Redes de Computadores

Internet Protocol (IP)

Princípios da camada de rede
Endereçamento global
Protocolos auxiliares (ARP, ICMP, DHCP)
Roteamento interno (RIP, OSPF)
Roteamento entre sistemas autônomos (BGP)
IPv6

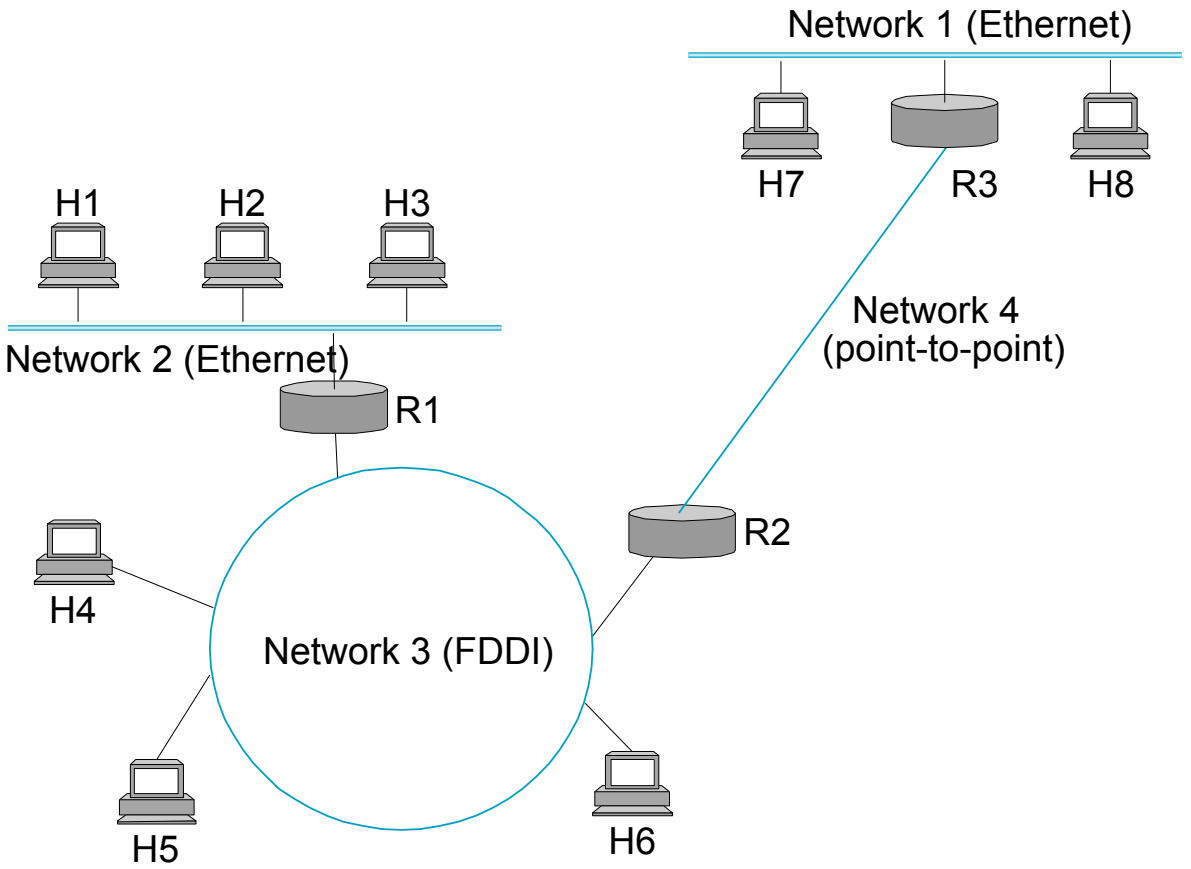
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Motivação para interconexão

- Diferentes tecnologias de rede possuem diferentes características:
  - LANs: alta velocidade, pequena distância
  - WANs: comunicação em uma grande área
  - Não existe uma única tecnologia que seja melhor em todos os casos

Motivação para interconexão

- É comum uma grande organização ter várias redes físicas, cada uma adequada para um determinado tipo de ambiente



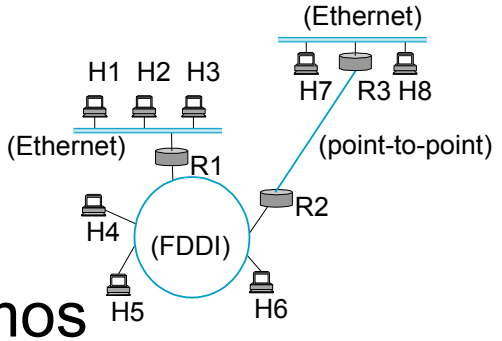
Funções básicas da camada de rede:

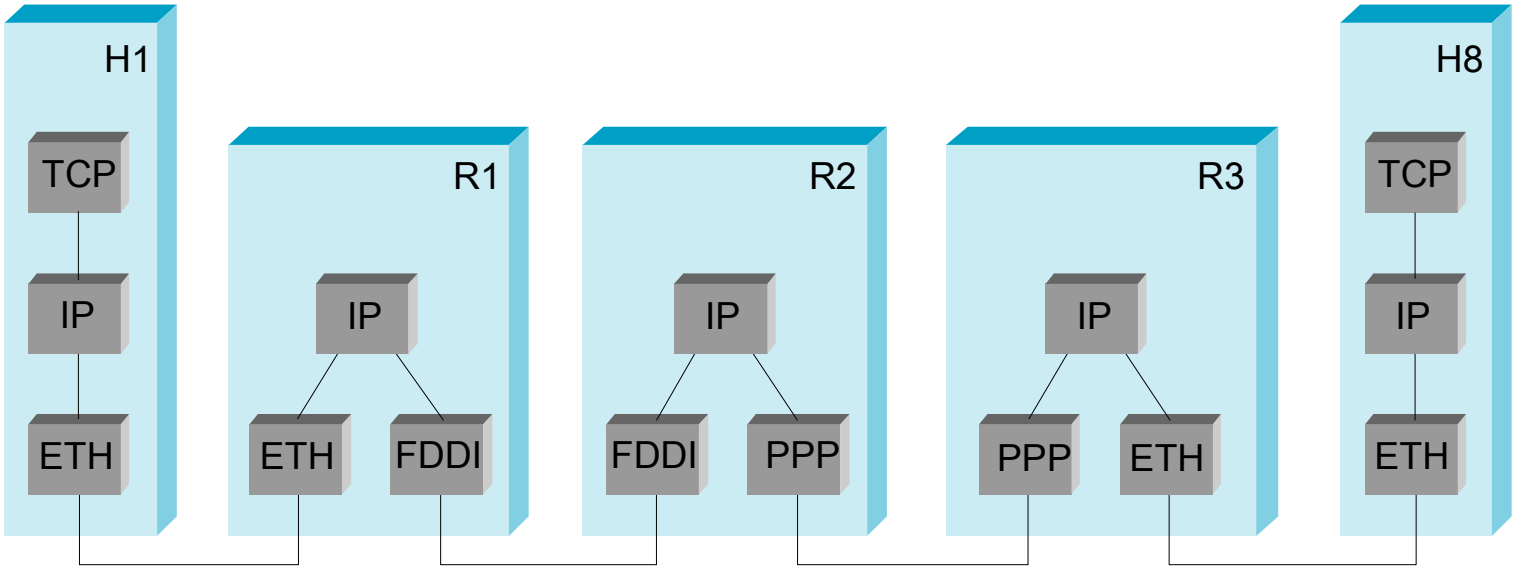
- Endereçamento
  - identificação de cada máquina,
    independente de sua localização ou da tecnologia
- Roteamento
  - determinação de um caminho
    entre duas máquinas quaisquer da internet

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Interconexão

- Conexão entre as redes é feita por roteadores
  - Computador de finalidade especial: interconexão
  - Deve ser capaz de lidar com diferentes tecnologias
- A rede virtual resultante deve oferecer os mesmos serviços em todos os seus pontos





UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Roteiro

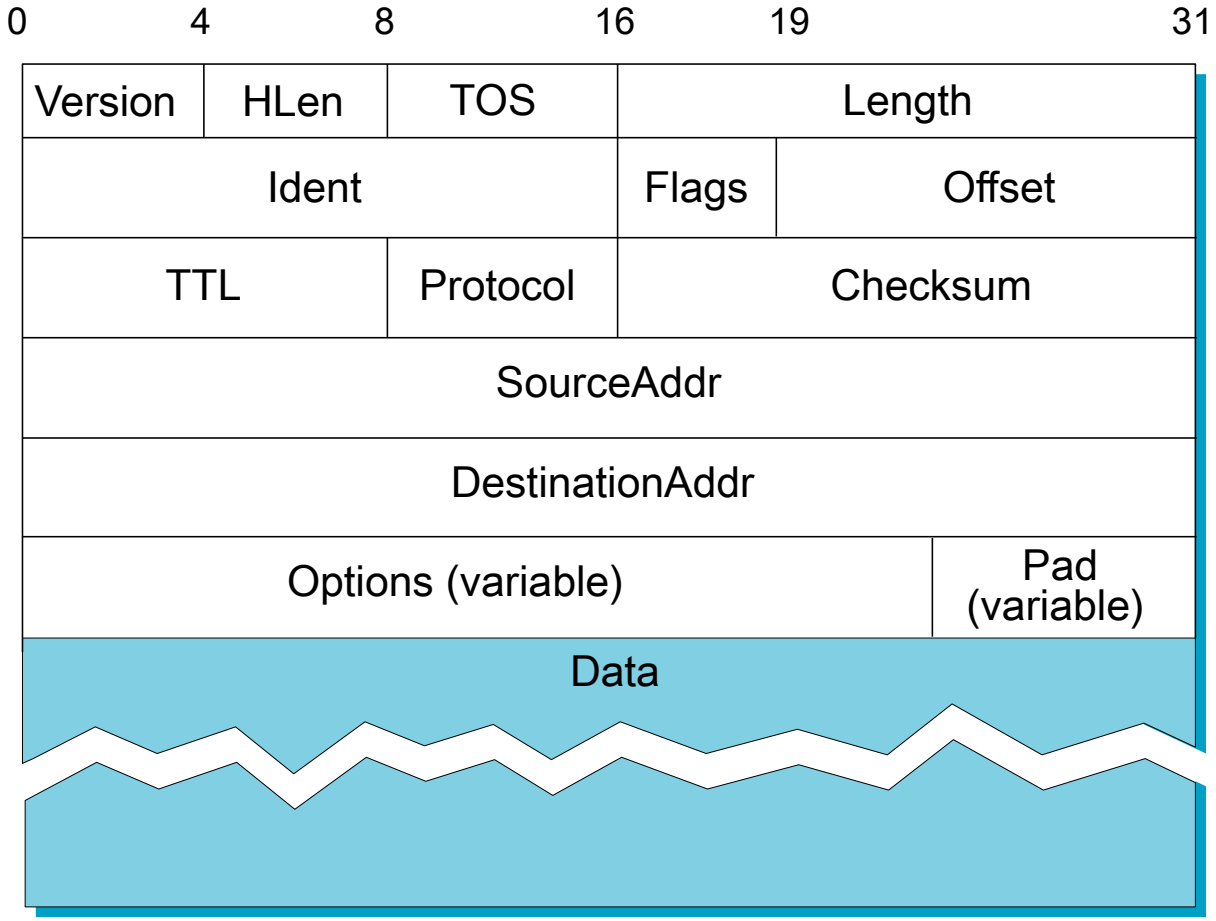
- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

Modelo de serviço

- Sem conexão (baseado em datagramas)
- Entrega segundo “melhor esforço possível” (*best-effort delivery*)
- Serviço não confiável:
  - Pacotes são perdidos
  - Pacotes são entregues fora de ordem
  - Várias cópias de um pacote podem ser entregues
  - Pacotes podem ser atrasados por muito tempo

Modelo de serviço

- Formato dos pacotes



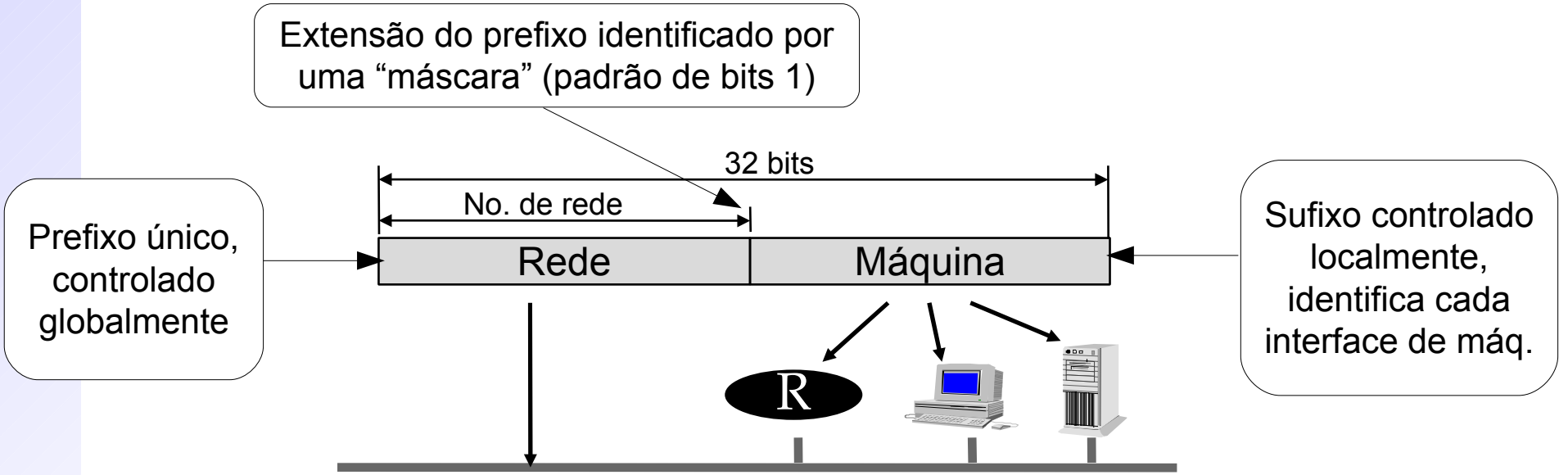
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Endereçamento global

- Componente crítico da abstração fornecida pela Internet
- Independente de endereços físicos, como os usados em redes locais
- Ajuda a criar a ilusão de uma rede única e integrada
- Usuários, aplicações e protocolos de alto nível usam endereços abstratos para se comunicar

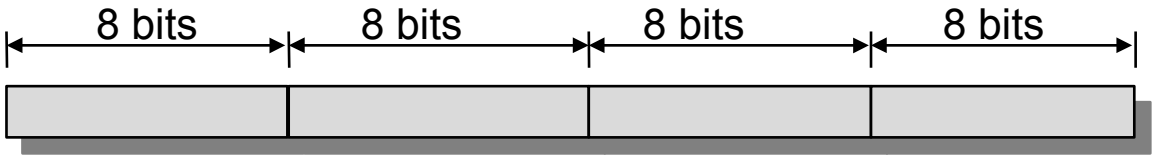
Endereço IP

- Dividido em endereço de rede (prefixo) e máquina



Endereços IP

- Notação de ponto decimal
  - 32 bits normalmente visualizados como quatro bytes
  - Máscara identificada pelo seu comprimento ou como padrão de 1's



  - Exemplos:

```
10000001 00110100 00000110 00000000 -> 129.52.6.0/24
11000000 00000101 00110000 00000011 -> 192.5.48.3/26
00001010 00000010 00000000 00100101 -> 10.2.0.37/20
10000000 00001010 00000010 00000011 -> 128.10.2.3 másc 255.255.0.0
10000000 10000000 11111111 00000000 -> 128.128.255.0/18
```

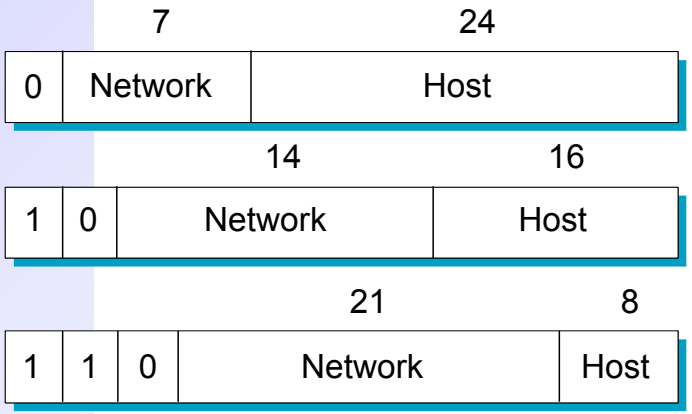
    :

Endereços IP

- Alguns endereços especiais reservados:
  - Sufixo todo em zeros: endereço identifica a rede
  - Sufixo todo em uns: endereço de *broadcast* (p/ todas as máquinas)
  - Outras faixas especiais para redes isoladas da Internet
    - P.ex., 10.x.x.x, 192.168.x.x

Classes de endereços (não mais usadas)

- Originalmente, tamanho dos prefixos era pré-definido



Classe A (máscara: 255.0.0.0) ----------> 126 redes, 16.777.214 hosts/rede

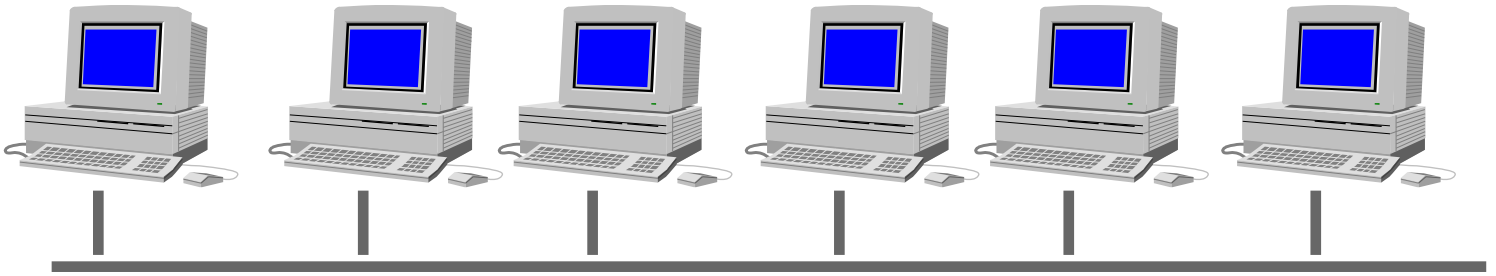
Classe B (máscara: 255.255.0.0) -------> 16.384 redes, 65.534 hosts/rede

Classe C (máscara: 255.255.255.0) --->2.097.152 redes, 254 hosts/rede

- Mais recentemente classes de prefixos foram abolidas
  - agora o comprimento do prefixo é um atributo explícito em cada caso
  - na maior parte dos casos, redes já alocadas mantiveram suas classes
    - p.ex., UFMG: 150.164.0.0 / 16

Segmentação de rede (*subnetting*)

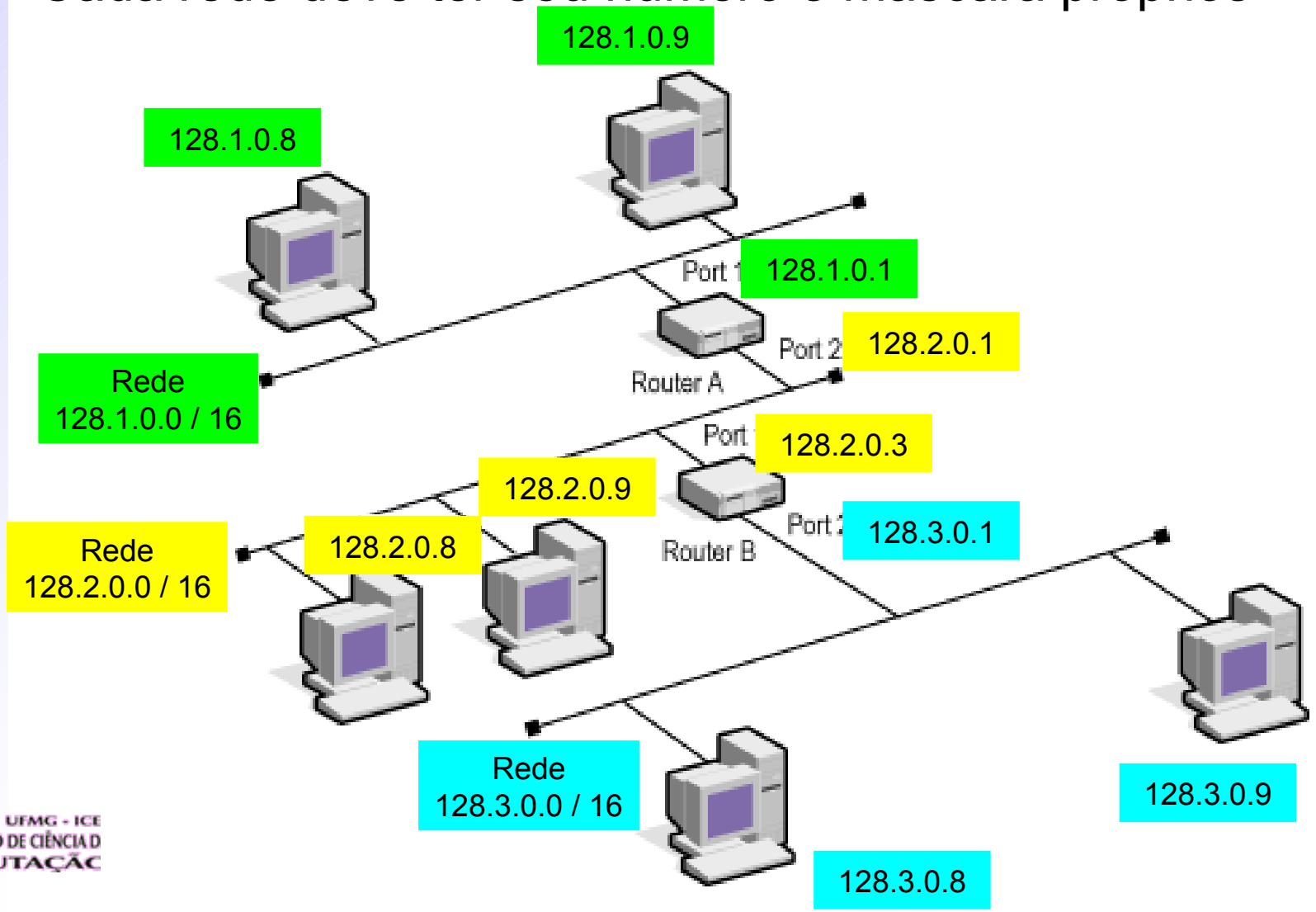
- Cada número de rede tem sua máscara padrão
- A máscara permite definir se um endereço é local ou não



