

Protocolos de Roteamento IP

Luiz Carlos Lobato



Protocolos de Roteamento IP

Luiz Carlos Lobato



Protocolos de Roteamento IP

Luiz Carlos Lobato

Rio de Janeiro
Escola Superior de Redes
2013

Copyright © 2013 – Rede Nacional de Ensino e Pesquisa – RNP
Rua Lauro Müller, 116 sala 1103
22290-906 Rio de Janeiro, RJ

Diretor Geral
Nelson Simões

Diretor de Serviços e Soluções
José Luiz Ribeiro Filho

Escola Superior de Redes

Coordenação
Luiz Coelho

Edição
Pedro Sangirardi

Coordenação Acadêmica de Administração e Projeto de Redes
Luiz Carlos Lobato

Equipe ESR (em ordem alfabética)
Celia Maciel, Cristiane Oliveira, Derlinéa Miranda, Edson Kowask, Elimária Barbosa, Lourdes Soncin, Luciana Batista, Luiz Carlos Lobato, Renato Duarte e Sergio de Souza

Capa, projeto visual e diagramação
Tecnodesign

Versão
1.0.0

Este material didático foi elaborado com fins educacionais. Solicitamos que qualquer erro encontrado ou dúvida com relação ao material ou seu uso seja enviado para a equipe de elaboração de conteúdo da Escola Superior de Redes, no e-mail info@esr.rnp.br. A Rede Nacional de Ensino e Pesquisa e os autores não assumem qualquer responsabilidade por eventuais danos ou perdas, a pessoas ou bens, originados do uso deste material.

As marcas registradas mencionadas neste material pertencem aos respectivos titulares.

Distribuição
Escola Superior de Redes

Rua Lauro Müller, 116 – sala 1103
22290-906 Rio de Janeiro, RJ
<http://esr.rnp.br>
info@esr.rnp.br

Dados Internacionais de Catalogação na Publicação (CIP)

L796p Lobato, Luiz Carlos Lobato
 Protocolo de Roteamento IP / Luiz Carlos Lobato. – Rio de Janeiro: RNP/ESR, 2013.
 142 p. : il. ; 28 cm..

Bibliografia: p. 127-129.
ISBN 978-85-63630-24-7

1. Redes de Computadores – Arquitetura. 2. Roteamento IP. 3. Rotas estáticas. 4. Rotas dinâmicas. 5. RIP (Protocolo de Roteamento de informação). 5. OSPF (Open Shortest Path First). 6. BGP (Protocolo de Roteamento Dinâmico). I. Título.

CDD 004.62

Sumário

1. Conceitos básicos de roteamento

Conceito de roteamento 1

Componentes do roteamento 2

Transporte dos pacotes 3

Roteamento IP 3

Tabela de rotas 5

Roteamento estático 7

Roteamento dinâmico 8

Exercício de fixação 1 – Tabela de rotas 9

Configuração básica de roteador 10

Modos de comando 10

Roteiro de Atividades 1 15

Atividade 1.1 – Comandos de configuração básica 15

Atividade 1.2 – Estudo de caso (parte 1) 20

Atividade 1.3 – Estudo de caso (parte 2) 23

2. Protocolo de roteamento RIP

Sistema Autônomo – AS 27

Classless Interdomain Routing (CIDR) 27

Classificação de protocolos de roteamento 28

Roteamento dinâmico 28

Algoritmo de roteamento 30

Tabela de roteamento Vetor-Distância 31



RIPv2 – Características 33

Contagem ao infinito 35

Implementações especiais do RIPv2 36

Pacote RIP 37

Configuração do protocolo RIP 38

Roteiro de Atividades 2 39

Atividade 2.1 – Configuração do protocolo RIP 39

Atividade 2.2 – Atualização de rotas do protocolo RIP 39

Atividade 2.3 – Projeto e configuração do protocolo RIP 40

3. Protocolo de roteamento OSPF

Open Shortest Path First (OSPF) 43

Comparação RIP x OSPF 44

Conceito de Estado do Enlace 45

Algoritmo SPF – Dijkstra 46

Funcionamento do protocolo OSPF 50

OSPF – Roteadores de borda e área 50

Pacotes de Estado de Enlace 52

OSPF – Resumo de funcionamento 53

Autenticação OSPF 54

Backbone OSPF 56

Layout dos pacotes OSPF 57

Roteiro de Atividades 3 61

Atividade 3.1 – Configuração do protocolo OSPF 61

Atividade 3.2 – Projeto e configuração do protocolo OSPF 62

4. Protocolo de roteamento BGP4 – Parte 1

Histórico 65

Border Gateway Protocol (BGP-4) 69

Routing Information Base (RIB) 70

Vizinhos e pares BGP 71

Atributos do BGP 72

Formato do Atributo de Caminho 73



Configuração BGP – roteadores Cisco 81

Configuração BGP – simulador Zebra 82

Roteiro de Atividades 4 85

Atividade 4.1 – Configuração do protocolo BGP 85

Atividade 4.2 – Configuração do protocolo BGP 85

5. Protocolo de roteamento BGP4 – Parte 2

Sessão BGP 87

Mensagens BGP 88

Tipos de mensagens BGP 89

Mensagem Open 89

Mensagem Notification 90

Mensagem Keep-Alive 90

Mensagem Update 91

Mensagem Route-Refresh 92

Mapas de rotas 93

Uso de mapas de rotas 96

Route Reflector 96

Cluster list 98

Pontos de Troca de Tráfego (PTT) 98

Troca de tráfego 99

Estrutura da internet 101

Pontos de acesso 102

Conexão do AS ao PTT 103

Anúncios de rotas 109

Roteiro de Atividades 5 113

Atividade 5.1 – Configuração do protocolo BGP 113

6. Resolução de problemas

Orientações gerais 115

Formação de grupos de trabalho 115

Problema 1 116

Problema 2 117

Problema 3 117

Problema 4 118



Problema 5 **118**

Problema 6 **119**

Apresentação das soluções **119**

Roteiro de Atividades 6 121

Atividade 6.1 – Configuração do protocolo BGP **121**

Atividade 6.2 – Configuração de sub-redes **122**

Atividade 6.3 – Configuração de sub-redes **123**

Atividade 6.4 – Projeto de endereçamento IP **124**

Atividade 6.5 – Configuração de rotas estáticas **125**

Atividade 6.6 – Configuração de OSPF e BGP **126**

Bibliografia 127

Escola Superior de Redes

A Escola Superior de Redes (ESR) é a unidade da Rede Nacional de Ensino e Pesquisa (RNP) responsável pela disseminação do conhecimento em Tecnologias da Informação e Comunicação (TIC).

A ESR nasce com a proposta de ser a formadora e disseminadora de competências em TIC para o corpo técnico-administrativo das universidades federais, escolas técnicas e unidades federais de pesquisa. Sua missão fundamental é realizar a capacitação técnica do corpo funcional das organizações usuárias da RNP, para o exercício de competências aplicáveis ao uso eficaz e eficiente das TIC.

A ESR oferece dezenas de cursos distribuídos nas áreas temáticas: Administração e Projeto de Redes, Administração de Sistemas, Segurança, Mídias de Suporte à Colaboração Digital e Governança de TI.

A ESR também participa de diversos projetos de interesse público, como a elaboração e execução de planos de capacitação para formação de multiplicadores para projetos educacionais como: formação no uso da conferência web para a Universidade Aberta do Brasil (UAB), formação do suporte técnico de laboratórios do Proinfo e criação de um conjunto de cartilhas sobre redes sem fio para o programa Um Computador por Aluno (UCA).

A metodologia da ESR

A filosofia pedagógica e a metodologia que orientam os cursos da ESR são baseadas na aprendizagem como construção do conhecimento por meio da resolução de problemas típicos da realidade do profissional em formação. Os resultados obtidos nos cursos de natureza teórico-prática são otimizados, pois o instrutor, auxiliado pelo material didático, atua não apenas como expositor de conceitos e informações, mas principalmente como orientador do aluno na execução de atividades contextualizadas nas situações do cotidiano profissional.

A aprendizagem é entendida como a resposta do aluno ao desafio de situações-problema semelhantes às encontradas na prática profissional, que são superadas por meio de análise, síntese, julgamento, pensamento crítico e construção de hipóteses para a resolução do problema, em abordagem orientada ao desenvolvimento de competências.

Dessa forma, o instrutor tem participação ativa e dialógica como orientador do aluno para as atividades em laboratório. Até mesmo a apresentação da teoria no início da sessão de aprendizagem não é considerada uma simples exposição de conceitos e informações. O instrutor busca incentivar a participação dos alunos continuamente.

As sessões de aprendizagem onde se dão a apresentação dos conteúdos e a realização das atividades práticas têm formato presencial e essencialmente prático, utilizando técnicas de estudo dirigido individual, trabalho em equipe e práticas orientadas para o contexto de atuação do futuro especialista que se pretende formar.

As sessões de aprendizagem desenvolvem-se em três etapas, com predominância de tempo para as atividades práticas, conforme descrição a seguir:

Primeira etapa: apresentação da teoria e esclarecimento de dúvidas (de 60 a 90 minutos).

O instrutor apresenta, de maneira sintética, os conceitos teóricos correspondentes ao tema da sessão de aprendizagem, com auxílio de slides em formato PowerPoint. O instrutor levanta questões sobre o conteúdo dos slides em vez de apenas apresentá-los, convidando a turma à reflexão e participação. Isso evita que as apresentações sejam monótonas e que o aluno se coloque em posição de passividade, o que reduziria a aprendizagem.

Segunda etapa: atividades práticas de aprendizagem (de 120 a 150 minutos).

Esta etapa é a essência dos cursos da ESR. A maioria das atividades dos cursos é assíncrona e realizada em duplas de alunos, que acompanham o ritmo do roteiro de atividades proposto no livro de apoio. Instrutor e monitor circulam entre as duplas para solucionar dúvidas e oferecer explicações complementares.

Terceira etapa: discussão das atividades realizadas (30 minutos).

O instrutor comenta cada atividade, apresentando uma das soluções possíveis para resolvê-la, devendo ater-se àquelas que geram maior dificuldade e polêmica. Os alunos são convidados a comentar as soluções encontradas e o instrutor retoma tópicos que tenham gerado dúvidas, estimulando a participação dos alunos. O instrutor sempre estimula os alunos a encontrarem soluções alternativas às sugeridas por ele e pelos colegas e, caso existam, a comentá-las.

Sobre o curso

O curso fornece uma visão geral dos conceitos básicos de roteamento e protocolos de roteamento IP. Roteamento estático e dinâmico; protocolos RIP, OSPF e BGP. Ao final do curso, o aluno será capaz de configurar protocolos de roteamento de uma rede TCP/IP e de conectá-la à internet.

A quem se destina

O público-alvo do curso é composto por profissionais de redes (segmento corporativo) e estudantes de informática (formandos em Ciência da Computação/Informática), interessados em obter um maior domínio de protocolos de roteamento da arquitetura TCP/IP, condição fundamental para a formação de especialistas em administração de redes de computadores.

Convenções utilizadas neste livro

As seguintes convenções tipográficas são usadas neste livro:

Itálico

Indica nomes de arquivos e referências bibliográficas relacionadas ao longo do texto.

Indica comandos e suas opções, variáveis e atributos, conteúdo de arquivos e resultado da saída de comandos. Comandos que serão digitados pelo usuário são grifados em negrito e possuem o prefixo do ambiente em uso (no Linux é normalmente # ou \$, enquanto no Windows é C:\).

Conteúdo de slide

Indica o conteúdo dos slides referentes ao curso apresentados em sala de aula.

Símbolo

Indica referência complementar disponível em site ou página na internet.

Símbolo

Indica um documento como referência complementar.

Símbolo

Indica um vídeo como referência complementar.

Símbolo

Indica um arquivo de áudio como referência complementar.

Símbolo

Indica um aviso ou precaução a ser considerada.

Símbolo

Indica questionamentos que estimulam a reflexão ou apresenta conteúdo de apoio ao entendimento do tema em questão.

Símbolo

Indica notas e informações complementares como dicas, sugestões de leitura adicional ou mesmo uma observação.

Permissões de uso

Todos os direitos reservados à RNP.

Agradecemos sempre citar esta fonte quando incluir parte deste livro em outra obra.

Exemplo de citação: LOBATO, Luiz Carlos Lobo. *Roteamento de Protocolos IP*. Rio de Janeiro: Escola Superior de Redes, 2013.

Comentários e perguntas

Para enviar comentários e perguntas sobre esta publicação:

Escola Superior de Redes RNP

Endereço: Av. Lauro Müller 116 sala 1103 – Botafogo

Rio de Janeiro – RJ – 22290-906

E-mail: info@esr.rnp.br

Sobre os autores

Luiz Carlos Lobato é formado em Engenharia Eletrônica pelo ITA, com pós-graduação em Negócios e Serviços de Telecomunicações pelo CEFET-RJ. Possui certificação de redes Cisco CCNA. Gerente da Divisão de Suporte Técnico da Telebrás até a privatização das Telecomunicações, sendo responsável pela operação e gerência da Rede de Dados do Sistema Telebrás. Após a privatização atuou como Coordenador de Cursos de Tecnologia de Redes (Graduação Superior) em diversas faculdades. É colaborador da Escola Superior de Redes desde 2008, tendo elaborado material de treinamento e lecionado diversos cursos na área de Redes. Atualmente é Coordenador Acadêmico de Redes da ESR.

1

Conceitos básicos de roteamento

objetivos

Neste capítulo veremos os conceitos básicos, para que possamos entender o que é roteamento. Para podermos desenvolver os demais tópicos do curso, precisaremos também do entendimento dos princípios de roteamento IP, tabela de rotas e configuração básica de roteador.

Roteamento IP, tabela de rotas, roteamento estático, roteamento dinâmico e configuração básica de roteador.

conceitos

Conceito de roteamento

Roteamento é a transferência de informação da origem até o destino através de uma rede.

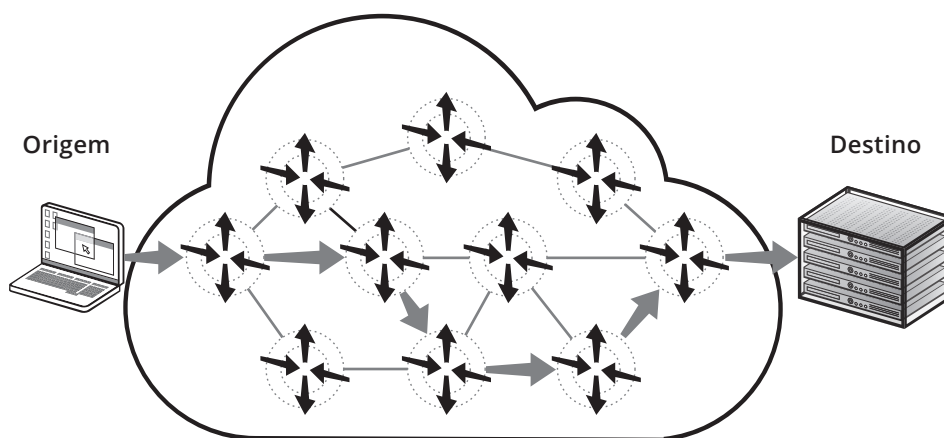


Figura 1.1
Conceito de roteamento.

Roteamento é a transferência de informação da fonte até o destino através de uma rede. Ao longo do caminho, tipicamente teremos pelo menos um nó intermediário. De acordo com essa definição, a função do roteador parece ser a mesma que a de uma ponte (switch/bridge). A principal diferença entre ambos é que a ponte opera na camada 2 (enlace de dados) do modelo OSI, enquanto os roteadores operam na camada 3 (rede). Assim, eles operam de maneiras diferentes, embora ambos executem operações de comutação.

Componentes do roteamento

- Determinação de rotas.
- Transporte dos pacotes (comutação).

Determinação de rotas.

- Métrica.
- Tabelas de roteamento.
- Troca de mensagens.



Para chegar à rede	Enviar para
10	Nó A
15	Nó B
20	Nó C
30	Nó A
25	Nó B
45	Nó A

Figura 1.2
Exemplo de
tabela de rotas.

O roteamento envolve duas atividades básicas:

- Determinação das rotas ótimas.
- Transporte da informação (pacotes) através da rede (processo de comutação – switching).

Os algoritmos de roteamento usam algum padrão de medida (chamado de métrica) para determinar a rota ótima para um dado destino. Para ajudar no processo de determinação de rotas, os algoritmos de roteamento inicializam e mantêm tabelas de roteamento, que contêm informações de rotas. Essas informações tipicamente são armazenadas no formato destino/próximo nó (destination/next hop). A tabela mostrada na figura anterior exemplifica o que foi dito.

Os roteadores se comunicam entre si, para terem conhecimento de seus vizinhos e manterem atualizadas as tabelas de rotas. A internet é uma rede em constante mudança e não pode parar; desse modo, as mudanças precisam ser feitas dinamicamente. Para isso, os roteadores trocam mensagens para a manutenção das tabelas.

Transporte dos pacotes

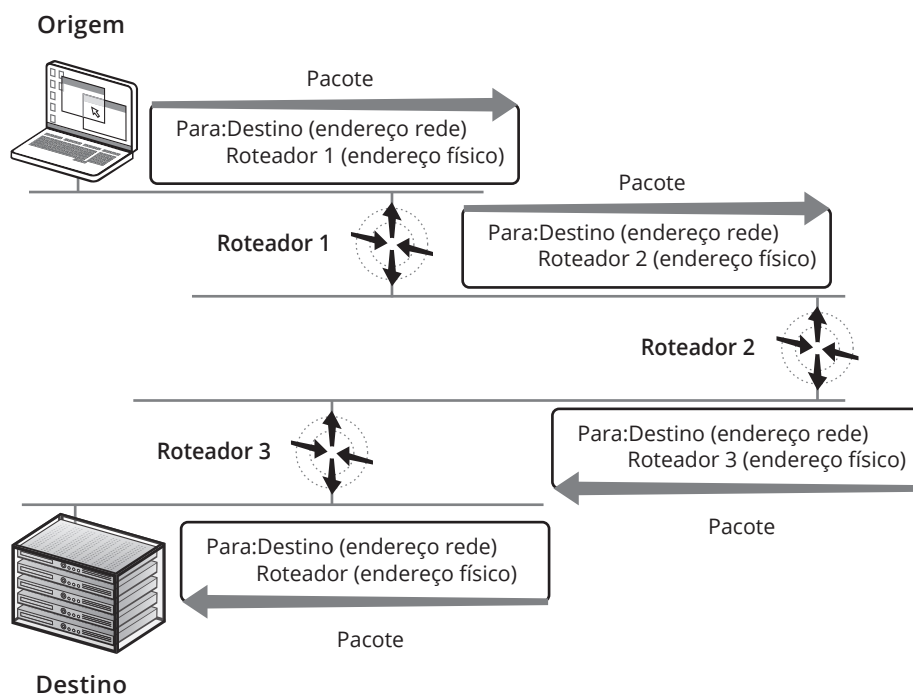


Figura 1.3
Transporte
de pacotes.

Algoritmos de comutação são relativamente simples e basicamente os mesmos para a maioria dos protocolos de roteamento. Tipicamente, um host determina que precisa enviar um pacote para outro host. Para isso, ele tem de saber, de alguma forma, o endereço do roteador que fará a ação (se não souber, não há como enviar o pacote).

O host envia o pacote para o roteador, colocando o endereço físico do roteador (normalmente estão na mesma rede local, portanto o endereço físico será o MAC address) e o endereço do protocolo de rede do host de destino. O roteador então examina o pacote e tenta encaminhá-lo para o host de destino, baseado no seu endereço de rede. Se o roteador tiver na sua tabela de rotas a rota adequada, ele encaminhará para o próximo nó, mudando o endereço físico para o endereço do próximo nó e mantendo o endereço de rede do host de destino. Se não tiver a rota na tabela, o roteador simplesmente descartará o pacote.

E o processo se repetirá até chegar ao roteador que está na mesma rede do host de destino, que entregará o pacote enviando-o para o endereço físico do host de destino. Assim, à medida que o pacote atravessa a rede, seu endereço físico vai mudando; porém, o endereço do protocolo de rede permanece igual (host de destino).

Roteamento IP

- Diretamente conectado.
- Gateway padrão.
- Configuração do host IP.
 - ▣ Somente como host.
 - ▣ Como host e roteador.



Conceitualmente, o roteamento do IP é bastante simples para um host. Se o destino estiver diretamente conectado ao host (enlace ponto-a-ponto) ou numa rede Ethernet compartilhada,

o datagrama IP é enviado diretamente para o destino. Caso contrário, o host envia o datagrama para um default router (gateway padrão) e deixa o roteador entregar o datagrama no seu destino.

O host poderá ser configurado para atuar como host ou como host e roteador. Se o host estiver configurado para atuar como um roteador, ele poderá encaminhar datagramas de uma de suas interfaces de rede para outra. Se não estiver configurado como roteador, ele só poderá encaminhar datagramas gerados pelas camadas superiores do protocolo nele residente (TCP, UDP, ICMP ou IGRP), não podendo encaminhar datagramas recebidos de suas interfaces de rede.

