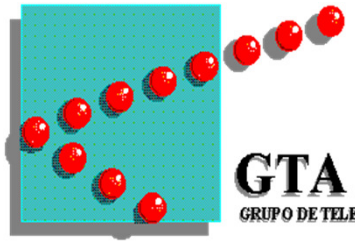


Roteamento em Redes de Computadores
CPE 825

Luís Henrique M. K. Costa

luish@gta.ufrj.br

Universidade Federal do Rio de Janeiro -PEE/COPPE
P.O. Box 68504 - CEP 21941-972 - Rio de Janeiro - RJ
Brasil - http://www.gta.ufrj.br

Roteiro

- Conceitos Básicos
- Roteamento Unicast
  - Intra-domínio
  - Inter-domínio
- Roteamento Multicast

Bibliografia

- **Christian Huitema, *Routing in the Internet*, Prentice Hall, 2nd. Edition**

- D. Medhi e K. Ramasamy, *Network Routing (Algorithms, Protocols, and Architectures)*, Morgan Kaufmann
- Costa, L. H. M. K. e Duarte, O. C. M. B., "*Roteamento Multicast na Internet",* Mini-curso SBRC'03
- Williamson, B. *Developing IP Multicast Networks*, Vol. 1, Cisco Press

- Artigos em alguns dos tópicos

Parte I
Conceitos Básicos:
A Internet

Padronização

- IETF (*Internet Engineering Task Force*)
  - Pequeno secretariado, financiado pelo governo americano
  - Grupos de Trabalho (*Working Groups*)
    - Organizados em áreas, cada área com um ou dois diretores
  - IESG (*Internet Engineering Steering Group*)
    - Formado pelos diretores de área e pelo IETF *chair*
    - Aprovam os padrões gerados pelos WGs
- IAB (*Internet Architecture Board*)
  - Estudos de longo prazo
  - Resolução de disputas
- IRTF (*Internet Research Task Force*)

Áreas do IETF

- *Applications Area*
- *General Area*
- *Internet Area*
- *Operations and Management Area*
- *Routing Area*
- *Security Area*
- *Sub-IP Area*
- *Transport Area*

Direitos Autorais

Note Well

All statements related to the activities of the IETF and addressed to the IETF are subject to all provisions of **Section 10 of RFC 2026**, which grants to the IETF and its participants certain licenses and rights in such statements. Such statements include verbal statements in IETF meetings, as well as written and electronic communications made at any time or place, which are addressed to:

-the IETF plenary session,

-any IETF working group or portion thereof,

-the IESG or any member thereof on behalf of the IESG,

-the IAB or any member thereof on behalf of the IAB,

-any IETF mailing list, including the IETF list itself, any working group or design team list, or any other list functioning under IETF auspices,

- the RFC Editor or the Internet-Drafts function

Statements made outside of an IETF meeting, mailing list or other function, that are clearly not intended to be input to an IETF activity, group or function, are not subject to these provisions.

Grupos de Trabalho do IETF

- Em geral, possuem vida curta
- São criados com um objetivo preciso
  - Especificado no *charter* do grupo

- *Internet-drafts*
  - Documentos de trabalho
  - Possuem vida limitada (6 meses)
  - Podem estar associados a um grupo de trabalho, ou não (submissões individuais)

Publicações

- *Request For Comments* (RFC)
- A série de documentos RFCs é o canal oficial de publicação dos padrões Internet, além de outras publicações do IESG, IAB e da comunidade Internet
- RFC 2026 – *The Internet Standards Process*
- Nem toda RFC define um padrão

Tipos de RFC

- *Internet Standard Specifications*
  - *Technical Specification* (TS)
    - Descrição de um protocolo, serviço, procedimento, convenção, ou formato
  - *Applicability Statement* (AS)
    - Especifica como, e em que circunstâncias, uma ou mais especificações técnicas (TSs) podem ser aplicadas para implementar uma aplicação Internet
- Best Current Practice (BCP)
  - Relata os melhores métodos de operação ou parâmetros de configuração utilizados na prática, em consenso geral

Processo de Padronização

- Os padrões evoluem seguindo níveis de maturidade, numa linha conhecida como "*standards track*"
- *Standards track*
  - *Proposed standard*
    - Especificação estável e aceita pela comunidade
    - O IESG decide quando a especificação atinge este nível
  - *Draft standard*
    - No mínimo duas implementações independentes e inter-operáveis
  - *Internet standard*
    - Padrão para o qual existem diversas implementações e sucesso operacional comprovado

A Trilha das RFCs Não-Padrão

- *Non-standards track*
  - especificações que não têm por objetivo se tornar padrões Internet
  - especificações que ainda não estão prontas para o *standards track*
  - especificações que foram revisadas, ou caíram em desuso
- *Experimental*
  - Experimento de pesquisa ou desenvolvimento
- *Informational*
  - Apenas informação, não representa consenso ou recomendação da comunidade Internet
- *Historic*
  - Especificações que foram revisadas ou deixaram de ser usadas

Princípios de Projeto da Internet

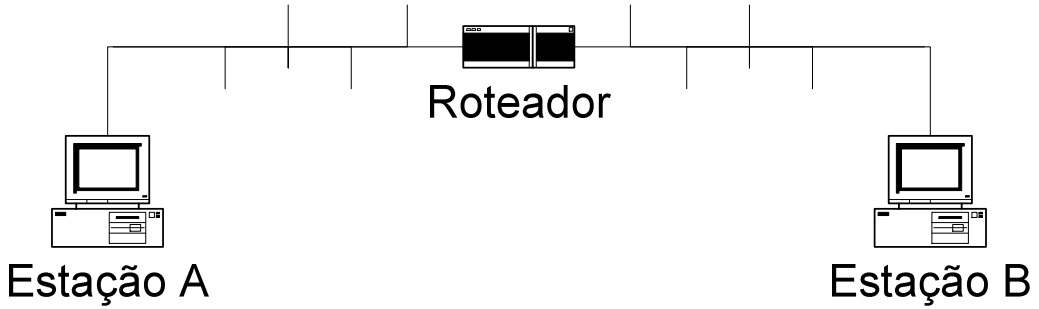
- Datagramas x circuitos virtuais
- Inteligência nos terminais
- A rede fornece a ***conectividade***, nada mais

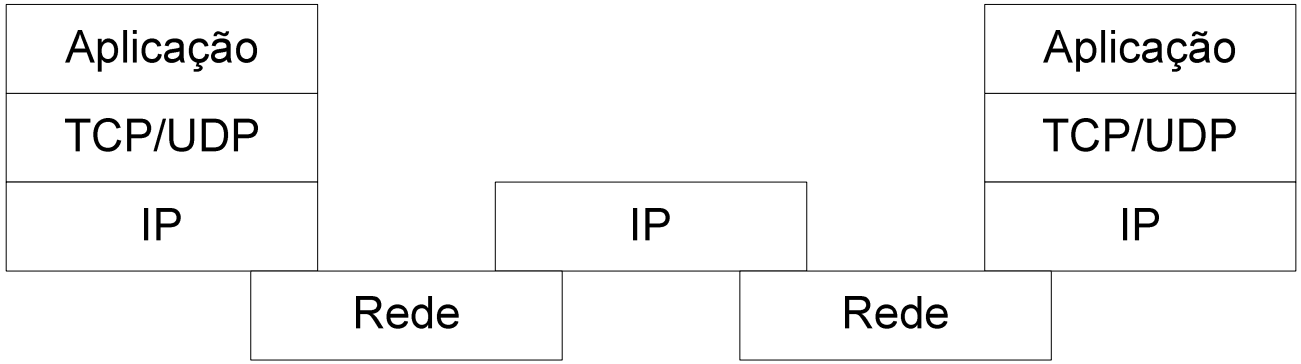
Envio de informação

- Internet Protocol – IP

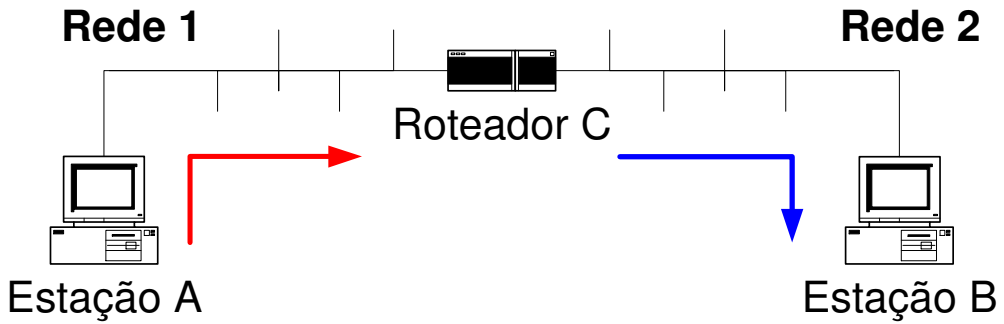
- *Internet Program*
  - Redes são interconectadas através de “programas inter-redes”
  - Cada sistema conectado à Internet executa uma instância deste programa inter-redes, ou internet
  - Aplicações geralmente acessam este programa através de um programa de transporte (ex. TCP, UDP)