131.108.0.0/16

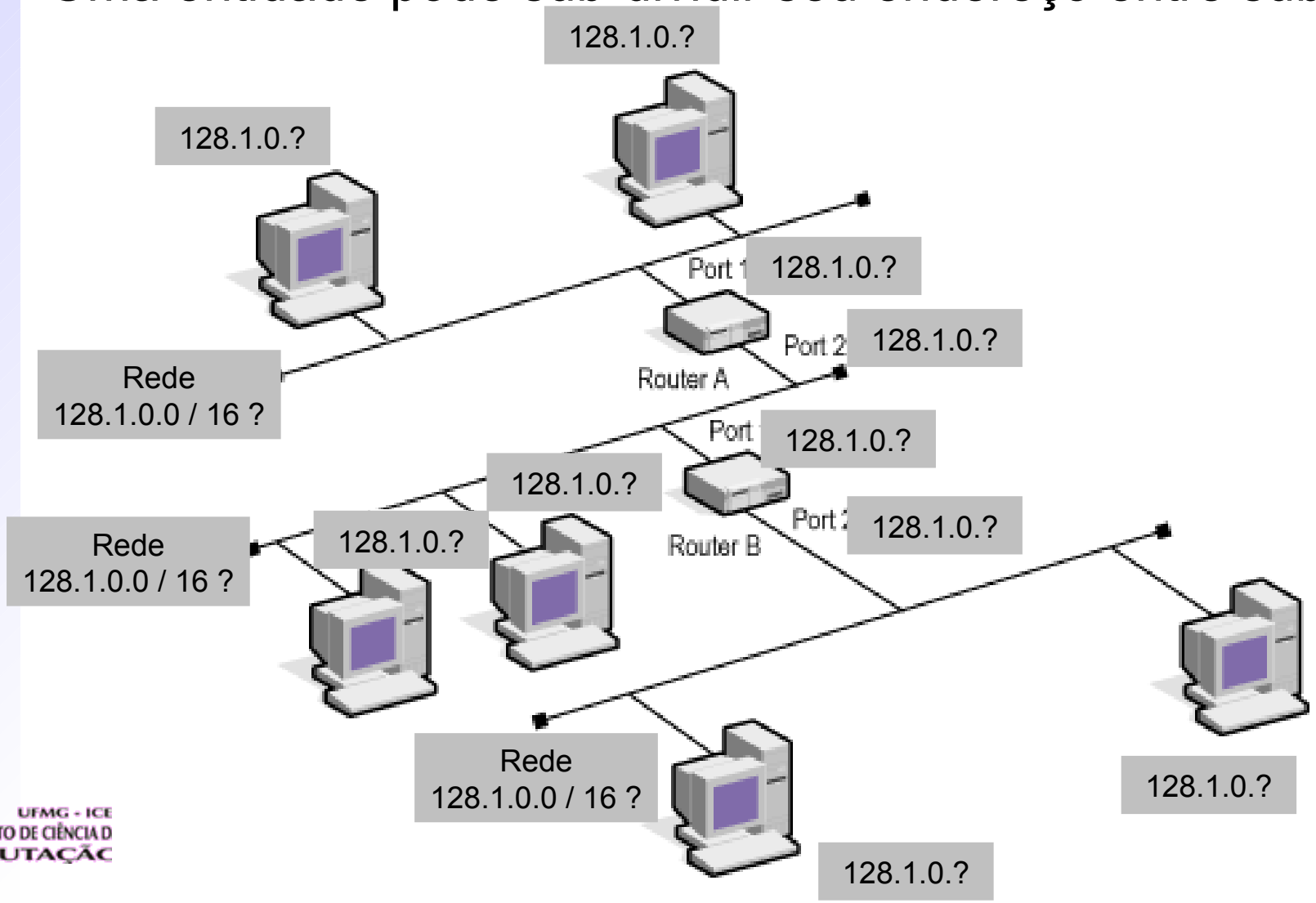
Segmentação de rede (*subnetting*)

- Cada rede deve ter seu número e máscara próprios



Segmentação de rede (*subnetting*)

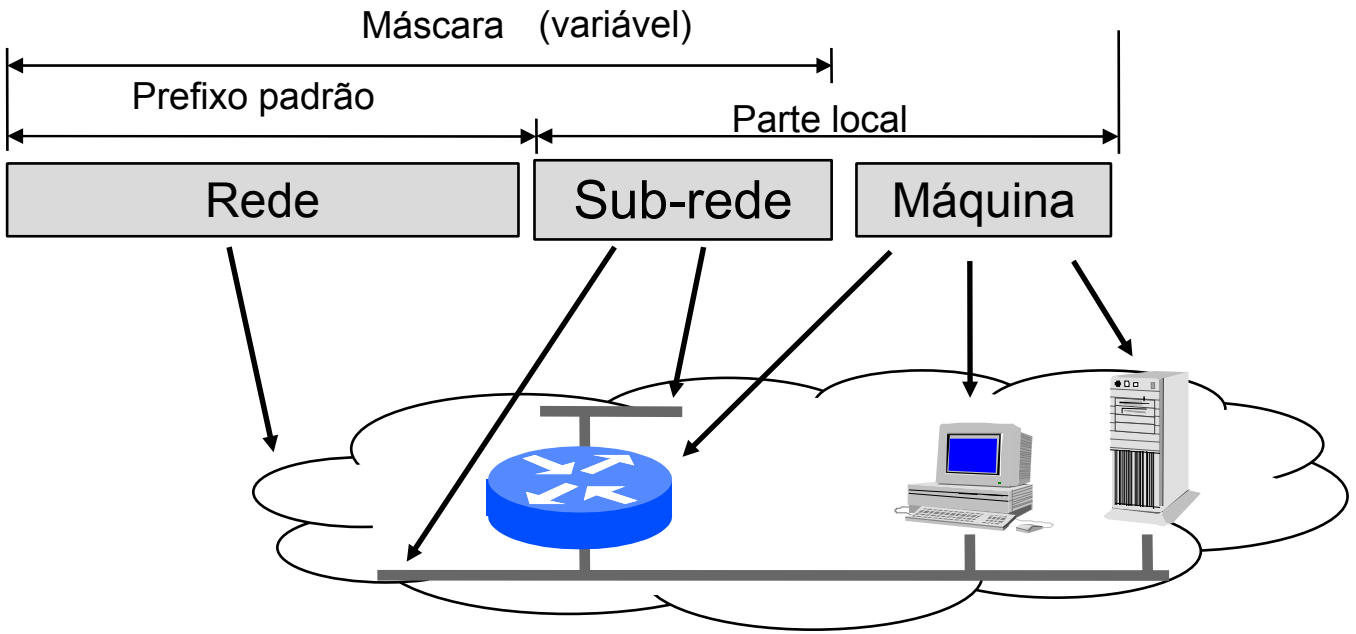
- Uma entidade pode sub-dividir seu endereço entre sub-redes



UFMG - ICE
DEPARTAMENTO DE CIÊNCIA D
COMPUTAÇÃC

Segmentação de rede (*subnetting*)

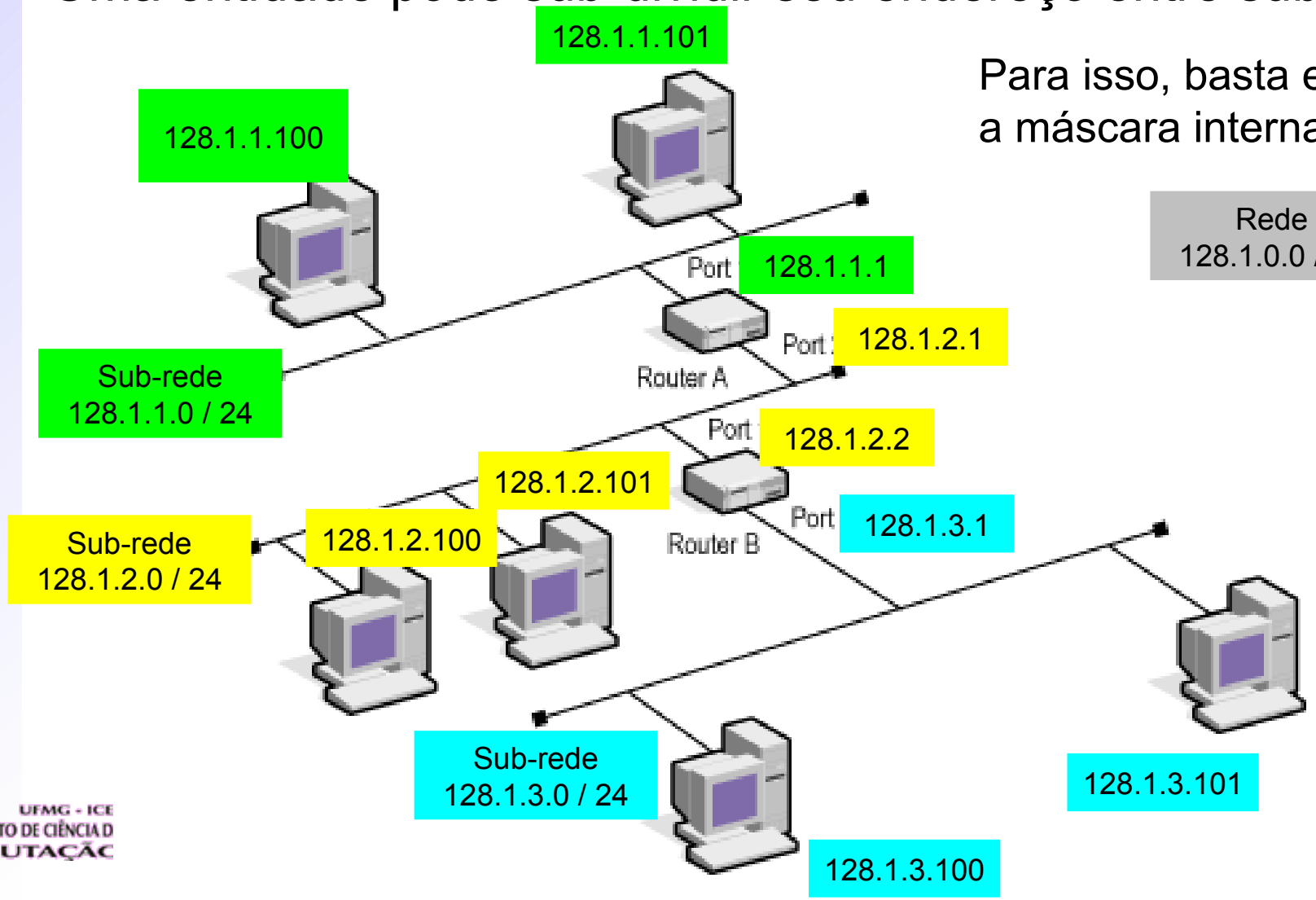
- *Máscara de sub-rede*
  - Também representada na forma a.b.c.d (p.ex.: 255.255.255.240)



Segmentação de rede (*subnetting*)

- Uma entidade pode sub-dividir seu endereço entre sub-redes

Para isso, basta estender a máscara internamente



UFMG - ICE
DEPARTAMENTO DE CIÊNCIA D
COMPUTAÇÃC

Roteiro

- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

Expedição de datagramas

- cada datagrama contém o endereço do destino
- cada interface tem seu endereço e sua máscara
- comportamento depende de o endereço de destino estar na mesma rede à que pertende a interface:

  SE ((end_destino & mascara) == (end_interface & mascara))

  envia datagrama diretamente ao host de destino

  SENÃO

  determina próximo passo (roteador) para o pacote

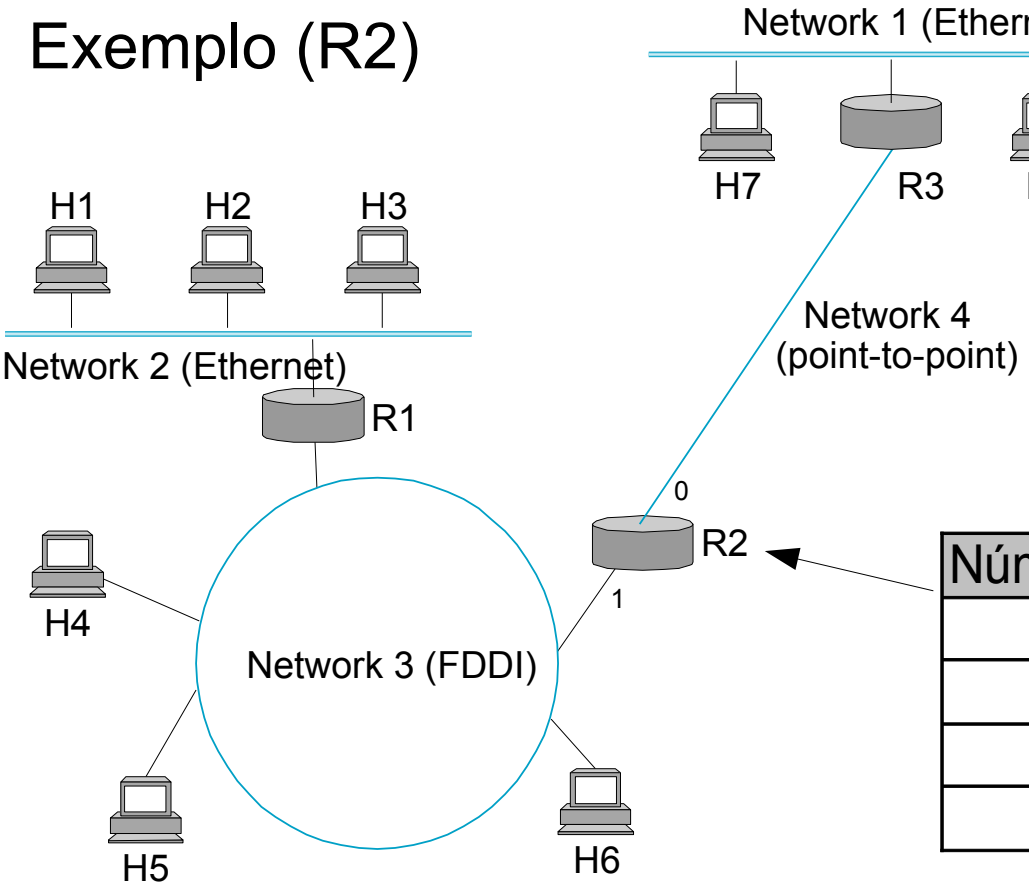
- tabela de expedição (tabela de roteamento)
  - mapeia endereços de rede para o próximo roteador
  - pode haver uma rota *default* – a ser usada na falta de uma rota específica

Estamos saltando: “Implementação do roteador”

UFMG - ICEx DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Expedição de datagramas

- Exemplo (R2)



Número da rede	Próximo passo
1	R3
2	R1
3	Interface 1
4	Interface 0

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Problemas associados à expedição

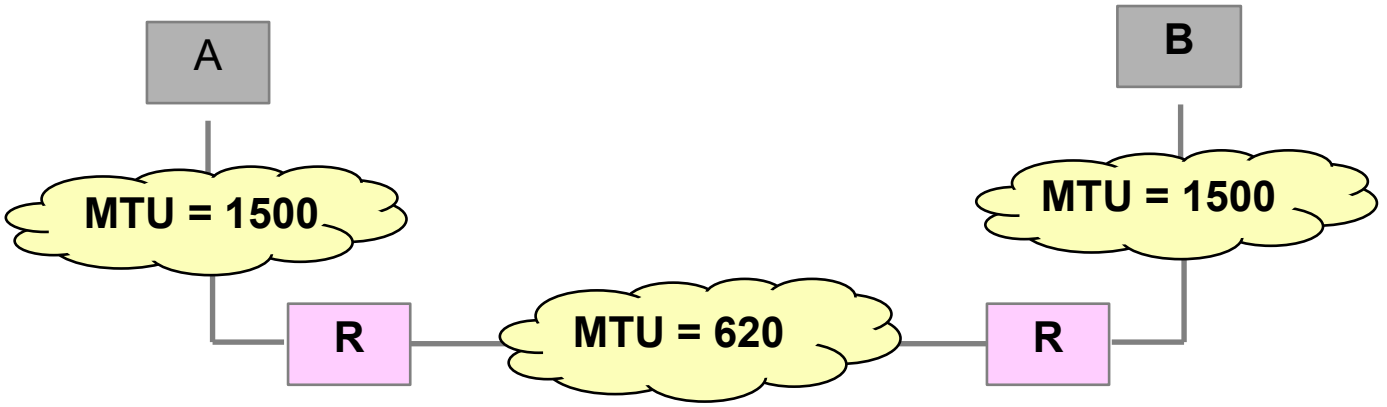
- Cada rede pode ter limites diferentes para o tamanho máximo de pacotes
  - É preciso ser capaz de enviar pacotes grandes em qualquer rede
  - Fragmentação e remontagem de pacotes

- Em uma rede, a entrega de pacotes depende dos endereços de enlace (rede local)
  - É preciso associar endereços IP locais a endereços físicos
  - Protocolo ARP

Fragmentação de pacotes

- A camada de rede de cada protocolo especifica um pacote máximo que pode ser enviado de cada vez:
  - MTU (*Maximum Transfer Unit*)

Qual o tamanho do datagrama neste caso?