O IP pesquisa uma tabela de roteamento na memória do host cada vez que ele recebe um datagrama de uma interface de rede para enviar. Os seguintes procedimentos serão executados:

- Primeiro o IP verifica se o endereço IP de destino é o seu próprio ou se é um endereço IP broadcasting; se for esse o caso, ele entrega o datagrama para o protocolo especificado no campo *protocolo* do cabeçalho do datagrama.
- Se o datagrama não se destina a ele, o IP verifica a sua configuração de host/router:
 1. Se ele estiver configurado como router, executará os procedimentos de roteamento IP baseado na sua tabela de roteamento residente na memória do host.
 2. Se ele não estiver configurado como router, o datagrama será simplesmente descartado.

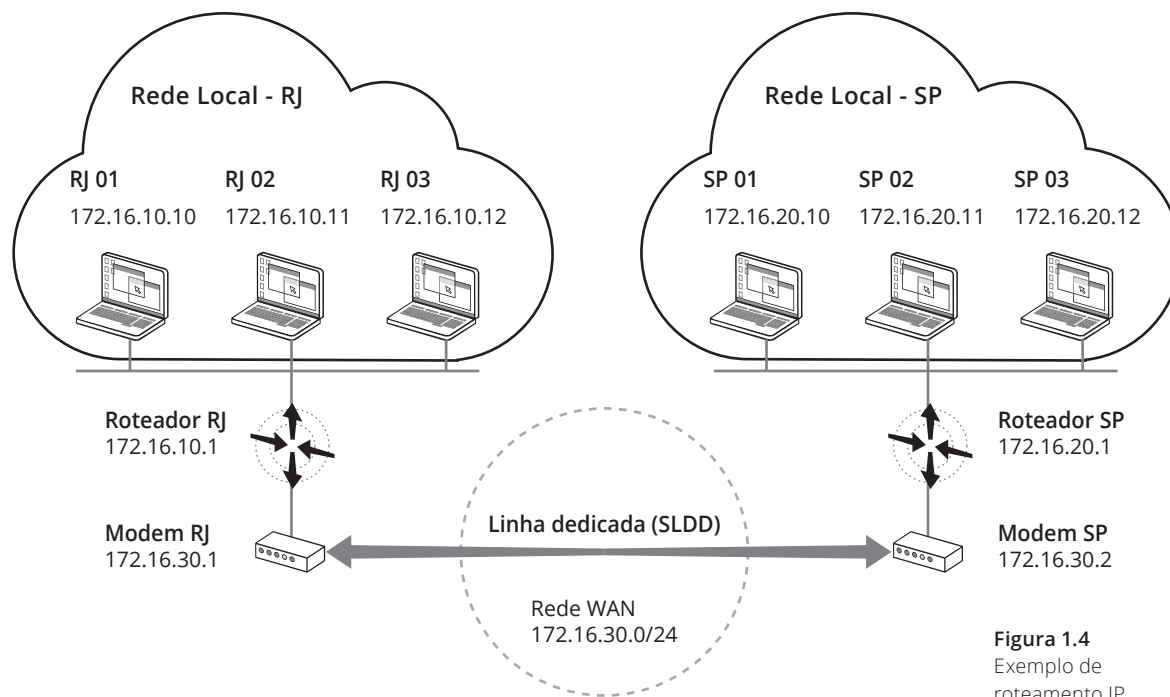


Figura 1.4
Exemplo de roteamento IP.

- Rede local RJ 172.16.10.0/24.
- Rede local SP 172.16.20.0/24.

Sejam as duas redes locais da figura 1.4, uma no Rio de Janeiro e outra em São Paulo. A rede local do Rio de Janeiro usa o endereço de rede 172.16.10.0/24 e a de São Paulo usa o endereço de rede 172.16.20.0/24.

Os respectivos roteadores usam na interface diretamente conectada às redes (interface Ethernet E0) um endereço válido de cada uma delas; no caso, no Rio de Janeiro o endereço 172.16.10.1 e, em São Paulo, o endereço 172.16.20.1. Esses endereços serão os gateways padrão das respectivas redes, tendo de ser configurados em todos os hosts das duas redes.

Para se comunicarem entre si, os roteadores usam uma linha dedicada conectada a uma interface serial (S0). Os endereços dessas interfaces têm de ser diferentes dos endereços das interfaces Ethernet, ou, em outras palavras, têm de ser de outra rede, mesmo porque essa linha dedicada é também uma rede física e nós já vimos que cada rede física tem de ter um prefixo de rede diferente.

Assim, os roteadores se comunicam através da rede 172.16.30.0/24, sendo que a interface serial do roteador Rio de Janeiro tem o endereço 172.16.30.1 e a de São Paulo, o endereço 172.16.30.2. Dessa forma, a rede 172.16.30.0/24 é uma “ponte” entre as duas redes locais.

Suponha que a máquina RJ 01 tenha de enviar um pacote para a máquina SP 03. Os respectivos endereços de origem e destino serão:

- Origem: 172.16.10.10.
- Destino: 172.16.20.22.

A máquina RJ01 conclui que o endereço de destino não é da rede dela e, nesse caso, envia para o gateway padrão porque o host não foi configurado como roteador. Trata-se de uma entrega indireta. Ao chegar ao roteador RJ (via interface 172.16.10.1), o roteador consulta sua tabela de rotas para saber como despachar o pacote. A sua tabela de rotas informa que, para chegar à rede de destino (172.16.20.0/24), ele precisa enviar o pacote para o roteador de SP no endereço 172.16.30.2 (nexthop), via interface serial que tem o endereço 172.16.30.1. E assim ele o faz.

O roteador de São Paulo consulta sua tabela de rotas e verifica que está diretamente conectado à rede de destino, logo ele entrega o pacote ao host 172.16.20.22 via interface 172.16.20.1.

Tabela de rotas

- Tabela com as rotas conhecidas do roteador.
- Formato padrão.
 - ▣ Identificação da rede de destino.
 - ▣ Máscara de sub-rede.
 - ▣ Gateway (nexthop).
 - ▣ Métrica.
 - ▣ Outras informações (depende do protocolo).

As rotas podem ser aprendidas através de:

- Administrador de rede.
- Protocolos de roteamento.

Quando um pacote chega a uma das interfaces do roteador, ele analisa a sua tabela de roteamento para verificar se nela existe uma rota para a rede de destino. Pode ser uma rota direta ou a indicação do roteador para o qual o pacote deve ser enviado. Esse processo continua até que o pacote seja entregue na rede de destino.



As informações da tabela de roteamento devem ser suficientes para que o roteador possa fazer isso. O formato padrão de uma entrada na tabela de roteamento é:

```
<Network id><Subnet mask><Gateway><Metric><outras informações>
```

Se a rede de destino não estiver na tabela, o datagrama será descartado sumariamente.

Para montar essa tabela, o roteador pode “aprender” as rotas de duas maneiras:

- Administrador de rede.
- Protocolos de roteamento.

A primeira maneira é manual e a segunda é automática. Mais adiante veremos em que situações elas se aplicam melhor.

Exemplo de tabela de rotas

```
C:\>route print

=====
=====

Lista de interfaces

0x1 ..... MS TCP Loopback interface
0x2 ...00 60 67 01 d3 06 ... Acer ALN-330 10/100M PCI Fast Ethernet
Adapter

=====
=====

Rotas ativas:

Endereço de rede      Máscara Ender.gateway  Interface Custo
0.0.0.0                0.0.0.0      189.6.12.1189.6.12.158    1
127.0.0.0              255.0.0.0      127.0.0.1  127.0.0.1    1
189.6.12.0              255.255.252.0189.6.12.158 189.6.12.158    1
189.6.12.158 255.255.255.255 127.0.0.1    127.0.0.1    1
189.6.255.255255.255.255.255 189.6.12.158 189.6.12.158    1
224.0.0.0              224.0.0.0      189.6.12.158189.6.12.158    1
255.255.255.255255.255.255.255 189.6.12.158 189.6.12.158    1

Gateway padrão:189.6.12.1

=====
=====

Rotas persistentes: Nenhuma

C:\>
```

Figura 1.5
Tabela de rotas
do Windows.

O comando *route print* do DOS lista a tabela de rotas atual aprendida pelo Windows.

Na primeira parte temos a lista de interfaces de rede atualmente ativas: a loopback (teste interno) e, no caso, uma interface Ethernet. Depois vêm as rotas ativas. Veja, por exemplo, a primeira entrada:

0.0.0.0	0.0.0.0	189.6.12.1	189.6.12.158	1
---------	---------	------------	--------------	---

Essa entrada é a chamada de rota padrão. Essa rota é indicada por uma identificação de rede 0.0.0.0 com uma máscara de sub-rede 0.0.0.0. Quando o TCP/IP tenta encontrar uma rota para um determinado destino, ele percorre todas as entradas da tabela de roteamento em busca de uma rota específica para a rede de destino. Caso não seja encontrada uma rota para a rede de destino, será utilizada a rota padrão. Em outras palavras, se não houver uma rota específica, envie através da rota padrão.

Observe que a rota padrão é justamente o default gateway da rede (189.6.12.1), ou seja, a interface de LAN do roteador da rede. O parâmetro Interface (189.6.12.158) é o número IP da placa de rede do próprio computador. Não havendo uma rota específica, deve-se enviar para a rota padrão, onde o próximo hop da rede deverá ser o 189.6.12.1, e o envio para esse hop é feito através da interface 189.6.12.158 (ou seja, a própria placa de rede do computador).

A próxima entrada define o endereço de loopback que, como já dissemos, é usado para a finalidade de testes internos.

A terceira entrada define a rota para a rede 189.6.12.0/22:

189.6.12.0	255.255.252.0	189.6.12.158	189.6.12.158	1
------------	---------------	--------------	--------------	---

Essa rota é conhecida como rota da rede local. Ela basicamente diz o seguinte: “Quando o endereço IP de destino for um endereço da minha rede local, envie as informações através da minha placa de rede” (observe que tanto o parâmetro gateway como o parâmetro Interface estão configurados com o número IP do próprio computador). Ou seja, “se for para uma das máquinas da minha rede local, envie através da placa de rede, não precisa enviar para o roteador”. É o caso de uma entrega direta.

As demais entradas não são relevantes para nosso estudo.

Quando um roteador é configurado com os endereços IP de cada interface, ele só pode enviar pacotes IP para as redes às quais está diretamente conectado. Se ele receber um pacote destinado a uma rede remota que não está na tabela de roteamento, ele simplesmente descarta o pacote (não envia em nenhuma hipótese um broadcasting para localizar a rede remota).

Para que o roteador seja capaz de enviar pacotes para redes remotas, é necessário configurar as rotas.

Podem ser usados os seguintes métodos:

- Roteamento estático.
- Roteamento dinâmico.

Roteamento estático

Vantagens:

- Sem overhead na CPU do roteador.
- Roteadores não usam a largura de banda.
- Segurança (administrador define as rotas).



Desvantagens:

- Exige maior conhecimento técnico.
- Cada mudança de configuração deve ser feita em todos os roteadores da rede.
- Inviável em grandes redes.

Nesse método, o administrador da rede configura manualmente todas as rotas em cada roteador da rede. Em redes pequenas é até relativamente simples, como veremos em nosso exemplo mais adiante. Porém, em redes grandes, esse procedimento é inviável, por causa do tempo necessário para atualizar todas as tabelas em todos os roteadores da rede a cada mudança de topologia (seja por adição de novo hardware ou por falha de algum componente). Suas vantagens são principalmente simplicidade, segurança e menor overhead de CPU do roteador e de largura de banda da rede.

Roteamento dinâmico

Vantagens:

- Configuração mais fácil que a da rota estática.
- Atualizações dinâmicas pelos roteadores.
- Usado em redes grandes.

Desvantagens:

- Overhead na CPU do roteador.
- Roteadores usam a largura de banda.

Esse método é usado normalmente em grandes redes, porque permite que os próprios roteadores construam e atualizem suas tabelas de roteamento, através de protocolos de roteamento: IPX (só em redes Novell), RIP, IGRP, OSPF etc. É mais simples de configurar do que rotas estáticas, porém à custa da CPU dos roteadores e da largura de banda da rede.

Para rotear pacotes, o roteador precisa conhecer:

- Endereço de destino.
- Roteadores vizinhos dos quais possa aprender rotas para as redes remotas.
- Rotas possíveis para todas as redes remotas.
- A melhor rota para cada rede remota.
- Como manter e verificar a informação das rotas.

Roteamento dinâmico é o processo pelo qual protocolos de roteamento executados no roteador se comunicam com os roteadores vizinhos. Os roteadores trocam informações entre si a respeito de todas as redes para as quais eles conhecem as rotas.

O roteador faz o roteamento do tráfego para todas as redes interconectadas. O roteador aprende as rotas para as redes remotas através dos roteadores vizinhos ou do administrador da rede. O roteador então constrói a tabela de roteamento, que descreve a forma de achar as redes remotas.

Se a rede estiver diretamente conectada a uma interface do roteador, então o roteador já sabe como chegar a ela. Se as redes não estiverem diretamente conectadas, o roteador precisará aprender como chegar a elas, seja através de rotas estáticas configuradas pelo administrador ou através de rotas dinâmicas aprendidas dos roteadores vizinhos.

Se ocorrer uma mudança de topologia na rede, os protocolos de roteamento dinâmico automaticamente informam todos os roteadores a respeito da mudança. Se, por outro lado, forem usadas rotas estáticas, é responsabilidade do administrador de rede atualizar as rotas em todos os roteadores da rede.

Exercício de fixação 1

Tabela de rotas

Considerando a rede da Figura 1.4, exemplo de roteamento IP, uma possível tabela de rotas (parcial) para as estações SP01 e RJ01 seria:

```
C:\>route print
```

Estacao RJ01

Rotas ativas:

Endereço de rede	Máscara	Ender. gateway	Interface	Custo
0.0.0.0	0.0.0.0	172.16.10.1	172.16.10.10	1
172.16.10.0	255.255.255.0	172.16.10.10	172.16.10.10	1
...				

```
C:\>route print
```

Estacao SP01

Rotas ativas:

Endereço de rede	Máscara	Ender. gateway	Interface	Custo
0.0.0.0	0.0.0.0	172.16.20.1	172.16.20.10	1
172.16.20.0	255.255.255.0	172.16.20.10	172.16.20.10	1
...				

Explique cada uma das entradas na tabela mostradas anteriormente.

Configuração básica de roteador

- Comandos de configuração variam de um fabricante para outro.
- Padrão de mercado: IOS, da Cisco.
- Interface de configuração – Linha de Comando (CLI).
 - Modos de comando.
 - Indicado pelo prompt do roteador.



Os comandos de configuração de roteador dependem do sistema operacional do fabricante do roteador. Os comandos que serão aqui mostrados são aqueles usados pelo simulador Core no console dos roteadores virtuais das redes que serão usadas nas atividades práticas. Esses comandos são baseados no sistema operacional IOS, da Cisco, e constituem um padrão de mercado de fato.

A interface de configuração do roteador é baseada na interface de linha de comando – Command-Line Interface (CLI). A CLI é dividida em vários modos de comando diferentes. Cada modo de comando tem seu próprio conjunto de comandos disponível para configuração, manutenção e monitoramento do roteador e das operações de rede. Os comandos disponíveis dependem do modo de comando em que você está. O modo de comando é indicado pelo prompt do roteador, como veremos adiante.

Modos de comando

Hierarquia para navegação entre os modos de comando:

- user EXEC mode
- privileged EXEC mode
- global configuration mode
- specific configuration modes
- configuration submodes
- configuration subsubmodes



Digitando o caractere “?” (interrogação, sem as aspas) você obterá a lista de comandos possíveis no modo de comando em que você está no momento. A hierarquia básica para navegação entre os modos de comando é:

```
user EXEC mode-> privileged EXEC mode-> global configuration
mode-> specific configuration modes-> configuration submodes->
configuration subsubmodes.
```

user EXEC mode

- *Router>*.
- Modo de início da sessão.
- Não permite configuração.
- *Router>enable* vai para o modo privilegiado.

privileged EXEC mode

- *Router#*.
- Comandos modo EXEC e CONFIGURE.
- *Router# conf t* vai para o modo de configuração global.



Quando você inicia uma sessão em um roteador, você inicia no modo *user EXEC mode*, que não permite configurar o roteador, apenas verificar o status, nada mais. Em nosso simulador esse modo não existe, mesmo porque não possui utilidade para os nossos propósitos. Para ter acesso aos comandos de configuração, você precisa navegar para o modo seguinte, que é o *privileged EXEC mode*, no qual você pode digitar comandos tipo EXEC, por exemplo, ou o comando *show*, que mostra o status da configuração corrente do roteador.

❗ Os comandos tipo EXEC não são salvos quando você dá reboot no roteador.

O modo seguinte é o *global configuration mode*, que permite a configuração das características gerais do roteador, fazendo modificações na configuração corrente do roteador (running configuration). Se você salvar a configuração modificada, os comandos terão efeito no próximo reboot do roteador.

❗ É preciso salvar a configuração para que as modificações sejam permanentes.

A partir do modo de configuração global, você pode entrar nos modos específicos de configuração e em seus submodos.

Modo privilegiado (privileged EXEC mode)

O modo privilegiado de comando inclui os comandos do modo EXEC e os do modo CONFIGURE, através do qual você pode acessar os demais modos de comando. O prompt do modo privilegiado de comando é (assumindo que o nome do roteador é *Router*):

```
Router#
```

Para retornar ao modo de usuário, digite o comando:

```
Router#disable  
Router>
```

O prompt do modo de usuário foi mostrado anteriormente.

Modo de configuração global (global configuration mode)

- *Router(config)# ip route*
- Permite a configuração global do roteador.
- A configuração corrente (running-config) é atualizada.
- Salvar a configuração corrente:
 - *copyrun star*
- Encerrar esse modo de configuração:
 - Comandos *end*, *exit* e *Ctrl + Z*.



O termo “global” é usado para indicar características que afetam o sistema como um todo. Esse modo é usado para configurar o roteador globalmente ou para entrar em modos específicos de configuração que servem para configurar certos elementos, tais como interfaces ou protocolos.



O comando privilegiado para entrar no modo de configuração global é:

```
Router# configure terminal
Enter configuration commands, one per line. Endwith CNTL/Z.
Router(config)#
```

Note que o prompt mudou, indicando que você está no modo de configuração global. Os comandos executados nesse modo atualizam a configuração corrente (running configuration) em tempo real. Porém, é importante notar que essa configuração será perdida no próximo reboot.

Para salvar a configuração corrente, use o comando:

```
copy running-config startup-config
```

Para encerrar o modo de configuração global, existem três opções de comandos:

```
Router(config)# end
Router(config)# exit
Router(config)# ^Z(Ctrl-z)
```

Modo de configuração de interface (interface configuration mode)

specific configuration modes

- Router(config)# int ser0 <Return>
- Router(config-if)# ip address 192.168.1.1 255.255.255.0
- Router(config-if)# no shut
- Router(config)# router rip
- Router(config-router)# network

Esse modo é um exemplo de modo de configuração específico. O comando é:

```
Router(config)# interface serial 0 <Return>
Router(config-if)#
```

Note que o prompt mudou, indicando que você está no modo de configuração de interface. Um dos comandos mais usados, que podem ser digitados agora, é o de atribuição de endereço IP:

```
Router(config-if)#ip address 192.168.1.1 255.255.255.0
Router(config-if)#no shutdown
```

A interface serial 0 recebeu o endereço IP acima e a máscara é a padrão classe C (/24). O comando seguinte é para habilitar a interface, porque todas as interfaces de um roteador que não estão configuradas ficam no estado shutdown por default. É preciso negar esse comando para “levantar” a interface. É que não existe um comando específico para “levantar” a interface.

Exemplos de comandos de roteamento:

```
Router(config)#router rip
Router(config-router)#network 192.168.1.0
Router(config-router)#network 10.0.0.0
```



Os comandos acima habilitam o protocolo de roteamento RIP e definem as redes que serão anunciadas pelo roteador para os seus vizinhos. Note a mudança do prompt.

```
Router(config)#iproute 192.168.0.0 255.255.255.0 192.168.1.1
```

O comando acima cria uma rota estática para a rede 192.168.0.0/24, apontando para o próximo salto (next hop) 192.168.1.1.

Comandos de verificação e diagnóstico

- Router(config)# show ?
- Router(config)# show arp
- Router(config)# shint
- Router(config)# shipint brief
- Router(config)# sh run
- Router(config)# sh star



```
Router#show ?
```

O comando *show ?* fornece uma lista dos comandos *show* disponíveis.

```
Router#show arp
```

Exibe a tabela ARP do roteador.

```
Router#sh interfaces
```

Verifica detalhadamente as configurações das interfaces.

```
Router#ship interface brief
```

Verifica resumidamente as configurações das interfaces.

```
Router#shiproute
```

Verifica a tabela de roteamento.

```
Router#shrrunning-config
```

Verifica as configurações ativas na memória RAM do roteador.

```
Router#sh startup-config
```

Verifica as configurações da NVRAM. Essas configurações serão carregadas na memória RAM do roteador no próximo reboot.