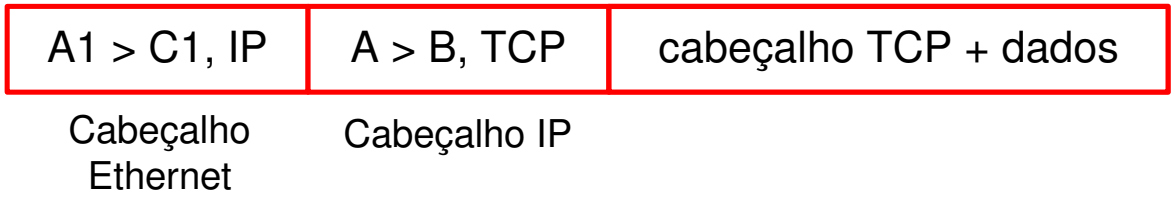
Operação do IP

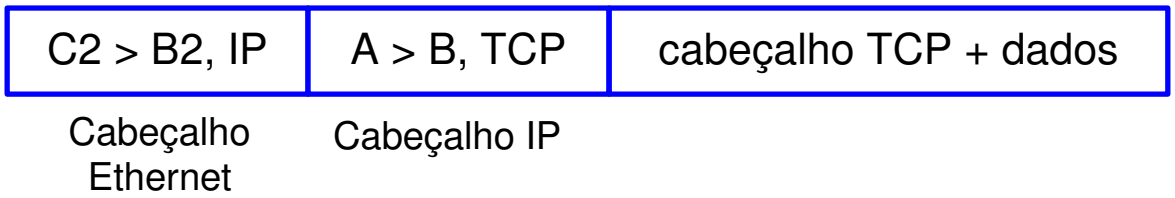




Transmissão de um Pacote IP







Endereçamento IP

- Cada **interface** de rede é identificada por um **endereço IP** de 32 bits
- Formato do Endereço IP
  - Dividido em duas partes, “identificador de rede” e “identificador de estação”
- 3 classes de “números de rede”, A, B e C
- Mais tarde, classe D definida para endereços multicast
- A classe E possui endereços reservados para utilização experimental

Classes de Endereços IP

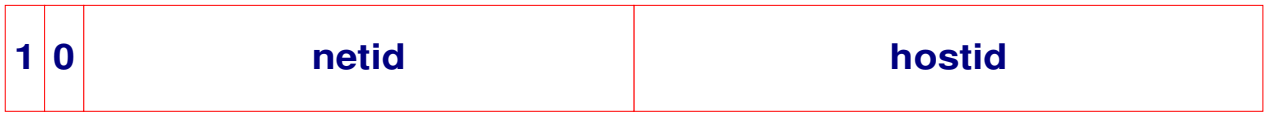
Classe	Bits mais significativos	Formato	
A	0	7 bits de redes	24 bits de estações
B	10	14 bits de redes	16 bits de estações
C	110	21 bits de redes	8 bits de estações
D	1110	28 bits de endereços de grupo multicast	
E	1111	reservados para testes	

Classes A, B e C

classe A

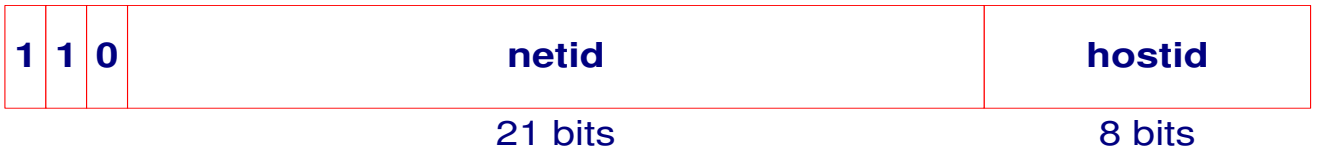
- 2^7 = 128 prefixos de classe A (0.x.x.x a 127.x.x.x)
- $(2^{24} - 2)$ = 16.777.214 estações em cada rede

classe B



- 2^{14} = 16.384 prefixos de classe B (128.x.x.x a 191.x.x.x)
- $(2^{16} - 2)$ = 65.534 estações em cada rede

classe C



- 2^{21} = 2.097.152 prefixos de classe C (192.x.x.x a 223.x.x.x)
- $(2^8 - 2)$ = 254 estações em cada rede

Estrutura de Endereçamento

- Quando o IP foi padronizado, em 1981

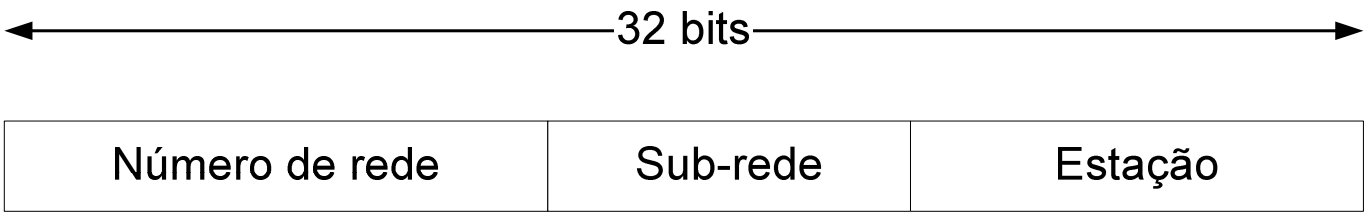
- *Números de rede* (*netid*) são alocados por uma autoridade de numeração Internet
- *Números de estação* (*hostid*) são alocados pelo gerente de rede

- A unicidade dos números de rede associada à unicidade dos números de estação garantem a ***unicidade global*** dos endereços IP unicast

Estrutura de Endereçamento

- Com a maior utilização de redes locais, estações de trabalho, e PCs, tornou-se necessário estruturar a rede dentro de cada organização

- Em 1984, o conceito de sub-rede (*subnet id*) foi adicionado ao endereço IP



Máscaras de sub-rede

Máscara	Endereço	Rede	Sub-rede	Estação
255.255.0.0 0x FF FF 00 00	10.27.32.100	A: 10	27	32.100
255.255.254.0 0x FF FF FE 00	136.27.33.100	B: 136.27	16 (32)	1.100
255.255.254.0 0x FF FF FE 00	136.27.34.141	B: 136.27	17 (34)	0.141
255.255.255.192 0x FF FF FF C0	193.27.32.197	C: 193.27.32	3 (192)	5

n-ésima rede da máscara (número da rede)

Endereços e Interfaces

- Endereços IP identificam ***interfaces de rede***, não identificam estações
- Uma estação com várias interfaces de rede possui vários endereços IP (a estação é dita *multi-homed*)
  - Ex. roteadores, estações que balanceiam o tráfego entre diversas redes
- Cada endereço pertence a uma sub-rede, que geralmente corresponde a uma "rede física"

Endereços e Interfaces

- ***Entradas na tabela de roteamento*** dos roteadores normalmente apontam para ***sub-redes***
  - Eventualmente, podem apontar para endereços de máquinas

- Porque não um endereço por estação?
  - Um endereço por interface permite ***escolher o caminho*** utilizado para chegar a uma estação
  - Endereços por interface permitem a ***agregação de endereços*** nas tabelas de roteamento
  - Se os endereços não fossem ligados à topologia, seria necessária uma entrada na tabela de roteamento para cada estação

Endereços e Interfaces

- Desvantagens
  - *Todos os endereços* de uma estação devem ser incluídos no *servidor de nomes*
  - O “melhor endereço” deve ser escolhido para uma conexão
  - O endereço fonte deve ser cuidadosamente escolhido pela aplicação
    - determina o caminho seguido pelos pacotes de resposta

Endereços Especiais

- "**0**" pode ser utilizado como endereço fonte, quando o número de rede é desconhecido, portanto:
  - **0.0.0.0** significa "esta estação nesta rede"
  - **0.X.Y.Z** significa "a estação **X.Y.Z** nesta rede"
  - utilizado por ex. quando uma estação está iniciando

- Difusão limitada (*limited broadcast*)
  - Formado por todos os bits em "**1**" – **255.255.255.255**
  - só pode ser utilizado como endereço destino
  - o pacote é enviado a todas as estações da sub-rede
  - não é retransmitido por um roteador

Endereços Especiais

- Difusão direcionada (*directed broadcast*)
  - Todos os bits da "parte estação" do endereço são colocados em "**1**"
  - Ex. "**A.255.255.255**", "**C.C.C.255**"
  - Com sub-redes a mesma regra é válida (todos os bits do complemento da máscara são colocados em "**1**")
- Conseqüências
  - Não existe sub-rede identificada apenas por **0**'s,
  - assim como não existe sub-rede identificada apenas por **1**'s
  - O tamanho da sub-rede é maior ou igual a 2 bits

Endereços Especiais

- Endereço de *loopback*
  - Na verdade, existe um número de rede de loopback:
    Rede Classe A **`127`**
- Qualquer endereço da forma **`127.X.Y.Z`** deve ser considerado local e não é transmitido para fora da estação

- Também existem diversos endereços de grupo multicast (classe D) reservados
  - Ex. **`224.0.0.1`** – todos os sistemas nesta sub-rede

Endereços Especiais

Endereço	Significado
0.0.0.0	Alguma estação desconhecida (fonte)
255.255.255.255	Qualquer estação (destino)
129.34.0.3	Estação número 3 na rede classe B 129.34
129.34.0.0	Alguma estação na rede 129.34 (fonte)
129.34.255.255	Qualquer estação na rede 129.34 (destino)
0.0.0.3	Estação número 3 “nesta rede”
127.0.0.1	Esta estação (*loop local*)

