Redes de Computadores

A Camada de Rede: Plano de Dados

Fabricio Breve
www.fabriciobreve.com

Chapter 4
Network Layer: Data Plane

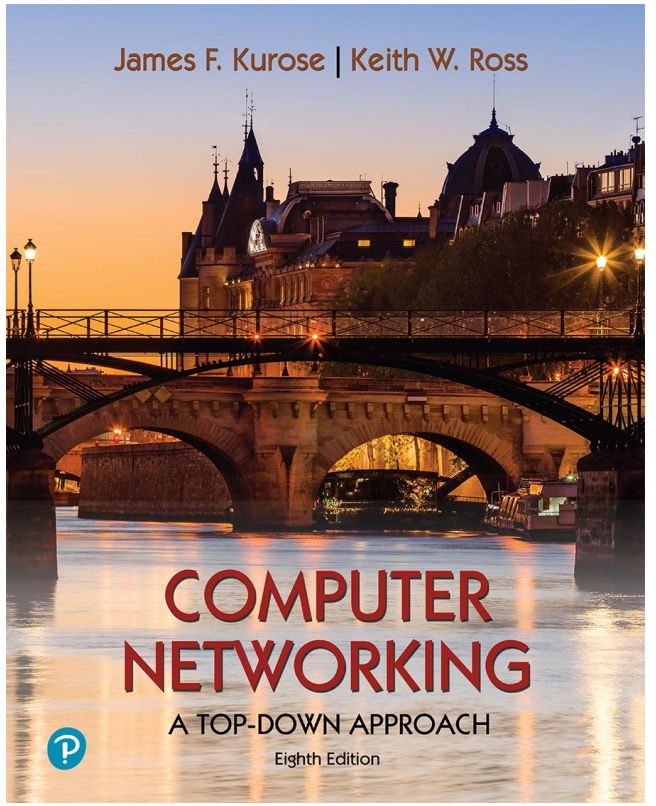
A note on the use of these PowerPoint slides:

We're making these slides freely available to all (faculty, students, readers). They're in PowerPoint form so you see the animations; and can add, modify, and delete slides (including this one) and slide content to suit your needs. They obviously represent a *lot* of work on our part. In return for use, we only ask the following:

- If you use these slides (e.g., in a class) that you mention their source (after all, we'd like people to use our book!)
- If you post any slides on a www site, that you note that they are adapted from (or perhaps identical to) our slides, and note our copyright of this material.

For a revision history, see the slide note for this page.

Thanks and enjoy! JFK/KWR



Computer Networking: A Top-Down Approach

8th edition
Jim Kurose, Keith Ross
Pearson, 2020

Camada de Rede: nossos objetivos

- entender princípios por trás dos serviços de camada de rede, focando 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 e implementação na Internet
  - protocolo IP
  - NAT, *middleboxes*

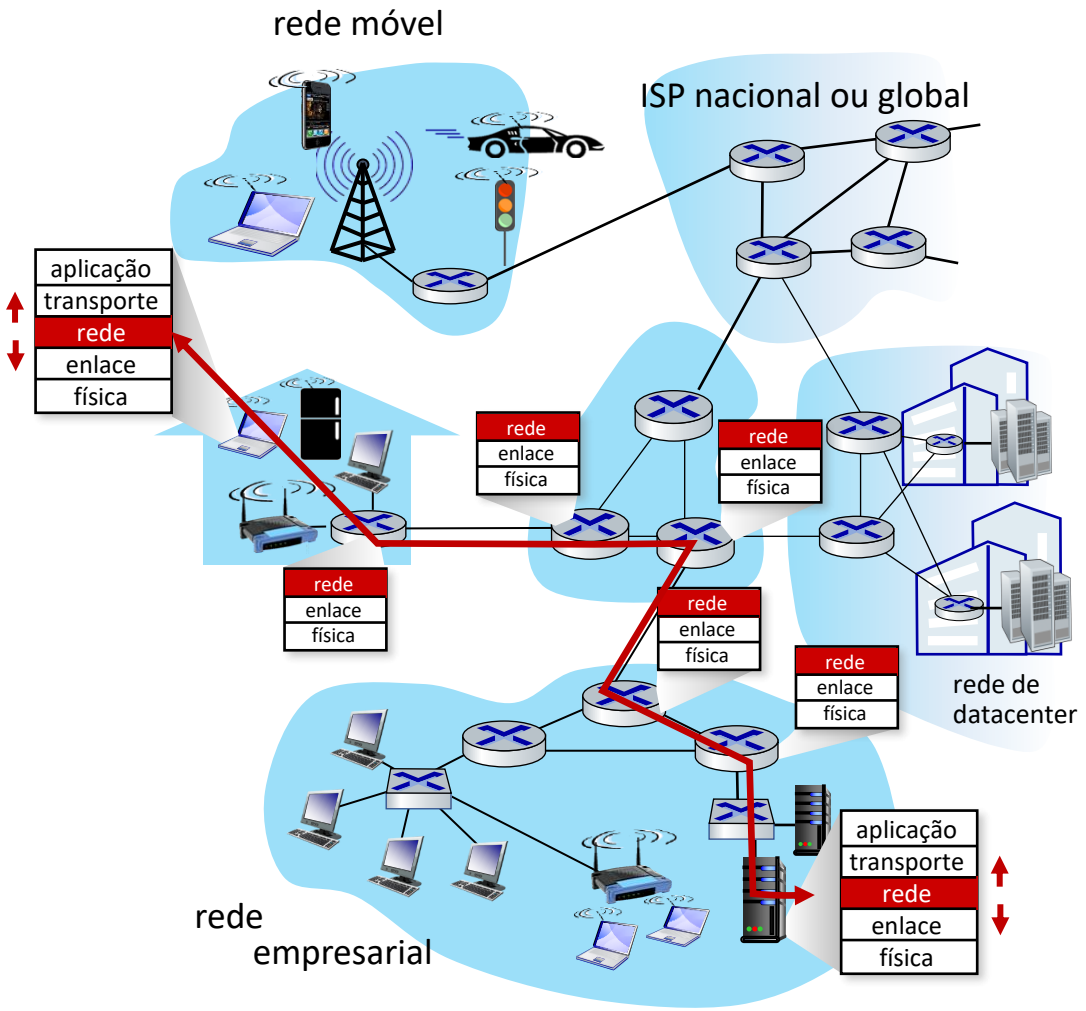
Camada de rede: roteiro do “plano de dados”

- 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 da Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereço de rede (NAT)
  - IPv6

- Encaminhamento Generalizado e SDN
  - Correspondência+ação
  - OpenFlow: correspondência+ação em ação
- Middleboxes

Serviços e protocolos da camada de rede

- transporta segmento do hospedeiro emissor ao receptor
  - emissor: encapsula segmentos em datagramas, passa para a camada de enlace
  - receptor: entrega segmentos para protocolo da camada de transporte
- protocolos de camada de rede em *todos os dispositivo de Internet*: hospedeiros e roteadores
- roteadores:
  - examinam campos de cabeçalho em todos os datagramas IP que passam por eles
  - move datagramas de portas de entrada para portas de saída para transferi-los ao longo do caminho final



Duas funções-chave da camada de rede

funções da camada de rede:

- *encaminhamento:* move pacotes do enlace de entrada de um roteador para o enlace de saída apropriado
- *roteamento:* determina rota tomada por pacotes da origem para o destino
  - *algoritmos de roteamento*

analogia: fazer uma viagem

- *encaminhamento:* processo de passar por um único cruzamento
- *roteamento:* processo de planejar a viagem da origem ao destino

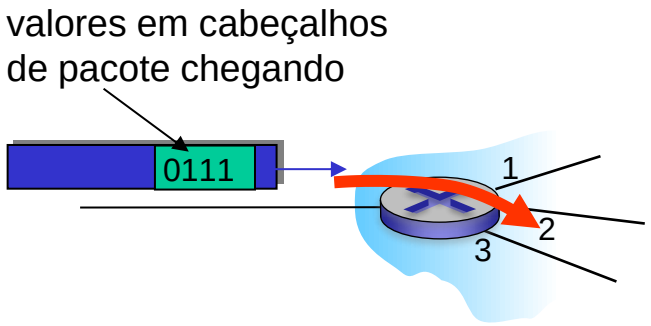
encaminhamento

roteamento

Camada de Rede: plano de dados e plano de controle

Plano de dados:

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

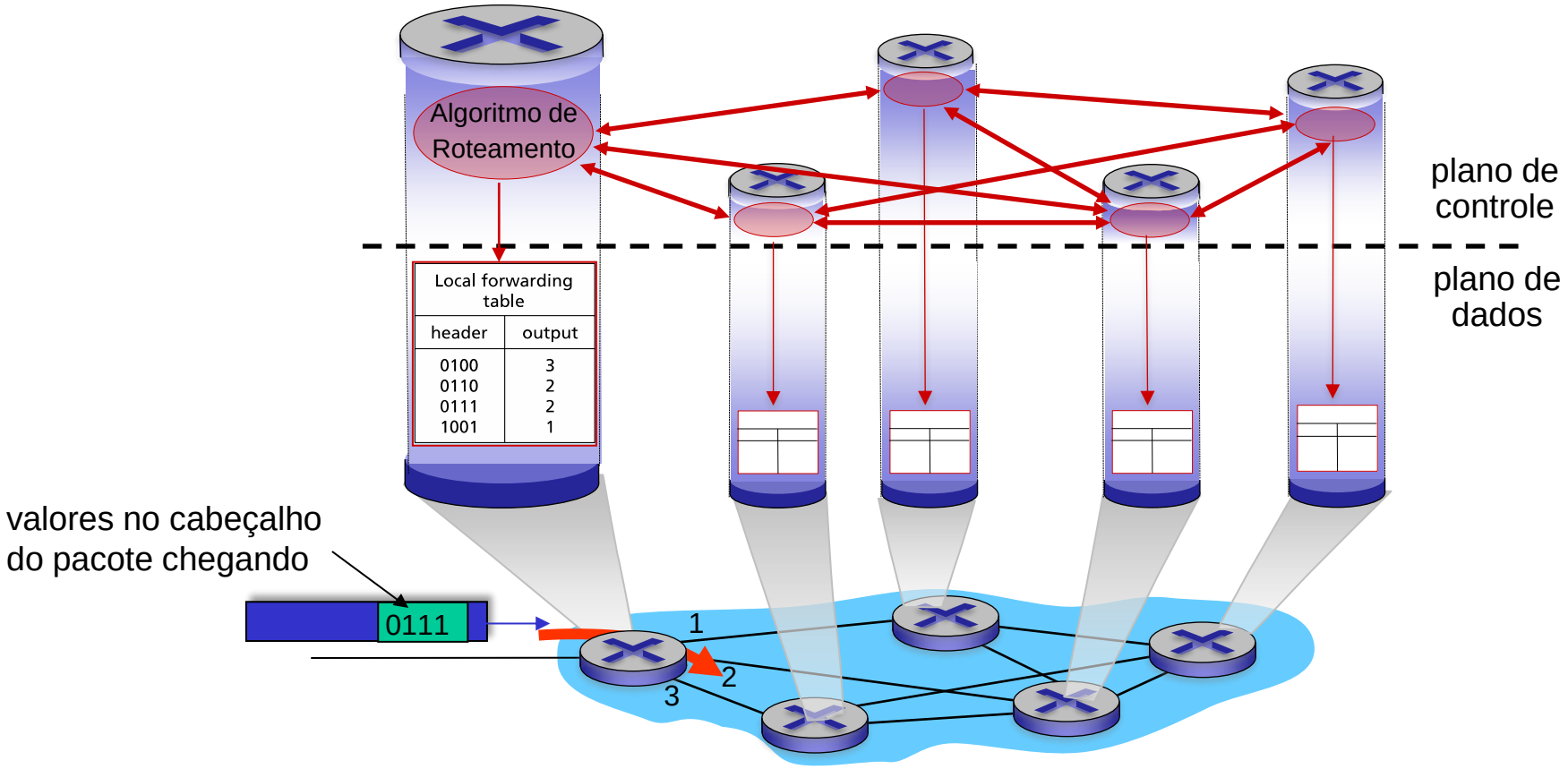


Plano de controle:

- lógica *em toda a rede*
- determina como o datagrama é roteado entre roteadores ao longo do caminho fim-a-fim do hospedeiro de origem ao de destino
- duas abordagens de plano de controle:
  - *algoritmos de roteamento tradicionais:* implementados em roteadores
  - *software-defined networking (SDN – rede definida por software)*: implementada em servidores (remotos)

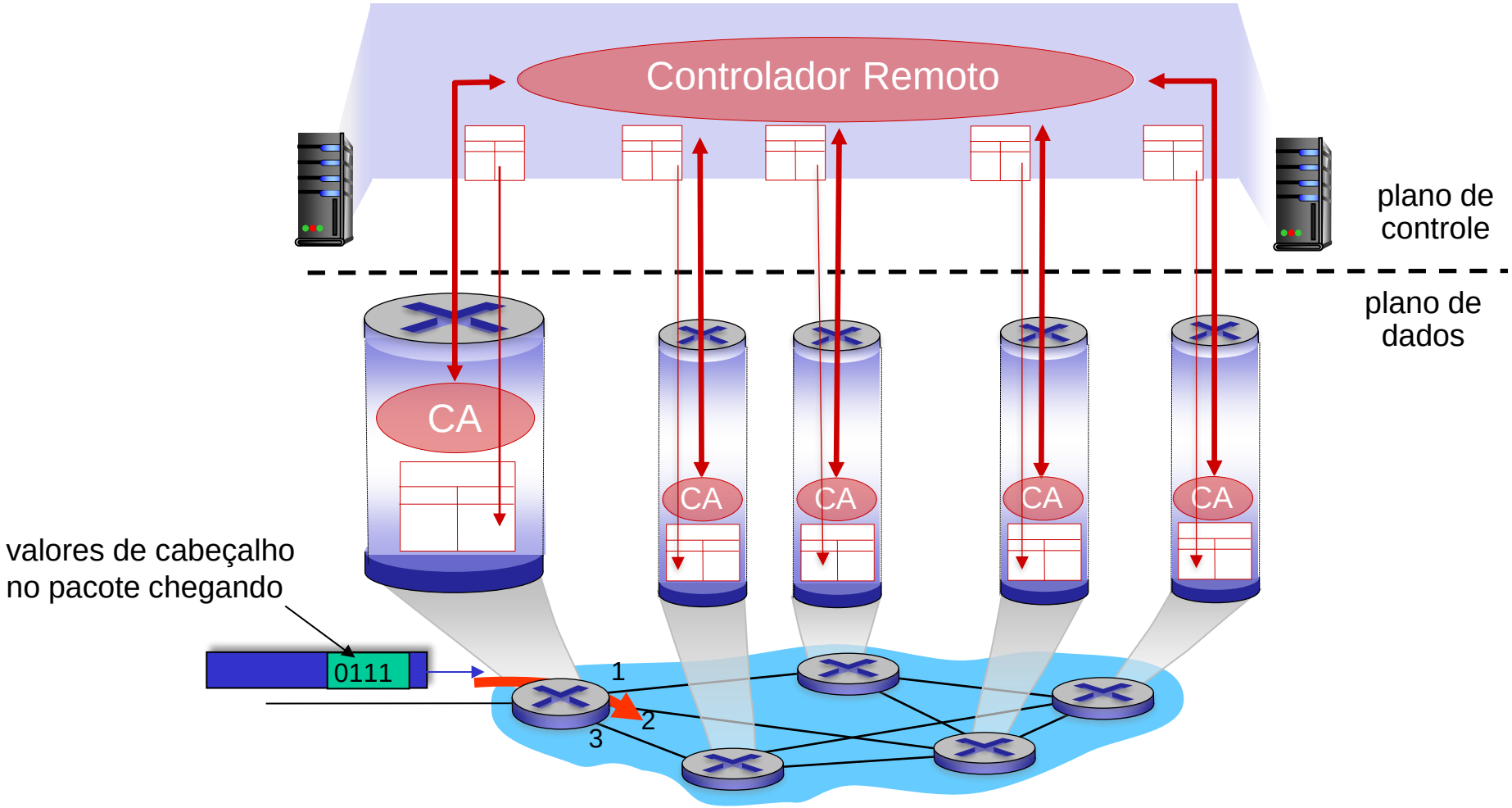
Plano de controle por roteador

Componentes individuais do algoritmo de roteamento *em todo e cada roteador* interagem no plano de controle



Plano de controle de Rede Defininida por Software (SDN)

Controlador remoto computa e instala tabelas de encaminhamento nos roteadores



Modelo de serviço de rede

Q: Qual *modelo de serviço* para "canal" transportando datagramas do emissor para o receptor?

exemplo de serviços para datagramas *individuais*:

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

exemplos de serviços para um *fluxo* de datagramas:

- entrega de datagramas em ordem
- largura de banda mínima garantida para fluxo
- restrições às alterações no espaçamento entre pacotes

Modelo de serviço da camada de rede

Arquitetura de rede	Modelo de serviço	Garantias de Qualidade de Serviço (QoS) ?			
		Banda	Perda	Ordem	Tempo
Internet	melhor esforço	nenhuma	não	não	não

Modelo de serviço de “melhor esforço” da Internet

Sem garantias sobre*:*

i. entrega de datagrama bem-sucedida para o destino
ii. tempo ou ordem de entrega
iii. largura de banda disponível para o fluxo fim-a-fim

Modelo de serviço da camada de rede

Arquitetura de rede	Modelo de serviço	Garantias de Qualidade de Serviço (QoS) ?			
		Banda	Perda	Ordem	Tempo
Internet	melhor esforço	nenhuma	não	não	não
ATM	taxa constante	taxa constante	sim	sim	sim
ATM	taxa disponível	tx. mín. garantida	não	sim	não
Internet	Intserv Guaranteed (RFC 1633)	sim	sim	sim	sim
Internet	Diffserv (RFC 2475)	possível	possivel-mente	possivel-mente	não

Reflexões sobre o serviço de melhor esforço:

- simplicidade do mecanismo permitiu que a Internet fosse amplamente implantada e adotada
- provisionamento de largura de banda suficiente permite que o desempenho de aplicações de tempo real (por exemplo, voz interativa, vídeo) seja "bom o suficiente" na "maior parte do tempo"
- serviços da camada de aplicação replicados e distribuídos (datacenters, redes de distribuição de conteúdo) conectando perto das redes dos clientes permitindo que os serviços sejam fornecidos a partir de múltiplos locais
- controle de congestionamento de serviços "elásticos" ajuda

É difícil argumentar com o sucesso do modelo de serviço de melhor esforço

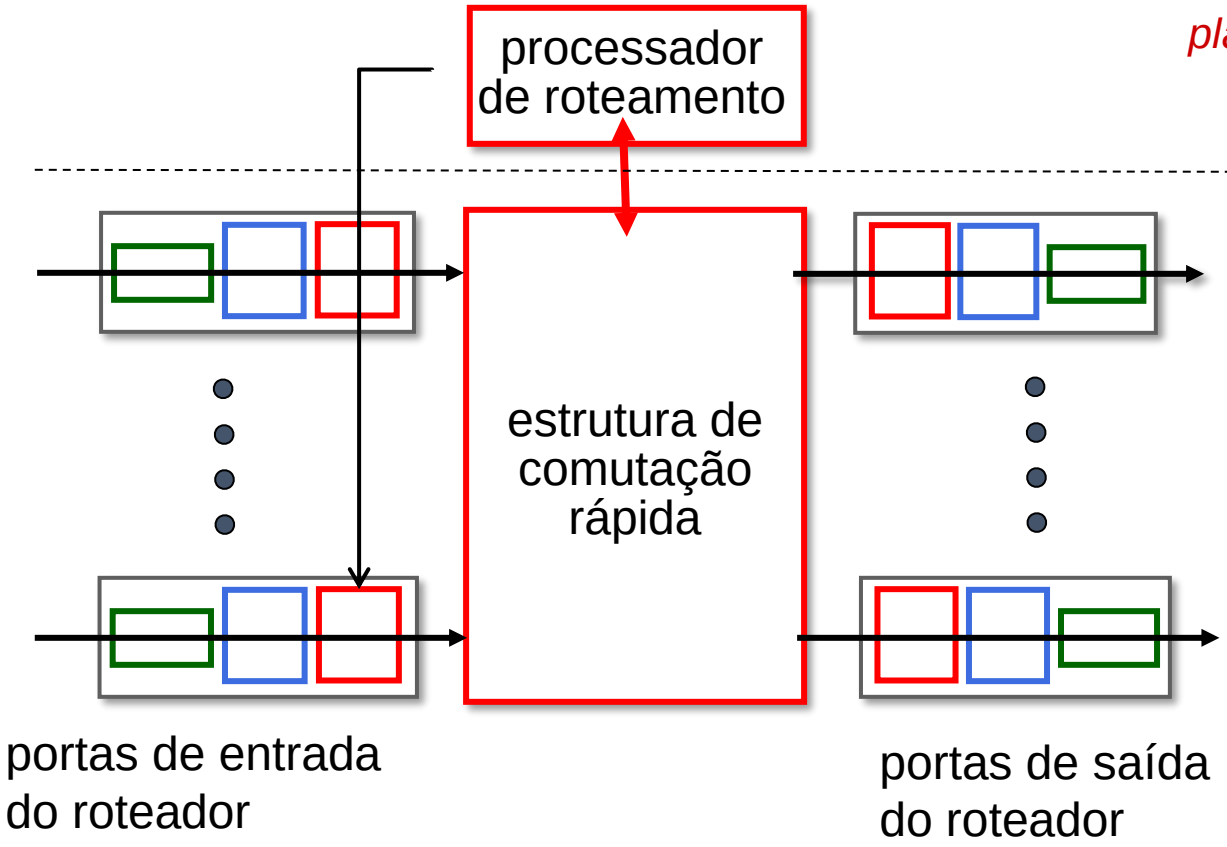
Camada de rede: roteiro do “plano de dados”

- 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 da Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereço de rede (NAT)
  - IPv6

- Encaminhamento Generalizado e SDN
  - Correspondência+ação
  - OpenFlow: correspondência+ação em ação
- Middleboxes

Visão geral da arquitetura do roteador

visão de alto nível de uma arquitetura de roteador genérica:

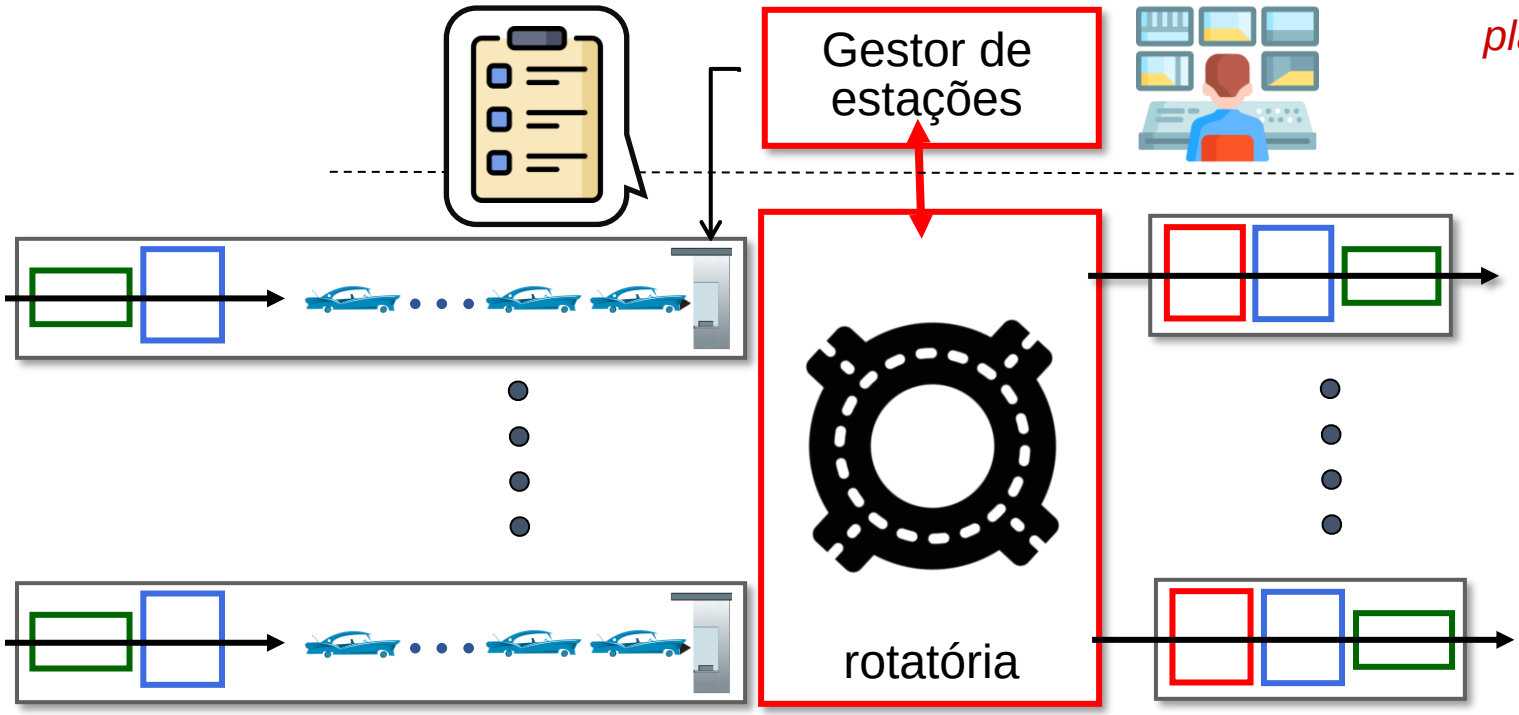


roteamento, gerenciamento plano de controle (software) opera em períodos de milissegundos

plano de encaminhamento de dados (hardware) opera em períodos de nanossegundos

Visão geral da arquitetura do roteador

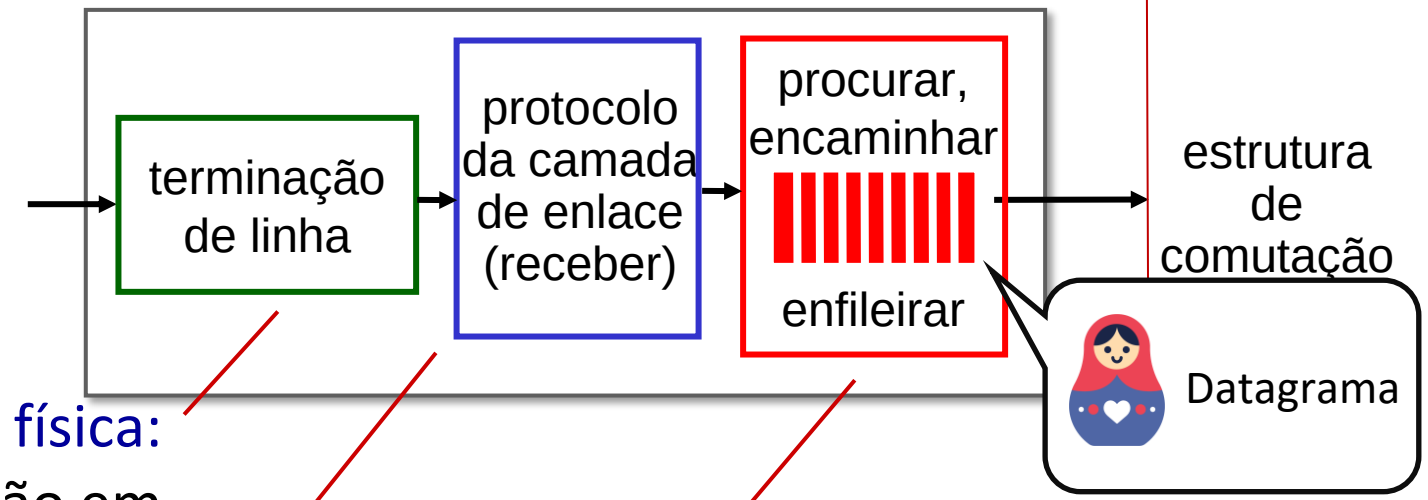
analogia da arquitetura genérica do roteador:



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

plano de encaminhamento de dados (hardware) opera em períodos de nanossegundos

Funções da porta de entrada



camada física: recepção em nível de bits

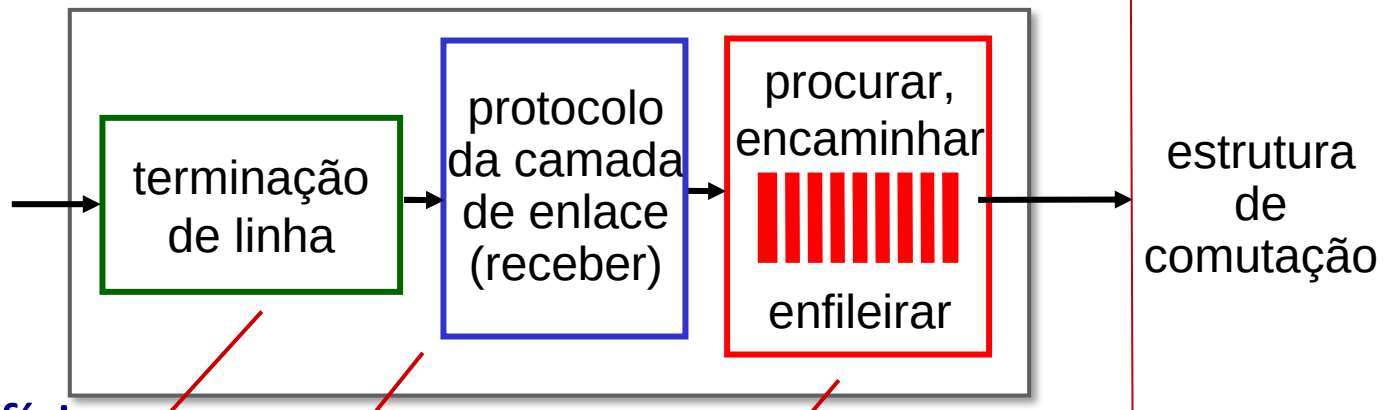
camada de enlace: ex.: Ethernet (capítulo 6

Quadro

comutação descentralizada:

- usando valores de campo de cabeçalho, procurar porta de saída usando tabela de encaminhamento na memória da porta de entrada *("correspondência mais ação")*
- objetivo: processamento completo da porta de entrada na 'velocidade da linha'
- enfileiramento na porta de entrada: ocorre se os datagramas chegarem mais rápido do que a taxa de encaminhamento para a estrutura de comutação

Funções da porta de entrada



camada física:
recepção em nível de bits

camada de enlace:
ex.: Ethernet (capítulo 6)

comutação descentralizada:

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

Q: mas o que acontece se os intervalos não se dividem tão bem?