Todos os comandos, salvo ambiguidade, aceitam uma forma resumida:

sh no lugar de *show*; *conf t* no lugar de *configure terminal*; *shint* no lugar de *show interface* etc.

A seguir um resumo dos modos de comando disponíveis na CLI.

Prefixo da CLI	Significado
Router>	Modo usuário.
Router#	Modo privilegiado.
Router(config)#	Modo de configuração global.
Router(config-if)#	Modo de configuração de interface.

Figura 1.6
Modos de comando
da CLI.

Prefixo da CLI	Significado
Router(config-subif)#	Modo de configuração de subinterface.
Router(config-line)#	Modo de configuração de linha (ex.: auxiliar, console e telnet).
Router(config-router)#	Modo de configuração de protocolo de roteamento.

Correção de comandos CLI

Os comandos digitados no modo console (CLI) não podem ser editados porque o roteador não tem interface gráfica que permita isso. Se o erro for detectado durante a digitação e antes de apertar a tecla <Enter>, basta retornar até a posição errada, usando a seta para a esquerda, apagar o(s) caracteres(s) errado(s) com a tecla <backspace> (<-->) e digitar o caractere correto. Não use a tecla <del>.

Se o erro for detectado depois de apertar a tecla <Enter>, o comando errado já está registrado na memória RAM do roteador, mais precisamente no arquivo “running-config”. Você pode verificar isso usando o comando *sh run* (forma reduzida de *show running-config*).

Nesse caso, não basta simplesmente digitar o comando correto que ele não irá apagar o comando errado anteriormente digitado. O que vai acontecer é ficarem na memória RAM os dois comandos. É preciso repetir o comando errado inserindo antes dele a palavra “no” (negar o comando).

Exemplo:

Suponha que você digitou:

```
Router(config-if)# ip address 19.168.1.1 255.255.255.0
```

e o comando correto era:

```
Router(config-if)# ip address 192.168.1.1 255.255.255.0
```

O procedimento para correção é o seguinte:

```
Router(config-if)# no ip address 19.168.1.1 255.255.255.0
Router(config-if)# ip address 192.168.1.1 255.255.255.0
```

Observe que, para negar o comando, é preciso redigitá-lo exatamente como foi digitado antes, com a palavra “no” na frente.



Roteiro de Atividades 1

Atividade 1.1 – Comandos de configuração básica

Vamos usar a rede da Figura 1.4 para configurar rotas estáticas. Siga o procedimento:

1. Inicie o VMWare Player e selecione a opção *Open a Virtual Machine*. Selecione a máquina virtual vcore-4.2 no diretório que o instrutor indicar e a inicie.
2. Aguarde a carga completa da máquina virtual Core.
3. Selecione *File* no menu suspenso, selecione a opção *Open* e localize o diretório onde se encontra a rede: *Rede1_Sessao1_ADR10.imn*, seguindo a orientação do instrutor.
4. A rede deverá ser idêntica à Figura 1.7, mostrada a seguir.

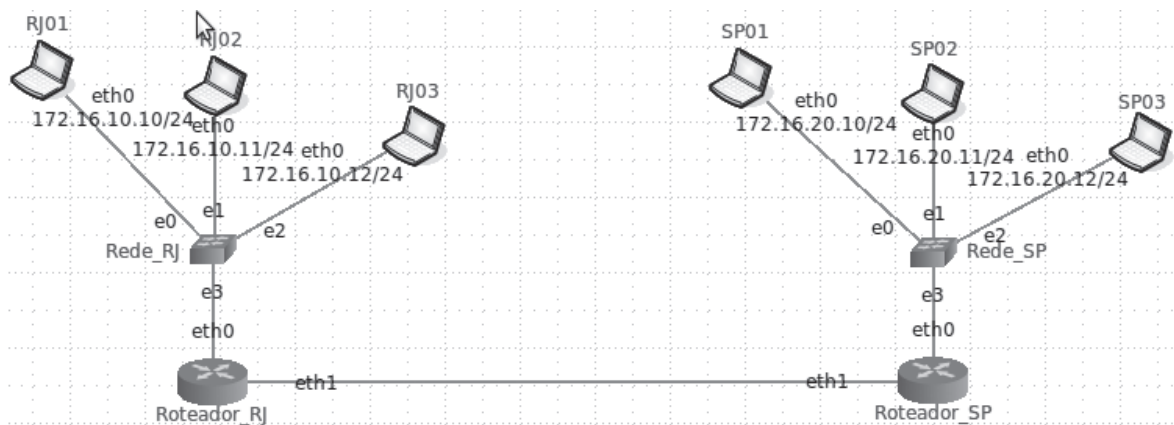


Figura 1.7
Rede usada
no exercício.

5. Os endereços IPv4 dos PCs já estão configurados, conforme mostrado na figura. Temos três redes físicas representadas pelos switches Rede_RJ, Rede_SP e o enlace serial entre os dois roteadores. Cada rede física tem um prefixo de rede diferente, conforme a tabela a seguir.

Rede	Endereço de rede	Gateway padrão
Rede_RJ	172.16.10.0/24	172.16.10.1
Rede_SP	172.16.20.0/24	172.16.20.1
Enlace serial	172.16.30.0/24	–

Figura 1.8
Endereços IPv4 da
rede da Figura 1.7.

6. Nenhum tipo de rota ou protocolo de roteamento está configurado, bem como nenhuma das interfaces dos dois roteadores estão configuradas. Nesta atividade vamos configurar completamente essa rede, de forma a deixá-la pronta para uso.

O modo inicial de operação do simulador é o Modo de Edição. Esse modo é utilizado para desenhar a rede e configurar o endereçamento IPv4. Para efetivamente executar os protocolos de roteamento e as aplicações, é necessário iniciar o experimento, que pode ser feito de duas maneiras:

- Clicando no ícone  à esquerda na barra de ferramentas.
- Selecionando no menu superior a opção *Experiment/Start*.



No Modo de Experimento, não é possível fazer edição da topologia da rede (note a barra de ferramentas modificada). Aguarde até que toda a rede seja iniciada (até desaparecerem os colchetes vermelho/verde em cada nó da rede).

7. Vamos configurar primeiro o roteador do Rio de Janeiro. Para configurar o Roteador_RJ é preciso abrir o console do roteador. Neste simulador o procedimento é:

- Aponte com o mouse para o roteador, clique no botão direito e selecione a opção: *Shell window/vtysh*, conforme mostrado na Figura 1.9.

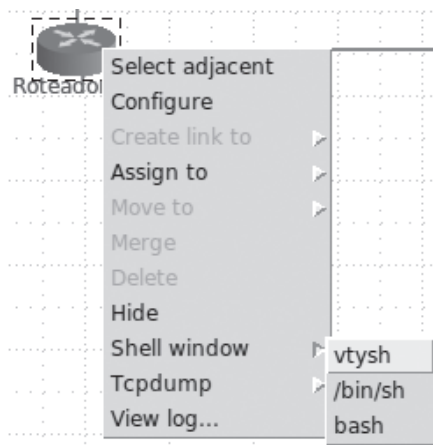


Figura 1.9
Abertura console
do Roteador_RJ.

8. O console deverá ser parecido com o da Figura 1.10.

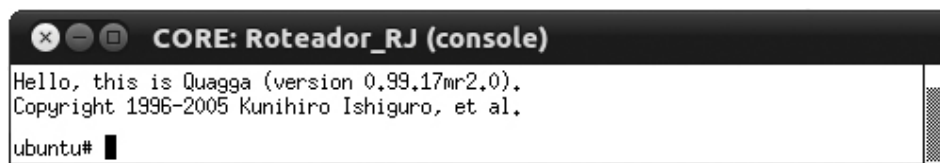


Figura 1.10
Console do
Roteador_RJ.

Se digitarmos o comando *shrun*, que lista a configuração corrente (running-config) do roteador, veremos o seguinte resultado:

```
ubuntu# sh run

Building configuration...

Current configuration:

!
interface eth0
ipv6nd suppress-ra
!
interface eth1
ipv6nd suppress-ra
!
interface lo
!
```

```

ip forwarding
ipv6 forwarding
!
linevty
!
end
ubuntu#

```

Observe que nenhuma interface (eth0 e eth1) foi configurada. A configuração das interfaces dos dois roteadores será feita de acordo com as informações a seguir.

Figura 1.11
Endereços IPv4
das interfaces dos
roteadores.

Roteador	Interface eth0	Interface eth1
Roteador_RJ	172.16.10.1/24	172.16.30.1
Roteador_SP	172.16.20.1/24	172.16.30.2

9. Para configurar as duas interfaces do Roteador_RJ, digite os comandos:

```

ubuntu# conf t
ubuntu(config)# int eth0
ubuntu(config-if)# ip address 172.16.10.1/24
ubuntu(config-if)# no shut
ubuntu(config-if)# int eth1
ubuntu(config-if)# ip address 172.16.30.1/24
ubuntu(config-if)# no shut
ubuntu(config-if)# ^Z
ubuntu#

```

O comando que atribui o endereço IP às interfaces tem uma sintaxe um pouco diferente daquela mostrada nos exemplos anteriores (a máscara é informada em contagem de bits). Note também que a configuração foi encerrada com um *Ctrl* + *Z*. Para verificar a configuração do roteador, digite o comando *sh run*, conforme mostrado na listagem a seguir.

```

ubuntu# sh run
Building configuration...
Current configuration:
!
interface eth0
ip address 172.16.10.1/24
ipv6nd suppress-ra
!

```

```

interface eth1
ip address 172.16.30.1/24
ipv6nd suppress-ra
!
interface lo
!
ip forwarding
ipv6 forwarding
!
linevty
!
end
ubuntu#

```

10. Vamos fazer a configuração do Roteador_SP, que é idêntica à do Roteador_RJ, que pode ser usada como referência. Anote os comandos necessários abaixo.

11. Feitas as configurações dos roteadores, vamos testar a comunicação entre as duas redes locais fazendo um ping do RJ01 para o SP03, por exemplo. Para isso, siga o seguinte procedimento:

- Aponte com o mouse para o RJ01, clique no botão direito e selecione a opção:
Shell window/bash.

Na console do RJ01, digite o seguinte comando (ping para o SP03):

```
root@RJ01:/tmp/pycore.38136/RJ01.conf# ping -c 2 172.16.20.12
```

O resultado deve ser semelhante ao listado a seguir.

```

PING 172.16.20.12 (172.16.20.12) 56(84) bytes of data.
From 172.16.10.1 icmp_seq=1 Destination Net Unreachable
From 172.16.10.1 icmp_seq=2 Destination Net Unreachable

--- 172.16.20.12 ping statistics ---
2 packets transmitted, 0 received, +2 errors, 100% packet loss,
time 1007ms

root@RJ01:/tmp/pycore.38136/RJ01.conf#

```

Por que não funcionou, se os roteadores foram configurados corretamente?

Responda no espaço a seguir.

12. Vamos precisar “ensinar” ao Roteador_RJ como encaminhar pacotes para a rede 172.16.20.0/24, que fica em SP e que só pode ser alcançada através do Roteador_SP. Para isso, os seguintes comandos devem ser digitados no console de Roteador_RJ:

```
ubuntu#conf t
ubuntu(config)#iproute 172.16.20.0/24 172.16.30.2
ubuntu(config)# ^Z
ubuntu#
```

Esse comando define uma rota estática que “ensina” ao Roteador_RJ como encaminhar pacotes para a rede 172.16.20.0/24, enviando-os para o próximo salto (next-hop), que é a interface eth1 do Roteador_SP (endereço IP 172.16.30.2).

13. Vamos fazer a mesma coisa para o Roteador_SP. Anote os comandos no espaço abaixo.

14. Agora, com as rotas configuradas, o *ping* deve funcionar, conforme mostrado a seguir.

```
root@RJ01:/tmp/pycore.38136/RJ01.conf# ping -c 2 172.16.20.12
PING 172.16.20.12 (172.16.20.12) 56(84) bytes of data.
64 bytes from 172.16.20.12: icmp_req=1 ttl=62 time=11.8 ms
64 bytes from 172.16.20.12: icmp_req=2 ttl=62 time=3.24 ms

--- 172.16.20.12 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1007ms
Rtt min/avg/max/mdev = 3.243/7.550/11.858/4.308 ms
root@RJ01:/tmp/pycore.38136/RJ01.conf#
```

15. Para confirmar que a rota foi configurada corretamente, basta digitar o comando abaixo, no console do Roteador_RJ.

```
ubuntu# sh ip route

Codes: K-kernel route, C-connected, S-static, R-RIP, O-OSPF
        I-ISIS, B-BGP, >-selected route, *-FIB route, o-OSPFv3
```



```

C>* 127.0.0.0/8 is directly connected, lo
C>* 172.16.10.0/24 is directly connected, eth0
S>* 172.16.20.0/24 [1/0] via 172.16.30.2, eth1
C>* 172.16.30.0/24 is directly connected, eth1
ubuntu#

```

Observe que a rota estática aparece com um S>* no início da linha e as redes diretamente conectadas, com um "C>".

Atividade 1.2 – Estudo de caso (parte 1)

- Planejamento de endereçamento IP.
- Escritório central.
- Quatro redes departamentais; um backbone corporativo.
- Disponível uma Classe C: 200.248.228.0/24.
- 20-30 computadores/rede.

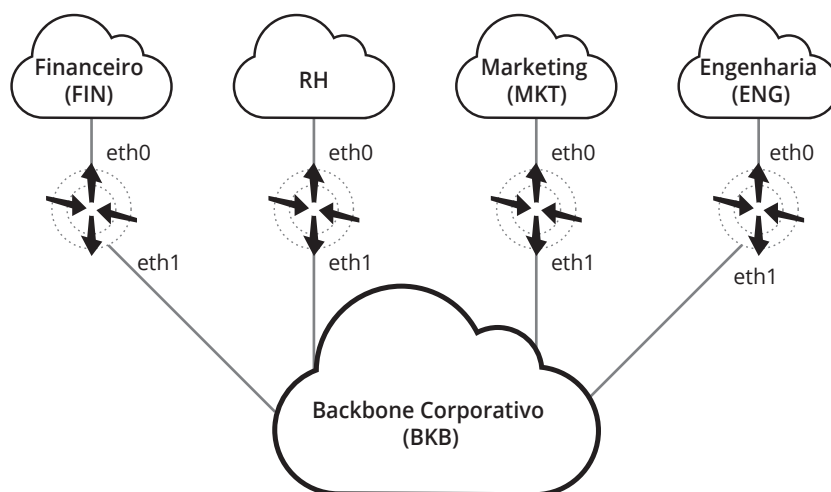


Figura 1.12
Estudo de caso
(parte 1).

Este estudo de caso deve ser resolvido em 15 minutos. Os alunos podem fazê-lo em duplas.

Seja a rede corporativa da Figura 1.12.

Cada departamento tem uma rede com cerca de 20 a 30 computadores. Essas redes estão interligadas ao backbone corporativo composto por um Switch Gigabit Ethernet que, por sua vez, interliga todas as redes departamentais.

Está disponível uma Classe C 200.248.228.0/24, que se deseja dividir em sub-redes que acomodem as necessidades de endereçamento de todas as redes descritas anteriormente. Usando o conceito de VLSM, defina:

1. Endereços IP de cada sub-rede (Endereço de sub-rede, Máscara e Faixa de Hosts).
2. Endereços IP das interfaces dos roteadores de cada sub-rede.

Preencha a tabela abaixo com os resultados pedidos.

Rede	Endereço de sub-rede	Faixa de Hosts	Interface eth0 roteador	Interface eth1 roteador
FIN				
RH				
MKT				
ENG				
BKB			-	-

Uma vez feito o planejamento, vamos configurar a rede, seguindo o procedimento descrito abaixo.

1. Carregue a máquina virtual do simulador e abra o arquivo *Rede2_Sessao1_ADR10.imn*, seguindo a orientação do instrutor. A rede deverá ser idêntica à da Figura 1.13, mostrada a seguir.

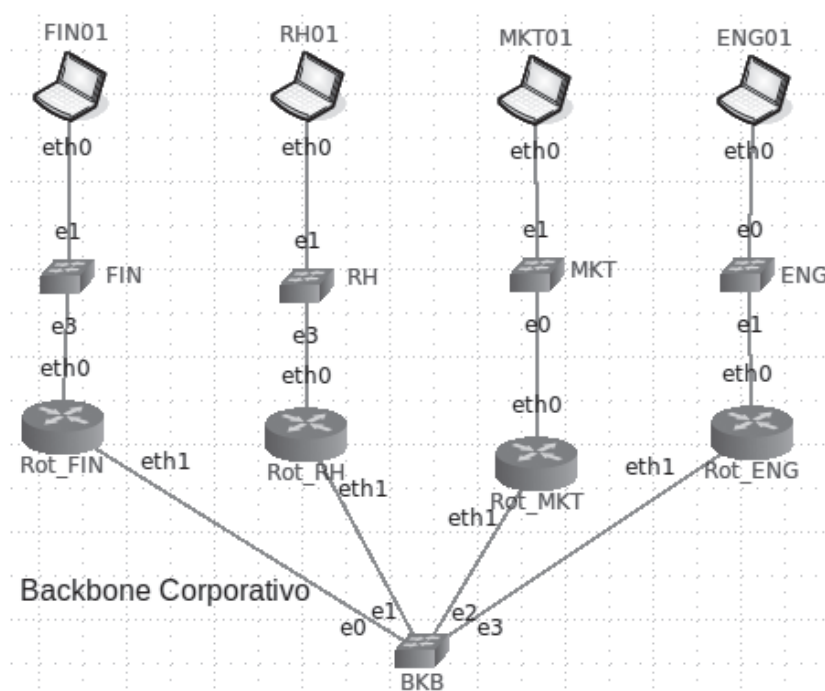


Figura 1.13
Rede do estudo de caso (parte 1).

2. Configure as interfaces dos roteadores (eth0 e eth1) de cada departamento. Configure um PC de cada departamento, conforme mostrado na Figura 1.10. Use o primeiro endereço disponível de cada sub-rede para o roteador e o último endereço para o PC.
3. Ainda falta a configuração de rotas estáticas, para que cada roteador possa acessar as redes remotas. Por exemplo, no caso do roteador Rot_FIN, precisamos “ensinar” as rotas para as redes remotas:
 - 200.248.228.32/27 via 200.248.228.130 (Rot_RH).
 - 200.248.228.64/27 via 200.248.228.131 (Rot_MKT).
 - 200.248.228.96/27 via 200.248.228.132 (Rot_ENG).

De forma semelhante para os demais roteadores (consulte a sua tabela de endereços IP).

4. Precisamos configurar os PCs.



Para configurar o endereço IP e a máscara de sub-rede do FIN01, o comando é:

```
# ifconfig eth0 200.248.228.30 netmask255.255.255.224.
```

Precisamos definir também o gateway padrão dos PCs, que será a interface eth0 do roteador correspondente a cada departamento. Por exemplo, na console do FIN01, se digitarmos o comando `route -n`, obteremos a seguinte resposta:

```
root@FIN01:/tmp/pycore.38996/FIN01.conf# route -n

Kernel IP routing table

Destination      Gateway          Genmask          Flags ...  Iface
200.248.228.0    0.0.0.0          255.255.255.224  U           etho

root@FIN01:/tmp/pycore.38996/FIN01.conf#
```

Aparece somente a rota para a rede 200.248.228.0/24, que é usada para entrega direta. Falta o gateway padrão para poder encaminhar pacotes para as outras redes. O comando é:

```
root@FIN01:/tmp/pycore.38996/FIN01.conf# route add -net default gw
200.248.228.1
```

Verificando:

```
root@FIN01:/tmp/pycore.38996/FIN01.conf# route -n

Kernel IP routing table

Destination      Gateway          Genmask          Flags ...  Iface
200.248.228.0    0.0.0.0          255.255.255.224  U           etho
0.0.0.0          200.248.228.1   0.0.0.0          UG          eth0

root@FIN01:/tmp/pycore.38996/FIN01.conf#
```

Os demais PCs são configurados de maneira semelhante.

5. Para verificar se os roteadores aprenderam todas as rotas para todas as sub-redes, use o comando `shiptoute`, conforme mostrado a seguir, para o roteador Rot_FIN.

```
ubuntu# shiptoute

Codes: K-kernel route, C-connected, S-static, R-RIP, O-OSPF
       I-ISIS, B-BGP, >-selected route, *-FIB route, o-OSPFv3

C>* 127.0.0.0/8 is directly connected, lo
C>* 200.248.228.0/27 is directly connected, eth0
S>* 200.248.228.32/27 [1/0] via 200.248.228.130, eth1
S>* 200.248.228.64/27 [1/0] via 200.248.228.131, eth1
S>* 200.248.228.96/27 [1/0] via 200.248.228.132, eth1
C>* 200.248.228.128/29 is directly connected, eth1

ubuntu#
```

6. Agora podemos testar a continuidade entre os PCs, por exemplo, entre o FIN01 e o ENG01.

```
root@FIN01:/tmp/pycore.38996/FIN01.conf# ping -c 2 200.248.228.126
ping 200.248.228.126 (200.248.228.126) 56(84) BYTES OF DATA.
64 bytes from 200.248.228.126: icmp_req=1 ttl=62 time=8.17 ms
64 bytes from 200.248.228.126: icmp_req=2 ttl=62 time=0.245 ms

--- 200.248.228.126 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1001 ms
Rtt min/avg/max/mdev = 0.245/4.210/8.175/3.965 ms
root@FIN01:/tmp/pycore.38996/FIN01.conf#
```

O *ping* deve funcionar entre todos os PCs.