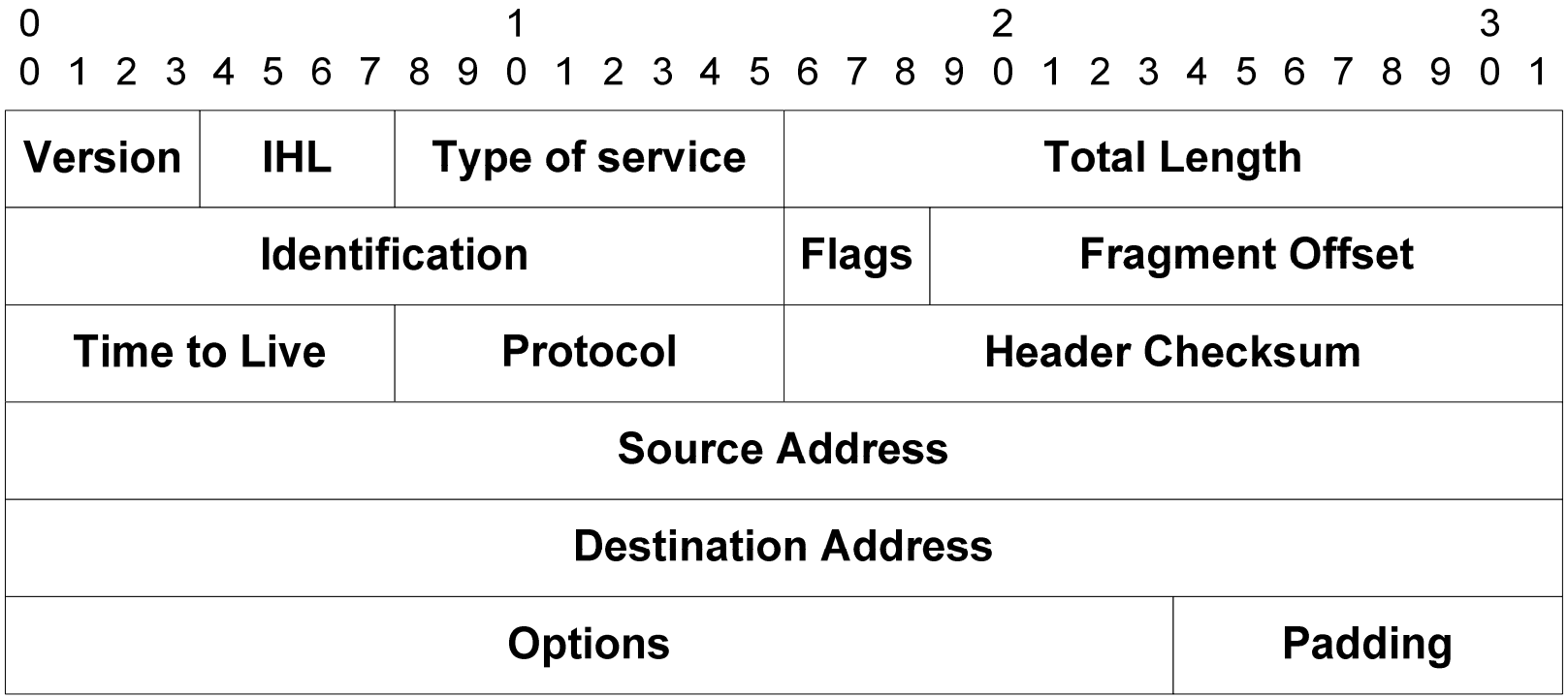
Alocação de Endereços IP

- IANA (*Internet Assigned Numbers Authority*)
- Os endereços IP são alocados através de delegações. Usuários recebem endereços IP de um provedor de serviço (ISP - *Internet Service Provider*). Os ISPs obtêm faixas de endereços IP de uma autoridade de registro local (LIR - *Local Internet Registry*), nacional (NIR - *National Internet Registry*), ou regional (RIR - *Regional Internet Registry*):
  - APNIC (*Asia Pacific Network Information Centre*) – Região Ásia/Pacífico
  - ARIN (American Registry for Internet Numbers) - América do Norte e África ao Sul do Saara
  - LACNIC (Regional Latin-American and Caribbean IP Address Registry) – América Latina e algumas Ilhas Caribenhas
  - RIPE NCC (Réseaux IP Européens) - Europa, Oriente Médio, Ásia Central e Áfrrica do Norte
- O papel do IANA é alocar faixas de endereços aos RIRs, de acordo com suas necessidades e a partir das faixas de endereços livres

Alocação de Endereços IP

- ICANN (*The Internet Corporation for Assigned Names and Numbers*)
- Organização sem fins lucrativos responsável pela
  - alocação do espaço de endereçamento IP,
  - atribuição de parâmetros de protocolos,
  - gerenciamento do sistema de nomes de domínios e
  - gerenciamento dos servidores raiz
- Estas funções eram atribuições do IANA e outras entidades através de contratos com o governo americano

IP - O Cabeçalho



- Todos os campos, exceto o de opções, são fixos

Campos do Cabeçalho IP

- Versão (4bits)
  - Versão atual = 4
  - Versão 5 = Protocolo ST-2
  - Versão 6 = "***A*** próxima geração"
  - Versões 7 e 8

- IHL (*Internet header's length*) (4bits)
  - comprimento do cabeçalho, em palavras de 32 bits
  - varia de 5 (quando não há opções) a 15
    - ou seja, podem haver 40 bytes de opções, no máximo

Campos do Cabeçalho IP

- Type of Service (8 bits)
  - Define a *precedência* e o *tipo de roteamento* desejado para o pacote
- Total Length (16 bits)
  - Comprimento total do pacote, incluindo o cabeçalho
  - Limita o tamanho do pacote a 65.535 bytes
- Identification, Flags e Fragment Offset
  - Utilizados no processo de fragmentação e remontagem

Campos do Cabeçalho IP

- Time to Live
  - ***Tempo de vida máximo*** do pacote na rede, em segundos
  - RFC-791: Um roteador deve sempre decrementar o TTL antes de retransmitir um pacote
    - O TTL deve ser decrementado de 1, se o tempo gasto nas filas e na transmissão ao próximo nó for menor que 1 segundo
    - Ou do número de segundos estimado
  - Na prática, estimar este tempo é difícil e o tempo de transmissão nos enlaces dificilmente ultrapassa 1 s
  - A maioria dos roteadores simplesmente decrementa o TTL de 1
  - Se o TTL atinge o valor 1, o pacote deve ser descartado
    - sinal de que o pacote já trafegou por mais tempo que o devido...

Campos do Cabeçalho IP

- Source Address e Destination Address (32bits cada)
  - Identificam a fonte e destino do pacote
- Protocol (8 bits)
  - Determina o programa para o qual o pacote é passado, no destino

Decimal	Sigla	Protocolo
0		Reservado
1	ICMP	Internet Control Message
2	IGMP	Internet Group Management
4	IP	IP em IP (encapsulação)
6	TCP	Transmission Control

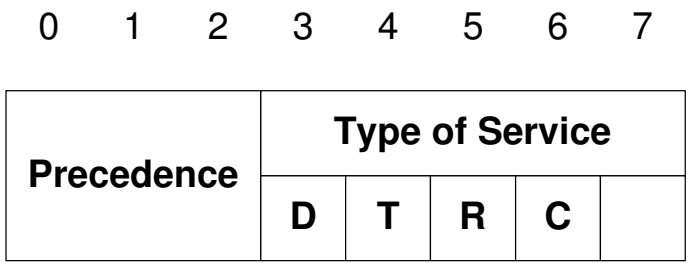
Decimal	Sigla	Protocolo
17	UDP	User Datagram
29	ISO-TP4	ISO Transport Prot Class 4
80	ISO-IP	ISO Internet Protocol (CLNP)
89	OSPF	Open Shortest Path First
255		Reservado

Campos do Cabeçalho IP

- Header Checksum (16 bits)
  - Proteção do cabeçalho contra erros
- Calculado como “o complemento a 1 da soma em complemento a 1 de todas as palavras de 16 bits do cabeçalho, considerando os bits do *checksum* em 0.”
- Não protege contra inserção de palavras em zero (16 bits iguais a zero) ou inversão de palavras
- Mas é de simples implementação

Precedência e Tipo de Serviço

- Precedence (3 bits)
  - Indica a prioridade do pacote
  - Valores maiores, maior prioridade
  - RFC791 diz que a precedência é válida apenas dentro de uma rede



- Type of Service (5 bits)
  - Indicação para o roteamento
  - Útil quando existem múltiplas rotas
    - Rota com o melhor
      - D – delay
      - T – throughput
      - R – reliability
      - C – cost

- Atualmente, o campo está sendo revisto, de acordo com a definição dos Serviços Diferenciados (DiffServ)

Serviços Diferenciados

- DS field

- DSCP field (6 bits)(RFC2474)
  - Differentiated Services Code Points
  - Diferentes classes de serviço no encaminhamento de pacotes

- ECN field (2 bits)(RFC3168)
  - Explicit Congestion Notification
  - Auxílio à camada de transporte para o controle de congestionamento

Nomenclatura do DSCP

Prioridade de tráfego

- CS: Class Selector
  - Equivalentes a 8 prioridades do IP Precedence

Classes de Serviço do DiffServ

- AF: Assured Forwarding
  - ***Garantia de entrega***, desde que não se exceda taxa contratada
  - Em caso de congestionamento, pacotes são descartados com diferentes probabilidades
- EF: Expedited Forwarding
  - ***Prioridade estrita de enfileiramento*** sobre todas as outras classes

Códigos do Campo DS

Nome do DSCP	Valor do campo DS	IP Precedence
CS0	0	0: best effort
CS1, AF11-AF13	8, 10, 12, 14	1: priority
CS2, AF21-AF23	16, 18, 20, 22	2: immediate
CS3, AF31-AF33	24, 26, 28, 30	3: flash
CS4, AF41-AF43	32, 34, 36, 38	4: flash override
CS5, EF	40, 46	5: critical
CS6	48	6: internetwork control
CS7	56	7: network control