Obs: em ATM usa-se a AAL, então não há fragmentação nesse nível

Fragmentação de pacotes

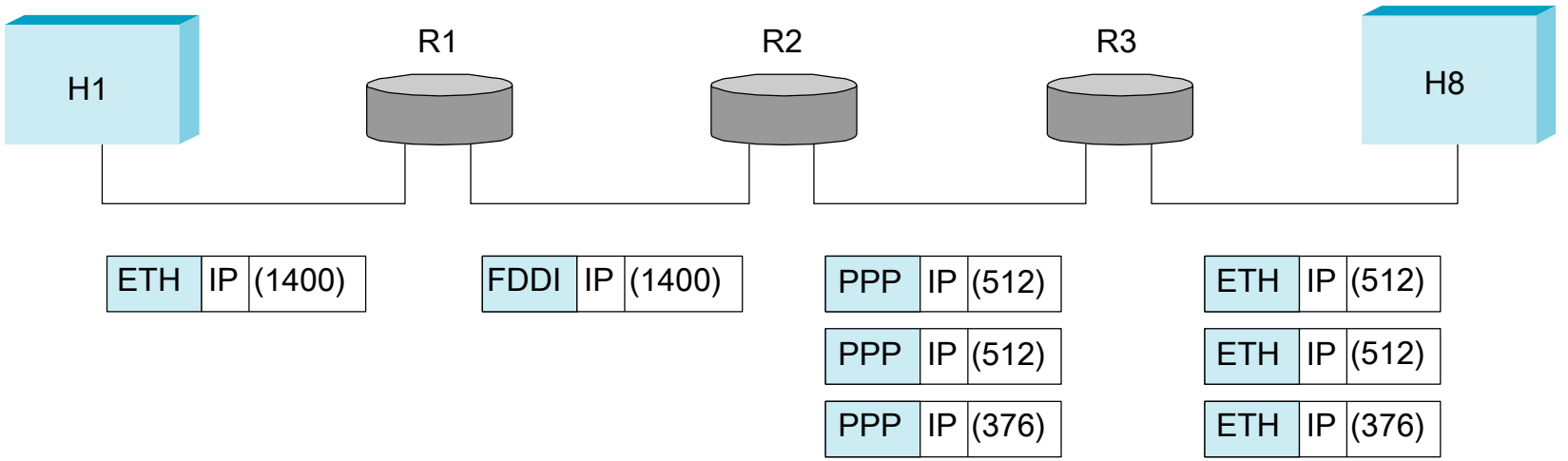
- Princípios básicos:
  - informação do cabeçalho original é mantida
  - fragmentar apenas se necessário (**MTU** < pacote)
  - tentar evitar fragmentação no nó de origem
  - é permitido re-fragmentar, se necessário
  - remontagem só no nó de destino
  - não tentar recuperar fragmentos perdidos

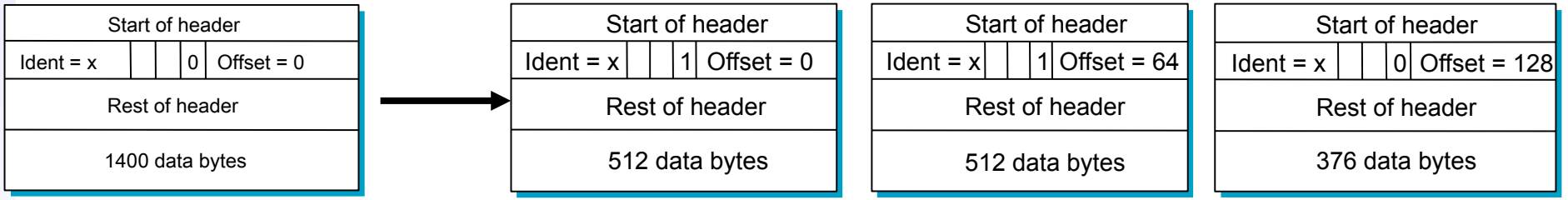
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Fragmentação de pacotes

- Campos *Flag*, Identificação e *Fragment Offset* são usados:
  - Identificação:
    - Fragmentos recebem identificação do pacote original
  - Flags:
    - DF (*don't fragment*): pacote não pode ser fragmentado
    - MF (*more fragments*): há outros fragmentos do mesmo pacote original
  - Fragment offset:
    - Indica posição do fragmento no pacote original

Exemplo


H1
R1
R2
R3
H8
ETH IP (1400)
FDDI IP (1400)
PPP IP (512)
PPP IP (512)
PPP IP (376)
ETH IP (512)
ETH IP (512)
ETH IP (376)


Start of header
Ident = x 0 Offset = 0
Rest of header
1400 data bytes
Start of header
Ident = x 1 Offset = 0
Rest of header
512 data bytes
Start of header
Ident = x 1 Offset = 64
Rest of header
512 data bytes
Start of header
Ident = x 0 Offset = 128
Rest of header
376 data bytes

Remontagem de pacotes

- Processo inverso ao da fragmentação
- Computador de destino é responsável
- O que ocorre se fragmentos são perdidos, chegam foram de ordem ou atrasados?
  - RX não tem como informar TX para re-enviar um fragmento, já que TX não conhece nada sobre fragmentação.
- Solução:
  - RX ao receber o primeiro fragmento inicializa um temporizador
  - Se todos os fragmentos não chegam antes do temporizador se esgotar então os fragmentos recebidos são descartados

Problemas associados à expedição

- Cada rede pode ter limites diferentes para o tamanho máximo de pacotes
  - É preciso ser capaz de enviar pacotes grandes em qualquer rede
  - Fragmentação e remontagem de pacotes

- Em uma rede, a entrega de pacotes depende dos endereços de enlace (rede local)
  - É preciso associar endereços IP locais a endereços físicos
  - Protocolo ARP

Como localizar o destino na rede física?

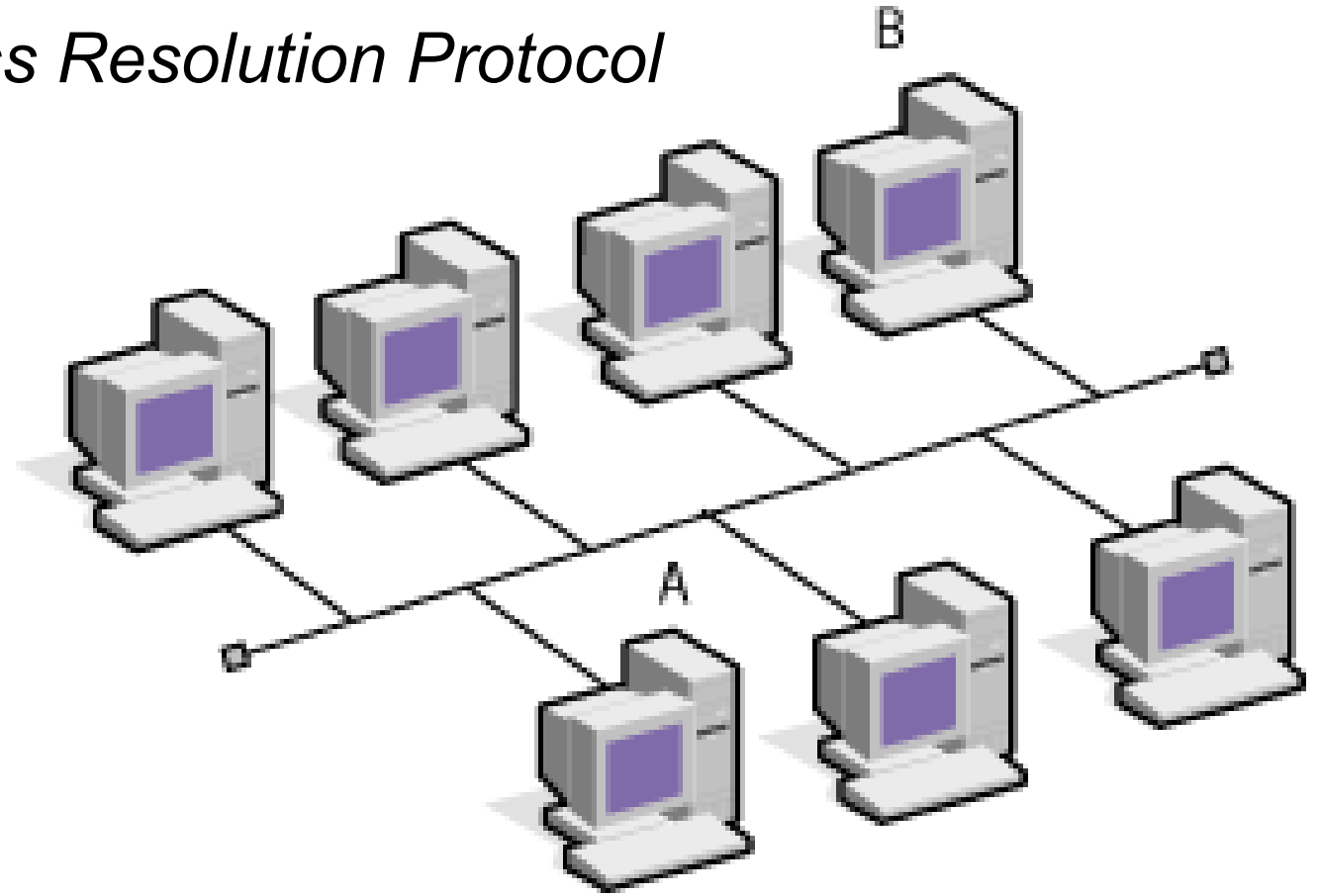
- Pacotes IP trazem endereços IP, mas para entregá-los ao destino é preciso conhecer o endereço físico
- Endereço físico não tem nada a ver com endereço IP, mas programas só usam IP
  - Em redes Ethernet, por exemplo, uma máquina só recebe um pacote se ele contém o seu endereço físico
- É preciso "perguntar" às máquinas da rede qual é o endereço físico da máquina que se deseja alcançar
- Protocolo utiliza mensagens *broadcast* para que todas as máquinas participem do processo

Tradução de endereços

- ARP (*Address Resolution Protocol*)
  - protocolo de tradução de endereços
  - gerencia cache de associações entre endereços IP e físicos
    - entradas são descartadas após aprox. 10 minutos sem utilização
    - tabela é atualizada mesmo se a entrada já existe

ARP

Address Resolution Protocol

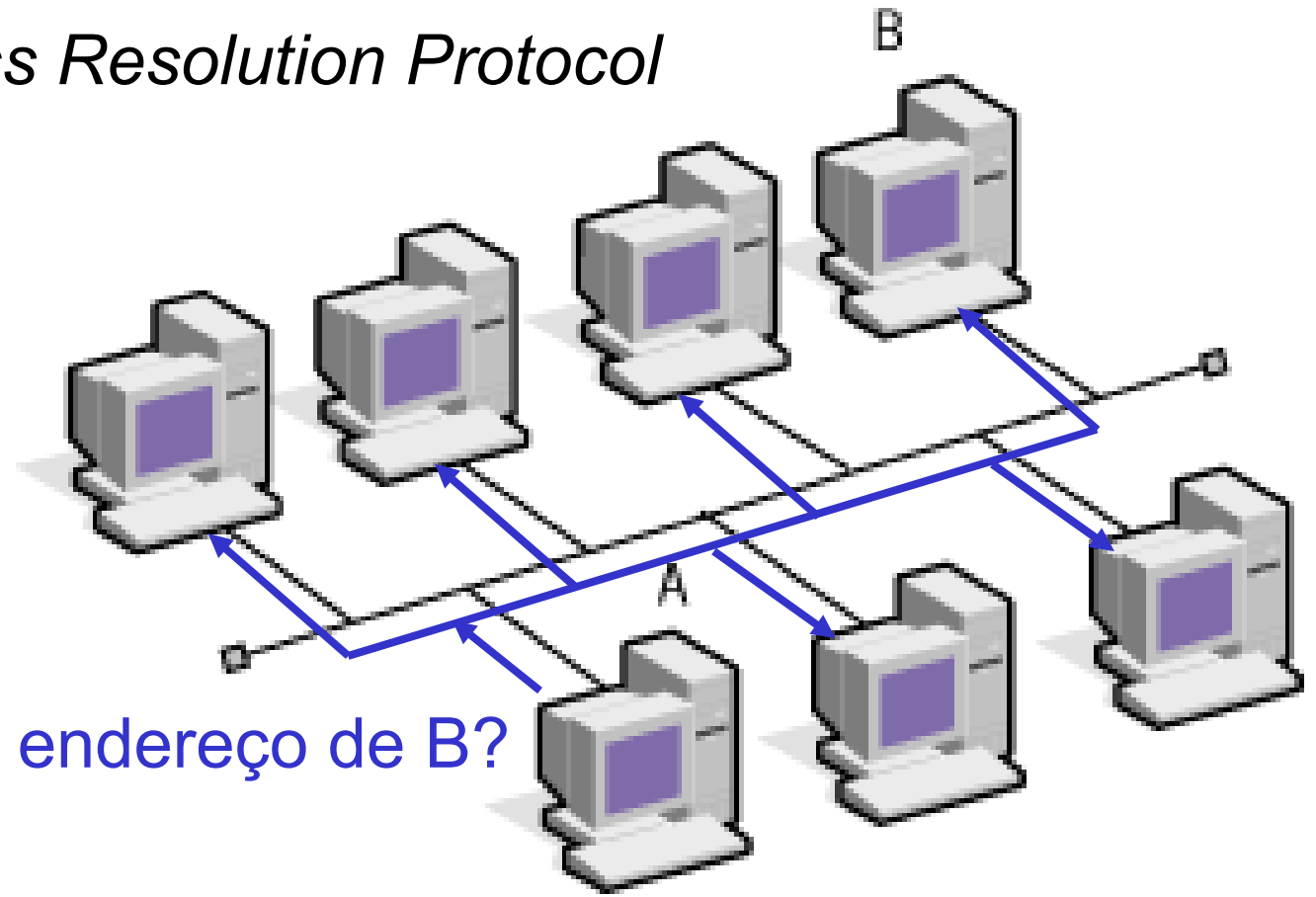


Como A determina endereço local (hardware) de B?

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

ARP

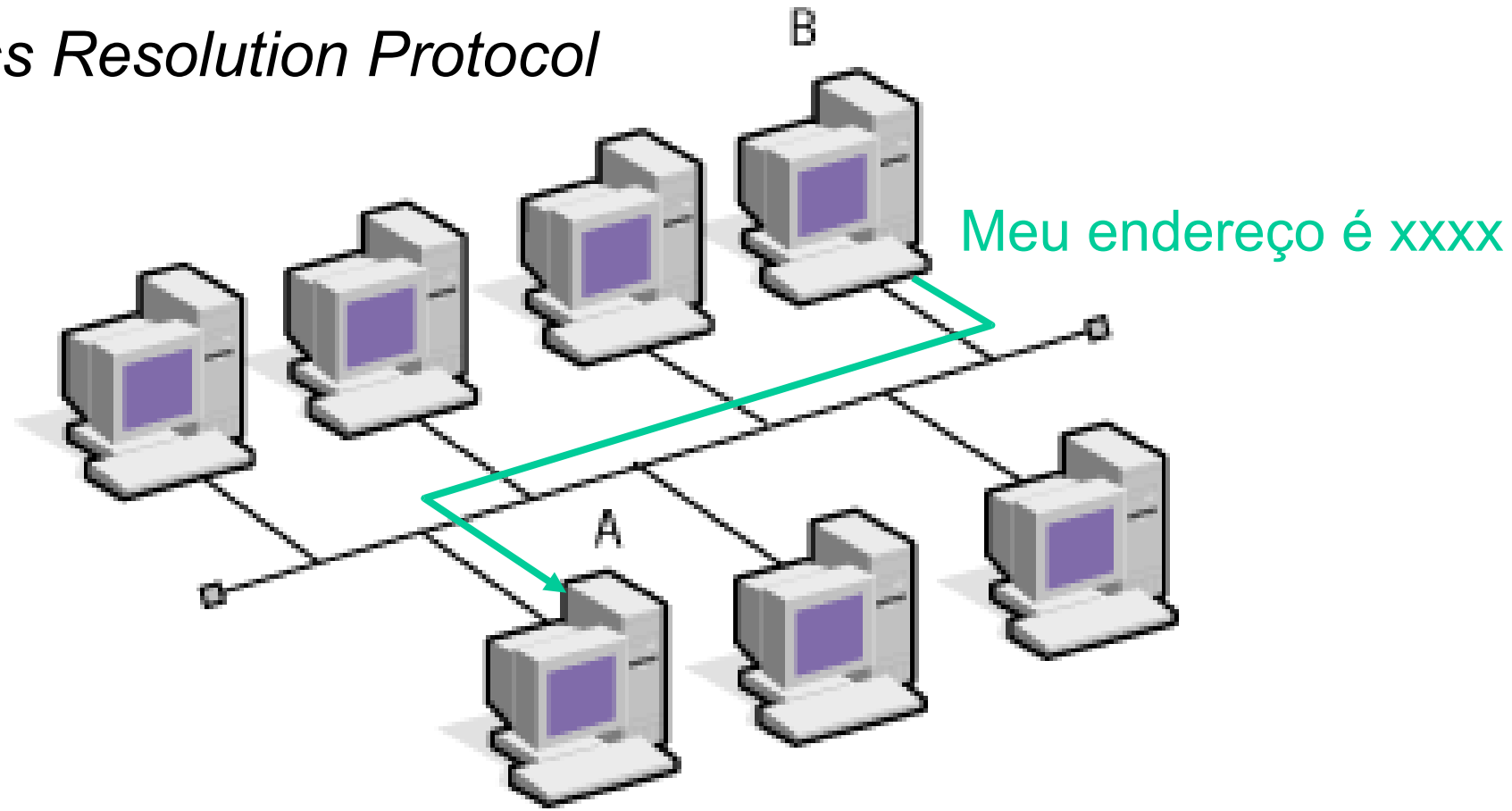
Address Resolution Protocol



Qual o endereço de B?

ARP

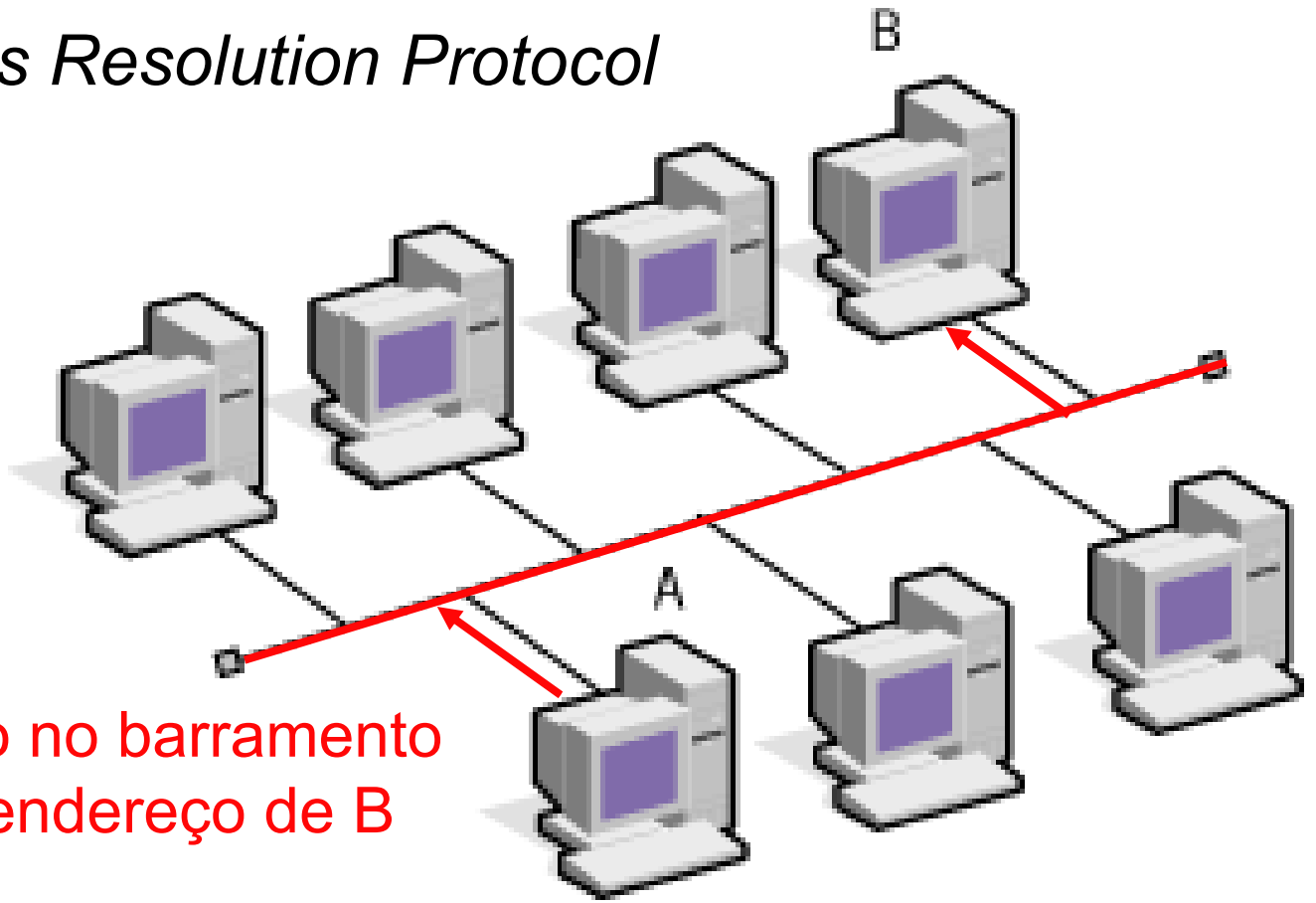
Address Resolution Protocol



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

ARP

Address Resolution Protocol



Difusão no barramento com endereço de B

Pode ser verificado pelo comando **`arp -n`**

Estamos saltando: “ATMARP”

Roteiro

- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

Protocolos auxiliares na operação da rede

- Gerência de configuração
  - Cada máquina deve receber informações básicas para operar
  - Isso pode ser feito manualmente, ou por um protocolo: DCHP
- Notificação de erros e controle
  - Problemas na operação da rede podem ser notificados
  - Protocolo de controle geral deve ser reconhecido: ICMP
- Transporte de pacotes sobre outras redes
  - Os pacotes de uma rede podem ter que passar por uma rede intermediária
  - Criam-se “túneis” onde pacotes entram e só aparecem em outro ponto
  - Princípio de redes virtuais (VPNs)
- Economia de endereços e ocultação de máquinas
  - Por diversos motivos, máquinas em uma rede podem ter end. inválidos
  - Tradução de endereços de rede (*Network Address Translation*, NAT)

Configuração de máquinas

Qual a informação mínima para uma máquina operar na rede?