Correspondência com prefixo mais longo

correspondência com prefixo mais longo

ao procurar a entrada da tabela de encaminhamento para determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponde ao endereço de destino.

Faixa de Endereços de Destino	Interface do enlace
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 com prefixo mais longo

correspondência com prefixo mais longo

ao procurar a entrada da tabela de encaminhamento para determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponde ao endereço de destino.

Faixa de Endereços de Destino	Interface do enlace
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001 011*** ********	2
caso contrário	3

correspondência!

exemplos:

11001000 00010111 00010110 10100001 qual interface?

11001000 00010111 00011000 10101010 qual interface?

Correspondência com prefixo mais longo

correspondência com prefixo mais longo

ao procurar a entrada da tabela de encaminhamento para determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponde ao endereço de destino.

Faixa de Endereços de Destino	Interface do enlace
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011*** ********	2
otherwise	3

correspondência!

exemplos:

11001000 00010111 00010110 10100001 qual interface?

11001000 00010111 00011000 10101010 qual interface?

Correspondência com prefixo mais longo

correspondência com prefixo mais longo

ao procurar a entrada da tabela de encaminhamento para determinado endereço de destino, use o prefixo de endereço *mais longo* que corresponde ao endereço de destino.

Faixa de Endereços de Destino	Interface do enlace
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011*** ********	2
otherw	3

correspondência!

exemplos:

11001000 00010111 00010110 10100001 qual interface?

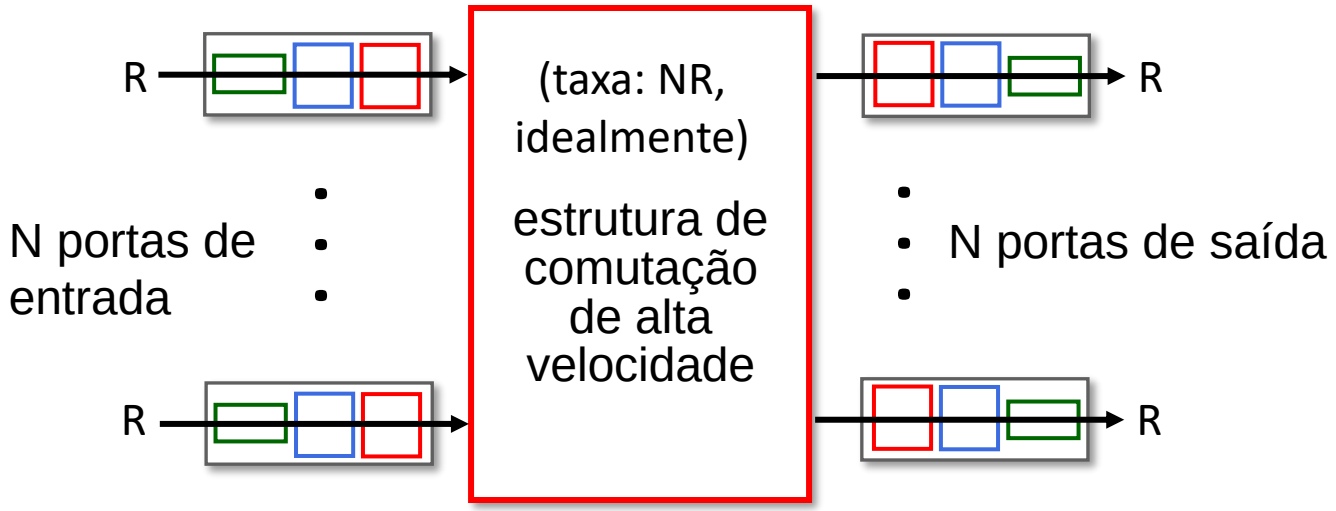
11001000 00010111 00011000 10101010 qual interface?

Correspondência com prefixo mais longo

- vamos ver *por que* a correspondência de prefixo mais longo é usada em breve, quando estudarmos endereçamento
- correspondência de prefixo mais longo: frequentemente realizada usando memórias endereçáveis de conteúdo ternário (*ternary content addressable memories* - TCAMs)
  - *conteúdo endereçável:* apresenta endereço para TCAM: recupera endereço em um ciclo de clock, independentemente do tamanho da tabela
  - Cisco Catalyst: ~1M entradas de tabela de roteamento em TCAM

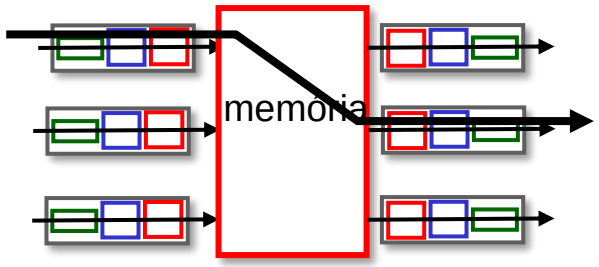
Estruturas de comutação

- transferem pacotes do enlace de entrada para o enlace de saída apropriado
- taxa de comutação: taxa em que pacotes podem ser transferidos de entradas para saídas
  - frequentemente medida como múltiplos da taxa de linha de entrada/saída
  - N entradas: taxa de comutação de N vezes a taxa de linha é desejável

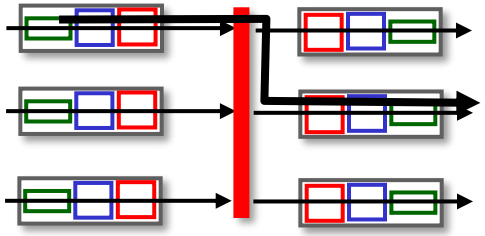


Estruturas de comutação

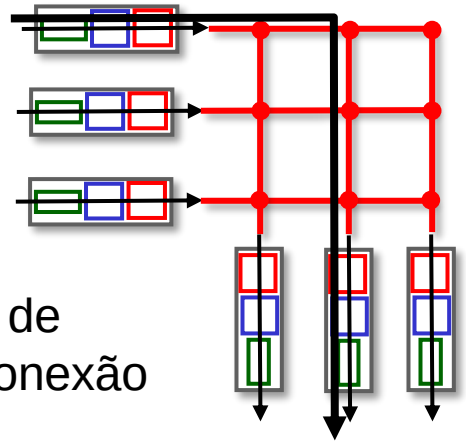
- transferem pacotes do enlace de entrada para o enlace de saída apropriado
- taxa de comutação: taxa em que pacotes podem ser transferidos de entradas para saídas
  - frequentemente medida como múltiplos da taxa de linha de entrada/saída
  - N entradas: taxa de comutação de N vezes a taxa de linha é desejável
- três principais tipos de estruturas de comutação:



memória



barramento

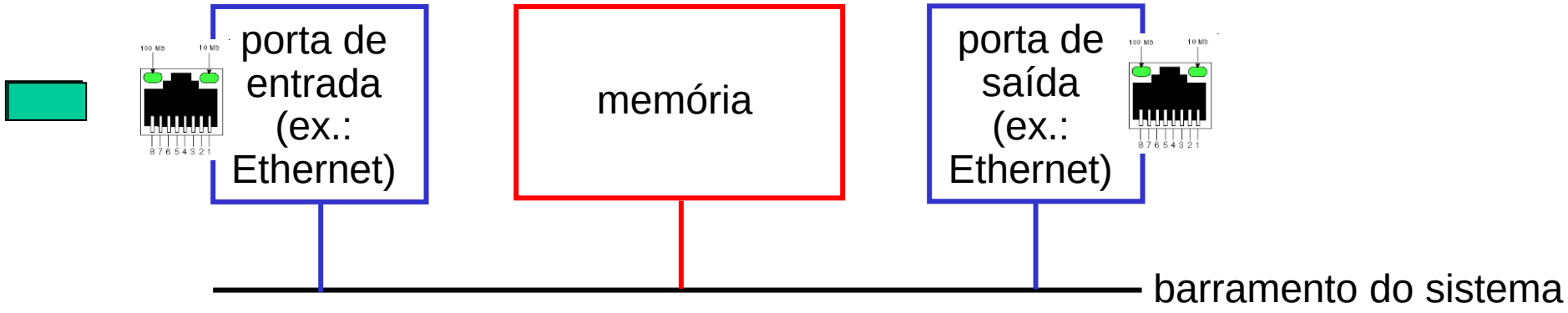


redes de interconexão

Comutação via memória

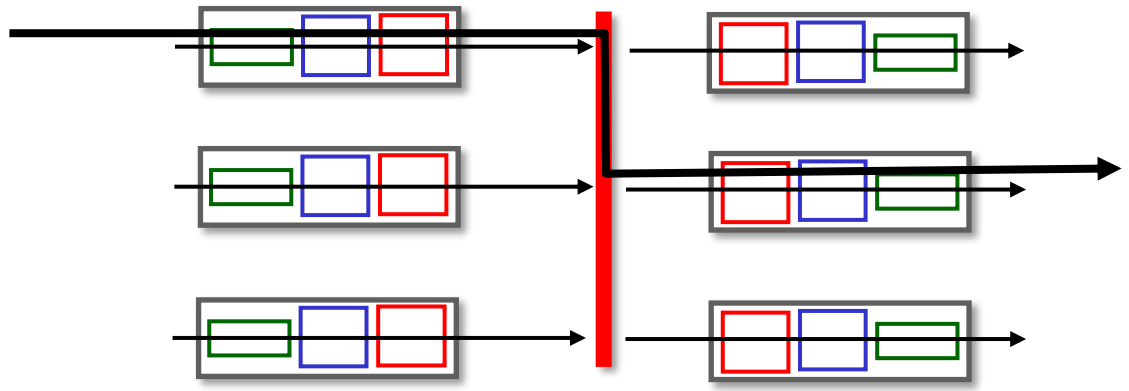
roteadores de primeira geração:

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



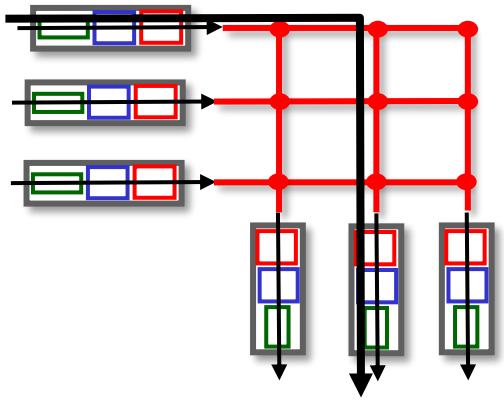
Comutação via barramento

- datagrama da memória da porta de entrada para a memória da porta de saída através 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

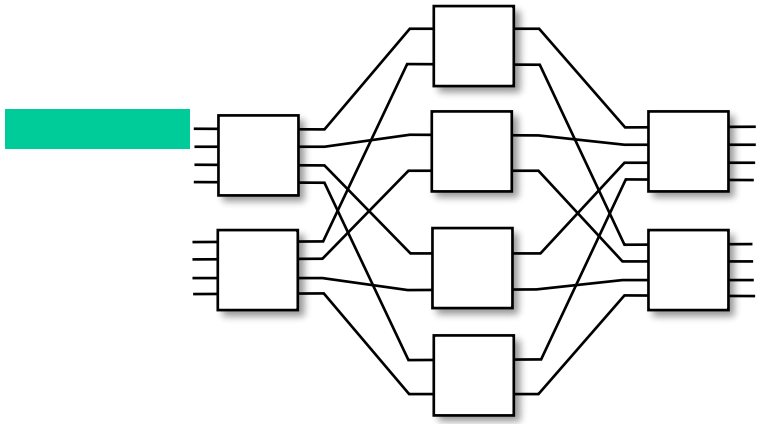


Comutação via rede de interconexão

- redes Crossbar, Clos e outras redes de interconexão foram inicialmente desenvolvidas para conectar processadores em multiprocessador
- comutador de vários estágios: comutador *nxn* composto por múltiplos estágios de comutadores menores
- explorando o paralelismo:
  - fragmenta datagrama em células de comprimento fixo na entrada
  - comuta células através da estrutura, remonta datagrama na saída



3x3 crossbar

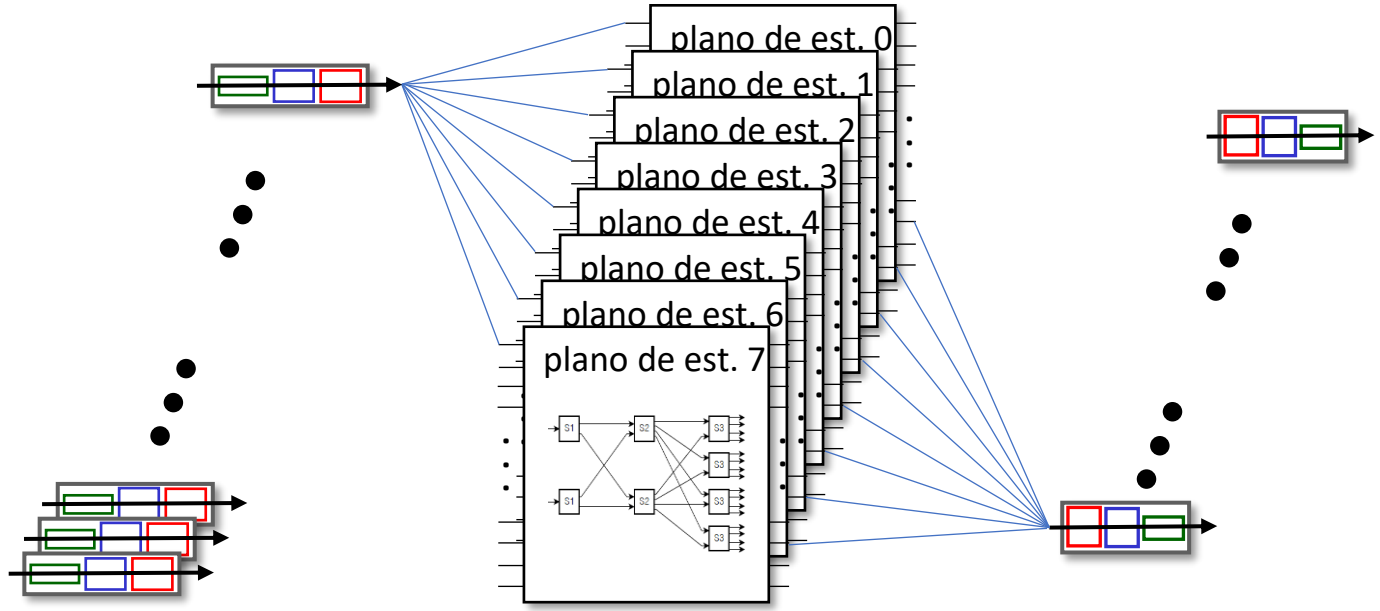


comutador multiestágio 8x8 construído a partir de comutadores de tamanho menor

Comutação via rede de interconexão

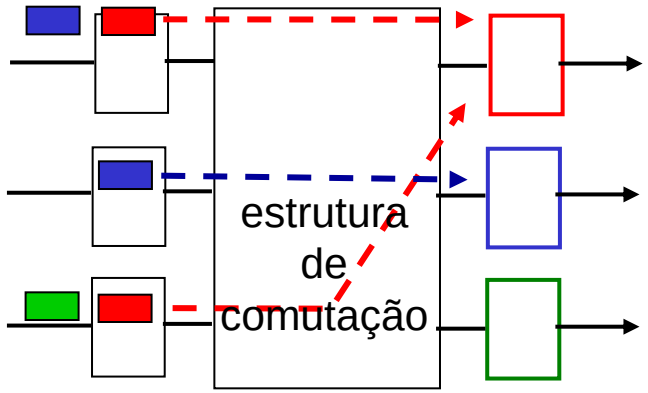
- escala usando vários “planos” de comutação em paralelo:
  - *speedup*, *scaleup* via paralelismo

- roteador Cisco CRS:
  - unidade básica: 8 planos de comutação
  - cada plano: rede de interconexão de 3 estágios
  - até centenas de Tbps de capacidade de comutação

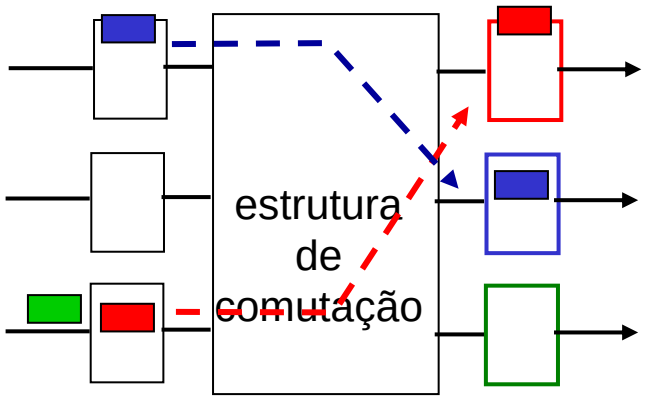


Fila da porta de entrada

- Se a estrutura de comutação for mais lenta do que as portas de entrada combinadas -> enfileiramento pode ocorrer nas filas de entrada
  - atraso na fila e perda devido ao estouro do buffer de entrada!
- Bloqueio de cabeça de linha (*Head-of-the-Line* - HOL): datagrama na primeira posição da fila impede que outros na fila avancem

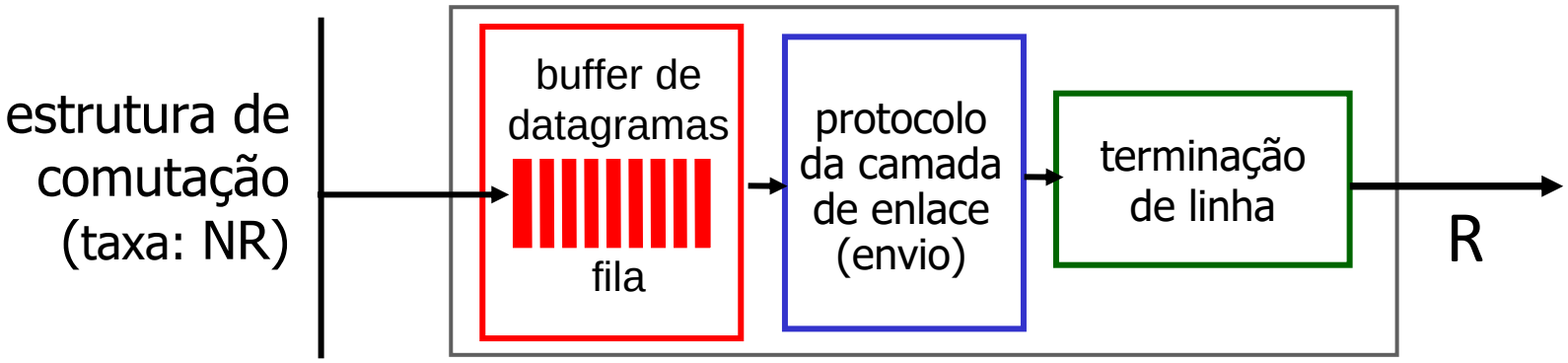


contenção da porta de saída: apenas um datagrama vermelho pode ser transferido. o pacote vermelho inferior está *bloqueado*



um pacote depois: pacote verde experimenta bloqueio de cabeça de linha

Fila da porta de saída



Este é um slide muito importante

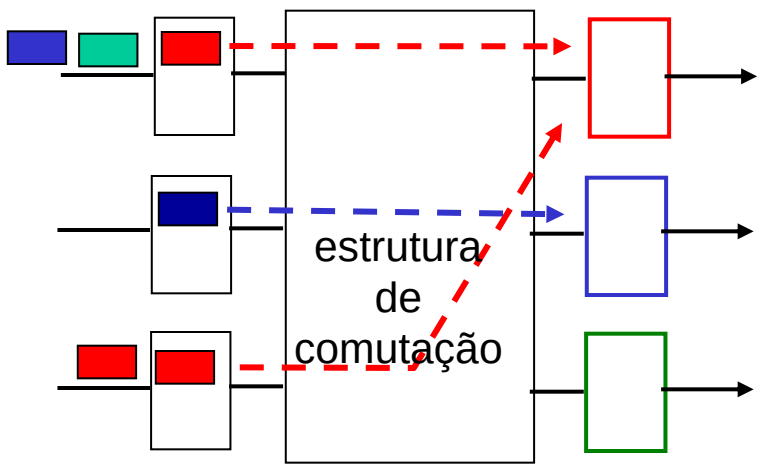
- *Buffering* necessário quando os datagramas chegam da estrutura de comutação mais rápido do que a taxa de transmissão de enlace. *Política de perda:* quais datagramas descartar se não houver buffers disponíveis?

  → Datagramas podem ser perdidos devido ao congestionamento, falta de buffers

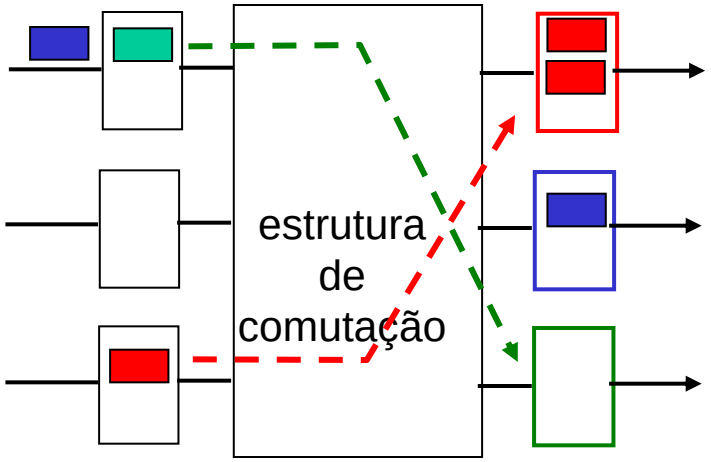
- *Disciplina de agendamento* escolhe entre datagramas enfileirados para transmissão

  → Agendamento prioritário – quem obtém melhor desempenho, neutralidade da rede

Fila da porta de saída



em *t*, mais que um pacote de entrada para a mesma saída



um tempo de pacote depois

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

Quanto espaço em buffer?

- Regra de ouro da RFC 3439: *buffering* médio igual a RTT "típico" (digamos 250 ms) vezes capacidade do enlace C
  - ex.: C = enlace de 10 Gbps: 2,5 Gbit de buffer
- recomendação mais recente: com *N* fluxos, *buffering* igual a

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

- mas buffer *demais* pode aumentar atrasos (particularmente em roteadores domésticos)
  - RTTs longos: desempenho ruim para aplicações em tempo real, resposta TCP lenta
  - lembre-se do controle de congestionamento baseado em retardo: "mantenha o enlace de gargalo cheio o suficiente (ocupado), mas não mais cheio que isso"

- Um caso: Peplink Balance One
  - Buffer padrão: 2000 pacotes
    - Considerando um pacote de 1500 bytes, seriam 22,89 Mbits de buffer.
    - Com upload de 100 Mbps, são 229 ms de buffer.
    - Com upload de 2 Mbps, são 11 segundos de buffer!
    - Com upload de 600 Kbps, são 38 segundos de buffer!
      - Um único upload (ex.: subir um vídeo para o Youtube) deixava a conexão inutilizável.

BUFFERBLOAT GRADE

A+

Your latency did not increase under load.