Classes do Serviço AF

	Classe 1	Classe 2	Classe 3	Classe 4
Prob. de descarte baixa	AF11 (DSCP 10)	AF21 (DSCP 18)	AF11 (DSCP 26)	AF11 (DSCP 34)
Prob. de descarte média	AF12 (DSCP 12)	AF22 (DSCP 20)	AF12 (DSCP 28)	AF12 (DSCP 36)
Prob. de descarte alta	AF13 (DSCP 14)	AF23 (DSCP 22)	AF13 (DSCP 30)	AF13 (DSCP 38)

- Classes 1 a 4
  - ***Possuem a mesma prioridade***
  - Em cada classe, três ***probabilidades*** de descarte crescentes
  - Se houver congestionamento entre ***tráfegos de diferentes classes:***
    - Tráfego na classe mais alta tem prioridade

Explicit Congestion Notification

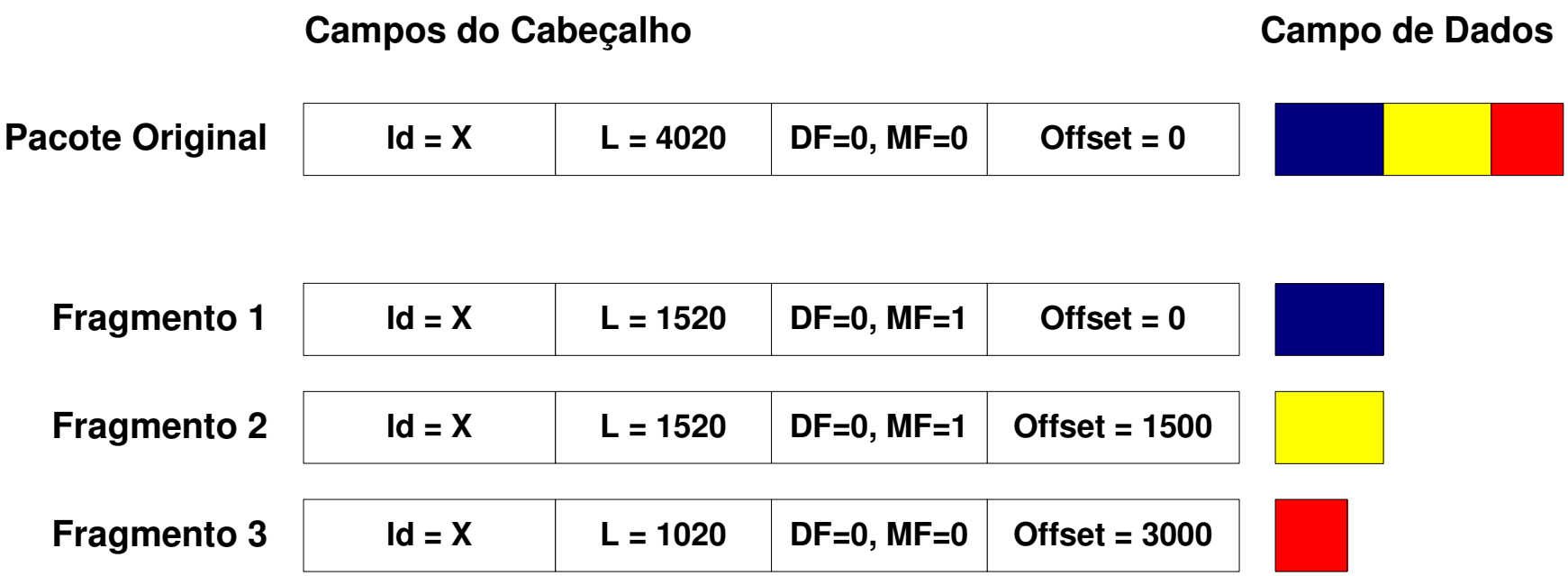
- Explicit Congestion Notification
  - ECT(0) ou ECT (1)
    - Os terminais utilizam um protocolo de transporte capaz de usar a notificação de congestionamento
    - Se o Transporte não souber diferenciar entre ECT(0) e ECT(1), usa-se o ECT(0)
  - CE
    - O pacote foi marcado com a indicação de que há congestionamento iminente (o roteador utiliza gerenciamento ativo de fila com RED (random early detection)

ECN field	Significado
00	Prot. de Transporte não capaz de ECN
01	ECN Capable Transport, ECT(1)
10	ECN Capable Transport, ECT(0)
11	Congestion Encountered, CE

Fragmentação e Remontagem

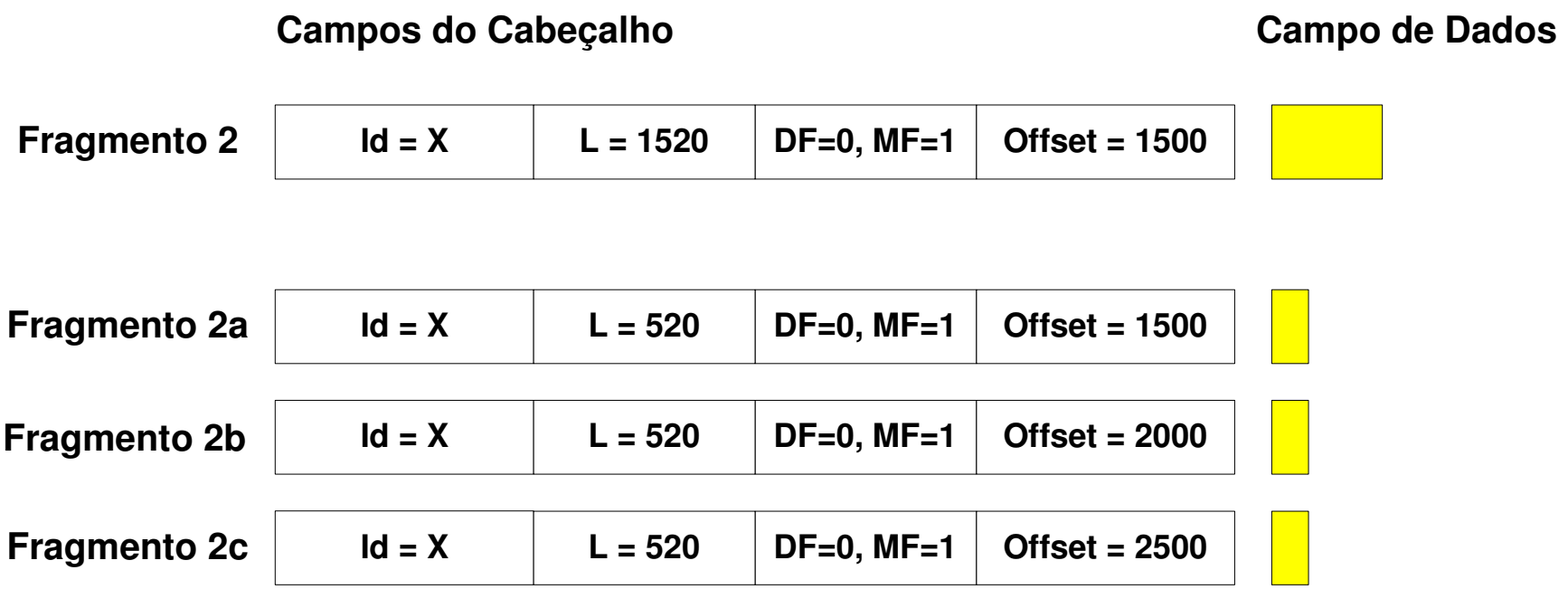
- A fragmentação é necessária quando um roteador conecta duas tecnologias de rede com tamanho máximo de quadro diferentes
- Identification (16 bits), Flags (3 bits) e Fragment Offset (13 bits)
- Flags
  - Bit 0 – reservado
  - Bit 1 – *don´t fragment* (DF)
  - Bit 2 – *more fragments* (MF)
- Cada fragmento possui um cabeçalho completo, igual ao do pacote original, exceto pelos campos de comprimento, offset e o bit MF

Fragmentação e Remontagem



- O bit MF é sempre 1, exceto no último fragmento

Fragmentação e Remontagem



- Os campos MF e offset são calculados com relação ao pacote original

Fragmentação e Remontagem

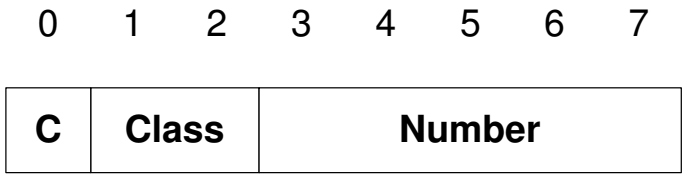
- O campo identificação (16 bits) associado ao endereço de origem identifica o pacote
- O receptor deve "expirar" pacotes parcialmente remontados, após um certo período de espera
  - Por ex., decrementando o campo TTL a cada segundo
- O emissor só pode reutilizar um identificador após o período igual ao TTL utilizado

Evitando a Fragmentação

- A reutilização dos identificadores limita a taxa de transmissão possível
  - 16 bits = 65.536 pacotes por TTL
  - TTL recomendado pelo TCP = 2 min
  - Limite de 544 pacotes por segundo
    - 17Mbps com pacotes de 4kbytes
- A fragmentação é ineficiente combinada com o TCP
  - Perda de um fragmento implica retransmissão do pacote inteiro
- O TCP implementa um mecanismo de descoberta da MTU (*Maximum Transmission Unit*) do caminho
  - Tentativas com diferentes tamanhos de pacote, com o DF em 1

Opções do IP

- Definido para criação de funcionalidades especiais, através do roteamento específico de *alguns* pacotes
- Options
  - Pode transportar vários parâmetros
  - Cada opção começa por um byte de 'tipo de opção"



