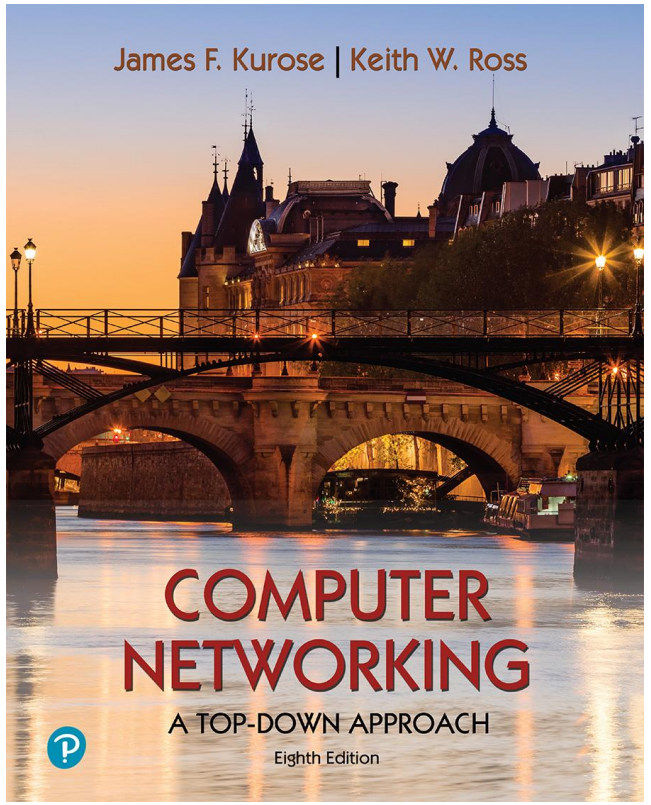
Capítulo 4
Camada de rede: Plano de dados

Uma observação sobre o uso desses slides do PowerPoint:
Estamos disponibilizando esses slides gratuitamente para todos (professores, alunos, leitores). Eles estão no formato PowerPoint para que você veja as animações e possa adicionar, modificar e excluir slides (inclusive este) e o conteúdo dos slides para atender às suas necessidades. Obviamente, eles representam *muito* trabalho de nossa parte. Em troca do uso, pedimos apenas o seguinte:

- Se você usar esses slides (por exemplo, em uma aula), mencione a fonte (afinal, gostaríamos que as pessoas usassem nosso livro!)
- Se você publicar algum slide em um site www, informe que ele foi adaptado de nossos slides (ou talvez idêntico a eles) e informe nossos direitos autorais sobre esse material.

Para obter um histórico de revisões, consulte a nota do slide desta página.

Obrigado e divirta-se! JFK/KWR



Redes de computadores: A Top-Down Approach (Uma abordagem de cima para baixo)
8th edition
Jim Kurose, Keith Ross
Pearson, 2020

Camada de rede: nossos objetivos

- compreender os princípios por trás dos serviços da camada de rede, com foco no plano de dados:
  - modelos de serviço da camada de rede
  - encaminhamento versus roteamento
  - Como funciona um roteador
  - endereçamento
  - encaminhamento generalizado
  - Arquitetura da Internet
- instanciação, implementação na Internet
  - Protocolo IP
  - NAT, middleboxes

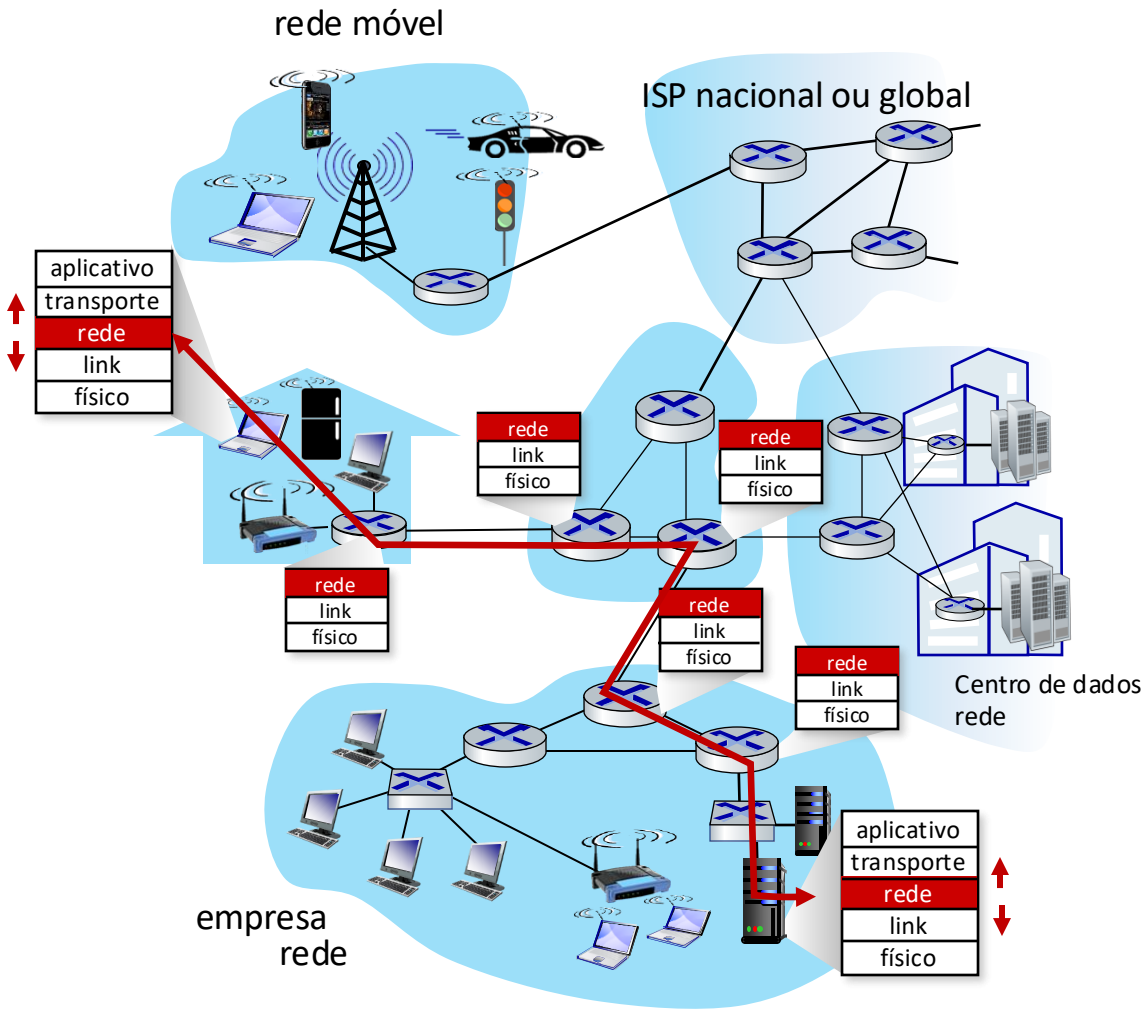
Camada de rede: Roteiro do "plano de dados"

- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo de Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereços de rede
  - IPv6

- Encaminhamento generalizado, SDN
  - Partida+ação
  - OpenFlow: match+action em ação
- Caixas intermediárias

Serviços e protocolos da camada de rede

- segmento de transporte do host de envio para o host de recebimento
  - remetente: encapsula segmentos em datagramas, passa para a camada de link
  - receptor: entrega segmentos ao protocolo da camada de transporte
- protocolos de camada de rede em *todos os dispositivos da Internet*: hosts, roteadores
- roteadores:
  - examina os campos de cabeçalho em todos os datagramas IP que passam por ele



Duas funções-chave da camada de rede

funções da camada de rede:

- *encaminhamento:* mover pacotes do link de entrada de um roteador para o link de saída apropriado do roteador
- *Roteamento:* determina a rota seguida pelos pacotes da origem ao destino
  - *algoritmos de roteamento*

analogia: fazer uma viagem

- *encaminhamento:* processo de passar por um único intercâmbio
- *roteamento:* processo de planejamento da viagem da origem ao destino

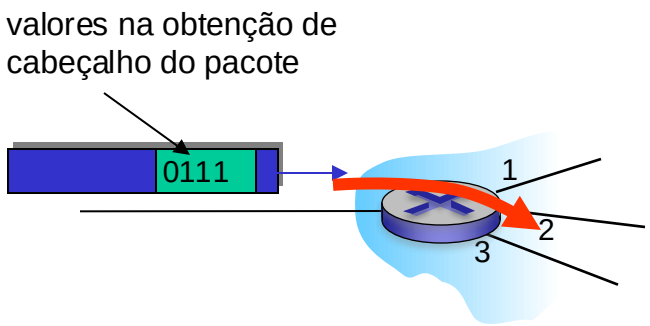
encaminhamento

roteamento

Camada de rede: plano de dados, plano de controle

Plano de dados:

- função *local*, por roteador
- determina como o datagrama que chega à porta de entrada do roteador é encaminhado para a porta de saída do roteador



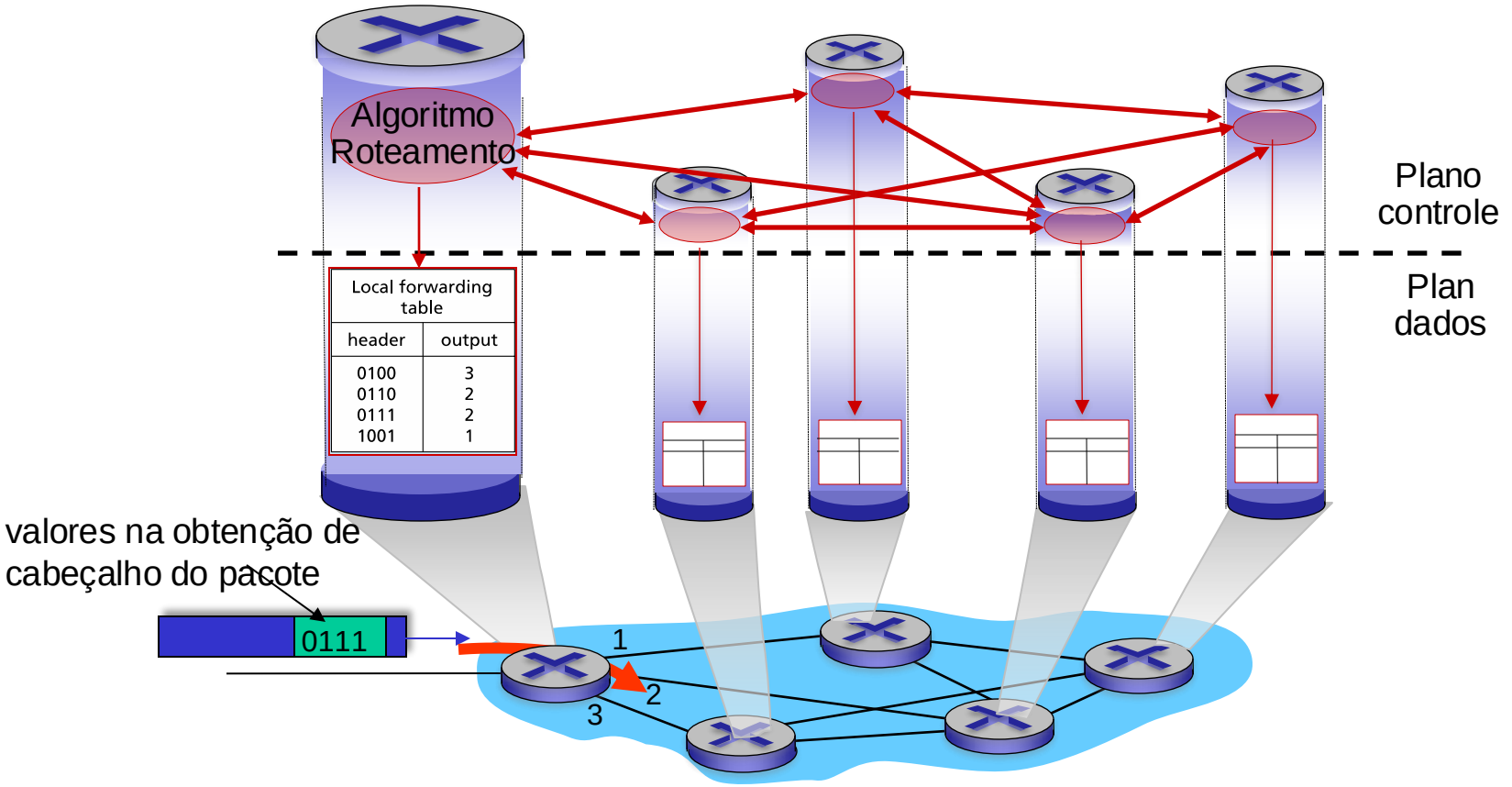
Plano de controle

- lógica *em toda a rede*
- determina como o datagrama é roteado entre os roteadores ao longo do caminho final do host de origem para o host de destino

- duas abordagens de plano de controle:
  - *algoritmos de roteamento tradicionais:* implementados em roteadores
  - *rede definida por software (SDN)*: implementada em servidores (remotos)

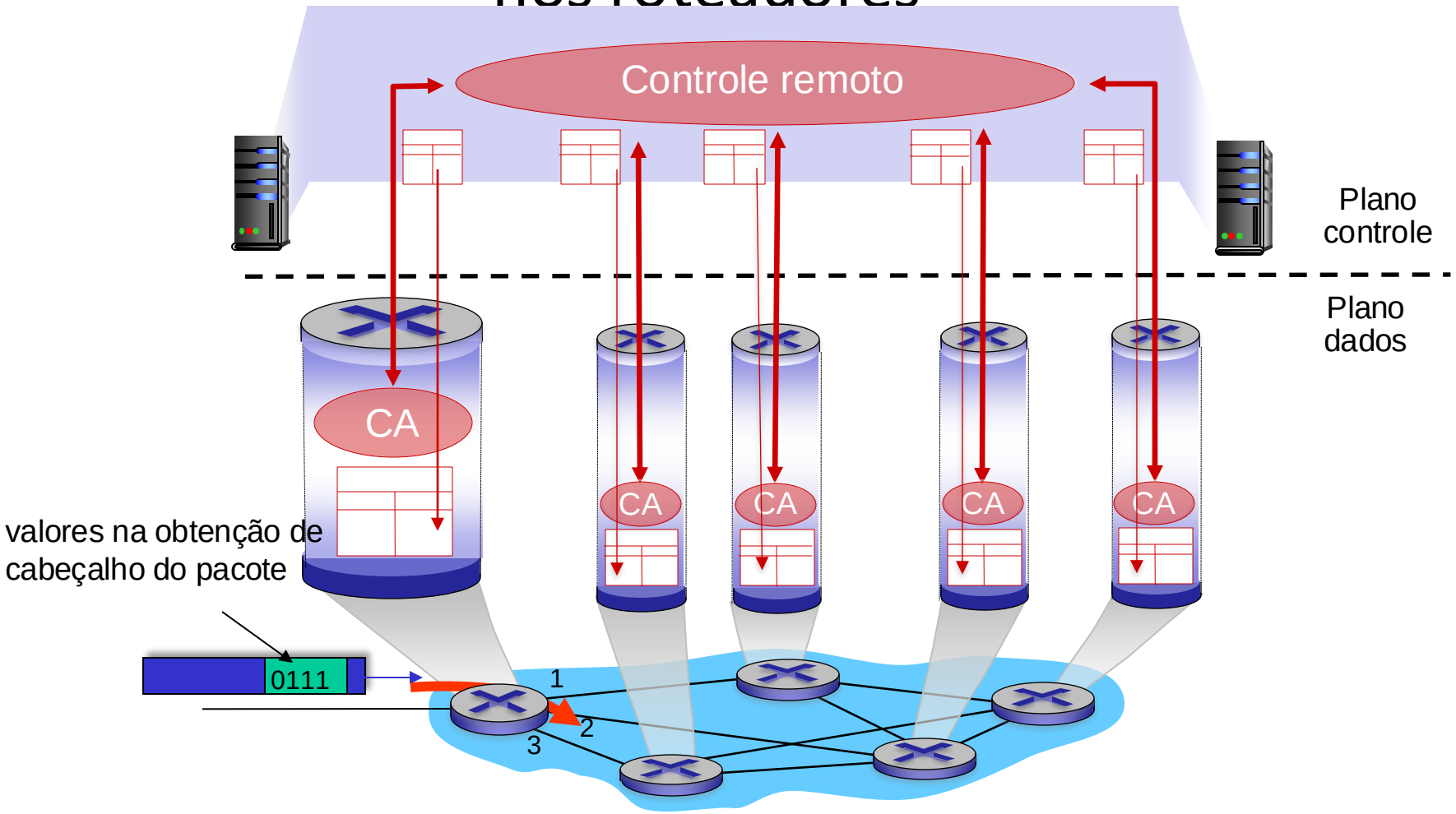
Plano de controle por roteador

Os componentes individuais do algoritmo de roteamento *em cada roteador* interagem no plano de controle



Plano de controle de SDN (Software-Defined Networking)

O controlador remoto calcula e instala tabelas de encaminhamento nos roteadores



Modelo de serviço de rede

P: Qual é o *modelo de serviço* para o "canal" que transporta datagramas do remetente para o receptor?

serviços de exemplo para datagramas *individuais*:

- entrega garantida
- entrega garantida com menos de 40 mseg de atraso

exemplo de serviços para um *fluxo* de datagramas:

- entrega de datagrama em ordem
- largura de banda mínima garantida para o fluxo
- restrições sobre alterações no espaçamento entre pacotes

Modelo de serviço da camada de rede

Rede Arquitetura	Serviço Modelo	Garantias de qualidade de serviço (QoS) ?			
		Largura de banda	Perda	Pedido	Cronograma
Internet	melhor esforço	nenhum	não	não	não

Modelo de serviço de "melhor esforço" da Internet

Não há garantias*:*

i. entrega bem-sucedida do datagrama ao destino
ii. prazo ou ordem de entrega
iii. largura de banda disponível para o fluxo final

Modelo de serviço da camada de rede

Rede Arquitetura	Serviço Modelo	Garantias de qualidade de serviço (QoS) ?			
		Largura de banda	Perda	Pedido	Cronograma
Internet	melhor esforço	nenhum	não	não	não
ATM	Taxa de bits constante	Taxa constante	sim	sim	sim
ATM	Taxa de bits disponível	Mínimo garantido	não	sim	não
Internet	Intserv Garantido (RFC 1633)	sim	sim	sim	sim
Internet	Diffserv (RFC 2475)	possível	possível	possível	não

Reflexões sobre o serviço de melhor esforço:

- A simplicidade do mecanismo permitiu que a Internet fosse amplamente implantada e adotada
- o provisionamento suficiente de largura de banda permite que o desempenho de aplicativos em tempo real (por exemplo, voz interativa, vídeo) seja "bom o suficiente" na "maior parte do tempo"
- Serviços replicados e distribuídos na camada de aplicativos (centros de dados, redes de distribuição de conteúdo) que se conectam perto das redes dos clientes e permitem que os serviços sejam fornecidos de vários locais
- o controle de congestionamento de serviços "elásticos" ajuda

É difícil contestar o sucesso do modelo de serviço de melhor esforço

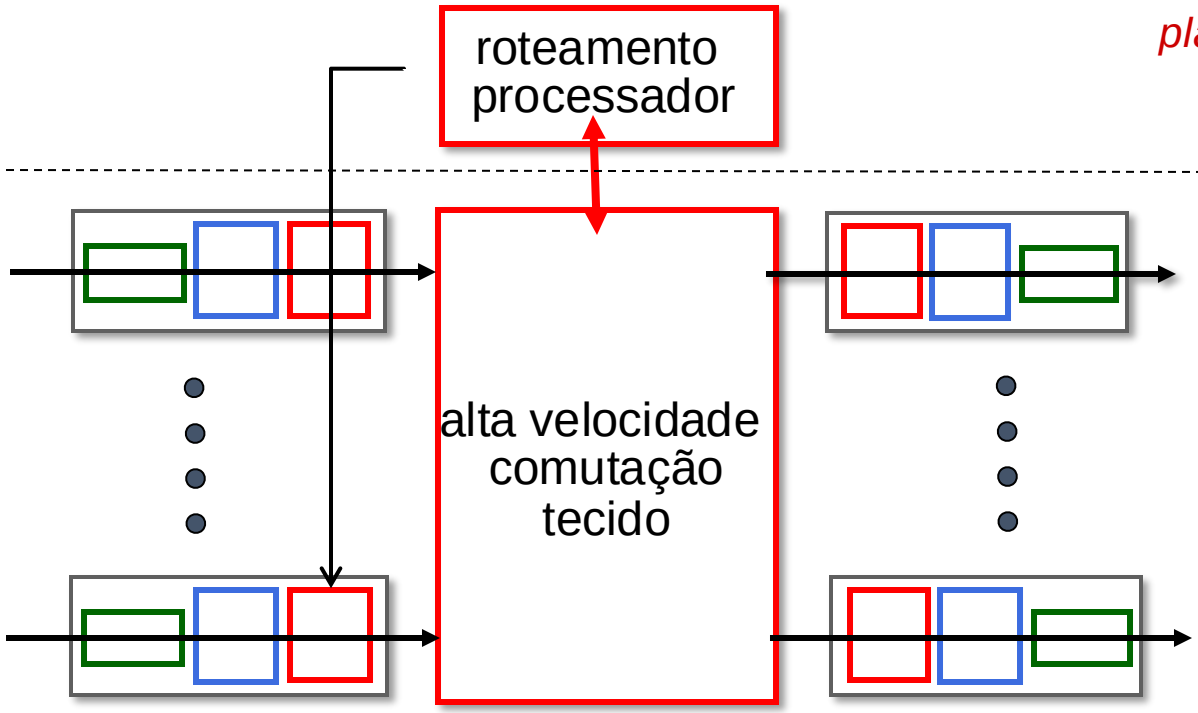
Camada de rede: Roteiro do "plano de dados"

- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo de Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereços de rede
  - IPv6

- Encaminhamento generalizado, SDN
  - Partida+ação
  - OpenFlow: match+action em ação
- Caixas intermediárias

Visão geral da arquitetura do roteador

visão de alto nível da arquitetura genérica do roteador:

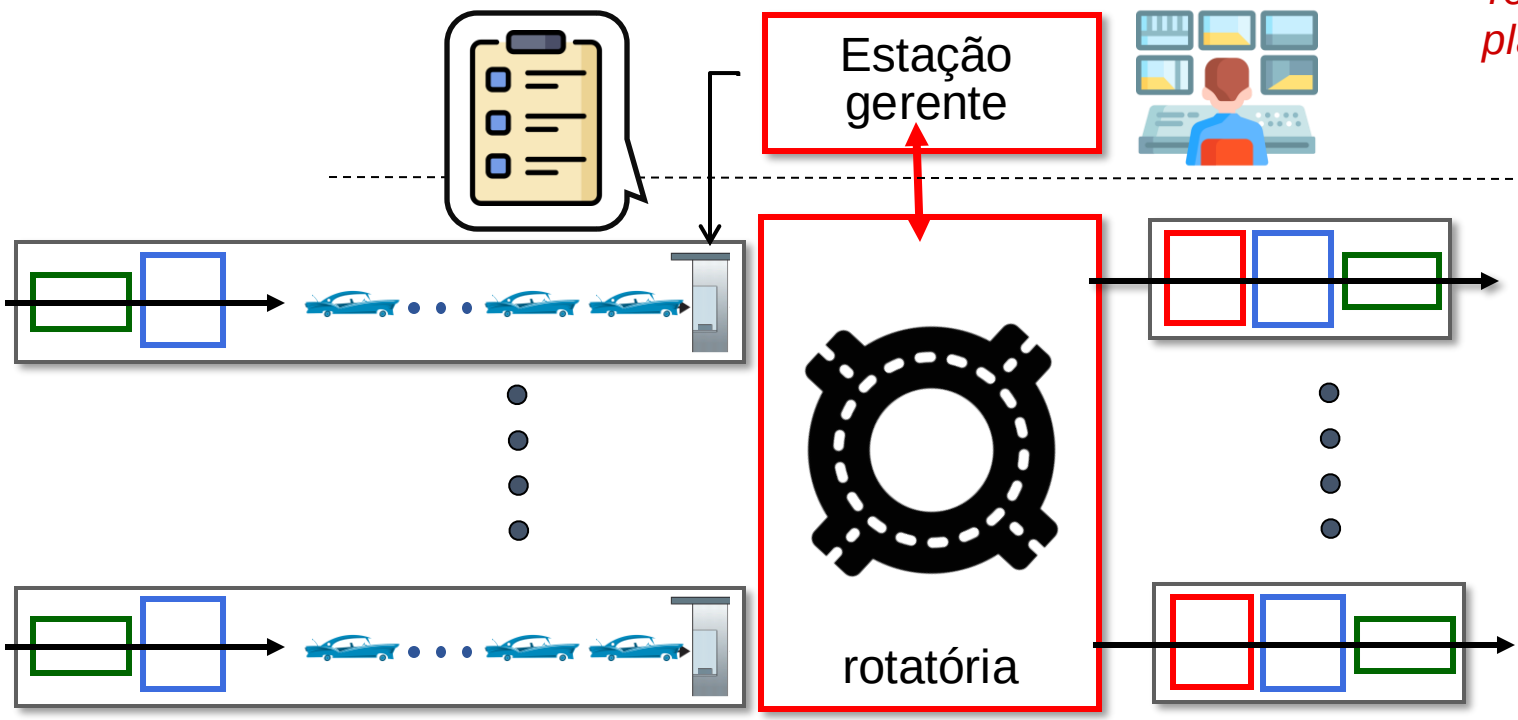


roteamento, gerenciamento plano de controle (software) opera em milissegundos prazo

o plano de dados de encaminhamento (hardware) opera em um intervalo de tempo de nanossegundos