LATENCY

Unloaded	Download Active	Upload Active
10 ms	+2 ms	+0 ms

YOUR CONNECTION

	Under Ideal Conditions	Currently, Due To Bufferbloat
Web Browsing	✓	✓
Audio Calls	✓	✓
4K Video Streaming	✓	✓
Video Conferencing	✓	✓
Low Latency Gaming	✓	✓

Read More

SPEED

↓ Download	↑ Upload
287.0 Mbps	158.6 Mbps

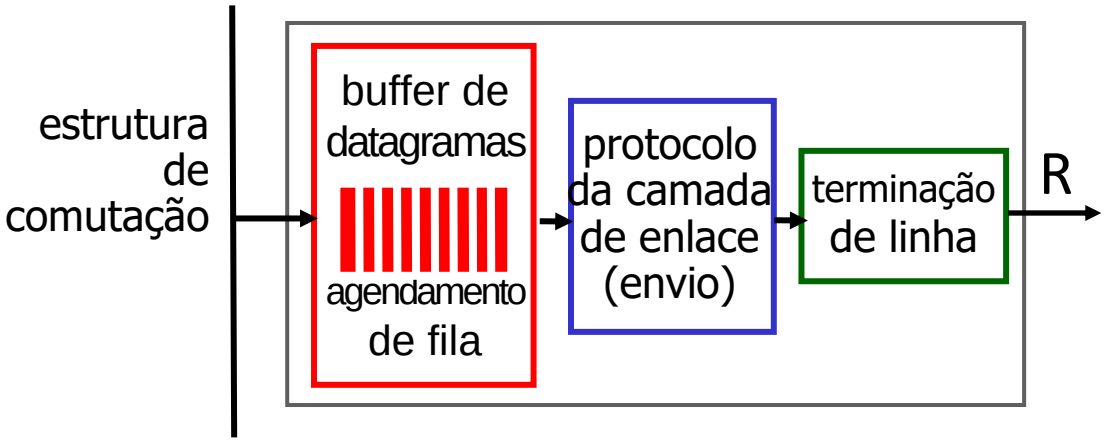
Unloaded	Min: 7.5 ms Median: 8.9 ms Max: 27.7 ms Mean: 10 ms	25th %ile: 8.4 ms 75th %ile: 10 ms 95th %ile: 18.6 ms Jitter: 2 ms
↓ Active	Min: 5.2 ms Median: 8 ms Max: 253.8 ms Mean: 11.9 ms	25th %ile: 6.9 ms 75th %ile: 10.4 ms 95th %ile: 28.7 ms Jitter: 6.7 ms
↑ Active	Min: 5.2 ms Median: 8.4 ms Max: 65.9 ms Mean: 10 ms	25th %ile: 7 ms 75th %ile: 9.6 ms 95th %ile: 19.6 ms Jitter: 3.7 ms

Share Your Results: https://www.waveform.com/tools/bufferbloat?test-id=258b5af8-49dd-48fe-8143-adeee! Copy

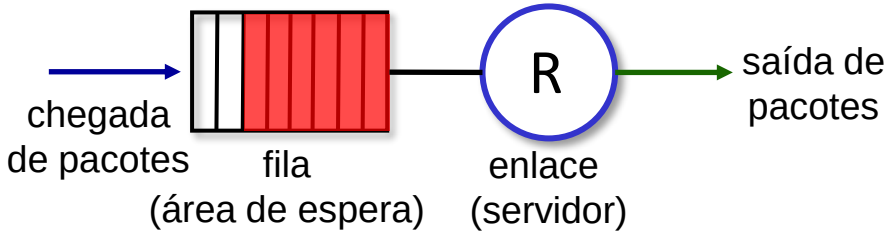
Test Again

https://www.waveform.com/tools/bufferbloat?test-id=258b5af8-49dd-48fe-8143-adeee5798d5f

Gerenciamento de buffer



Abstração: fila



gerenciamento de buffer:

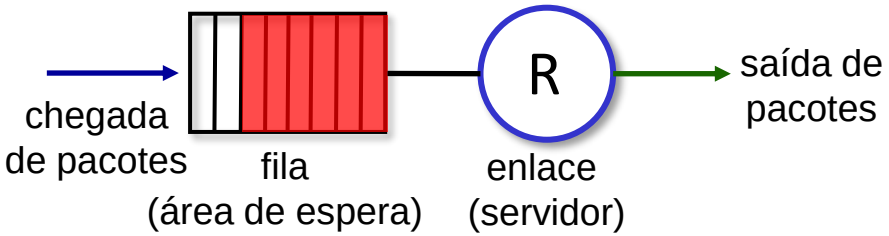
- perda: quais pacotes adicionar/descartar quando buffers estão cheios
  - tail drop: descarta pacote que está chegando
  - priority: descarta/remove com base em prioridades
- marcação: quais pacotes marcar para sinalizar congestionamento (ECN, RED)

Agendamento de pacotes: FCFS

agendamento de pacotes:
decidir qual pacote enviar a seguir no enlace

- first come, first served (primeiro a chegar, primeiro a ser servido)
- prioridade
- round robin (rodízio)
- weighted fair queueing (fila justa ponderada)

Abstração: fila



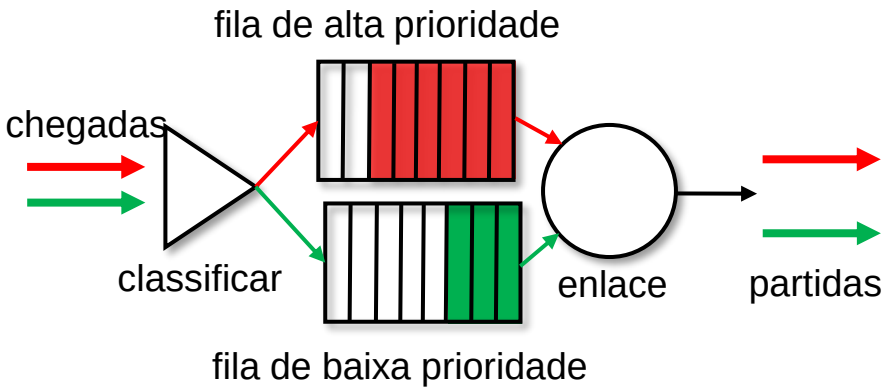
FCFS: pacotes transmitidos em ordem de chegada à porta de saída

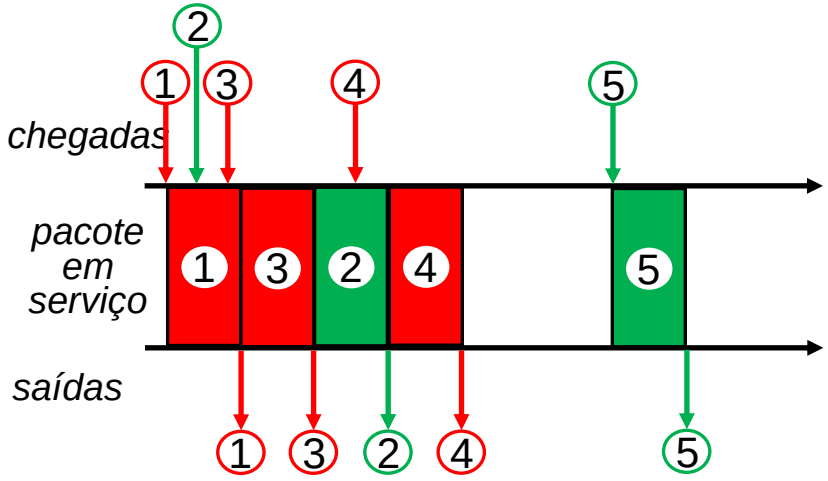
- também conhecido como: First-in-first-out (FIFO)
- exemplos do mundo real?

Políticas de agendamento: prioridade

Agendamento prioritário:

- tráfego chegando é classificado e enfileirado por classe
  - quaisquer campos de cabeçalho podem ser usados para classificação
- envia pacote da fila de prioridade mais alta que tem pacotes em buffer
  - FCFS é usado dentro da classe prioritária

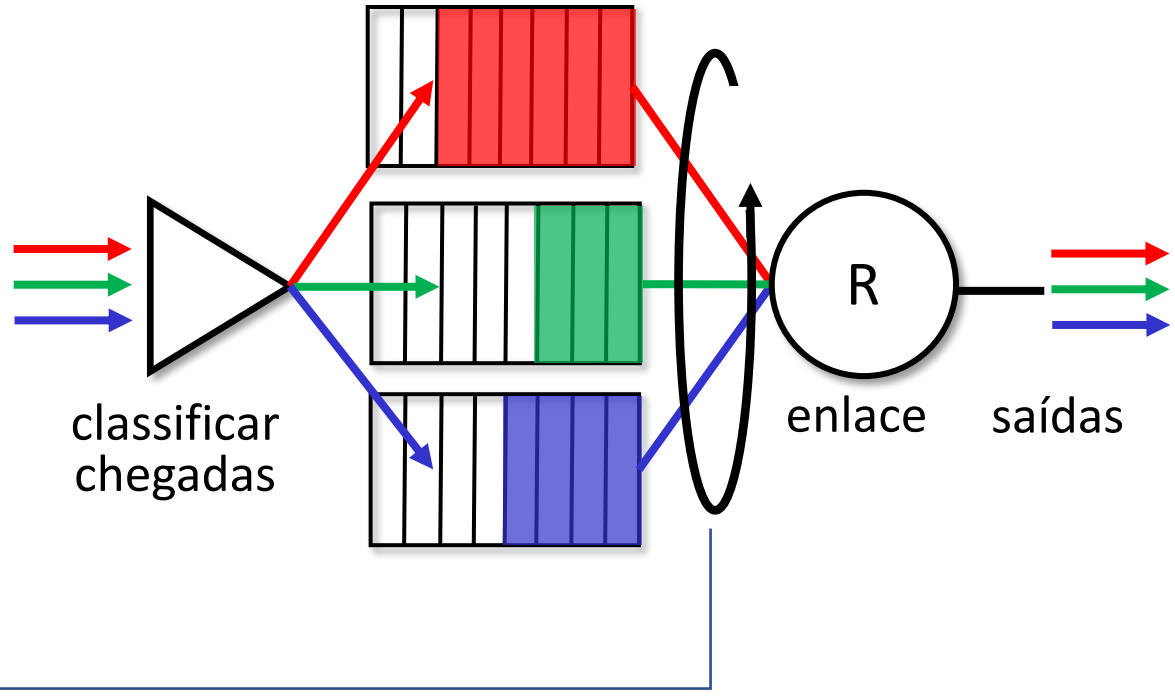




Políticas de agendamento: round robin

agendamento Round Robin (RR - rodízio):

- tráfego chegando é classificado e enfileirado por classe
  - quaisquer campos de cabeçalho podem ser usados para classificação
- o servidor verifica ciclicamente e repetidamente as filas de classes, enviando um pacote completo de cada classe (se disponível) por vez



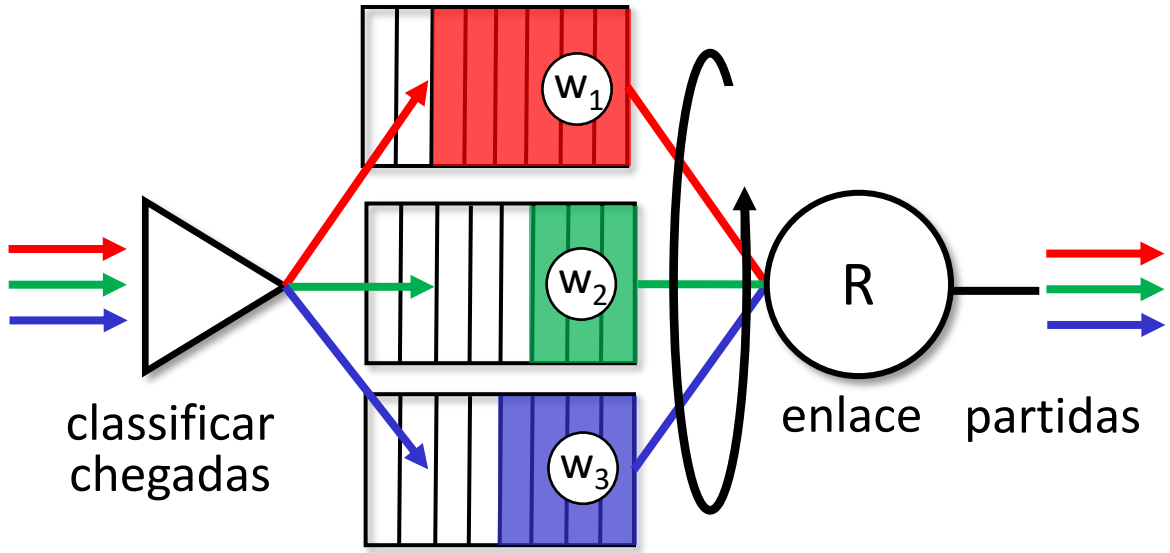
Políticas de agendamento: weighted fair queueing

Weighted Fair Queuing (WFQ – fila justa ponderada):

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

$$\frac{w_i}{\sum_j w_j}$$

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



QoS Adaptativo

No QoS adaptativos, os aplicativos com maior prioridade obterão melhor largura de banda que outros quando os APPs estão competindo entre si pela largura de banda limitada, arraste e solte os aplicativos de nível de categoria superior para maior prioridade.

Mais alto

- Jogos
- Transmissão de Vídeo e Áudio
- Work-From-Home
- Learn-From-Home
- Surfar na Web
- Transferência de Arquivo
- Outros

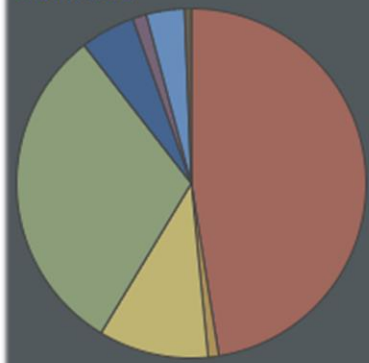
Mais baixo

Cancelar | Gravar

Traffic classification

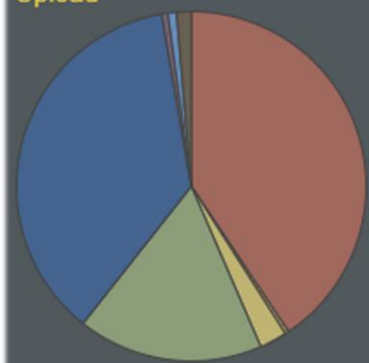
Automatically refresh data every: 3 seconds

Download



Class	Total	Rate	Packet rate
Rede	67.82 GB	40,304 bit	13 pps
Jogos	1.40 GB	16 bit	0 pps
Transmissão de Vídeo e Áudio	14.54 GB	663,472 bit	61 pps
Work-From-Home	44.07 GB	153,960 bit	20 pps
Learn-From-Home *(Default)*	7.24 GB	47,584 bit	7 pps
Surfar na Web	1.84 GB	368 bit	0 pps
Transferência de Arquivo	5.04 GB	0 bit	0 pps
Outros	908.32 MB	3,264 bit	2 pps

Upload



Class	Total	Rate	Packet rate
Rede	4.77 GB	47,720 bit	15 pps
Jogos	42.33 MB	24 bit	0 pps
Transmissão de Vídeo e Áudio	304.37 MB	8,752 bit	8 pps
Work-From-Home	2.02 GB	26,216 bit	13 pps
Learn-From-Home *(Default)*	4.31 GB	96,736 bit	11 pps
Surfar na Web	67.69 MB	320 bit	0 pps
Transferência de Arquivo	95.47 MB	0 bit	0 pps
Outros	162.92 MB	2,896 bit	2 pps

Neutralidade de Rede

O que é neutralidade de rede?

- *técnico:* 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
  - encorajamento da inovação e da competição
- regras e políticas *legais* aplicadas

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

Neutralidade de Rede

Em 2015, a FCC dos EUA publicou o documento *Order on Protecting and Promoting an Open Internet* com três regras claras:

- sem bloqueio … "não bloqueará conteúdo legal, aplicações, serviços ou dispositivos não prejudiciais, sujeitos a um gerenciamento razoável de rede."
- sem estrangulamento … "não prejudicará ou degradará o tráfego legal da Internet com base no conteúdo, aplicação ou serviço da Internet ou no uso de um dispositivo não prejudicial, sujeito a um gerenciamento razoável de rede."
- nenhuma priorização paga. … "não deve se envolver em priorização paga"

Em 2017 um novo documento da FCC removeu as proibições, substituindo-as por princípios de transparência.

ISP: serviço de telecomunicações ou informações?

Um provedor de internet é um “serviço de telecomunicações” ou um provedor de “serviço de informação”?

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

Lei de Telecomunicações dos EUA de 1934 e 1996:

- *Título II:* impõe “deveres comuns de transporte” em *serviços de telecomunicações*: taxas razoáveis, não discriminação e requer regulação
- *Título I:* aplica-se a *serviços de informação:*
  - não há deveres comuns de transporte (não regulamentado)
  - mas concede autoridade à FCC “... conforme seja necessário na execução de suas funções”

Neutralidade da Rede no Brasil

- Marco Civil da Internet, sancionado em 2014:

Art. 3º A disciplina do uso da internet no Brasil tem os seguintes princípios:
I - garantia da liberdade de expressão, comunicação e manifestação de pensamento, nos termos da Constituição Federal;
II - proteção da privacidade;
III - proteção dos dados pessoais, na forma da lei;
IV - preservação e garantia da neutralidade de rede;
V - preservação da estabilidade, segurança e funcionalidade da rede, por meio de medidas técnicas compatíveis com os padrões internacionais e pelo estímulo ao uso de boas práticas;
VI - responsabilização dos agentes de acordo com suas atividades, nos termos da lei;
VII - preservação da natureza participativa da rede;
VIII - liberdade dos modelos de negócios promovidos na internet, desde que não conflitem com os demais princípios estabelecidos nesta Lei.
Parágrafo único. Os princípios expressos nesta Lei não excluem outros previstos no ordenamento jurídico pátrio relacionados à matéria ou nos tratados internacionais em que a República Federativa do Brasil seja parte.

Art. 9º O responsável pela transmissão, comutação ou roteamento tem o dever de tratar de forma isonômica quaisquer pacotes de dados, sem distinção por conteúdo, origem e destino, serviço, terminal ou aplicação.
§ 1º A discriminação ou degradação do tráfego será regulamentada nos termos das atribuições privativas do Presidente da República previstas no inciso IV do art. 84 da Constituição Federal, para a fiel execução desta Lei, ouvidos o Comitê Gestor da Internet e a Agência Nacional de Telecomunicações, e somente poderá decorrer de:
I - requisitos técnicos indispensáveis à prestação adequada dos serviços e aplicações; e
II - priorização de serviços de emergência.
§ 2º Na hipótese de discriminação ou degradação do tráfego prevista no § 1º , o responsável mencionado no **caput** deve:
I - abster-se de causar dano aos usuários, na forma do art. 927 da Lei nº 10.406, de 10 de janeiro de 2002 - Código Civil;
II - agir com proporcionalidade, transparência e isonomia;
III - informar previamente de modo transparente, claro e suficientemente descritivo aos seus usuários sobre as práticas de gerenciamento e mitigação de tráfego adotadas, inclusive as relacionadas à segurança da rede; e
IV - oferecer serviços em condições comerciais não discriminatórias e abster-se de praticar condutas anticoncorrenciais.
§ 3º Na provisão de conexão à internet, onerosa ou gratuita, bem como na transmissão, comutação ou roteamento, é vedado bloquear, monitorar, filtrar ou analisar o conteúdo dos pacotes de dados, respeitado o disposto neste artigo.

http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm

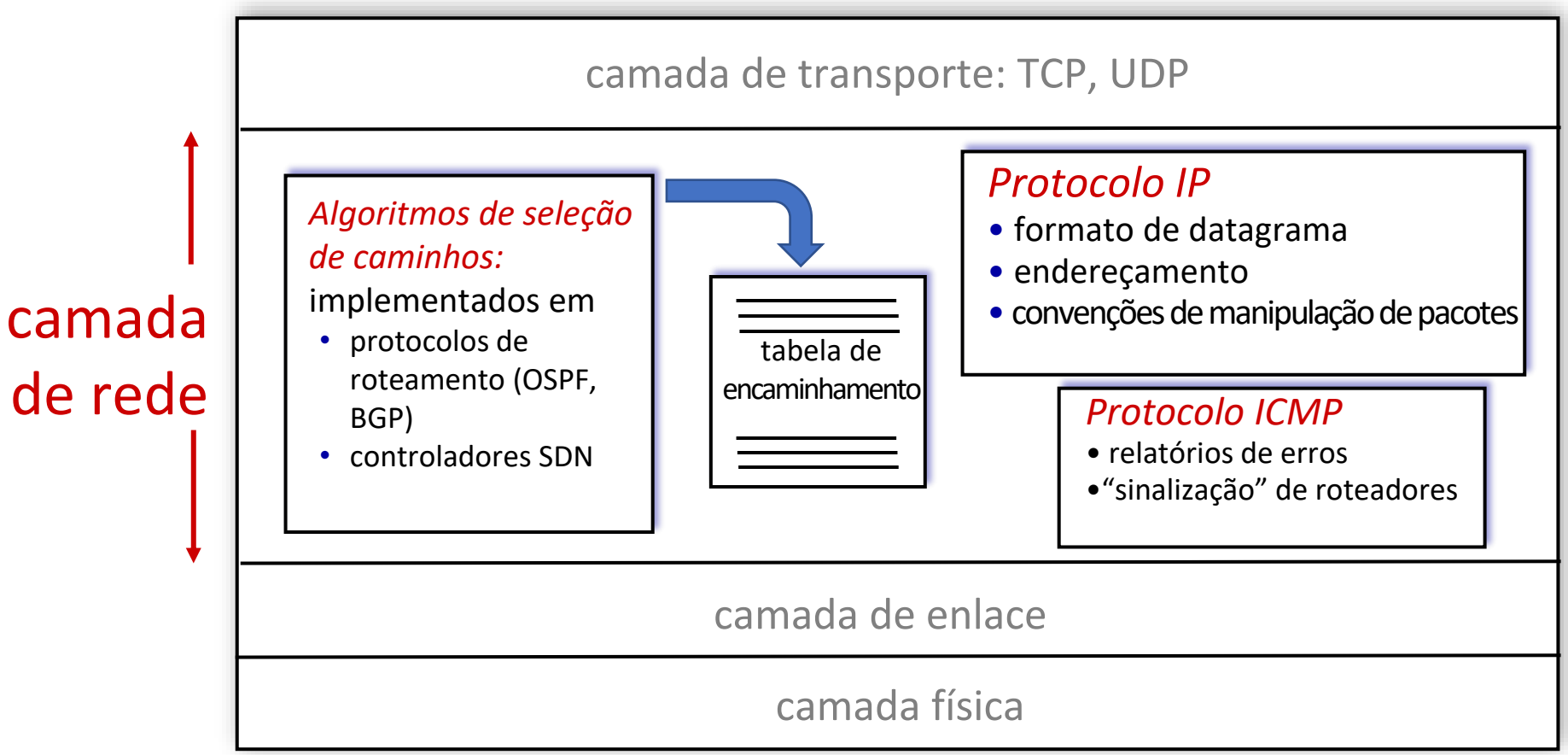
Camada de rede: roteiro do “plano de dados”

- 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 da Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereço de rede (NAT)
  - IPv6

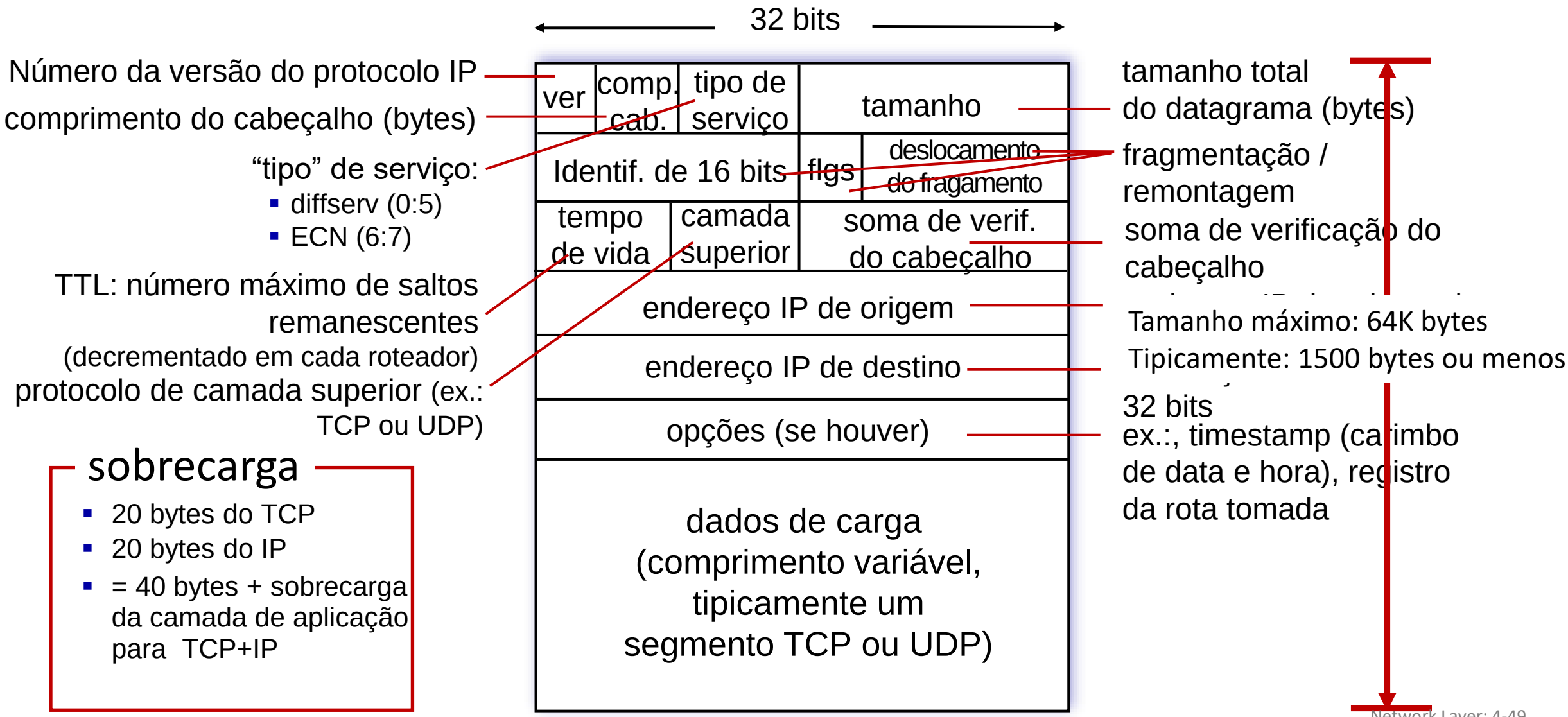
- Encaminhamento Generalizado e SDN
  - Correspondência+ação
  - OpenFlow: correspondência+ação em ação
- Middleboxes

Camada de Rede: Internet

funções de camada de rede de hospedeiros e roteadores:



Formato do Datagrama IP

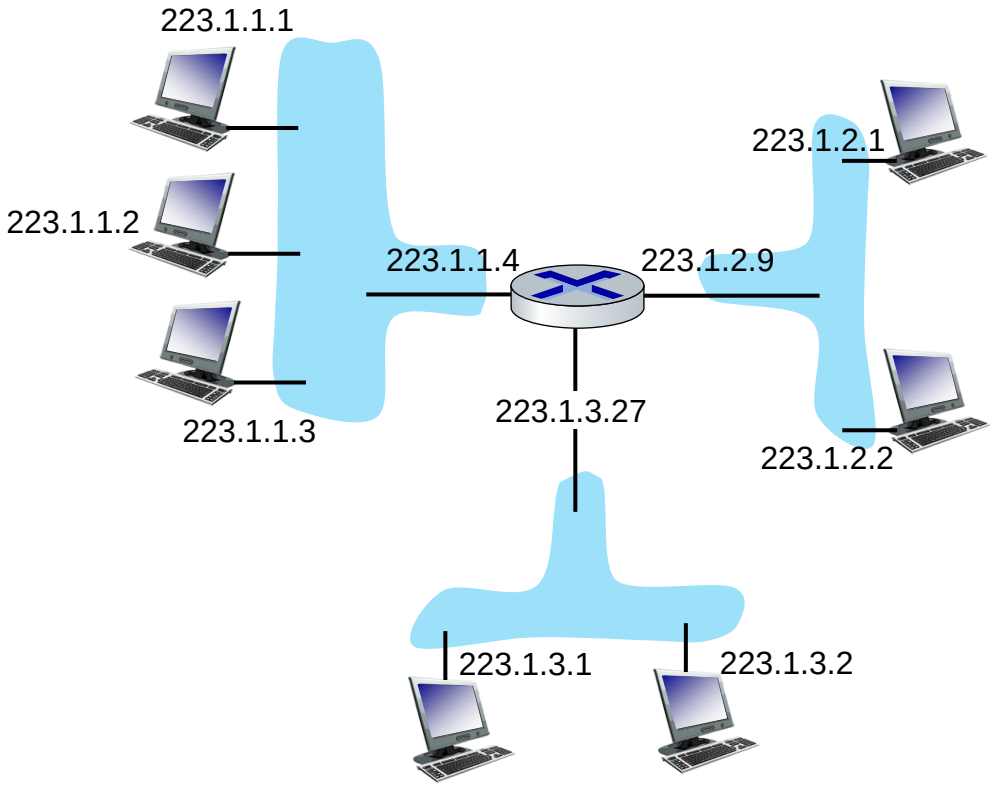


sobrecarga

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

Endereçamento IP: introdução

- endereço IP: identificador de 32 bits associado a cada *interface* de hospedeiro ou roteador
- interface: conexão entre hospedeiro / roteador e enlace físico
  - roteador normalmente tem várias interfaces
  - hospedeiro tipicamente tem uma ou duas interfaces (ex.: Ethernet cabeada, 802.11 sem fio)