    - Flag C (*Copied*)
      - Indica que a opção deve ser copiada em todos os fragmentos
    - Class
      - 0 – opções de controle
      - 2 – opções de *debug* e medidas
    - Number
      - Identifica uma opção dentro de cada classe
  - O segundo byte normalmente indica o comprimento da opção

Opções do IP

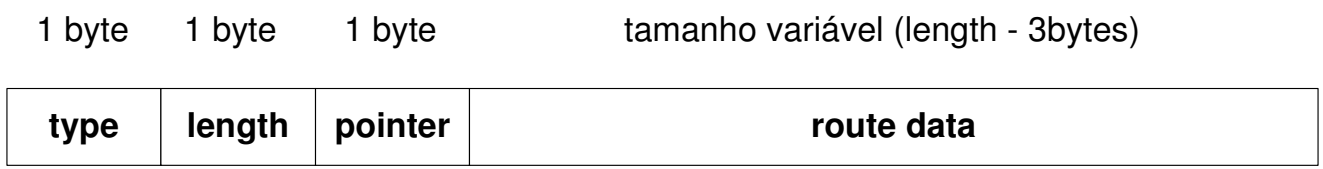
Classe	Número	Compr.	Significado
0	0	-	**End of Option list.** Indica o fim da lista de opções, possui apenas 1 byte. Não há byte de comprimento.
0	1	-	**No Operation.** Possui apenas 1 byte. Não há byte de comprimento.
0	2	11	**Security.** Utilizada para carregar parâmetros de segurança definidos pelo dep. de defesa americano.
0	3	var.	**Loose Source Routing.** Utilizada para rotear o pacote IP de acordo com a informação fornecida pela fonte.
0	7	var.	**Record Route.** Utilizada para registrar a rota atravessada pelo pacote IP.
0	8	4	**Stream ID.** Utilizada para carregar o identificador do stream.
0	9	var.	**Strict Source Routing.** Utilizada para rotear o pacote IP de acordo com a informação fornecida pela fonte.
2	4	var.	**Internet Timestamp.**

Opções do IP

- No operation
  - Utilizada para enchimento entre opções, de forma que o início da opção está alinhado em 32 bits
- End of option
  - Indica o ponto onde a opção termina, mesmo se o campo comprimento total indicar mais espaço alocado para opções
- A maioria das opções não é usada
  - Stream ID foi usada apenas no experimento Satnet
  - Security codifica necessidades militares do final dos anos 70
  - Timestamp e route record visavam serviços que o programa `traceroute` implementa
- Apenas loose e strict source routing foram utilizadas

Loose e Strict Source Routing

- Sintaxe



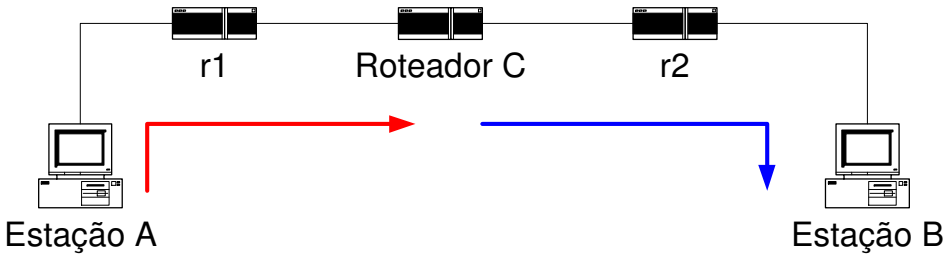
- Route data
  - Contém a lista de endereços pelos quais o pacote deve passar
- Funcionamento
  - O campo destination possui o próximo nó pelo qual o pacote deve passar
  - Quando este destino é atingido, a opção é examinada
  - O pointer indica um número de octetos a partir do início da opção, de onde deve ser lido o próximo endereço
  - Se pointer > comprimento da opção, o destino final foi atingido
  - No strict source routing, o próximo endereço deve ser um roteador vizinho, enquanto no loose source routing, não

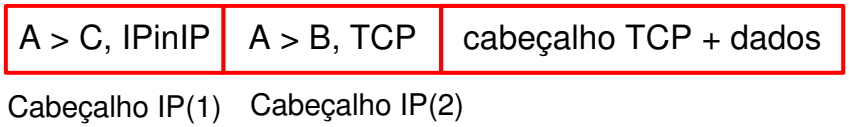
Processamento do Cabeçalho IP

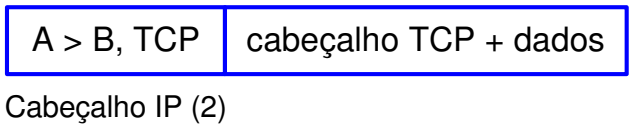
- Operações
  - Verificação da versão, do *checksum*, tamanho do pacote, e leitura das opções (se houver)
  - Consultar a tabela de roteamento para o destino e tipo de serviço do pacote, obter a interface e endereço no meio físico
- Roteadores otimizam as operações mais comuns (*fast-path*)
  - Ex. caches com rotas mais utilizadas
- Pacotes ***sem opções*** possuem cabeçalho de tamanho fixo, passam pelo *fast-path*
- Pacotes ***com opções*** seguem o caminho "normal"
  - Além disso, em alguns roteadores, pacotes com opções possuem menos prioridade para aumentar o desempenho global

Evitando a opção Source Routing

- Envio de um pacote de A para B, passando pelo roteador C





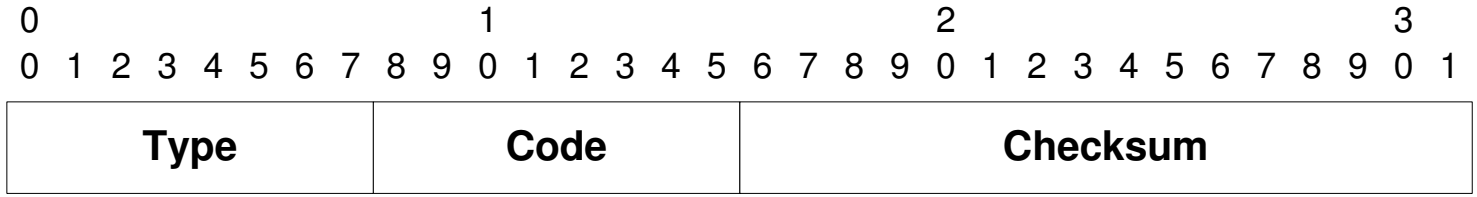


Internet Control Message Protocol

- Objetivo
  - Diagnóstico de condições de erro da rede
- Executado em cima do IP
    - Protocol type = 1
- Parte integrante do *Internet Program*
  - Todo sistema que roda IP deve rodar o ICMP
- Não provê confiabilidade, apenas informação sobre problemas na rede
- Erros de transmissão de pacotes IP geram mensagens ICMP
  - Exceto erros nas próprias mensagens ICMP
  - Evita-se a recursividade e avalanche de mensagens de controle

Mensagens ICMP

- Cabeçalho
  - Toda mensagem ICMP possui uma parte do cabeçalho em comum

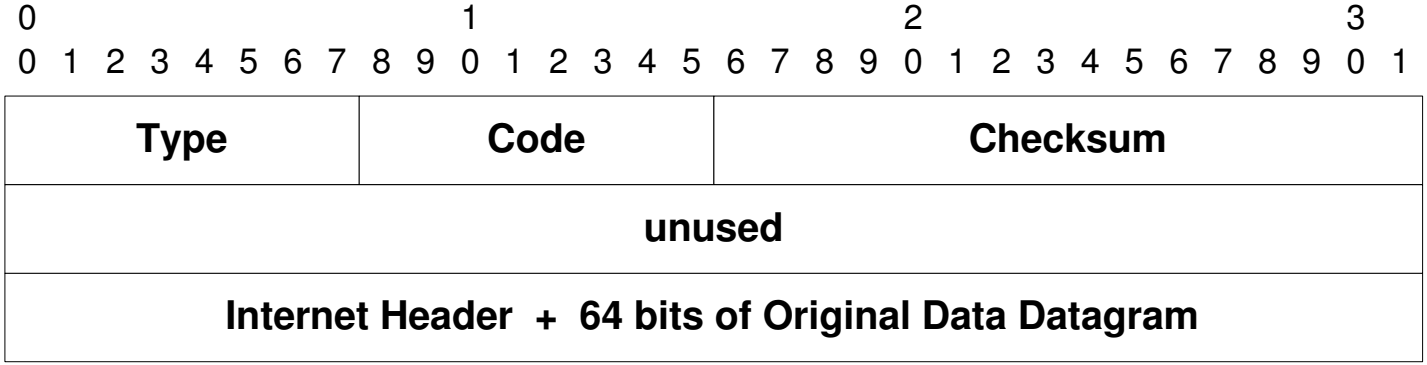


  - O *checksum* do cabeçalho é calculado como para o IP

Tipo	Significado
0	Echo Reply
3	Destination Unreachable
4	Source Quench
5	Redirect
8	Echo
9	Router Advertisement
10	Router Solicitation
11	Time Exceeded
12	Parameter Problem
13	Timestamp
14	Timestamp Reply
15	Information Request
16	Information Reply

Diagnóstico com o ICMP

- Problemas operacionais
  - Time Exceeded
  - Destination Unreachable
  - Source Quench