Atividade 1.3 – Estudo de caso (parte 2)

Planejamento de endereçamento IP: mais duas filiais, com 8 a 10 computadores cada.

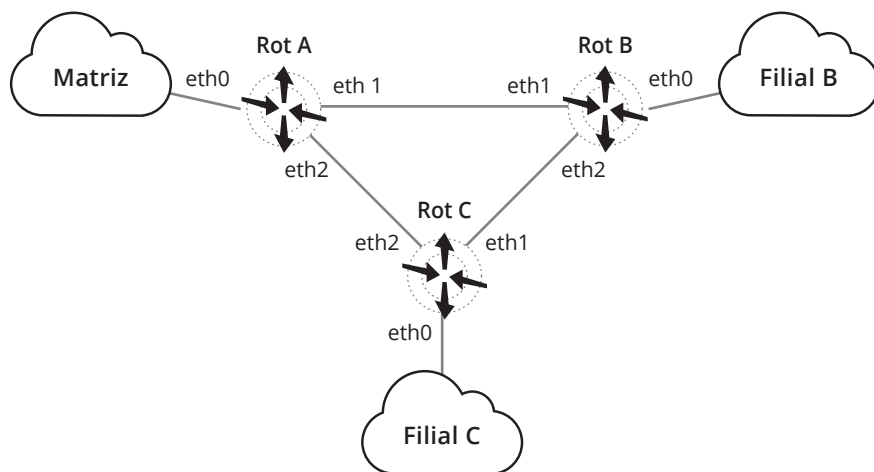


Figura 1.14
Estudo de caso
(parte 2).

Suponhamos agora que seja necessário acrescentar a essa rede do Escritório Central (Site A) mais duas filiais (Sites B e C), cada uma com cerca de 8 a 10 computadores, conforme mostrado na Figura 1.14. Note que o Site A continua como está e o roteador RotA será acrescentado ao backbone corporativo (via interface eth0), respeitando o endereçamento previamente definido.

Só dispomos da Classe C 200.248.228.0. Deve-se manter, tanto quanto possível, os endereços IP atribuídos à rede do Escritório Central, para evitar transtornos aos usuários.

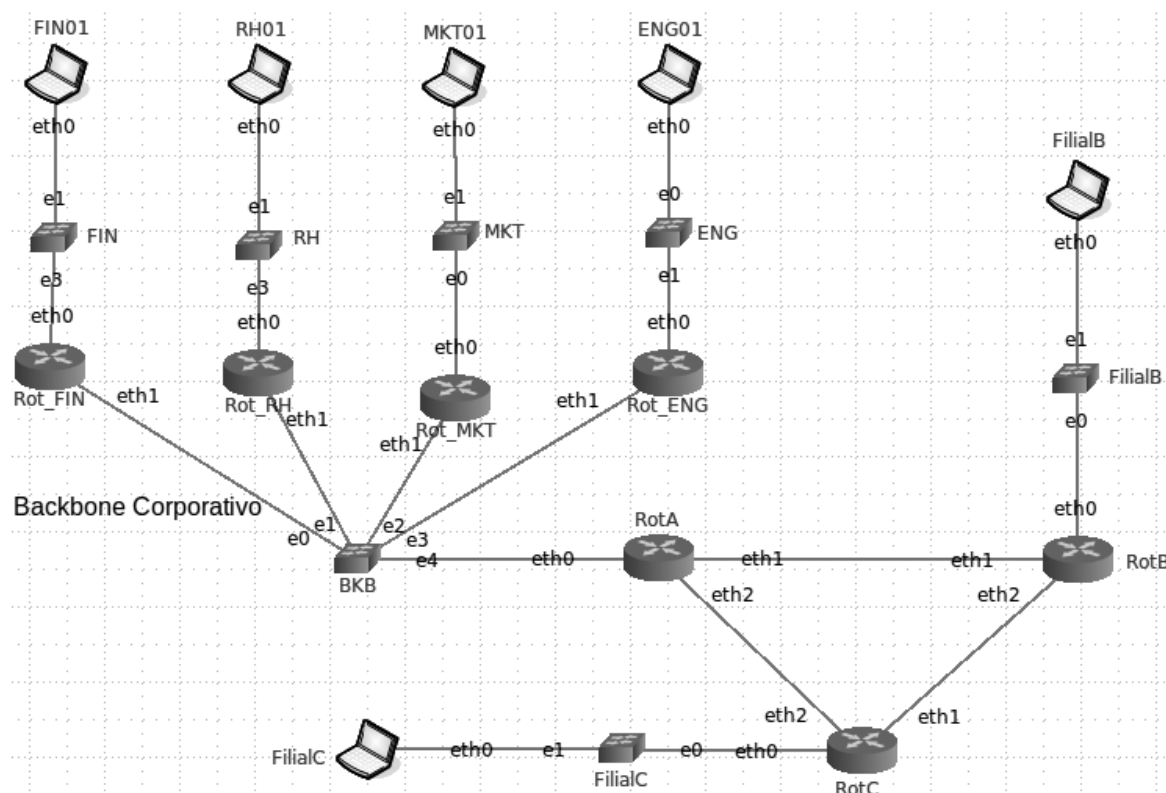
Definir:

1. Endereços IP de cada sub-rede (Endereço de sub-rede, Máscara e Faixa de Hosts).
2. Endereços IP das interfaces dos roteadores de cada sub-rede.

Preencha a tabela abaixo com os resultados pedidos.

Rede Física	Endereço de sub-rede	Faixa de Hosts	Interface eth0 roteador	Interface eth1 roteador	Interface eth2 roteador
BKB Matriz					
RotA-RotB			-	-	-
RotA-RotC			-	-	-
RotB-RotC			-	-	-
Filial B					
Filial C					

1. Carregue a máquina virtual do simulador e abra o arquivo *Rede3_Sessao1_ADR8.imn*, seguindo a orientação do instrutor. A rede deverá ser idêntica à da Figura 1.12, mostrada a seguir.



2. Configure as interfaces eth0, eth1 e eth2 dos roteadores RotA, RotB e RotC. Configure um PC de cada filial, conforme mostrado na Figura 1.15. Use o primeiro endereço disponível de cada sub-rede para o roteador e o último endereço para o PC.
3. Ainda falta a configuração de rotas estáticas, para que cada roteador possa acessar as redes remotas. Por exemplo, no caso do roteador Rot_A, precisamos “ensinar” as rotas para as redes remotas:
 - 200.248.228.0/27 via 200.248.228.129 (Rot_FIN).
 - 200.248.228.32/27 via 200.248.228.130 (Rot_RH).
 - 200.248.228.64/27 via 200.248.228.131 (Rot_MKT).
 - 200.248.228.96/27 via 200.248.228.132 (Rot_ENG).
 - 200.248.228.160/28 via 200.248.228.138 (RotB).
 - 200.248.228.176/28 via 200.248.228.146 (RotC).

Figura 1.15
Rede do estudo de caso (parte 2).

De forma semelhante para os roteadores RotB e RotC (consulte a sua tabela de endereços IP). Os roteadores Rot_FIN, Rot_RH, Rot_MKT e Rot_ENG já tiveram as interfaces e as rotas estáticas configuradas na atividade anterior.

Após a configuração de todas as interfaces, podemos testar a continuidade entre a matriz (PC FIN01) e a filial C (PC FilialC), por exemplo, como mostrado a seguir.

```
root@FIN01:/tmp/pycore.43287/FIN01.conf# ping -c 2 200.248.228.190
PING 200.248.228.190 (200.248.228.190) 56(84) bytes of data.
64 bytes from 200.248.228.190: icmp_req=1 ttl=61 time=0.19 ms
64 bytes from 200.248.228.190: icmp_req=2 ttl=61 time=0.315 ms

--- 200.248.228.190 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1004 ms
rtt min/avg/max/mdev = 0.245/4.210/8.175/3.965 ms
root@FIN01:/tmp/pycore.343287/FIN01.conf#
```

Deve haver continuidade entre todos os PCs.

Conclusão

Este estudo de caso simula a atividade de planejamento de endereçamento de sub-redes:

- Cálculo de endereços e máscaras de sub-redes.
- Acréscimo de novas redes sem refazer a configuração.
- Configuração de interfaces de roteadores.
- Configuração de rotas estáticas.

2

Protocolo de roteamento RIP

objetivos

Conhecer os conceitos básicos sobre protocolos de roteamento interiores e o funcionamento do protocolo RIP.

Sistema Autônomo, CIDR, métricas de protocolo, algoritmos de roteamento e protocolo RIPv2.

conceitos

Sistema Autônomo – AS

Conceito de Sistema Autônomo (Autonomous System – AS):

- Um grupo de redes e roteadores controlados por uma única autoridade administrativa.

Segundo RFC 1930 (definição formal):

- Um conjunto de roteadores controlados por uma única administração técnica, usando um protocolo interior e métricas comuns para rotear pacotes dentro do AS, e usando um protocolo exterior para rotear pacotes para os outros ASs.
- Requisito básico: política de roteamento única.
- A política de roteamento define como são tomadas as decisões de roteamento na internet.

Uma definição comumente adotada diz apenas que o AS está sujeito a administração única.

Uma definição mais rigorosa acrescenta que deve haver uma política de roteamento única no AS. Esse é o requisito básico para ter um AS.

Por política de roteamento entendemos a maneira como as decisões de roteamento são tomadas na internet. Está claro na definição formal que as decisões de roteamento internas ao AS são tão importantes quanto as decisões externas, ou seja, a maneira como o AS se comunica com os outros ASs.

Classless Interdomain Routing (CIDR)

Prefixo IP (Prefix) ou bloco CIDR.

- Bloco de redes Classes A, B ou C.
- Identificação das redes inicia e termina em múltiplos de 2.
- O bloco é identificado por um prefixo e uma máscara.
- Um AS é um grupo conectado de um ou mais prefixos IP controlados por operadores de redes, que têm uma política de roteamento única e bem definida.



Segundo o RFC 1930, a definição clássica de AS é um tanto ambígua, porque não define corretamente o comportamento de um AS.

O conceito de AS está intimamente relacionado ao conceito de CIDR. Um bloco CIDR é um conjunto de redes classful (Classes A, B ou C) cujos números são sequenciais e iniciam e terminam em múltiplos de 2, conforme o exemplo no slide.

Exemplo:

192.168.0.0/24
192.168.1.0/24
192.168.2.0/24 → 192.168.0.0/2
192.168.3.0/24

Assim, um AS pode ser definido em função dos prefixos que o compõem.

O RFC 1930 enfatiza a importância de uma política de roteamento única, relaciona os erros mais comuns na definição de ASs e também discute os critérios de decisão para definir a necessidade de um AS.

O AS é identificado por um número inteiro de dois octetos; portanto, é um número na faixa de 1 a 65535. Na época da emissão do RFC 1930 existiam 5.100 ASs autorizados, porém menos de 600 eram ativamente roteados na internet global.

Os ASs são controlados pelo Internet Assigned Numbers Authority – IANA (<http://www.iana.org>).



Veja em <http://www.iana.org/numbers> como registrar um AS.

Classificação de protocolos de roteamento

Protocolos de roteamento podem ser:

- Interiores (Interior Gateway Protocol – IGP): usados para comunicação entre roteadores de um mesmo AS.
 - Exemplos: RIP – RFC2453, OSPF – RFC2328.
- Exteriores (Exterior Gateway Protocol – EGP): usados para comunicação entre roteadores de ASs diferentes.
 - Exemplos: EGP (obsoleto), BGP-4 – RFC4271.



Os roteadores na internet são organizados hierarquicamente. Alguns roteadores são usados para enviar informação através de um grupo particular de redes sob uma mesma autoridade administrativa e de controle (AS). Roteadores usados para troca de informação no interior dos ASs são chamados roteadores interiores (interior routers) e usam uma variedade de protocolos chamada Interior Gateway Protocols (IGPs). Roteadores usados para troca de informação entre os ASs são chamados roteadores exteriores (exterior routers) e usam protocolos chamados Exterior Gateway Protocols (EGPs).

Os principais protocolos padronizados usados atualmente são RIP, OSPF e BGP-4. Existem outros protocolos proprietários de fabricantes, como por exemplo o IGRP e o EIGRP, da Cisco.



É recomendável usar protocolos padronizados, sempre que possível. Os padrões podem ser encontrados na listagem de RFCs, disponível em: <http://www.rfc-editor.org/rfc-index2.html>

Roteamento dinâmico

Métrica dos protocolos de roteamento:

- Contador de hops.
- Bandwidth (largura de banda).
- Delay (retardo).



- Custo.
- Confiabilidade.
- Carga.
- MTU.
- Ticks.

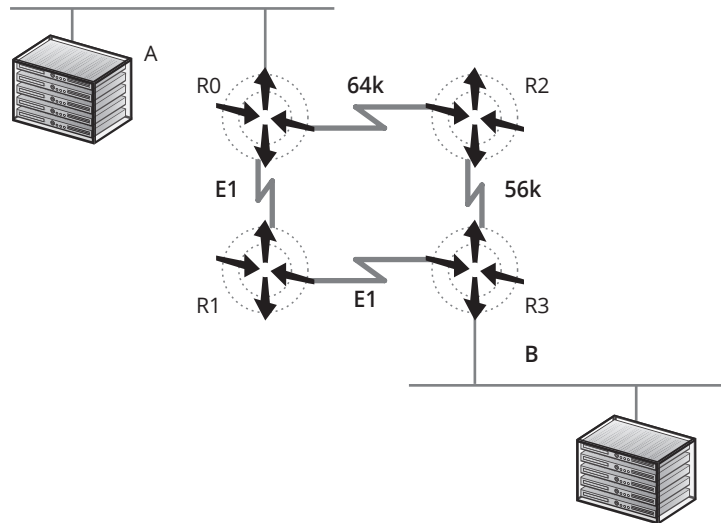


Figura 2.1
Métrica dos
protocolos de
roteamento.

Roteadores usam um protocolo em comum para trocar informações de roteamento. Atualizações de informações de roteamento são mandadas quando a topologia da rede muda ou em intervalos fixos. As informações de roteamento atualizadas contêm as redes acessíveis acrescidas de um valor de métrica associado a cada caminho possível.

O melhor caminho entre redes ou sub-redes é determinado por uma métrica de roteamento. As métricas são importantes porque, além de determinarem uma rota para o destino, os roteadores têm de determinar a melhor rota para cada destino. Na Figura 2.1, por exemplo, no caminho de A para B é evidente que a rota que passa por R1 é mais rápida que a rota que passa por R2. Assim, o roteador R0 deve usar essa informação para escolher a melhor rota.

Variáveis usadas para métricas incluem:

- **Contador de hops** (saltos): o número de paradas intermediárias que um pacote faz em um caminho para seu destino. Passando através de um roteador/gateway, conta-se um hop.
- **Largura de banda** (bandwidth): a capacidade de transportar dados de um meio. Usualmente medido em Mbps ou alguma fração dessa medida.
- **Atraso** (delay): a quantidade de tempo associado ao uso de um meio em particular. Usualmente medido em ms (10^{-3} seg).
- **Confiabilidade**: um valor associado a cada meio, indicando a probabilidade de os dados serem entregues. Usualmente expresso como um valor fracionário; algum número dividido por 255.
- **Carga**: um valor dinâmico indicando a utilização de um meio. Usualmente expresso como um valor fracionário; algum número dividido por 255.
- **MTU**: unidade máxima de transmissão. O maior tamanho do pacote para um meio em particular.

- **Custo:** um valor arbitrário indicando o custo para usar essa interface. Usualmente expressa como um valor inteiro; definido para uma interface de saída.
- **Ticks:** um valor arbitrário associado a um delay quando links e interfaces são usados. O valor usualmente adotado é 1/18 de segundo.

Algoritmo de roteamento

- Vetor-Distância (Bellman-Ford):
 - Cada roteador mantém uma lista de rotas conhecidas.
 - Cada roteador divulga sua tabela para seus vizinhos.
 - Cada roteador seleciona os melhores caminhos dentre as rotas conhecidas e divulgadas.
- A escolha do melhor caminho é baseada na métrica.
 - Normal: menor caminho, melhor rota.
- Processo de montagem da tabela de rotas.
 - Vide anotações.

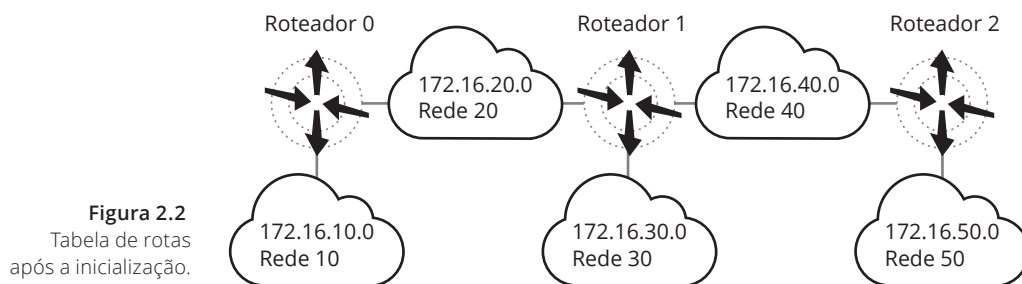


Processo de montagem da tabela de rotas:

1. Quando o roteador executa o boot, ele armazena na tabela informações sobre cada uma das redes que estão diretamente conectadas a ele. Cada entrada na tabela indica uma rede de destino, o gateway para a rede e a sua métrica.
2. Periodicamente cada roteador envia uma cópia da sua tabela para qualquer outro roteador que seja diretamente alcançável (seus vizinhos).
3. Cada roteador que recebe uma cópia da tabela verifica as rotas divulgadas e suas métricas. O roteador soma à métrica divulgada o custo do enlace entre ele e o roteador que fez a divulgação. Depois, compara cada uma das entradas da tabela divulgada com as entradas da sua tabela de roteamento. Rotas novas são adicionadas e rotas existentes são selecionadas pela sua métrica.
 - 3.1. Se a rota já existe na tabela e a métrica calculada é menor do que a da rota conhecida, então remove a entrada anterior e adiciona a nova rota divulgada.
 - 3.2. Se a rota já existe na tabela e a métrica calculada é igual a da rota conhecida, então não altera a entrada, desprezando a rota divulgada.
 - 3.3. Se a rota já existe na tabela e a métrica divulgada é maior do que a da rota conhecida, então verifica se o gateway para essa rota é o mesmo que está fazendo a nova divulgação:
 - 3.3.1...Se o gateway é o mesmo, altera a métrica para essa rota.
 - 3.3.2...Se o gateway não é o mesmo, não altera a rota conhecida, desprezando a rota anunciada.

Tabela de roteamento Vetor-Distância

Router0			Router1			Router2		
Destino	Next Hop	Métrica	Destino	Next Hop	Métrica	Destino	Next Hop	Métrica
Rede 10	-	1	Rede 20	-	1	Rede 40	-	1
Rede 20	-	1	Rede 30	-	1	Rede 50	-	1
			Rede 40	-	1			



Considere a rede exemplo da Figura 2.2 com três roteadores e cinco sub-redes. Todas as sub-redes são /24. O mecanismo de cálculo da tabela de rotas através do algoritmo Vetor-Distância (Bellman-Ford) funciona conforme explicado a seguir.

Na inicialização, antes de trocar informações com seus vizinhos, cada roteador só conhece as redes às quais está diretamente conectado. Então as respectivas tabelas, mostradas na figura, só têm as seguintes redes diretamente conectadas:

- router0: redes 10 e 20.
- router1: redes 20, 30 e 40.
- router2: redes 40 e 50.

Todas as redes têm métrica =1, porque não há nenhum roteador entre as redes e os respectivos roteadores. Segundo a recomendação do RFC 2453, usualmente adota-se a métrica =1 nesse caso, embora não haja nenhum hop intermediário para redes diretamente conectadas. Teoricamente a métrica seria zero para redes diretamente conectadas.

Router0			Router1			Router2		
Destino	Next Hop	Métrica	Destino	Next Hop	Métrica	Destino	Next Hop	Métrica
Rede 10	-	1	Rede 20	-	1	Rede 40	-	1
Rede 20	-	1	Rede 30	-	1	Rede 50	-	1
Rede 30	Router1	2	Rede 40	-	1	Rede 20	Router1	2
Rede 40	Router1	2	Rede 10	Router0	2	Rede 30	Router1	2
			Rede 50	Router2	2			

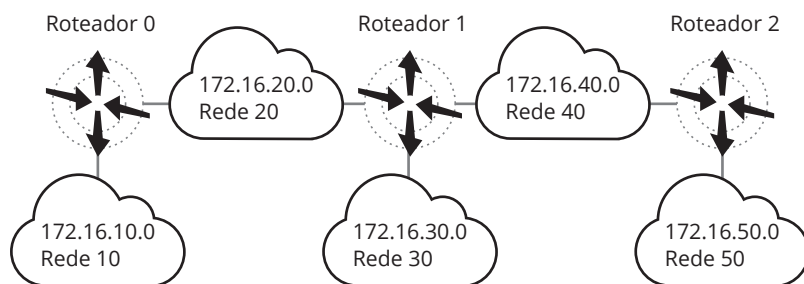


Figura 2.3
Tabela de rotas
após o primeiro
anúncio de rotas.