Visão geral da arquitetura do roteador

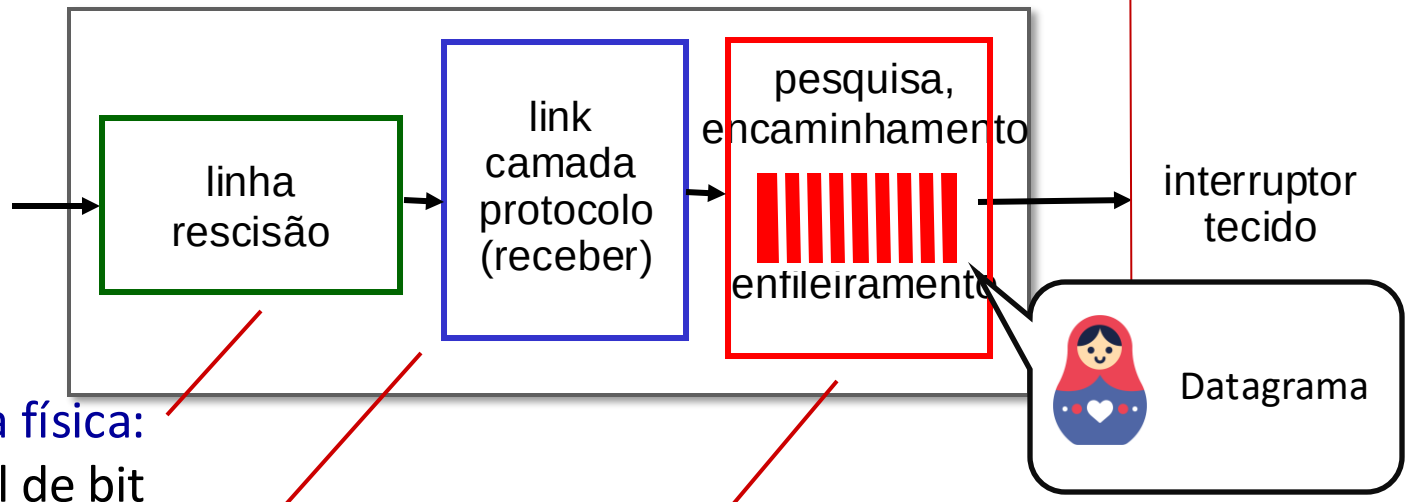
Visão analógica da arquitetura genérica do roteador:



roteamento, gerenciamento plano de controle (software) opera em milissegundos prazo

o plano de dados de encaminhamento (hardware) opera em um intervalo de tempo de nanossegundos

Funções da porta de entrada



camada física:
recepção em nível de bit

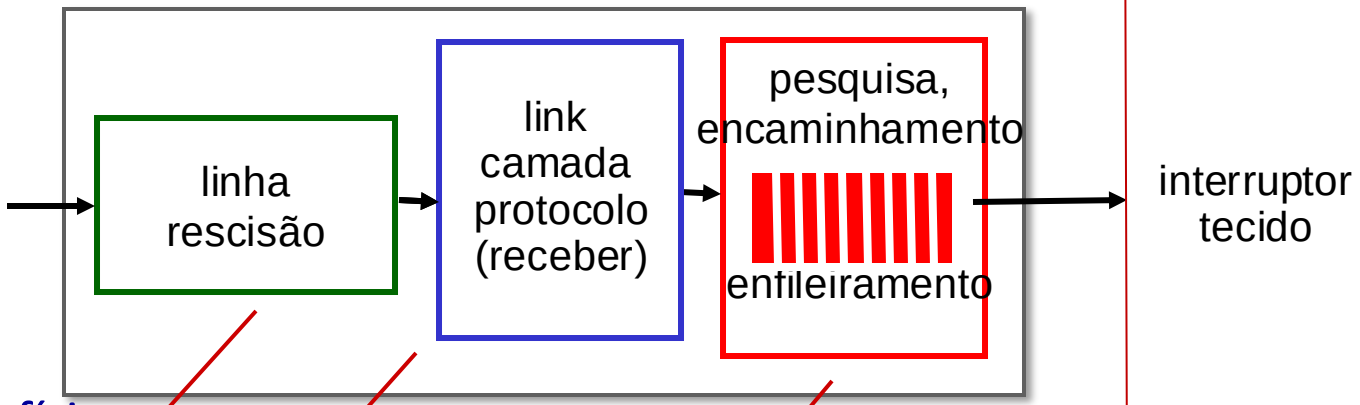
camada de link:
Por exemplo, Ethernet
(capítulo 6)

Moldura

comutação descentralizada:

- usando valores de campo de cabeçalho, porta de saída de pesquisa usando tabela de encaminhamento na memória da porta de entrada *("correspondência mais ação")*
- meta: concluir o processamento da porta de entrada na "velocidade da linha
- Enfileiramento da porta de entrada: se os datagramas chegarem mais rápido do que a taxa de encaminhamento para a estrutura do switch

Funções da porta de entrada



camada física:
recepção em nível de bit

camada de link:
Por exemplo, Ethernet
(capítulo 6)

comutação descentralizada:

- usando valores de campo de cabeçalho, porta de saída de pesquisa usando tabela de encaminhamento na memória da porta de entrada *("correspondência mais ação")*
- encaminhamento baseado em destino: encaminhar com base apenas no endereço IP de destino (tradicional)
- encaminhamento generalizado: encaminhar com base em qualquer conjunto de valores de campo de cabeçalho

Encaminhamento baseado em destino

forwarding table

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through 11001000 00010111 00010000 00000100	0
11001000 00010111 00010000 00000101 through 11001000 00010111 00010000 00000111	3
11001000 00010111 00011000 11111111	
11001000 00010111 00011001 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

P: mas o que acontece se os intervalos não se dividirem tão bem?

Correspondência do prefixo mais longo

correspondência de ~~prefixo mais longa~~

Ao procurar uma entrada na tabela de encaminhamento para um determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponda ao endereço de destino.

Intervalo de endereços de destino	Interface de link
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
caso contrário	3

exemplos:

11001000 00010111 00010110 10100001 Qual interface?

11001000 00010111 00011000 10101010 Qual interface?

Correspondência do prefixo mais longo

correspondência de prefixo mais longo

Ao procurar uma entrada na tabela de encaminhamento para um determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponda ao endereço de destino.

Intervalo de endereços de destino	Interface de link
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
caso contrário	3

Combine!

exemplos:

11001000 00010111 00010110 10100001 Qual interface?

11001000 00010111 00011000 10101010 Qual interface?

Correspondência do prefixo mais longo

correspondência de prefixo mais longo

Ao procurar uma entrada na tabela de encaminhamento para um determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponda ao endereço de destino.

Intervalo de endereços de destino	Interface de link
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
caso contrário	3

jogo!

exemplos:

11001000 00010111 00010110 10100001 Qual interface?

11001000 00010111 00011000 10101010 Qual interface?

Correspondência do prefixo mais longo

correspondência de prefixo mais longo

Ao procurar uma entrada na tabela de encaminhamento para um determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponda ao endereço de destino.

Faixa de endereços de destino	Interface de link
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
caso contrário	3

jogo!

exemplos:

11001000 00010111 00010110 10100001 Qual interface?

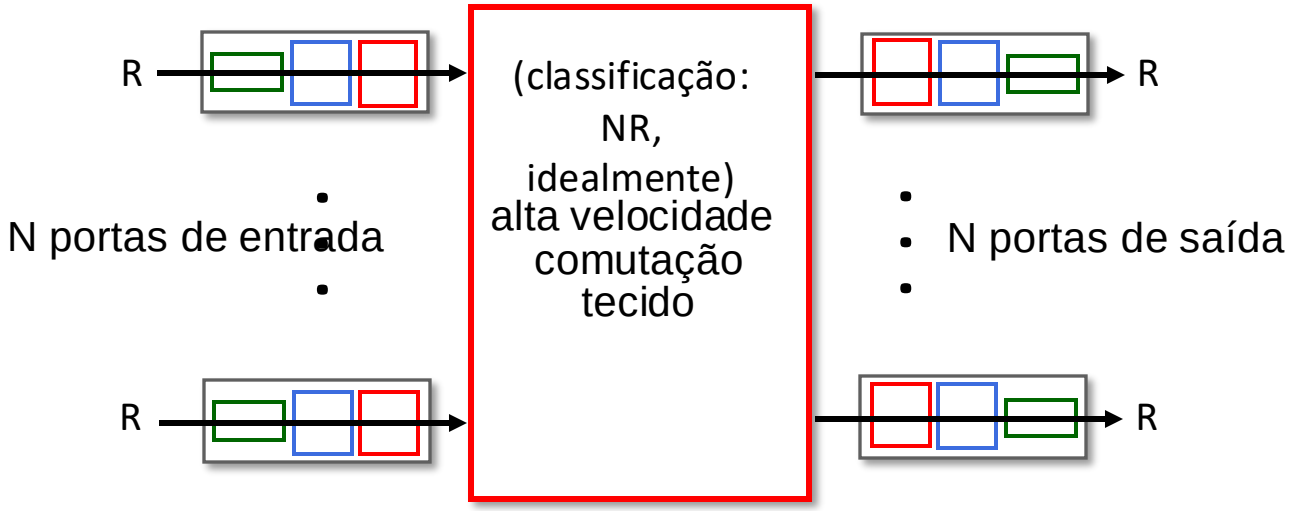
11001000 00010111 00011000 10101010 Qual interface?

Correspondência do prefixo mais longo

- Veremos *por que* a correspondência de prefixo mais longo é usada em breve, quando estudarmos o endereçamento
- Correspondência do prefixo mais longo: geralmente realizada com o uso de memórias ternárias endereçáveis por conteúdo (TCAMs)
  - *conteúdo endereçável:* endereço presente no TCAM: recupera o endereço em um ciclo de clock, independentemente do tamanho da tabela
  - Cisco Catalyst: ~1 milhão de entradas de tabela de roteamento no TCAM

Elementos de comutação

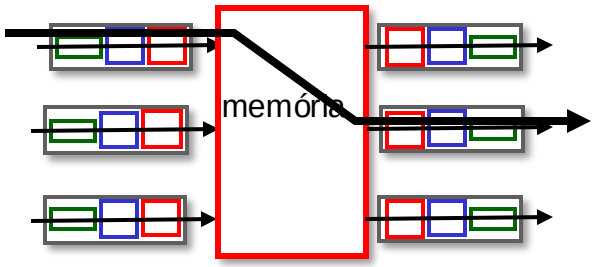
- transferir o pacote do link de entrada para o link de saída apropriado
- taxa de comutação: taxa na qual os pacotes podem ser transferidos das entradas para as saídas
  - geralmente medido como múltiplo da taxa de linha de entrada/saída
  - N entradas: taxa de comutação N vezes a taxa de linha desejável



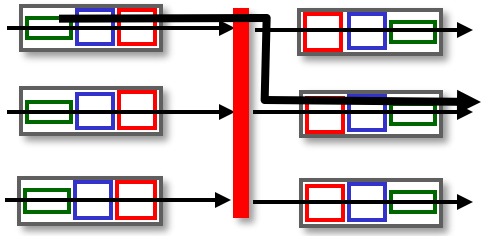
Elementos de comutação

- transferir o pacote do link de entrada para o link de saída apropriado
- taxa de comutação: taxa na qual os pacotes podem ser transferidos das entradas para as saídas
  - geralmente medido como múltiplo da taxa de linha de entrada/saída
  - N entradas: taxa de comutação N vezes a taxa de linha desejável

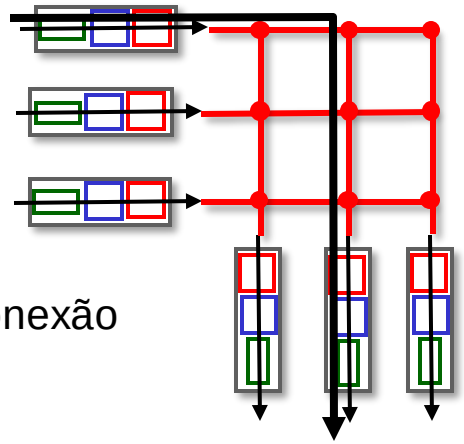
- três tipos principais de telas de comutação:



memória



barramento

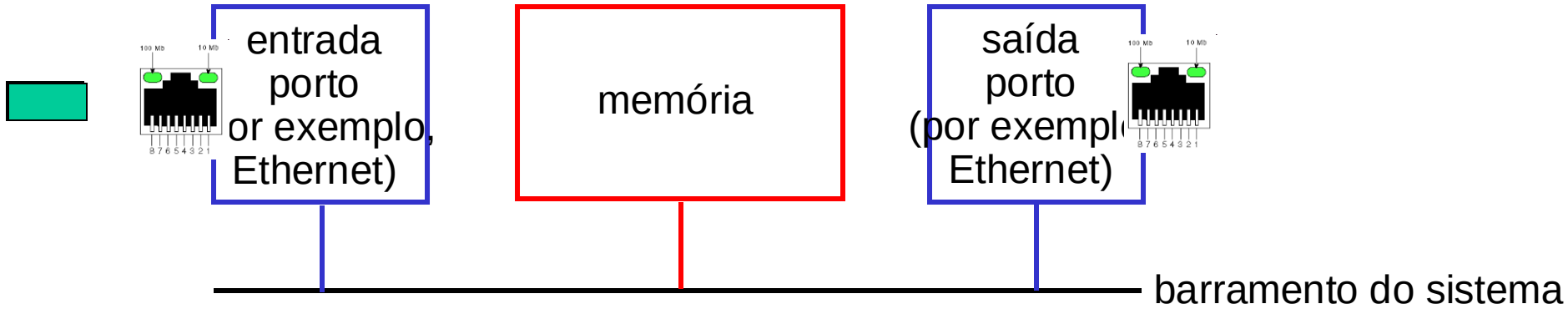


interconexão
rede

Comutação via memória

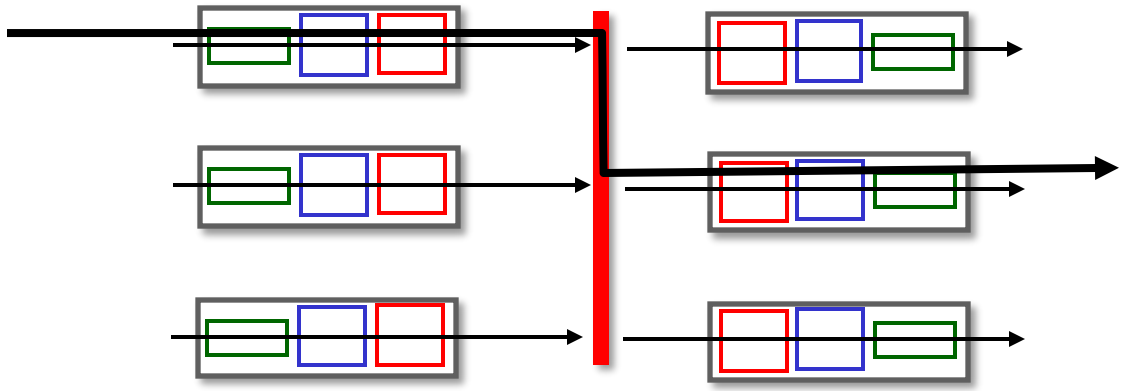
roteadores de primeira geração:

- computadores tradicionais com comutação sob controle direto da CPU
- pacote copiado para a memória do sistema
- velocidade limitada pela largura de banda da memória (2 passagens de barramento por datagrama)



Comutação por meio de um barramento

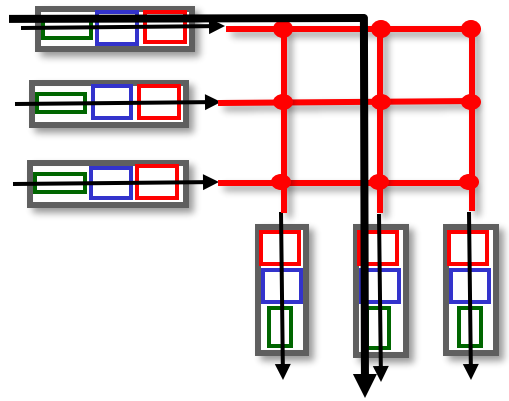
- datagrama da memória da porta de entrada para a memória da porta de saída por meio de um barramento compartilhado
- *contenção de barramento:* velocidade de comutação limitada pela largura de banda do barramento
- Barramento de 32 Gbps, Cisco 5600: velocidade suficiente para roteadores de acesso



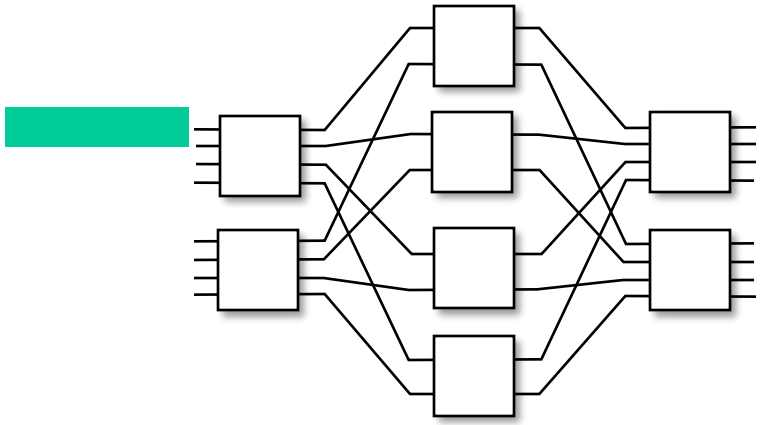
Comutação via rede de interconexão

- Crossbar, redes Clos, outras redes de interconexão inicialmente desenvolvidas para conectar processadores em multiprocessadores

- interruptor de vários estágios: interruptor *nxn* de vários estágios de interruptores menores
- explorando o paralelismo:
  - fragmentar o datagrama em células de comprimento fixo na entrada
  - células de switch através da malha, remontar o datagrama na saída



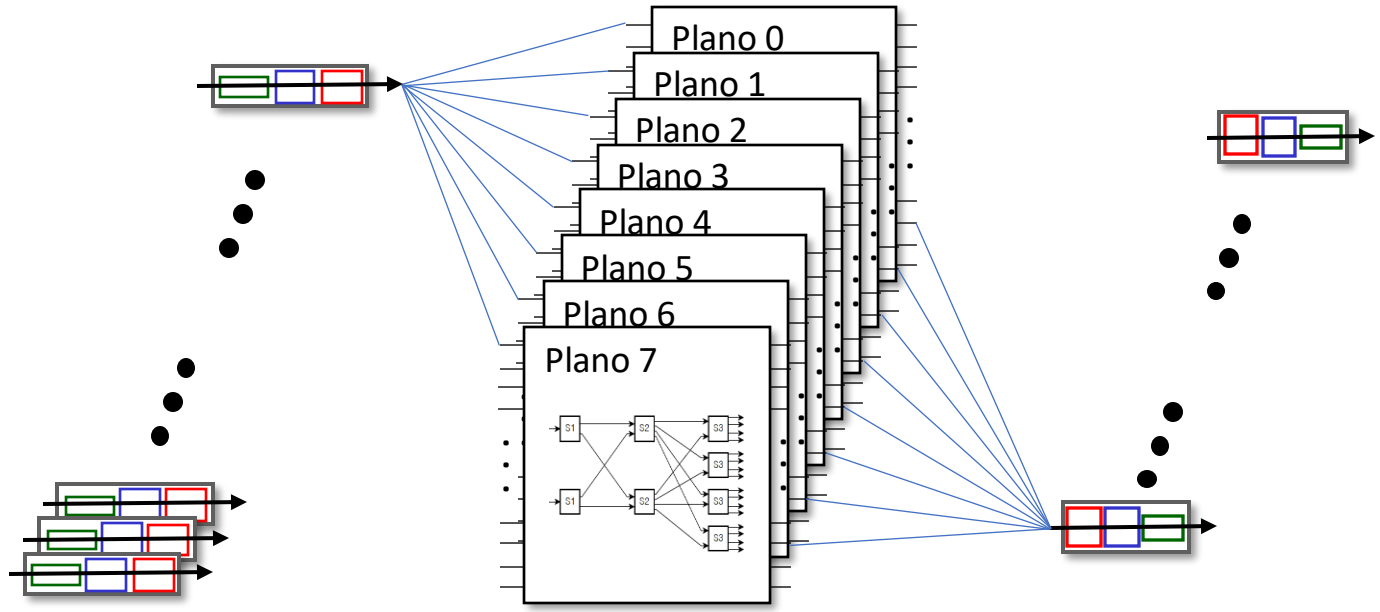
Barra transversal 3x3



Chave multiestágio 8x8
construído a partir de switches de tamanho menor

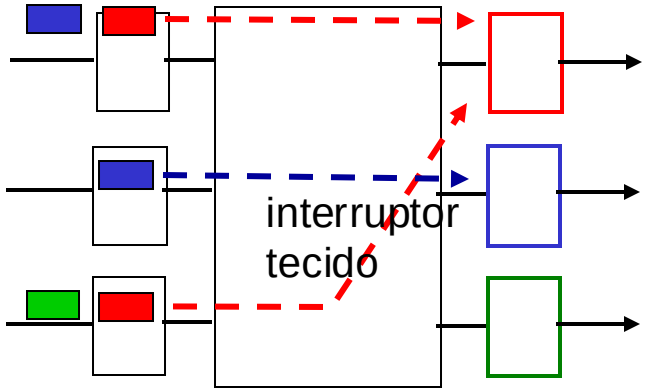
Comutação via rede de interconexão

- escalonamento, usando vários "planos" de comutação em paralelo:
  - aumento de velocidade, aumento de escala via paralelismo
- Roteador Cisco CRS:
  - unidade básica: 8 planos de comutação
  - cada plano: Rede de interconexão de 3 estágios
  - Capacidade de comutação de até 100 Tbps

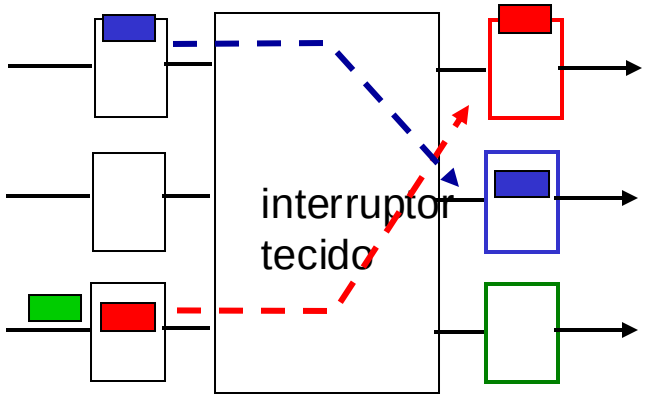


Enfileiramento de portas de entrada

- Se a malha do switch for mais lenta do que as portas de entrada combinadas, poderá ocorrer enfileiramento nas filas de entrada
  - atraso e perda na fila devido ao estouro do buffer de entrada!
- Bloqueio de cabeça de fila (HOL): o datagrama enfileirado na frente da fila impede que outros na fila avancem

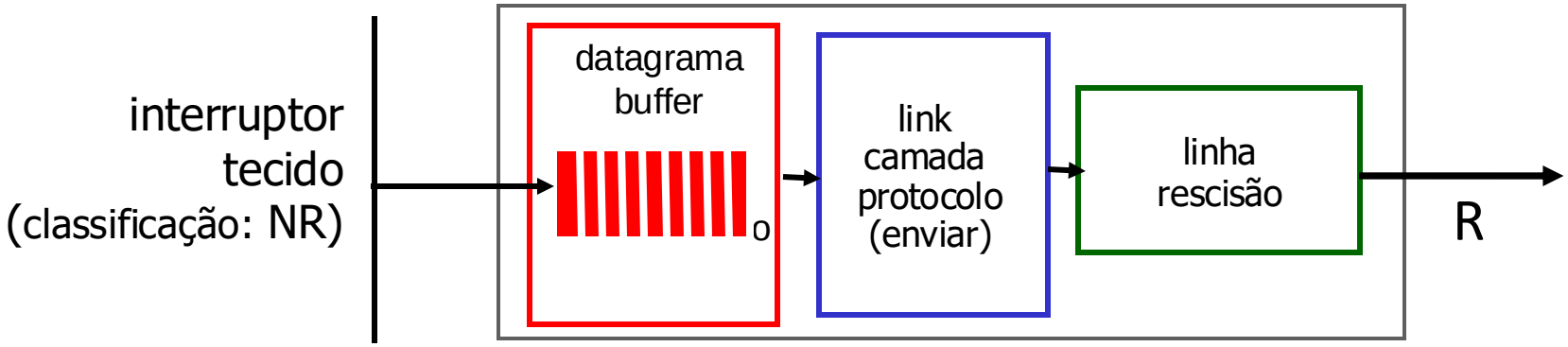


contenção da porta de saída: somente um datagrama vermelho pode ser transferido. o pacote vermelho inferior é *bloqueado*



um pacote de tempo depois: o pacote verde sofre bloqueio HOL

Enfileiramento de portas de saída



Este é um slide muito importante

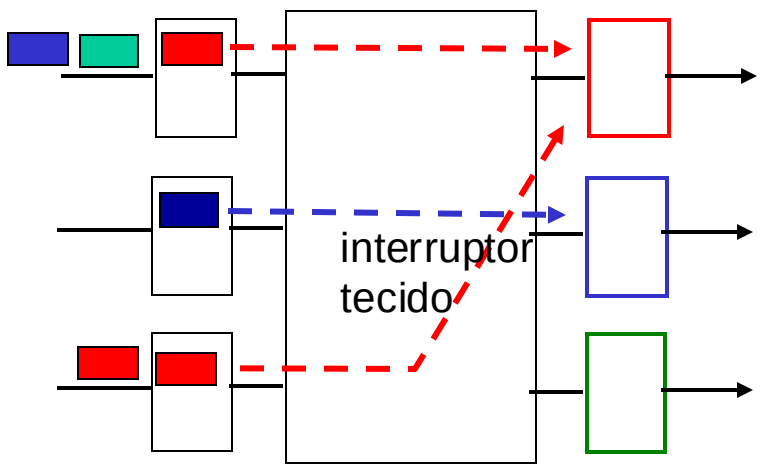
- *O buffer* é necessário quando os datagramas chegam da malha mais rapidamente do que a taxa de transmissão do link. *Política de descarte:* quais datagramas devem ser descartados se não houver buffers livres?