- Quem cô sô?
  - Endereço e máscara locais
- Proncovô?
  - Caminho *default* de saída dos pacotes destinados a outras redes
- Cadê usôtro?
  - Processo de descobrimento de endereços de outras máquinas

DHCP

Dynamic Host Configuration Protocol:

- Permite a obtenção dinâmica de parâmetros de configuração para máquinas da rede (por exemplo, endereços IP)
  - Host envia mensagem "DHCP discover" (*broadcast*)
  - Servidor DHCP se identifica com mensagem "DHCP offer"
  - Host pede endereço IP com mensagem : "DHCP request"
  - Servidor DHCP envia endereço com a mensagem: "DHCP ack"
- Substitui o protocolo RARP (*Reverse Address Resolution Protocol*)

Relato de erros (ICMP)

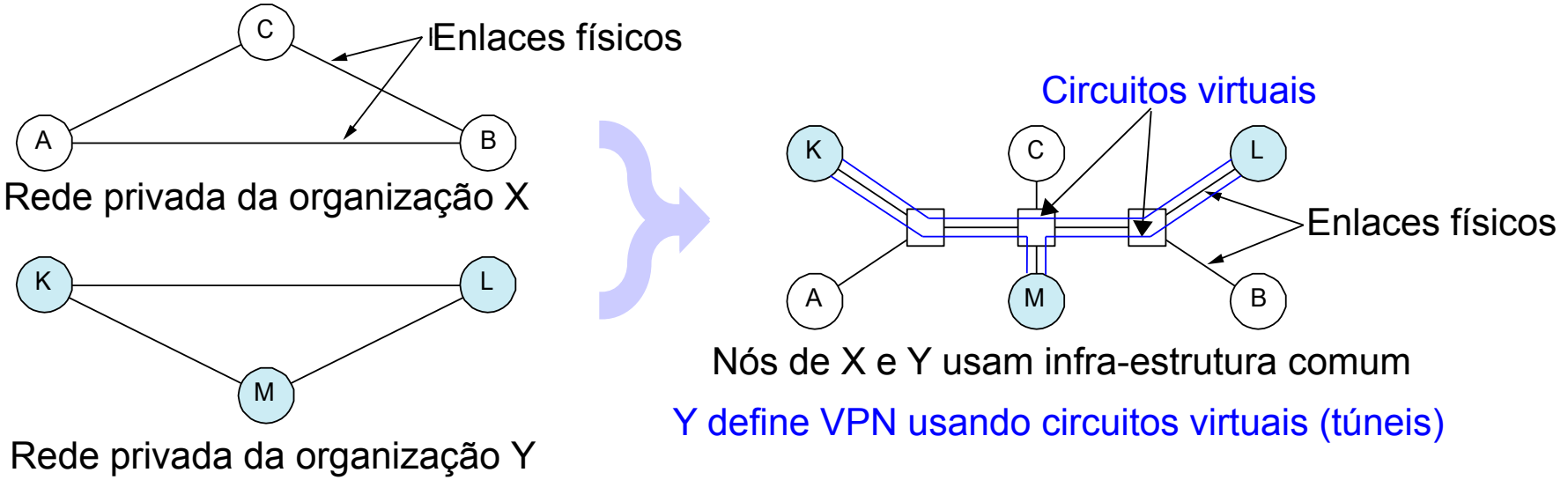
Internet Control Message Protocol

- Troca de mensagens entre elementos da rede IP para controle da transmissão e roteamento
  - Controle de fluxo (source quench)
  - Notificação de falhas
    - *TTL exceeded*
    - *Destination (port, protocol, host, network) unreachable*
    - *Checksum failed*
    - *Reassembly failed*
    - *Cannot fragment*
  - Redirecionamento de rotas
  - Requisição de informações (ping)

Programas: ping, traceroute, etc.

Redes virtuais (VPNs)

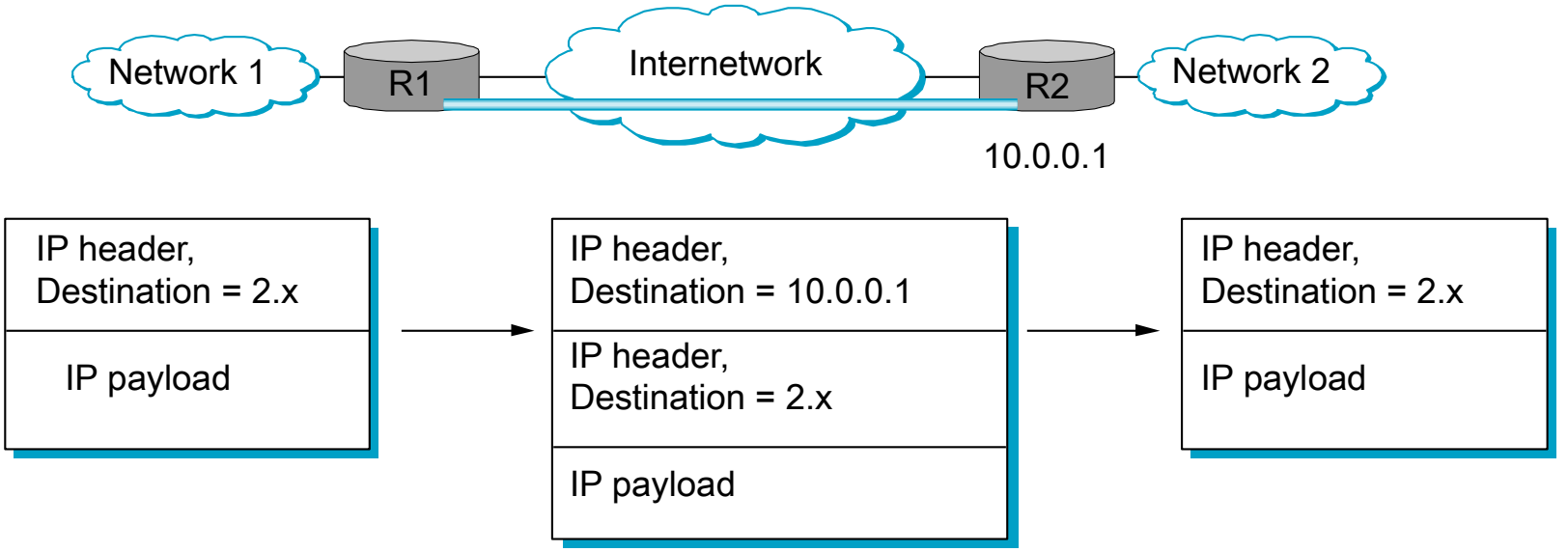
- Organizações podem ter políticas de segurança/acesso definidas em termos de suas redes “privativas”
- Na prática, partes de cada organização podem estar em pontos diferentes da Internet
- VPN: *Virtual Private Network*



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Redes virtuais e túneis (tunelamento)

- Pacote IP pode trafegar dentro de um outro pacote IP
  - Máquina origem (na rede 1) gera pacote como se estivesse na rede 2
  - Roteador empacota o pacote dentro de outro IP e envia para a rede 2
  - Na rede 2, pacote original é desempacotado e navega normalmente



  - Para todos os efeitos, máquina origem parece estar na rede 2
    - Controle de acesso, por exemplo

NAT: *Network Address Translation*

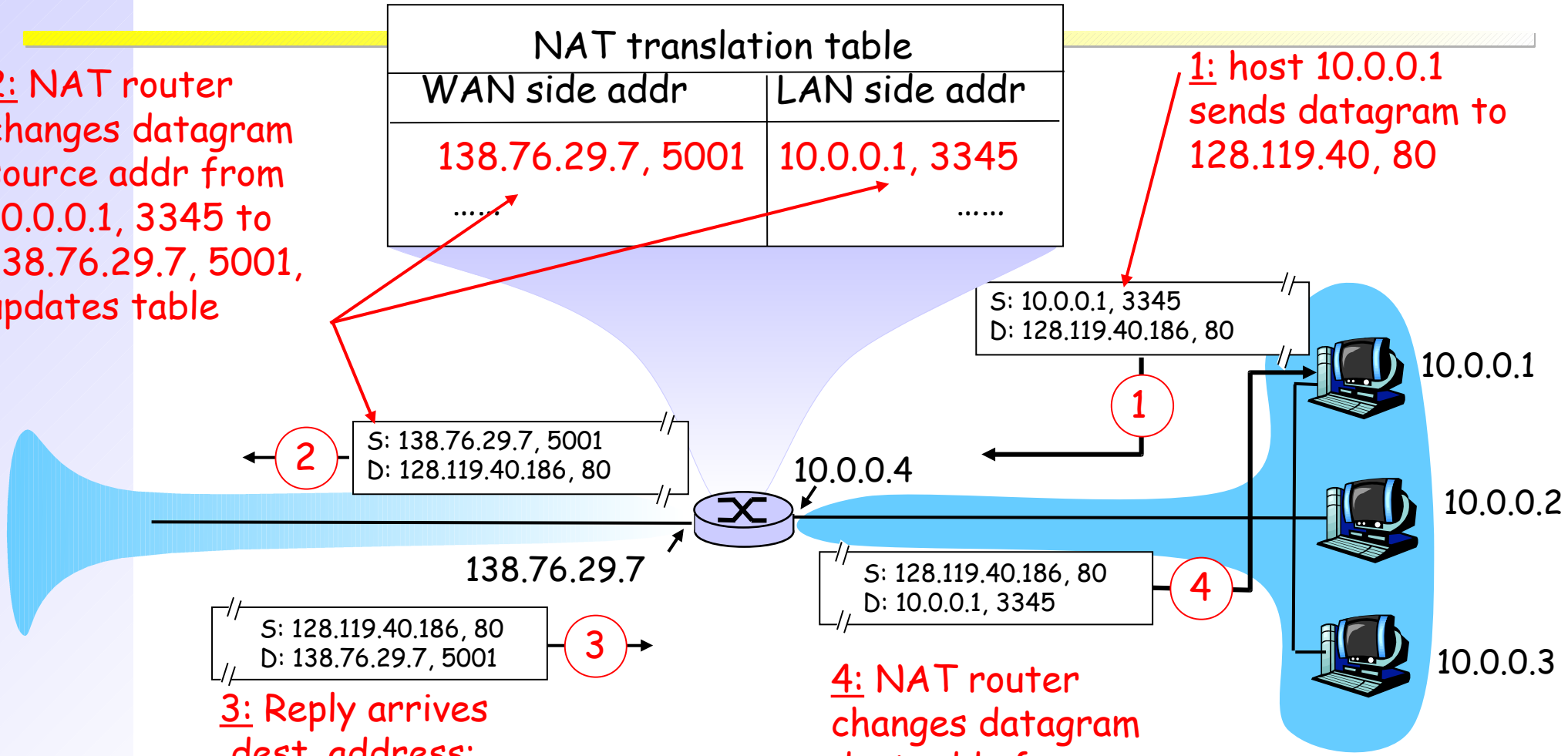
- Motivação: por vários motivos, uma rede usa apenas um endereço IP visível para o resto do mundo
- Na prática, rede interna usa endereços inválidos
- Efeitos:
  - Reduz a necessidade de obtenção de uma faixa de ends.
  - Facilita a configuração dos endereços na rede local
  - Permite a troca de provedor Internet sem reconfiguração
  - Esconde dispositivos da rede local de acessos externos
- Implementação: roteador/*gateway* NAT

NAT: *Network Address Translation*

- Princípio básico:
  - utilizar o campo de número de porto, contido no cabeçalho TCP (e UDP)
  - a princípio, apenas comunicações originadas por máquinas internas precisam passar pela fronteira da rede

- Um roteador/gateway NAT deve:
  - re-escrever pacotes que saem com o endereço válido
  - guardar a informação sobre o pacote original e o re-escrito
  - re-escrever pacotes que entram com o endereço inválido (interno)

NAT: *Network Address Translation*



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Fonte: Kurose e Ross

NAT: *Network Address Translation*

- Algumas vezes, é preciso tornar uma máquina interna permanentemente visível para o mundo externo
  - p.ex., um servidor Web geral
- NAT reverso:
  - Uma ou mais entradas na tabela de tradução NAT preenchidos a priori
  - Associam um endereço e porto externos a um par interno pré-definido
  - p.ex.:

NAT translation table	
WAN side addr	LAN side addr
138.76.29.2, 80	10.0.0.10, 8080
……	……

Conexões externas para o servidor Web serão entregues ao porto 8080 da máquina com endereço 10.0.0.10

NAT: *Network Address Translation*

- Campo de número de porto tem 16 bits:
  - Mais de 60.000 conexões simutâneas com um end. único!
- NAT é controverso:
  - roteadores deveriam processar apenas até a camada de rede (IP)
  - viola o princípio de projeto fim-a-fim
    - projetistas de aplicações devem considerar a inviabilidade de conectar certas máquinas (p.ex., P2P)
  - a falta de endereços deveria ser resolvida com o uso de IPv6

Roteiro

- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

Roteamento

- Expedição (*forwarding*) x roteamento
  - *Forwarding*: selecionar um porto de saída baseado no endereço de destino e na tabela de rotas

    SE ((end_destino & mascara) == (end_interface & mascara))

        envia datagrama diretamente ao host de destino

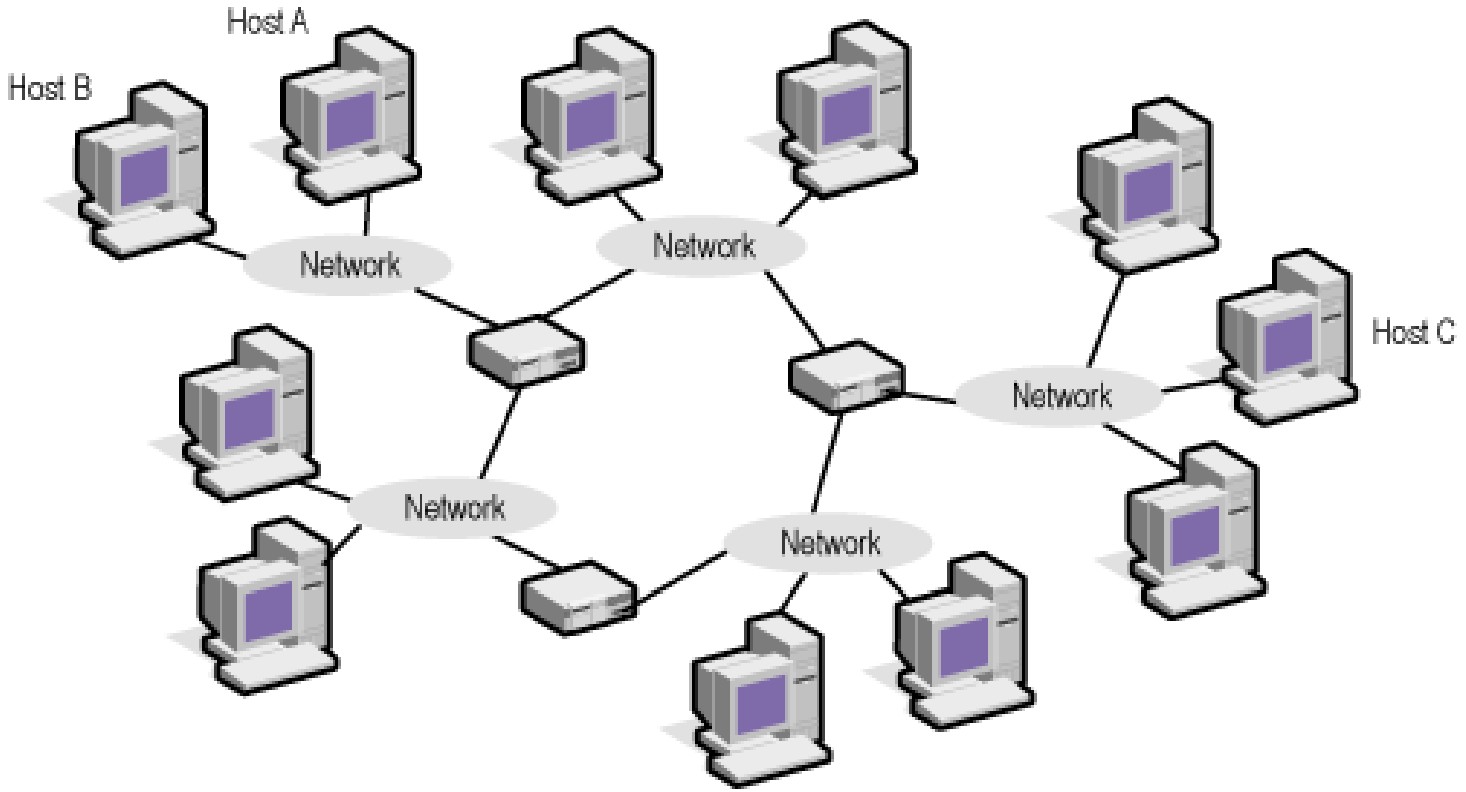
    SENÃO

        determina próximo passo (roteador) pela tabela de rotas

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Roteamento

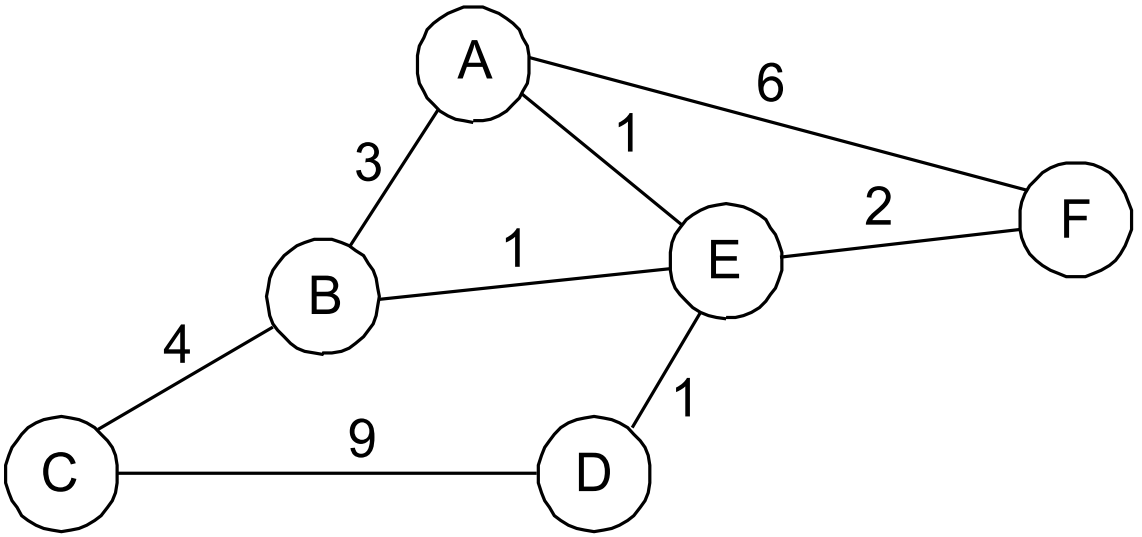
- Expedição (*forwarding*) x roteamento
  - Roteamento: processo de construção da tabela de rotas



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Roteamento

- Expedição (*forwarding*) x roteamento
  - Roteamento: processo de construção da tabela de rotas
  - Para isso, a rede deve ser vista como um grafo:



  - Problema: encontrar o caminho de menor custo entre nós do grafo
  - Fatores relevantes: estáticos (topologia) e dinâmicos (carga)

Determinação de rotas

- Responsabilidade de cada entidade ligada à rede: Sistema autônomo (*autonomous system*, AS)
  - corresponde a um domínio administrativo
  - tem controle absoluto sobre caminhos internos
  - exemplos: universidades, empresas, backbones
  - cada AS recebe um número de 16 bits

- Hiearquia de propagação de rotas em dois níveis
  - protocolo interior (*interior gateway protocol*), cada AS pode escolher o seu
  - protocolo exterior (*exterior gateway protocol*), padrão comum a toda a Internet

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

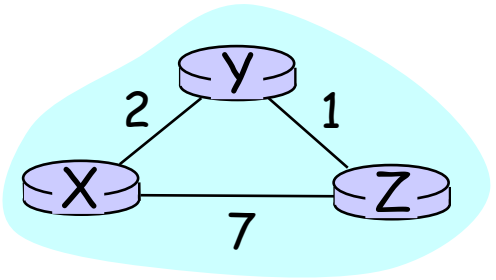
Protocolos interiores populares

- RIP: *Route Information Protocol*
  - desenvolvido para o XNS (rede da Xerox)
  - distribuído com o Unix
  - algoritmo de vetor de distâncias (*distance vector*)
  - baseado na contagem de roteadores (*hop-count*)
- OSPF: *Open Shortest Path First*
  - padrão Internet mais recente
  - algoritmo de estado dos links (*link-state*)
  - permite balanceamento de carga
  - suporta autenticação de roteadores

Distance vector (RIP)

- Cada nó mantém um conjunto de tuplas
  (destino, custo, próximo passo/*next hop*)
- Troca informações com vizinhos imediatos
  - periodicamente (vários segundos)
  - quando sua tabela muda (*triggered update*)
- Cada atualização é uma lista de pares: (destino, custo)
- Tabela local é atualizada se surge rota “melhor”
  - menor custo
  - atualização vinda do nó já escolhido como próximo
- Rotas existentes são refrescadas a cada atualização
- Se uma entrada temporiza sem ser refrescada é removida