As tabelas vistas na figura anterior serão anunciadas pelos roteadores a seus vizinhos, da seguinte forma:

- router0 informa ao router1 que tem acesso às redes 10 e 20.
- router1 informa ao router0 e ao router2 que tem acesso às redes 20, 30 e 40.
- router2 informa ao router1 que tem acesso às redes 40 e 50.

Vejam agora o que cada roteador faz com essas informações:

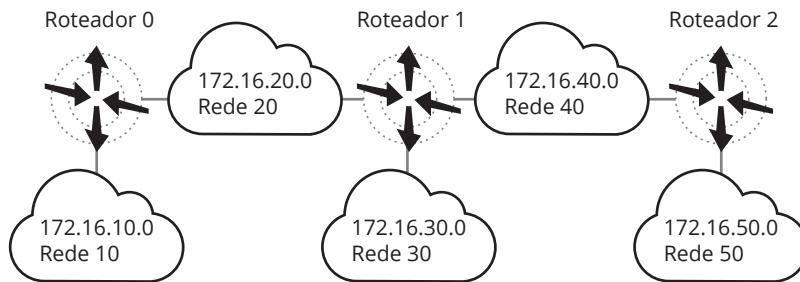
- router0 recebe as informações do router1 e ignora a rota da rede 20, porque ele já a tem na tabela com métrica = 1; as rotas para as redes 30 e 40 são acrescentadas na tabela com métrica = 2, porque elas passam pelo router1, que informou essas rotas com métrica = 1.
- router1 recebe as informações do router0 e do router2; quanto ao router0, ele ignora a rota da rede 20, mas atualiza a rota da rede 10, com métrica = 2, que passa pelo router0; quanto ao router2, ele ignora a rota da rede 40, mas atualiza a rota da rede 50, com métrica = 2, que passa pelo router2.
- router2 recebe as informações do router1 e ignora a rota da rede 40, porque ele já a tem na tabela com métrica = 1; as rotas para as redes 20 e 30 são acrescentadas na tabela com métrica = 2, porque elas passam pelo router1, que informou essas rotas com métrica = 1.

Nesse ponto, a tabela do router1 está completa, mas as tabelas dos roteadores router0 e router2 ainda não estão.

No router0 falta a rota para a rede 50 e, no router2, falta a rota para a rede 10.

Router0			Router1			Router2		
Destino	Next Hop	Métrica	Destino	Next Hop	Métrica	Destino	Next Hop	Métrica
Rede 10	-	1	Rede 20	-	1	Rede 40	-	1
Rede 20	-	1	Rede 30	-	1	Rede 50	-	1
Rede 30	Router1	2	Rede 40	-	1	Rede 20	Router1	2
Rede 40	Router1	2	Rede 10	Router0	2	Rede 30	Router1	2
Rede 50	Router1	3	Rede 50	Router2	2	Rede 10	Router1	3

Figura 2.4
Tabela de rotas
após o segundo
anúncio de rotas.



Finalmente, os roteadores router0 e router2 receberão do router1 as informações de rotas que faltavam.

- O router1 vai ignorar as informações de rotas dos outros dois roteadores, porque a tabela dele está completa.
- router0 atualiza a rota para a rede 50 com métrica = 3, porque o router1 informou que sua métrica era 2 e essa rota passa pelo router1.
- router2 atualiza a rota para a rede 10 com métrica = 3, porque o router1 informou que sua métrica era 2 e essa rota passa pelo router1.

Fim de atualização. Finalmente as tabelas estão completas. Dizemos que o protocolo convergiu. O tempo de convergência vai depender do tempo de cada atualização.

RIPv2 – Características

- Distribuído em 1982 junto com BSD Unix (v1).
- RFC 2453 (standard 56) (v2).
- Protocolo interior (IGP).
- Algoritmo Vetor-Distância (contagem de hops).
- Limite de hops: 15 (16 = destino inalcançável).
- Administrador pode definir métricas das rotas.
- A divulgação é por multicast para os vizinhos.

O protocolo RIP foi projetado inicialmente para a arquitetura Xerox Network Systems (XNS). Em 1982, a versão RIP-IP (v1) foi distribuída junto com o BSD Unix, formalmente definida pelo RFC 1058 de 1988. A versão atual (v2) foi definida pelo RFC 2453.



A tabela de roteamento do RIP fornece várias informações sobre as rotas, tais como: métrica, máscara de sub-rede, temporizadores etc. A métrica usada indica o número de hops até o destino, por default. Em algumas implementações de RIP, o administrador pode definir uma métrica diferente de 1, para um determinado enlace. O RIP mantém apenas a melhor rota para cada destino, sem rotas alternativas. Quando chega nova informação sobre rotas, a antiga é desprezada. Cada roteador, ao perceber modificações nos seus enlaces, manda informação de atualização de rotas para os outros roteadores e assim por diante, propagando as mudanças ao longo da rede.

Na versão 1 (RIPv1), essa atualização era feita através de mensagens broadcast. Na versão 2 são usadas mensagens multicast com endereço multicast padrão: 224.0.0.9. Como outros protocolos de roteamento, o RIP usa certos temporizadores (timers) para regular sua performance. Os temporizadores utilizados pelo RIP são os seguintes:

- **Routing-update timer:** intervalo entre os anúncios de atualização de rota. Cada roteador enviará uma cópia completa de sua tabela de rotas a seus vizinhos a cada 30 segundos, através de uma mensagem de resposta não solicitada.
- **Route-holddown timer:** tempo em que uma rota permanece na tabela após ser declarada inválida por um outro roteador. Quando outro roteador informa que uma rota não é mais válida, esse roteador inicia esse temporizador e aguarda anúncios de atualização para essa rota. Enquanto isso, essa rota é marcada como “possivelmente inválida”. Se vierem anúncios válidos para essa rota, esse temporizador é desligado e essa rota volta a ser anunciada normalmente. Geralmente configurado para 180 segundos.
- **Route-timeout timer:** tempo máximo em que a rota pode ficar sem anúncio de atualização de rota. Esse temporizador controla quando uma rota não está mais disponível. O timeout, normalmente configurado para 180 segundos, determina quanto tempo precisa decorrer sem que um roteador receba qualquer informação sobre uma rota antes que a rota seja declarada inválida. A rota também pode ser declarada inválida se tem métrica = 16.
- **Route-flush timer:** tempo para remoção da rota da tabela. Esse temporizador controla quanto tempo a rota ainda permanece na tabela de rotas depois que vence o temporizador Route-timeout timer. Quando uma rota é declarada inválida, ela permanece na tabela para que os vizinhos possam ser notificados do fato. Essa notificação tem de ocorrer antes do término desse temporizador, normalmente configurado para 120 segundos. Quando esse temporizador expira, a rota é removida da tabela.

Vantagens:

- Simples de configurar.
- Funciona bem em redes pequenas.
- Baixo consumo de largura de banda.

Desvantagens:

- Limitado a 15 hops, sendo inviável em redes grandes.
- Não suporta rotas alternativas.
- Problemas de estabilidade.
 - ▣ Tempo de convergência alto.
 - ▣ Loops.

Com o advento dos protocolos OSPF e IS-IS, parecia que o protocolo RIP se tornaria obsoleto. Embora os novos protocolos sejam superiores ao RIP, este ainda tem algumas vantagens interessantes. Considerando as pequenas redes, o overhead do RIP é muito pequeno,



tanto em termos do uso de largura de banda, como em termos de simplicidade de configuração e implementação. Além disso, existem muito mais implementações de RIP nas redes atuais do que nos outros dois protocolos combinados.

Como desvantagens, podemos citar o fato de que ele é limitado a 15 hops, tornando-o inviável em redes grandes. Por outro lado, com grande número de roteadores, teremos muitas mensagens de anúncio de rotas. Também não suporta rotas alternativas, mantendo apenas a melhor rota para cada destino na tabela. Essas limitações são, na realidade, consequências da concepção do protocolo. Ainda como consequência da concepção do RIP, existem problemas de estabilidade e convergência de tabelas de rotas. A convergência das tabelas dos diversos roteadores é lenta, devido ao tempo de atualização. Definimos tempo de convergência como o tempo necessário para que todas as tabelas dos roteadores fiquem atualizadas, quando há mudança de topologia.

Quanto aos problemas de estabilidade, podemos citar a contagem ao infinito. Veremos que esse problema pode ser minimizado com as técnicas de horizonte dividido, horizonte dividido com inversão envenenada e atualizações imediatas.

Contagem ao infinito

Suponha que a rede 10 esteja fora (caiu o link).

- router0 anuncia que a rota tem métrica 3 (via router1).
- router1 atualiza a métrica para 4 (3+1) (via router0).
- router0 atualiza a métrica para 5 (4+1) (via router1).
- E assim por diante, até atingir a métrica 16.



Destino	Next Hop	Métrica	Destino	Next Hop	Métrica
Rede 10	-	1	Rede 10	Router0	2
Rede 20	-	1	Rede 20	-	1
Rede 30	Router1	2	Rede 30	-	1

Figura 2.5
Problema da contagem ao infinito.



O problema da contagem ao infinito (count to infinity) pode ser demonstrado na Figura 2.5.

- O router0 está diretamente conectado às redes 10 e 20, e o router1 às redes 20 e 30. Cada um anuncia sua tabela para o outro.
- Quando o router0 recebe a tabela do router1, ele atualiza sua tabela incluindo a rede 30, com métrica = 2 via router1.
- Quando o router1 recebe a tabela do router0, ele atualiza sua tabela incluindo a rede 10, com métrica = 2 via router0.

As tabelas ficam como estão na figura. Imagine agora que o link do router0 para a rede 10 caiu (representado pelo X). O router0 verifica que o router1 anuncia que tem rota para a rede 10 com métrica = 2. O router0 então atualiza sua tabela para a rede 10 com métrica = 3, via router1.

O que o router0 não percebe (e aí está o problema) é que a rota anunciada pelo router1 passa por ele mesmo (router0), deixando de ser uma rota válida.

O router1 por sua vez atualiza sua tabela colocando métrica = 4 para a rede 10. O router0, baseado na informação do router1, atualiza a sua tabela para métrica = 5 e assim por diante, até atingir a métrica = 16, que significa rede inatingível.

Considerando que as atualizações são feitas a cada 30 segundos, vai demorar muito para as tabelas convergirem.

Algumas implementações foram feitas no RIPv2 para resolver ou contornar esse problema.

Implementações especiais do RIPv2

Solução do problema de contagem ao infinito:

- **Horizonte dividido:**
 - ▣ Não retorna informações de uma rota ao roteador do qual aprendeu essa rota.
- **Horizonte dividido com inversão envenenada:**
 - ▣ Retorna informação de uma rota com métrica = 16 para o roteador que aprendeu essa rota.
- **Atualizações imediatas:**
 - ▣ Informa imediatamente modificações de rota, sem aguardar o próximo período de anúncio.
- **Horizonte dividido** (Split horizon): com essa técnica, o roteador registra a interface através da qual recebeu informações sobre uma rota, e não difunde informações sobre essa rota através dessa mesma interface. No nosso exemplo, o router1 receberia informações sobre a rota para a rede 10, a partir do router0, logo o router1 não iria enviar informações sobre rotas para a rede 10 de volta para o router0. Com isso já seria evitado o problema do count to infinity. Em outras palavras, essa característica pode ser resumida assim: eu aprendi uma rota para a rede X através de você, logo você não pode aprender uma rota para a mesma rede X através de minhas informações.
- **Inversão envenenada** (Split horizon with poison reverse): nessa técnica, quando um roteador aprende o caminho para uma determinada rede, ele anuncia o seu caminho de volta para essa rede, com uma métrica = 16. No exemplo da figura anterior, o router1 recebe a informação do router0 de que a rede 10 está a 1 hop de distância. O router1 anuncia para o router0 que a rede 10 está a 16 hops de distância. Com isso, jamais o router0 vai tentar achar um caminho para a rede 10 através do router1, o que faz sentido, já que o router0 está diretamente conectado à rede 10.
- **Atualizações instantâneas** (Triggered updates): com essa técnica, os roteadores podem anunciar mudanças na métrica de uma rota imediatamente, sem esperar o próximo período de anúncio. Nesse caso, redes que se tornem indisponíveis podem ser anunciadas imediatamente com um hop de 16, ou seja, inalcançável. Essa técnica é utilizada em combinação com a técnica de inversão envenenada, para tentar diminuir o tempo de convergência da rede em situações onde houve indisponibilidade de um roteador ou de um link. Essa técnica diminui o tempo necessário para convergência da rede, porém gera mais tráfego na rede.

As implementações de RIP têm de suportar *split horizon* e *split horizon with poison reverse*, embora essa última possa ser desabilitada pelo administrador, se ele o desejar. Normalmente isso é feito por motivo de tráfego.



Por outro lado, as atualizações imediatas geram muito tráfego, mas aumentam a velocidade de convergência.

Pacote RIP

- RIP usa protocolo UDP porta 520 para enviar e receber mensagens de atualização de rotas.
- Formato da mensagem.

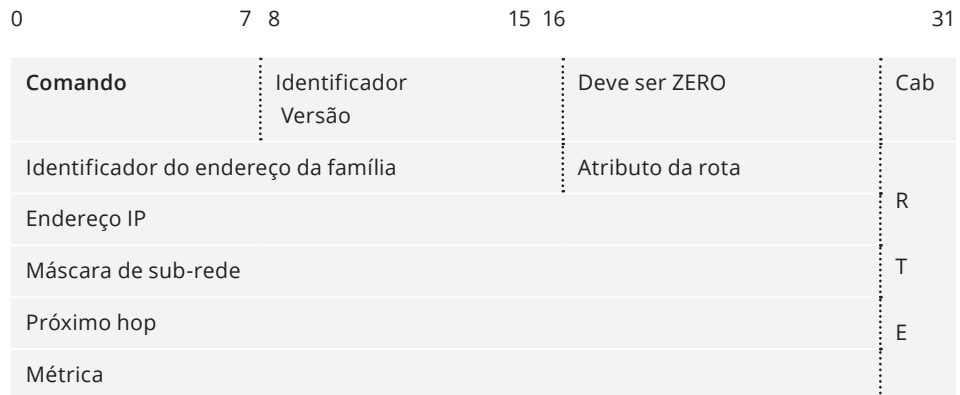


Figura 2.6
Layout do pacote do protocolo RIP.

O protocolo RIP (v1 ou v2) utiliza o UDP, port 520, para enviar e receber datagramas. Todas as mensagens de atualização de rotas usam o port UDP520.

Mensagens de atualização em resposta a um pedido são enviadas ao port de onde veio o pedido. Isso quer dizer que consultas específicas podem ser enviadas de qualquer port, mas o destino delas sempre será o port UDP520.

O layout do pacote RIP está mostrado na figura. O cabeçalho tem 4 octetos (32 bits) e cada anúncio de rota (RouTe Entry – RTE) tem 20 octetos.

São permitidas até 25 RTE por pacote. Se o roteador tiver de anunciar mais de 25 rotas, terá de enviar um segundo pacote.

Descrição dos campos do cabeçalho:

- **Comando:** especifica o propósito dessa mensagem: pedido (1) ou resposta (2).
- **Identificador de versão:** versão 1 ou 2.

Descrição dos campos do anúncio das rotas (RTE):

- **Identificador do endereço da família:** tipo de endereço; na prática RIP só é usado para endereços IP.
- **Atributo da rota** (route tag): deve ser usado para diferenciar as rotas RIP internas do domínio de outras rotas aprendidas de outros AS, via BGP, ou ainda de outros protocolos IGP do domínio como, por exemplo, OSPF.
- **Endereço IP:** identificação da rede para a qual a rota está sendo anunciada.
- **Máscara de sub-rede** (subnet mask): deve ser aplicada ao endereço IP para separar a parte de endereço de rede da parte de endereço de host.
- **Próximo hop** (next hop): endereço IP do próximo hop imediato para o qual os pacotes destinados à rede anunciada devem ser encaminhados.
- **Métrica:** deve conter um valor entre 1 e 15; o valor de 16 indica destino inalcançável.

Configuração do protocolo RIP

No simulador Core os comandos de configuração do protocolo RIP obedecem ao padrão Cisco IOS, conforme mostrado na tabela abaixo.

Comando	Função
# Configure terminal (conf t)	Entra no modo de configuração. Para poder entrar nesse modo, o prompt do roteador tem de terminar com o caractere #, que indica o modo privilegiado de configuração.
(config)# router rip	Habilita o protocolo de roteamento RIP. Note a mudança do prompt do roteador, devido ao comando anterior.
(config-router)# Network <endereço rede>/<máscara>	Define a rede (endereço e máscara) que será anunciada pelo protocolo RIP. Tem de ser uma rede que esteja diretamente conectada às interfaces do roteador. Observe a mudança do prompt.
(config-router)# ^Z	(Ctrl + Z) Encerra o modo de configuração e retorna ao modo privilegiado de configuração.

Figura 2.7
Comandos de
configuração do
protocolo RIP.

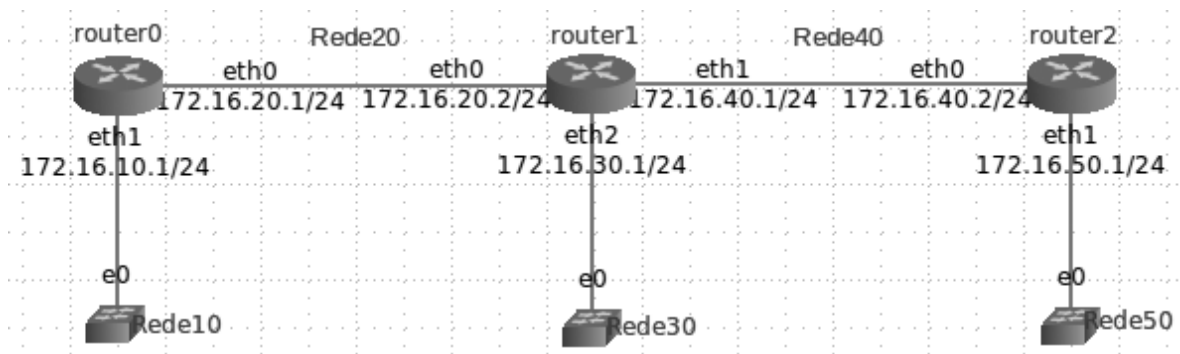


Roteiro de Atividades 2

Atividade 2.1 – Configuração do protocolo RIP

- Carregar a rede *Rede1_Sessao2_ADR8.imn* no Core.
- Iniciar o Wireshark na interface 172.16.20.1 do router0.
- Configurar o protocolo RIP nos três roteadores.
- Analisar o fluxo de pacotes.
- Verificar as tabelas de rotas dos roteadores.

Figura 2.8
Rede1_Sessao2_
ADR8.imn.



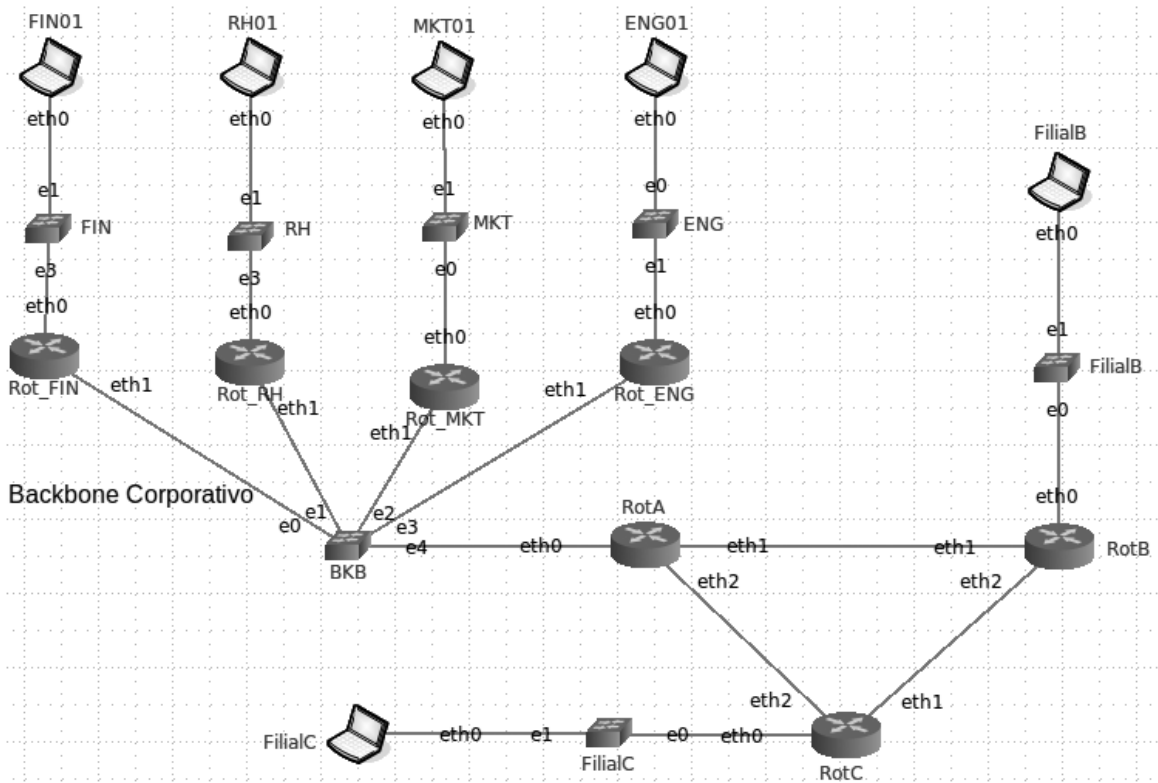
Atividade 2.2 – Atualização de rotas do protocolo RIP

- Iniciar nova captura do Wireshark na mesma interface.
- Colocar em *Off* a interface do router1 para a rede 30.
- Analisar o fluxo de pacotes.
- Verificar as tabelas de rotas dos roteadores.
- Colocar em *On* a interface do router1 para a rede 30.