  → Os datagramas podem ser perdidos devido a congestionamento, falta de buffers

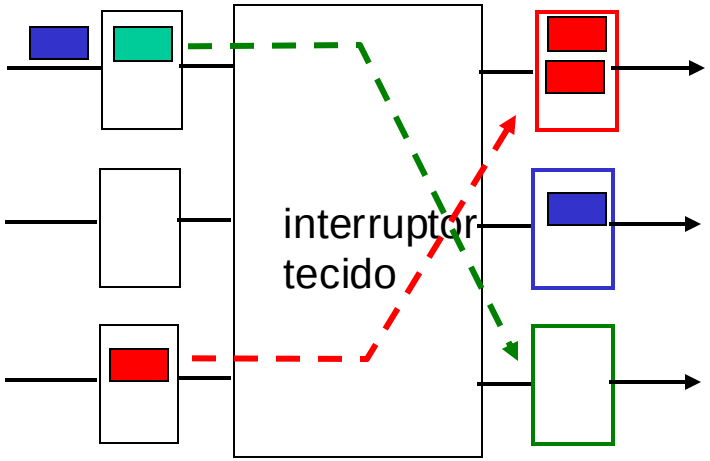
- *A disciplina de agendamento* escolhe entre datagramas enfileirados para transmissão

  → Programação de prioridades - quem obtém o melhor desempenho, neutralidade da rede

Enfileiramento de portas de saída



em *t*, pacotes mais da entrada para a saída



um pacote de tempo depois

- buffering quando a taxa de chegada via chave excede a velocidade da linha de saída
- *enfileiramento (atraso) e perda devido ao estouro do buffer da porta de saída!*

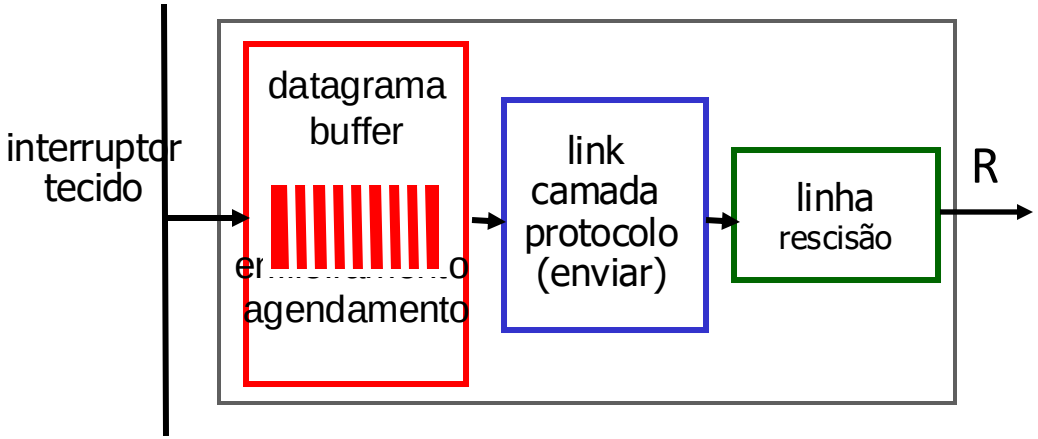
Qual é a quantidade de buffer?

- Regra geral da RFC 3439: buffer médio igual ao RTT "típico" (digamos 250 mseg) vezes a capacidade do link C
  - Por exemplo, C = link de 10 Gbps: buffer de 2,5 Gbit
- Recomendação mais recente: com *N* fluxos, buffering igual a

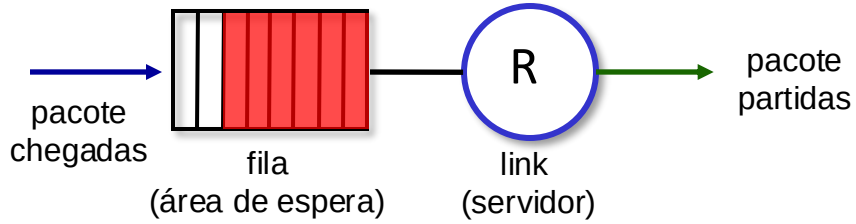
$$\frac{RTT \cdot C}{\sqrt{N}}$$

- mas o excesso de buffer pode aumentar os atrasos (principalmente em roteadores domésticos)
  - RTTs longos: desempenho ruim para aplicativos em tempo real, resposta lenta do TCP
  - lembre-se do controle de congestionamento baseado em atraso: "manter o link de gargalo cheio o suficiente (ocupado), mas não mais cheio"

Gerenciamento de buffer



Abstração: fila



gerenciamento de buffer:

- drop: qual pacote adicionar, descartar quando os buffers estiverem cheios
  - drop tail: drop do pacote que está chegando
  - prioridade: descartar/remover com base na prioridade
- marcação: quais pacotes devem ser marcados para sinalizar congestionamento (ECN, RED)

Programação de pacotes: FCFS

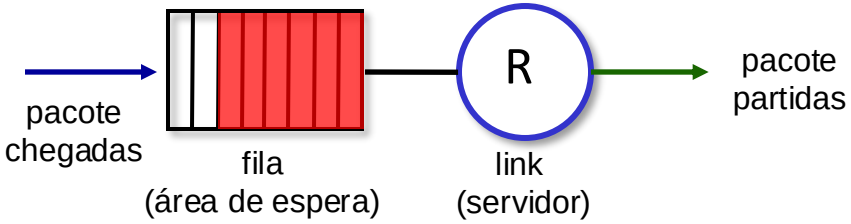
programação de pacotes: decidir qual pacote enviar em seguida no link

- primeiro a chegar, primeiro a ser servido
- prioridade
- round robin
- enfileiramento justo ponderado

FCFS: pacotes transmitidos na ordem de chegada à porta de saída

- também conhecido como: Primeiro a entrar, primeiro a sair (FIFO)
- exemplos do mundo real?

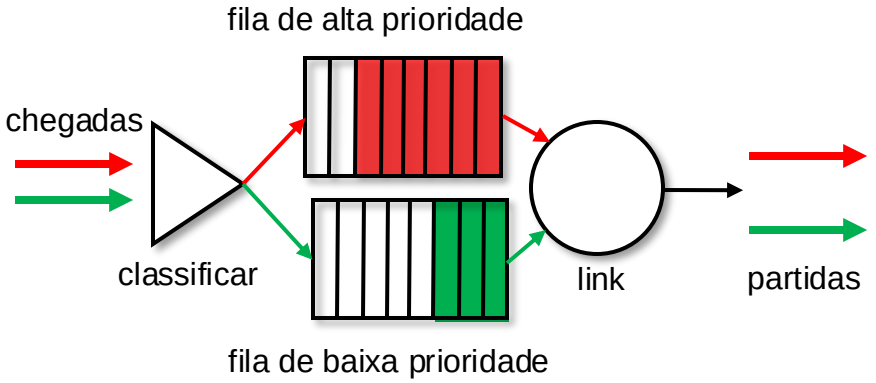
Abstração: fila

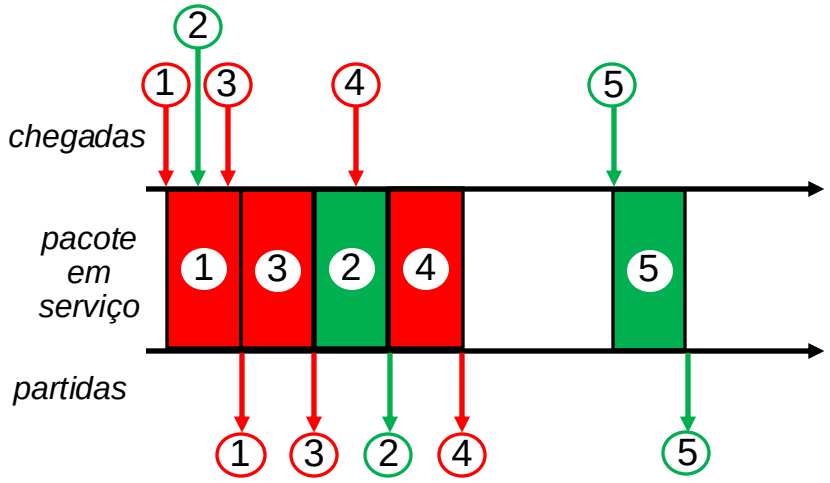


Políticas de agendamento: prioridade

Agendamento de prioridades:

- tráfego de chegada classificado e enfileirado por classe
  - qualquer campo de cabeçalho pode ser usado para classificação
- enviar pacote da fila de prioridade mais alta que tem pacotes em buffer
  - FCFS dentro da classe de prioridade

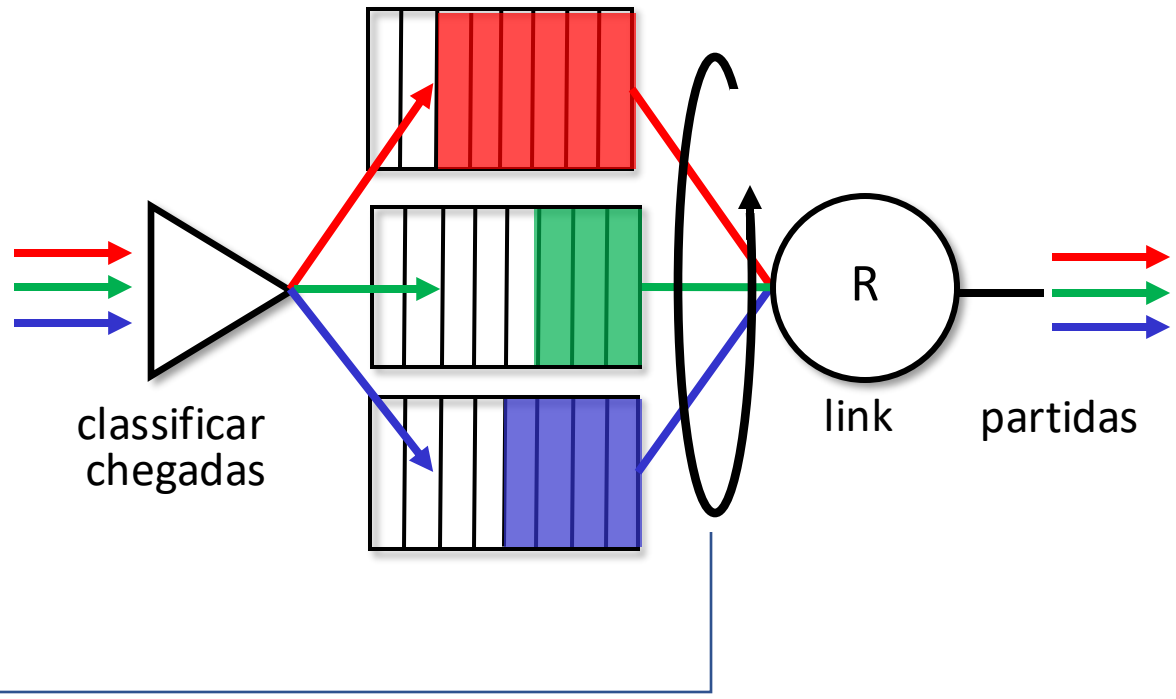




Políticas de agendamento: round robin

Programação de Round Robin (RR):

- tráfego de chegada classificado e enfileirado por classe
  - qualquer campo de cabeçalho pode ser usado para classificação
- O servidor varre ciclicamente e repetidamente as filas de classe, enviando um pacote completo de cada classe (se disponível) por vez



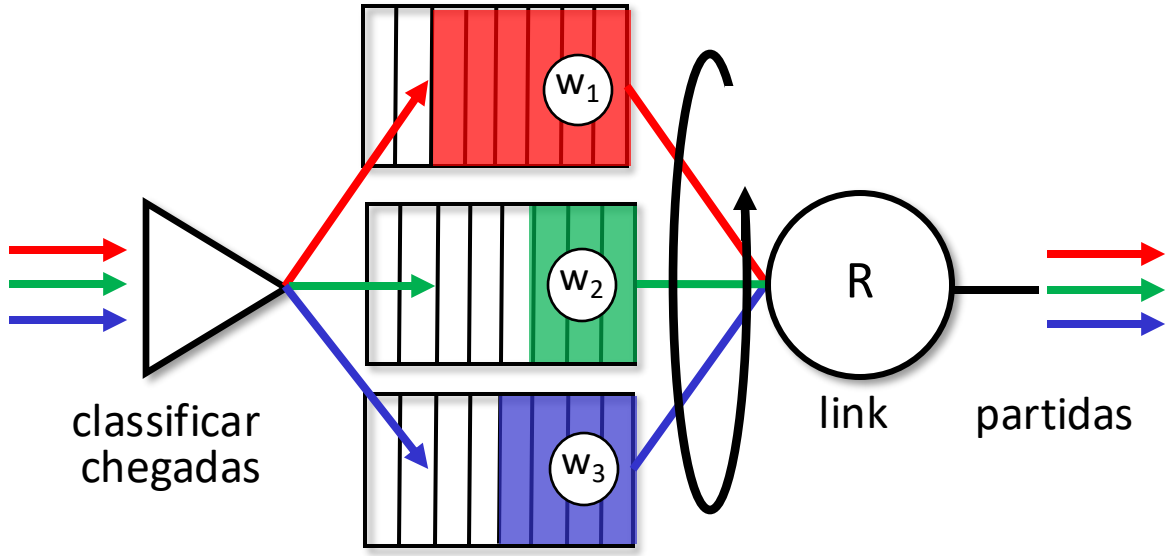
Políticas de agendamento: fila justa ponderada

Enfileiramento justo ponderado (WFQ):

- Round Robin generalizado
- Cada classe, *i,* tem um peso, w_i , e recebe uma quantidade ponderada de serviço em cada ciclo:

$$\frac{w_i}{\sum \omega_{jj}}$$

- garantia de largura de banda mínima (por classe de tráfego)



Barra lateral: Neutralidade da rede

O que é neutralidade da rede?

- *técnica:* como um ISP deve compartilhar/alocar seus recursos
  - agendamento de pacotes e gerenciamento de buffer são os *mecanismos*
- princípios *sociais e econômicos*
  - proteção da liberdade de expressão
  - incentivo à inovação e à concorrência
- regras e políticas *legais* aplicadas

Diferentes países têm diferentes "visões" sobre a neutralidade da rede

Barra lateral: Neutralidade da rede

Ordem da FCC dos EUA de 2015 *sobre a proteção e a promoção de uma Internet aberta:* três regras "claras e de linhas claras":

- sem bloqueio ... "não deve bloquear conteúdo, aplicativos, serviços ou dispositivos não prejudiciais legais, sujeito a um gerenciamento de rede razoável".
- sem limitação ... "não deve prejudicar ou degradar o tráfego legal da Internet com base no conteúdo, aplicativo ou serviço da Internet ou no uso de um dispositivo não prejudicial, sujeito a um gerenciamento de rede razoável".
- nenhuma priorização paga. ... "não se envolverá em priorização paga"

ISP: serviço de telecomunicações ou de informações?

Um ISP é um provedor de "serviço de telecomunicações" ou de "serviço de informações"?

- a resposta *é realmente* importante do ponto de vista regulatório!

Lei de Telecomunicações dos EUA de 1934 e 1996:

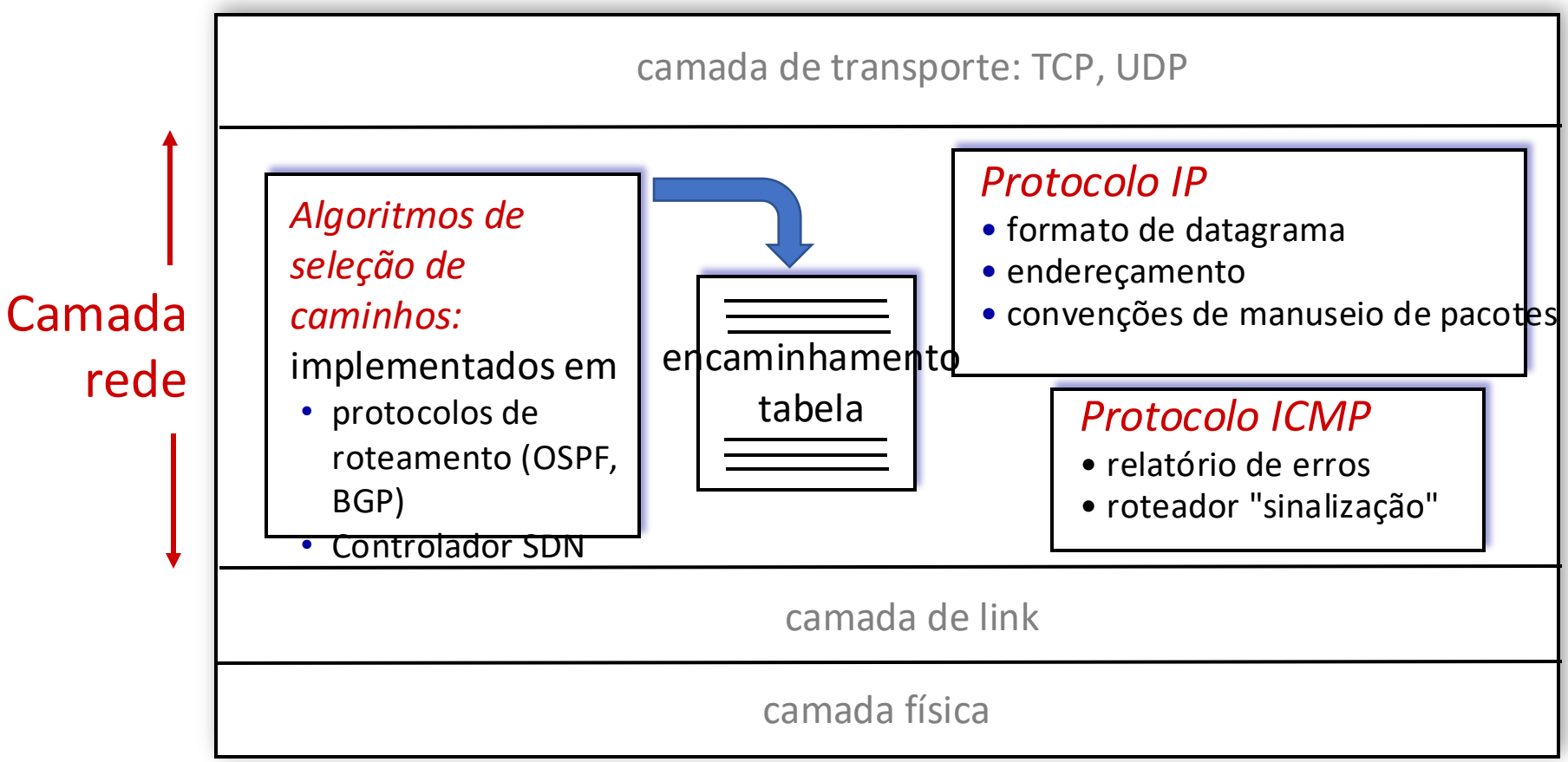
- *Título II:* impõe "deveres de transportadora comum" aos *serviços de telecomunicações*: tarifas razoáveis, não discriminação e *exige regulamentação*
- *Título I:* aplica-se a *serviços de informação:*
  - não há taxas para transportadoras comuns (*não regulamentadas*)
  - mas concede à FCC autoridade "... conforme necessário para a execução de suas funções "[4]

Camada de rede: Roteiro do "plano de dados"

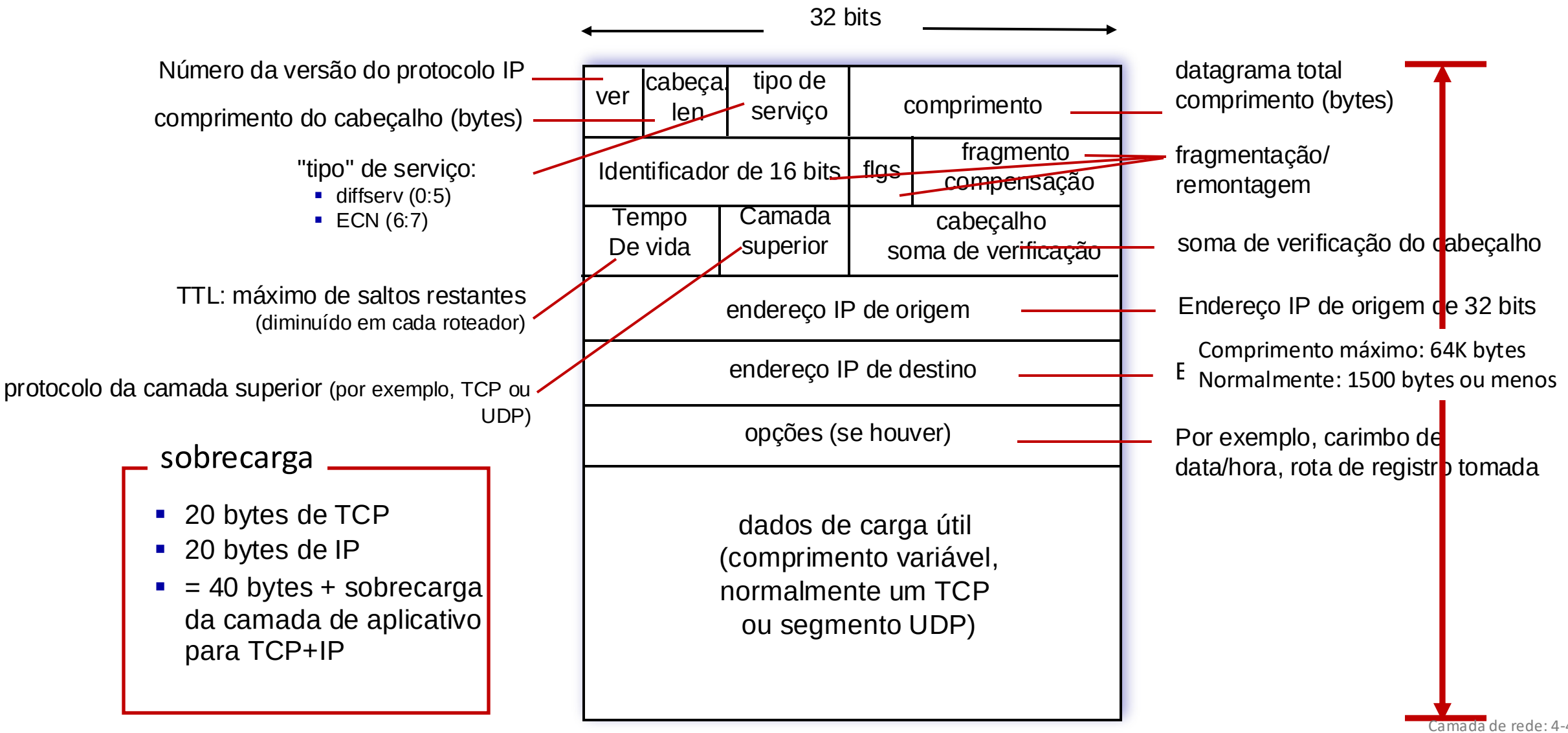
- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo de Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereços de rede
  - IPv6
- Encaminhamento generalizado, SDN
  - correspondência+ação
  - OpenFlow: match+action em ação
- Caixas intermediárias

Camada de rede: Internet

funções da camada de rede do host e do roteador:



Formato do datagrama IP

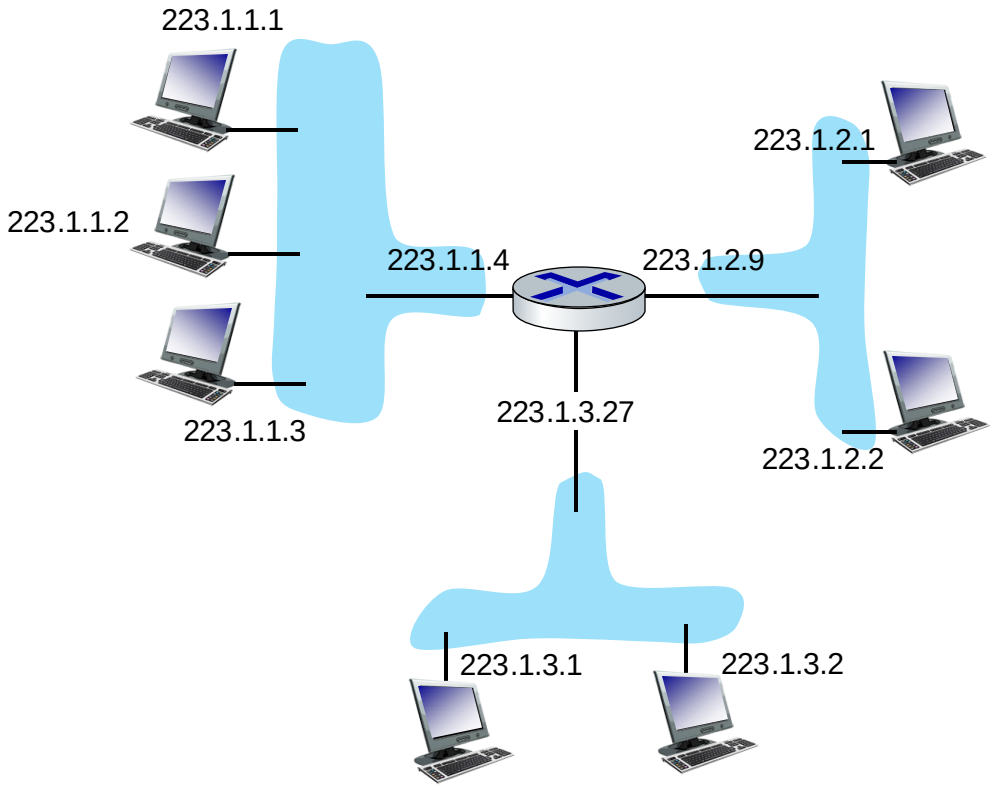


sobrecarga

- 20 bytes de TCP
- 20 bytes de IP
- = 40 bytes + sobrecarga da camada de aplicativo para TCP+IP