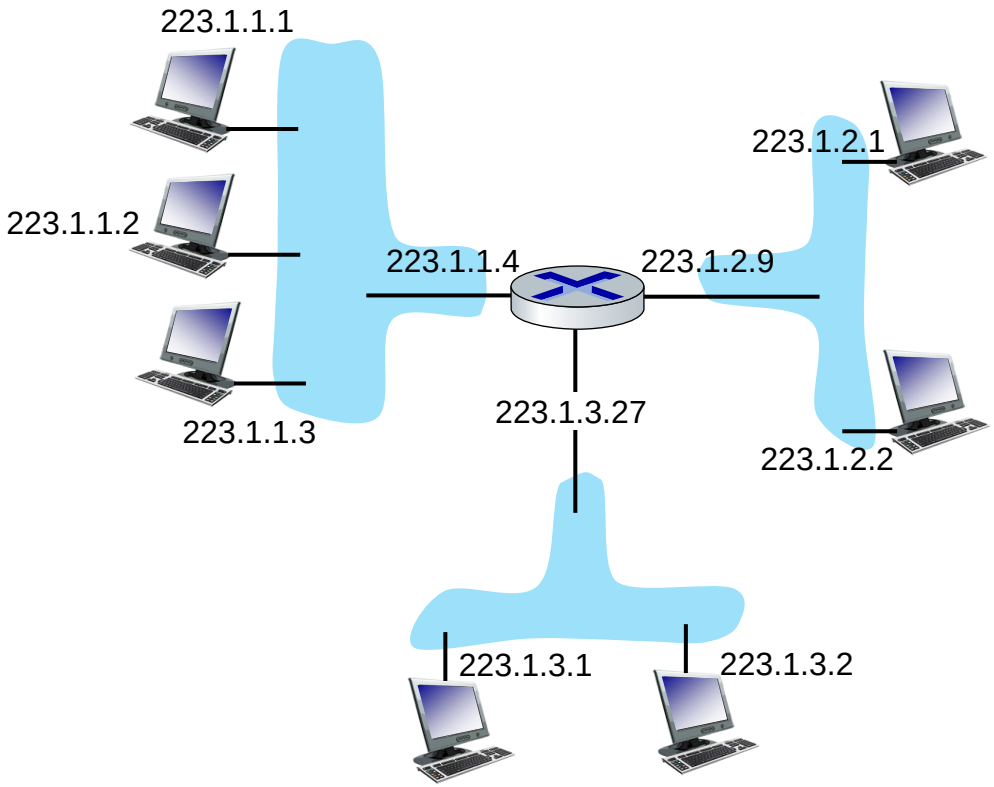
notação decimal com pontos de endereço IP:

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 hospedeiro ou roteador
- interface: conexão entre hospedeiro / roteador e enlace físico
  - roteador normalmente tem várias interfaces
  - hospedeiro tipicamente tem uma ou duas interfaces (ex.: Ethernet cabeada, 802.11 sem fio)



notação decimal com pontos de endereço IP:

223.1.1.1 = 11011111 00000001 00000001 00000001

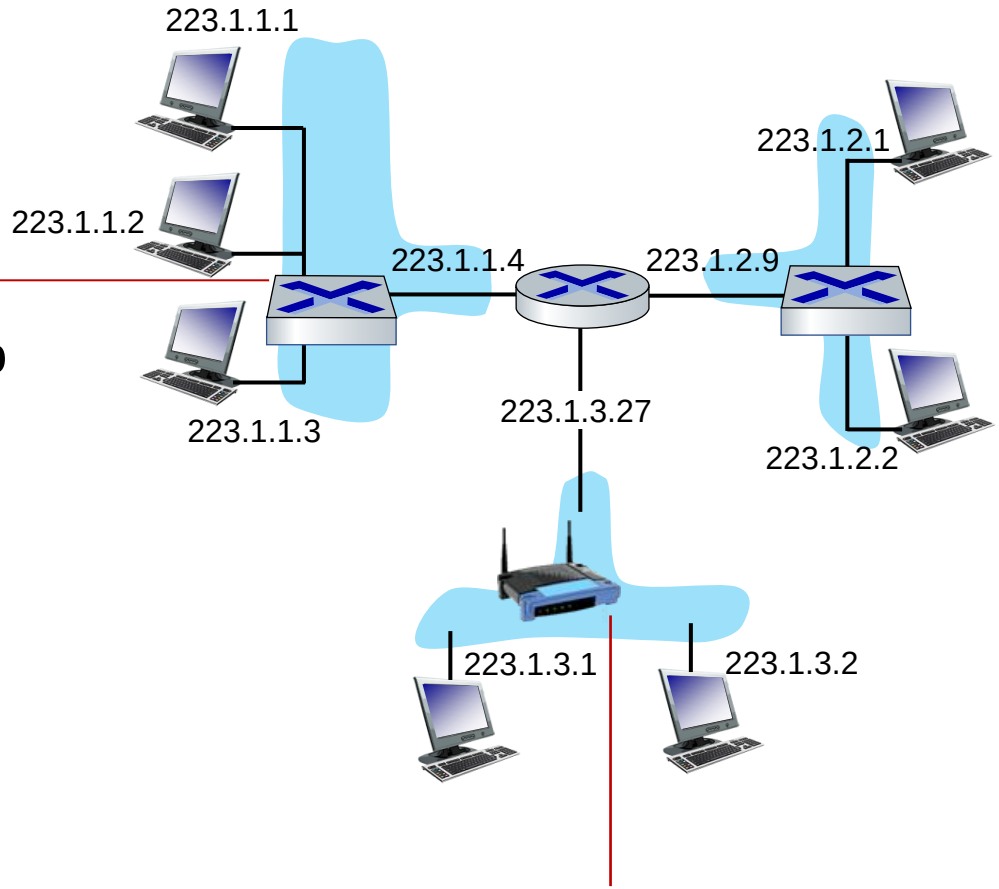
223 1 1 1

Endereçamento IP: introdução

Q: como as interfaces estão realmente conectadas?

R: vamos aprender sobre isso nos capítulos 6, 7

R: interfaces Ethernet com fio conectadas por comutadores Ethernet

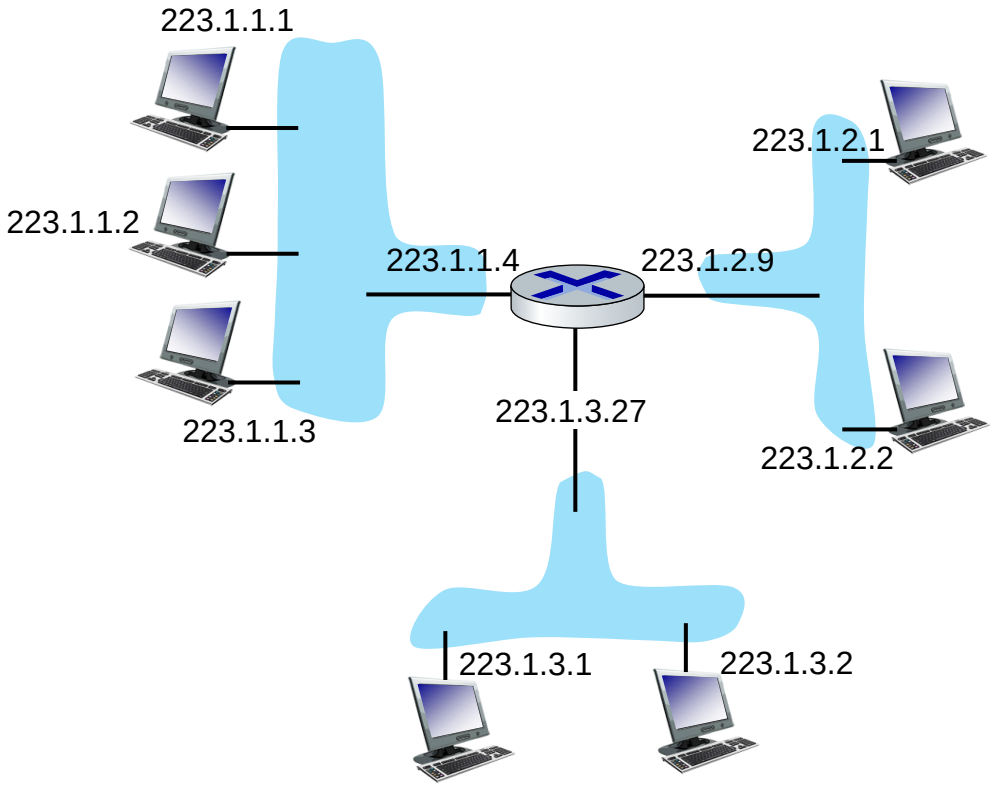


Por enquanto: não se preocupe sobre como uma interface está conectada a outra (sem nenhum roteador intervindo)

R: interfaces sem fio WiFi conectadas por estação base WiFi

Sub-redes

- *O que é uma sub-rede?*
  - interfaces de dispositivo que podem se alcançar fisicamente sem passar por um roteador intermediário
- Endereços IP têm a estrutura:
  - parte da sub-rede: dispositivos na mesma sub-rede têm bits comuns de alta ordem
  - parte do hospedeiro: bits de baixa ordem remanescentes

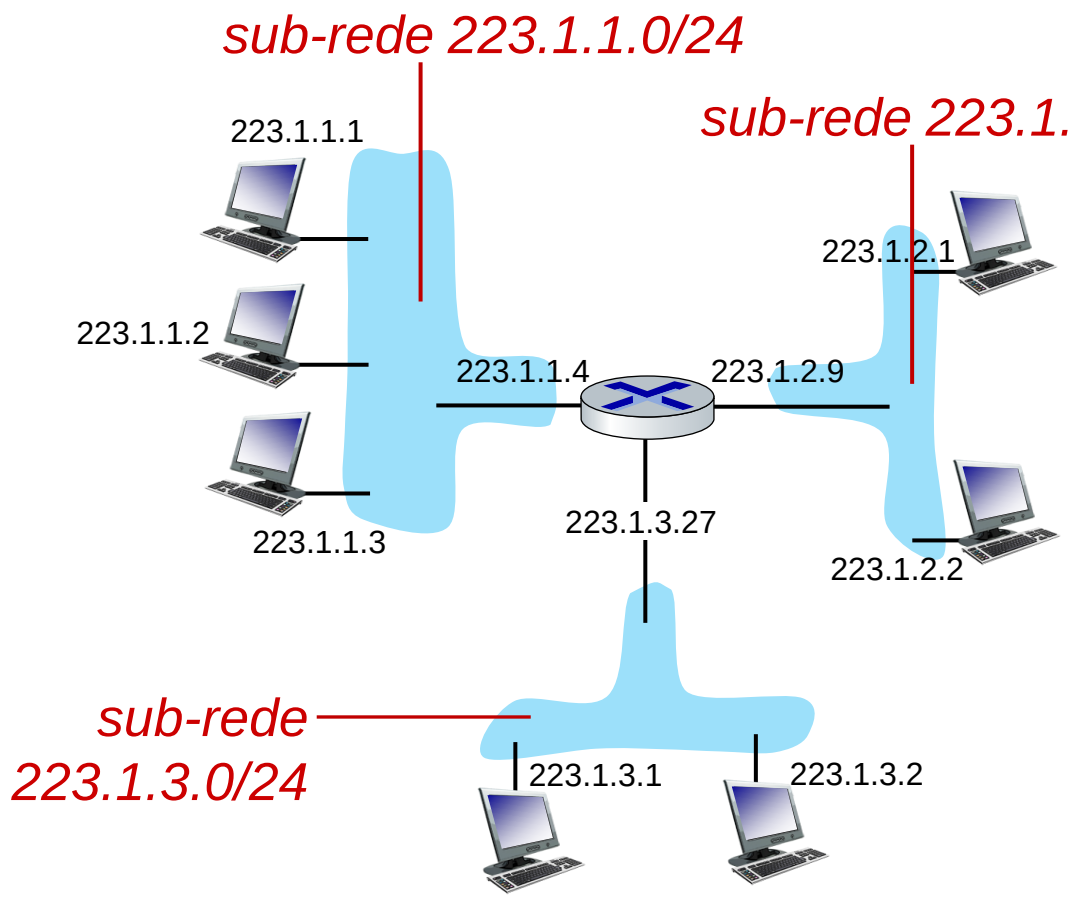


rede composta por 3 sub-redes

Sub-redes

Receita para definição de sub-redes:

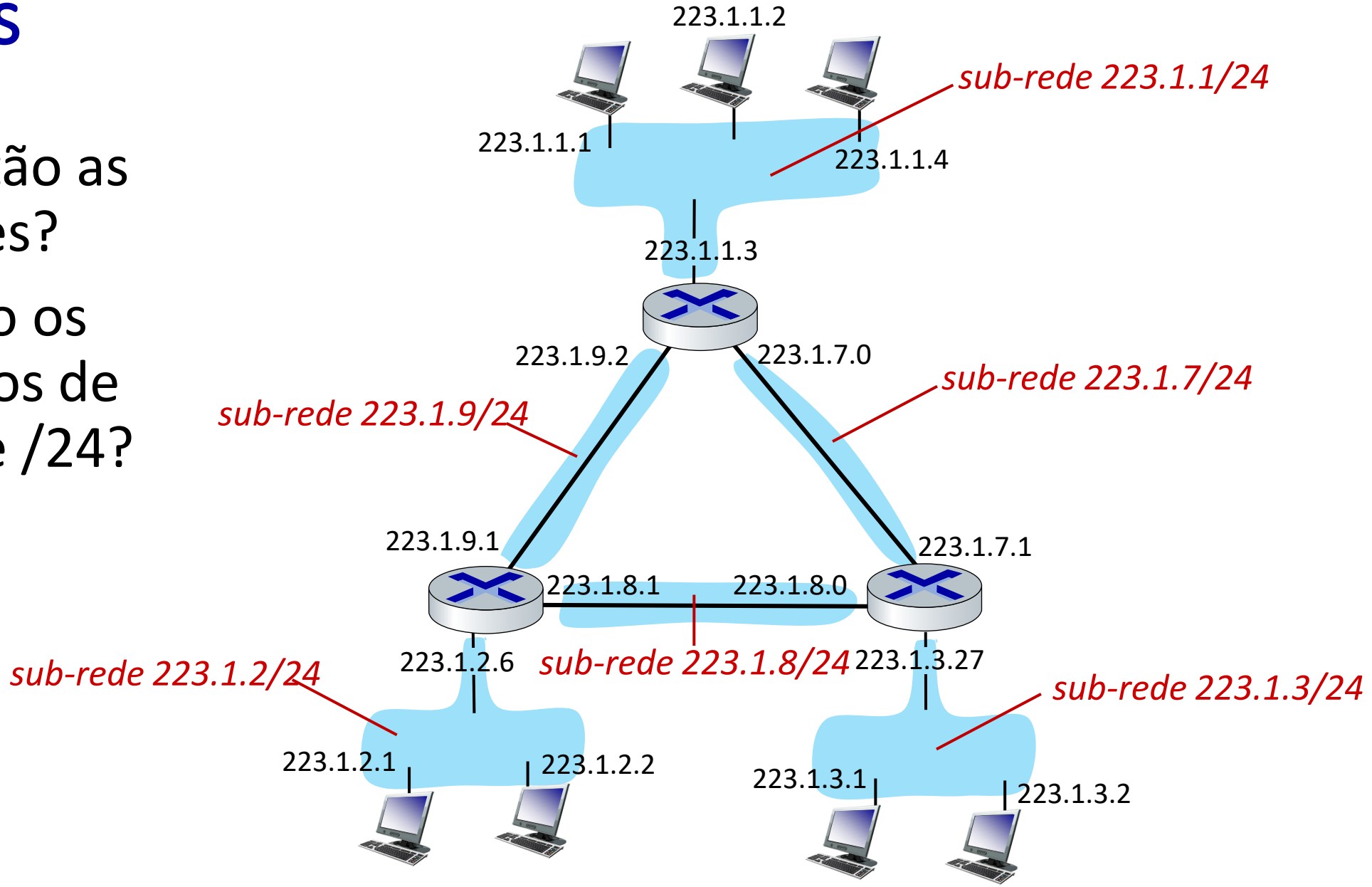
- desconecte cada interface de seu hospedeiro ou roteador, criando "ilhas" de redes isoladas
- cada rede isolada é chamada de *sub-rede*



Máscara de sub-rede: /24
(24 bits de alta ordem: parte da sub-rede do endereço IP)

Sub-redes

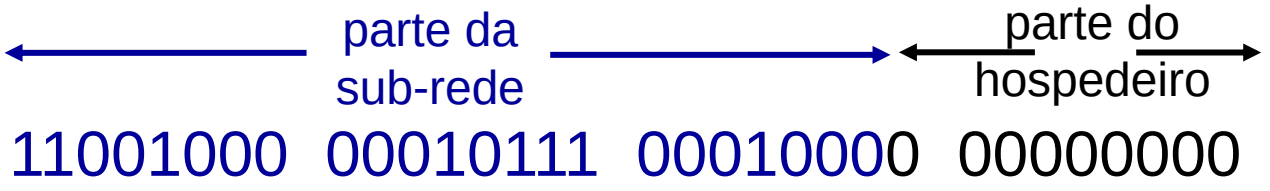
- onde estão as sub-redes?
- quais são os endereços de sub-rede /24?



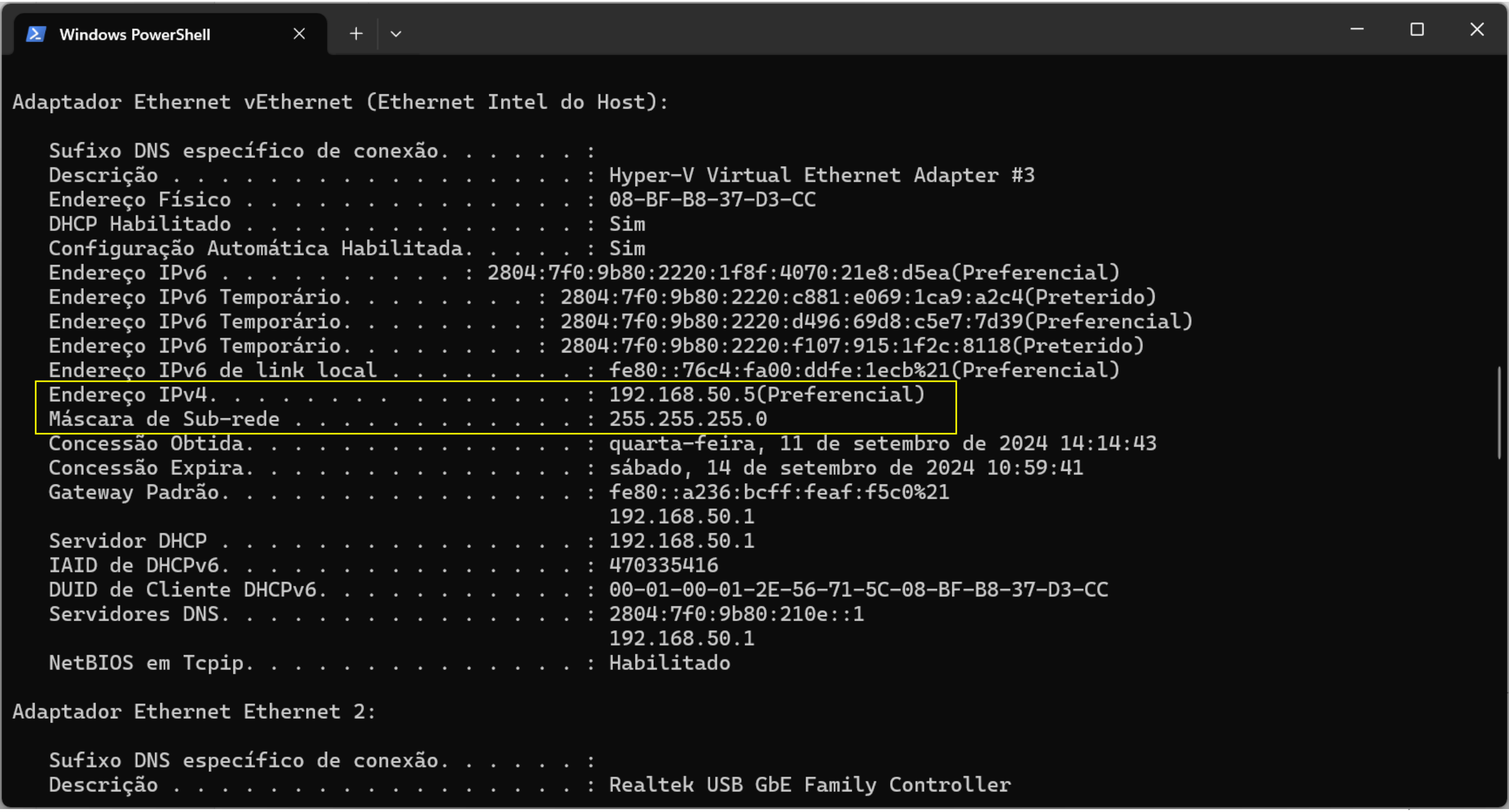
Endereçamento IP: CIDR

CIDR: Classless InterDomain Routing (pronuncia-se “cider”)

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



200.23.16.0/23


Windows PowerShell
Adaptador Ethernet vEthernet (Ethernet Intel do Host):
Sufixo DNS específico de conexão. . . . . :
Descrição . . . . . : Hyper-V Virtual Ethernet Adapter #3
Endereço Físico . . . . . : 08-BF-B8-37-D3-CC
DHCP Habilitado . . . . . : Sim
Configuração Automática Habilitada. . . . . : Sim
Endereço IPv6 . . . . . : 2804:7f0:9b80:2220:1f8f:4070:21e8:d5ea(Preferencial)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:c881:e069:1ca9:a2c4(Preterido)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:d496:69d8:c5e7:7d39(Preferencial)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:f107:915:1f2c:8118(Preterido)
Endereço IPv6 de link local . . . . . : fe80::76c4:fa00:ddfe:1ecb%21(Preferencial)
Endereço IPv4. . . . . : 192.168.50.5(Preferencial)
Máscara de Sub-rede . . . . . : 255.255.255.0
Concessão Obtida. . . . . : quarta-feira, 11 de setembro de 2024 14:14:43
Concessão Expira. . . . . : sábado, 14 de setembro de 2024 10:59:41
Gateway Padrão. . . . . : fe80::a236:bcff:feaf:f5c0%21
192.168.50.1
Servidor DHCP . . . . . : 192.168.50.1
IAID de DHCPv6. . . . . : 470335416
DUID de Cliente DHCPv6. . . . . : 00-01-00-01-2E-56-71-5C-08-BF-B8-37-D3-CC
Servidores DNS. . . . . : 2804:7f0:9b80:210e::1
192.168.50.1
NetBIOS em Tcpip. . . . . : Habilitado
Adaptador Ethernet Ethernet 2:
Sufixo DNS específico de conexão. . . . . :
Descrição . . . . . : Realtek USB GbE Family Controller

Endereços IP: como conseguir um?

Na verdade, são duas questões:

1. Q: Como um *hospedeiro* obtém o endereço IP em sua rede (parte do hospedeiro do endereço)?
2. Q: Como uma *rede* obtém um endereço IP para si mesma (parte da rede do endereço)?

Como um *hospedeiro* obtém o endereço IP?

- codificado pelo administrador do sistema em um arquivo de configuração (ex.: /etc/rc.config no UNIX)
- DHCP: Dynamic Host Configuration Protocol: obtém endereços dinamicamente de um servidor
  - "plug-and-play"

DHCP: Dynamic Host Configuration Protocol

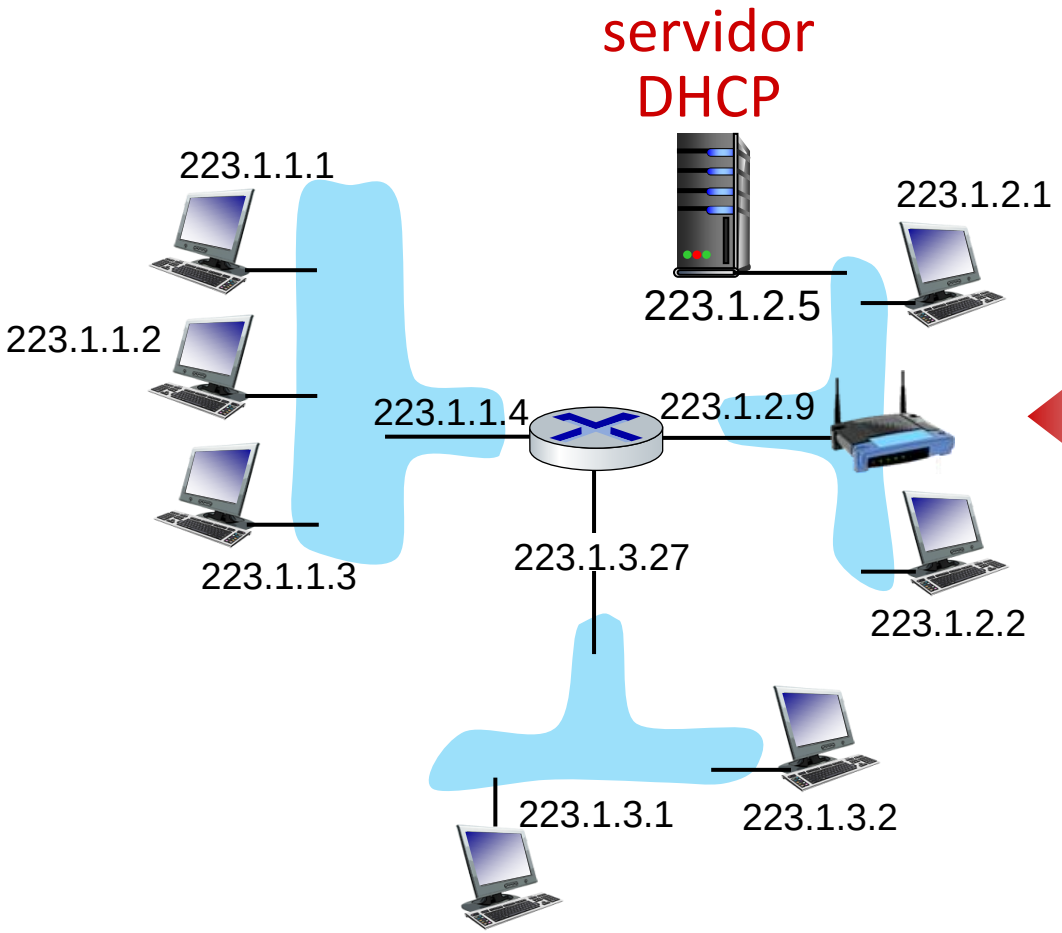
objetivo: o hospedeiro obtém *dinamicamente* o endereço IP do servidor de rede quando "entra" na rede

- pode renovar sua concessão sobre o endereço em uso
- permite a reutilização de endereços (apenas mantem o endereço enquanto conectado / ligado)
- suporte para usuários móveis que ingressam / saem da rede

visão geral do DHCP:

- hospedeiro difunde (*broadcasts*) mensagem DHCP discover [opcional]
- servidor DHCP responde com mensagem, DHCP offer [opcional]
- hospedeiro requisita endereço IP: mensagem DHCP request
- servidor DHCP envia endereço: mensagem DHCP ack