- Verificar novamente as tabelas de rotas dos roteadores.

Atividade 2.3 – Projeto e configuração do protocolo RIP



As seguintes redes foram atribuídas às respectivas redes físicas, conforme a tabela abaixo.

Figura 2.9
Rede2_Sessao2_ ADR8.imn.

Rede física	Identificação da rede	Máscara de rede
FIN	200.130.24.0/24	255.255.255.0
RH	200.130.25.0/24	255.255.255.0
MKT	200.130.26.0/24	255.255.255.0
ENG	200.130.27.0/24	255.255.255.0
Filial B	200.130.28.0/24	255.255.255.0
Filial C	200.130.29.0/24	255.255.255.0
Backbone Matriz	192.168.0.0/24	255.255.255.0
RotA – RotB	192.168.1.0/30	255.255.255.252
RotA – RotC	192.168.1.4/30	255.255.255.252
RotB – RotC	192.168.1.8/30	255.255.255.252

- Configurar o protocolo RIP em todos os roteadores.

- Verificar as tabelas de rotas dos roteadores.
- Verificar a conectividade entre as diversas redes.

Conclusão

Nestas atividades práticas aprendemos a:

- Planejar endereçamento IP de redes que usam o protocolo RIP;
- Configurar protocolo RIP;
- Analisar fluxo de mensagens RIP;
- Verificar convergência das tabelas de rotas;
- Verificar como os roteadores atualizam dinamicamente suas tabelas de rotas.

Esse processo de atualização de rotas também acontece exatamente assim em redes grandes. Dependendo da quantidade de atualizações de rotas e da quantidade de roteadores envolvidos, teremos um grande volume de tráfego de mensagens RIP.



3

Protocolo de roteamento OSPF

objetivos

Realizar uma comparação entre os protocolos RIP e OSPF, assim como entre suas características básicas e configurações.

Protocolo OSPF, Estado de Enlace, Algoritmo SPF, Arquitetura OSPF e Pacotes OSPF.

conceitos

Open Shortest Path First (OSPF)

- Protocolo de roteamento IP tipo IGP (Interior).
- Substituiu o protocolo RIP.
- RFC 2328 – STD54 – OSPF v2.
- Algoritmo SPF (Shortest Path First).
- Protocolo de Estado de Enlace (Link State).
- Métrica: custo de saída da interface.
- Convergência rápida das tabelas de rotas.
- Utilizado em redes de médio e grande porte.



O OSPF é um protocolo de roteamento desenvolvido para redes IP pelo grupo de trabalho do Interior Gateway Protocol (IGP) do Internet Engineering Task Force (IETF). Esse grupo de trabalho foi formado em 1988 para projetar um protocolo IGP baseado no algoritmo Shortest Path First (SPF), para uso na internet.

O OSPF foi criado pela mesma razão que o IGRP: o protocolo RIP estava se tornando incapaz de operar em redes grandes e heterogêneas. O OSPF é padronizado pelo RFC 2328, sendo, portanto, um padrão aberto e bastante difundido entre todos os fabricantes de roteadores.

O OSPF é um protocolo do tipo Estado do Enlace (Link State). Como tal, ele solicita aos roteadores dentro da mesma área hierárquica que enviem Anúncios do Estado do Enlace (Link State Advertisements – LSAs), que contêm informações sobre métricas usadas, interfaces conectadas e outras variáveis. À medida que os roteadores acumulam informações sobre o estado dos enlaces, eles usam o algoritmo SPF para calcular a menor trajetória para cada nó.

O OSPF contrasta com os protocolos RIP e IGRP, que usam algoritmo de Vetor de Distância (Distance Vector). Estes últimos enviam parte ou toda a tabela de roteamento em mensagens

de atualização de rotas, mas somente para seus vizinhos. Diferentemente do RIP, o OSPF pode operar dentro de uma hierarquia. A entidade de nível mais alto é o Sistema Autônomo (Autonomous System – AS), que é uma coleção de redes sob uma administração comum, compartilhando uma estratégia de roteamento comum. OSPF é um protocolo de roteamento intra-AS (Interior Gateway), embora seja capaz de receber e enviar rotas para outros ASs.

O algoritmo SPF é a base para operação do OSPF. Quando um roteador OSPF é ligado, ele inicializa suas estruturas de dados do protocolo de roteamento e espera que os protocolos da camada inferior indiquem que suas interfaces estão funcionais. Uma vez assegurado do funcionamento de suas interfaces, o roteador usa mensagens de Hello para conhecer seus vizinhos, que são roteadores com interfaces para uma rede comum.

As métricas, por default, são calculadas segundo a velocidade do enlace, usando a fórmula:

■ $\text{Custo} = 10^8 / \text{velocidade de enlace}$.

Exemplo: um enlace de 100Mbps terá o custo = $100000000 / 100000000 = 1$.

Um enlace E1 de 2,048 Mbps terá o custo = $100000000 / 2048000 = 48$.

Enquanto o RIP converge proporcionalmente ao número de nós da rede, o OSPF converge em uma proporção logarítmica ao número de enlaces. Isso torna a convergência do OSPF muito mais rápida. Além disso, no protocolo RIP, a mensagem é proporcional ao número de destinos; portanto, se a rede é muito grande, cada mensagem terá de ser subdividida em vários pacotes, diminuindo mais ainda a velocidade de convergência. Por sua concepção, o OSPF é adequado a redes de médio e grande porte.

Comparação RIP x OSPF

Característica	RIP	OSPF
Limite de hops	15	Não
Suporta Variable Length Subnet Mask (VLSM)	Sim	Sim
Broadcasting periódico da tabela de roteamento	Sim	Não
Broadcasting somente quando a tabela é atualizada	Não	Sim
Atualização de tabelas com mensagens IP multicast	Sim	Sim
Convergência das tabelas de roteamento	Lenta	Rápida
Decisão de roteamento baseada somente nos hops	Sim	Não
Decisão de roteamento baseada em várias métricas	Não	Sim
Rotas alternativas para o mesmo destino	Não	Sim
Hierarquia de roteamento (divisão em áreas)	Não	Sim
Autenticação das mensagens de atualização de rotas	Sim	Sim
Comunicação com protocolos exteriores ao AS	Não	Sim

Comparação das características dos protocolos RIP e OSPF:

- RIP é limitado a 15 hops, acima disso é inalcançável; o OSPF não tem limite no número de hops.
- RIPv1 não suporta VLSM, que é um recurso muito útil para o aproveitamento de endereçamento IP; o RIPv2 suporta VLSM, bem como o OSPF, que faz um uso inteligente desse recurso.
- Broadcasts periódicos da tabela de roteamento completa consomem grandes quantidades de largura de banda, especialmente em redes maiores, e são críticos em enlaces seriais lentos e redes WAN; OSPF só faz broadcast quando há alteração na tabela de roteamento.
- RIPv1 não suporta mensagens multicast de atualização de tabelas, apenas o RIPv2, bem como o OSPF.
- RIP tem uma convergência mais lenta do que o OSPF, porque os roteadores RIP temporizam hold-down e garbage collection, e demoram para perceber o timeout de informações; o OSPF propaga instantaneamente as informações da tabela de roteamento, e não periodicamente, como o RIP.
- RIP usa somente a métrica de número de hops, sem considerar retardo dos enlaces e custos das rotas, que são parâmetros importantes em grandes redes; links com menos hops são sempre preferidos em detrimento de links com mais hops, embora esses últimos possam ser mais rápidos.
- OSPF considera o custo de cada rota e faz um melhor balanceamento de carga considerando o uso de rotas alternativas; isso é possível porque o OSPF tem um banco de dados da topologia da rede e não apenas dados de cada rota que ele conhece; em consequência disso, o OSPF calcula rotas livres de loops.
- Redes RIP não têm hierarquia (são ditas flat — planas), não permitindo a definição de áreas; OSPF permite a divisão do domínio de roteamento (AS) em várias áreas, reduzindo a propagação de informações de roteamento.
- RIPv1 não autentica mensagens de atualização de tabelas, apenas o RIPv2, bem como o OSPF.
- RIP não permite a comunicação com protocolos exteriores ao AS (como o BGP, por exemplo); o OSPF permite a introdução de rotas externas oriundas do BGP.

Conceito de Estado do Enlace

- O Estado do Enlace pode ser considerado como uma descrição da interface do roteador.

Link State Routing Protocol.

- Substituiu o protocolo Vetor-Distância.

Características:

- Descobrir seus vizinhos e seus endereços de rede.
- Calcular o retardo ou custo para cada um dos vizinhos.
- Construir um pacote informando tudo o que aprendeu.
- Propagar o pacote para todos os roteadores.
- Calcular o menor caminho para todos os roteadores.

Podemos considerar um enlace (link) como uma interface do roteador. O Estado do Enlace (Link State) é uma descrição da interface e de seu relacionamento com os seus roteadores vizinhos. Uma descrição da interface pode incluir, por exemplo, o endereço IP da interface,



a máscara de sub-rede, o tipo de rede a qual está conectada, os roteadores conectados àquela rede e assim por diante.

O conjunto de todos os Estados de Enlace forma um banco de dados de Estado de Enlace.

Esse protocolo com abordagem dinâmica é um dos mais populares algoritmos empregados em redes modernas, substituindo o protocolo Vetor-Distância pelas vantagens apresentadas na comparação RIP x OSPF.

O protocolo se baseia nos cinco pontos relacionados a seguir:

- Descobrir seus vizinhos e seus endereços de rede.
- Calcular o retardo ou custo para cada um dos vizinhos.
- Construir um pacote informando tudo o que aprendeu.
- Propagar o pacote para todos os roteadores.
- Calcular o menor caminho para todos os roteadores.

Algoritmo SPF – Dijkstra

- Cada nó é etiquetado com a distância desde o nó fonte, ao longo do melhor caminho conhecido.
- Inicialmente todos os nós são etiquetados com “ ∞ ”.
- À medida que o algoritmo vai calculando os caminhos mais curtos, as etiquetas vão mudando.
- Uma etiqueta pode ser experimental ou permanente.
- Inicialmente todas as etiquetas são experimentais.
- Quando uma etiqueta representa o menor caminho possível entre a fonte e o nó específico, ela se torna permanente e não será mais alterada.



O roteamento pelo caminho mais curto se baseia no algoritmo de Dijkstra, cujos passos estão representados anteriormente. Basicamente consiste em construir um grafo da rede, onde cada nó representa um roteador, e o arco que interliga um par de nós representa uma linha de comunicação (link).

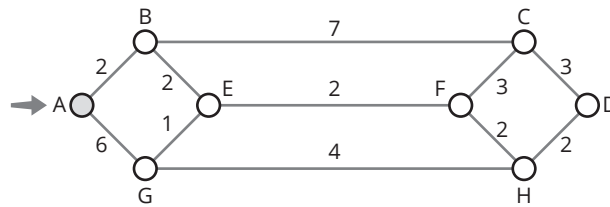
Para escolher uma rota entre um dado par de roteadores, o algoritmo precisa apenas determinar o caminho mais curto entre os roteadores no grafo. Para isso, pode se basear em diferentes métricas:

- Número de hops entre fonte e destino.
- Distância física (geográfica).
- Fila média e atraso de transmissão associado a cada linha de comunicação no caminho.

O administrador pode definir um valor único que representa a ponderação desses fatores e que se chama custo da rota. O caminho mais curto será o caminho mais rápido, não necessariamente o de menor número de hops ou de quilômetros.

Algoritmo SPF (passo 1)

Figura 3.1
Algoritmo SPF
(passo 1).

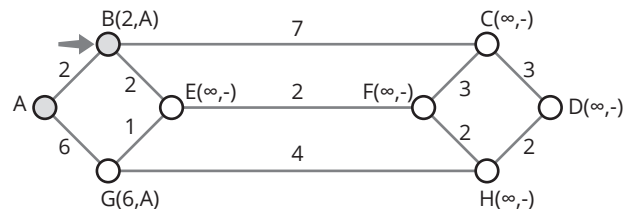


Exemplo de funcionamento do algoritmo SPF:

- Os números representam o custo das rotas.
 - Desejamos determinar o menor caminho de A até D.
1. Nó A é marcado como nó permanente (círculo cheio).
 2. Nó A é chamado Nó de Trabalho.
 3. Cada um dos nós adjacentes ao nó A é examinado e reetiquetado com a distância entre ele e o nó A.

Algoritmo SPF (passo 2)

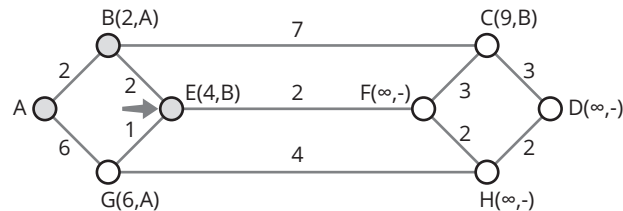
Figura 3.2
Algoritmo SPF
(passo 2).



4. Sempre que um nó é reetiquetado, é marcado com a identificação do nó a partir do qual o cálculo foi feito, para permitir a reconstrução do caminho final.
5. Tendo sido examinados todos os nós adjacentes ao nó A, aquele com o menor valor é feito nó permanente, passando a ser o novo Nó de Trabalho, no caso, o nó B.

Algoritmo SPF (passo 3)

Figura 3.3
Algoritmo SPF
(passo 3).



6. A partir do novo Nó de Trabalho (nó B), cada um dos nós adjacentes é examinado. Se a soma da etiqueta em B com a distância entre o nó B e o nó que está sendo examinado for menor que o valor da etiqueta naquele nó, está definido o menor caminho, e o nó é reetiquetado.
7. Após todos os nós adjacentes ao nó de trabalho terem sido examinados, o nó com o menor valor se torna permanente e passa a ser o novo nó de trabalho (nó E).

Algoritmo SPF (passo 4)

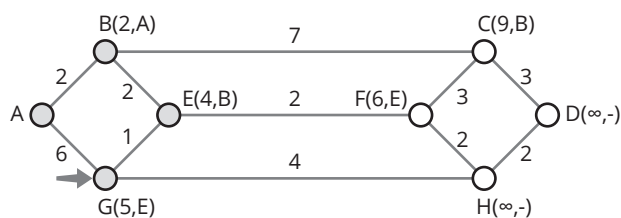


Figura 3.4
Algoritmo SPF
(passo 4).

8. A partir do novo Nó de Trabalho (nó E), cada um dos nós adjacentes é examinado. Se a soma da etiqueta em E com a distância entre o nó E e o nó que está sendo examinado for menor que o valor da etiqueta naquele nó, está definido o menor caminho e o nó é reetiquetado.
9. Após todos os nós adjacentes ao Nó de Trabalho terem sido examinados, o nó com o menor valor se torna permanente e passa a ser o novo Nó de Trabalho (nó G).

Algoritmo SPF (passo 5)

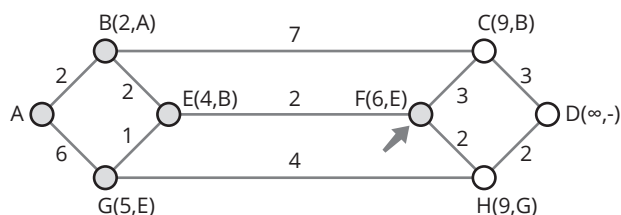


Figura 3.5
Algoritmo SPF
(passo 5).

10. A partir do novo Nó de Trabalho (nó G), cada um dos nós adjacentes é examinado.
11. Após todos os nós adjacentes ao Nó de Trabalho terem sido examinados, o nó com o menor valor se torna permanente e passa a ser o novo Nó de Trabalho (nó F).

Algoritmo SPF (passo 6)

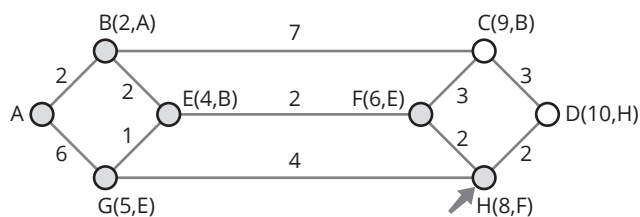


Figura 3.6
Algoritmo SPF
(passo 6).

12. A partir do novo Nó de Trabalho (nó F), cada um dos nós adjacentes é examinado.
13. Após todos os nós adjacentes ao Nó de Trabalho terem sido examinados, o nó com o menor valor se torna permanente e passa a ser o novo Nó de Trabalho (nó H).
14. A partir do nó H, determina-se o nó D (10,H).

Algoritmo SPF (resumo)

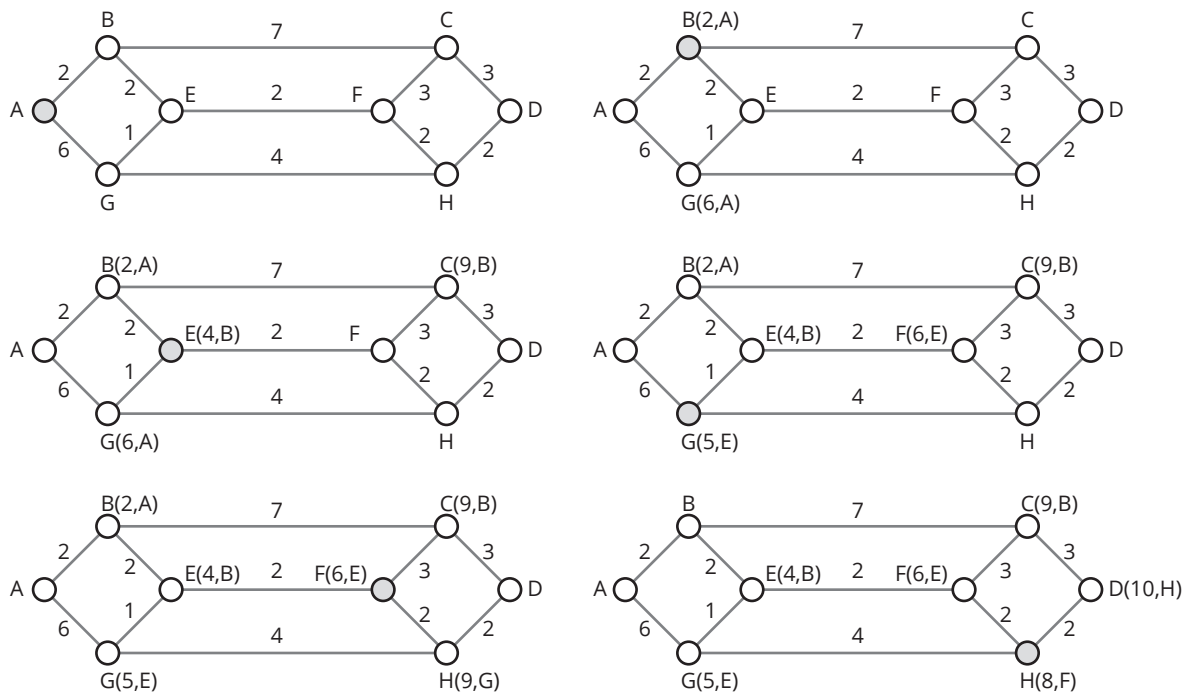


Figura 3.7
Algoritmo SPF
(Resumo).

A Figura 3.7 resume os passos explicados anteriormente.

- **Passo 1:** o nó A é o Nó de Trabalho (círculo azul cheio); a etiqueta torna-se permanente e os nós adjacentes a ele são examinados para determinar a distância dele até o nó em exame; os nós B e G ganham etiquetas, mas apenas o nó B é permanente, porque tem o menor valor; todas as etiquetas são marcadas com a identificação do nó a partir do qual o cálculo foi feito, para permitir a reconstrução do caminho.
- **Passo 2:** como o nó B é de menor valor, ele passa a ser o novo Nó de Trabalho; a partir dele são calculadas as distâncias relativas aos nós C e E; o nó E é tornado permanente porque tem o menor valor.
- **Passo 3:** o nó E é o novo Nó de Trabalho; a partir dele são calculadas as distâncias relativas aos nós G e F; note que o nó G tem seu valor redefinido; era 6,A e ficou 5,E, porque o valor é menor a partir de E do que a partir de A; o nó G é tornado permanente porque tem o menor valor.
- **Passo 4:** o nó G é o novo Nó de Trabalho; a partir dele é calculada a distância relativa ao nó H; o nó F é tornado permanente porque tem o menor valor.
- **Passo 5:** o nó F é o novo Nó de Trabalho; a partir dele são calculadas as distâncias relativas aos nós C e H; o nó H é tornado permanente porque tem o menor valor; sua distância foi recalculada porque foi encontrado um caminho menor; o nó C permanece com o mesmo valor, pois as duas rotas calculadas são iguais.
- **Passo 6:** o nó H é o novo Nó de Trabalho; a partir dele é calculada a distância relativa ao nó D (10,H).