Endereçamento IP: introdução

- Endereço IP: Identificador de 32 bits associado a cada interface de host ou roteador
- interface: conexão entre o host/roteador e o link físico
  - Os roteadores geralmente têm várias interfaces
  - o host normalmente tem uma ou duas interfaces (por exemplo, Ethernet com fio, 802.11 sem fio)



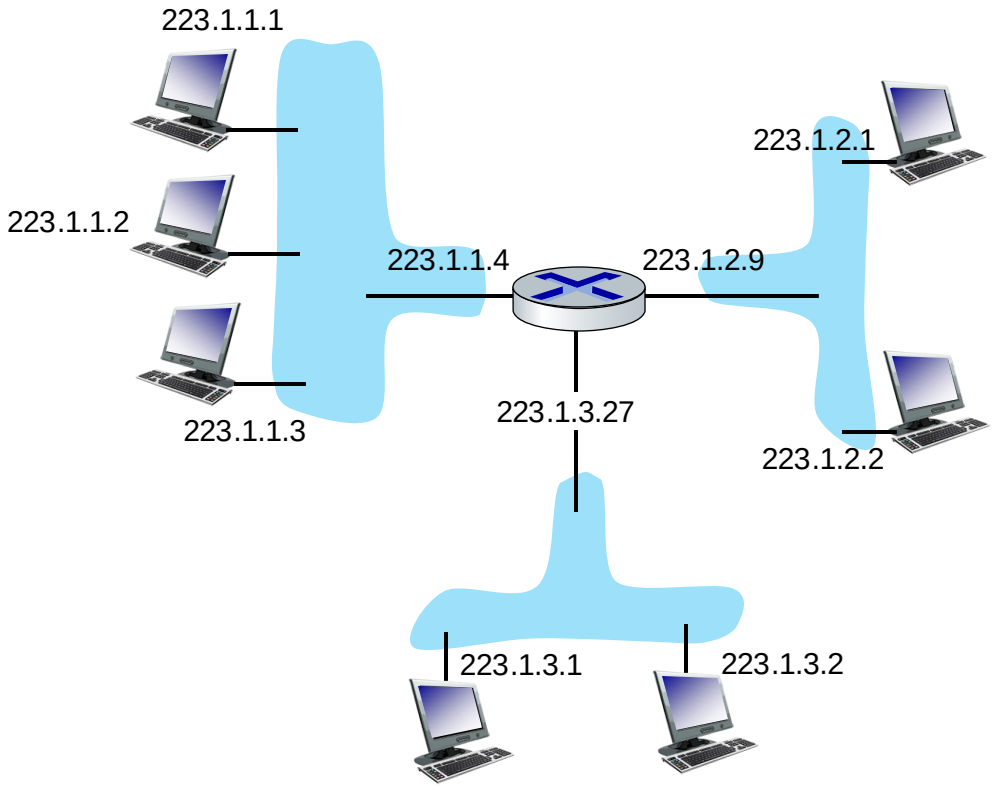
notação de endereço IP decimal com pontos:

223.1.1.1 = 11011111 00000001 00000001 00000001

223 1 1 1

Endereçamento IP: introdução

- **Endereço IP:** Identificador de 32 bits associado a cada interface de host ou roteador
- **interface:** conexão entre o host/roteador e o link físico
  - Os roteadores geralmente têm várias interfaces
  - o host normalmente tem uma ou duas interfaces (por exemplo, Ethernet com fio, 802.11 sem fio)



notação de endereço IP decimal com pontos:

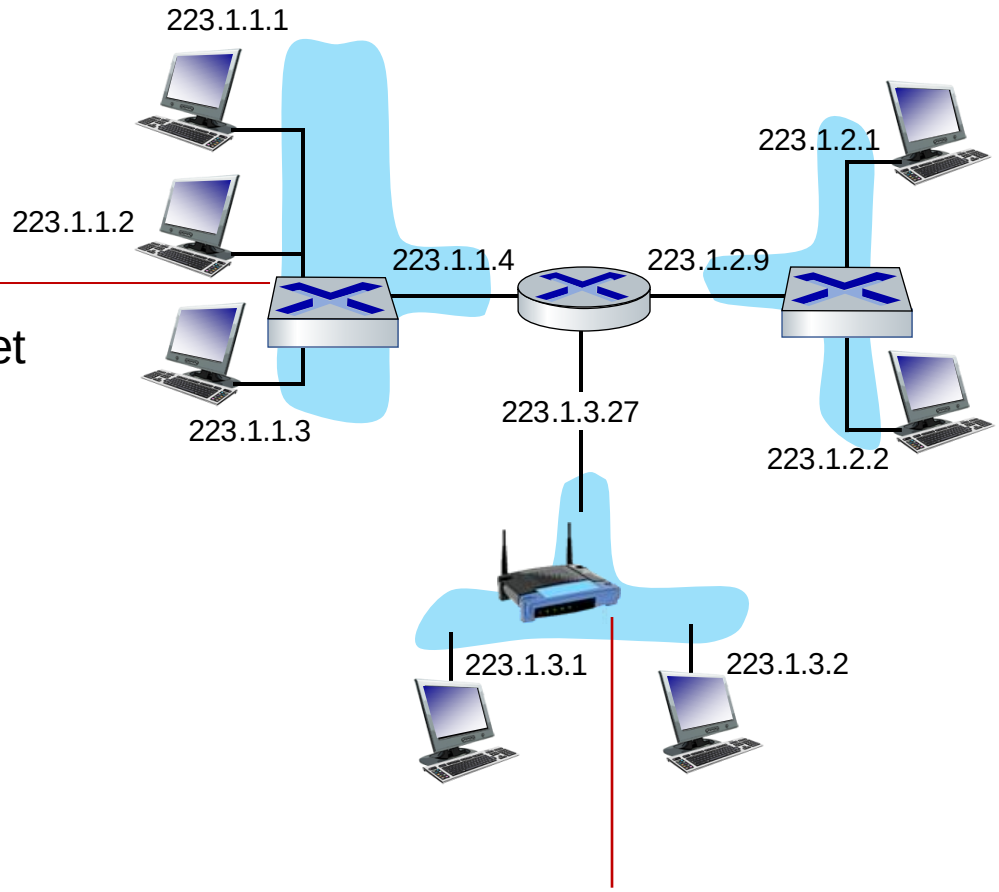
223.1.1.1 = 11011111 00000001 00000001 00000001

223 1 1 1

Endereçamento IP: introdução

P: Como as interfaces são realmente conectadas?

R: aprenderemos sobre isso nos capítulos 6 e 7



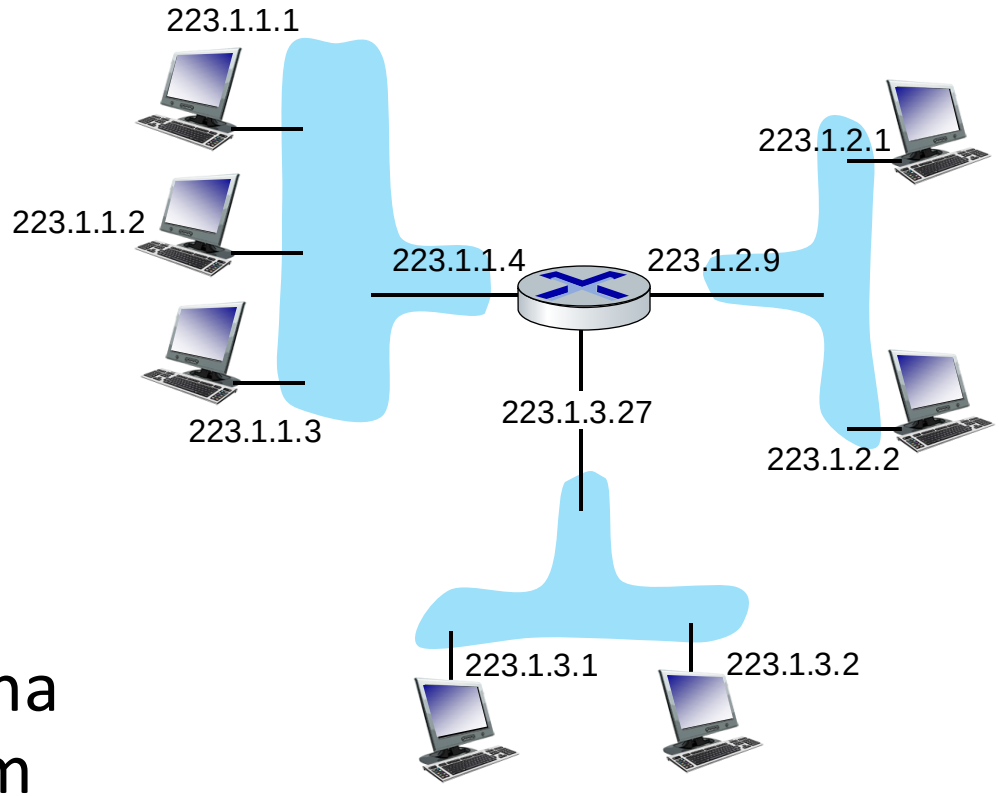
R: com fio Interfaces Ethernet conectadas por switches Ethernet

A: interfaces WiFi sem fio conectadas por uma estação base WiFi

Por enquanto: não é necessário se preocupar com a forma como uma interface está conectada a outra (sem roteador intermediário)

Sub-redes

- *O que é uma sub-rede?*
  - interfaces de dispositivos que podem alcançar fisicamente uns aos outros sem passar por um roteador intermediário
- Os endereços IP têm estrutura:
  - parte da sub-rede: os dispositivos na mesma sub-rede têm bits de ordem alta comuns
  - Parte do host: bits de baixa ordem restantes

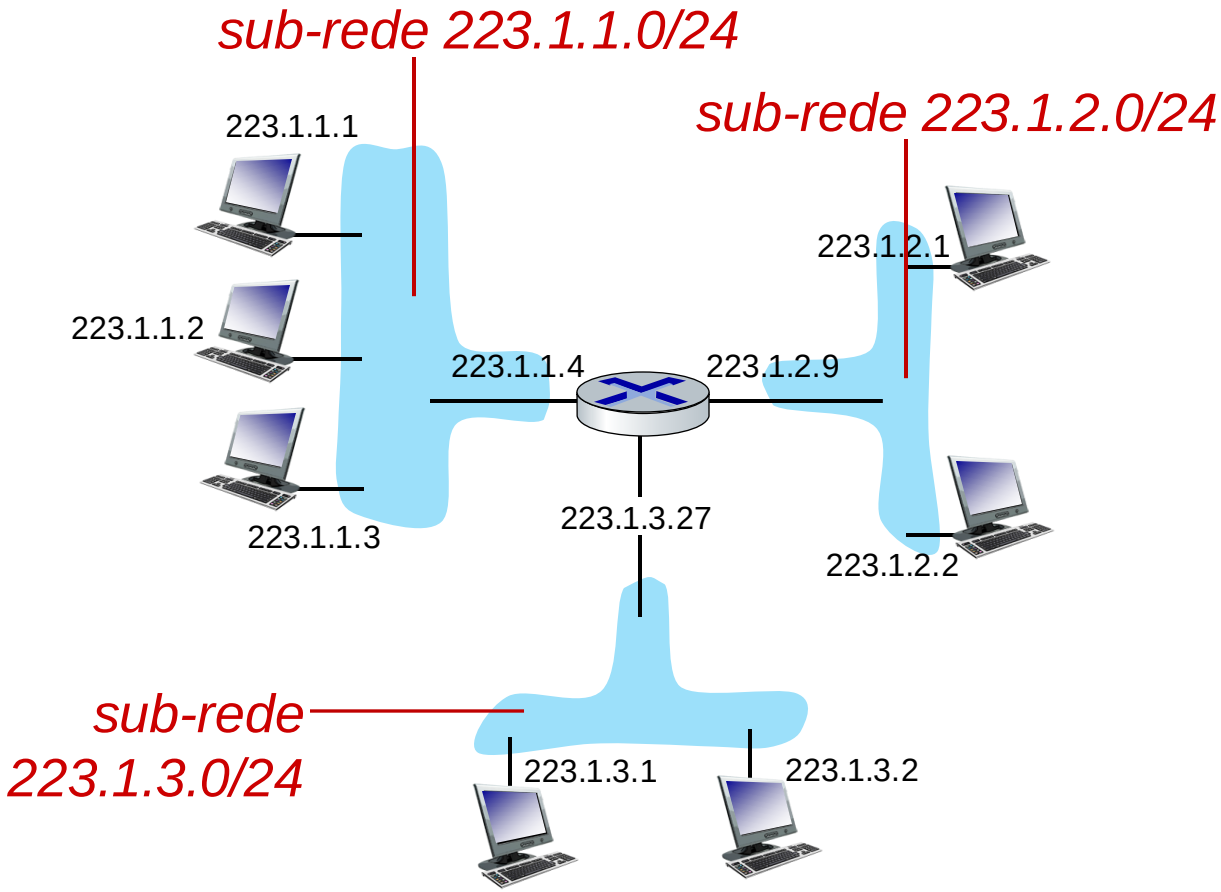


rede composta por 3 sub-redes

Sub-redes

Receita para definir sub-redes:

- separar cada interface de seu host ou roteador, criando "ilhas" de redes isoladas
- cada rede isolada é chamada de *sub-rede*

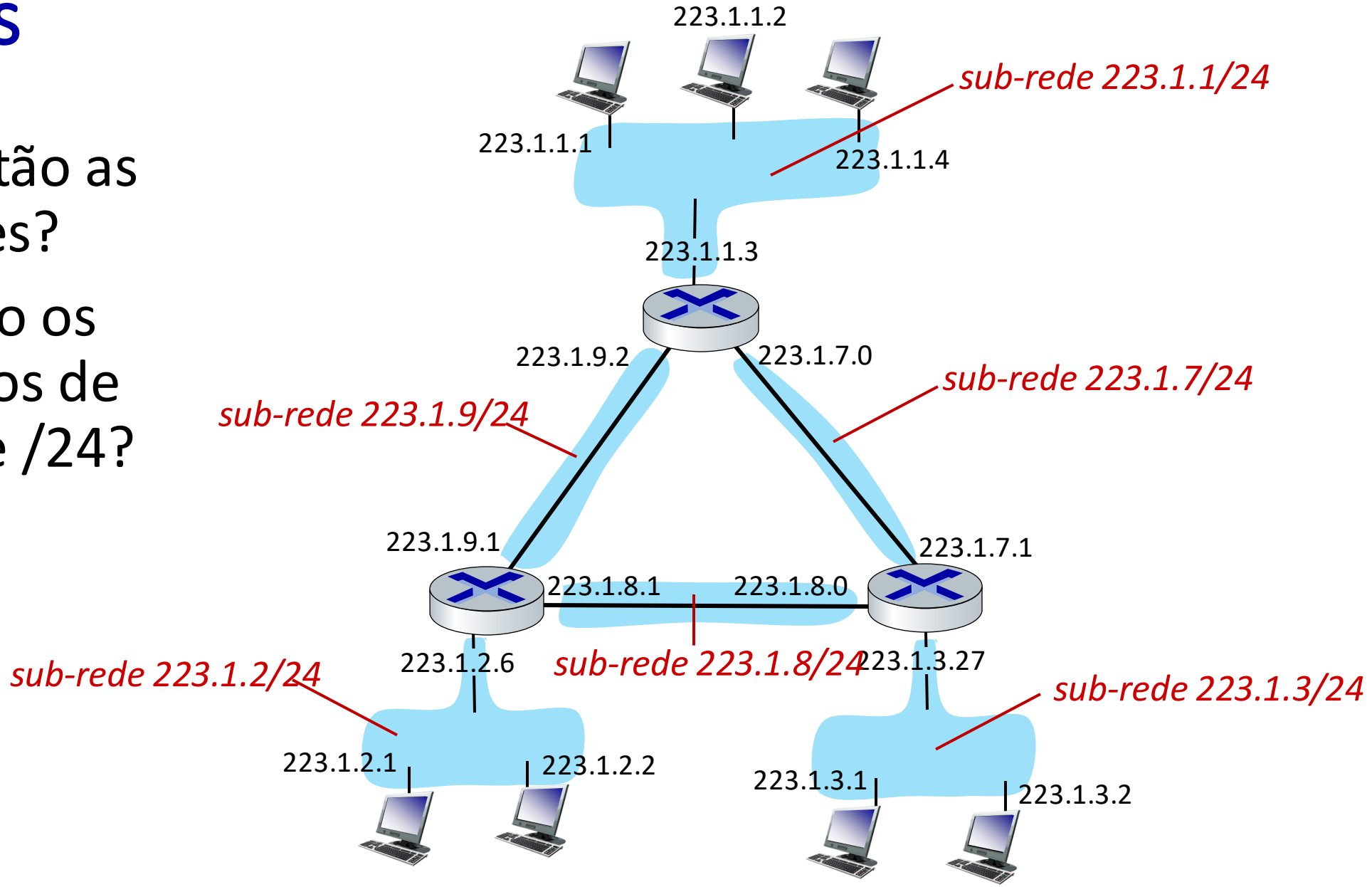


máscara de sub-rede: /24

(24 bits de ordem superior: parte da sub-rede do endereço IP)

Sub-redes

- Onde estão as sub-redes?
- Quais são os endereços de sub-rede /24?



Endereçamento IP: CIDR

CIDR: Classless InterDomain Routing (pronuncia-se "cider")

- parte da sub-rede do endereço de comprimento arbitrário
- formato de endereço: a.b.c.d/x, em que x é o número de bits na parte de sub-rede do endereço

11001000 00010111 00010000 00000000

200.23.16.0/23

Endereços IP: como obter um?

Na verdade, são duas perguntas:

1. P: Como um *host* obtém o endereço IP em sua rede (parte do host do endereço)?
2. P: Como uma *rede* obtém um endereço IP para si mesma (parte de rede do endereço)?

Como *o host* obtém o endereço IP?

- codificado pelo administrador do sistema no arquivo de configuração (por exemplo, /etc/rc.config no UNIX)
- DHCP: Dynamic Host Configuration Protocol (protocolo de configuração dinâmica de host): obtém dinamicamente o endereço de um servidor
  - "plug-and-play"

DHCP: Protocolo de configuração dinâmica de host

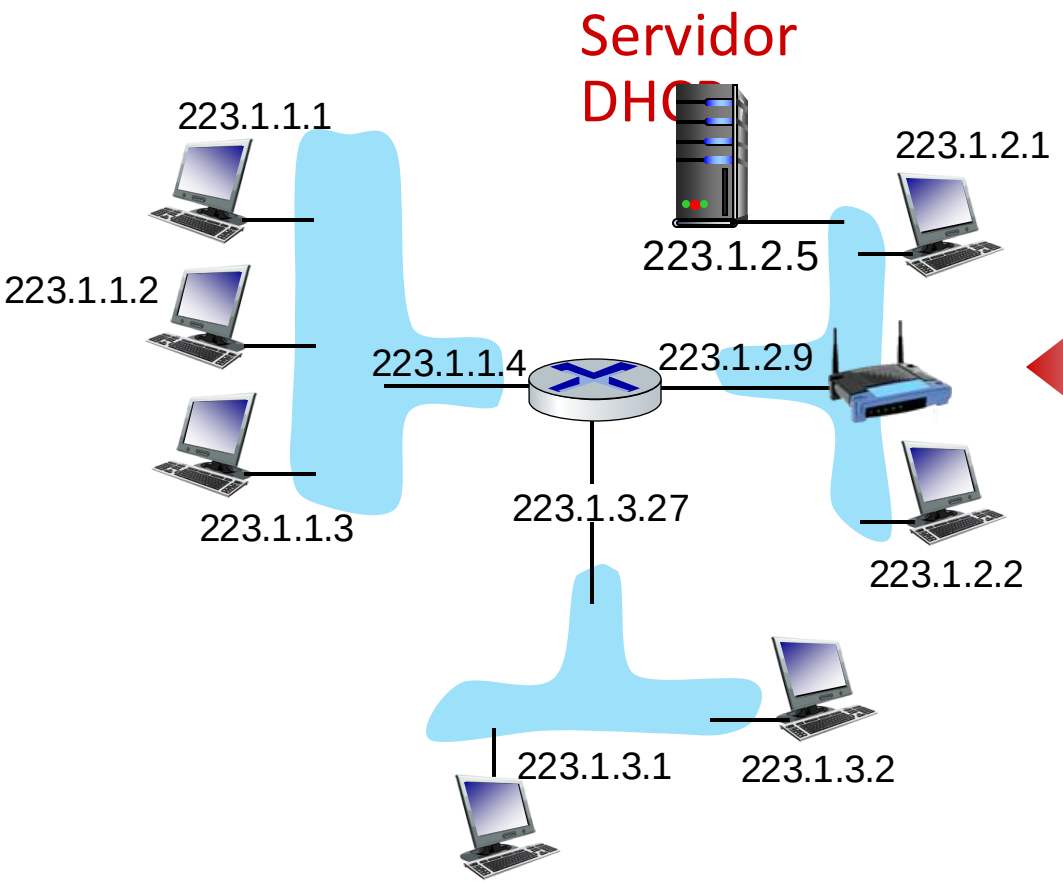
Objetivo: o host obtém *dinamicamente* o endereço IP do servidor de rede quando "entra" na rede

- pode renovar sua concessão no endereço em uso
- permite a reutilização de endereços (só mantém o endereço enquanto estiver conectado/ligado)
- suporte para usuários móveis que entram/saem da rede

Visão geral do DHCP:

- o host transmite a mensagem de descoberta de DHCP [opcional]
- O servidor DHCP responde com a mensagem de oferta DHCP [opcional]
- host solicita endereço IP: Mensagem de solicitação de DHCP
- O servidor DHCP envia o endereço: DHCP ack msg

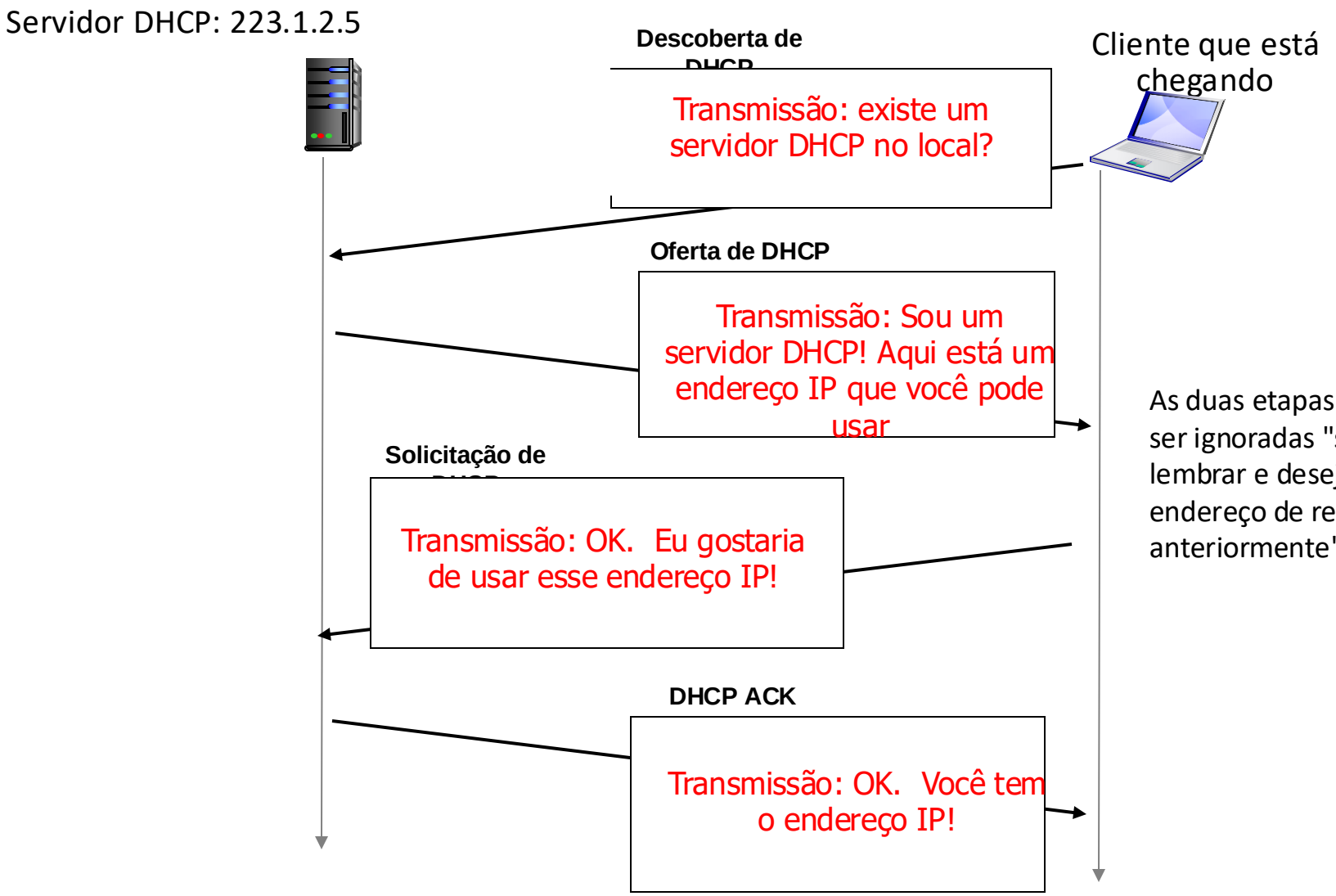
Cenário cliente-servidor DHCP



Normalmente, o servidor DHCP estará localizado no roteador, atendendo a todas as sub-redes às quais o roteador está conectado

chegando às necessidades do cliente DHCP endereço nesta rede

Cenário cliente-servidor DHCP



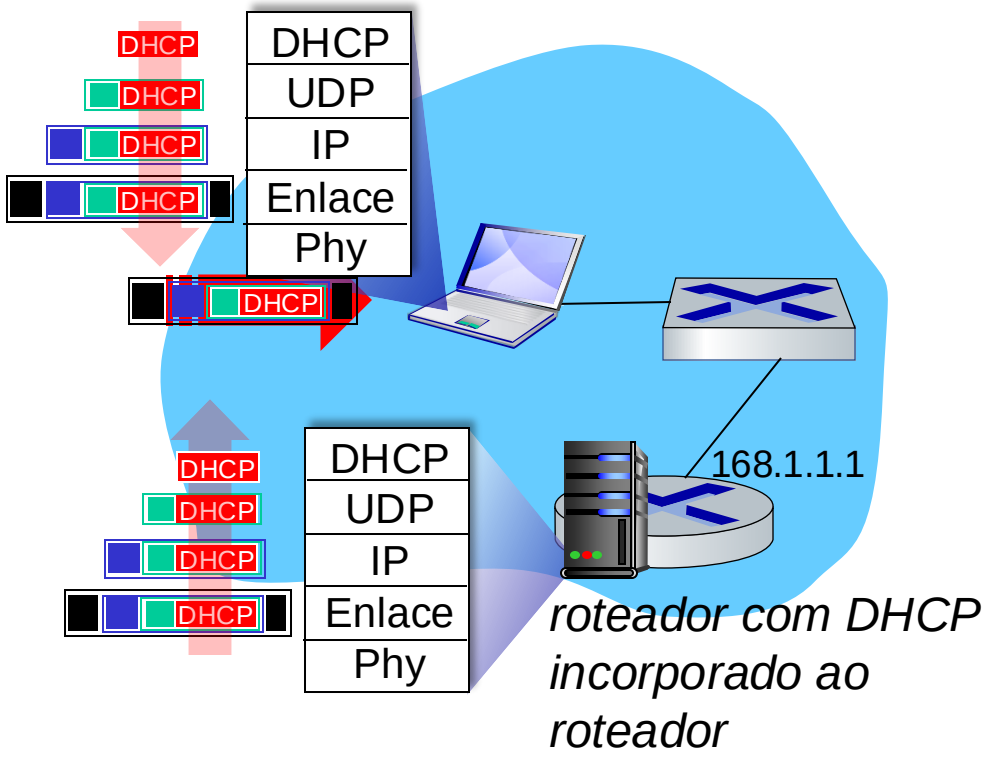
As duas etapas acima podem ser ignoradas "se um cliente se lembrar e desejar reutilizar um endereço de rede alocado anteriormente" [RFC 2131]