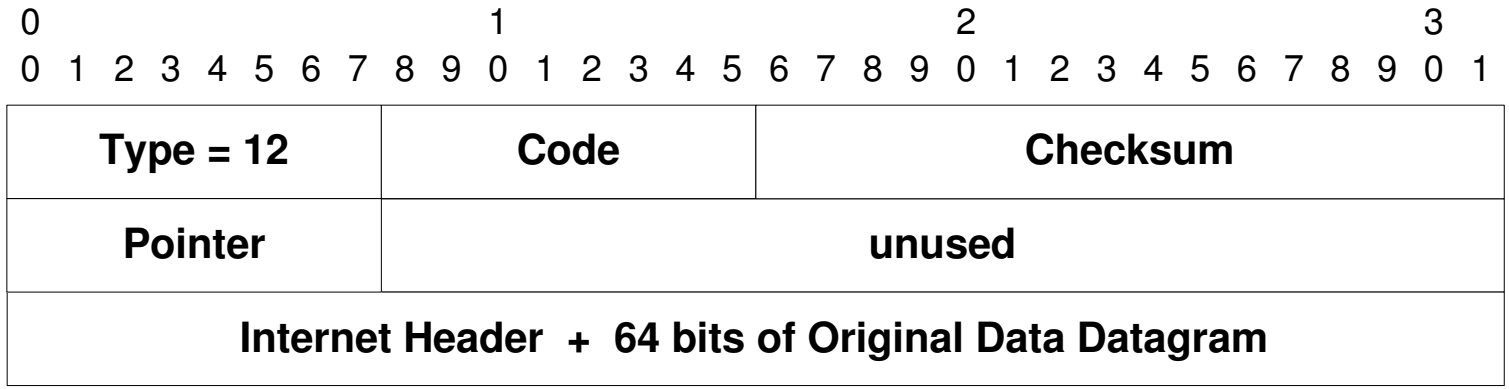
- Formato comum
  - Cabeçalho básico do ICMP +
  - 32 bits de enchimento +
  - Primeiros bytes do pacote que causou o envio do ICMP

Diagnóstico com o ICMP

- Destination Unreachable
  - Código
    - 0 = net unreachable
    - 1 = host unreachable
    - 2 = protocol unreachable
    - 3 = port unreachable
    - 4 = fragmentaion needed but DF set
    - 5 = source route failed
- Time Exceeded
  - TTL estourado
  - Código
    - 0 = em trânsito
    - 1 = durante remontagem
- Source Quench
  - Enviado pelo roteador para sinalizar congestionamento
  - Não utiliza código (code = 0)

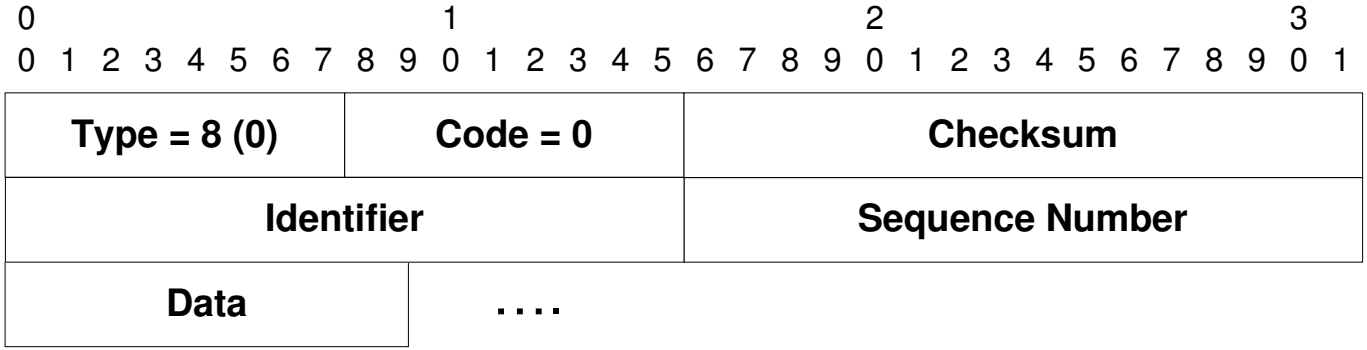
Diagnóstico com o ICMP

- Parameter Problem
  - Enviado por um roteador ao encontrar um erro de codificação *no cabeçalho* do pacote IP
  - O ***ponteiro*** identifica o byte no datagrama original onde foi encontrado o erro



Ping

- Testa se uma estação está “viva”
- Utiliza a função echo do ICMP
  - **Type = 8 – Echo**
  - **Type = 0 – Echo Reply**



  - Resposta
    - Endereços fonte e destino são trocados
    - Troca do valor do tipo da mensagem
    - *Checksums* IP e ICMP recalculados
    - Dados inalterados

Ping

- Campos identificação e número de seqüência possibilitam estatísticas
- Outras mensagens ICMP com funcionalidade semelhante
  - Type = 15 – Information Request
  - Type = 16 – Information Reply

Exemplo de Ping

```
PING angra (146.164.69.1) from 146.164.69.2 : 56(84) bytes of data.
recreio::luish [ 31 ] ping angra
64 bytes from angra (146.164.69.1): icmp_seq=1 ttl=64 time=0.471 ms
64 bytes from angra (146.164.69.1): icmp_seq=2 ttl=64 time=0.404 ms
64 bytes from angra (146.164.69.1): icmp_seq=3 ttl=64 time=0.544 ms
64 bytes from angra (146.164.69.1): icmp_seq=4 ttl=64 time=0.388 ms
64 bytes from angra (146.164.69.1): icmp_seq=5 ttl=64 time=0.398 ms
64 bytes from angra (146.164.69.1): icmp_seq=6 ttl=64 time=0.398 ms
64 bytes from angra (146.164.69.1): icmp_seq=7 ttl=64 time=0.495 ms
64 bytes from angra (146.164.69.1): icmp_seq=8 ttl=64 time=0.436 ms
64 bytes from angra (146.164.69.1): icmp_seq=9 ttl=64 time=0.413 ms
64 bytes from angra (146.164.69.1): icmp_seq=10 ttl=64 time=0.407 ms
64 bytes from angra (146.164.69.1): icmp_seq=11 ttl=64 time=0.393 ms
64 bytes from angra (146.164.69.1): icmp_seq=12 ttl=64 time=0.391 ms

--- angra ping statistics ---
12 packets transmitted, 12 received, 0% loss, time 11109ms
rtt min/avg/max/mdev = 0.388/0.428/0.544/0.049 ms
```

Traceroute

- Identificação dos roteadores entre uma fonte e um destino

- Funcionamento
  - Envio sucessivo de pacotes para o destino, variando o TTL
    - UDP numa porta não utilizada
  - TTL inicial = 1
  - Primeiro roteador decrementa o TTL, descarta o pacote, e envia uma mensagem ICMP TTL Exceeded
    - Endereço fonte identifica o roteador
  - A fonte continua o processo incrementando o TTL de 1
  - Até chegar ao destino, ou um enlace com problemas ser identificado
  - O destino é identificado, pois este envia uma mensagem ICMP port unreachable

Exemplo - Traceroute

```
recreio::luish [ 38 ] traceroute sphinx.lip6.fr
traceroute to sphinx.lip6.fr (132.227.74.253), 30 hops max, 38 byte packets
 1  angra (146.164.69.1)  0.596 ms  0.349 ms  0.341 ms
 2  rt-ct-bloco-H.ufrj.br (146.164.5.193)  175.723 ms  203.553 ms  30.226 ms
 3  rt-nce2.ufrj.br (146.164.1.5)  51.432 ms  3.994 ms  4.137 ms
 4  rederio2-atm-cbpf.rederio.br (200.20.94.58)  3.495 ms  4.421 ms  4.664 ms
 5  200.143.254.66 (200.143.254.66)  4.184 ms  12.224 ms 200.143.254.78
    (200.143.254.78)  13.372 ms
 6  rj7507-fast6_1.bb3.rnp.br (200.143.254.93)  4.473 ms  4.135 ms  4.550 ms
 7  ds3-rnp.ampath.net (198.32.252.237)  110.658 ms  106.239 ms  107.241 ms
 8  abilene.ampath.net (198.32.252.254)  125.393 ms  135.971 ms  127.111 ms
 9  washng-atla.abilene.ucaid.edu (198.32.8.66)  143.388 ms  154.348 ms  144.619 ms
10  abilene.de2.de.geant.net (62.40.103.253)  234.914 ms  235.300 ms  239.316 ms
11  de2-1.de1.de.geant.net (62.40.96.129)  234.644 ms  238.821 ms  236.147 ms
12  de.fr1.fr.geant.net (62.40.96.50)  231.422 ms  232.743 ms  232.437 ms
13  renater-gw.fr1.fr.geant.net (62.40.103.54)  234.984 ms  234.233 ms  231.723 ms
14  jussieu-a1-1-580.cssi.renater.fr (193.51.179.154)  230.906 ms  231.090 ms
    233.714 ms
15  rap-jussieu.cssi.renater.fr (193.51.182.201)  232.602 ms  232.125 ms  238.066 ms
16  cr-jussieu.rap.prd.fr (195.221.126.77)  235.182 ms  239.903 ms  276.221 ms
17  jussieu-rap.rap.prd.fr (195.221.127.182)  234.955 ms  237.264 ms  234.210 ms
18  r-scott.reseau.jussieu.fr (134.157.254.10)  233.992 ms  238.306 ms  239.047 ms
19  olympe-gw.lip6.fr (132.227.109.1)  236.396 ms !N  235.261 ms !N  234.322 ms !N
```