Funcionamento do protocolo OSPF

Os passos do algoritmo do Estado de Enlace:

1. O roteador gera um anúncio de Estado de Enlace.
2. Os roteadores trocam informações entre si usando o protocolo flooding (inundação).
3. Após completar o banco de dados, cada roteador calcula o caminho mais curto para os demais.
4. Se nenhuma alteração de topologia ocorrer, nenhuma informação será trocada entre os roteadores; se ocorrer mudança, o caminho mais curto será recalculado.

O protocolo OSPF usa o algoritmo do Estado de Enlace para construir e calcular o caminho mais curto para todos os destinos conhecidos. Um resumo dos passos do algoritmo pode ser descrito como:

- Na inicialização ou devido a alguma mudança de topologia na rede, o roteador gera um anúncio de Estado de Enlace. Esse anúncio contém o conjunto de todos os estados de enlace do roteador.
- Todos os roteadores trocam informações do Estado de Enlace através do protocolo de inundação (flooding). Cada roteador que recebe uma atualização do Estado de Enlace armazena uma cópia no seu banco de dados de Estado de Enlace, e depois propaga a atualização para os outros roteadores.
- Depois que cada roteador completa o seu banco de dados, ele calcula a árvore de trajetórias mais curta (Shortest Path Tree) para todos os destinos. O roteador usa o algoritmo de Dijkstra para fazer esse cálculo. Os destinos, o seu respectivo custo associado e o próximo hop para atingir aquele destino formam a tabela de roteamento IP.
- Se nenhuma alteração na rede OSPF ocorrer, como, por exemplo, o custo de um link ou novas redes adicionadas ou excluídas, OSPF permanecerá em silêncio. Quaisquer mudanças que ocorram serão informadas via pacotes de Estado de Enlace e o algoritmo de Dijkstra é recalculado para encontrar o caminho mais curto.

OSPF – Roteadores de borda e área

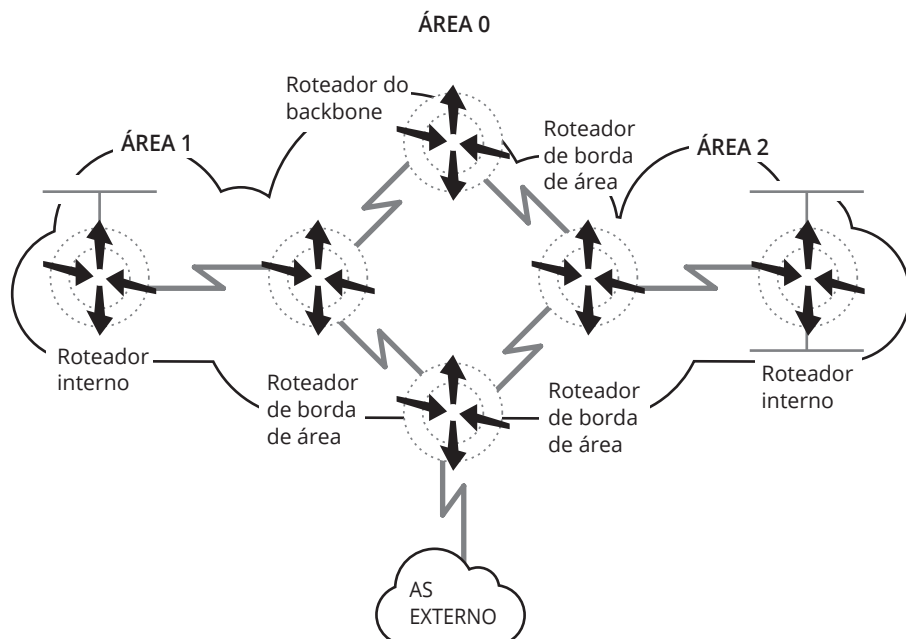
Rede é dividida em áreas.

- Objetivo: reduzir o tráfego.
- Hierarquia de roteamento dentro do AS.
- Área 0 (zero): Backbone OSPF.
- Todas as áreas se conectam ao backbone.
- Função do roteador depende da localização.
 - ▣ Internal Router.
 - ▣ Backbone Router.
 - ▣ Area Border Router.
 - ▣ AS Border Router.

Um AS pode ser dividido em áreas, que são um grupo de redes contíguas e seus hosts conectados. Roteadores com múltiplas interfaces podem participar em múltiplas áreas e são chamados roteadores de fronteira de área (Area Border Routers).

Eles mantêm bancos de dados topológicos para cada área. Um banco de dados topológico é essencialmente uma visão geral de redes relativamente a roteadores, contendo a coleção

Os roteadores responsáveis pela distribuição de informações para outros ASs são os AS Border Routers. Os roteadores de fronteira no AS Border Routers (AS) que executam o OSPF aprendem as rotas exteriores por meio de protocolos Exterior Gateway Protocols (EGPs), tais como o Border Gateway Protocol (BGP).



- **Backbone router:** um roteador que possui interface(s) para o backbone.
- **Area Border Router:** está conectado a múltiplas áreas e executa várias cópias do algoritmo de roteamento para cada área em que está conectado.
- **Internal Router:** um roteador que está conectado a redes pertencentes à mesma área; executa apenas uma única cópia do algoritmo de roteamento.

- **Autonomous System Border Router:** um roteador que troca informações de roteamento com roteadores pertencentes a outros sistemas autônomos.

Pacotes de Estado de Enlace

- Router Links.
- Network Links.
- Summary Links.
- External Links.

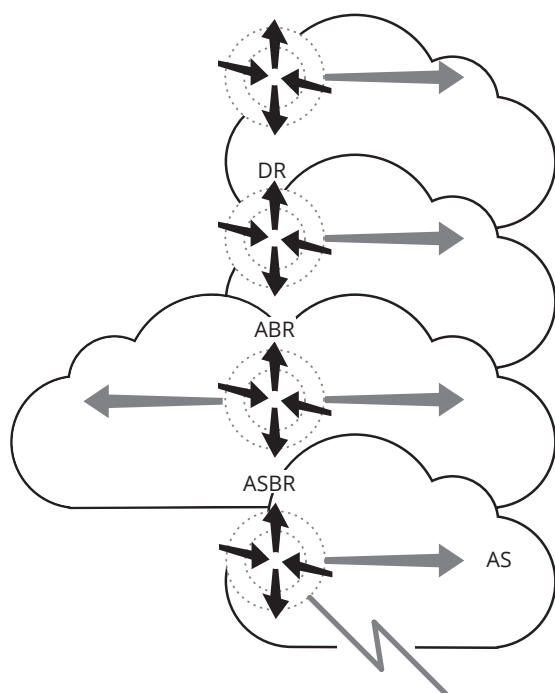


Figura 3.9
Pacotes de Estado de Enlace.

Os diferentes tipos de pacotes de Estado de Enlace mais comuns estão ilustrados na figura anterior. São eles:

- **Router links:** descrevem o estado e o custo dos enlaces do roteador (interfaces) para a área (intra-área).
- **Network links:** originados por um Designated Router (DR) para segmentos multiacesso com mais de um roteador conectado. Descreve todos os roteadores conectados ao segmento específico.
- **Summary links:** originados somente por ABRs (Area Border Routers). Descrevem as redes no AS fora de uma área (inter-área). Também descrevem a localização do Autonomous System Border Router (ASBR).
- **External links:** originados por um ASBR. Descreve os destinos externos ao AS ou uma rota padrão para fora do AS.

Como mostrado anteriormente, os pacotes router links são uma indicação do estado das interfaces de um roteador que pertence a certa área. Cada roteador gera um router link para todas as suas interfaces.

Pacotes summary links são gerados por ABRs. Essa é a maneira como a informação de rotas das redes é disseminada entre áreas. Normalmente, toda a informação é enviada para o backbone (área 0) e, por sua vez, o backbone a propaga para as outras áreas.

ABRs também têm a tarefa de propagar a rota para o ASBR. Essa é a maneira como os roteadores conhecem as rotas externas para outros ASs.

Pacotes network links são gerados por um DR num segmento de rede. Essa informação é uma indicação de todos os roteadores conectados a um particular segmento multiacesso, tal como redes locais Ethernet, Token ring e FDDI, além de redes NBMA.

Pacotes external links são uma indicação de redes fora do AS. Essas redes são informadas ao OSPF via redistribuição. O ASBR tem a tarefa de informar essas rotas para o AS.

OSPF – Resumo de funcionamento



Cada área tem uma cópia independente do OSPF.

Quando ligado, o roteador executa a seguinte sequência:

- Inicializa as estruturas de dados do protocolo.
- Determina as interfaces funcionais.
- Executa o protocolo Hello para conhecer seus vizinhos.
 - ▣ Em redes multiacesso (broadcast e NBMA) é eleito um DR.
- Tenta formar adjacências com seus novos vizinhos.
- Periodicamente anuncia o estado de seus enlaces.
- LSAs são propagados na área OSPF usando o algoritmo de flooding para sincronizar os bancos de dados.
- A partir do banco de dados calcula as rotas para as redes.

Uma cópia independente do algoritmo de roteamento básico do OSPF roda em cada área. Roteadores que têm interfaces para múltiplas áreas rodam múltiplas cópias do algoritmo. A seguir um resumo do funcionamento do algoritmo de roteamento.

Quando um roteador é ligado, ele primeiro inicializa as estruturas de dados do protocolo de roteamento. O roteador então aguarda a indicação dos protocolos de camadas inferiores (enlace de dados e física) de que suas interfaces estão funcionais.

O roteador usa o protocolo Hello para conhecer seus vizinhos. O roteador envia pacotes de Hello para seus vizinhos e, em troca, recebe deles pacotes de Hello. Em redes ponto a ponto e broadcast, o roteador detecta dinamicamente seus vizinhos enviando pacotes de Hello para o endereço multicast 224.0.0.5. Em redes NBMA e broadcast o protocolo de Hello elege um DR para a rede.

Dois roteadores não se tornarão vizinhos, a menos que concordem com as seguintes condições:

- **Area-id:** para dois roteadores no mesmo segmento, suas interfaces têm de pertencer à mesma área no segmento.
- **Authentication:** OSPF permite a configuração de uma senha para a área específica. Roteadores que querem ser vizinhos têm de ter a mesma senha num segmento em particular.
- **Hello and Dead Intervals:** OSPF troca pacotes Hello em cada segmento; isso é uma forma de keep alive usada pelos roteadores para conhecer a existência deles num segmento e também para eleição de um DR.

O intervalo de Hello especifica o tempo, em segundos, entre os pacotes Hello que um roteador envia numa interface OSPF. O intervalo chamado dead interval especifica o tempo (em segundos) que os pacotes Hello podem deixar de ser enviados antes que um



roteador seja declarado inativo (down). OSPF exige que esses dois intervalos sejam exatamente iguais entre dois vizinhos. Se não, os roteadores não se tornarão vizinhos.

- **Stub area flag:** dois roteadores também têm de concordar com o stub area flag (caso pertençam ambos a uma stub area) nos pacotes Hello, para que possam ser vizinhos.

O roteador tentará criar adjacências com alguns de seus novos vizinhos conhecidos. Os bancos de dados de Estado de Enlace são sincronizados entre pares de roteadores adjacentes. Em redes broadcast e NBMA, o DR determina que os roteadores se tornarão adjacentes. Adjacências controlam a distribuição de informação de roteamento. Atualizações de rotas são enviadas e recebidas somente entre roteadores adjacentes.

Um roteador periodicamente anuncia seu estado, também chamado de Estado de Enlace. O Estado de Enlace também é anunciado quando o estado de um roteador muda. As adjacências de um roteador estão descritas no conteúdo dos LSAs. Esse relacionamento entre adjacências e Estado de Enlace permite que o protocolo detecte rotas inativas de tempos em tempos.

LSAs são propagadas na área por flooding (inundação). O algoritmo de flooding é confiável, garantindo que todos os roteadores numa área tenham exatamente o mesmo banco de dados de Estado de Enlace. Esse database consiste em um conjunto de LSAs originados de cada roteador pertencente à área.

A partir desse banco de dados, cada roteador calcula os caminhos mais curtos (shortest-path tree), sendo ele mesmo a raiz. Esses caminhos mais curtos permitem a montagem da tabela de roteamento do protocolo OSPF.

Autenticação OSPF

- Autenticação nula, por default.
- Autenticação com senha simples (simple password):
 - Senha (chave) por área.
 - Todos os roteadores da área têm a mesma senha.
 - Método vulnerável a sniffers.
- Autenticação Message Digest (MD-5):
 - Autenticação criptográfica.
 - Uma key (senha) e uma key-id por roteador.
 - Algoritmo baseado no pacote OSPF, key e key-id.
 - Gera uma message digest adicionada ao pacote.



É possível autenticar os pacotes OSPF de forma que os roteadores possam participar do roteamento baseado em senhas (password) pré-definidas. Por default, um roteador usa autenticação nula, que significa que as informações de roteamento trocadas na rede não são autenticadas. Isso torna a rede vulnerável a ataques que informem rotas falsas.

Existem dois métodos de autenticação: Simple Password e Message Digest (MD-5).

Simple Password Authentication

A autenticação com senha simples permite que uma senha (chave) seja configurada por área. Roteadores na mesma área que quiserem participar do roteamento terão de ser configurados com a mesma chave. A desvantagem desse método é que ele é vulnerável a ataques passivos, porque qualquer um com um sniffer pode capturar a senha da rede.

Para habilitar autenticação com senha, use os comandos:

```
ip ospf authentication-key <key> (no modo de configuração da interface conectada à área)
```

```
area area-id authentication (após o comando router ospf <process-id>)
```

Exemplo:

```
interface Ethernet0  
  
ip address 10.10.10.10 255.255.255.0  
ip ospf authentication-key mypassword  
router ospf 10
```

```
network 10.10.0.0 0.0.255.255 area 0 (note que aqui não é subnet mask, mas a máscara que indica a rede 10.10.0.0/16).
```

```
area 0 authentication
```

Message Digest Authentication

Autenticação Message Digest é uma autenticação criptográfica. Uma key (senha) e uma key-id são configuradas em cada roteador. O roteador usa um algoritmo baseado no pacote OSPF, a key e a key-id para gerar uma message digest, que é adicionada ao pacote no final dele.

Ao contrário da autenticação simples, a chave não é trocada entre os roteadores. Um número de sequência é incluído em cada pacote OSPF para proteção contra ataques de repetição (replay attacks). Esse método também permite transições sem interrupção entre chaves. Isso é muito útil para administradores que desejam trocar a senha OSPF sem interromper o funcionamento da rede. Se uma interface for configurada com uma nova chave, o roteador enviará múltiplas cópias do mesmo pacote, cada uma com uma chave de autenticação diferente. O roteador vai parar de enviar pacotes duplicados assim que detectar que todos os seus vizinhos já adotaram a nova chave.

Para habilitar autenticação com message digest, use os comandos:

```
ip ospf message-digest-key key-id md5 key (no modo de configuração da interface conectada à área)  
  
area area-id authentication message-digest (após o comando router ospf <process-id>)
```

Exemplo:

```
interface Ethernet0  
  
ip address 10.10.10.10 255.255.255.0  
ip ospf message-digest-key 10 md5 mypassword  
router ospf 10  
  
network 10.10.0.0 0.0.255.255 area 0  
area 0 authentication message-digest
```

Backbone OSPF

- Todas as áreas devem estar conectadas ao backbone.
- O backbone deve ser o ponto de partida do projeto.

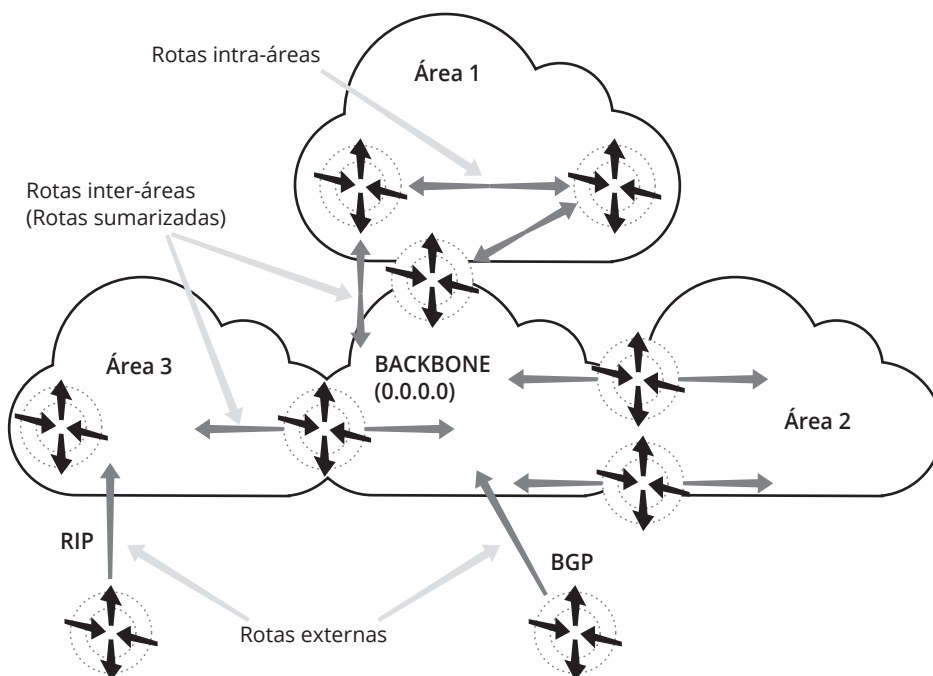


Figura 3.10
Backbone OSPF.

OSPF tem restrições especiais quando múltiplas áreas estão envolvidas. Se mais de uma área for configurada, uma dessas áreas tem de ser a área 0 (zero). Ela é chamada backbone (espinha dorsal). No projeto de redes, uma boa prática é iniciar com a área 0 e expandir para as outras áreas depois. O backbone tem de estar no centro de todas as áreas, ou seja, todas as áreas têm de estar fisicamente conectadas ao backbone, porque o OSPF espera que todas as áreas propaguem informações de roteamento para o backbone e este, por sua vez, propagará essas informações para as outras áreas. A figura acima ilustra o fluxo de informação numa rede OSPF.

Observe que todas as áreas estão conectadas ao backbone. Quando acontece de uma área nova não poder ser fisicamente conectada ao backbone, um enlace virtual terá que ser configurado.

As rotas que têm origem e destino na mesma área são chamadas intra-area routes e são representadas pela letra O na tabela de roteamento IP. Rotas originadas de outras áreas são chamadas inter-area routes ou summary routes e são representadas por O IA na tabela de roteamento IP.

Rotas originadas de outros protocolos de roteamento (ou de diferentes processos OSPF) e que são propagadas para o OSPF através de redistribuição são chamadas external routes e são representadas por O E2 ou O E1 na tabela de roteamento IP.

Se existirem múltiplas rotas para o mesmo destino, a ordem de preferência é: intra-area, inter-area, external E1 e external E2.



Para pensar

Enlaces virtuais (virtual links) são usados para conectar áreas que não têm conexão física com o backbone. O enlace virtual tem de passar por outra área que sirva de “ponte” entre a área em questão e o backbone. Também podem ser usados para particionamento do backbone.

Rotas externas do tipo 2 são aquelas nas quais o custo é sempre o custo externo, independente do custo interno para aquela rota.

Rotas externas do tipo 1 são aquelas nas quais o custo é a soma do custo externo com o custo interno para aquela rota. Uma rota do tipo 1 tem sempre preferência sobre uma rota do tipo 2 para o mesmo destino.

Layout dos pacotes OSPF

Todo pacote OSPF começa com um cabeçalho (header) de 24 bytes. O cabeçalho contém toda a informação necessária para determinar se o pacote deve ser aceito para processamento.

0	7 8	15 16	31
Version	Type	Packet length	
Router ID			
Area ID			
Checksum		AU type	
Authentication			
Authentication			

Figura 3.11
Cabeçalho do
pacote OSPF.