DHCP: mais do que endereços IP

O DHCP pode retornar mais do que apenas o endereço IP alocado na sub-rede:

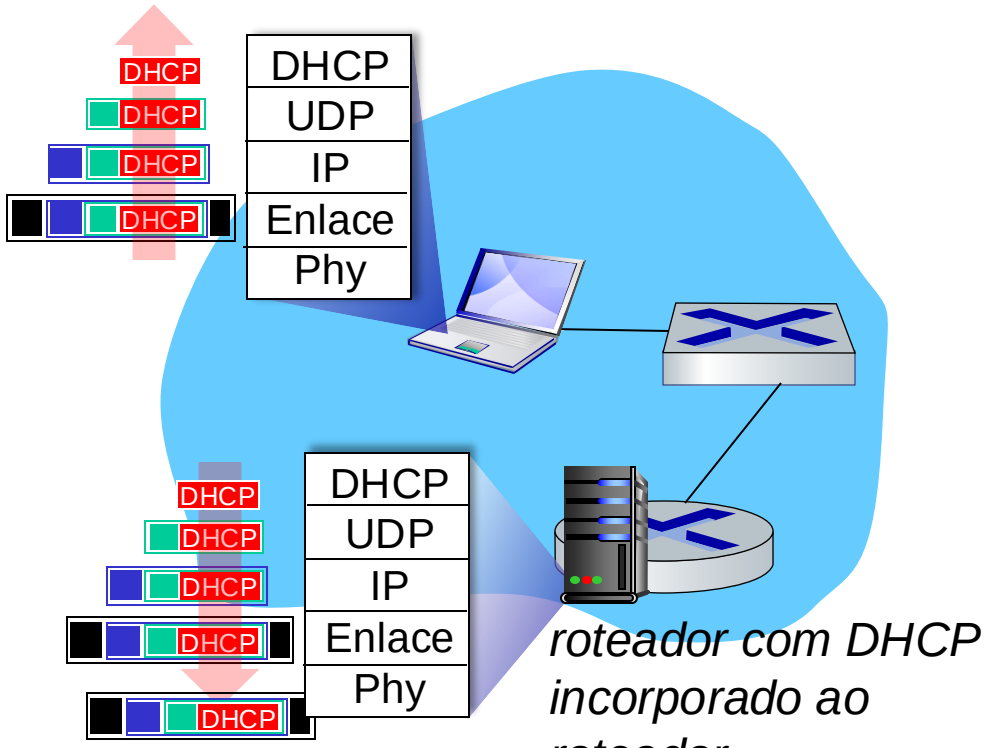
- endereço do roteador de primeiro salto para o cliente
- nome e endereço IP do servidor DNS
- máscara de rede (indicando a parte do endereço que é rede ou host)

DHCP: exemplo



- O laptop conectado usará o DHCP para obter o endereço IP, o endereço do roteador de primeiro salto e o endereço do servidor DNS.
- Mensagem DHCP REQUEST encapsulada em UDP, encapsulada em IP, encapsulada em Ethernet
- Broadcast de quadro Ethernet (destino: FFFFFFFFFFFF) na LAN, recebido no roteador que está executando o servidor DHCP
- Ethernet de-muxada para IP de-muxada, UDP de-muxada para DHCP

DHCP: exemplo



- O servidor DCP formula o DHCP ACK contendo o endereço IP do cliente, o endereço IP do roteador de primeiro salto para o cliente, o nome e o endereço IP do servidor DNS
- resposta encapsulada do servidor DHCP encaminhada para o cliente, com desmultiplicação até o DHCP no cliente
- O cliente agora sabe seu endereço IP, o nome e o endereço IP do servidor DNS, o endereço IP de seu roteador de primeiro salto

Endereços IP: como obter um?

P: Como *a rede* obtém a parte de sub-rede do endereço IP?

A: recebe uma parte alocada do espaço de endereços de seu provedor ISP

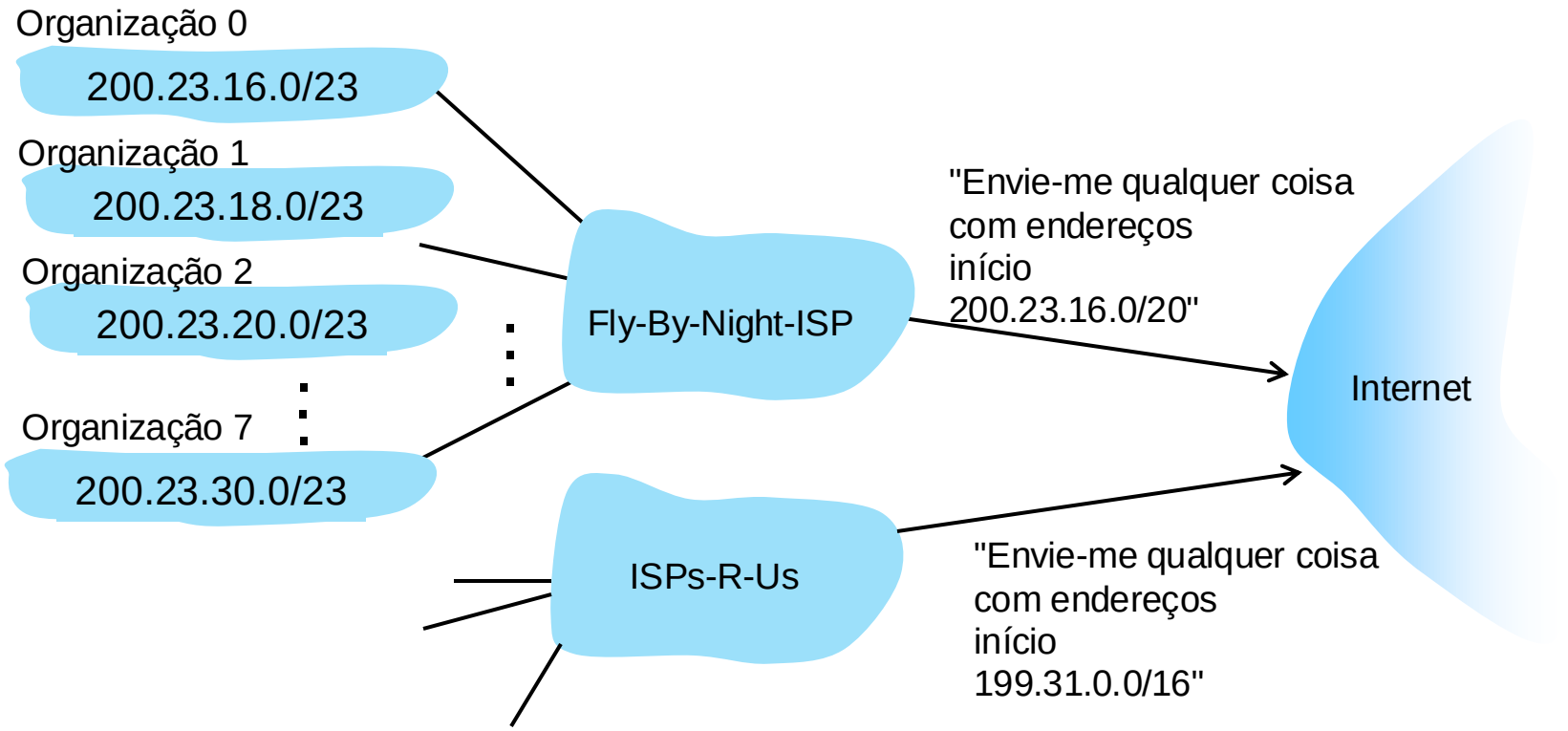
Bloqueio do ISP 11001000 00010111 00010000 00000000 200.23.16.0/20

O ISP pode então alocar seu espaço de endereço em 8 blocos:

Organização 0 11001000 00010111 00010000 00000000 200.23.16.0/23
Organização 1 11001000 00010111 00010010 00000000 200.23.18.0/23
Organização 2 11001000 00010111 00010100 00000000 200.23.20.0/23
... .
Organização 7 11001000 00010111 00011110 00000000 200.23.30.0/23

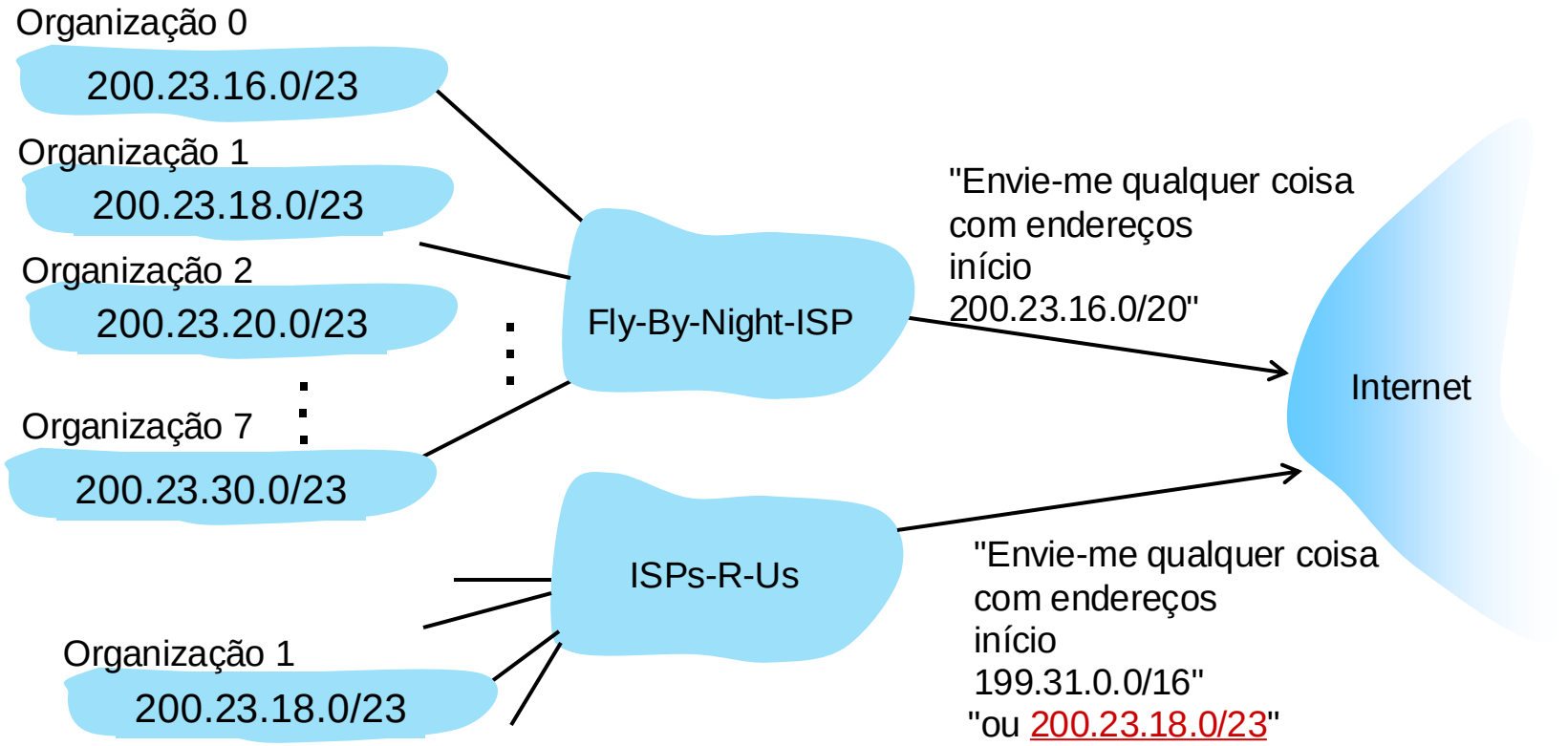
Endereçamento hierárquico: agregação de rotas

O endereçamento hierárquico permite o anúncio eficiente de informações de roteamento:



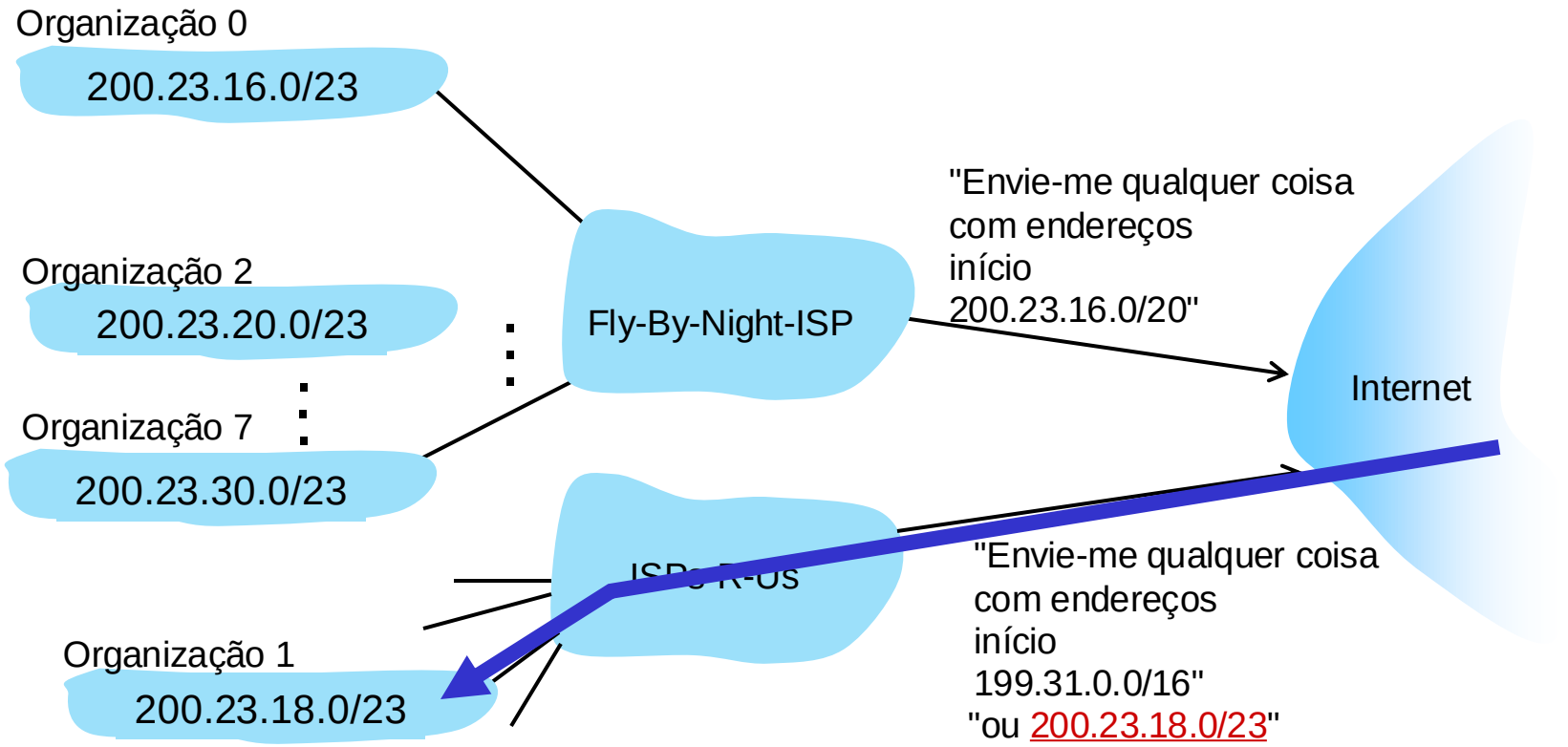
Endereçamento hierárquico: rotas mais específicas

- A organização 1 passa de um ISP que não funciona à noite para um ISPs-R-Us
- O ISPs-R-Us agora anuncia uma rota mais específica para a Organização 1



Endereçamento hierárquico: rotas mais específicas

- A organização 1 passa de um ISP que não funciona à noite para um ISPs-R-Us
- O ISPs-R-Us agora anuncia uma rota mais específica para a Organização 1



Endereçamento IP: últimas palavras ...

P: Como um ISP obtém um bloco de endereços?

A: ICANN: Corporação da Internet para Atribuição de Nomes e Números http://www.icann.org/

- aloca endereços IP, por meio de 5 registros regionais (RRs) (que podem então alocar para registros locais)
- gerencia a zona raiz do DNS, incluindo a delegação do gerenciamento de TLDs individuais (.com, .edu , ...)

P: Há endereços IP de 32 bits suficientes?

- ICANN alocou a última parcela de endereços IPv4 para RRs em 2011
- O NAT (a seguir) ajuda a esgotar o espaço de endereços IPv4
- O IPv6 tem um espaço de endereço de 128 bits

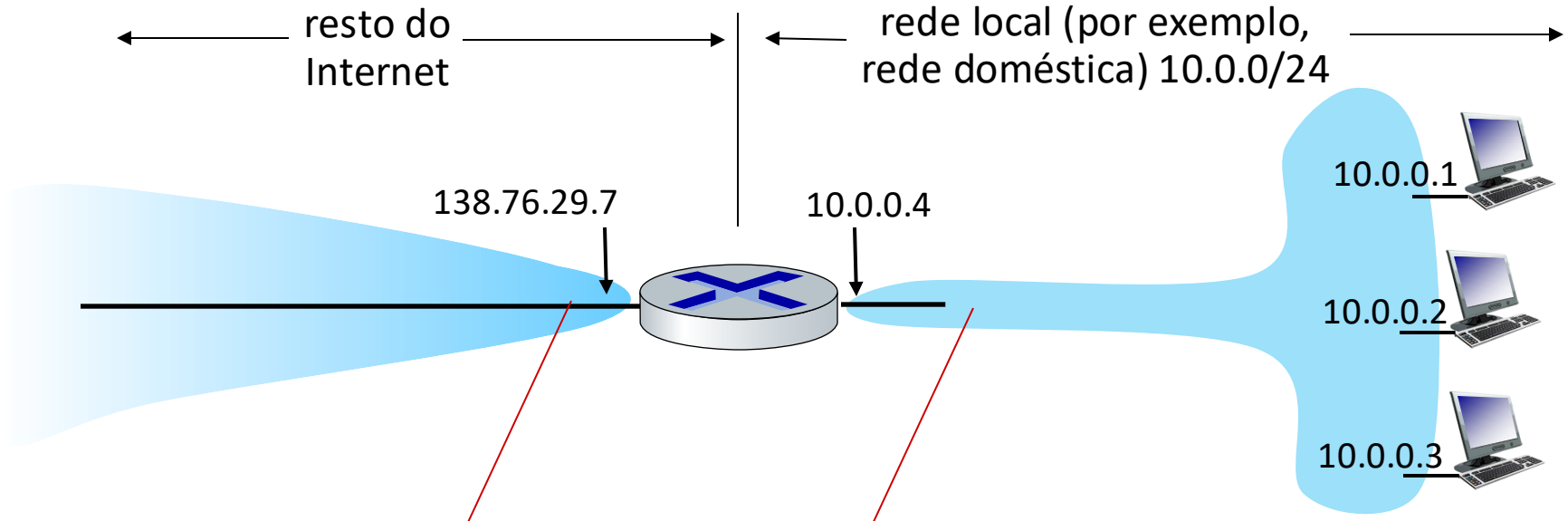
"Quem diabos sabia quanto espaço de endereço precisávamos?" Vint Cerf (refletindo sobre a decisão de tornar o endereço IPv4 com 32 bits de comprimento)

Camada de rede: Roteiro do "plano de dados"

- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo de Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereços de rede
  - IPv6
- Encaminhamento generalizado, SDN
  - correspondência+ação
  - OpenFlow: match+action em ação
- Caixas intermediárias

NAT: tradução de endereços de rede

NAT: todos os dispositivos da rede local compartilham apenas um endereço IPv4 em relação ao mundo externo



todos os datagramas *que saem da* rede local têm *o mesmo* endereço IP NAT de origem: 138.76.29.7, mas números de porta de origem *diferentes*

Os datagramas com origem ou destino nessa rede têm o endereço 10.0.0/24 como origem e destino (como de costume)

NAT: tradução de endereços de rede

- todos os dispositivos na rede local têm endereços de 32 bits em um espaço de endereço IP "privado" (prefixos 10/8, 172.16/12, 192.168/16) que só pode ser usado na rede local
- vantagens:
  - apenas um endereço IP necessário do provedor ISP para *todos os* dispositivos
  - pode alterar os endereços do host na rede local sem notificar o mundo externo
  - pode alterar o ISP sem alterar os endereços dos dispositivos na rede local
  - segurança: dispositivos dentro da rede local não diretamente endereçáveis, visíveis pelo mundo externo

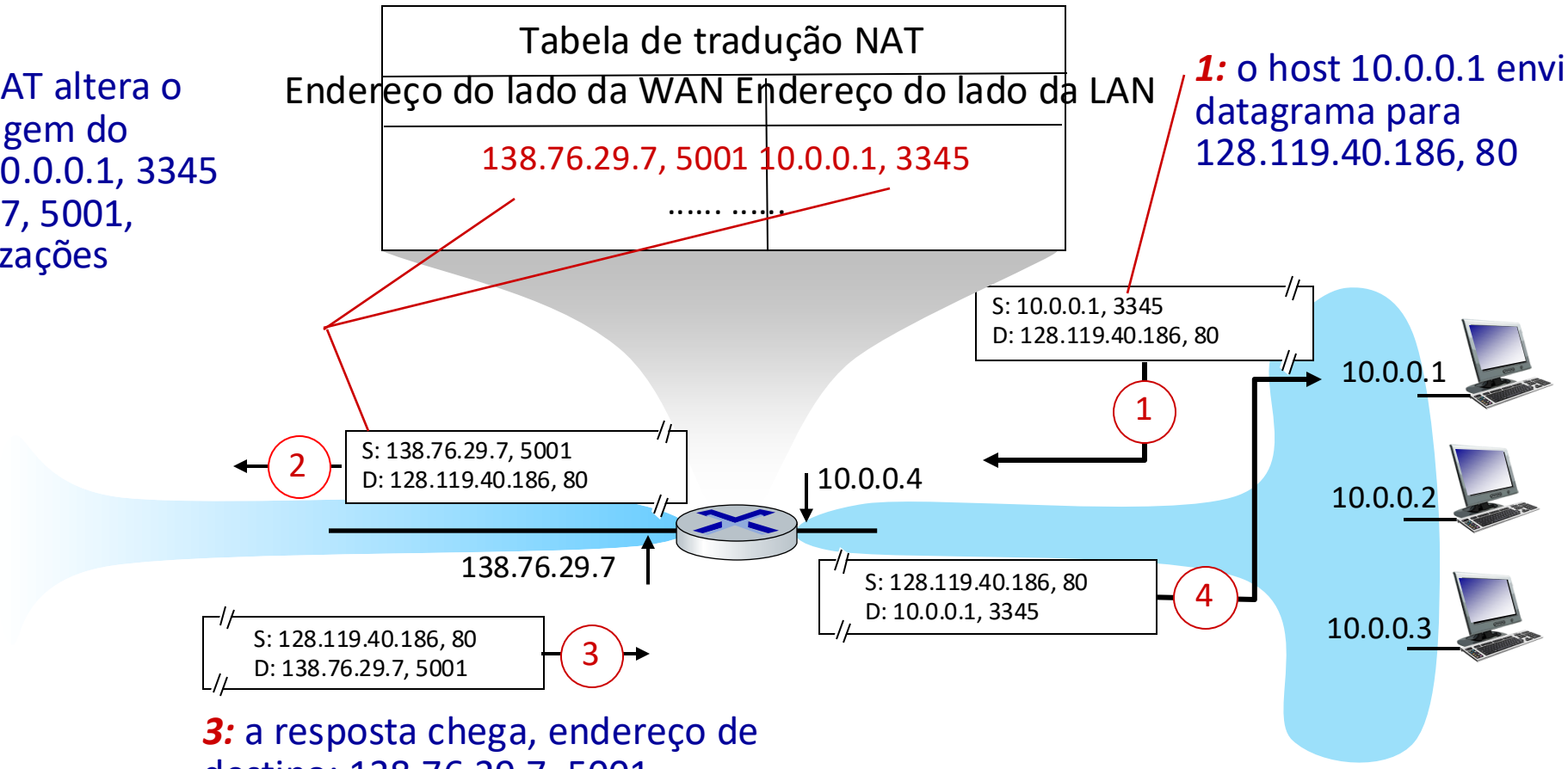
NAT: tradução de endereços de rede

implementação: O roteador NAT deve (de forma transparente):