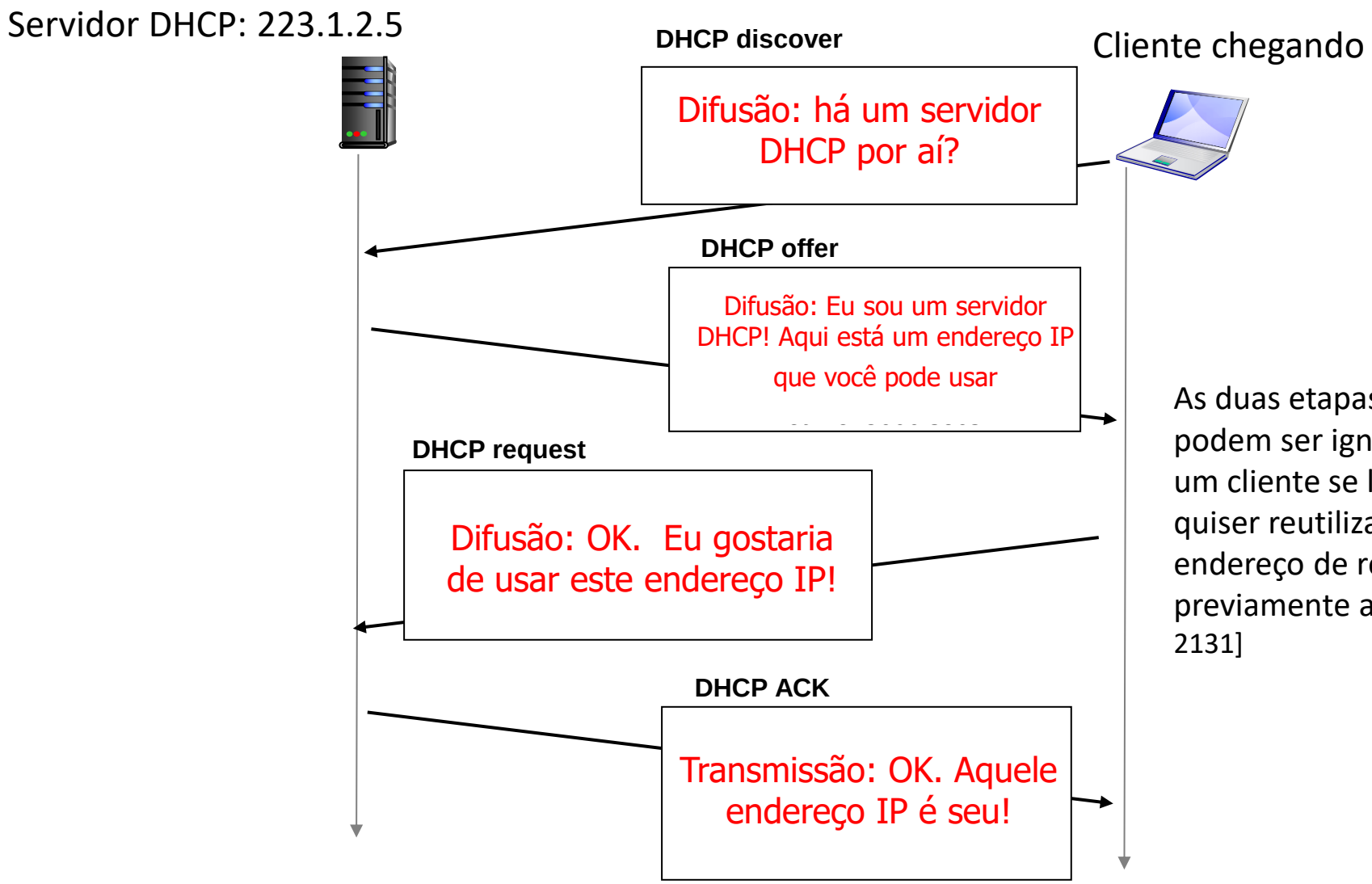
Cenário cliente-servidor DHCP



Normalmente, o servidor DHCP será co-localizado no roteador, servindo todas as sub-redes às quais o roteador está conectado

cliente DHCP chegando precisa de endereço nesta rede

Cenário cliente-servidor DHCP



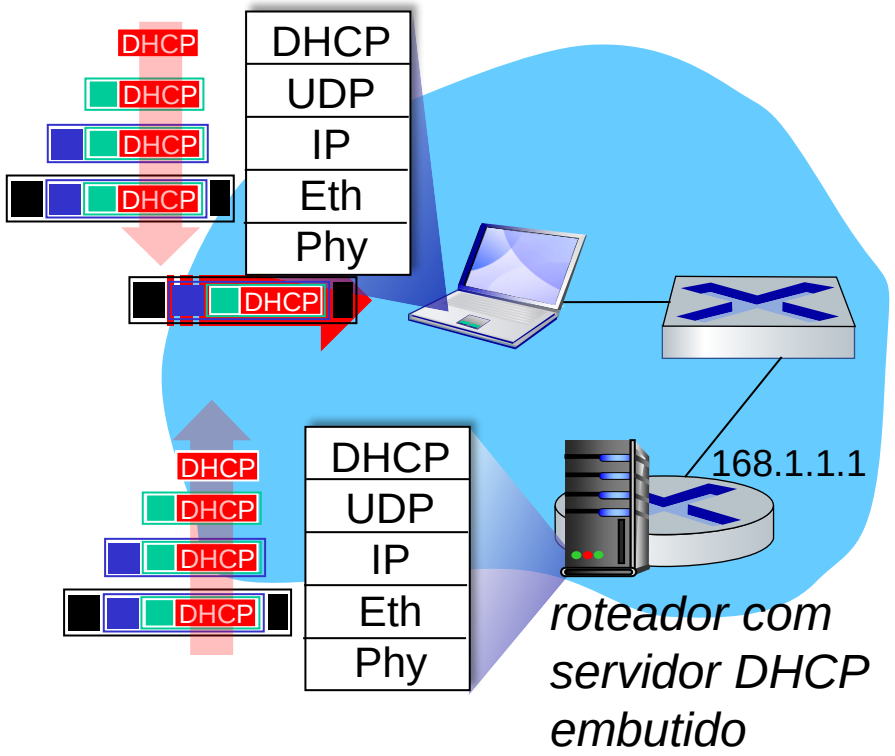
As duas etapas acima podem ser ignoradas "se um cliente se lembrar e quiser reutilizar um endereço de rede previamente alocado" [RFC 2131]

DHCP: mais que endereços IP

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

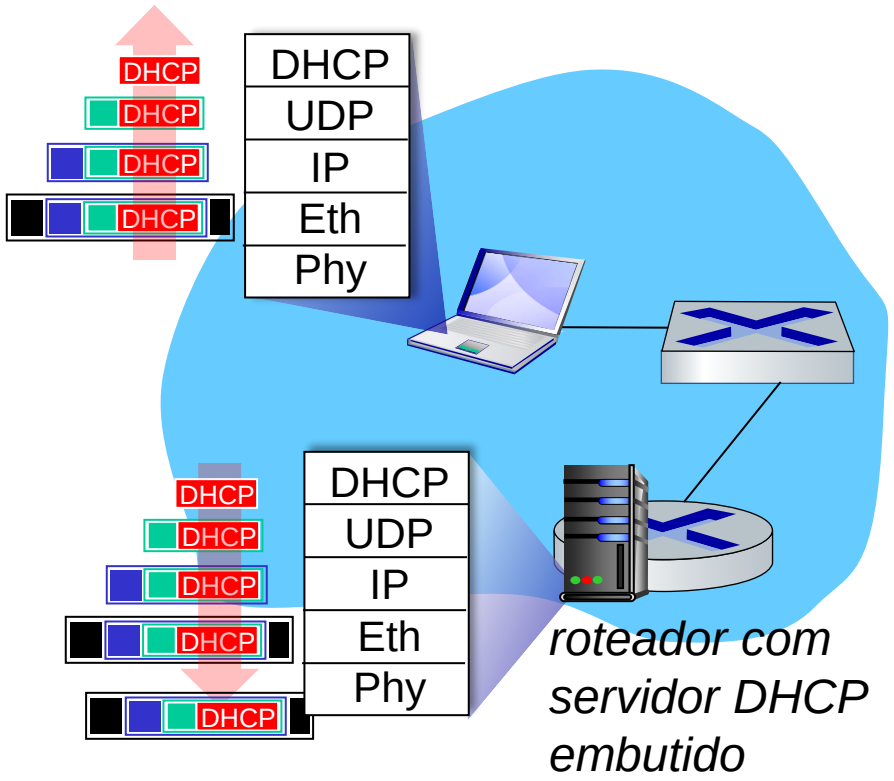
- endereço do roteador de primeiro salto para o cliente
- nome e endereço IP de servidor DNS
- máscara de sub-rede (indicando porção de rede versus porção de hospedeiro do endereço)

DHCP: exemplo

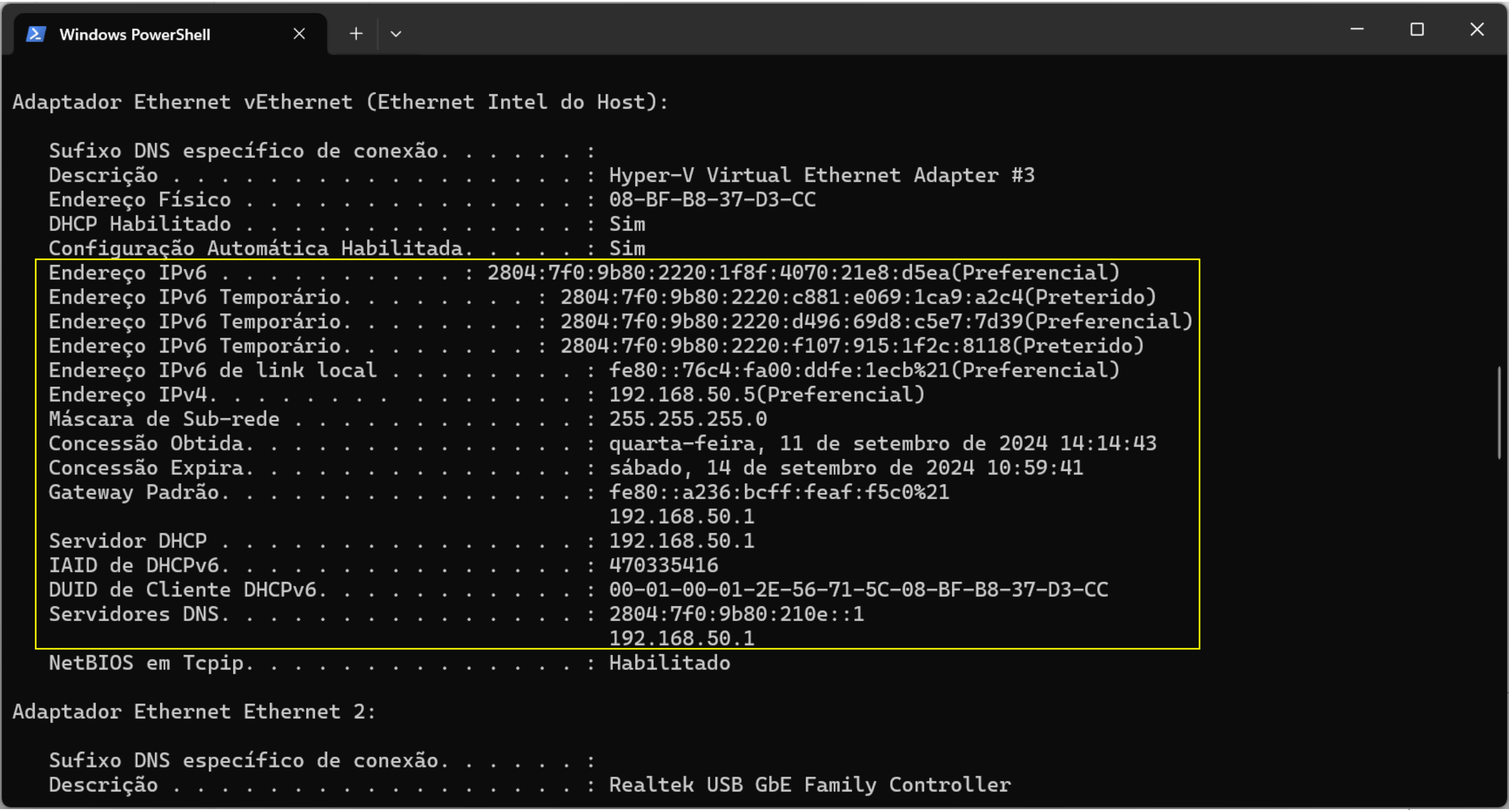


- Um laptop se conectando usará o DHCP para obter o endereço IP, o endereço do roteador de primeiro salto e o endereço do servidor DNS.
- Mensagem DHCP REQUEST encapsulada em UDP, encapsulada em IP, encapsulada em Ethernet
- Difusão do quadro Ethernet (dest: FFFFFFFFFFFF) na LAN, recebido no roteador executando o servidor DHCP
- Ethernet demultiplexada para IP, demultiplexada para UDP, demultiplexada para DHCP

DHCP: exemplo



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


Windows PowerShell
Adaptador Ethernet vEthernet (Ethernet Intel do Host):
Sufixo DNS específico de conexão. . . . . :
Descrição . . . . . : Hyper-V Virtual Ethernet Adapter #3
Endereço Físico . . . . . : 08-BF-B8-37-D3-CC
DHCP Habilitado . . . . . : Sim
Configuração Automática Habilitada. . . . . : Sim
Endereço IPv6 . . . . . : 2804:7f0:9b80:2220:1f8f:4070:21e8:d5ea(Preferencial)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:c881:e069:1ca9:a2c4(Preterido)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:d496:69d8:c5e7:7d39(Preferencial)
Endereço IPv6 Temporário. . . . . : 2804:7f0:9b80:2220:f107:915:1f2c:8118(Preterido)
Endereço IPv6 de link local . . . . . : fe80::76c4:fa00:ddfe:1ecb%21(Preferencial)
Endereço IPv4. . . . . : 192.168.50.5(Preferencial)
Máscara de Sub-rede . . . . . : 255.255.255.0
Concessão Obtida. . . . . : quarta-feira, 11 de setembro de 2024 14:14:43
Concessão Expira. . . . . : sábado, 14 de setembro de 2024 10:59:41
Gateway Padrão. . . . . : fe80::a236:bcff:feaf:f5c0%21
192.168.50.1
Servidor DHCP . . . . . : 192.168.50.1
IAID de DHCPv6. . . . . : 470335416
DUID de Cliente DHCPv6. . . . . : 00-01-00-01-2E-56-71-5C-08-BF-B8-37-D3-CC
Servidores DNS. . . . . : 2804:7f0:9b80:210e::1
192.168.50.1
NetBIOS em Tcpip. . . . . : Habilitado
Adaptador Ethernet Ethernet 2:
Sufixo DNS específico de conexão. . . . . :
Descrição . . . . . : Realtek USB GbE Family Controller

Endereços IP: como conseguir um?

Q: como a *rede* obtem a parte de sub-rede de endereço IP?

R: obtém parte alocada do espaço de endereços de seu provedor

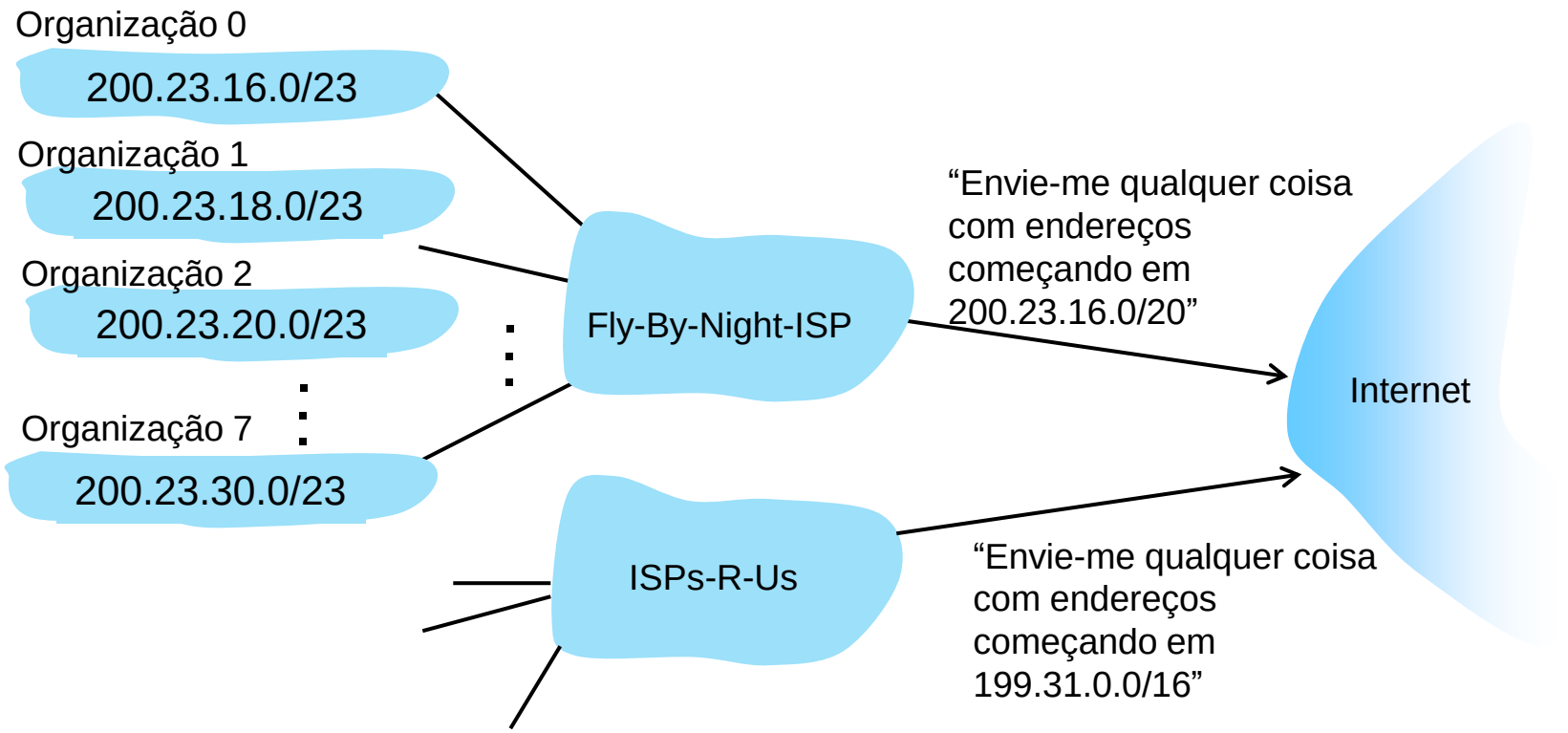
Bloco do ISP	11001000 00010111 00010000 00000000	200.23.16.0/20

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

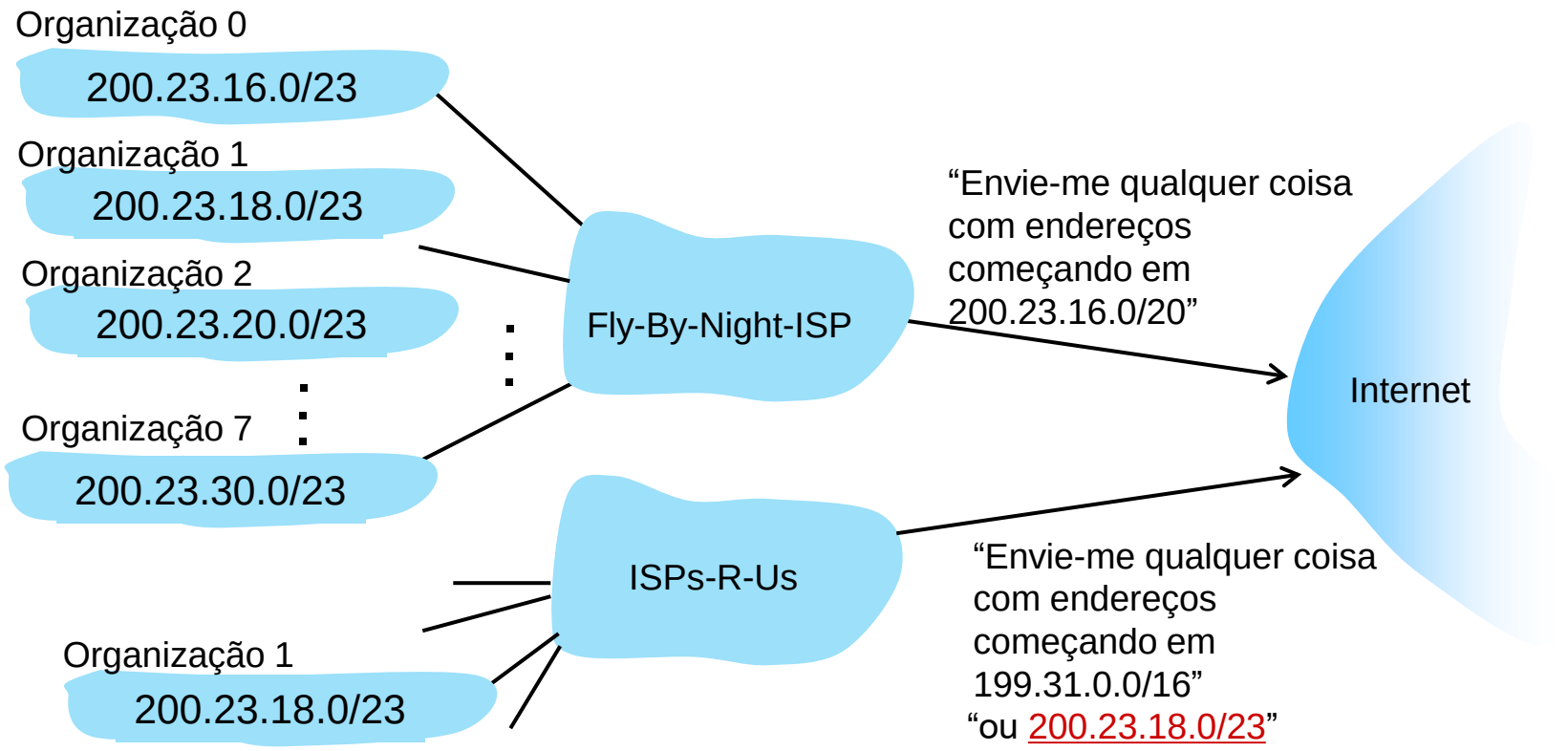
Endereçamento hierárquico: agregação de rotas

o endereçamento hierárquico permite a divulgação eficiente de informações de roteamento:



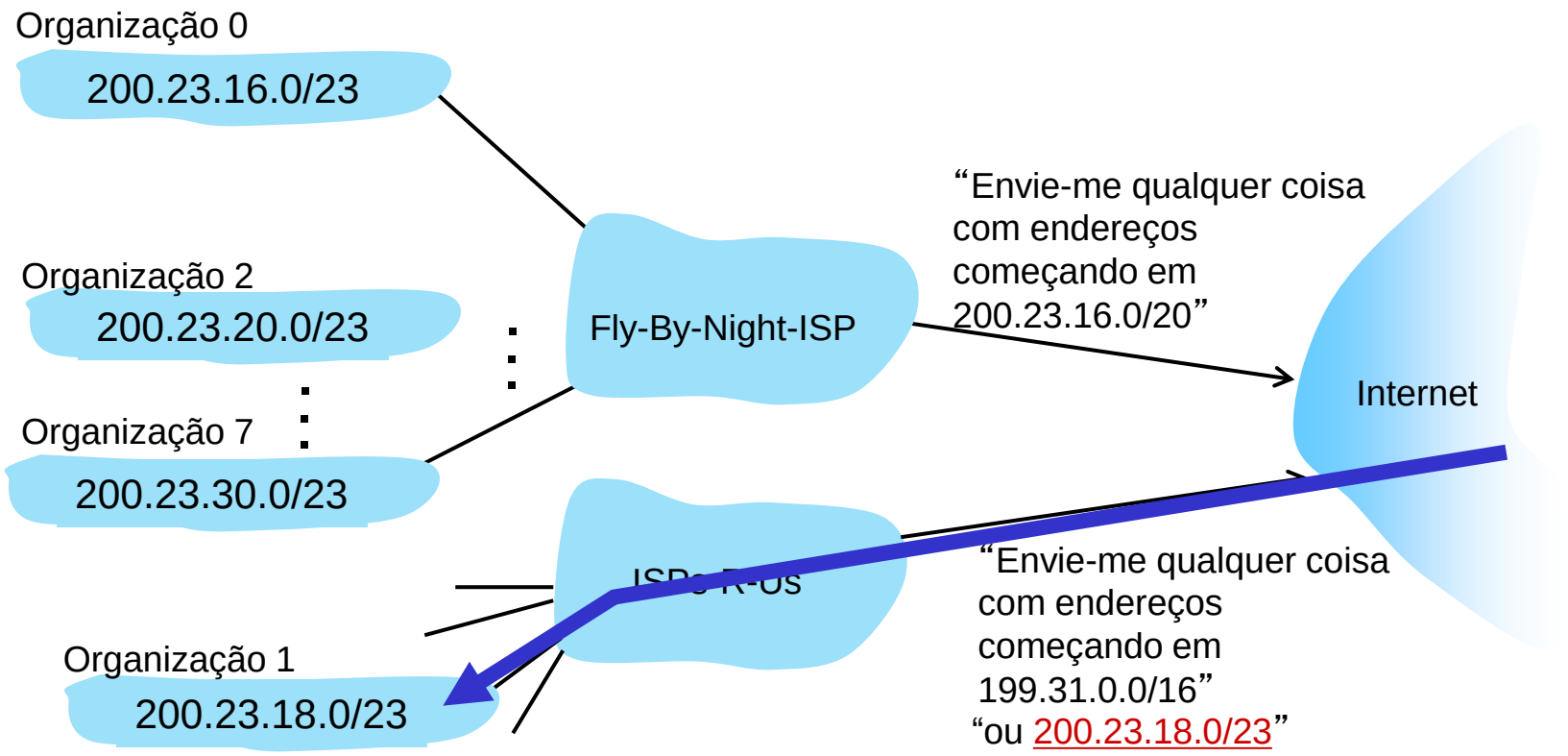
Endereçamento hierárquico: rotas mais específicas

- Organização 1 passa de Fly-By-Night-ISP para ISPs-R-Us
- ISPs-R-Us anuncia agora uma rota mais específica para a Organização 1



Endereçamento hierárquico: rotas mais específicas

- Organização 1 passa de Fly-By-Night-ISP para ISPs-R-Us
- ISPs-R-Us anuncia agora uma rota mais específica para a Organização 1



Endereçamento IP: últimas palavras ...

Q: como um ISP recebe bloco de endereços?

R: ICANN: Internet Corporation for Assigned Names and Numbers http://www.icann.org/

- aloca endereços IP, através de 5 regional registries (RRs) (que podem, então, alocar para registros locais)
- gerencia a zona raiz do DNS, incluindo gerenciamento de delegação de TLD individual (.com, .edu , ...)

Q: existem endereços IP suficientes de 32 bits?