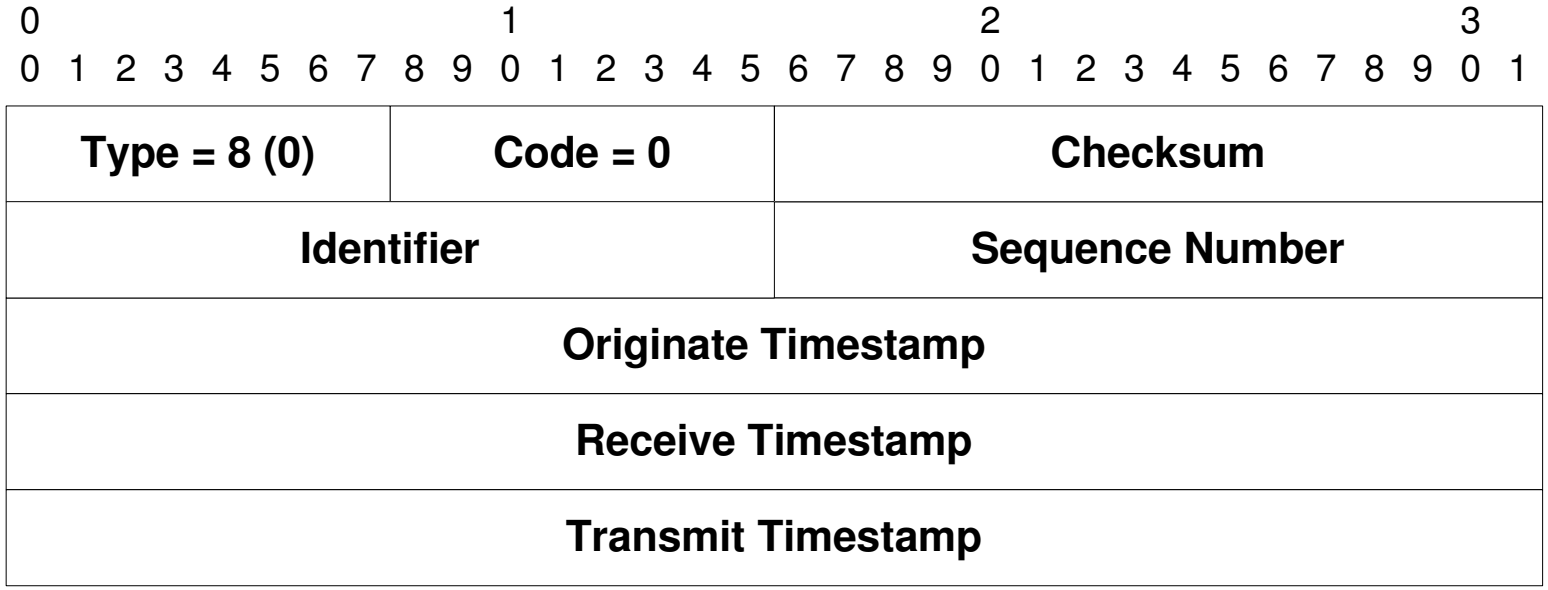
Exemplo – Ping -R

```
recreio::luish [ 35 ] ping -R sphinx.lip6.fr
PING sphinx.lip6.fr (132.227.74.253) from 146.164.69.2 : 56(124) bytes of data.
64 bytes from sphinx.lip6.fr (132.227.74.253): icmp_seq=1 ttl=237 time=252 ms
RR:     recreio (146.164.69.2)
        gtagw (146.164.5.210)
        rt-ct2.ufrj.br (146.164.1.3)
        ufrj-atm.rederio.br (200.20.94.9)
        200.143.254.65
        rj-fast4_1.bb3.rnp.br (200.143.254.94)
        rnp.ampath.net (198.32.252.238)
        abilene-oc3.ampath.net (198.32.252.253)
        atla-washng.abilene.ucaid.edu (198.32.8.65)
64 bytes from sphinx.lip6.fr (132.227.74.253): icmp_seq=2 ttl=237 time=289 ms
RR:     recreio (146.164.69.2)
        ...
64 bytes from sphinx.lip6.fr (132.227.74.253): icmp_seq=3 ttl=237 time=247 ms
RR:     recreio (146.164.69.2)
        ...
--- sphinx.lip6.fr ping statistics ---
3 packets transmitted, 3 received, 0% loss, time 2021ms
rtt min/avg/max/mdev = 247.821/263.167/289.150/18.477 ms
```

Gerenciamento de Tempo

- Mensagens
  - Type = 13 – Timestamp
  - Type = 14 – Timestamp reply

  - Tempos expressos em ms desde 0:00hs Greenwich time



Cálculo da defasagem entre 2 estações

- Funcionamento
  - Estação A preenche o tempo de origem (To)
  - Na recepção, a estação B preenche o tempo de recepção (Tr)
    - A estação B prepara a resposta
  - Antes do envio da resposta, B preenche o tempo de transmissão (Tt)
  - Ao receber a resposta, A armazena o tempo de chegada (Tc)

- **Defasagem** = Diferença medida de relógios – tempo de transmissão
- Tempo de transmissão = RTT/2 (*Round Trip Time*)
  - RTT = Tc – To – (Tt – Tr)

- **Defasagem** = Tr – To – RTT/2

Envio de Pacotes IP

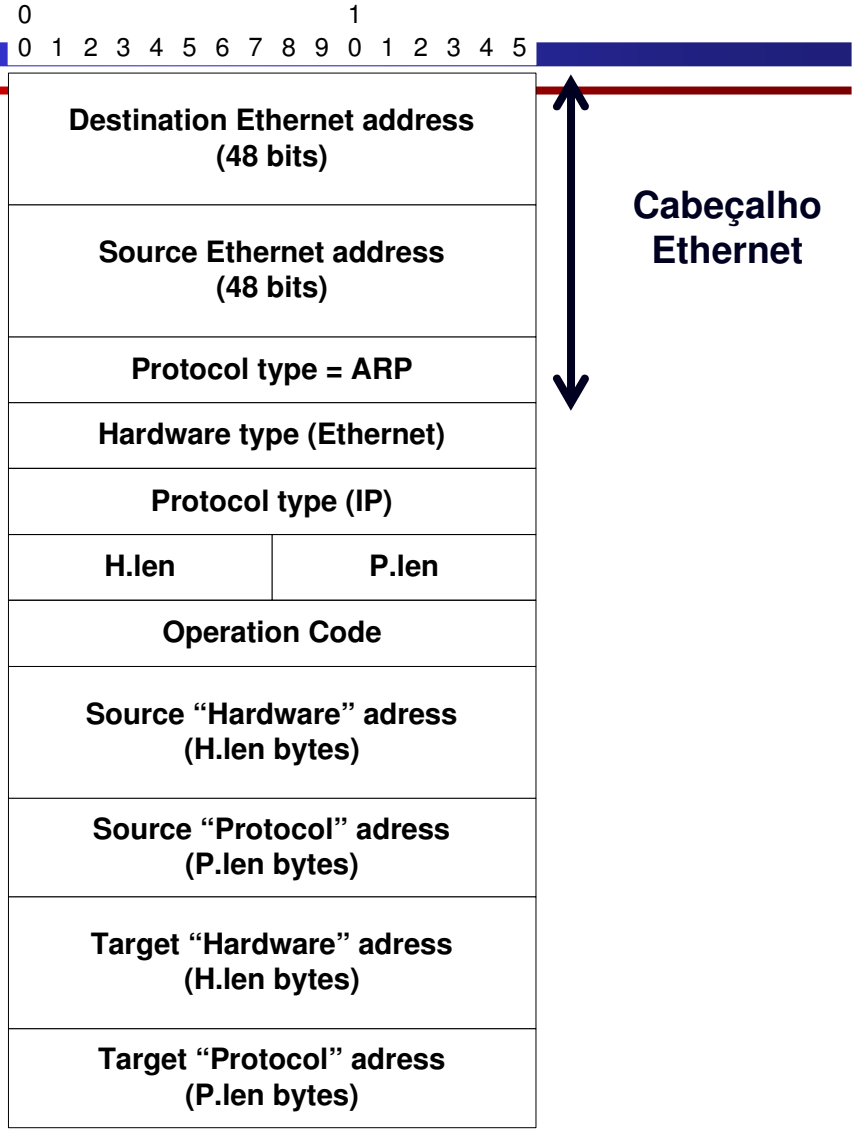
- No IP, existem
  - Roteadores (executam um protocolo de roteamento)
  - Estações (não, necessariamente, executam um protocolo de roteamento)
- Porque...
  - Complexidade de protocolos de roteamento modernos
  - Variedade de protocolos de roteamento
  - Poderia-se apenas “ouvir” as mensagens de roteamento
    - Algumas vezes este processo pode não ser fácil
      - Ex. mecanismos de segurança (autenticação, criptografia)
- Para enviar pacotes, a estação deve
  - ***Descobrir*** um roteador de saída
  - Ouvir mensagens de ***redirecionamento***

Descoberta do próximo salto

- Dado um pacote IP a transmitir, a quem enviar?
  - Estação destino na rede
    - envio direto
  - Estação destino distante
    - envio a um roteador, que encaminhará o pacote
- Dado o endereço IP destino
  - Teste da máscara de rede diz se a estação está na sub-rede
- Próximo passo
  - Descoberta do endereço “físico” do próximo salto

Address Resolution Protocol (ARP)

- A estação envia um ARP request (op. code 1) em broadcast
- A máquina que reconhece seu IP no request envia um ARP response (op. code 2)
- As máquinas implementam um *cache* para evitar o envio freqüente de ARPs