- datagramas de saída: substitui (endereço IP de origem, porta nº) de cada datagrama de saída para (endereço IP NAT, nova porta nº)
  - clientes/servidores remotos responderão usando (endereço IP NAT, nova porta #) como endereço de destino
- lembrar (na tabela de tradução NAT) cada par de tradução (endereço IP de origem, porta nº) para (endereço IP NAT, nova porta nº)
- datagramas de entrada: substitua (endereço IP NAT, nova porta #) nos campos de destino de cada datagrama de entrada pelo correspondente (endereço IP de origem, porta #) armazenado na tabela NAT

NAT: tradução de endereços de rede

2: O roteador NAT altera o endereço de origem do datagrama de 10.0.0.1, 3345 para 138.76.29.7, 5001, tabela de atualizações



3: a resposta chega, endereço de destino: 138.76.29.7, 5001

NAT: tradução de endereços de rede

- O NAT tem sido controverso:
  - os roteadores "devem" processar somente até a camada 3
  - a "escassez" de endereços deve ser resolvida pelo IPv6
  - viola o argumento de ponta a ponta (manipulação de porta # pelo dispositivo de camada de rede)
  - NAT traversal: e se o cliente quiser se conectar ao servidor por trás do NAT?
- mas o NAT veio para ficar:
  - amplamente utilizado em redes domésticas e institucionais, redes celulares 4G/5G

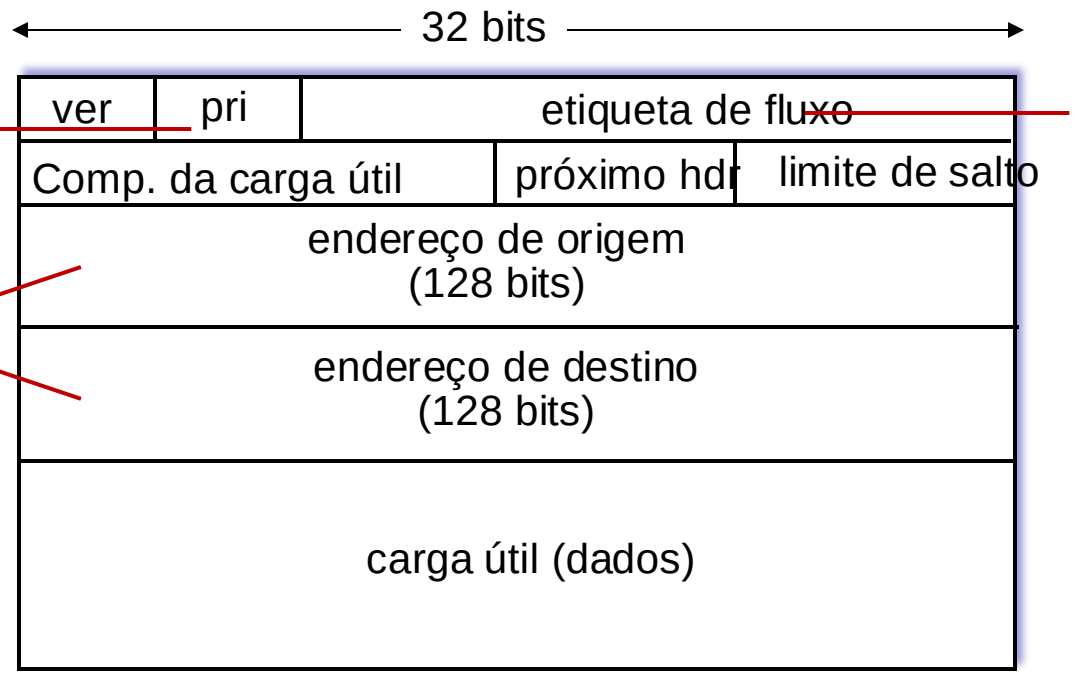
IPv6: motivação

- motivação inicial: O espaço de endereços IPv4 de 32 bits seria completamente alocado
- motivação adicional:
  - velocidade de processamento/encaminhamento: Cabeçalho de comprimento fixo de 40 bytes
  - permitem um tratamento diferente de "fluxos" na camada de rede

Formato do datagrama IPv6

prioridade: identifica a prioridade entre os datagramas no fluxo

128 bits
Endereços IPv6



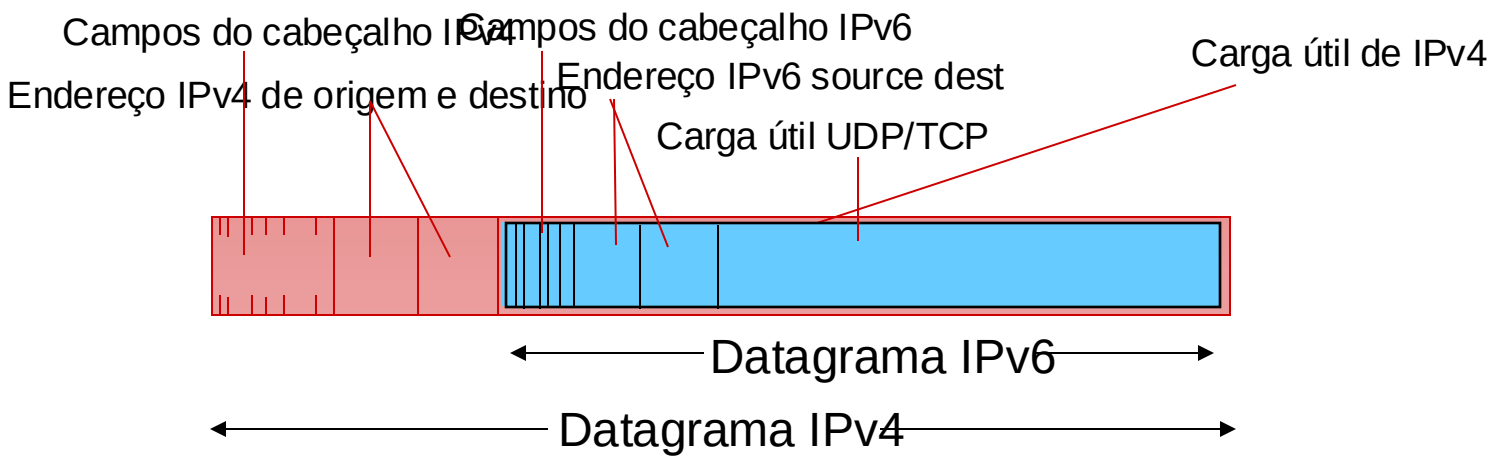
rótulo de fluxo: identifica datagramas no mesmo "fluxo". (o conceito de "fluxo" não está bem definido).

O que está faltando (em comparação com o IPv4):

- sem soma de verificação (para acelerar o processamento nos roteadores)
- sem fragmentação/montagem
- sem opções (disponível como camada superior, protocolo de próximo cabeçalho no roteador)

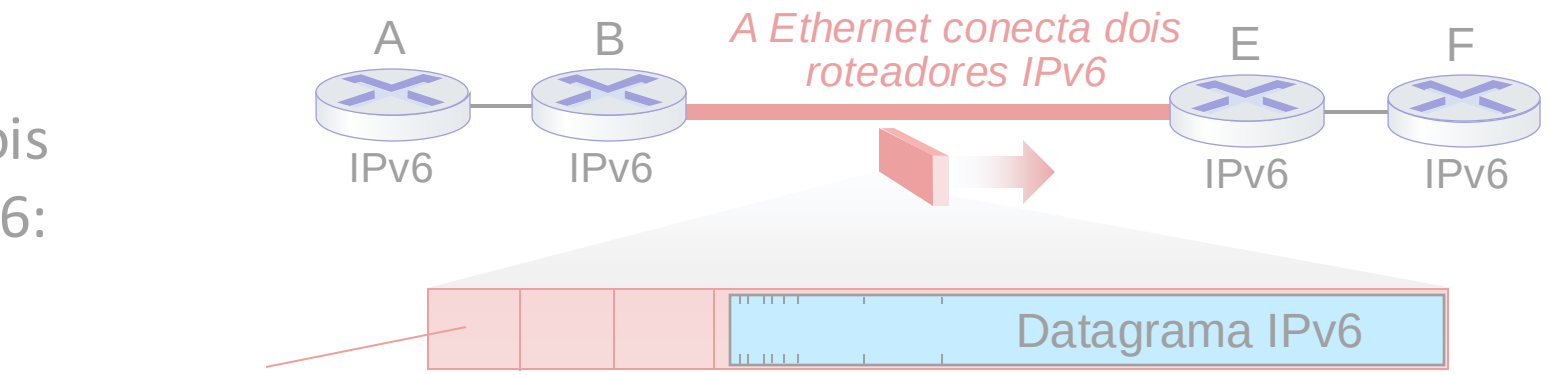
Transição do IPv4 para o IPv6

- nem todos os roteadores podem ser atualizados simultaneamente
  - sem "dias de bandeira"
  - Como a rede funcionará com roteadores IPv4 e IPv6 mistos?
- tunelamento: datagrama IPv6 transportado como *carga útil* em um datagrama IPv4 entre roteadores IPv4 ("pacote dentro de um pacote")
  - tunelamento usado extensivamente em outros contextos (4G/5G)



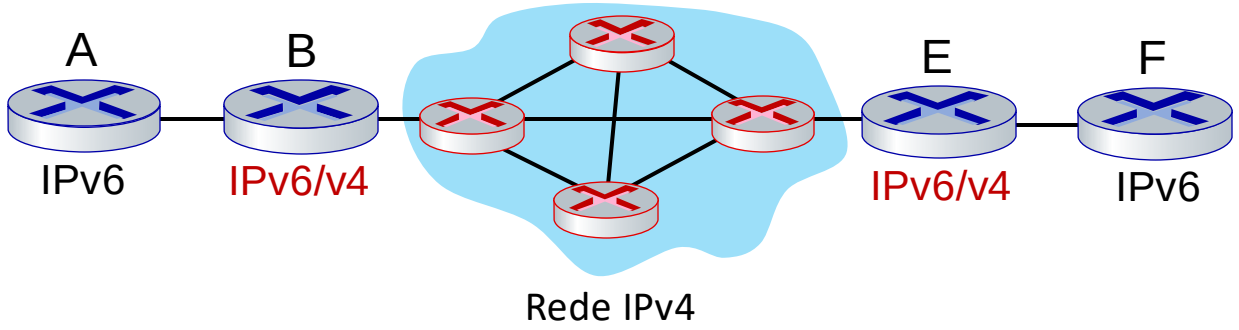
Tunelamento e encapsulamento

Ethernet conectando dois roteadores IPv6:



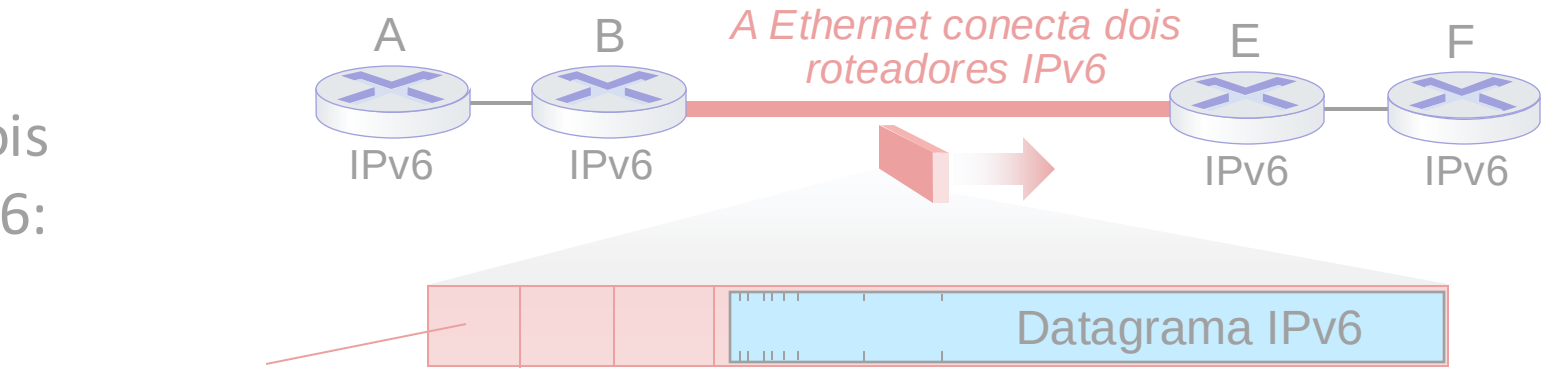
O usual: datagrama como carga útil no quadro da camada de enlace

Rede IPv4 conectando dois roteadores IPv6



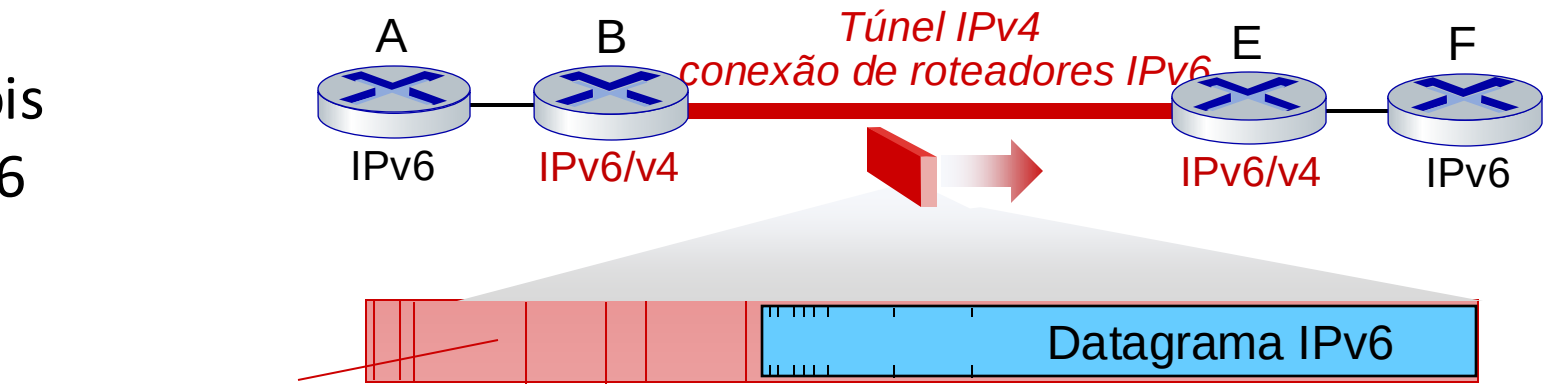
Tunelamento e encapsulamento

Ethernet conectando dois roteadores IPv6:



O usual: datagrama como carga útil no quadro da camada de enlace

Túnel IPv4 conectando dois roteadores IPv6

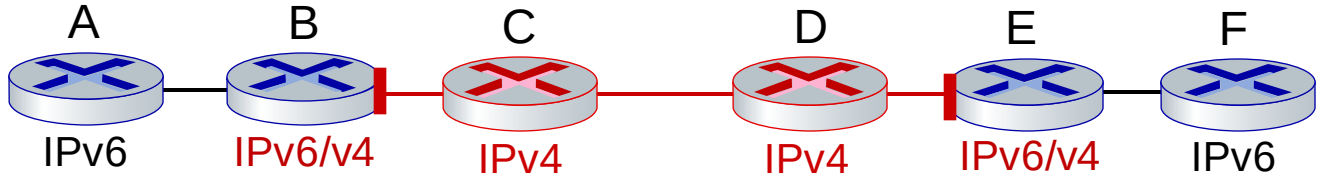


tunelamento: datagrama IPv6 como carga útil em um datagrama IPv4

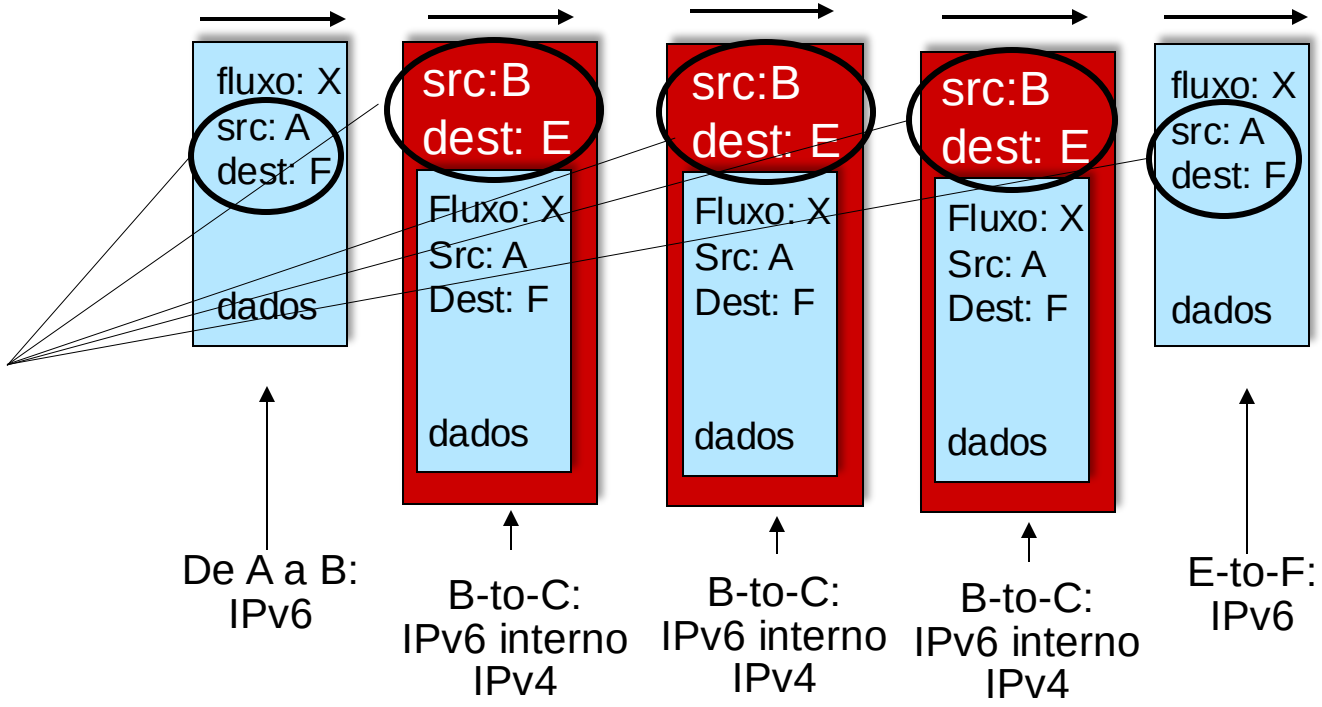
Tunelamento

visão lógica:

visão física:

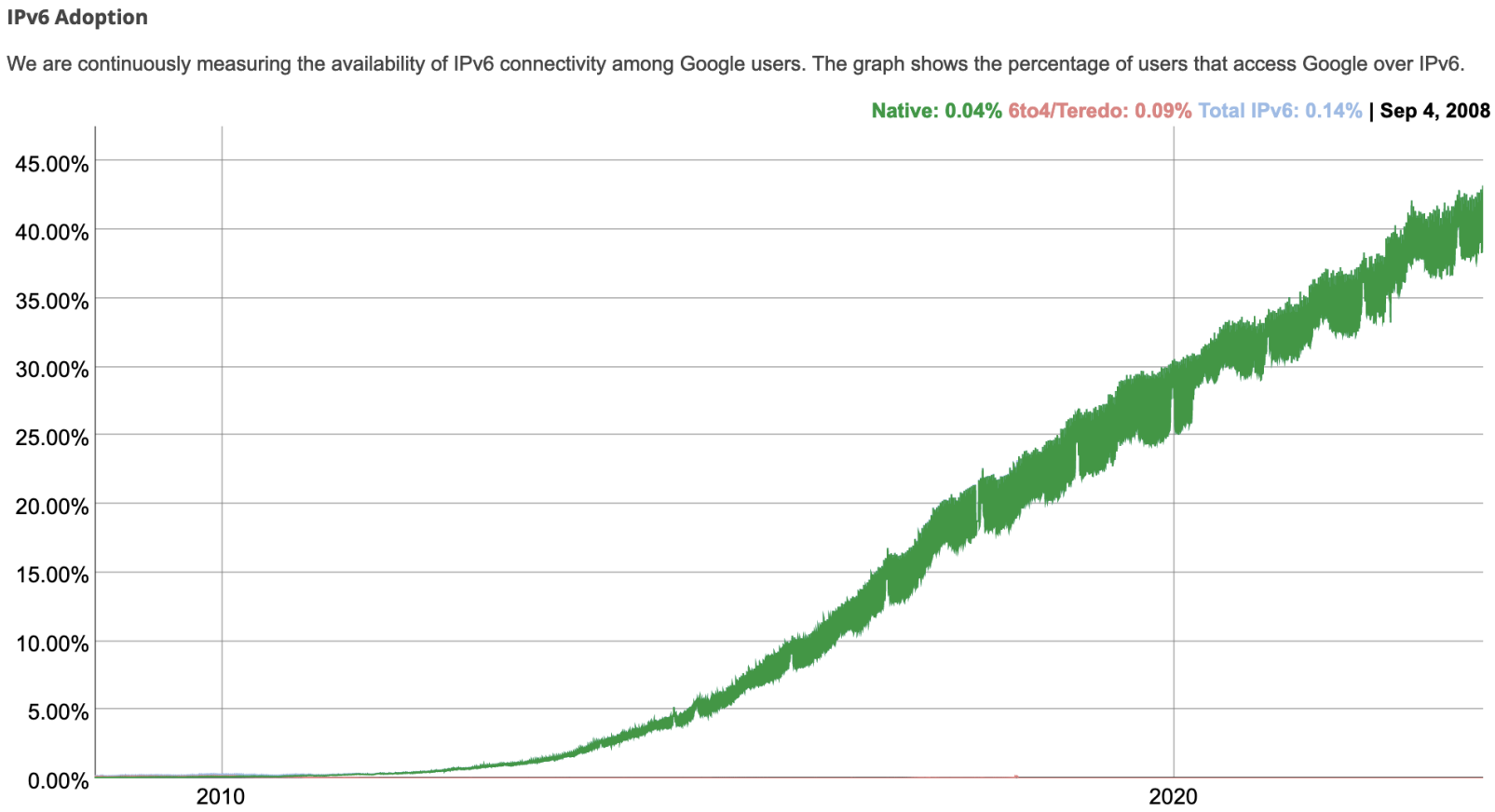


Observe os endereços de origem e destino!