- ICANN alocou último pedaço de endereços IPv4 para RRs em 2011
- NAT (a seguir) ajuda o IPv4 a lidar com o esgotamento do espaço
- IPv6 tem espaço de endereço de 128 bits

"Quem diabos sabia quanto espaço de endereço precisávamos?" - Vint Cerf (refletindo sobre a decisão de fazer endereços IPv4 de 32 bits de comprimento)

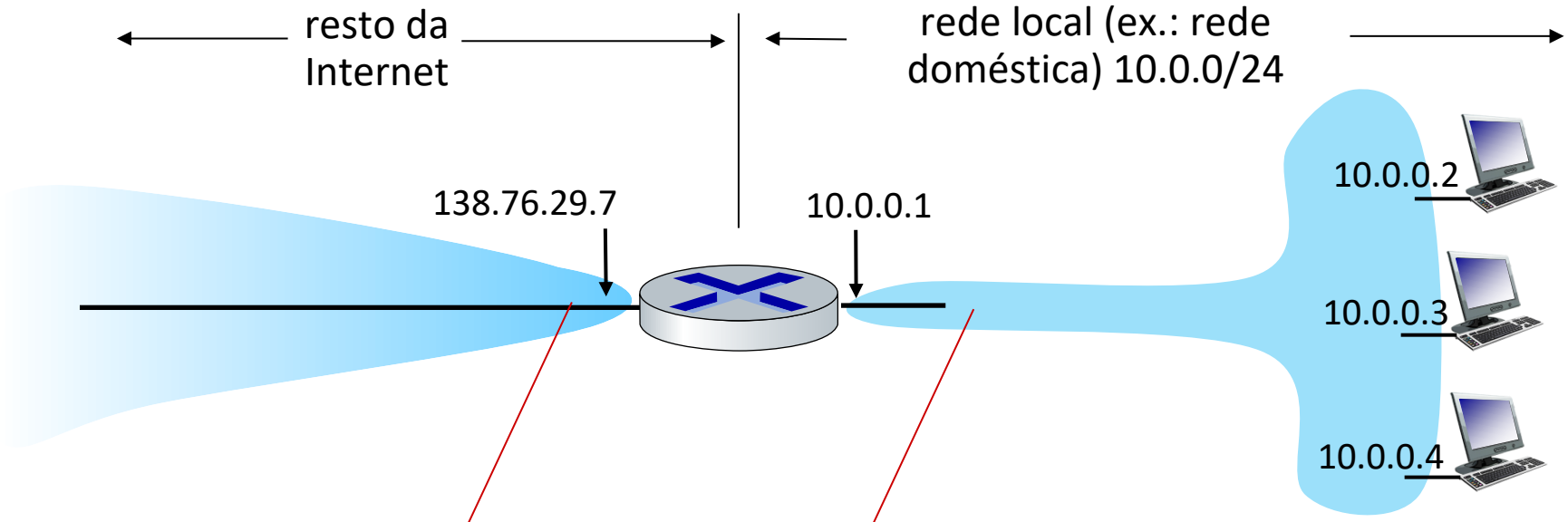
Camada de rede: roteiro do “plano de dados”

- 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 da Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereço de rede (NAT)
  - IPv6

- Encaminhamento Generalizado e SDN
  - Correspondência+ação
  - OpenFlow: correspondência+ação em ação
- Middleboxes

NAT: network address translation

NAT: todos os dispositivos em uma rede local compartilham apenas um endereço IPv4 no que diz respeito ao mundo exterior



todos os datagramas *deixando* a rede local tem o *mesmo* endereço IP NAT fonte: 138.76.29.7, mas *diferentes* números de porta de origem

datagramas com origem ou destino nesta rede tem endereços 10.0.0/24 para origem e destino (como de costume)

NAT: network address translation

- 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 podem ser usados em redes locais
- vantagens:
  - apenas um endereço IP do provedor é necessário para *todos* os dispositivos
  - é possível mudar endereços de hospedeiro em redes locais sem notificar o mundo exterior
  - pode alterar o provedor sem alterar endereços de dispositivos na rede local
  - segurança: dispositivos dentro da rede local não são diretamente endereçáveis ou visíveis pelo mundo exterior

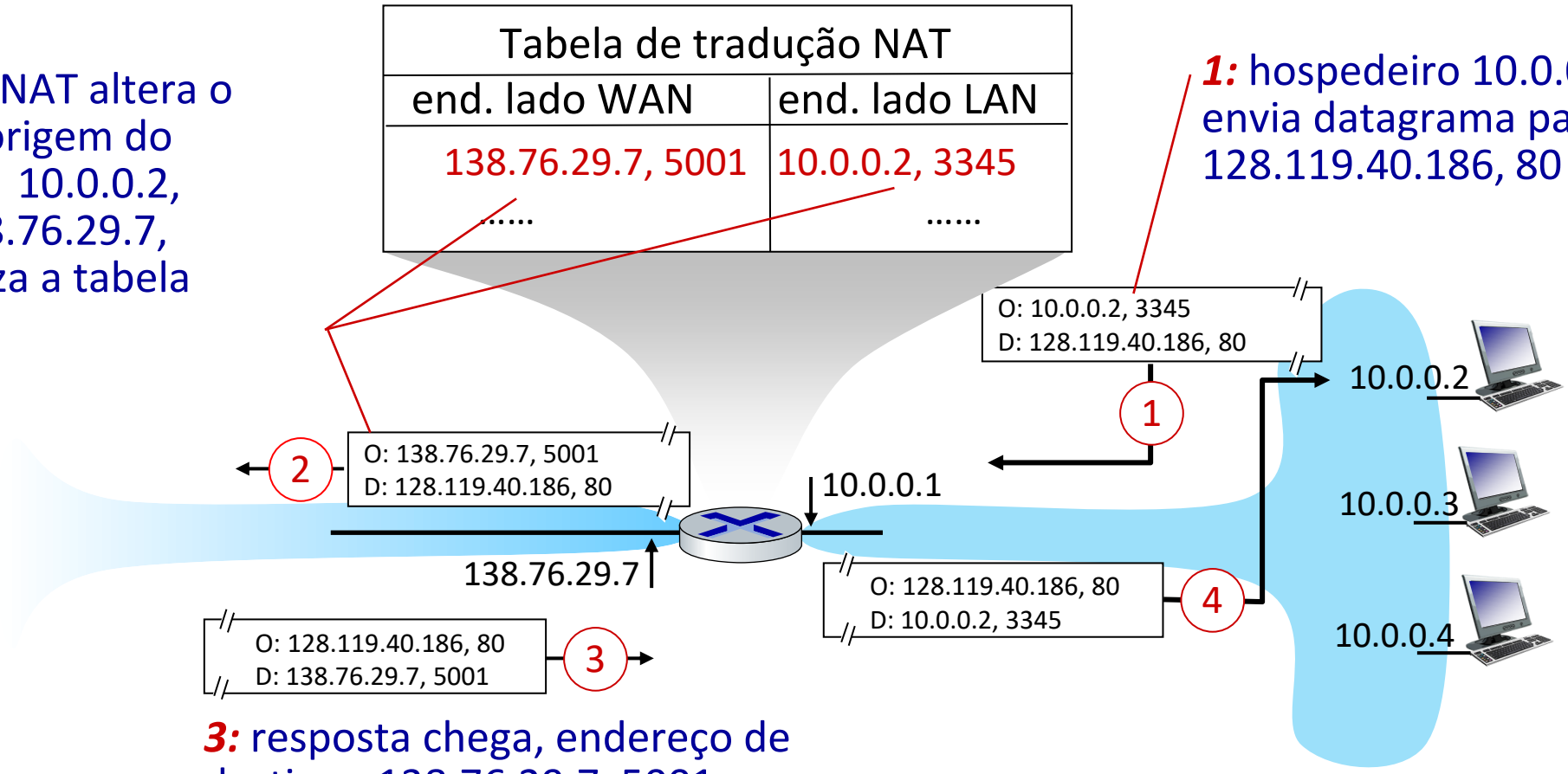
NAT: network address translation

implementação: roteador NAT precisa (transparentemente):

- datagramas de saída: substituir (endereço IP de origem, número de porta) de cada datagrama de saída para (endereço IP NAT, novo número de porta)
  - clientes/servidores remotos responderão usando (endereço IP NAT, novo número de porta) como endereço de destino
- lembrar (na tabela de tradução NAT) todo par de tradução (endereço IP de origem, número de porta) para (endereço IP NAT, novo número de porta)
- datagramas de entrada: substituir (endereço IP NAT, novo número de porta) em campos de destino de cada datagrama de entrada com os correspondentes (endereço IP de origem, número de porta) armazenados na tabela NAT

NAT: network address translation

2: O roteador NAT altera o endereço de origem do datagrama de 10.0.0.2, 3345 para 138.76.29.7, 5001, e atualiza a tabela



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

NAT: network address translation

- NAT tem sido controverso:
  - roteadores “deveriam” processar apenas até a camada 3
  - “escassez” de endereços deve ser resolvida pelo IPv6
  - viola argumento fim-a-fim (manipulação de número de porta por dispositivo da camada de rede)
  - Travessia de NAT: e se o cliente quiser se conectar a um servidor atrás de um NAT?
- mas o NAT está aqui para ficar:
  - amplamente utilizado em redes domésticas e institucionais e redes celulares 4G/5G

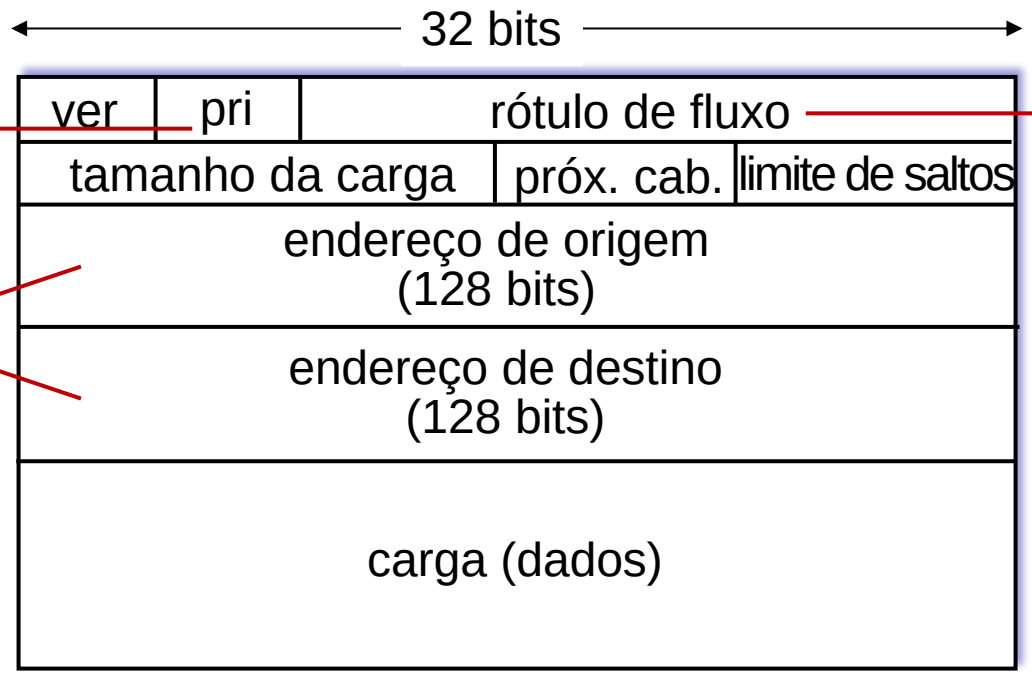
IPv6: motivação

- motivação inicial: espaço de endereço IPv4 de 32 bits seria completamente alocado
- motivação adicional:
  - processamento/encaminhamento de alta velocidade: cabeçalho de comprimento fixo de 40 bytes
  - permitir diferentes tratamentos de “fluxos” na camada de rede

Formato de datagrama IPv6

prioridade: identifica prioridade entre datagramas no fluxo

endereços IPv6 de 128 bits



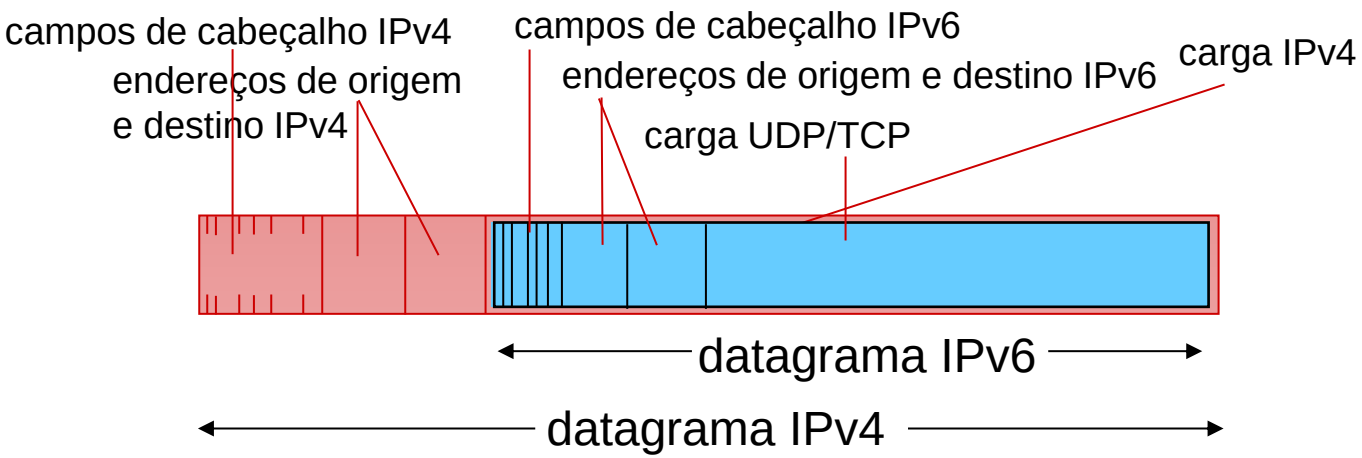
rótulo de fluxo: identifica datagramas no mesmo "fluxo". (conceito de "fluxo" não é bem definido).

O que está faltando (comparado com IPv4):

- sem soma de verificação (para acelerar o processamento em roteadores)
- sem fragmentação/remontagem
- sem opções (disponíveis como protocolo de próximo cabeçalho de camada superior no roteador)

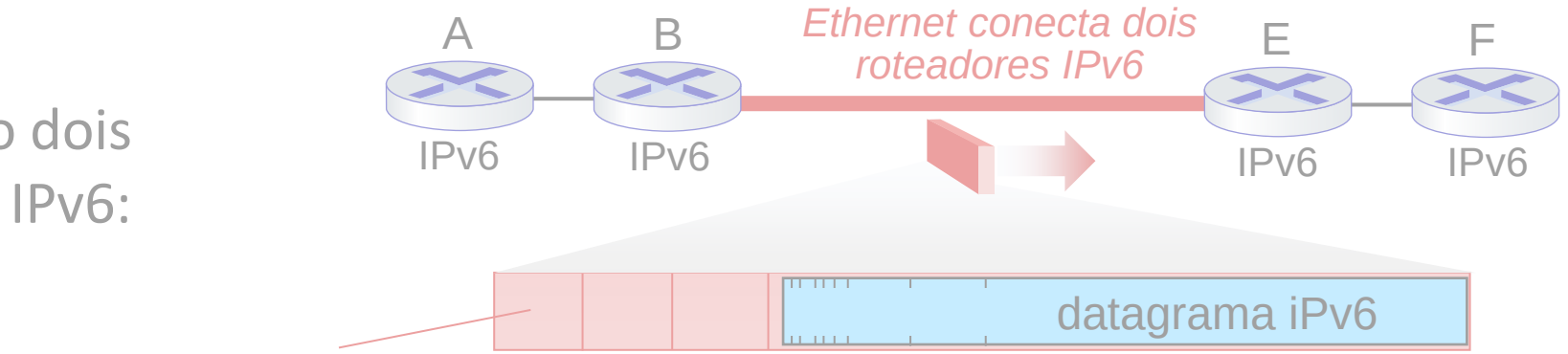
Transição de IPv4 para IPv6

- nem todos os roteadores podem ser atualizados simultaneamente
  - sem data marcada para a mudança
  - como a rede funcionará com roteadores IPv4 e IPv6 misturados?
- tunelamento: datagrama IPv6 carregado como *carga* em datagrama IPv4 entre roteadores IPv4 ("pacote dentro de um pacote")
  - tunelamento é usado extensivamente em outros contextos (4G / 5G)



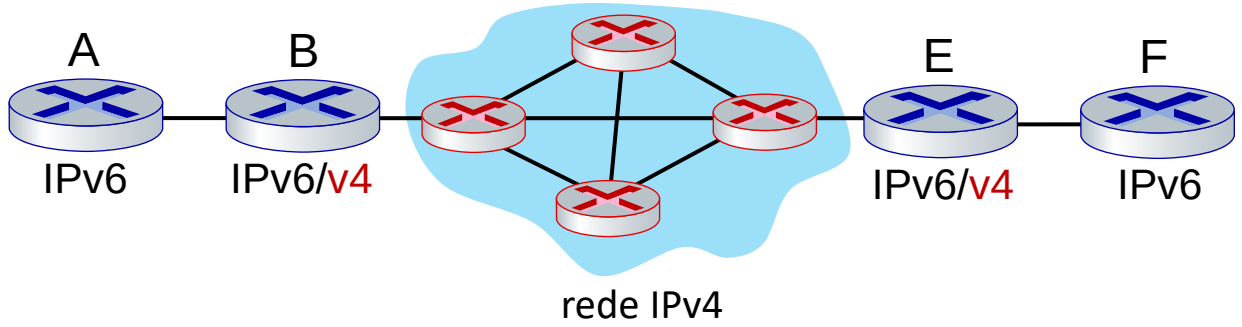
Tunelamento e encapsulamento

Ethernet conectando dois roteadores IPv6:



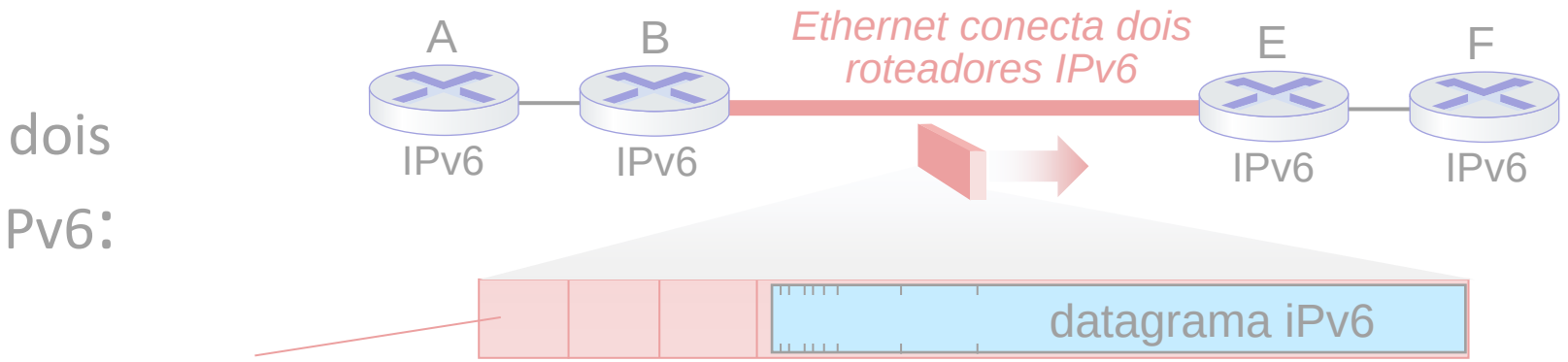
O usual: datagrama como carga útil no quadro da camada de enlace

Rede IPv4 conectando dois roteadores IPv6



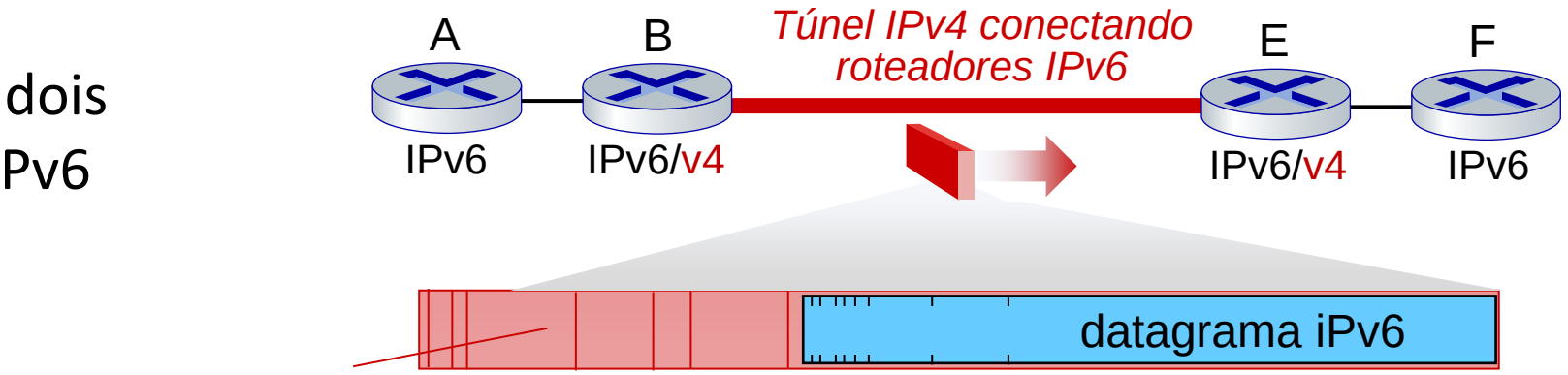
Tunelamento e encapsulamento

Ethernet conectando dois roteadores IPv6:



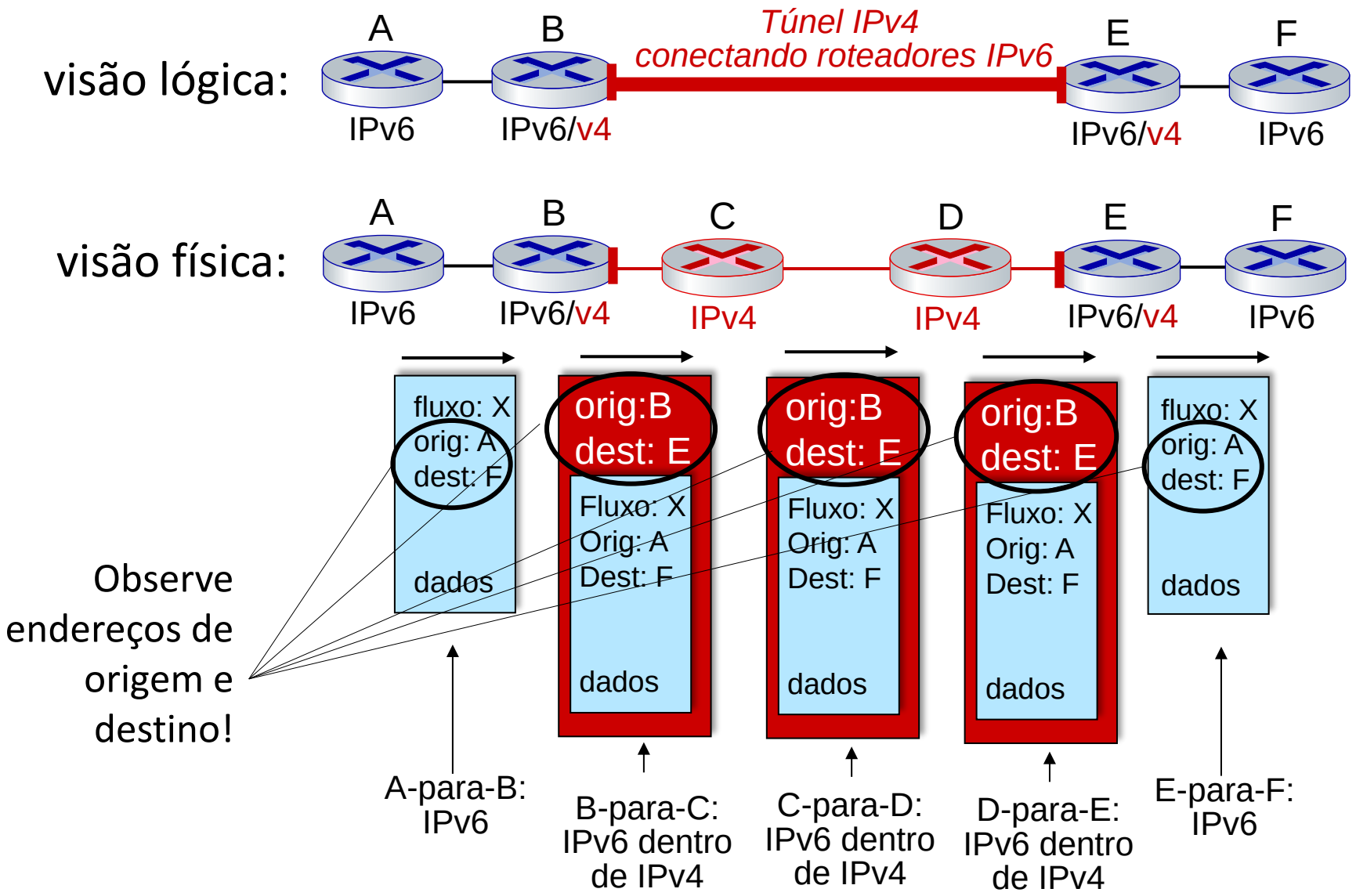
O usual: datagrama como carga útil no quadro da camada de enlace

Túnel IPv4 conectando dois roteadores IPv6



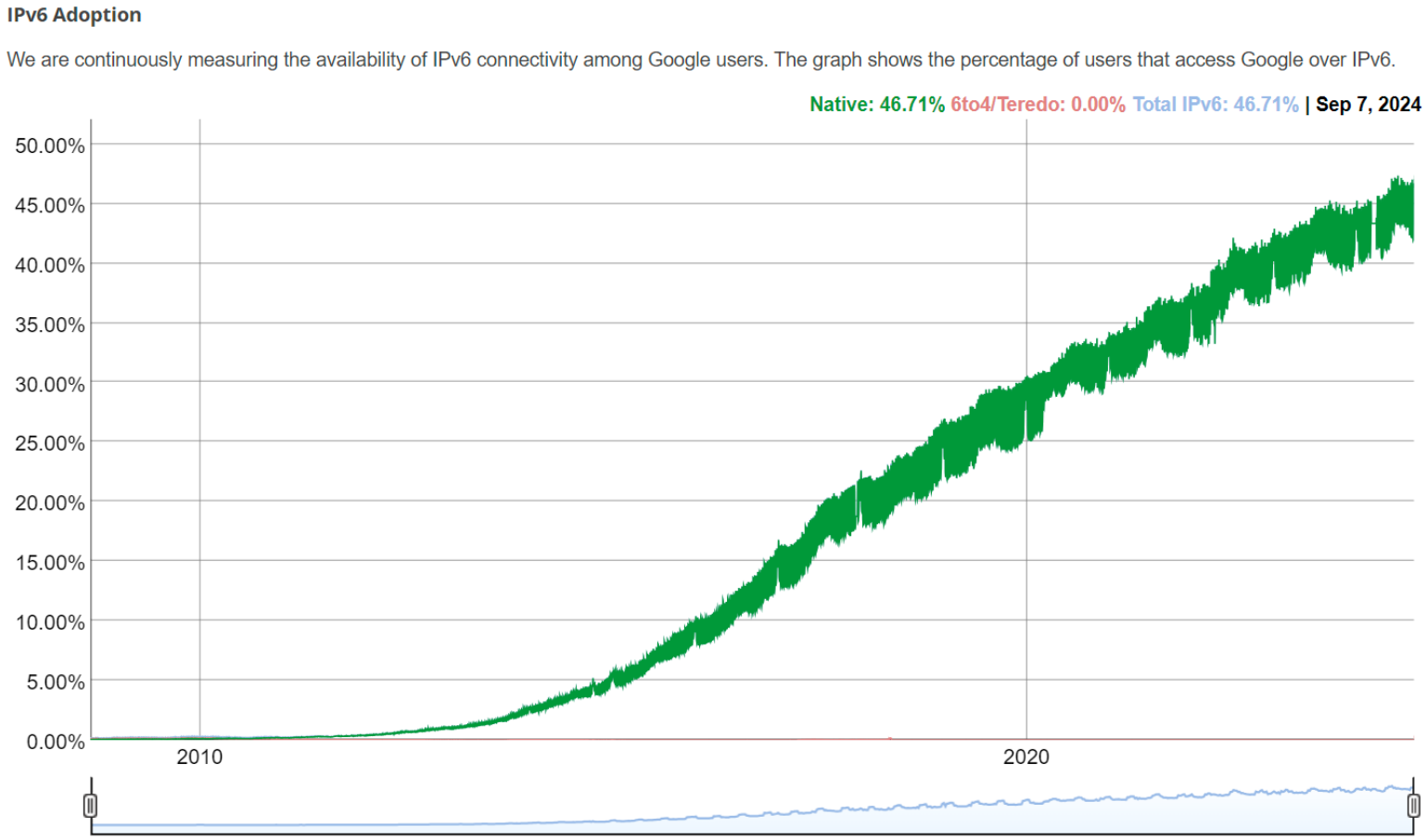
tunelamento: datagrama IPv6 como carga útil em um datagrama IPv4

Tunelamento



IPv6: adoção

- Google[1]: ~ 47% de clientes acessam serviços via IPv6 (2024)
- NIST: 1/3 de todos os domínios do governo dos EUA são compatíveis com IPv6



1

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

IPv6: adoção

- Google[1]: ~ 47% de clientes acessam serviços via IPv6 (2024)
- NIST: 1/3 de todos os domínios do governo dos EUA são compatíveis com IPv6
- Longo (longo!) tempo para implantação e uso
  - 29 anos e contando!
  - pense nas mudanças no nível de aplicação nos últimos 29 anos: WWW, mídias sociais, transmissão de mídia, jogos, telepresença, ...
  - *Por que?*

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

Camada de rede: roteiro do “plano de dados”