Distance Vector (RIP): exemplo



D^X

dest	cost via Y	cost via Z
Y	2	∞
Z	∞	7

D^Y

dest	cost via X	cost via Z
X	2	∞
Z	∞	1

D^Z

dest	cost via X	cost via Y
X	7	∞
Y	∞	1

D^X

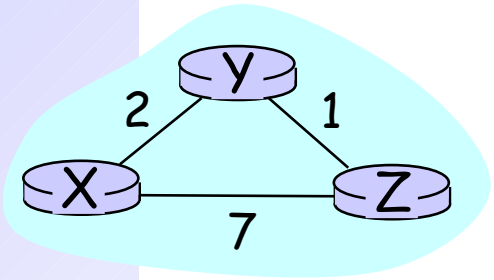
dest	cost via Y	cost via Z
Y	2	8
Z	3	7

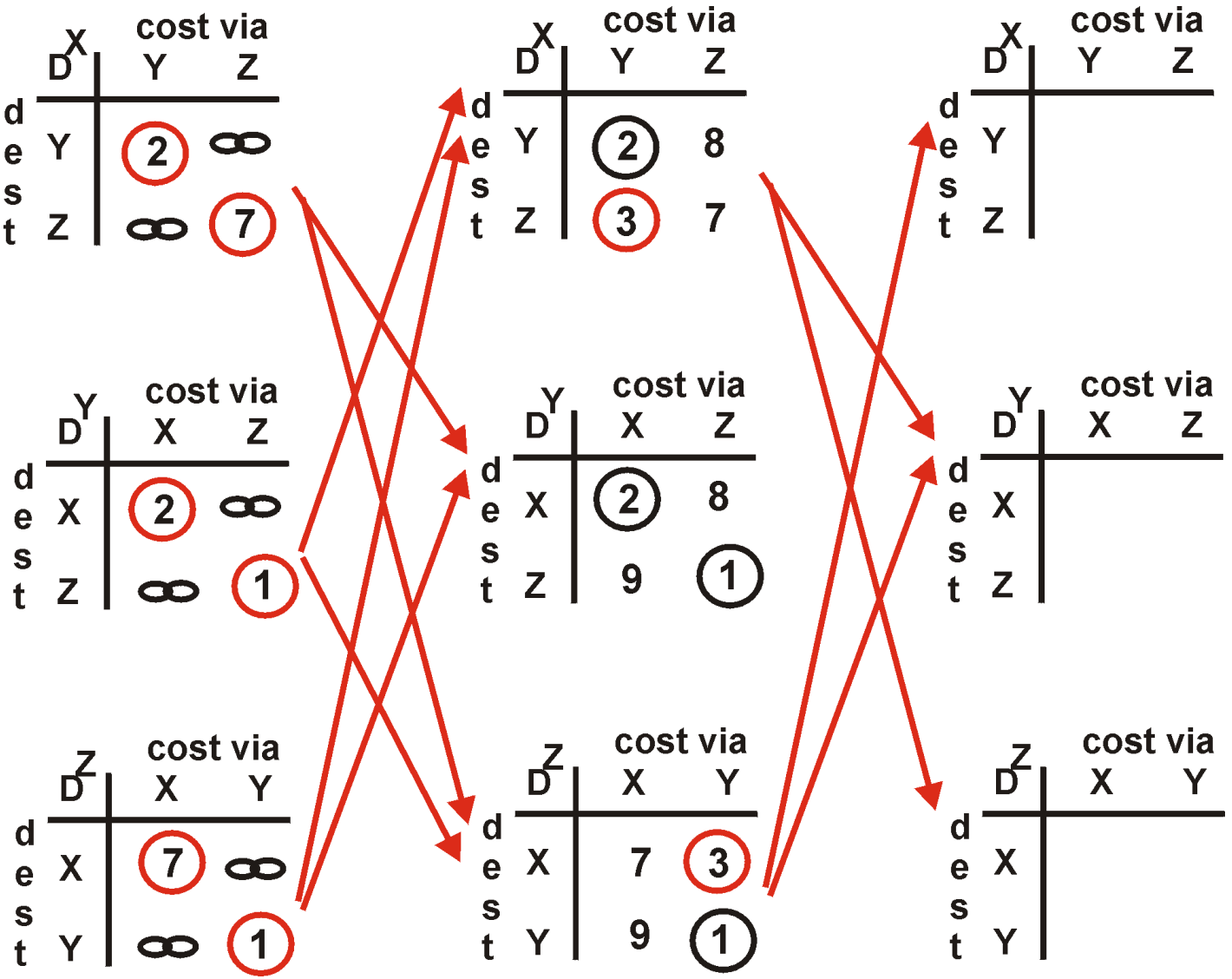
$$D^X(Y,Z) = c(X,Z) + \min_w \{D^Z(Y,w)\}$$
$$= 7+1 = 8$$

$$D^X(Z,Y) = c(X,Y) + \min_w \{D^Y(Z,w)\}$$
$$= 2+1 = 3$$

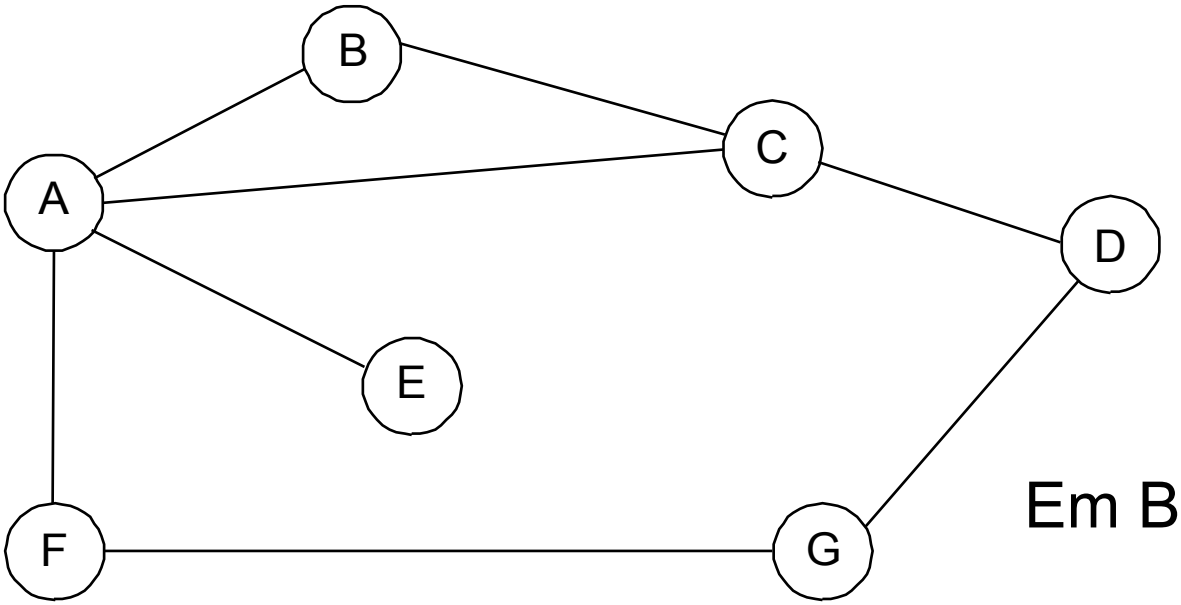
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Distance Vector (RIP): exemplo





Exemplo de aplicação: *distance vector*



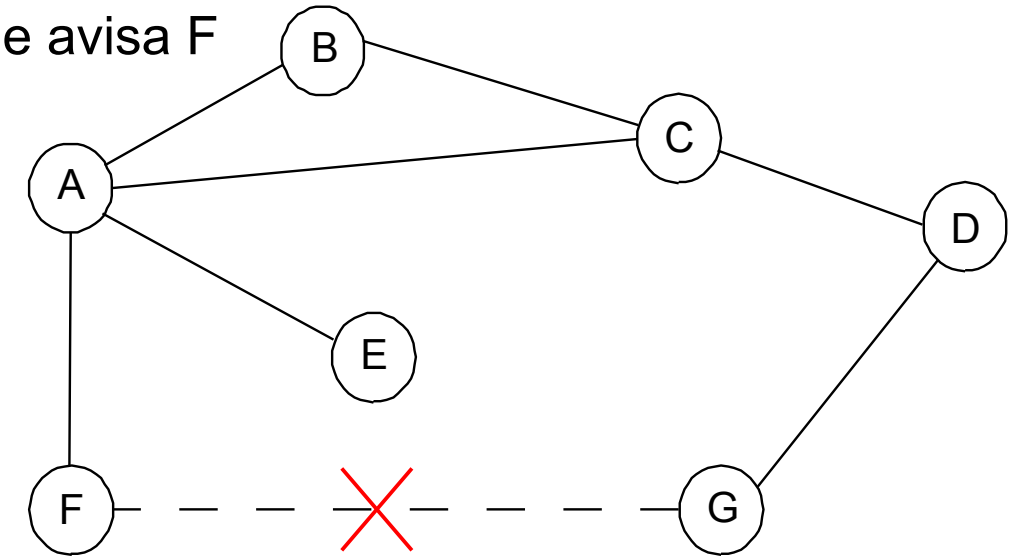
Em B:

Destino	Custo	Next Hop
A	1	A
C	1	C
D	2	C
E	2	A
F	2	A
G	3	A

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Operação de *distance vector* sob falhas

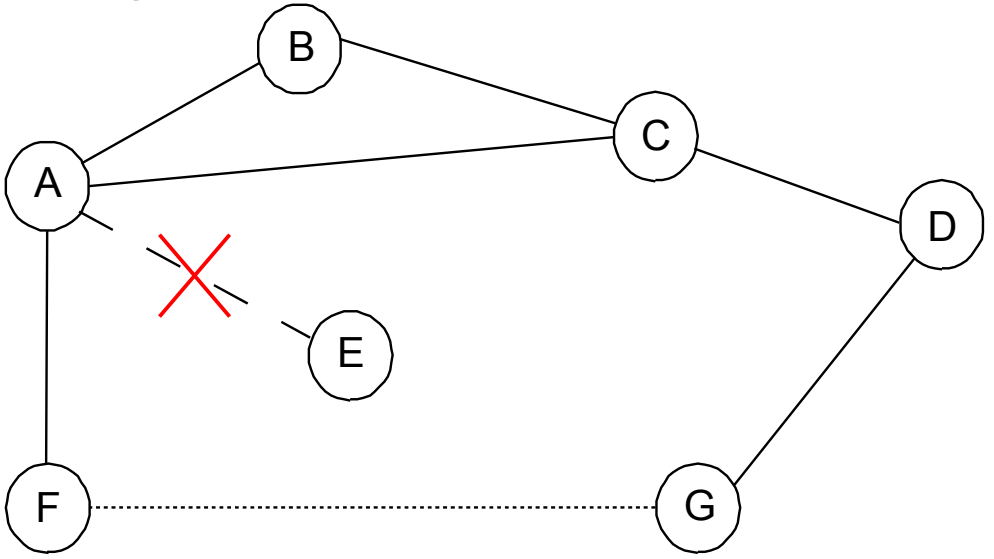
- Exemplo 1
  - F detecta que ligação com G falhou
  - F anota como infinita a distância para G e avisa A
  - A anota a distância para G como infinita pois F é atualmente o próximo passo para G
  - A recebe atualização periódica de C com caminho para G com dois passos
  - A anota distância 3 para G e avisa F



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Operação de *distance vector* sob falhas

- Exemplo 2
  - Canal de A para E falha
  - A anuncia distância infinita para E
  - B e C anunciam distância 2 para E
  - B decide que pode atingir E por C com distância 3 e avisa A
  - A decide que pode atingir E por B com 4 passos e avisa C
  - C decide que pode atingir E em 5 passos...
  - ***Loop* de roteamento!!!**



A seguir: algoritmo *link state*

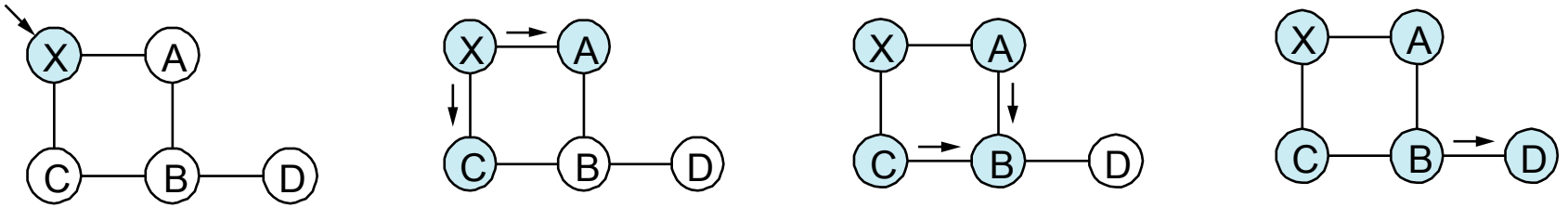
UFMG - ICEx DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Link state (OSPF)

- Envie:
  - para todos os nós (não apenas vizinhos)
  - apenas informações sobre ligações imediatas
- Pacote de estado do canal (*Link State Packet* - LSP)
  - Identificador do nó criador do pacote
  - Custo dos canais com vizinhos imediatos do mesmo
  - Número de seqüência (NSEQ)
  - Validade do pacote (*time-to-live* - TTL)

Link state (OSPF)

- "Envie para todos os vizinhos": alagamento confiável (*reliable flooding*)
  - inicie NSEQ em zero ao ligar
  - crie novo LSP (novo NSEQ) periodicamente
  - armazene LSP mais recente de cada nó
  - decremente TTL de cada pacote recebido
    - descarte LSP se TTL=0
  - redistribua LSP recebido em todos os canais, exceto o de recepção
    - envio aos vizinhos é confiável: confirmações e retransmissões



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Link state: cálculo das rotas

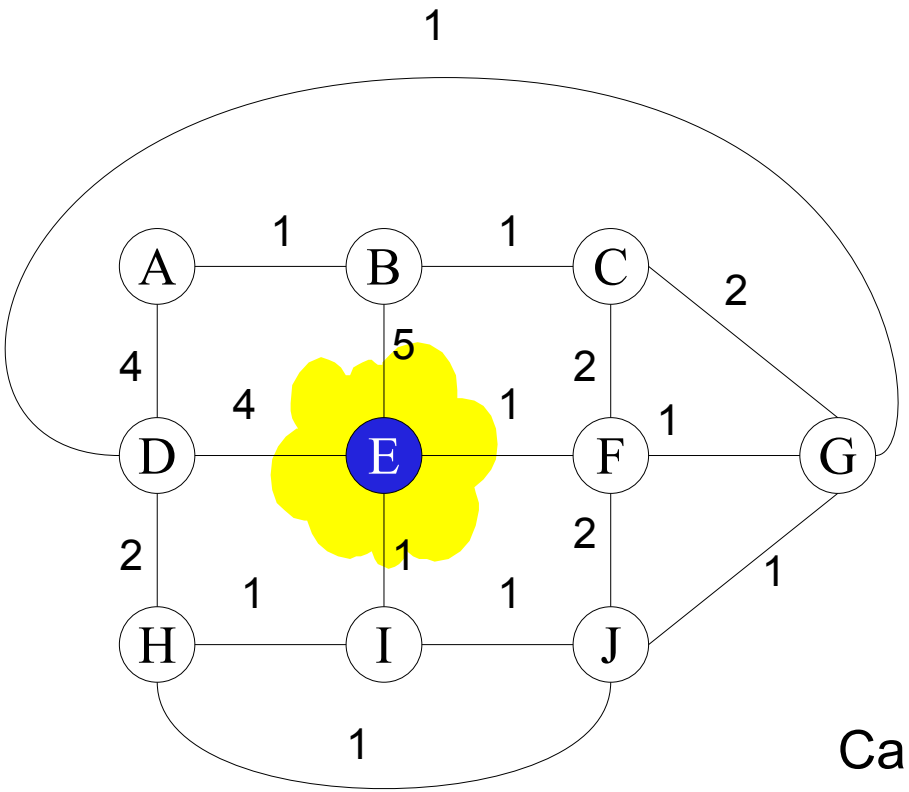
- Algoritmo de caminho mínimo de Dijkstra
- Nós (N) são separados em dois conjuntos:
  - Nós cujo caminho mínimo à origem já é conhecido (M)
  - Os demais, com distâncias não menores que os anteriores (N-M)
  - A cada passo, identifique os vizinhos de nós em M que não estão em M
  - Para esses nós, determine aquele de menor distância à raiz (fácil)
    - Acrescente esse nó a M e repita até que N=M

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Link state

Distâncias conhecidas:

E – 0



Candidatos:

B – 5

D – 4

F – 1

I – 1

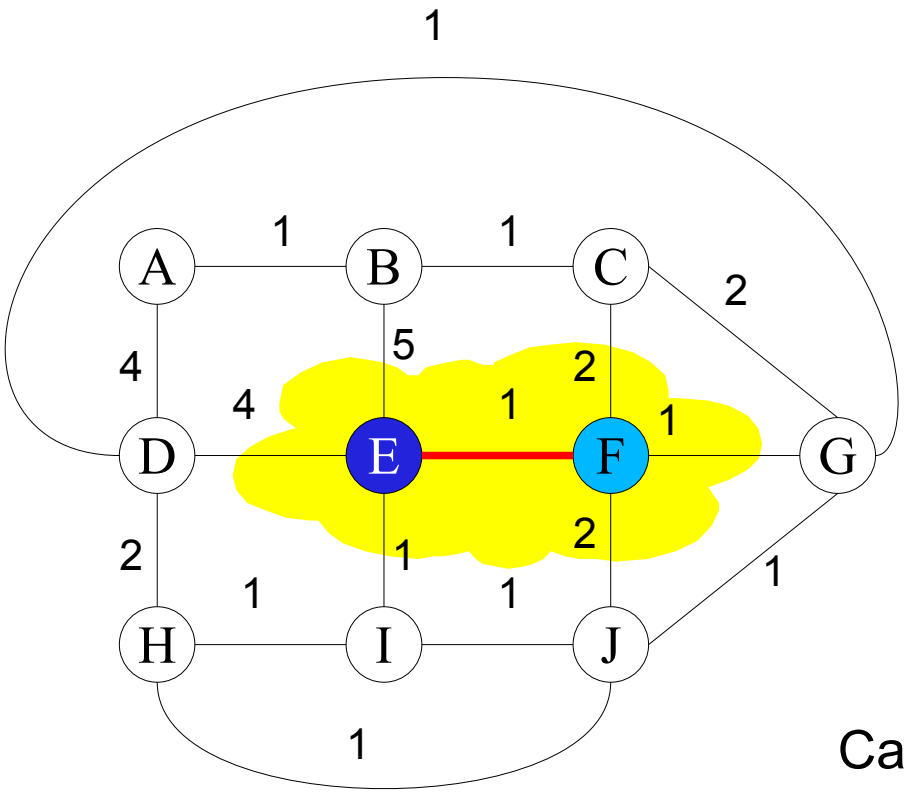
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Link state

Distâncias conhecidas:

E – 0

F – 1



Candidatos:

B – 5

D – 4

C – 3

I – 1

G – 2

J – 3

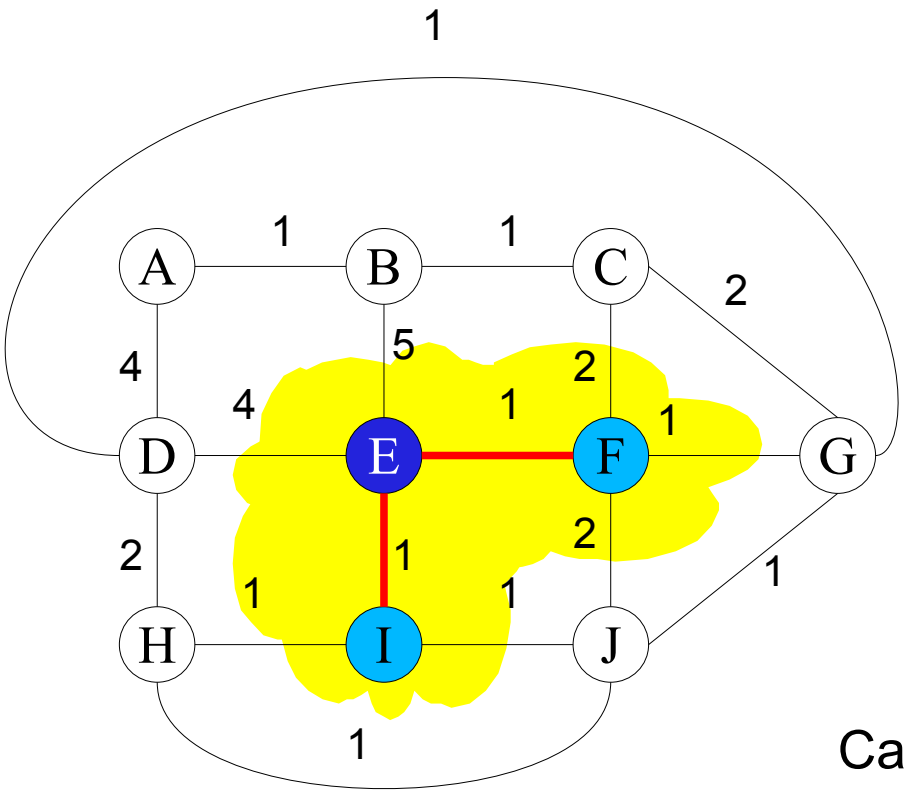
Link state

Distâncias conhecidas:

E – 0

F – 1

I – 1



Candidatos:

B – 5

D – 4

C – 3

G – 2

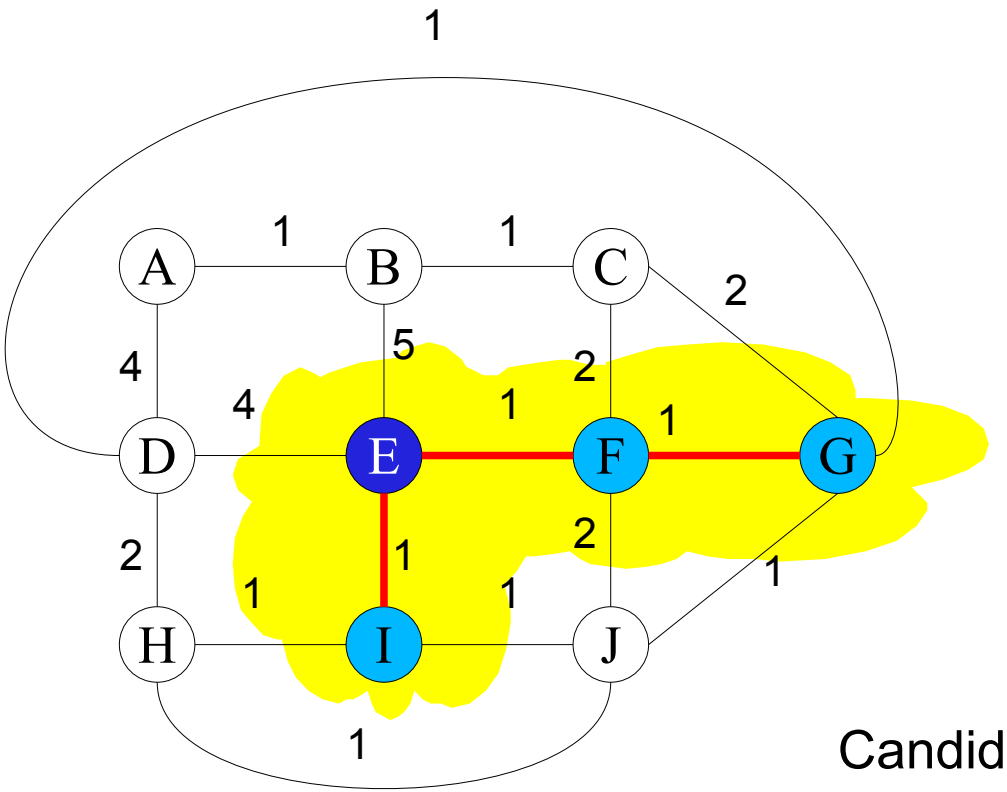
H – 2

J – 2

Link state

Distâncias conhecidas:

E – 0
F – 1
I – 1
G – 2



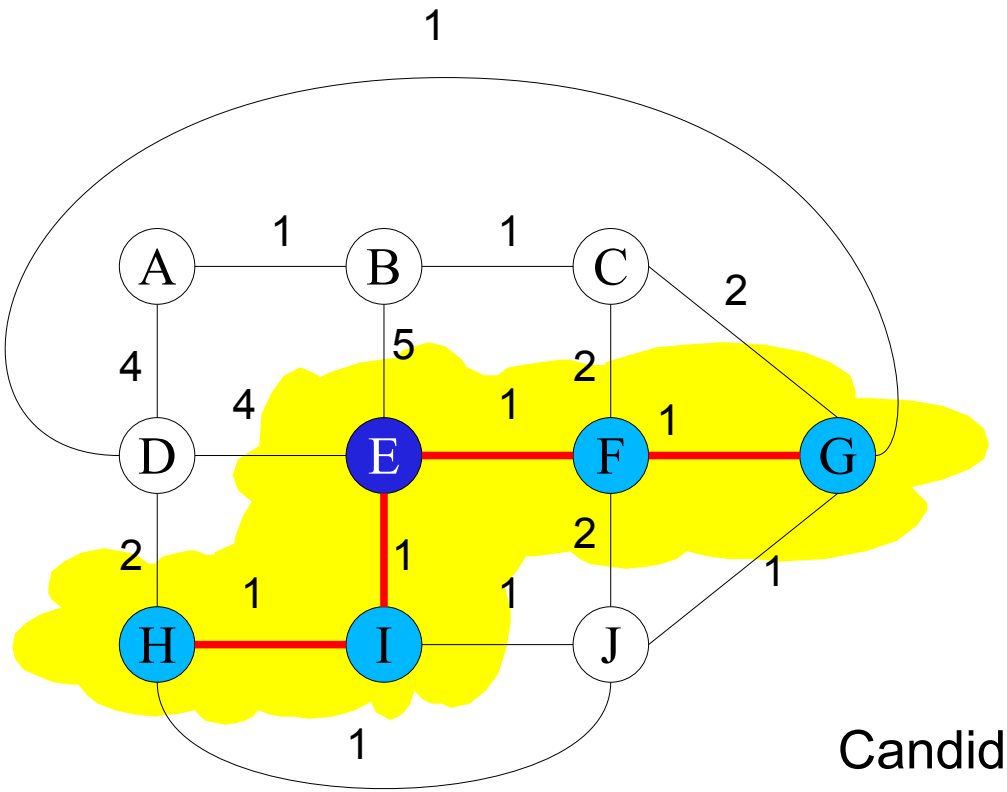
Candidatos:

B – 5
D – 3
C – 3
H – 2
J – 2

Link state

Distâncias conhecidas:

E – 0
F – 1
I – 1
G – 2
H – 2



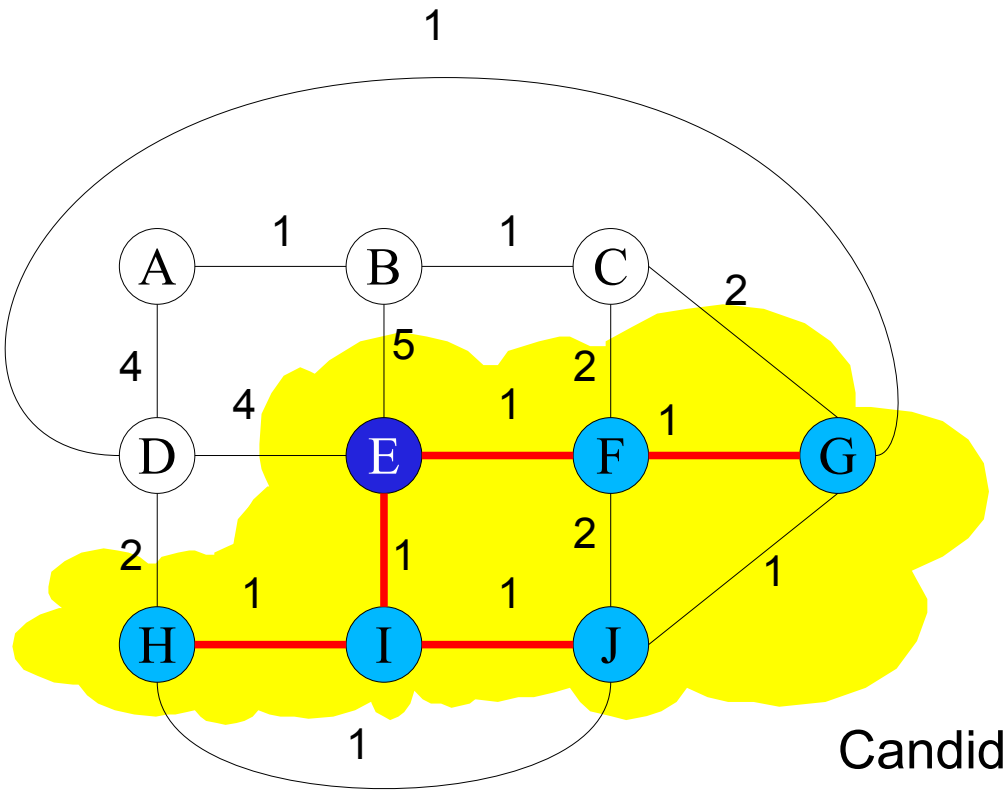
Candidatos:

B – 5
D – 3
C – 3
J – 2

Link state

Distâncias conhecidas:

- E – 0
- F – 1
- I – 1
- G – 2
- H – 2
- J – 2



Candidatos:

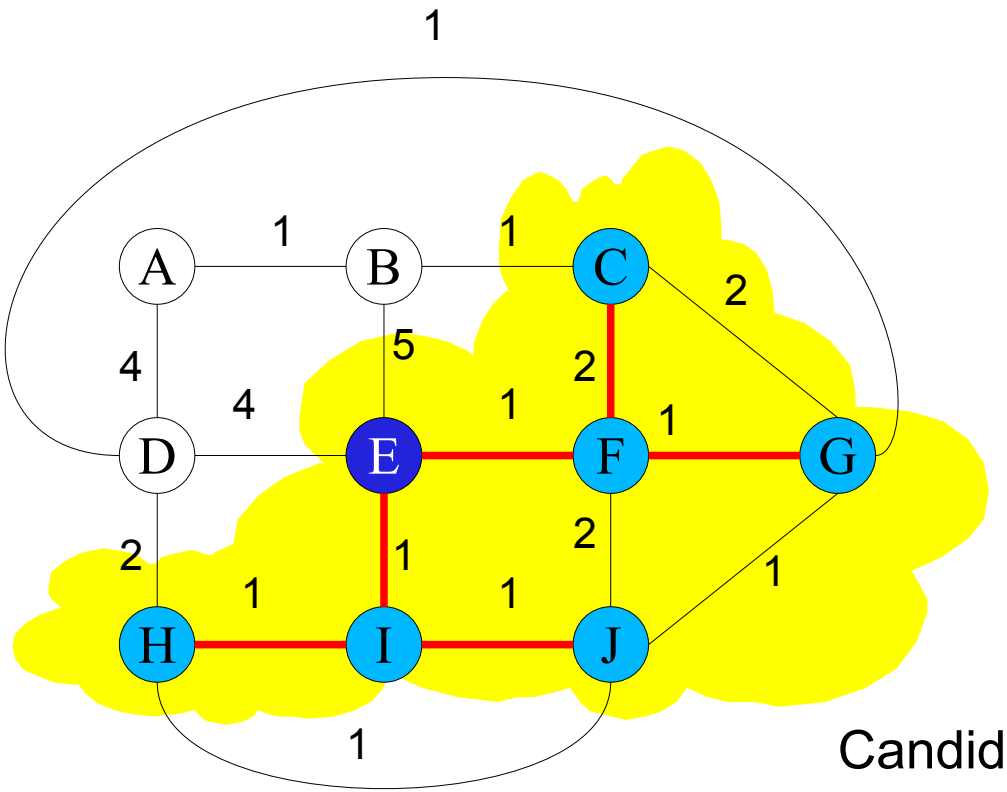
- B – 5
- D – 3
- C – 3

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Link state

Distâncias conhecidas:

- E – 0
- F – 1
- I – 1
- G – 2
- H – 2
- J – 2
- C – 3



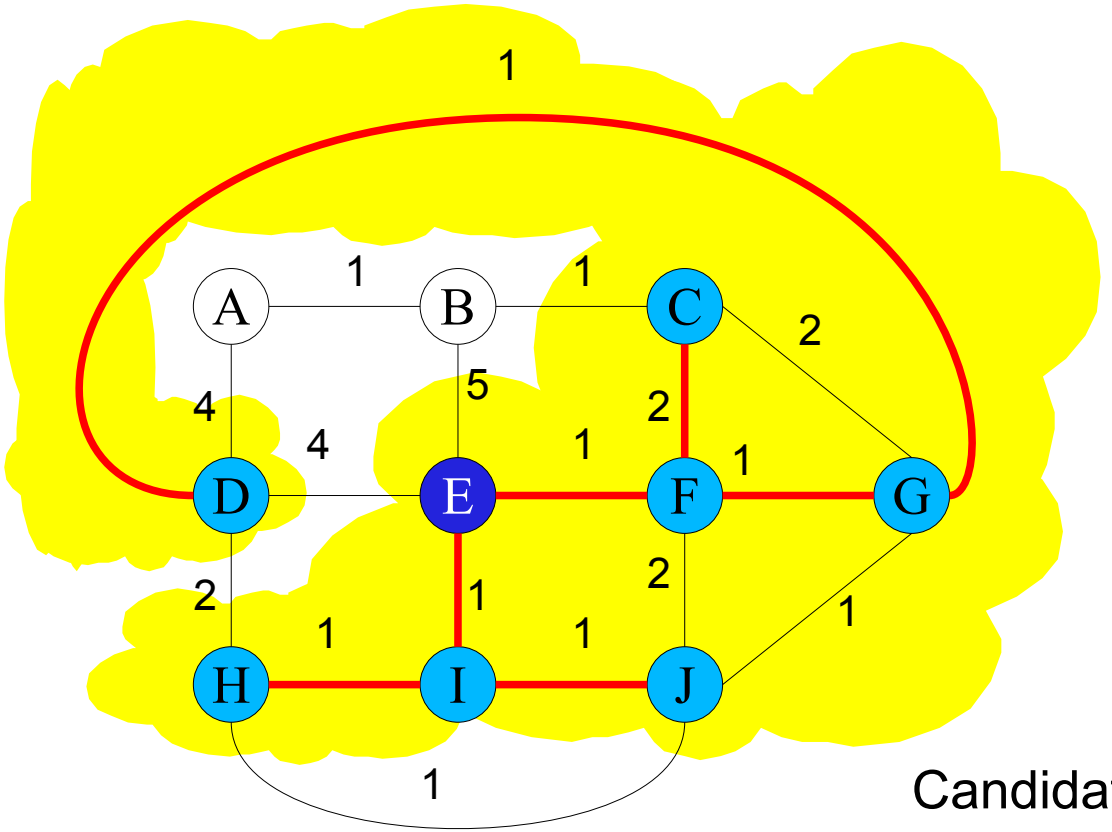
Candidatos:

- B – 4
- D – 3

Link state

Distâncias conhecidas:

- E – 0
- F – 1
- I – 1
- G – 2
- H – 2
- J – 2
- C – 3
- D – 3



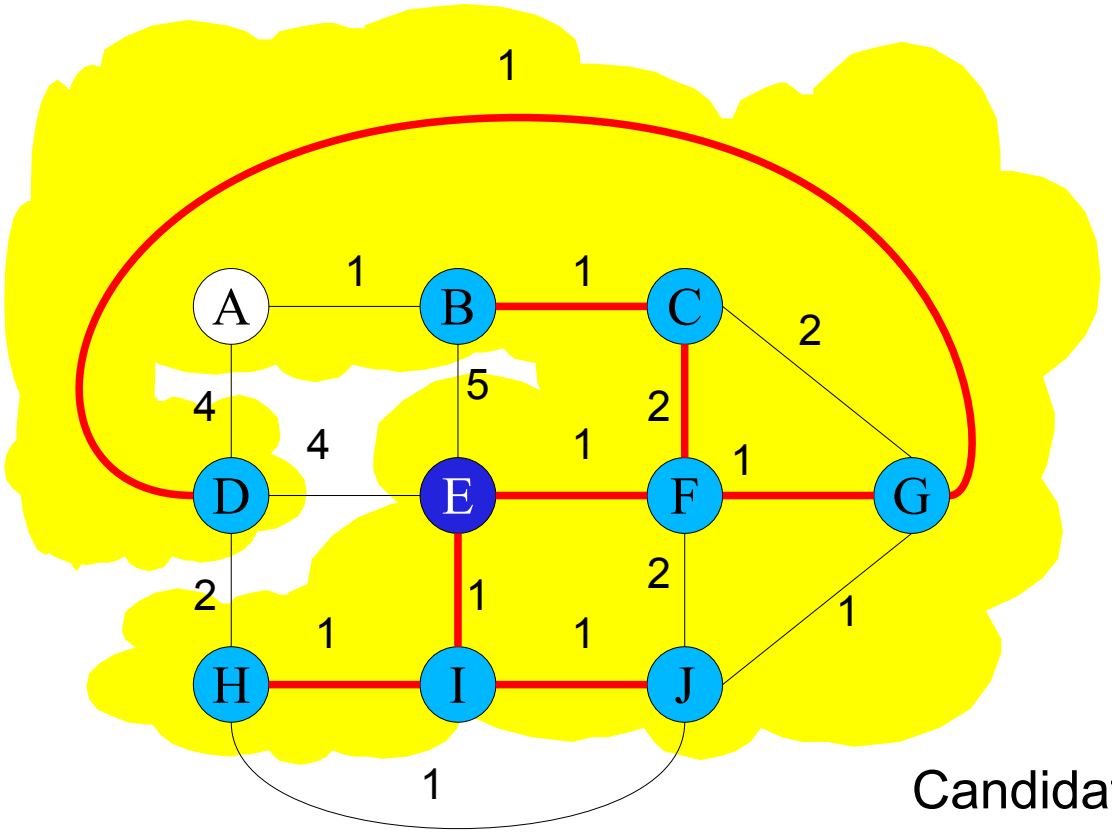
Candidatos:

- B – 4
- A – 7

Link state

Distâncias conhecidas:

E – 0
F – 1
I – 1
G – 2
H – 2
J – 2
C – 3
D – 3
B – 4

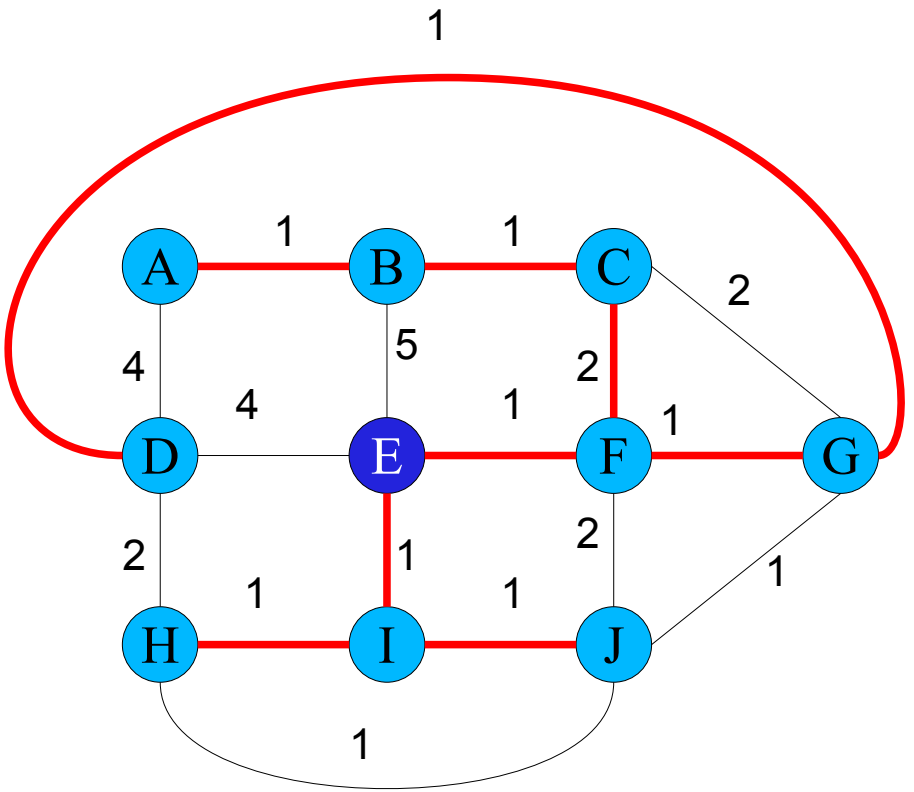


Candidatos:
A – 5

Link state

Distâncias conhecidas:

E – 0
F – 1
I – 1
G – 2 (F)
H – 2 (I)
J – 2 (I)
C – 3 (F)
D – 3 (F)
B – 4 (F)
A – 5 (F)