IPv6: adoção

- Google[1] : ~ 40% dos clientes acessam serviços via IPv6 (2023)
- NIST: 1/3 de todos os domínios do governo dos EUA são compatíveis com IPv6



IPv6: adoção

- Google[1] : ~ 40% dos clientes acessam serviços via IPv6 (2023)
- NIST: 1/3 de todos os domínios do governo dos EUA são compatíveis com IPv6
- Tempo longo (longo!) para implantação, use
  - 25 anos e contando!
  - Pense nas mudanças no nível dos aplicativos nos últimos 25 anos: WWW, mídia social, mídia de streaming, jogos, telepresença, ...
  - *Por quê?*

[1] https://www.google.com/intl/en/ipv6/statistics.html

Camada de rede: Roteiro do "plano de dados"

- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo de Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereços de rede
  - IPv6

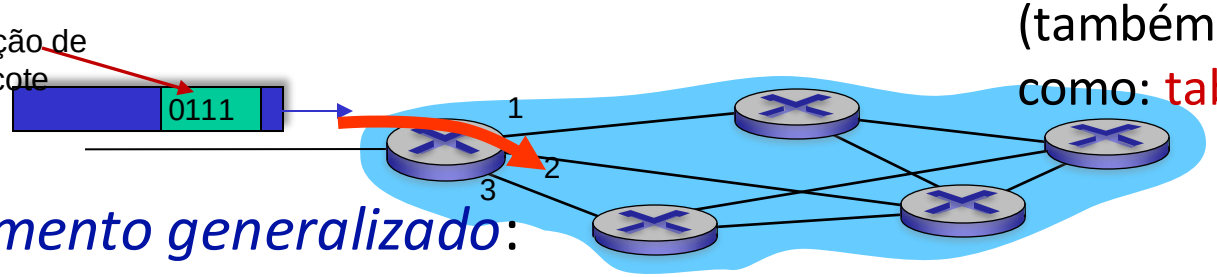
- Encaminhamento generalizado, SDN
  - Partida+ação
  - OpenFlow: match+action em ação
- Caixas intermediárias

Encaminhamento generalizado: correspondência mais ação

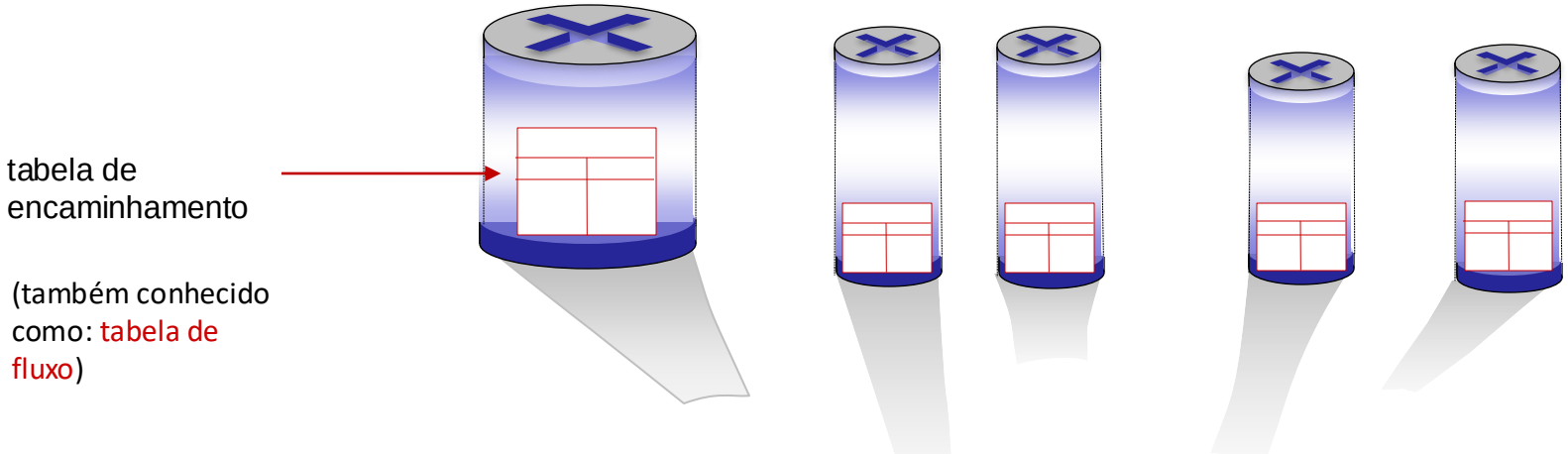
Revisão: cada roteador contém uma tabela de encaminhamento

- Abstração de "correspondência mais ação": correspondência de bits no pacote que chega, ação

  - *encaminhamento baseado em destino:* encaminhar com base no endereço IP de destino. endereço IP (também conhecido como: tabela de fluxo)

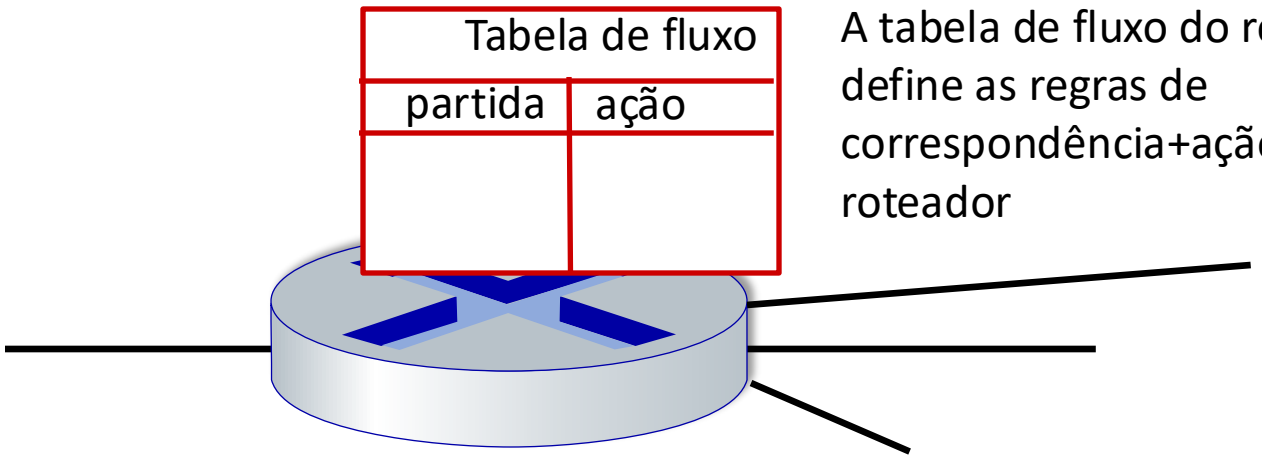


  - *encaminhamento generalizado*:
    - Muitos campos de cabeçalho podem determinar a ação
    - várias ações possíveis: descartar/copiar/modificar/registrar o pacote



Abstração da tabela de fluxo

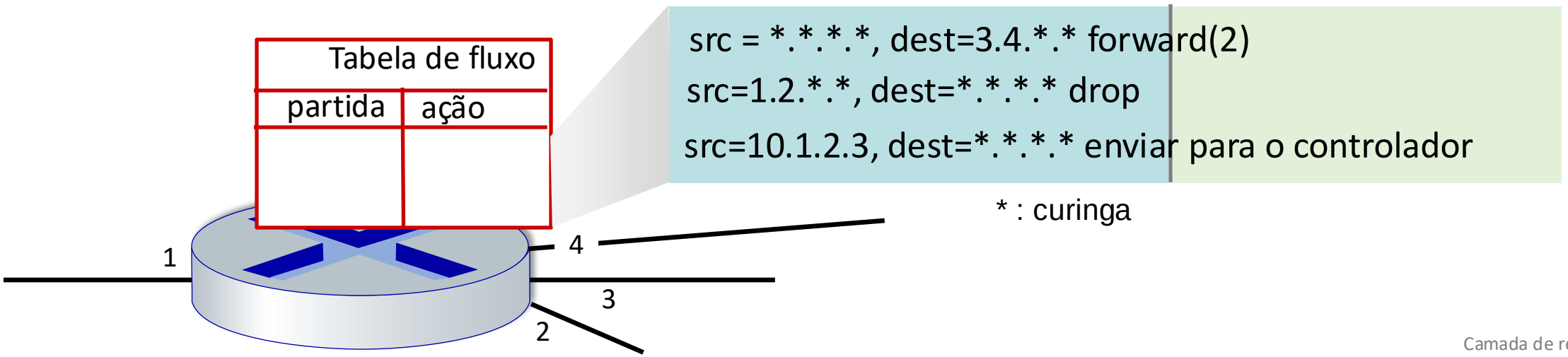
- fluxo: definido pelos valores do campo de cabeçalho (nos campos de camada de transporte, rede e link)
- encaminhamento generalizado: regras simples de manuseio de pacotes
  - match: valores de padrão nos campos do cabeçalho do pacote
  - ações: para o pacote correspondente: descartar, encaminhar, modificar o pacote correspondente ou enviar o pacote correspondente ao controlador
  - prioridade: desambiguação de padrões sobrepostos
  - contadores: #bytes e #packets



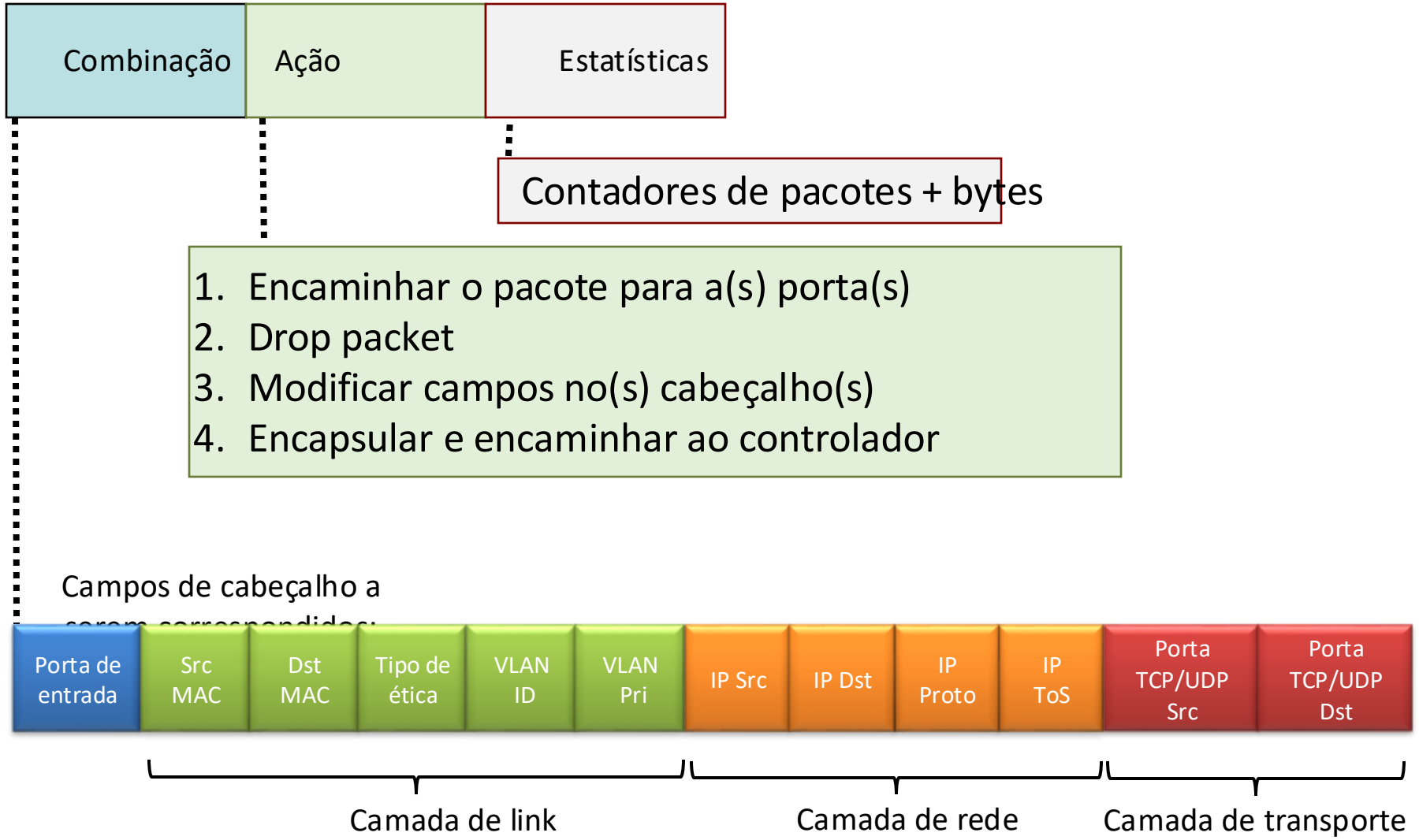
A tabela de fluxo do roteador define as regras de correspondência+ação do roteador

Abstração da tabela de fluxo

- fluxo: definido por campos de cabeçalho
- encaminhamento generalizado: regras simples de manuseio de pacotes
  - match: valores de padrão nos campos do cabeçalho do pacote
  - ações: para o pacote correspondente: descartar, encaminhar, modificar o pacote correspondente ou enviar o pacote correspondente ao controlador
  - prioridade: desambiguação de padrões sobrepostos
  - contadores: #bytes e #packets



OpenFlow: entradas da tabela de fluxo



OpenFlow: exemplos

Encaminhamento baseado em destino:

Interruptor Porto	MAC src	MAC dst	Ética tipo	VLAN ID	VLAN Pri	IP Src	IP Dst	IP Prot	IP ToS	TCP porta s	porta d	Ação
*	*	*	*	*	*	*	51.6.0.8	*	*	*	*	porta6

Os datagramas IP destinados ao endereço IP 51.6.0.8 devem ser encaminhados para a porta de saída 6 do roteador

Firewall:

Interruptor Porto	MAC src	MAC dst	Ética tipo	VLAN ID	VLAN Pri	IP Src	IP Dst	IP Prot	IP ToS	TCP porta s	porta d	Ação
*	*	*	*	*	*	*	*	*	*	*	22	queda

Bloquear (não encaminhar) todos os datagramas destinados à porta TCP 22 (porta ssh nº)

Interruptor Porto	MAC src	MAC dst	Ética tipo	VLAN ID	VLAN Pri	IP Src	IP Dst	IP Prot	IP ToS	TCP porta s	porta d	Ação
*	*	*	*	*	*	128.119.1.1	*	*	*	*	*	queda

Bloquear (não encaminhar) todos os datagramas enviados pelo host 128.119.1.1

OpenFlow: exemplos

Encaminhamento baseado em destino da camada 2:

Interruptor Porto	MAC src	MAC dst	Ética tipo	VLAN ID	VLAN Pri	IP Src	IP Dst	IP Prot	IP ToS	TCP porta s	TCP porta d	Ação
*	*	22:A7:23:11:E1:02	*	*	*	*	*	*	*	*	*	porta3

Os quadros de camada 2 com endereço MAC de destino 22:A7:23:11:E1:02 devem ser encaminhados para a porta de saída 3

Abstração do OpenFlow

- match+action: a abstração unifica diferentes tipos de dispositivos

Roteador

- *match:* prefixo IP de destino mais longo
- *ação:* encaminhar um link

Interruptor

- *match:* endereço MAC de destino
- *ação:* avançar ou inundar

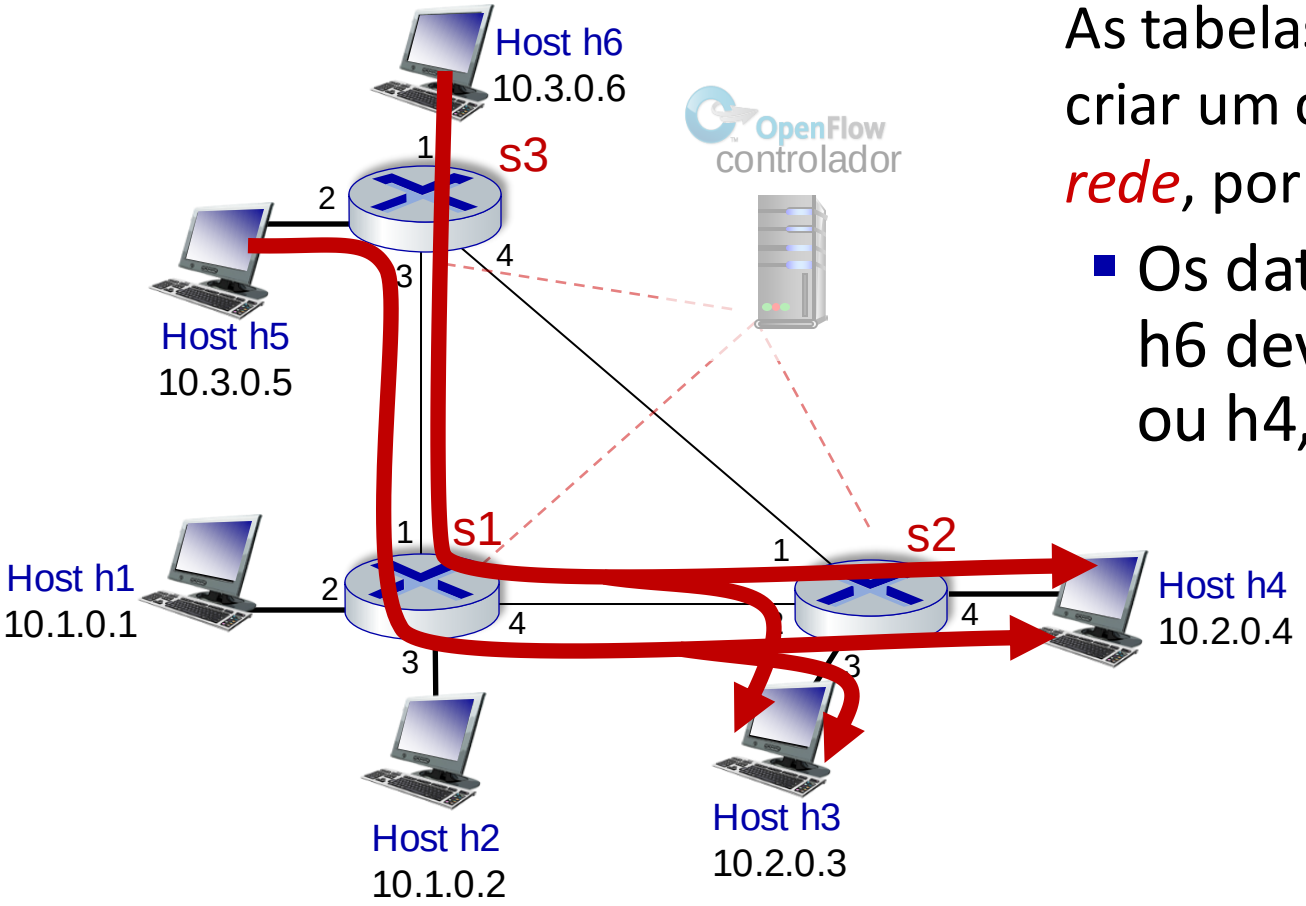
Firewall

- *corresponder*: Endereços IP e números de porta TCP/UDP
- *ação:* permitir ou negar

NAT

- *corresponder:* Endereço IP e porta
- *ação:* reescrever endereço e porta

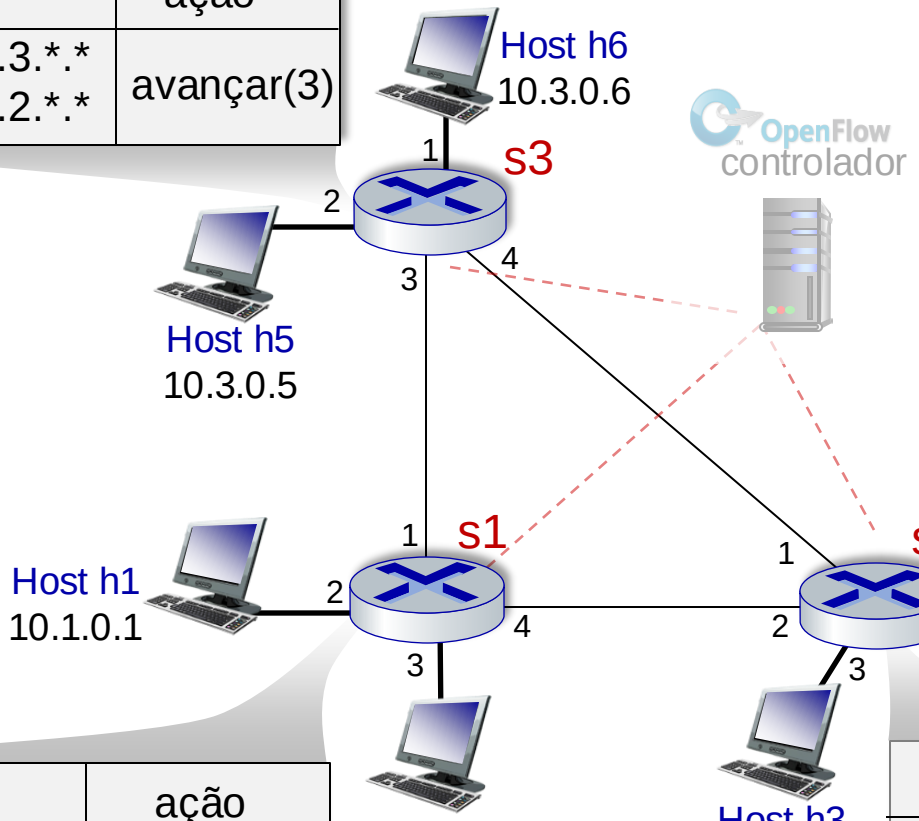
Exemplo de OpenFlow



As tabelas orquestradas podem criar um comportamento *em toda a rede*, por exemplo:

- Os datagramas dos hosts h5 e h6 devem ser enviados para h3 ou h4, via s1, e de lá para s2

Exemplo de OpenFlow



partida	ação
IP Src = 10.3.*.* IP Dst = 10.2.*.*	avançar(3)

partida	ação
porta de entrada = 1 IP Src = 10.3.*.* IP Dst = 10.2.*.*	avançar(4)

partida	ação
porta de entrada = 2 IP Dst = 10.2.0.3	avançar(3)
porta de entrada = 2 IP Dst = 10.2.0.4	avançar(4)

As tabelas orquestradas podem criar um comportamento *em toda a rede*, por exemplo:

- Os datagramas dos hosts h5 e h6 devem ser enviados para h3 ou h4, via s1, e de lá para s2

Encaminhamento generalizado: resumo

- Abstração de "correspondência mais ação": corresponder bits no(s) cabeçalho(s) do pacote que chega(m) em qualquer camada, tomar medidas
  - correspondência em vários campos (camada de link, rede, transporte)
  - ações locais: descartar, encaminhar, modificar ou enviar o pacote correspondente ao controlador
  - "programar" comportamentos *em toda a rede*
- forma simples de "programabilidade de rede"
  - "Processamento" programável e por pacote
  - *raízes históricas:* rede ativa
  - *hoje:* programação mais generalizada: P4 (consulte p4.org).

Camada de rede: Roteiro do "plano de dados"

- Camada de rede: visão geral
- O que há dentro de um roteador
- IP: o Protocolo de Internet
- Encaminhamento generalizado
- Caixas intermediárias
  - funções do middlebox
  - evolução, princípios arquitetônicos da Internet

Caixas intermediárias

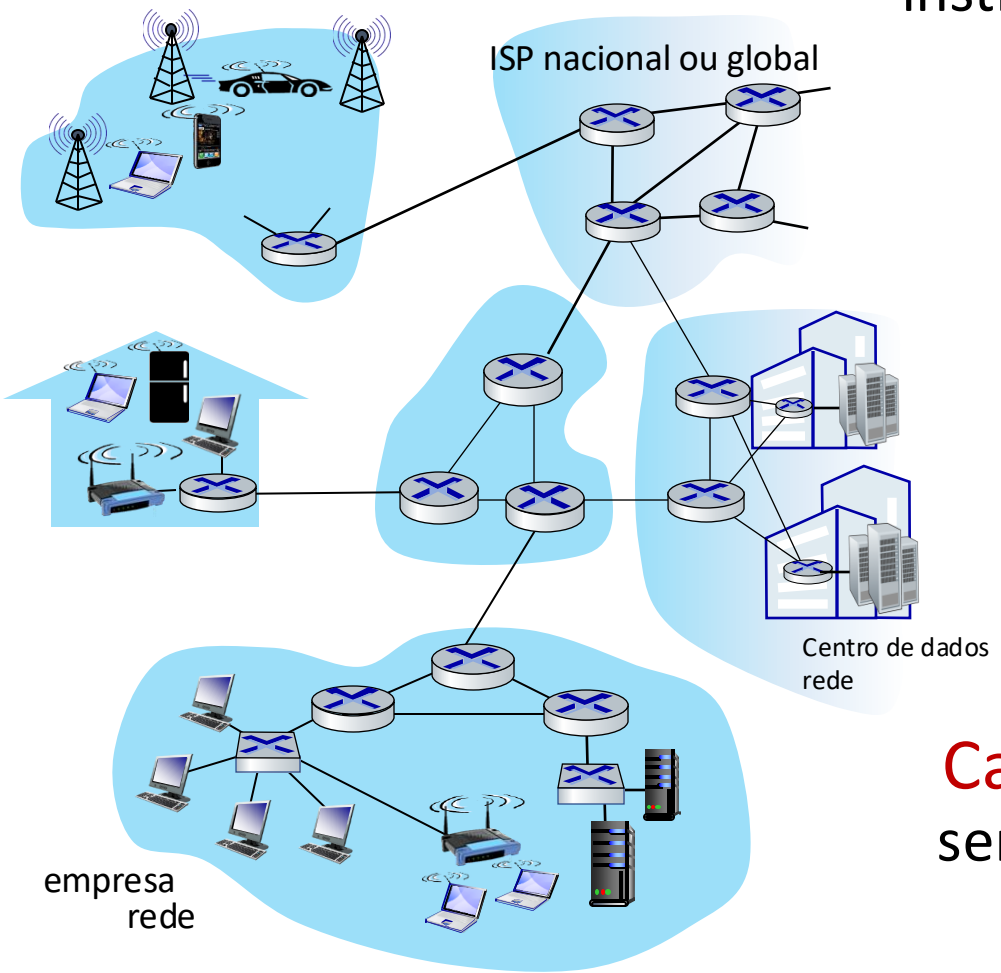
Middlebox (RFC 3234)

"qualquer caixa intermediária que execute funções além das funções normais e padrão de um roteador IP no caminho de dados entre um host de origem e um host de destino"

Middleboxes em todos os lugares!

