadores internos ao AS.
  3. Determinar “boas” rotas para as sub-redes baseado em informações de alcance e política.
- Permite que uma subrede comunique sua existência para o resto da Internet: **“Estou aqui”**

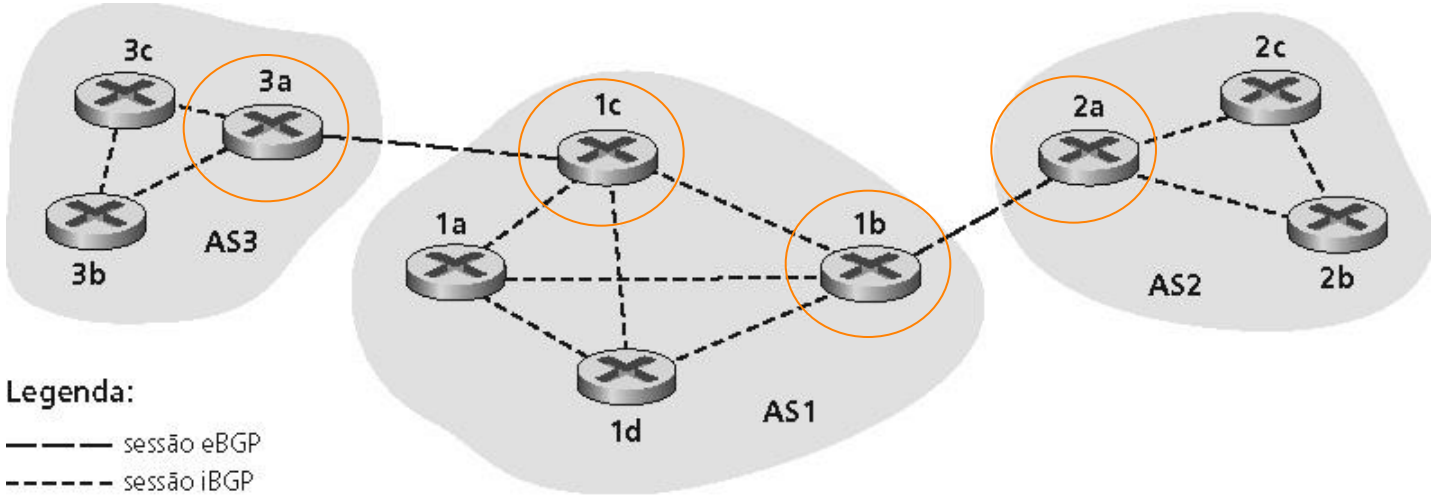
4 Roteamento inter-AS da Internet: BGP

- Pares de roteadores (BGP peers) trocam informações de roteamento por conexões TCP semipermanentes: sessões BGP.
- As sessões BGP não correspondem aos links físicos.
- Quando AS2 comunica um prefixo ao AS1, AS2 está **prometendo** que irá encaminhar todos os datagramas destinados a esse prefixo em direção ao prefixo.



4 Distribuindo informações de alcance

- Em cada sessão eBGP entre 3a e 1c, AS3 envia informações de alcance de prefixo para AS1.
- 1c pode então usar iBGP para distribuir essa nova informação de alcance de prefixo para todos os roteadores em AS1.
- 1b pode recomunicar essa nova informação para AS2 por meio da sessão eBGP 1b-para-2a.
- Quando um roteador aprende um novo prefixo, ele cria uma entrada para o prefixo em sua tabela de roteamento.