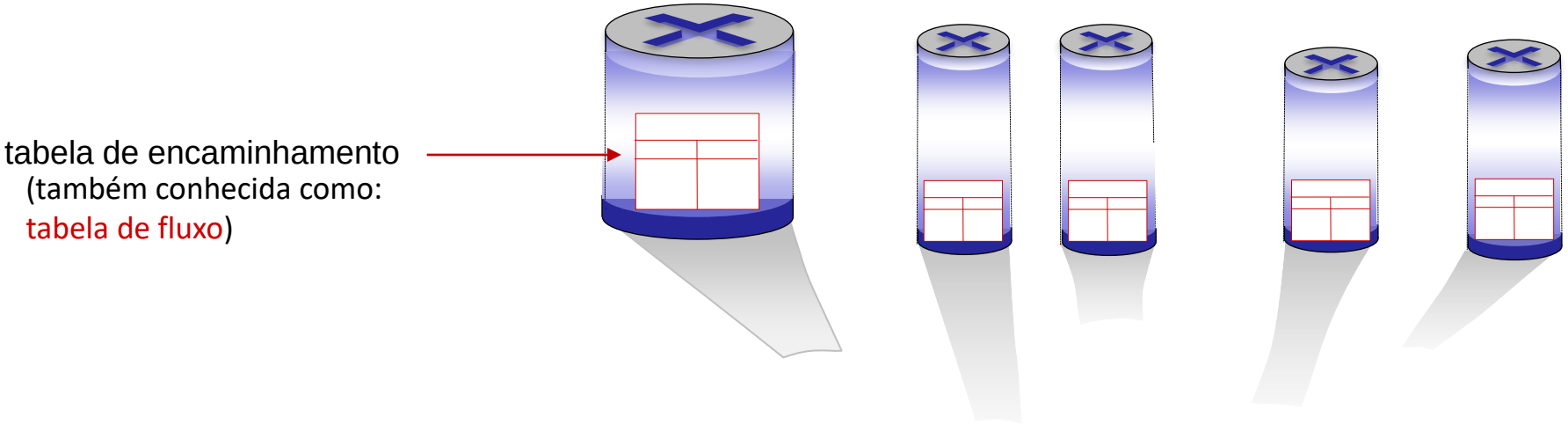
- Camada de rede: visão geral
  - plano de dados
  - plano de controle
- O que há dentro de um roteador
  - portas de entrada, comutação, portas de saída
  - gerenciamento de buffer, agendamento
- IP: o Protocolo da Internet
  - formato de datagrama
  - endereçamento
  - tradução de endereço de rede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

- Encaminhamento Generalizado e SDN
  - Correspondência+ação
  - OpenFlow: correspondência+ação em ação
- Middleboxes

Encaminhamento generalizado: correspondência mais ação (*match plus action*)

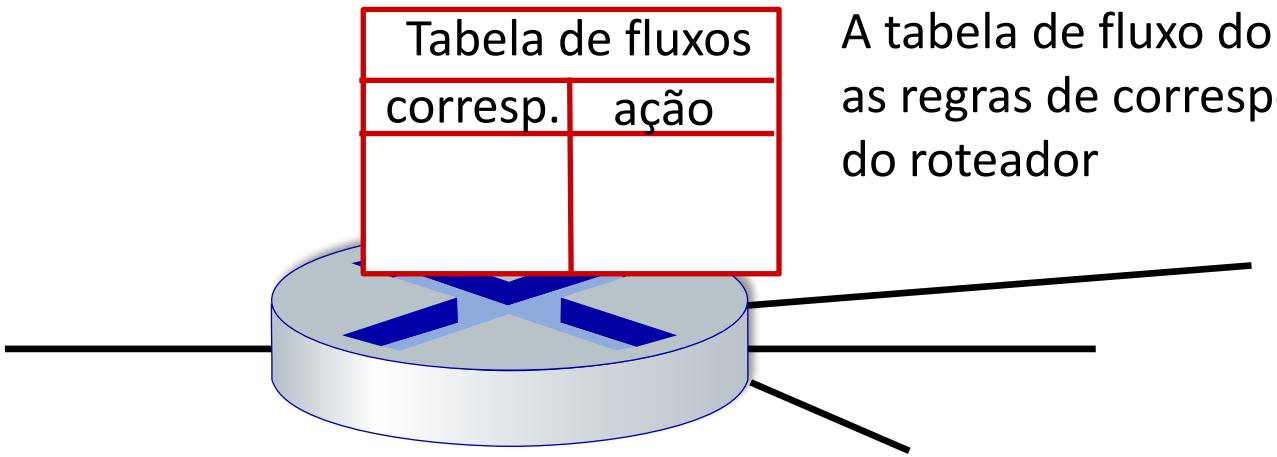
Revisão: cada roteador contém uma tabela de encaminhamento (também conhecida como: tabela de fluxo)

- abstração "correspondência mais ação": corresponde bits no pacote chegando, toma ação
  - *encaminhamento baseado em destino:* encaminha com base no endereço IP de destino
  - *encaminhamento generalizado:*
    - muitos campos de cabeçalho podem determinar ação
    - muitas ações possíveis: descartar/copiar/modificar/logar pacote



Abstração da tabela de fluxo

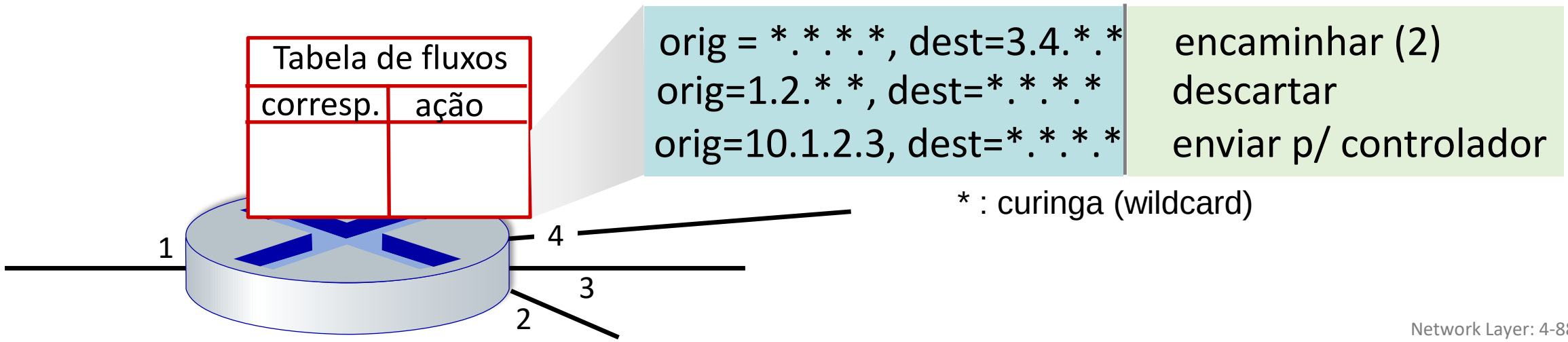
- fluxo: definido por valores de campos de cabeçalho (em campos das camadas de enlace, rede, e transporte)
- encaminhamento generalizado: regras simples de manipulação de pacotes
  - correspondência: padrões de valores em campos de cabeçalho do pacote
  - ações: para pacote correspondente: descartar, encaminhar, modificar ou enviar para o controlador
  - prioridade: desambiguação de padrões sobrepostos
  - contadores: número de bytes ou número de pacotes



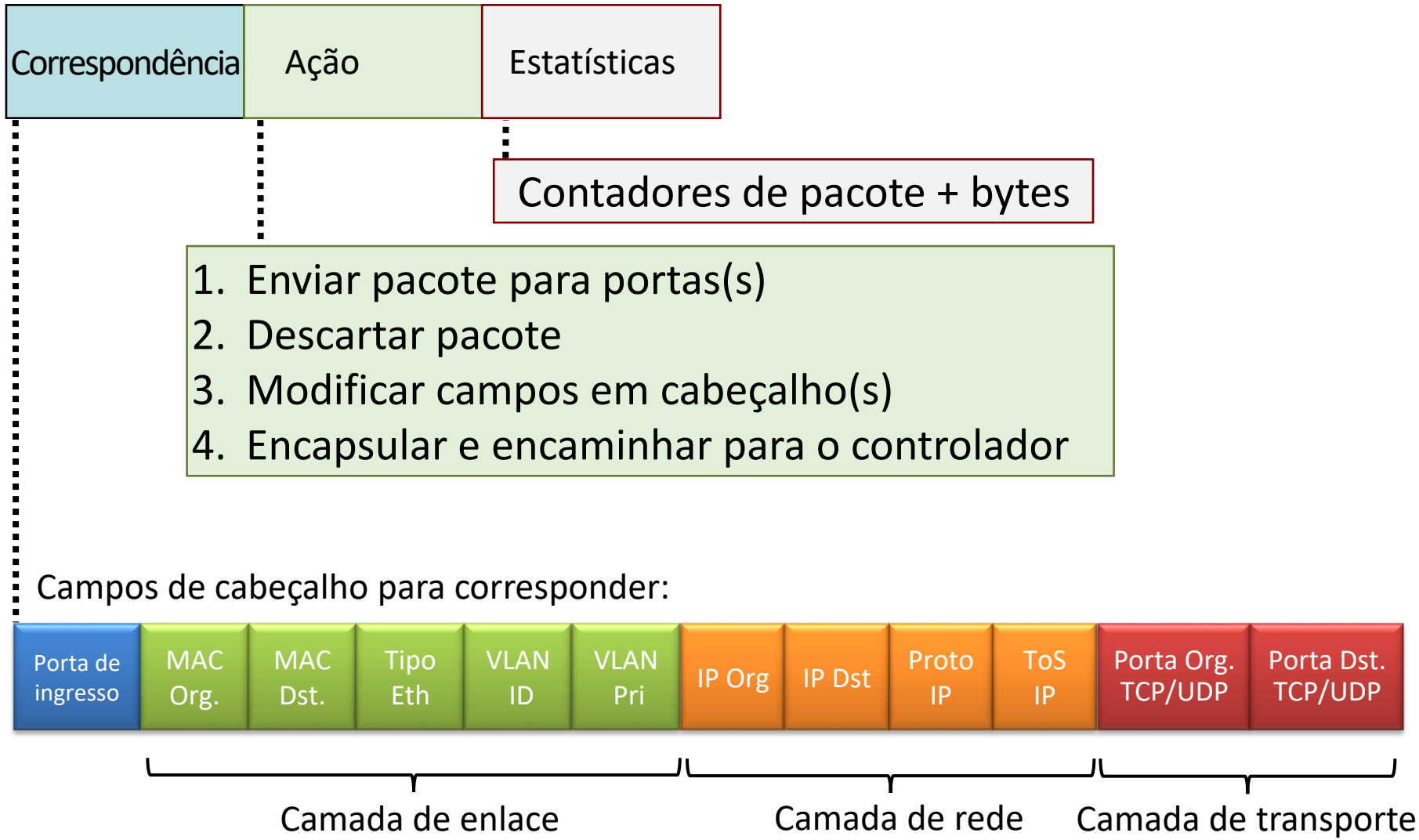
A tabela de fluxo do roteador define as regras de correspondência + ação do roteador

Abstração da tabela de fluxo

- fluxo: definido por valores de campos de cabeçalho
- encaminhamento generalizado: regras simples de manipulação de pacotes
  - correspondência: padrões de valores em campos de cabeçalho do pacote
  - ações: para pacote correspondente: descartar, encaminhar, modificar ou enviar para o controlador
  - prioridade: desambiguação de padrões sobrepostos
  - contadores: número de bytes ou número de pacotes



OpenFlow: entradas da tabela de fluxo



OpenFlow: exemplos

Encaminhamento baseado em destino:

Porta Switch	MAC org	MAC dst	Eth type	VLAN ID	VLAN Pri	IP Org	IP Dst	IP Prot	IP ToS	TCP o-port	TCP d-port	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:

Porta Switch	MAC org	MAC dst	Eth type	VLAN ID	VLAN Pri	IP Org	IP Dst	IP Prot	IP ToS	TCP o-port	TCP d-port	Ação
*	*	*	*	*	*	*	*	*	*	*	22	descartar

Bloqueie (não encaminhe) todos os datagramas destinados à porta TCP 22 (número de porta do SSH)

Porta Switch	MAC org	MAC dst	Eth type	VLAN ID	VLAN Pri	IP Org	IP Dst	IP Prot	IP ToS	TCP o-port	TCP d-port	Ação
*	*	*	*	*	*	128.119.1.1	*	*	*	*	*	descartar

Bloqueie (não encaminhe) todos os datagramas enviados pelo hospedeiro 128.119.1.1

OpenFlow: exemplos

Encaminhamento baseado em destino da camada 2:

Porta Switch	MAC org	MAC dst	Eth type	VLAN ID	VLAN Pri	IP Org	IP Dst	IP Prot	IP ToS	TCP o-port	TCP d-port	Ação
*	*	22:A7:23: 11:E1:02	*	*	*	*	*	*	*	*	*	porta3

quadros da 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

- correspondência+ação: abstração unifica diferentes tipos de dispositivos

Roteador

- *correspondência:* prefixo IP de destino mais longo
- *ação:* encaminhar por um enlace

Comutador

- *correspondência:* endereço MAC de destino
- *ação:* encaminhar ou inundar

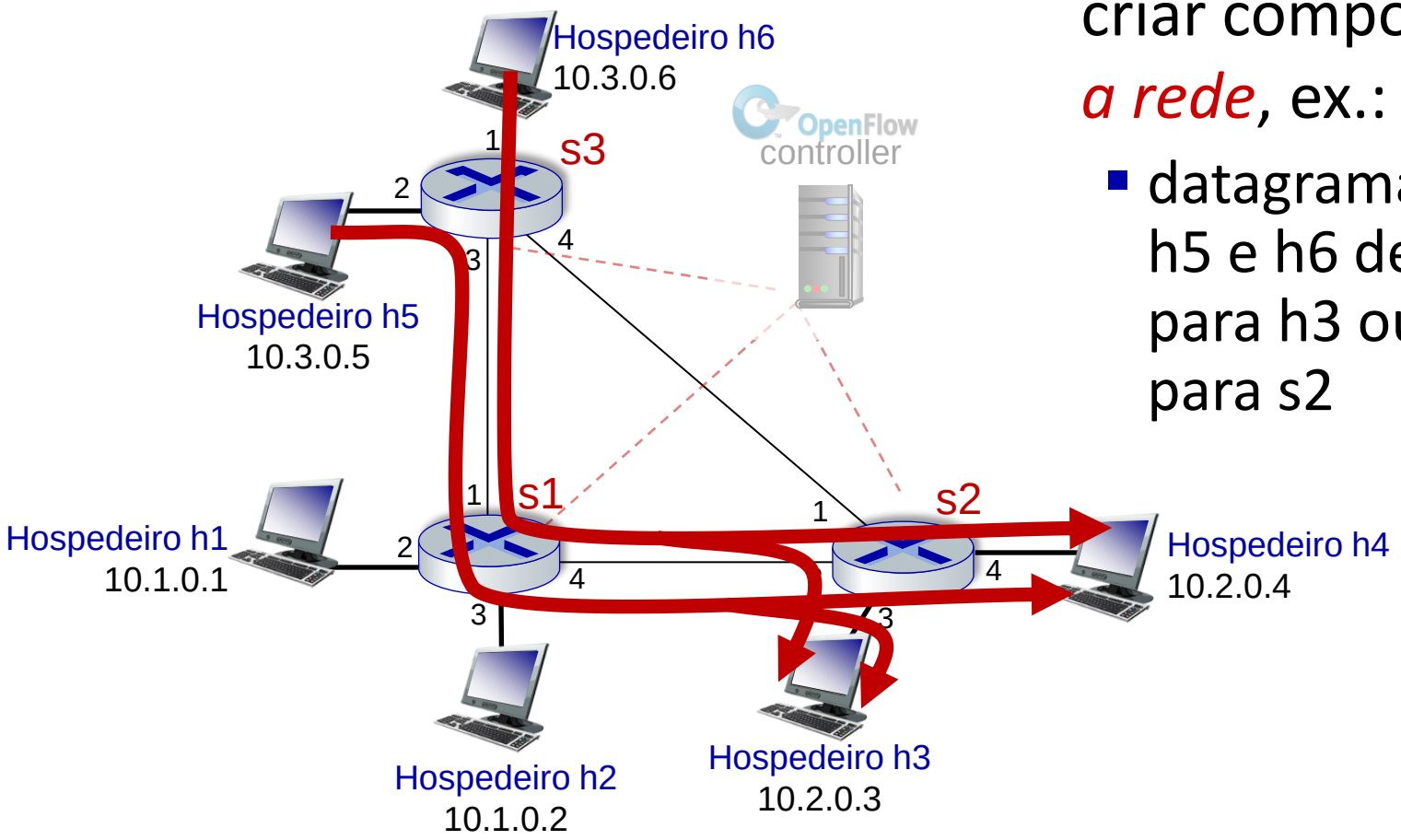
Firewall

- *correspondência:* endereços IP e números de porta TCP/UDP
- *ação:* permitir ou negar

NAT

- *correspondência:* endereço IP e porta
- *ação:* reescrever endereço e porta

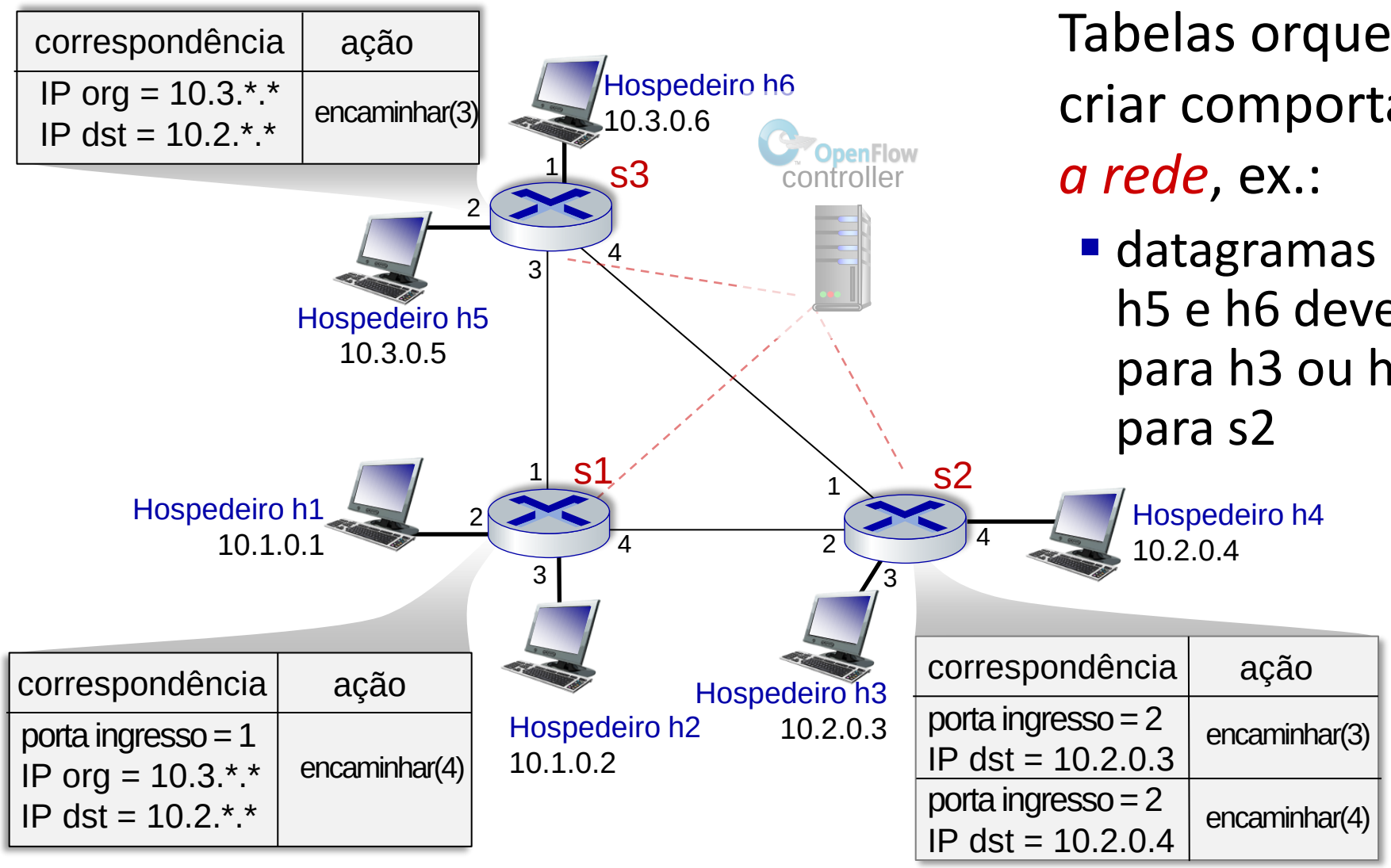
Exemplo de OpenFlow



Tabelas orquestradas podem criar comportamento *em toda a rede*, ex.:

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

Exemplo de OpenFlow



Tabelas orquestradas podem criar comportamento *em toda a rede*, ex.:

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

Encaminhamento generalizado: sumário

- abstração de "correspondência mais ação": corresponder bits nos cabeçalhos de qualquer camada dos pacotes chegando e tomar ação
  - correspondência sobre muitos campos (camadas de enlace, rede e transporte)
  - ações locais: descartar, encaminhar, modificar, ou enviar o pacote correspondente para o controlador
  - "programação" de comportamentos *na rede toda*
- forma simples de "programação da rede"
  - "processamento" por pacote programável
  - *raízes históricas:* rede ativa
  - *hoje:* programação mais generalizada: P4 (veja 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 da Internet
- Encaminhamento Generalizado e SDN

- **Middleboxes**
  - funções de middlebox
  - evolução, princípios arquitetônicos da Internet

Middleboxes

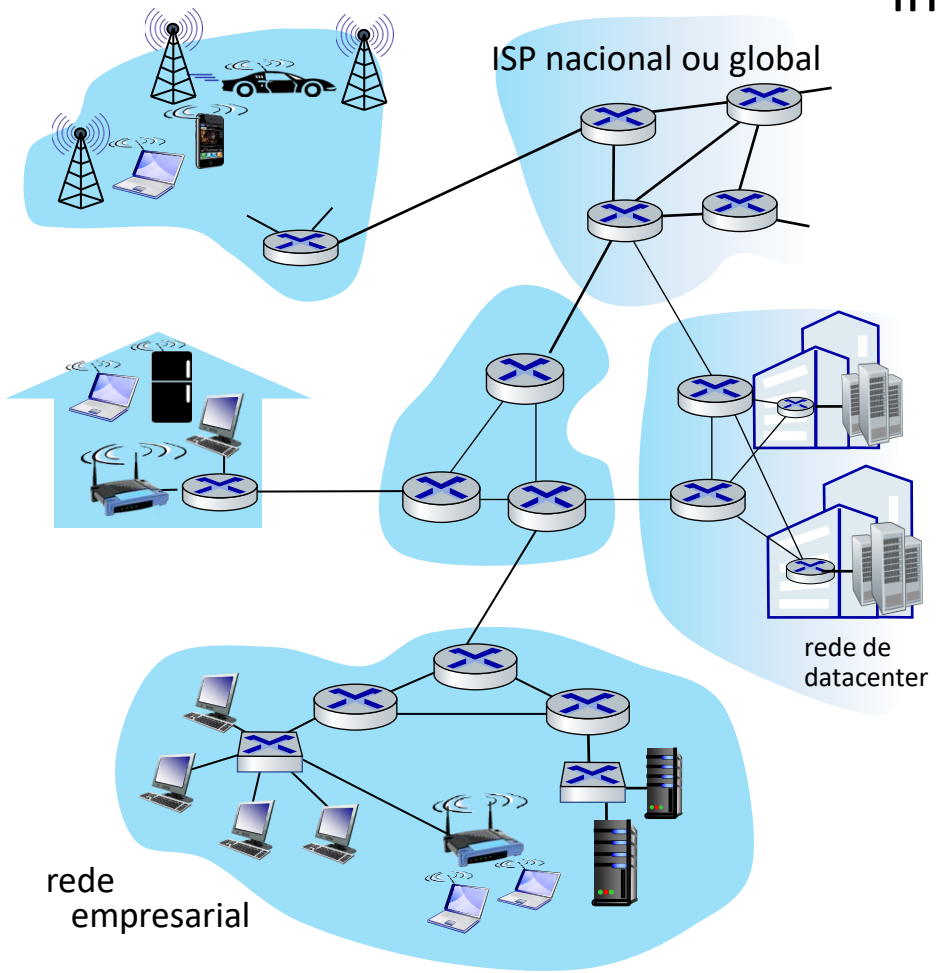
Middlebox (RFC 3234)

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

Middleboxes em toda parte!

NAT: casa, celular, institucional

Específico de aplicações: provedores de serviços, institucionais, CDN



Firewalls, IDS: corporativos, institucionais, provedores de serviços, ISPs

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

Caches: provedor de serviços, móvel, CDNs

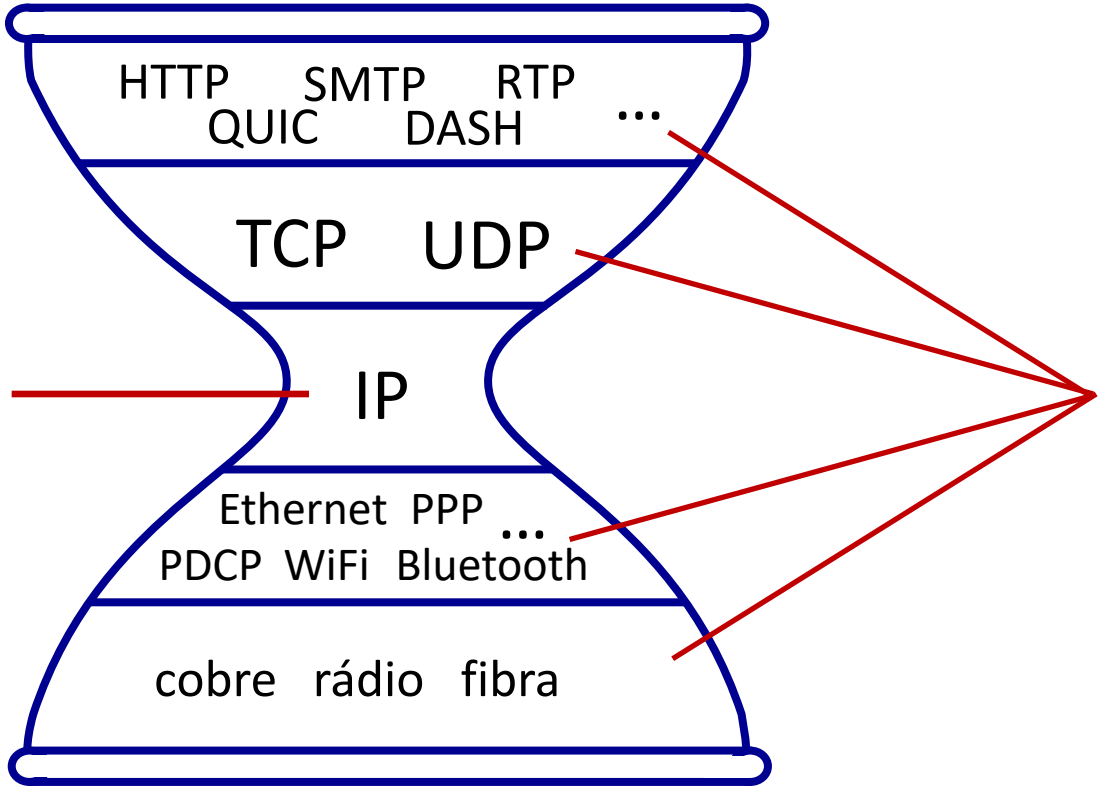
Middleboxes

- inicialmente: soluções de hardware proprietárias (fechadas)
- mudando para hardware “caixa branca” implementando API aberta
  - afastando-se de soluções de hardware proprietárias
  - ações locais programáveis via correspondência+ação
  - movimento em direção à inovação/diferenciação em software
- SDN: controle e gerenciamento de configuração centralizados (logicamente), muitas vezes em nuvem privada / pública
- network functions virtualization (NFV – virtualização de funções de rede): serviços programáveis sobre rede, computação, e armazenamento de caixa branca

A ampulheta IP

A “cintura fina” da Internet:

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

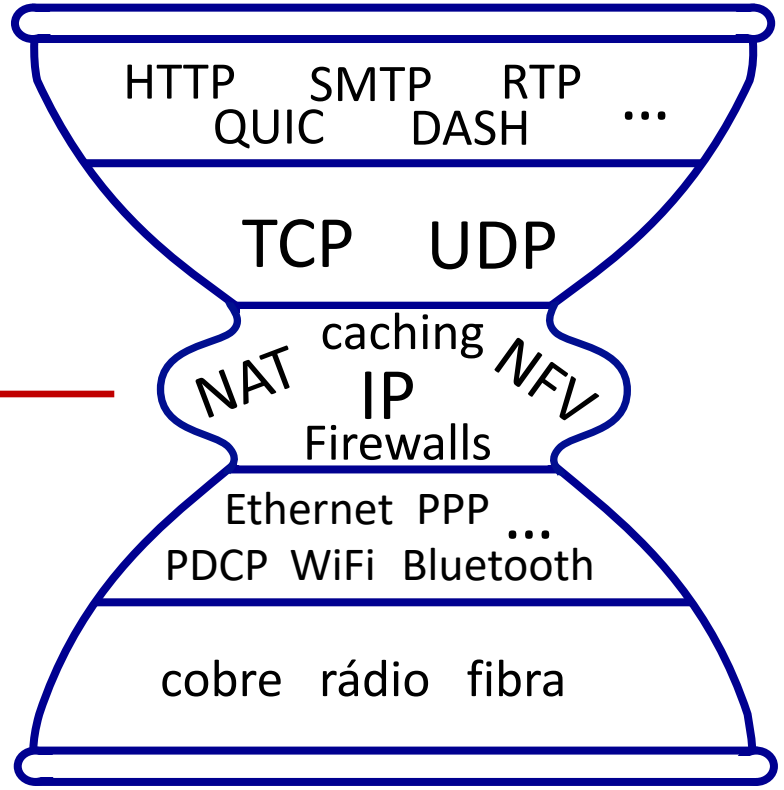


muitos protocolos nas camadas física, de enlace, transporte e aplicação

A ampulheta IP, na meia-idade

"alças do amor" da Internet de meia idade?