NAT: residencial, celular, institucional

Específico do aplicativo: provedores de serviços, institucional, CDN



Firewalls, IDS: corporativo, institucional, provedores de serviços, ISPs

Balanceadores de carga: corporativo, provedor de serviços, data center, redes móveis

Caches: provedor de serviços, celular, CDNs

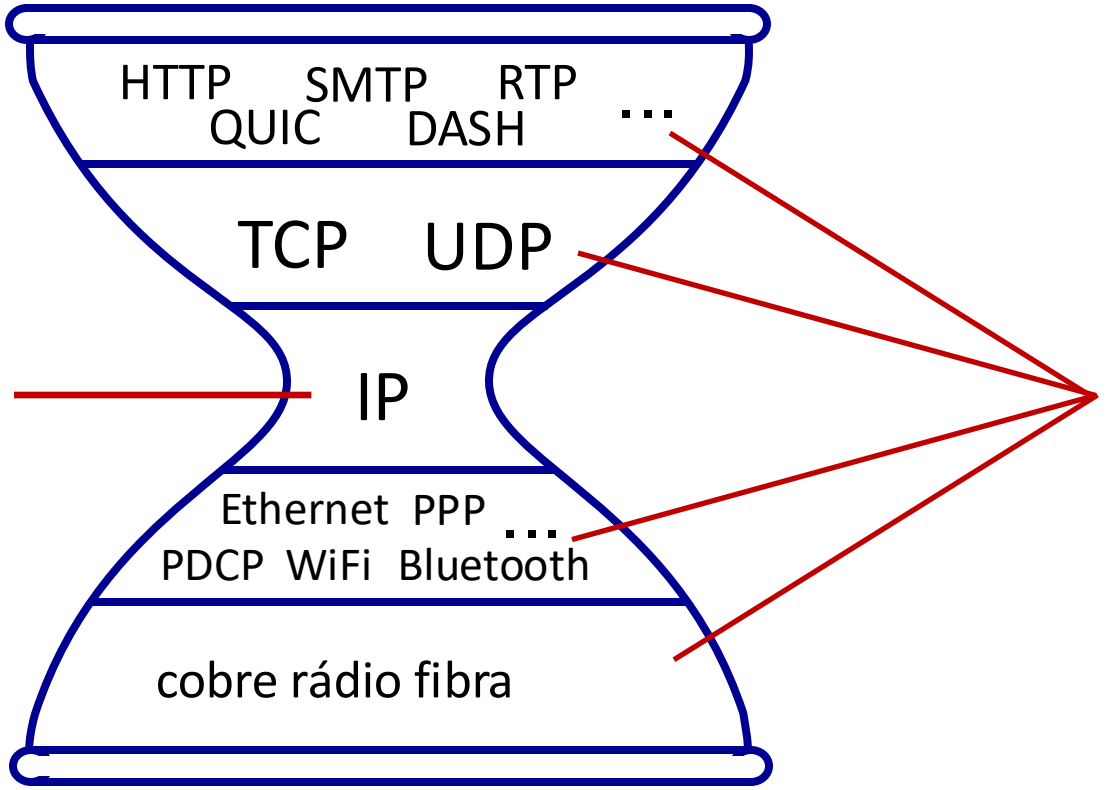
Caixas intermediárias

- inicialmente: soluções de hardware proprietárias (fechadas)
- mudança para hardware "whitebox" que implementa API aberta
  - abandonar as soluções de hardware proprietário
  - ações locais programáveis via match+action
  - mudança para inovação/diferenciação em software
- SDN: controle centralizado (logicamente) e gerenciamento de configuração, geralmente em nuvem privada/pública
- virtualização de funções de rede (NFV): serviços programáveis em rede, computação e armazenamento de caixa branca

A ampulheta de IP

A "cintura fina" da Internet:

- *um* protocolo de camada de rede: IP
- *deve* ser implementado por todos os (bilhões) de dispositivos conectados à Internet

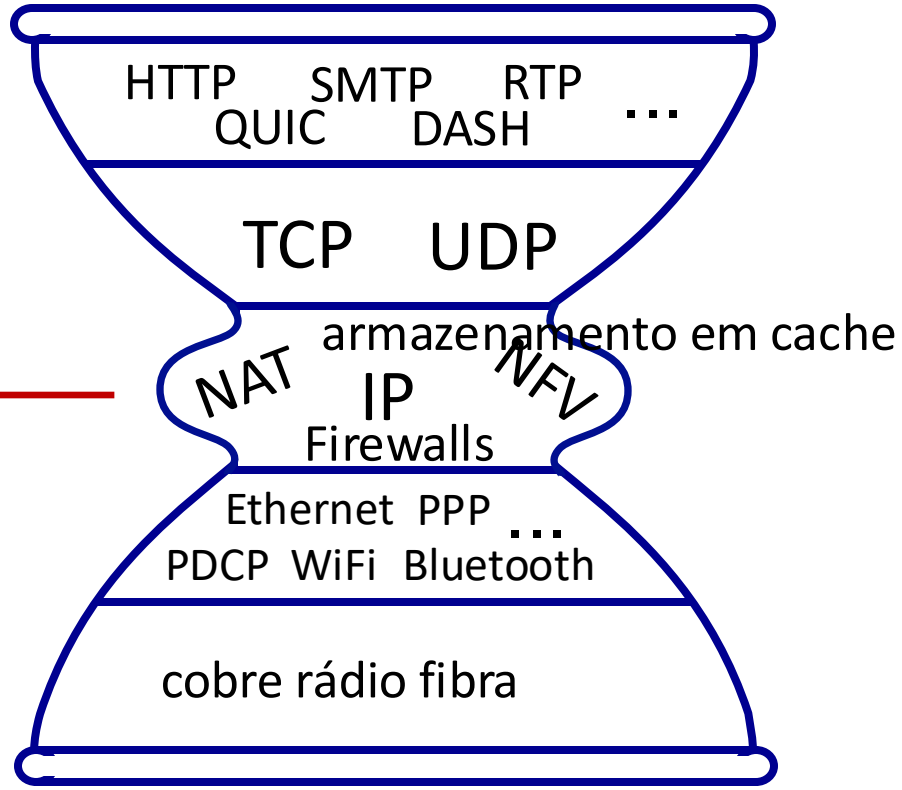


vários protocolos nas camadas física, de link, de transporte e de aplicativos

A ampulheta IP, na meia-idade

Os "pneuzinhos" da meia-idade na Internet?

- middleboxes, operando dentro da rede



Princípios arquitetônicos da Internet

RFC 1958

"Muitos membros da comunidade da Internet argumentariam que não existe uma arquitetura, mas apenas uma tradição, que não foi escrita nos primeiros 25 anos (ou pelo menos não pelo IAB). Entretanto, em termos muito gerais, a comunidade acredita que

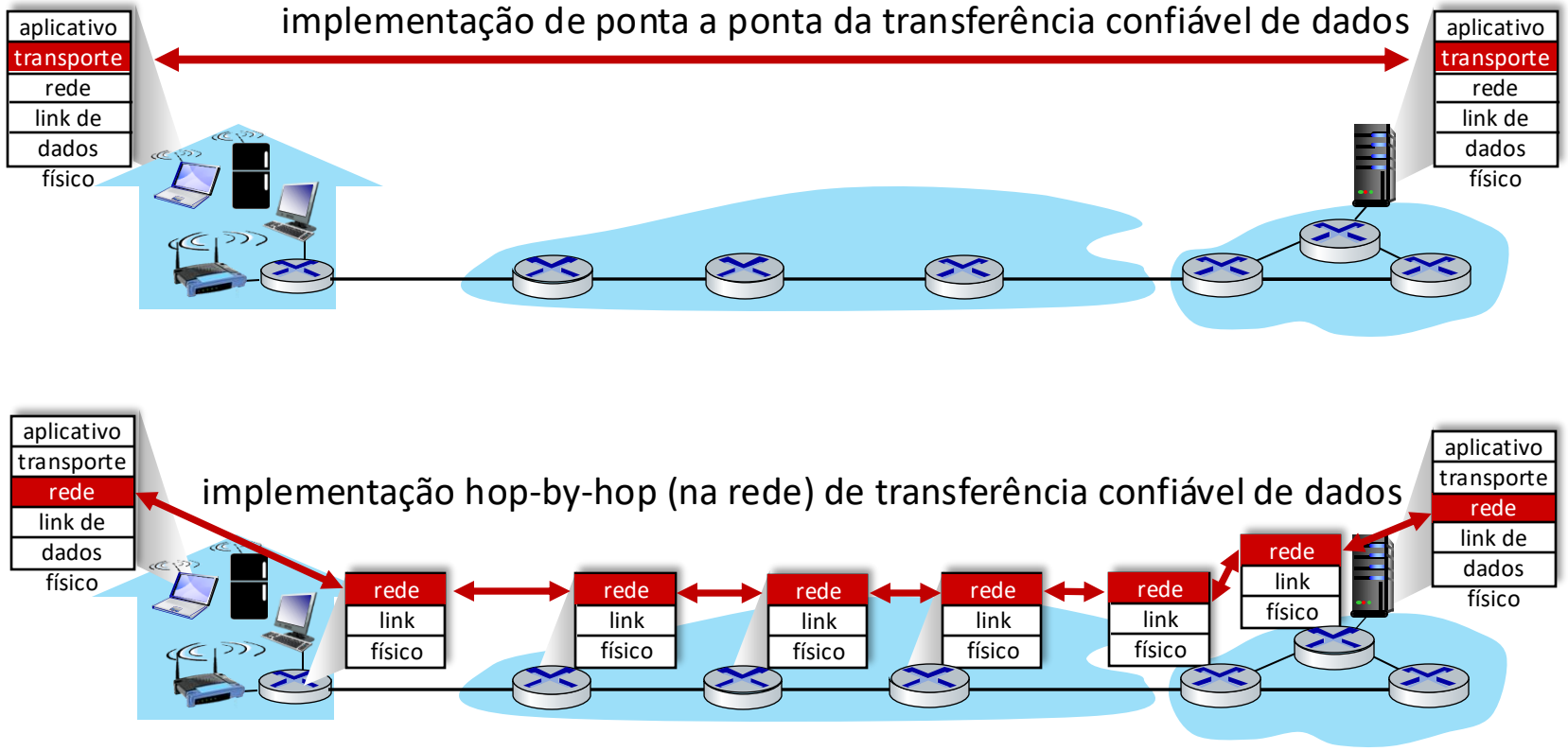
o objetivo é a conectividade, a ferramenta é o Protocolo de Internet e a inteligência é de ponta a ponta, e não oculta na rede".

Três crenças fundamentais:

- conectividade simples
- Protocolo IP: essa cintura estreita
- inteligência e complexidade na borda da rede

O argumento da extremidade final

- algumas funcionalidades da rede (por exemplo, transferência confiável de dados, congestionamento) podem ser implementadas na rede ou na borda da rede



O argumento da extremidade final

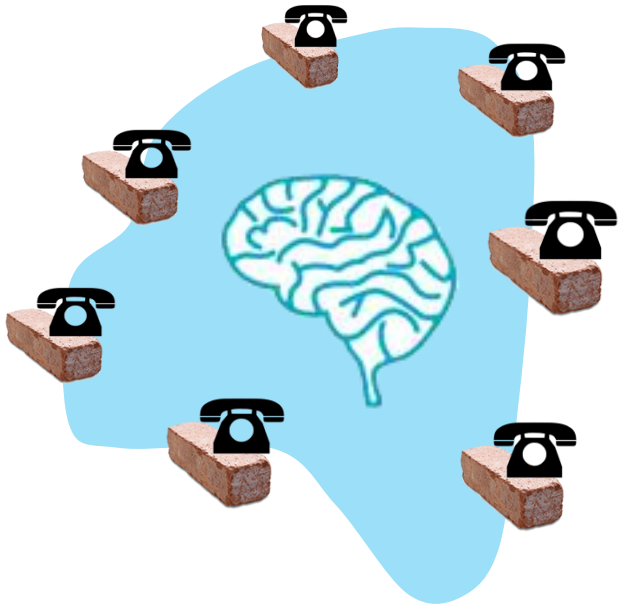
- algumas funcionalidades da rede (por exemplo, transferência confiável de dados, congestionamento) podem ser implementadas na rede ou na borda da rede

"A função em questão pode ser implementada de forma completa e correta somente com o conhecimento e a ajuda do aplicativo que está nos pontos finais do sistema de comunicação. Portanto, não é possível fornecer essa função questionada como um recurso do próprio sistema de comunicação. (Às vezes, uma versão incompleta da função fornecida pelo sistema de comunicação pode ser útil como um aprimoramento de desempenho).

Chamamos essa linha de raciocínio contra a implementação de funções de baixo nível de "argumento de ponta a ponta".

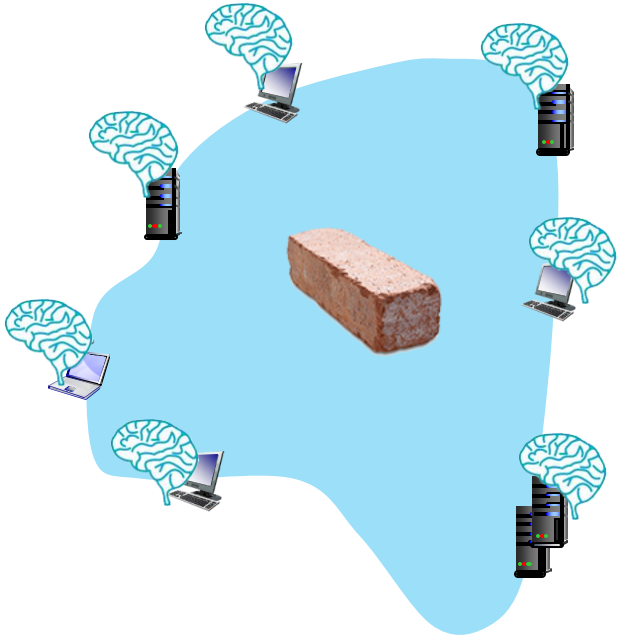
Saltzer, Reed, Clark 1981

Onde está a inteligência?



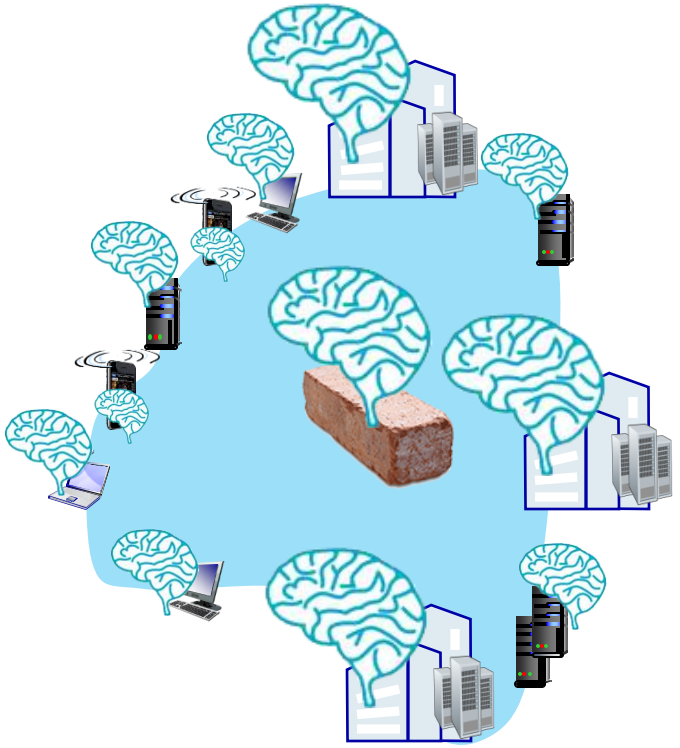
20th century phone net:

- inteligência/computação em switches de rede



Internet (antes de 2005)

- inteligência, computação na borda



Internet (após 2005)

- dispositivos de rede programáveis
- inteligência, computação, infraestrutura massiva em nível de aplicativo na borda

Capítulo 4: concluído!

- Camada de rede: visão geral
- O que há dentro de um roteador
- IP: o Protocolo de Internet
- Encaminhamento generalizado, SDN
- Caixas intermediárias

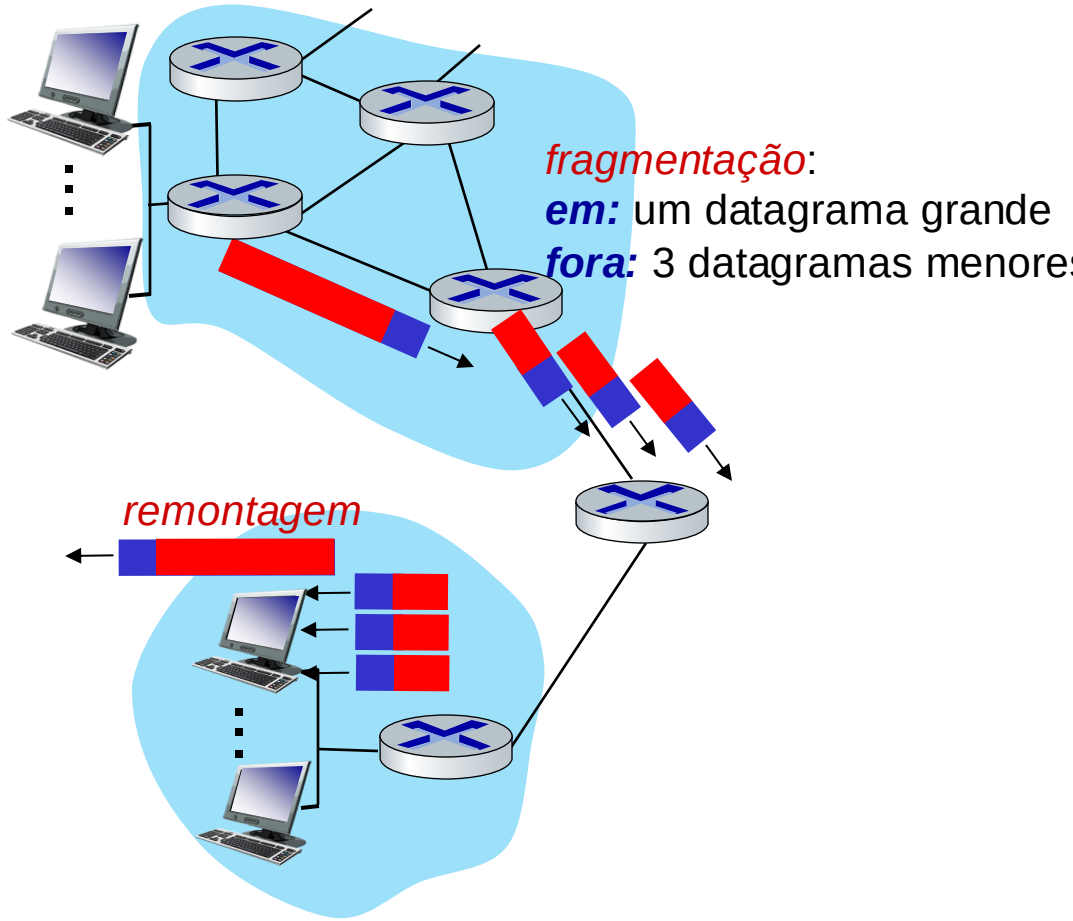
Pergunta: como são computadas as tabelas de encaminhamento (encaminhamento baseado em destino) ou as tabelas de fluxo (encaminhamento generalizado)?

Resposta: pelo plano de controle (próximo capítulo)

Slides adicionais do Capítulo 4

Fragmentação/montagem de IP

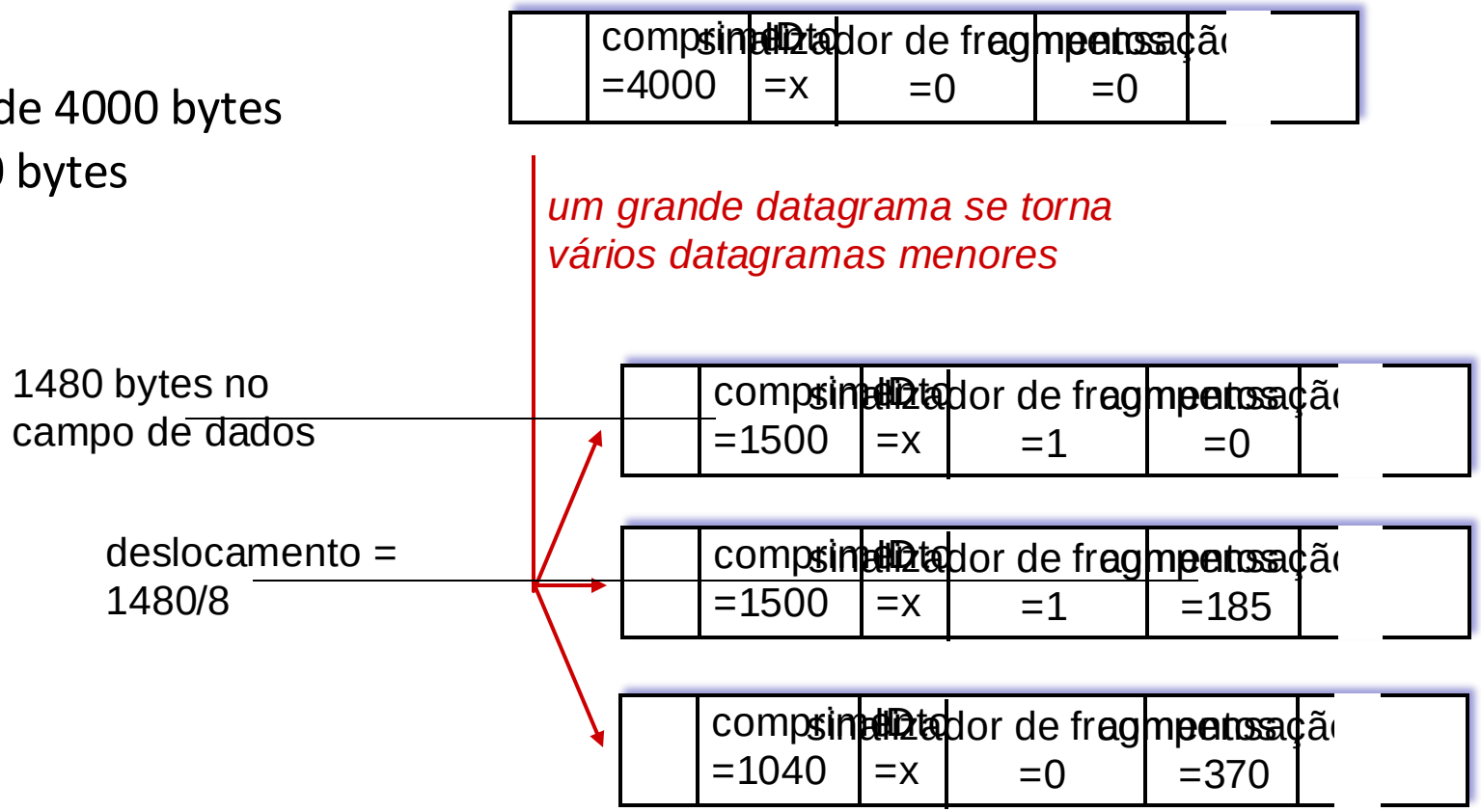
- os links de rede têm MTU (tamanho máximo de transferência) - o maior quadro possível em nível de link
  - diferentes tipos de links, diferentes MTUs
- datagrama IP grande dividido ("fragmentado") dentro da rede
  - um datagrama se transforma em vários datagramas
  - "remontado" somente no *destino*
  - Bits de cabeçalho IP usados para identificar e ordenar fragmentos relacionados



Fragmentação/montagem de IP

exemplo:

- Datagrama de 4000 bytes
- MTU = 1500 bytes



DHCP: saída do Wireshark (LAN doméstica)

solicitação

Tipo de mensagem: **Solicitação de inicialização (1)**
Tipo de hardware: Ethernet
Comprimento do endereço de hardware: 6
Lúpulo: 0
ID da transação: 0x6b3a11b7
Segundos decorridos: 0
Sinalizadores de bootp: 0x0000 (Unicast)
Endereço IP do cliente: 0.0.0.0 (0.0.0.0)
Seu endereço IP (cliente): 0.0.0.0 (0.0.0.0)
Endereço IP do próximo servidor: 0.0.0.0 (0.0.0.0)
Endereço IP do agente de retransmissão: 0.0.0.0 (0.0.0.0)
Endereço MAC do cliente: Wistron_23:68:8a (00:16:d3:23:68:8a)
Nome do host do servidor não fornecido
Nome do arquivo de inicialização não fornecido
Cookie mágico: (OK)
Opção: (t=53,l=1) **Tipo de mensagem DHCP = Solicitação DHCP**
Opção: (61) Identificador do cliente
    Comprimento: 7; Valor: 010016D323688A;
    Tipo de hardware: Ethernet
    Endereço MAC do cliente: Wistron_23:68:8a (00:16:d3:23:68:8a)
Opção: (t=50,l=4) Endereço IP solicitado = 192.168.1.101
Opção: (t=12,l=5) Nome do host = "nomad"
Opção: (55) Lista de solicitação de parâmetros
    Comprimento: 11; Valor: 010F03062C2E2F1F21F92B
    1 = Máscara de sub-rede; 15 = Nome de domínio
    3 = Roteador; 6 = Servidor de nomes de domínio
    44 = Servidor de nomes NetBIOS sobre TCP/IP
    ……

resposta

Tipo de mensagem: **Resposta de inicialização (2)**
Tipo de hardware: Ethernet
Comprimento do endereço de hardware: 6
Lúpulo: 0
ID da transação: 0x6b3a11b7
Segundos decorridos: 0
Sinalizadores de bootp: 0x0000 (Unicast)
Endereço IP do cliente: 192.168.1.101 (192.168.1.101)
Seu endereço IP (cliente): 0.0.0.0 (0.0.0.0)
Endereço IP do próximo servidor: 192.168.1.1 (192.168.1.1)
Endereço IP do agente de retransmissão: 0.0.0.0 (0.0.0.0)
Endereço MAC do cliente: Wistron_23:68:8a (00:16:d3:23:68:8a)
Nome do host do servidor não fornecido
Nome do arquivo de inicialização não fornecido
Cookie mágico: (OK)
Opção: (t=53,l=1) Tipo de mensagem DHCP = DHCP ACK
Opção: (t=54,l=4) Identificador do servidor = 192.168.1.1
Opção: (t=1,l=4) Máscara de sub-rede = 255.255.255.0
Opção: (t=3,l=4) Roteador = 192.168.1.1
Opção: (6) Servidor de nomes de domínio
    Comprimento: 12; Valor: 445747E2445749F244574092;
    Endereço IP: 68.87.71.226;
    Endereço IP: 68.87.73.242;
    Endereço IP: 68.87.64.146
Opção: (t=15,l=20) Nome de domínio = "hsd1.ma.comcast.net."