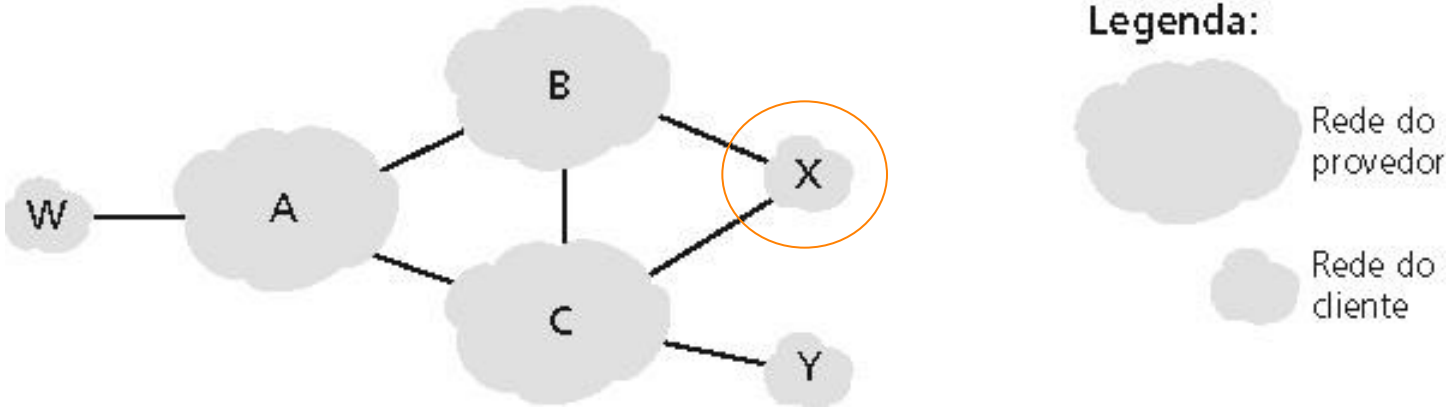
4 Atributos de caminho e rotas BGP

- Quando se comunica um prefixo, o comunicado inclui os atributos do BGP.
  - Prefixo + atributos = “rota”
- Dois atributos importantes:
  - AS-PATH: contém os ASs pelos quais o comunicado para o prefixo passou: AS 67 AS 17.
  - NEXT-HOP: Indica o roteador específico interno ao AS para o AS do próximo salto (next-hop). (Pode haver múltiplos links do AS atual para o AS do próximo salto.)

4 BGP: seleção de rota

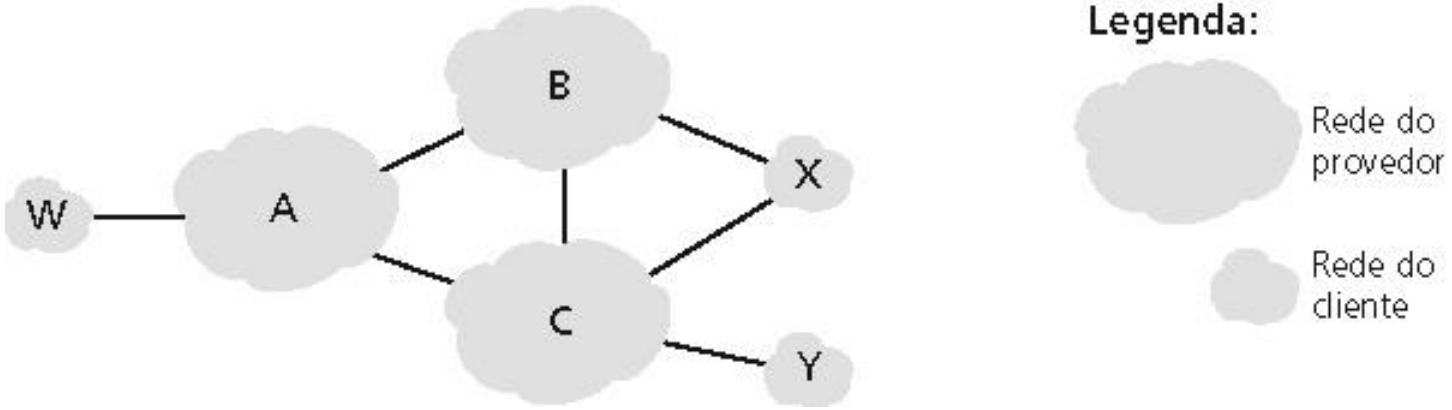
- Um roteador pode aprender mais do que 1 rota para o mesmo prefixo. O roteador deve selecionar uma rota.
- Regras de eliminação:
  - Atributo de valor de preferência local: decisão de política.
  - AS-PATH (caminho) mais curto
  - Roteador do NEXT-HOP (próximo salto) mais próximo: roteamento da "batata quente".
  - Critérios adicionais .