Métricas usadas para o roteamento

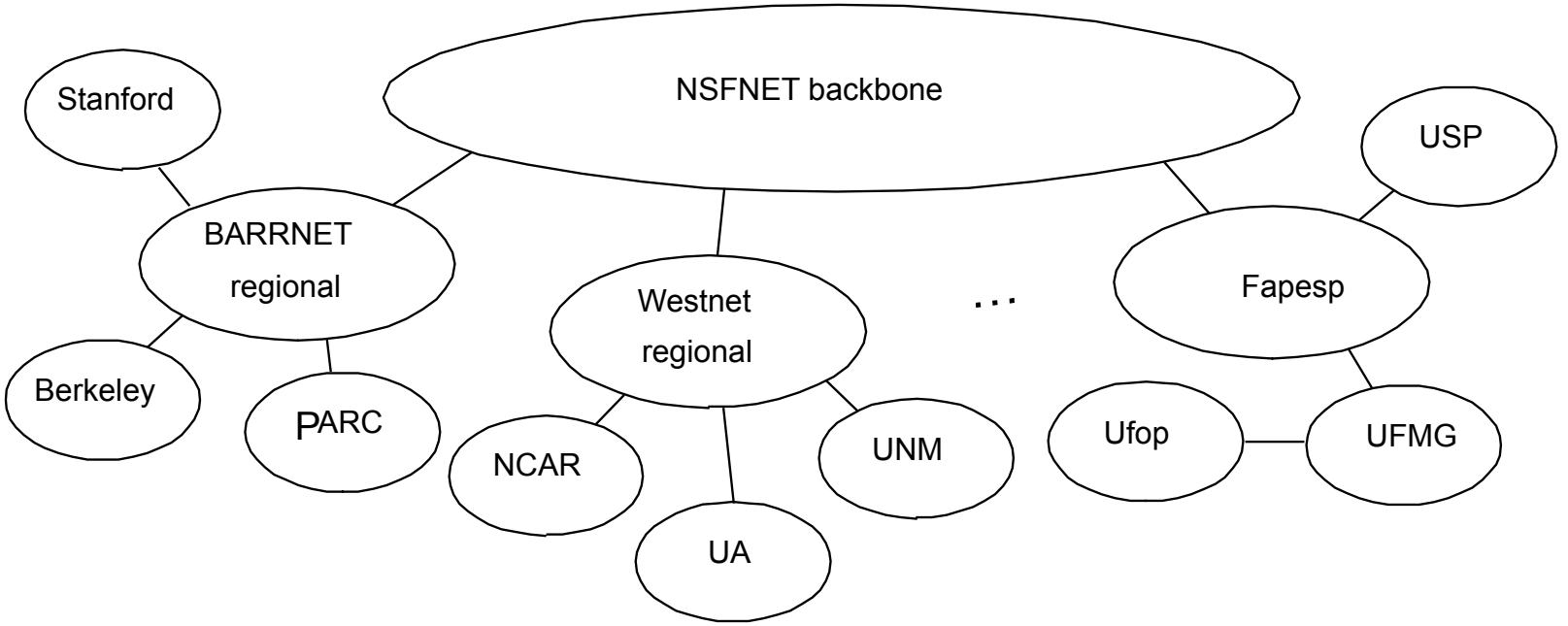
- Protocolos de roteamento simples (RIP)
  - contagem de links/roteadores no caminho
- Métrica original da ARPANET
  - medição do número de pacotes enfileirados em cada link
  - não considerava latência nem banda
- Nova métrica da ARPANET
  - atraso = tempo na fila + tempo de transmissão + latência
  - custo do link = atraso médio por algum período de tempo
  - sintonia fina: faixa de valores limitada, inclui utilização do link
- Em suma: problema complicado, sem resposta trivial

Roteiro

- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

Estrutura da Internet

Em um passado recente (1990):

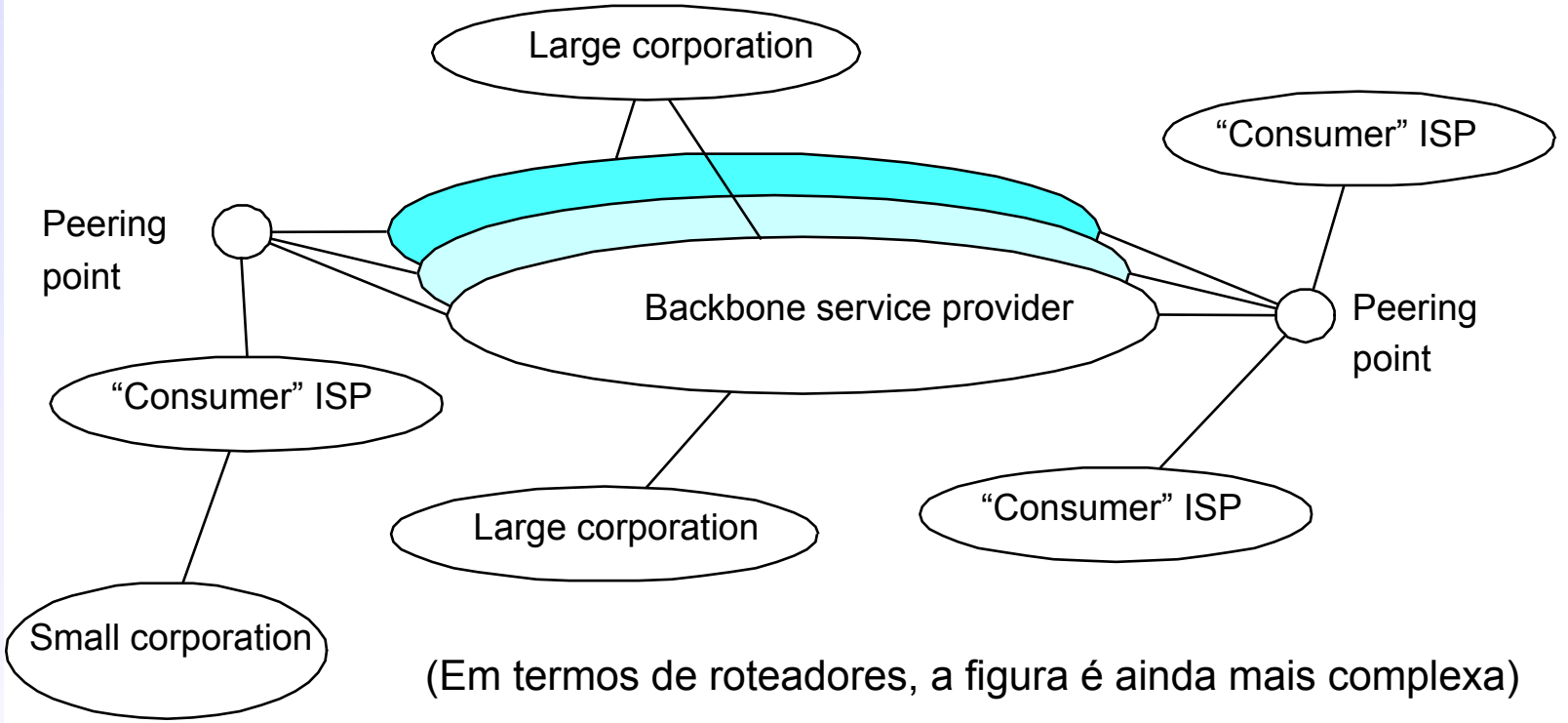


Muita coisa mudou desde então...

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Estrutura da Internet

Hoje:

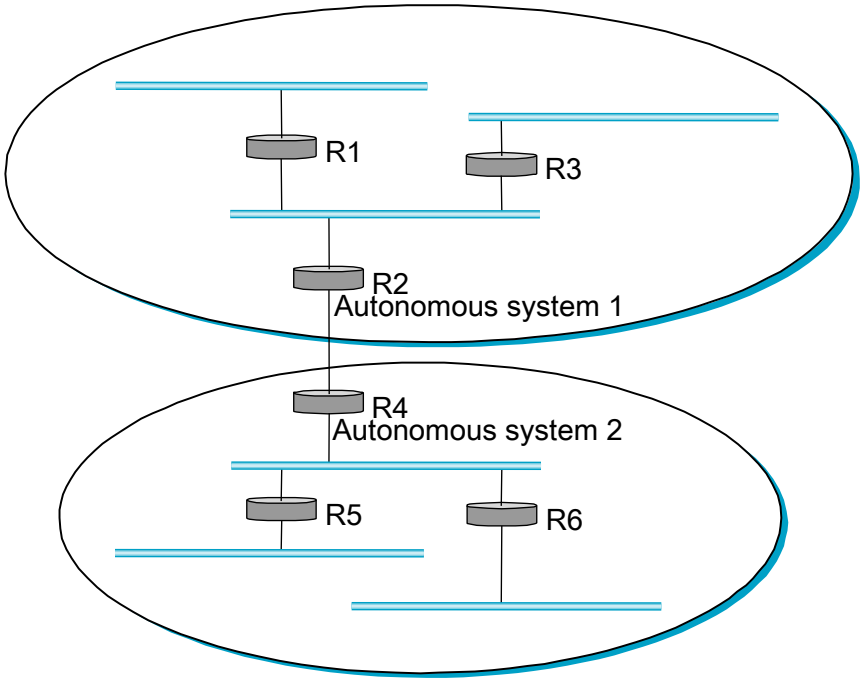


(Em termos de roteadores, a figura é ainda mais complexa)

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

BGP-4: *Border Gateway Protocol*

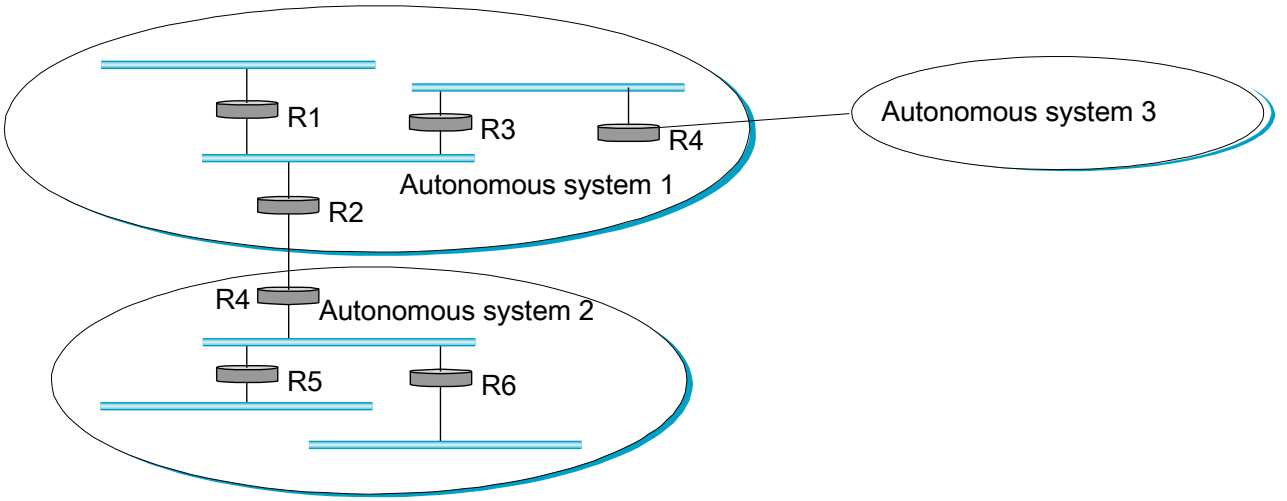
- Substituiu o EGP: *Exterior Gateway Protocol*
  - só tratava estruturas em árvore (Internet antiga)
- Protocolo para controle de alcance (*reachability*) entre AS
  - focado em determinar pelo menos um caminho (pode não ser ótimo)



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

BGP-4: *Border Gateway Protocol*

- Cada AS tem:
  - um ou mais roteadores de borda
  - um porta-voz BGP (*BGP speaker*) que anuncia:
    - redes locais internas ao AS
    - outras redes alcançáveis (no caso de *transit AS*)
    - informações sobre caminhos conhecidos
  - porta-voz pode também revogar caminhos previamente anunciados



BGP-4: *Border Gateway Protocol*

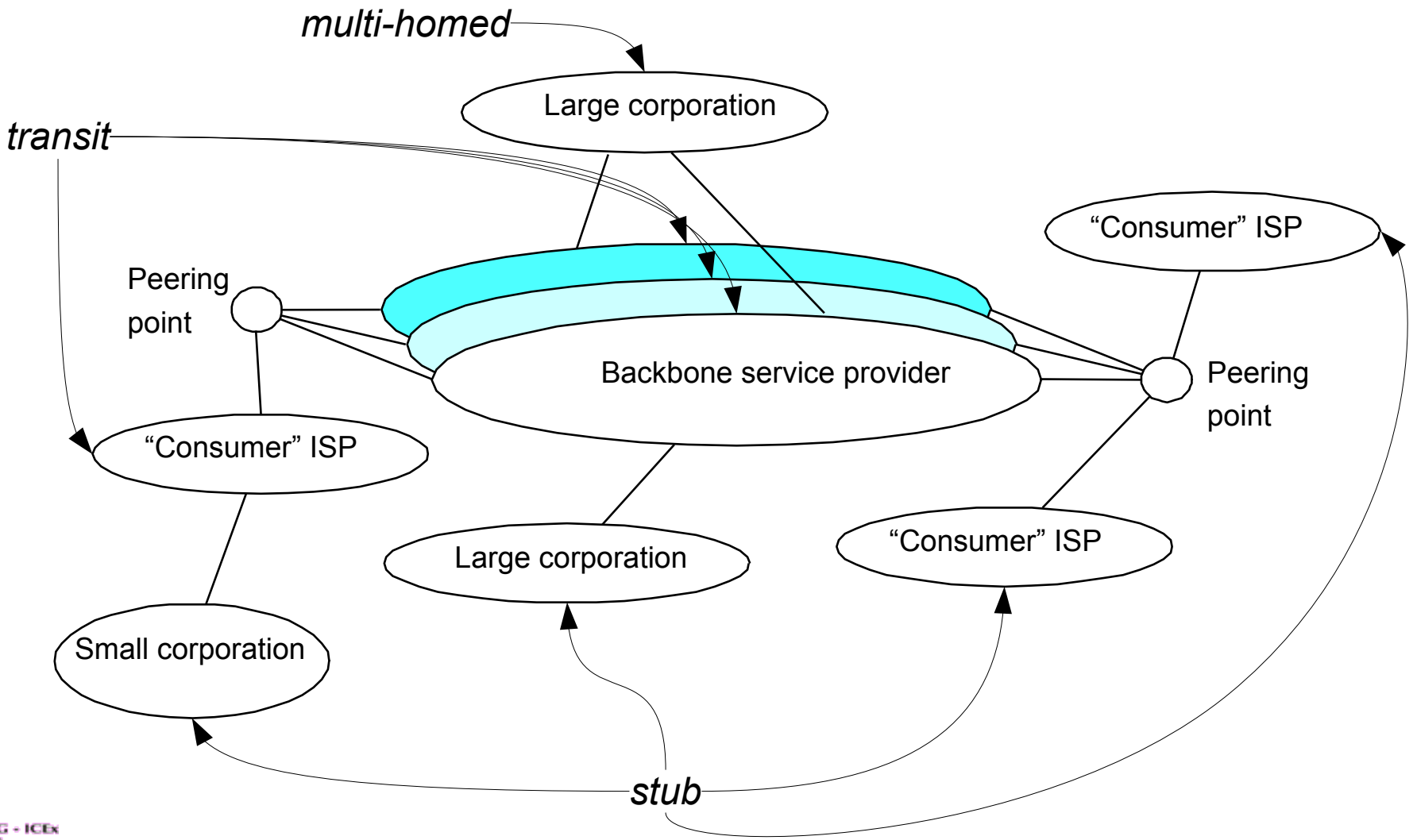
- Mensagens usadas pelos *speakers*
  - aquisição de vizinhos: roteadores se reconhecem e trocam informações
  - alcance de vizinhos: roteador testa se pares são alcançáveis (HELLO/ACK, k-em-n respostas)
  - atualizações de rotas: pares periodicamente trocam tabelas (a la *distance-vector*)
  - informação sobre o caminho também é trocada (evita *loops*)

BGP-4: *Border Gateway Protocol*

- Diferencia vários tipos de AS
  - *stub AS*: uma só conexão para outro AS
    - conduz apenas tráfego local
  - *multihomed AS*: tem conexões com mais de um AS
    - pode ter várias rotas para si mas se recusa a aceitar tráfego em trânsito
  - *transit AS*: também tem conexões com vários AS
    - conduz tanto tráfego local quanto em trânsito

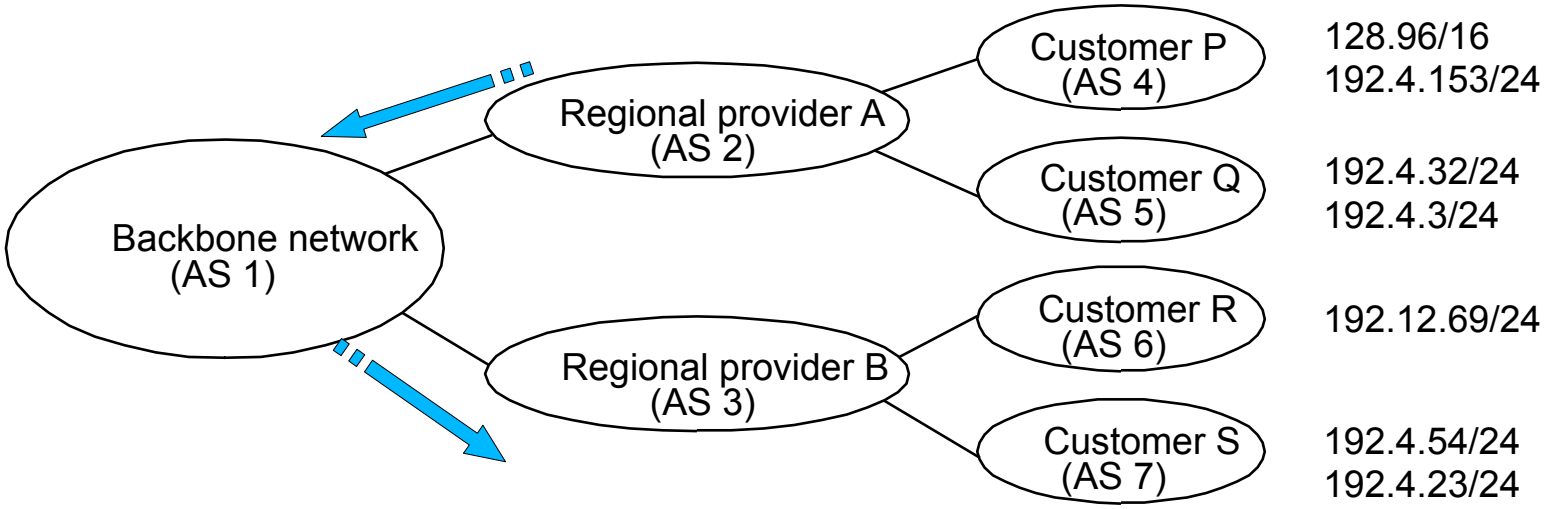
UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

BGP-4: *Border Gateway Protocol*



Exemplo com BGP

- *Speaker* do AS2 anuncia p/ AS1 o alcance a P e Q
  - redes 128.96/16, 192.4.153/24, 192.4.32/24 e 192.4.3/24 podem ser alcançadas diretamente a partir de AS2



- *Speaker* do *backbone* anuncia os caminhos
  - redes 128.96/16, 192.4.153/24, 192.4.32/24 e 192.4.3/24 podem ser alcançadas pelo caminho (AS1, AS2).