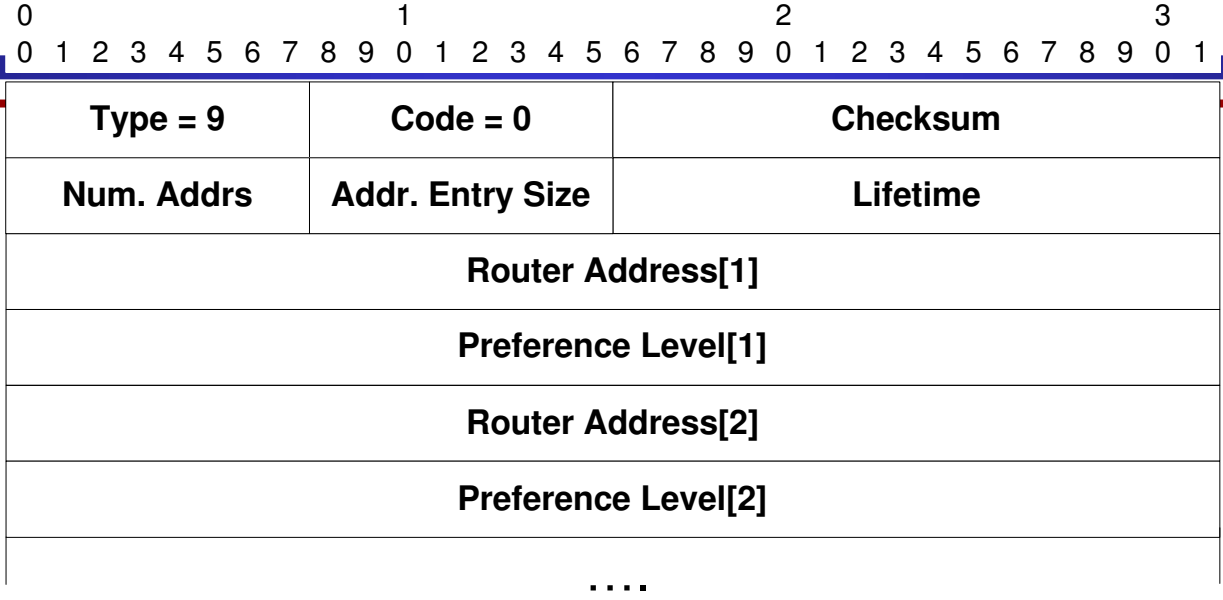


Descoberta do Roteador

- Por configuração
- Usando o ICMP
  - Roteadores enviam mensagens ICMP router advertisement (type = 10) ***periodicamente***
  - Estações podem disparar o envio de anúncios utilizando mensagens de solicitação (ICMP router solicitation, type = 9)

- O objetivo do procedimento é descobrir ***um*** roteador de saída, não necessariamente ***o melhor*** roteador de saída...
  - Mensagens ICMP redirect podem ser utilizadas para informar as estações de rotas melhores

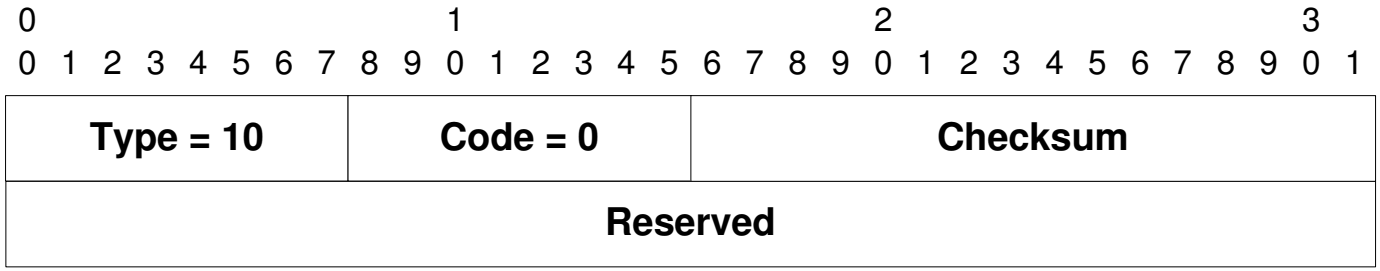
Anúncios (Router advertisements)



- Podem conter diversos endereços para o mesmo roteador
  - Várias interfaces conectadas à mesma rede
  - Uma interface de rede com dois endereços IP
    - Ex. duas sub-redes IP na mesma rede física (segmento Ethernet p. ex.)
  - Preference - prioridade de escolha entre vários roteadores
    - Configurado pelo administrador da rede
  - Addr. Entry Size = 2

Anúncios (Router advertisements)

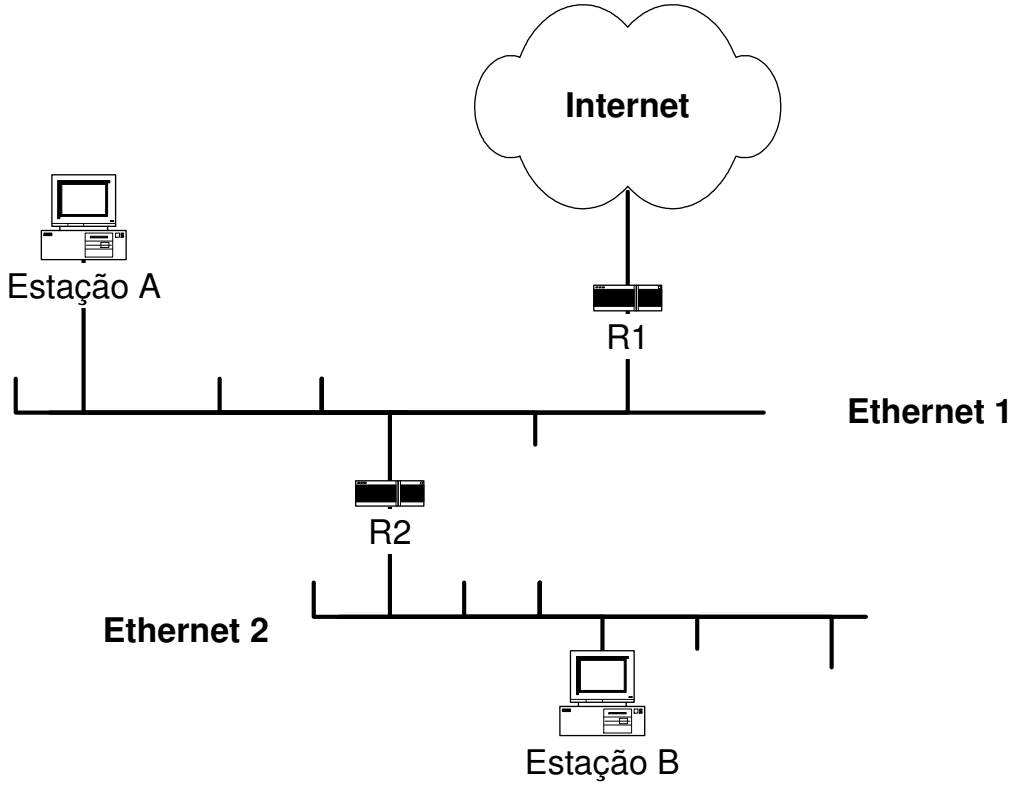
- São enviados ao endereço **224.0.0.1** (todas as máquinas) ou a **255.255.255.255**
- Informação sobre o roteador de saída
  - Deve ser volátil para evitar o envio de dados a rotas “mortas”
  - Tempo de vida - Lifetime
    - 30 min.
- Anúncios (router advertisements) enviados a cada 7 min.
  - Evitar congestionamento da rede
  - Como o período é longo, estações podem enviar solicitações



Escolha do Roteador

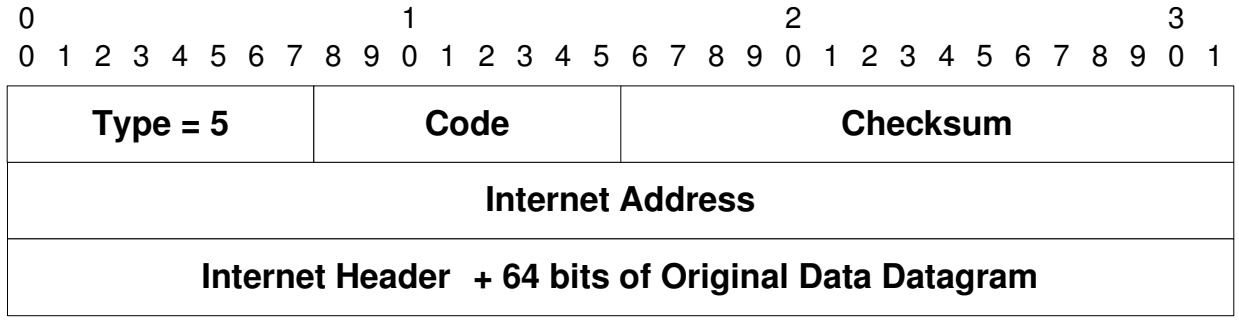
- Router solicitation
  - Enviadas a **`224.0.0.2`** (*"todos os roteadores"*) ou **`255.255.255.255`**
- O roteador envia a resposta
  - à estação, ou
  - a todas as estações, se o momento do anúncio estiver próximo
- Estações podem receber várias respostas
  - Devem considerar apenas os roteadores em sua sub-rede
  - Selecionar o de maior valor de preferência
  - Enviar todo o tráfego para este roteador

Redirecionamento ICMP



- Como evitar que o tráfego destinado a Estação B passe por R1? (e duas vezes no segmento Ethernet 1)

Redirecionamento ICMP



- **Code**

0: redirecionar pacotes para a Rede

1: redirecionar pacotes para a Estação

2: Rede e ToS

3: Estação e ToS

- Primeiro pacote é para B é enviado a R1
- R1 envia uma mensagem ICMP redirect à estação A

- Ao receber o redirect, a estação A deve mudar sua tabela de roteamento
  - Para o endereço contido no campo Internet Header, o próximo salto é dado por Internet Address
- O redirecionamento pode ser para uma rede
  - Indicado no campo código
    - Mas não existe espaço para uma máscara, portanto não é possível redirecionar o tráfego para uma sub-rede