4 BGP: política de rotear



- A,B,C são redes do provedor.
- X,W,Y são clientes (das redes do provedor).
- X é dual-homed: anexados a duas redes.
  - X não quer rotear de B via X para C
  - … então X não comunicará ao B uma rota para C.

4 BGP: política de roteamento (2)



- A comunica ao B o caminho AW
- B comunica ao X o caminho BAW
- B deveria comunicar ao C o caminho BAW?
  - De jeito nenhum! B não obtém nenhum "rendimento" em rotear CBAW pois nem W nem C são seus clientes.
  - B quer forçar C a rotear para W via A
  - B quer rotear **somente** de/para seus clientes!

4 Por que os protocolos intra- e inter-AS são diferentes?

Políticas:

- Inter-AS: a administração quer ter controle sobre como seu tráfego é roteado e sobre quem roteia através da sua rede.
- Intra-AS: administração única, então não são necessárias políticas de decisão.

Escalabilidade

- O roteamento hierárquico poupa espaço da tabela de rotas e reduz o tráfego de atualização.

Desempenho:

- Intra-AS: preocupação maior é desempenho.
- Inter-AS: políticas podem ser dominantes em relação ao desempenho.

4 Disciplinas de serviço

Tipos de filas:

- DropTail (FIFO)
- Round-Robin (RR)
- Weighted Round-Robin (WRR)
- Stochastic Fair Queuing (SFQ)
- Random Early Detection (RED)
- Hierárquica

4 Disciplinas de serviço - DropTail

É o tipo clássifo de fila, configurado por padrão nos roteadores

- FIFO (First In, First Out)
- A memória que deve ser alocada para a fila depende da capacidade (bps) e do atraso (s) de propagação do enlace:
  - $M = C$A
  - Exemplo: um enlace de 1Gbps com atraso de 20 ms deve ter uma fila com tamanho de de 1Gbps x 0,02s = 20 Mb = 2,5 MB

4 Disciplinas de serviço - RR

Round-Robin é uma fila que tenta dar justiça

- As conexões são divididas em classes e cada classe tem o mesmo tipo tempo de processamento. Pode ser injusta com classes com maior volume de tráfego

4 Disciplinas de serviço - WRR

Weighted Round-Robin é outra fila que tenta dar justiça

- As conexões são divididas em classes, agora com pesos diferentes, e cada classe tem o mesmo tempo de processamento vezes o peso da casse. Precisa ser bem configurada.

4 Disciplinas de serviço - SFQ

Stochastic Fair Queuing é uma fila que tenta dar justiça

- Tenta diversificar os descartes nas conexões de forma proporcional ao número de pacotes da conexão, evitando o descarte de vários pacotes de uma mesma conexão e poucos pacotes de outra

Disciplinas de serviço - RED

Random Early Detection é uma fila que tenta evitar a sincronização global do TCP

- São configurados dois limiares de ocupação da fila (por exempo, 10% e 90%) e dois percentuais de descarte de pacotes (por exempo, 1% e 20%). Caso o uso da fila esteja abaixo do primeiro limiar (10%), nenhum pacote é descartado. Caso o uso da fila esteja entre o primeiro e o segundo limiar de ocupação, os pacotes são descartados no primeiro percentual (1% dos pacotes). Caso o uso da fila esteja acima do segundo limiar de ocupação (90%), os pacotes são descartados no segundo porcentual (20%). Quando a fila está cheia, todos os pacotes que chegam são descartados.

4 Disciplinas de serviço - Hierárquica

Fila hierárquica é a junção de outras filas para fazer uma fila mais complexa

- Você pode usar diferentes filas para diferentes tipos de tráfego
- Por exemplo: usar a fila RED para tráfego web e de streaming, a fila SFQ para tráfego de jogos e a fila DropTail para tráfego de console remoto

Dúvidas?

Fontes: 1. Redes de Computadores e a Internet. James Kurose;
2. Comunicação de Dados e Redes de Computadores. Forouzan.