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Aproveitamento do espaço de endereços

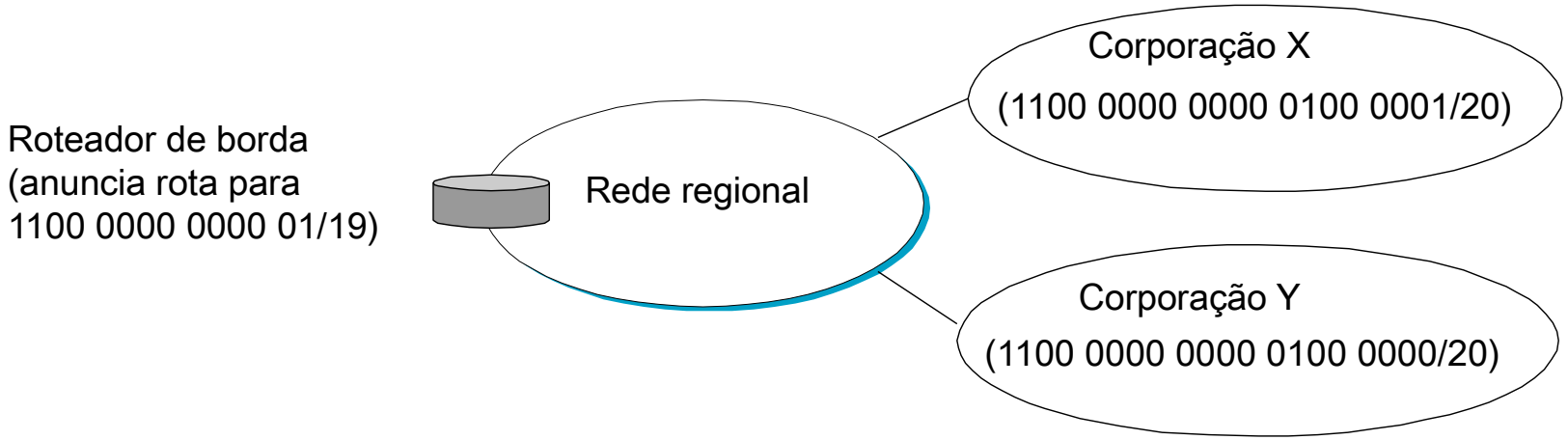
- Originalmente, havia apenas o esquema de classes A, B e C
- Com o tempo, as classes foram abolidas
  - As redes da classe A se mostraram grandes demais
  - Mesmo as classe B costumam ser sub-utilizadas
  - Em muitos casos, as classes C eram muito pequenas
- Nova solução: CIDR (*classless inter-domain routing*)
  - Além de flexibilizarem a alocação final, permitem que roteadores agrupem endereços na hora dos anúncios

Super-redes (CIDR)

- *Classless Inter-Domain Routing*
  - Redes topologicamente próximas ganham faixas de endereços contíguas
  - Roteadores distantes usam uma só entrada com o prefixo comum
  - Representa blocos de endereços com um par
    (prefixo_do_endereço_de_rede/número_de_bits)
  - Máscara de n bits (*CIDR mask*) identifica bloco (potência de 2)

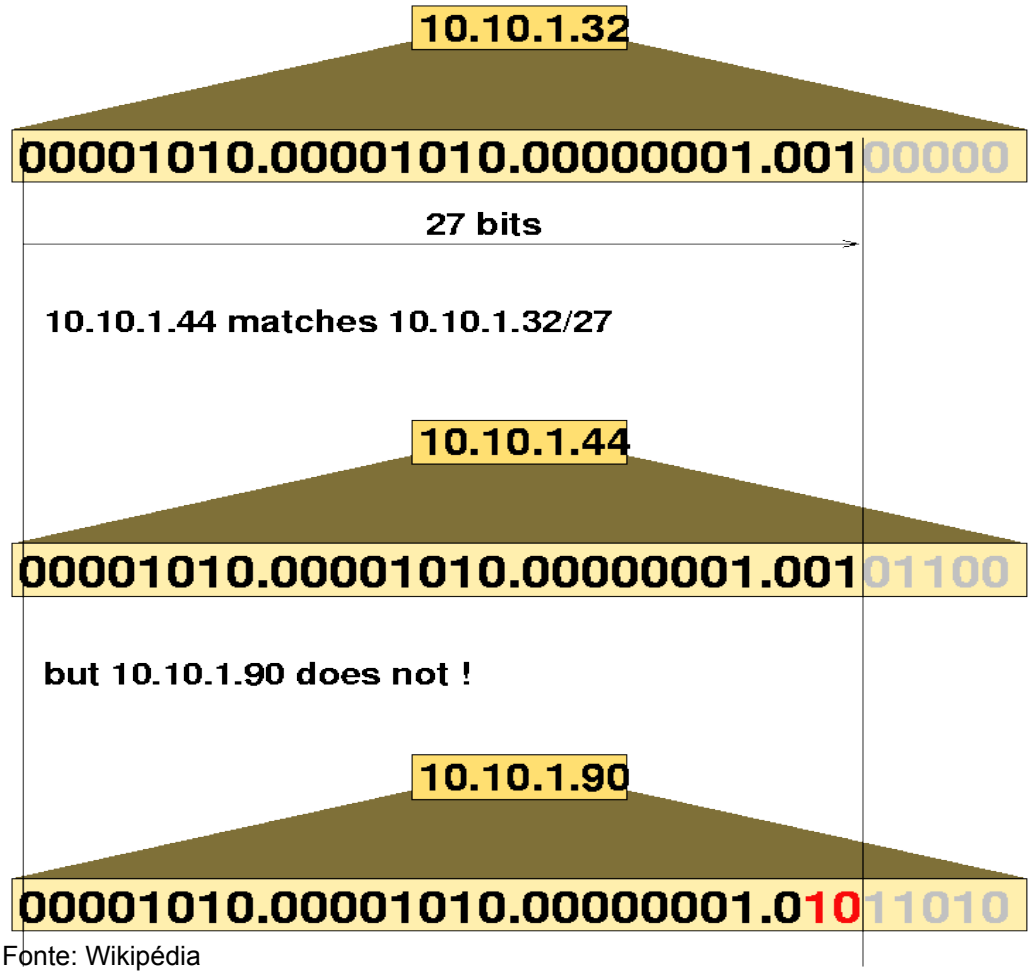
Super-redes (CIDR)

- Todos os roteadores de backbone devem entender CIDR
  - Tratamento obrigatório da máscara
  - Anúncio de rotas inclui a máscara associada



Super-redes (CIDR)

- Tratamento da máscara: suponha rota para 10.10.1.32/27:



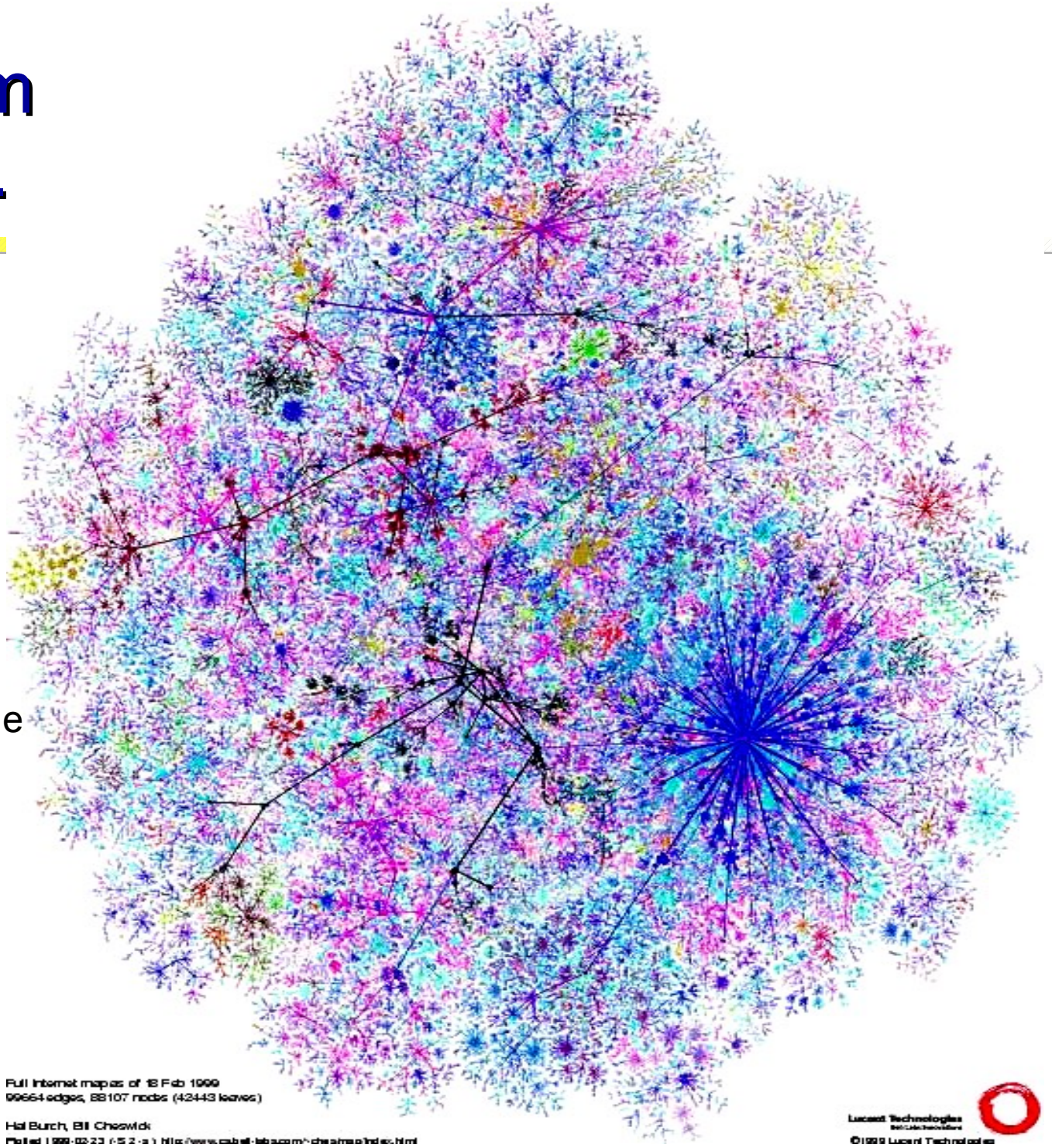
Fonte: Wikipédia

Integração entre BGP e IGPs

- Conheça um roteador mais “esperto”
- Roteadores de borda do AS injetam rotas externas nos IGPs
  - *Stubs*: basta divulgar rota *default*
  - *Multi-homed*: injetam rotas segundo políticas locais
  - *Backbones (transit)*: não há como reduzir a informação
    - volume elevado de rotas exige protocolo especial: IBGP (*Interior BGP)*
    - rotas entre roteadores internos definida com IGP convencional
    - IBGP anuncia para todos os roteadores rotas de cada roteador de borda
- CIDR pode ser usado para agregar informação
  - Ainda assim há redes demais
    - tabelas de roteamento não escalam
    - protocolos para propagação de rotas não escalam

Mesmo com tudo isso...

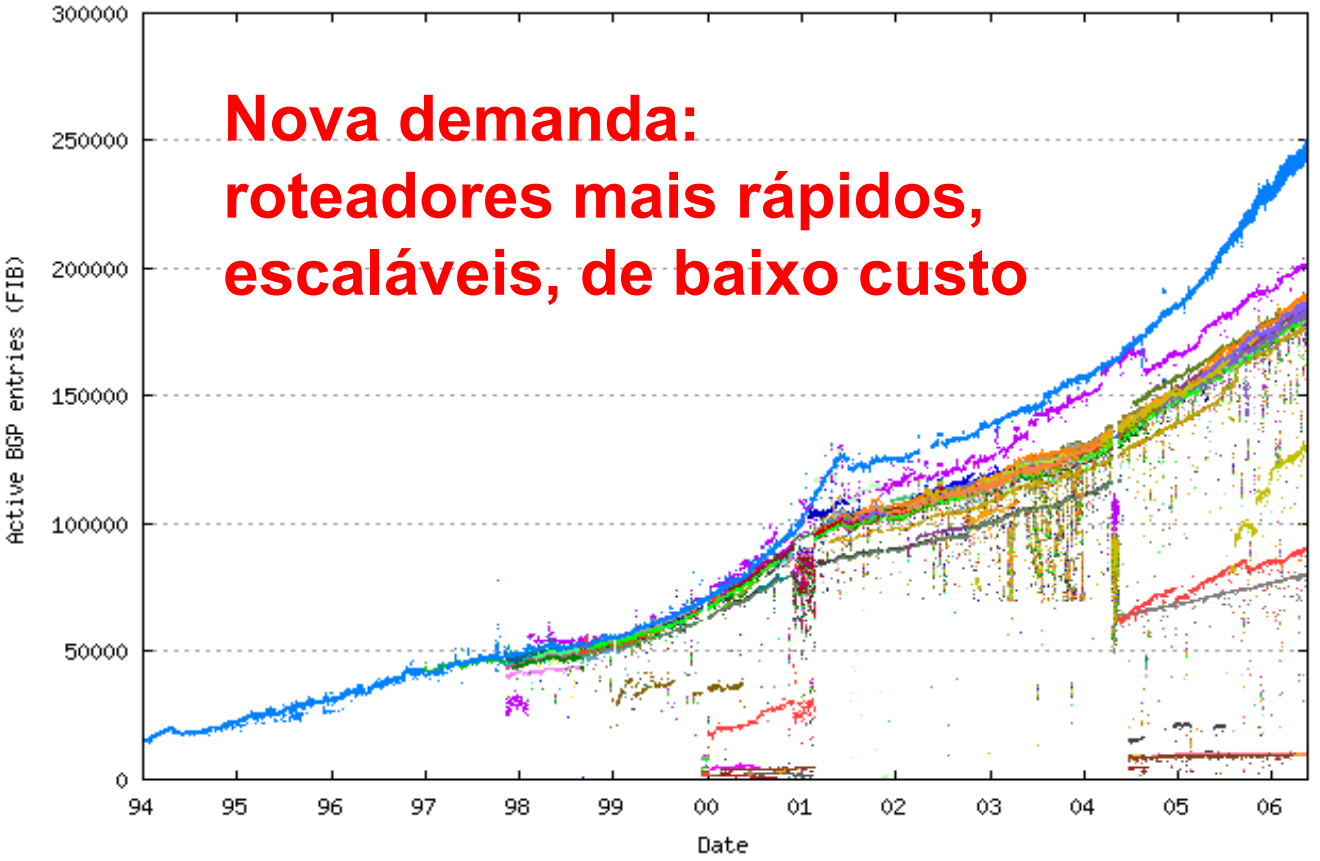
Representação de rotas baseado em anúncios BGP observados em pontos chave da rede



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Obtido de http://www.geog.ucl.ac.uk/casa/martin/atlas/topology.html (1999)

Mesmo com tudo isso...



Apresentação de Geoff Huston, 2006, evento: ???

- 01/01/2006
  - Table Size: 176,000 prefixes
  - Update Rate: 0.7M prefix / day
  - Withdrawal Rate 0.4M prefix / day
  - 250 Mbyte memory
  - 30% of a 1.5Ghz processor
- In 3-5 years (2009-2011)
  - 500,000 entries in the RIB
  - Update rate of up to 6M prefix /day
    - Short term peaks: 7000 prefix /sec
  - 2 Gbyte+ route processor memory
  - 5 GHz CPU for route processing
  - + Security processing overheads

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

Roteiro

- Interconexão de redes
- Modelo de serviço, endereços, máscaras, sub-redes
- Expedição de pacotes
  - fragmentação
  - entrega na rede local (ARP)
- Protocolos e técnicas auxiliares: ICMP, DHCP, VLANs, NAT
- Roteamento
  - RIP
  - OSPF
  - BGP
- IPv6

IPv6

- Motivação básica para criar nova versão de IP: esgotamento dos endereços
  - Tamanho do campo de endereço foi aumentado: 32 128 bits
  - Mas...
    - As pressões por mais endereços diminuíram por um tempo
      - uso de *firewalls*, IP *forwarding*, NAT, etc.
    - Mas agora há pressões no sentido inverso
      - P2P, multimídia pessoa-a-pessoa
    - Há discussões sobre a melhor maneira de implantação na rede como um todo
      - Mas o tempo está correndo!!! -> 2010? 2013?

IPv6

- Motivação secundária: suportar novas aplicações, como vídeo sob demanda e voz-sobre-IP
  - Cabeçalho inclui identificação de fluxo para roteadores com QoS
    - Utilização depende de políticas de cada backbone
    - As vantagens relativas ao tratamento de novas aplicações podem justificar adoção

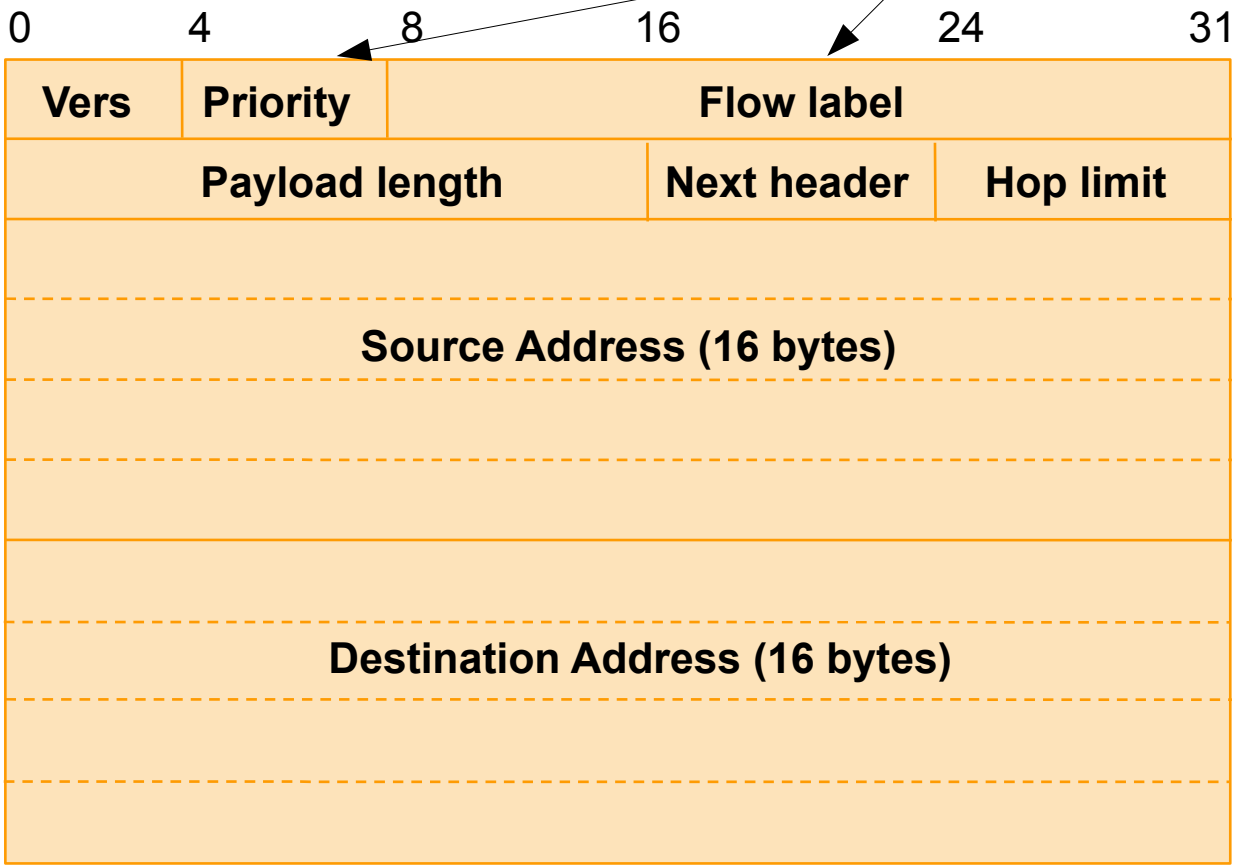
IPv6

- Características
  - endereços sem classes, com 128 bits
  - previsão para uso eficiente de *multicast*
  - serviço de tempo real
  - extensões do protocolo surgem como cabeçalhos extras:
    - fragmentação (fim-a-fim)
    - roteamento baseado na origem (*source routing*)
    - autenticação e segurança

IPv6

- Quase todos os campos mudaram
  - Cabeçalho básico:

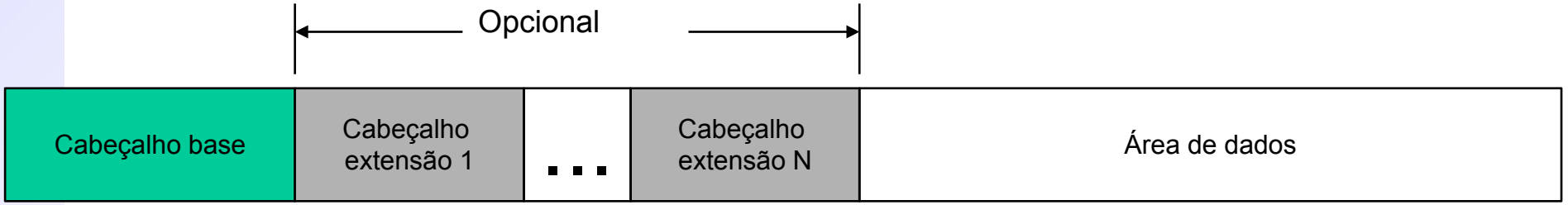
é possível associar pacotes a um **caminho definido** com uma **qualidade específica**



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

IPv6

- Novo formato simplifica o processamento
  - cabeçalho base contém apenas a informação essencial
  - o cabeçalho base pode ser seguido de cabeçalhos de extensão



UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

IPv6

- Espaço de endereçamento divido em um grande número de faixas

Prefix (binary)	Usage	Fraction
0000 0000	Reserved (including IPv4)	1/256
0000 0001	Unassigned	1/256
0000 001	OSI NSAP addresses	1/128
0000 010	Novell NetWare IPX addresses	1/128
0000 011	Unassigned	1/128
0000 1	Unassigned	1/32
0001	Unassigned	1/16
001	Unassigned	1/8
010	Provider-based addresses	1/8
011	Unassigned	1/8
100	Geographic-based addresses	1/8
101	Unassigned	1/8
110	Unassigned	1/8
1110	Unassigned	1/16
1111 0	Unassigned	1/32
1111 10	Unassigned	1/64
1111 110	Unassigned	1/128
1111 1110 0	Unassigned	1/512
1111 1110 10	Link local use addresses	1/1024
1111 1110 11	Site local use addresses	1/1024
1111 1111	Multicast	1/256

Estamos saltando:

- Roteamento para *hosts* móveis (mobile IP, seção 4.2.5)
  - Basicamente, utiliza-se túneis IP para ligar *host* móvel à rede de origem
- *Multicast* a nível de IP (seção 4.4)
  - Usando endereços classe D, roteadores podem fazer transmissões por multicast, atingindo diversos hosts sem sobrecarregar os links
  - Mas processamento da transmissão se torna oneroso para roteadores
  - Resultado: raramente disponível nas redes locais
    - Desabilitado pelos administradores
- MPLS (*Multiprotocolo Label Switching,* seção 4.5)
  - Proposta de combinar datagramas com circuitos
  - Cada pacote carrega um label (vide IPv6)
  - Roteadores podem identificar um circuito na rede e rotear pelo label


UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA
COMPUTAÇÃO