- middleboxes, operando dentro da rede



Princípios arquitetônicos da Internet

RFC 1958

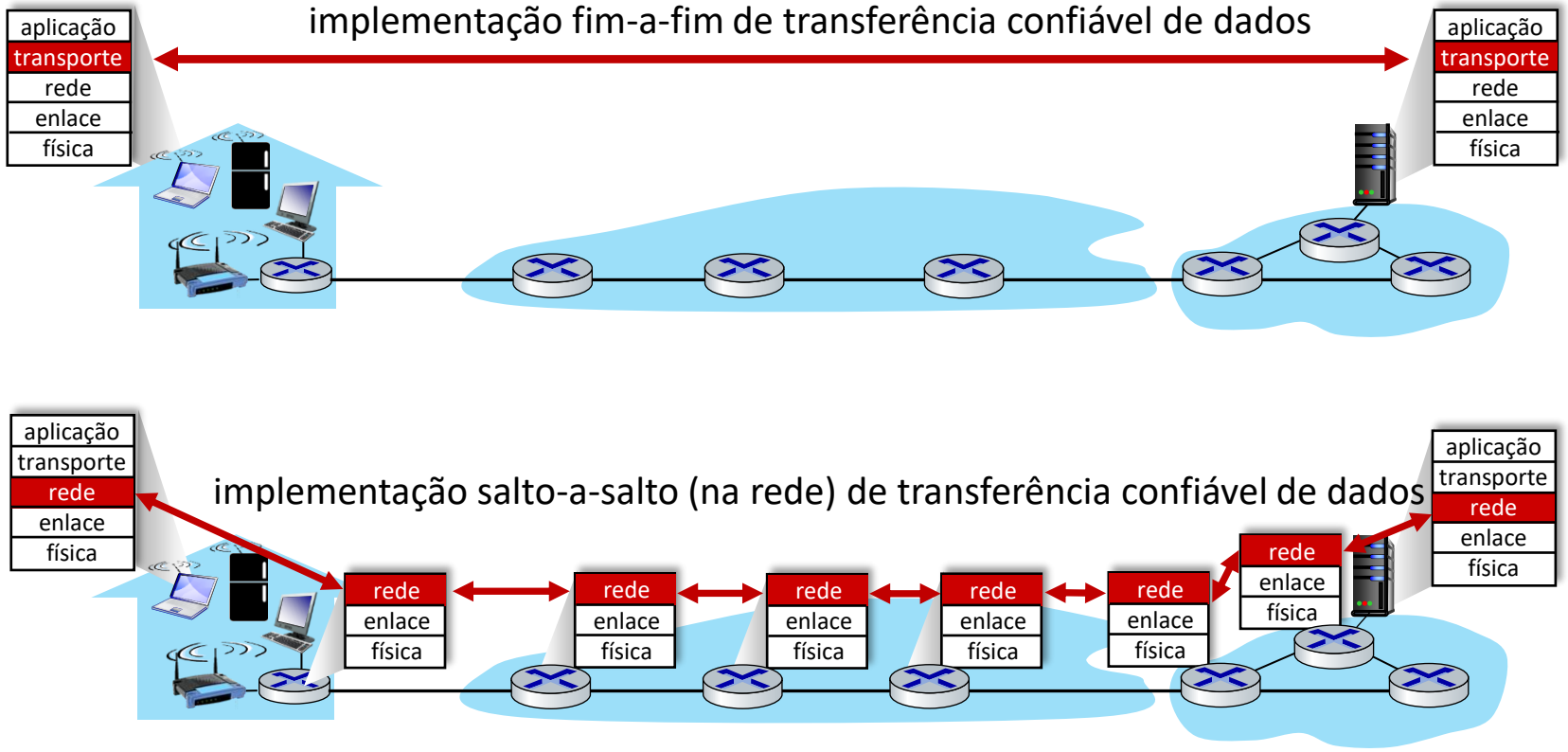
"Muitos membros da comunidade da Internet argumentariam que não existe arquitetura, mas apenas uma tradição, que não foi escrita nos primeiros 25 anos (ou pelo menos não pelo IAB). No entanto, em termos muito gerais, a comunidade acredita que o objetivo é a conectividade, a ferramenta é o Protocolo da Internet (IP) e a inteligência é ponta a ponta, e não escondida na rede

Três crenças fundamentais:

- conectividade simples
- protocolo IP: aquela cintura estreita
- inteligência, complexidade na borda da rede

O argumento fim-a-fim

- alguma funcionalidade de rede (por exemplo, transferência de dados confiável, congestionamento) pode ser implementada na rede, ou na borda da rede



O argumento fim-a-fim

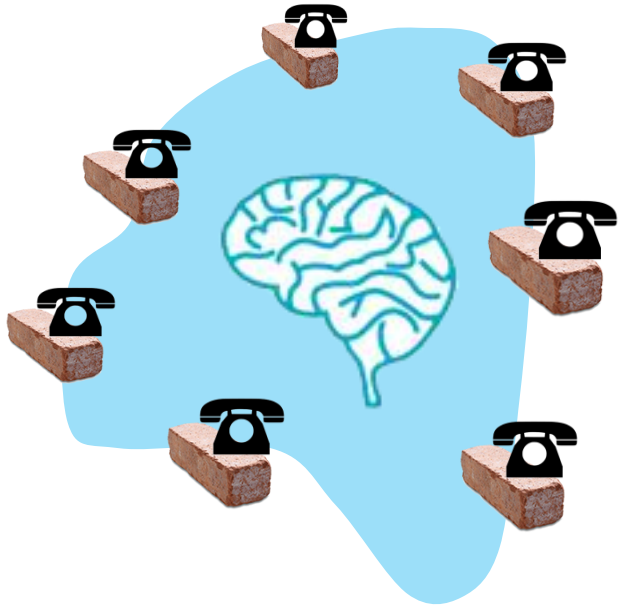
- alguma funcionalidade de rede (por exemplo, transferência de dados confiável, congestionamento) pode ser implementada na rede, ou na borda da rede

> "A função em questão pode ser implementada completa e corretamente apenas com o conhecimento e a ajuda da aplicação que está nos pontos finais do sistema de comunicação. Portanto, fornecer essa função questionada como uma característica do próprio sistema de comunicação não é possível. (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 fim-a-fim.'"

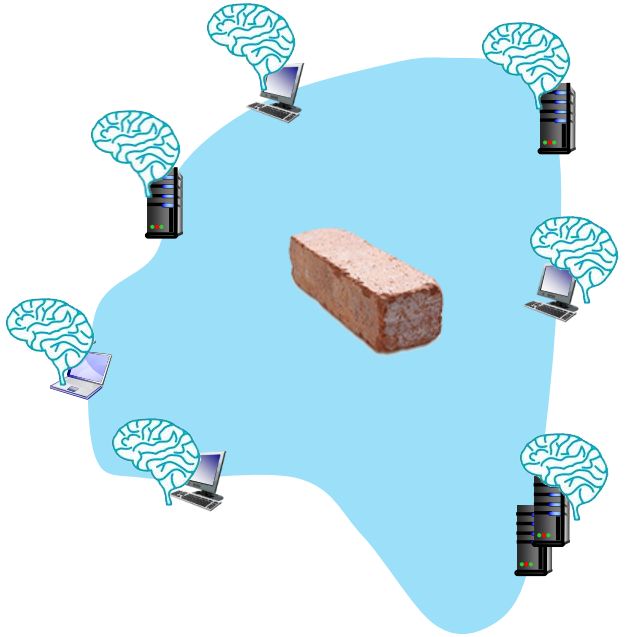
Saltzer, Reed, Clark 1981

Onde está a inteligência?



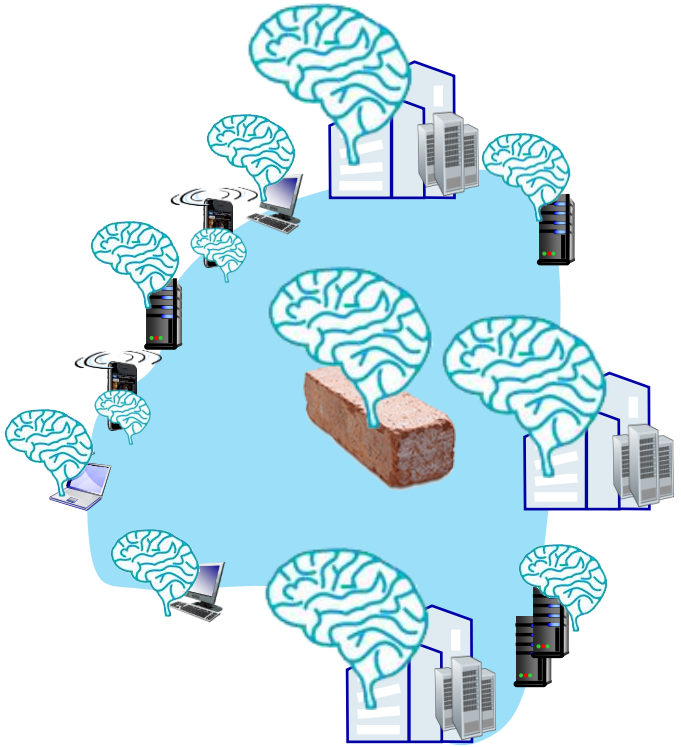
Rede telefônica do século 20:

- inteligência/computação em comutadores de rede



Internet (pré-2005)

- inteligência, computação na borda



Internet (pós-2005)

- dispositivos de rede programáveis
- inteligência, computação, e infraestrutura maciça de nível de aplicação na borda

Capítulo 4: concluído!

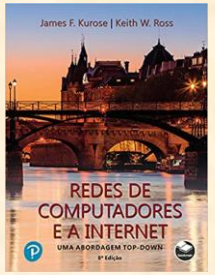
- Camada de rede: visão geral
- O que há dentro de um roteador
- IP: o Protocolo da Internet
- Encaminhamento Generalizado, SDN
- Middleboxes

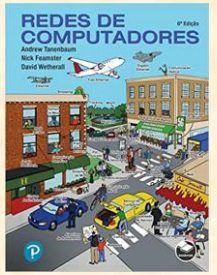
Questão: como as tabelas de encaminhamento (encaminhamento baseado em destino) ou tabelas de fluxo (encaminhamento generalizado) são calculadas?

Resposta: pelo plano de controle (próximo capítulo)

Leitura Recomendada e Complementar

- Leitura Recomendada
  - KUROSE, James F. e ROSS, Keith W. Redes de computadores e a Internet: Uma abordagem top-down. 8ª Edição. Bookman, 2021.
    - Capítulo 4 – Camada de Rede: O Plano de Dados
  - TANENBAUM, Andrew S., FEAMSTER, Nick e WETHERALL, David. Redes de Computadores. 5ª Edição. São Paulo: Bookman, 2021.
    - Capítulo 5 – A Camada de Rede.
- Leitura Complementar
  - FOUROUZAN, Behrouz A. e FIROUZ, Mosharraf. Redes de Computadores: uma abordagem top-down. Porto Alegre: AMGH, 2013.
    - Capítulo 4 – Camada de Rede.
  - TORRES, Gabriel. Redes de Computadores: Curso Completo. Axcel Books, 2001.
    - Capítulo 3 – TCP/IP e Capítulo 18 – Roteadores.
  - COMER, Douglas E. Interligação de Redes com TCP/IP. Volume 1: Princípios, protocolos e arquitetura. 6ª Edição. Rio de Janeiro: Elsevier, 2015.
    - Capítulos 6 a 9

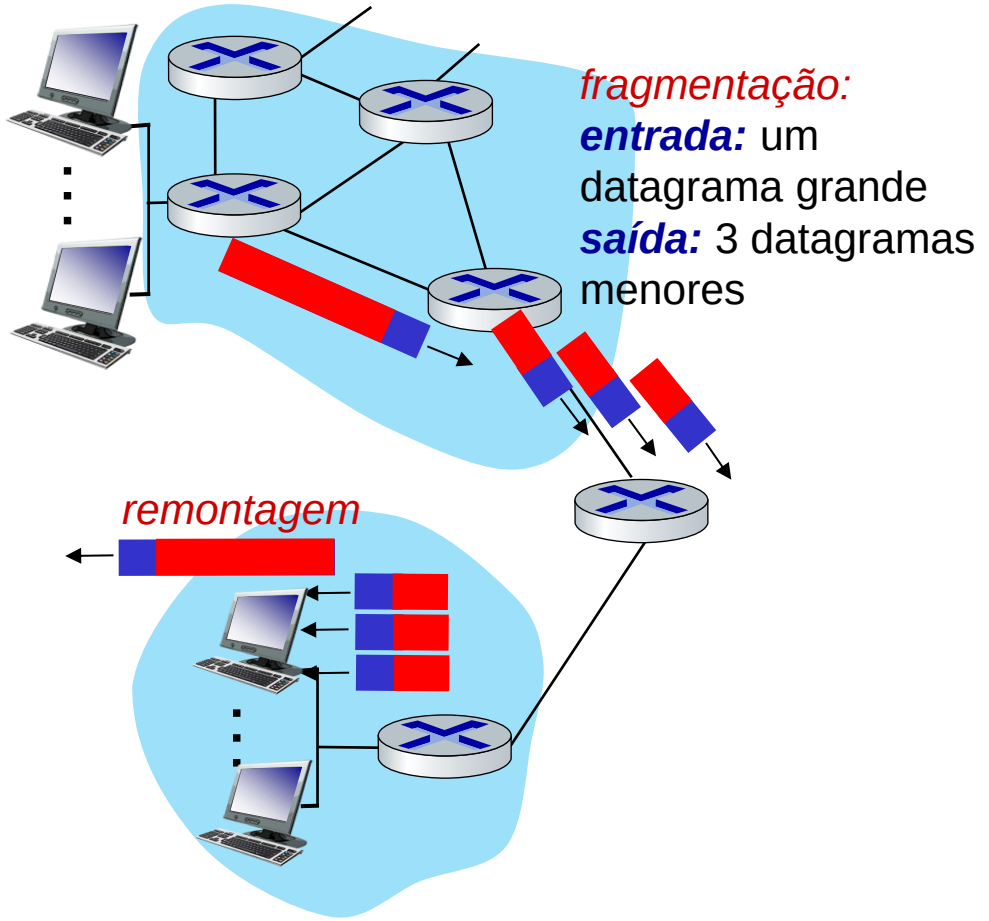


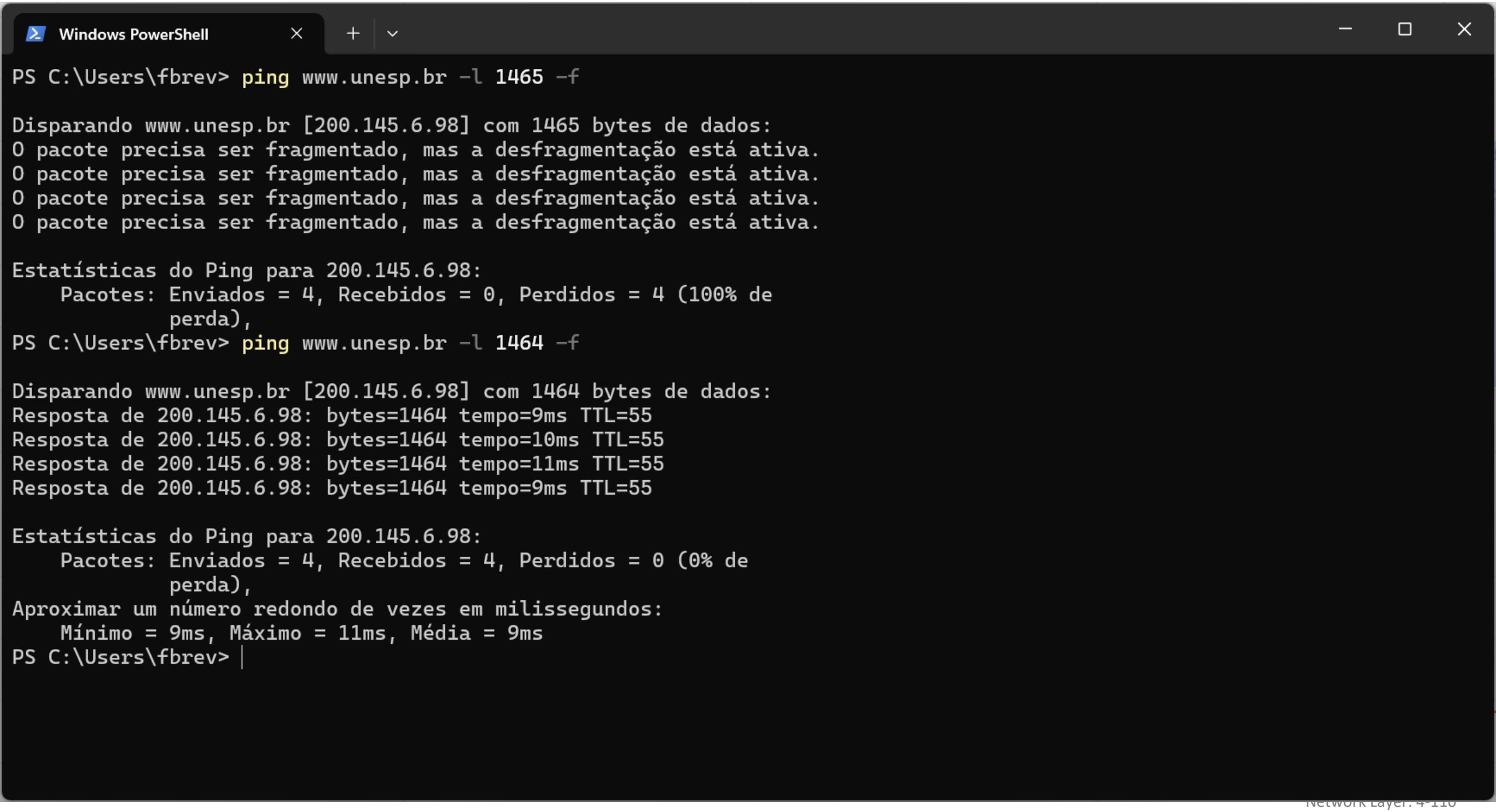


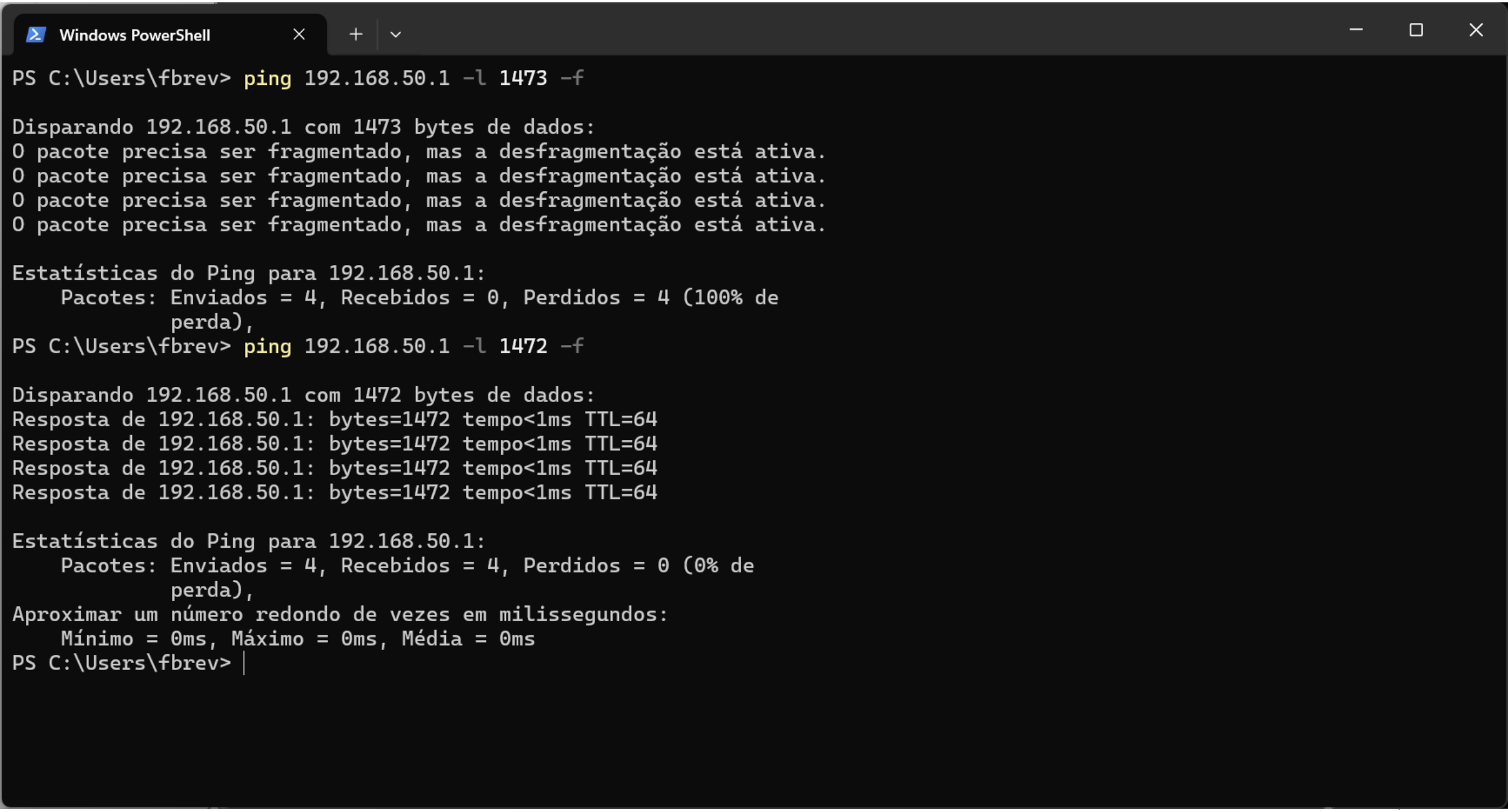
Slides adicionais do Capítulo 4

Fragmentação / remontagem do IP

- enlaces de rede tem MTU (tamanho máximo de transferência) - maior quadro de nível de enlace possível
  - diferentes tipos de enlace, diferentes MTUs
- datagrama IP grande dividido ("fragmentado") dentro da rede
  - um datagrama se torna vários datagramas
  - "remontados" apenas no destino
  - Bits de cabeçalho IP usados para identificar e ordenar fragmentos relacionados




Windows PowerShell
PS C:\Users\fbrev> ping www.unesp.br -l 1465 -f
Disparando www.unesp.br [200.145.6.98] com 1465 bytes de dados:
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
Estatísticas do Ping para 200.145.6.98:
Pacotes: Enviados = 4, Recebidos = 0, Perdidos = 4 (100% de
perda),
PS C:\Users\fbrev> ping www.unesp.br -l 1464 -f
Disparando www.unesp.br [200.145.6.98] com 1464 bytes de dados:
Resposta de 200.145.6.98: bytes=1464 tempo=9ms TTL=55
Resposta de 200.145.6.98: bytes=1464 tempo=10ms TTL=55
Resposta de 200.145.6.98: bytes=1464 tempo=11ms TTL=55
Resposta de 200.145.6.98: bytes=1464 tempo=9ms TTL=55
Estatísticas do Ping para 200.145.6.98:
Pacotes: Enviados = 4, Recebidos = 4, Perdidos = 0 (0% de
perda),
Aproximar um número redondo de vezes em milissegundos:
Mínimo = 9ms, Máximo = 11ms, Média = 9ms
PS C:\Users\fbrev>


Windows PowerShell
PS C:\Users\fbrev> ping 192.168.50.1 -l 1473 -f
Disparando 192.168.50.1 com 1473 bytes de dados:
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
O pacote precisa ser fragmentado, mas a desfragmentação está ativa.
Estatísticas do Ping para 192.168.50.1:
Pacotes: Enviados = 4, Recebidos = 0, Perdidos = 4 (100% de
perda),
PS C:\Users\fbrev> ping 192.168.50.1 -l 1472 -f
Disparando 192.168.50.1 com 1472 bytes de dados:
Resposta de 192.168.50.1: bytes=1472 tempo<1ms TTL=64
Resposta de 192.168.50.1: bytes=1472 tempo<1ms TTL=64
Resposta de 192.168.50.1: bytes=1472 tempo<1ms TTL=64
Resposta de 192.168.50.1: bytes=1472 tempo<1ms TTL=64
Estatísticas do Ping para 192.168.50.1:
Pacotes: Enviados = 4, Recebidos = 4, Perdidos = 0 (0% de
perda),
Aproximar um número redondo de vezes em milissegundos:
Mínimo = 0ms, Máximo = 0ms, Média = 0ms
PS C:\Users\fbrev>

DHCP: saída do Wireshark (LAN doméstica)

requisição

Message type: **Boot Request (1)**
Hardware type: Ethernet
Hardware address length: 6
Hops: 0
Transaction ID: 0x6b3a11b7
Seconds elapsed: 0
Bootp flags: 0x0000 (Unicast)
Client IP address: 0.0.0.0 (0.0.0.0)
Your (client) IP address: 0.0.0.0 (0.0.0.0)
Next server IP address: 0.0.0.0 (0.0.0.0)
Relay agent IP address: 0.0.0.0 (0.0.0.0)
Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)
Server host name not given
Boot file name not given
Magic cookie: (OK)
Option: (t=53,l=1) **DHCP Message Type = DHCP Request**
Option: (61) Client identifier
    Length: 7; Value: 010016D323688A;
    Hardware type: Ethernet
    Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)
Option: (t=50,l=4) Requested IP Address = 192.168.1.101
Option: (t=12,l=5) Host Name = "nomad"
Option: (55) Parameter Request List
    Length: 11; Value: 010F03062C2E2F1F21F92B
    1 = Subnet Mask; 15 = Domain Name
    3 = Router; 6 = Domain Name Server
    44 = NetBIOS over TCP/IP Name Server
    ……

resposta

Message type: **Boot Reply (2)**
Hardware type: Ethernet
Hardware address length: 6
Hops: 0
Transaction ID: 0x6b3a11b7
Seconds elapsed: 0
Bootp flags: 0x0000 (Unicast)
Client IP address: 192.168.1.101 (192.168.1.101)
Your (client) IP address: 0.0.0.0 (0.0.0.0)
Next server IP address: 192.168.1.1 (192.168.1.1)
Relay agent IP address: 0.0.0.0 (0.0.0.0)
Client MAC address: Wistron_23:68:8a (00:16:d3:23:68:8a)
Server host name not given
Boot file name not given
Magic cookie: (OK)
Option: (t=53,l=1) DHCP Message Type = DHCP ACK
Option: (t=54,l=4) Server Identifier = 192.168.1.1
Option: (t=1,l=4) Subnet Mask = 255.255.255.0
Option: (t=3,l=4) Router = 192.168.1.1
Option: (6) Domain Name Server
    Length: 12; Value: 445747E2445749F244574092;
    IP Address: 68.87.71.226;
    IP Address: 68.87.73.242;
    IP Address: 68.87.64.146
Option: (t=15,l=20) Domain Name = "hsd1.ma.comcast.net."