Os campos do cabeçalho são os seguintes:

- **Version #:** número da versão do OSPF; o RFC2328 especifica a versão 2.
- **Type:** tipo do pacote OSPF; tipo 1) Hello; tipo 2) descrição do banco de dados (database description); tipo 3) pedido de Estado de Enlace (Link State Request); tipo 4) atualização de Estado de Enlace (Link State Update); tipo 5) reconhecimento de Estado de Enlace (Link State Acknowledgement).
- **Packet Length:** tamanho do pacote OSPF em bytes, incluindo o cabeçalho padrão.
- **Router ID:** identificação do roteador que originou o pacote.
- **Area ID:** número de 32 bits que identifica a área à qual esse pacote pertence. Todos os pacotes OSPF são associados a uma única área. A maioria atravessa um único hop somente. Pacotes que atravessam um enlace virtual são identificados com 0.0.0.0, que é a identificação da área do backbone (backbone Area ID).
- **Checksum:** soma de verificação padrão do IP a partir do início do cabeçalho do pacote OSPF, inclusive o campo de autenticação de 64 bits. A soma verificadora é considerada parte do processo de autenticação do pacote.
- **AuType:** identifica o procedimento de autenticação a ser aplicado ao pacote OSPF.
- **Authentication:** campo de 64 bits para uso do esquema de autenticação.

Pacotes Hello são pacotes OSPF tipo 1. Esses pacotes são enviados periodicamente para todas as interfaces (incluindo enlaces virtuais) com o objetivo de estabelecer e manter



relacionamentos de vizinhança. Além disso, pacotes Hello usam endereço multicast nas redes que têm capacidade multicast/broadcast, permitindo a descoberta dinâmica de roteadores vizinhos.

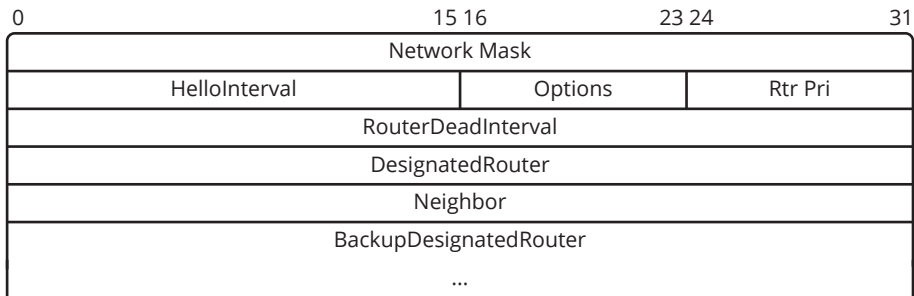


Figura 3.12
Pacote Hello
(sem cabeçalho).

Todos os roteadores conectados numa mesma rede têm de negociar certos parâmetros (Network mask, HelloInterval e RouterDeadInterval). Esses parâmetros são incluídos nos pacotes Hello, para que eventuais diferenças não possam impedir a formação de relacionamentos de vizinhança.

Os campos do pacote Hello são:

- **Network Mask:** máscara de sub-rede associada à essa interface.
- **HelloInterval:** número de segundos decorridos entre os pacotes Hello.
- **Options:** opções suportadas pelo roteador.
- **Rtr Pri:** prioridade desse roteador. Usado na eleição de DR e BDR; se inicializada com 0 (zero), esse roteador será inelegível para DR ou BDR.
- **RouterDeadInterval:** número de segundos decorridos para que se declare um roteador silencioso fora do ar (down).
- **Designated Router:** identificação do DR dessa rede, do ponto de vista do roteador que está enviando o pacote. O DR é identificado aqui pelo endereço IP da sua interface nessa rede. Se for 0.0.0.0, não existe DR.
- **Backup Designated Router:** identificação do BDR dessa rede, do ponto de vista do roteador que está enviando o pacote. O BDR é identificado aqui pelo endereço IP da sua interface nessa rede. Se for 0.0.0.0, não existe BDR.
- **Neighbor:** as identificações (Router IDs) de cada roteador que recebeu pacotes Hello recentemente na rede. Recentemente significa tempo menor do que o RouterDeadInterval.

Pacotes de descrição do banco de dados são pacotes OSPF tipo 2. Esses pacotes são trocados quando uma adjacência está sendo inicializada. Eles descrevem o conteúdo do banco de dados de Estado de Enlace. O procedimento é do tipo master slave, onde o roteador designado como master envia os pacotes e o outro (slave) recebe e envia reconhecimentos como resposta; as respostas são relacionadas aos pacotes enviados via DD sequence number.

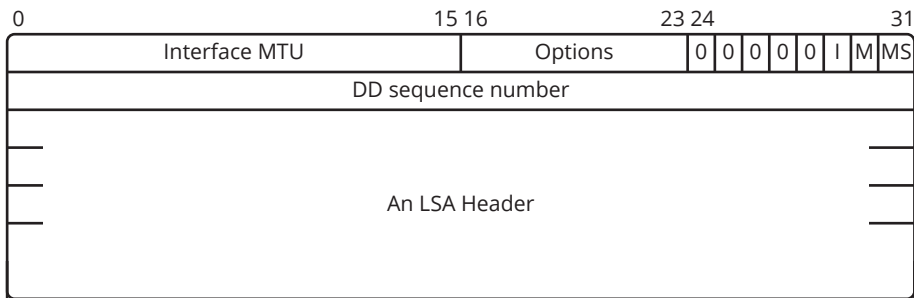


Figura 3.13
Pacote de descrição
do banco de dados
(sem cabeçalho).

Os campos do pacote Database description são:

- **Interface MTU:** o tamanho em bytes do maior datagrama IP que pode ser enviado nessa interface, sem fragmentação. No RFC1191, tabela 7-1, há uma lista das MTUs em uso na internet.
- **Options:** opções suportadas pelo roteador.
- **I-bit:** o bit de Init; quando ligado (valor 1), indica que esse pacote é o primeiro da sequência de pacotes de descrição do banco de dados; os 5 bits anteriores a este precisam ter valor zero.
- **M-bit:** o bit de More; quando ligado (valor 1), indica que mais pacotes de descrição do banco de dados virão em seguida.
- **MS-bit:** o bit master/slave; quando ligado (valor 1), indica que esse roteador é o master no processo de troca de informações do banco de dados; caso contrário, esse roteador é o slave.
- **DD sequence number:** usado para numerar o conjunto de pacotes de descrição do banco de dados; ele é incrementado a partir do valor único que vai no pacote que tem o bit-I ligado; a partir daí todos os pacotes são numerados até que toda a descrição do banco de dados tenha sido enviada.

O resto do pacote (An LSA Header) é uma lista das partes do banco de dados de Estado de Enlace. Cada LSA existente no banco de dados é descrito pelo seu cabeçalho.

0	15	16	23	24	31
LS age		Options		LS type	
Link State ID					
Advertising Router					
LS sequence number					
LS checksum			Length		

Figura 3.14
Cabeçalho do LSA.

Campos do cabeçalho do LSA:

- **LS age:** o tempo em segundos desde que o LSA foi criado.
 - **Options:** as opções suportadas por essa parte do domínio de roteamento.
 - **LS type:** o tipo do LSA; cada LSA tem um formato específico. Os tipos são: 1) Router LSAs; 2) Network LSAs; 3) Summary LSAs (IP network); 4) Summary LSAs (ASBR); 5) AS-External LSAs.
 - **Link State ID:** esse campo identifica a porção do ambiente internet (domínio de roteamento) que está sendo descrita nesse LSA. O conteúdo depende do tipo de LSA; por exemplo, nos Network LSAs esse campo contém o endereço IP da interface do DR da rede.
 - **Advertising Router:** a Router ID do roteador que originou o LSA; por exemplo, nos Network LSAs esse campo contém a Router ID do DR da rede.
 - **LS sequence number:** numera os LSAs para detectar os mais antigos ou duplicados.
 - **LS checksum:** a soma verificadora Fletcher do conteúdo completo do LSA, incluindo o cabeçalho LSA, mas excluindo o campo *LS age*.
 - **Length:** o tamanho em bytes do LSA, incluindo os 20 bytes do cabeçalho.
- Pacote Link State Request (sem cabeçalho).
 - Pacote Link State Update (sem cabeçalho).
 - Pacote Link State Acknowledgement (sem cabeçalho).



Link State Request (Pedido de Estado de Enlace) são pacotes OSPF do tipo 3. Após a troca de pacotes de descrição de banco de dados com um roteador vizinho, o roteador pode achar

que partes do seu banco de dados de Estado de Enlace estão desatualizadas. Um ou mais pacotes de Pedido de Estado de Enlace são usados para solicitar as partes do banco de dados do vizinho que estejam mais atualizadas.

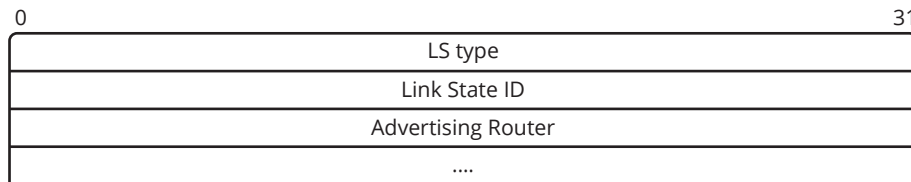


Figura 3.15
Pacote Link State Request (sem cabeçalho).

O roteador que solicita essas partes sabe exatamente do que está precisando. Cada parte é identificada pelos campos *LS sequence number*, *LS checksum* e *LS age*.

Cada pedido enviado é identificado pelos campos *LS type*, *Link State ID* e *Advertising Router*, que identifica o LSA, mas não a sua instância. Os pacotes Link State Request são entendidos como pedidos da instância mais recente (seja ela qual for). Os campos desse pacote já foram descritos.

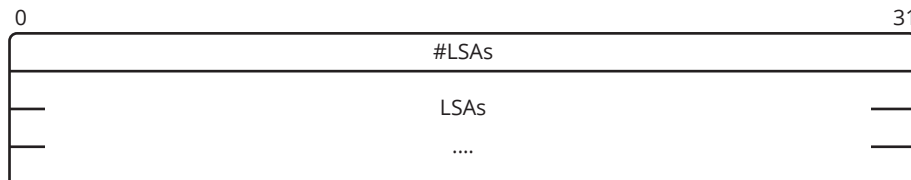


Figura 3.16
Pacote Link State Update (sem cabeçalho).

Pacotes Link State Update (Atualização de Estado de Enlace) são pacotes OSPF tipo 4. Esses pacotes implementam o algoritmo de flooding (inundação) de LSAs. Cada pacote transporta um conjunto de LSAs, um hop além da sua origem. Um pacote pode conter vários LSAs.

Campos do pacote Link State Update:

- **# LSAs:** número de LSAs contidos no pacote. O conteúdo do pacote é uma lista de LSAs, cada um iniciando com um cabeçalho de 20 bytes descrito anteriormente.
- **LSA:** lista de LSAs.

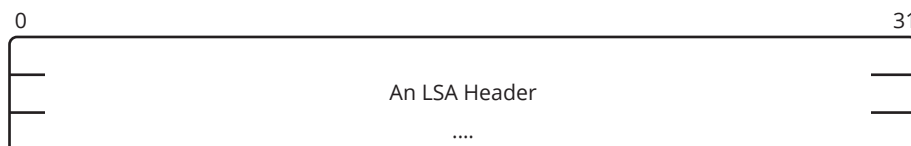


Figura 3.17
Pacote Link State Acknowledgement (sem cabeçalho).

Link State Acknowledgement (Reconhecimento de Estado de Enlace) – Pacotes OSPF tipo 5. Para garantir a confiabilidade do processo de flooding, os LSAs enviados são explicitamente reconhecidos. O conteúdo desse pacote é simplesmente uma lista de cabeçalhos LSAs.



Consulte o OSPF Design Guide – Cisco Systems.



Roteiro de Atividades 3

Atividade 3.1 – Configuração do protocolo OSPF

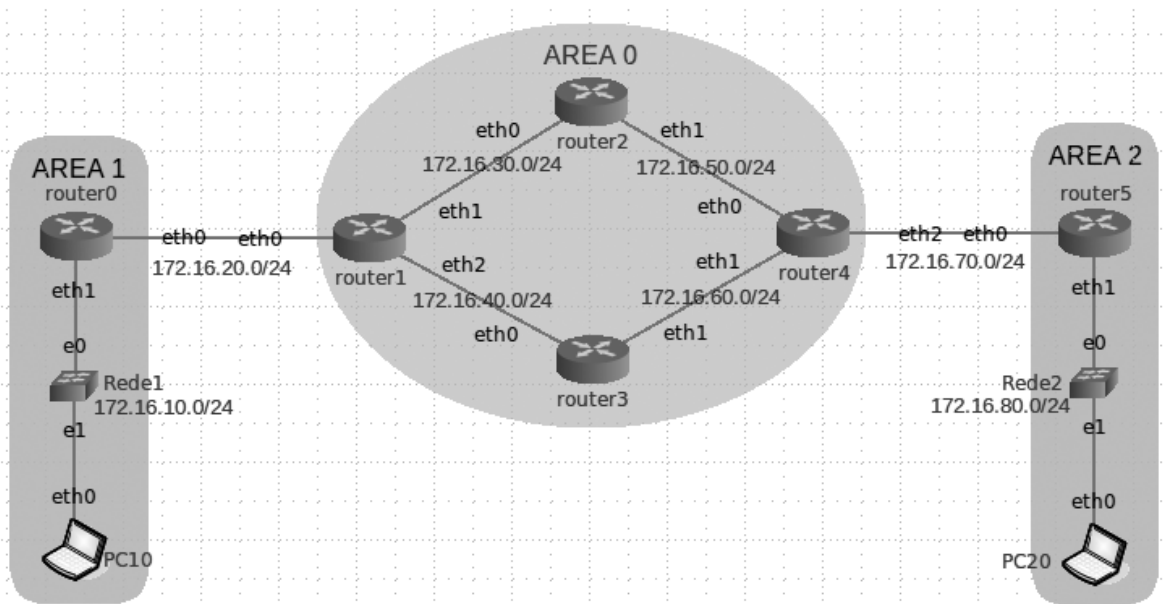
Nesta atividade vamos configurar uma rede OSPF com 3 áreas, inclusive o backbone. Serão revistos os conceitos teóricos sobre a arquitetura OSPF.

- Carregar a rede *Rede1_Sessao3_ADR8.imn* no Core.
- Iniciar o Wireshark na interface eth0 do router0.
- Configurar o protocolo OSPF nos roteadores.
- Analisar o fluxo de pacotes.
- Verificar as tabelas de rotas dos roteadores.
- Verificar a continuidade entre os PCs.

Para esta atividade utilizaremos a *Rede1_Sessao3_ADR8.imn*, composta por três áreas: área 0 (backbone), área 1 e área 2.

- A área 1 tem as redes 172.16.10.0/24 e 172.16.20.0/24.
- A área 0 tem as redes 172.16.30.0/24, 172.16.40.0/24, 172.16.50.0/24 e 172.16.60.0/24.
- A área 2 tem as redes 172.16.80.0/24 e 172.16.90.0/24.
- Os roteadores router1 e router4 são Area Border Router (ABR), os roteadores router0 e router5 são Internal Router e os roteadores router2 e router3 são Backbone Router.

Figura 3.18
Rede1_Sessao3_
ADR8.imn.



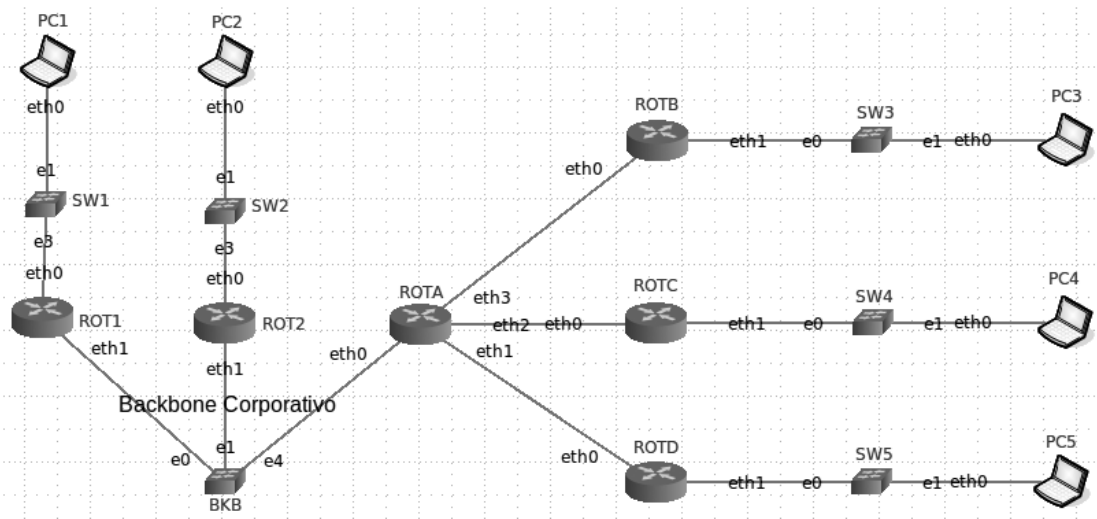
Atividade 3.2 – Projeto e configuração do protocolo OSPF

Objetivos desta atividade:

- Fazer o projeto das áreas OSPF.
- Configurar o protocolo OSPF em todos os roteadores.
- Verificar as tabelas de rotas dos roteadores.
- Verificar a conectividade entre as diversas redes.

A rede a seguir representa uma empresa com uma matriz e três filiais remotas. A matriz tem três redes físicas, representadas pelos switches SW1, SW2 e BKB (Backbone Corporativo). As filiais remotas possuem uma rede física cada uma, representadas pelos switches SW3, SW4 e SW5. As interfaces de todos os equipamentos já estão configuradas, de acordo com a tabela fornecida no enunciado do problema. Falta definir as áreas OSPF e configurar o protocolo OSPF nos roteadores.

Figura 3.19
Rede2_Sessao3_
ADR8.imn.



As seguintes redes foram atribuídas às respectivas redes físicas, conforme a tabela a seguir.

Rede física	Identificação da rede	Máscara de rede
SW1	200.130.24.0/24	255.255.255.0
SW2	200.130.25.0/24	255.255.255.0
BKB	192.168.0.0/24	255.255.255.0
ROTA – ROTB	192.168.1.0/24	255.255.255.0
ROTA – ROTC	192.168.2.0/24	255.255.255.0
ROTA – ROTD	192.168.3.0/24	255.255.255.0
SW3	200.130.26.0/24	255.255.255.0
SW4	200.130.27.0/24	255.255.255.0
SW5	200.130.28.0/24	255.255.255.0

Conclusão

Nestas atividades práticas aprendemos a:

- Planejar endereçamento IP de redes que usam o protocolo OSPF;
- Configurar o protocolo OSPF;
- Analisar o fluxo de mensagens OSPF;
- Verificar a conectividade da rede.

4

Protocolo de roteamento BGP4 – Parte 1

objetivos

Conhecer os conceitos básicos do protocolo Border Gateway Protocol versão 4 (BGP-4) e sua configuração.

Protocolos Exteriores (EGP), Routing Information Base (RIB), Vizinhos e pares BGP e Atributos BGP.

conceitos

Histórico

- Roteamento interno – Interior Gateway Protocol (IGP).
 - ▣ RIP, OSPF, IS-IS, IGRP e EIGRP.
- Roteamento externo – Exterior Gateway Protocol (EGP).
 - ▣ BGP-4 (RFC 4271).
- Backbone Arpanet.
 - ▣ Core router.
 - ▣ Non-core router.
 - ▣ Gateway to Gateway Protocol (GGP).

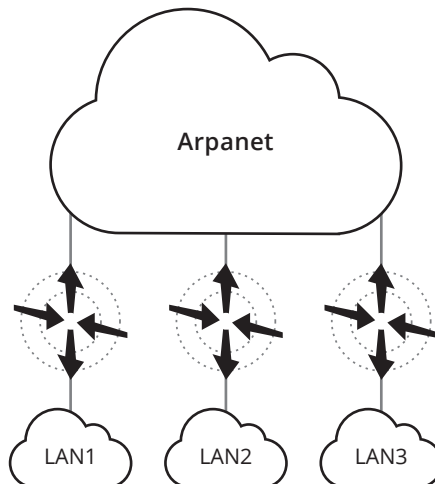


Figura 4.1
Backbone ARPANet.

Os roteadores utilizados para trocar informações dentro de Sistemas Autônomos são chamados de roteadores internos (interior routers) e podem utilizar uma variedade de protocolos de roteamento interno (Interior Gateway Protocols – IGP). Entre eles estão: RIP, OSPF, IS-IS, IGRP e EIGRP.

Os dois primeiros são padronizados pelo IETF (RFC2453 e RFC2328) e já foram vistos nas unidades anteriores. O IS-IS (ISO 10589) é um protocolo padrão do modelo OSI. Os protocolos IGRP e EIGRP são proprietários da Cisco.

Roteadores que trocam dados entre Sistemas Autônomos são chamados de roteadores externos (exterior routers) e utilizam o Exterior Gateway Protocol (EGP) ou o BGP (Border Gateway Protocol). Para esse tipo de roteamento são consideradas basicamente coleções de prefixos Classless Inter Domain Routing (CIDR), identificados pelo número de um Sistema Autônomo.

O BGP (RFCs 4271, 1772, 1773, 1930, 1997, 2119, 2858, 3065, 4456), assim como o EGP, é um protocolo de roteamento interdomínios, criado para uso nos roteadores principais da internet.

O BGP foi projetado para evitar loops de roteamento em topologias arbitrárias, que era o problema mais sério de seu antecessor, o EGP (Exterior Gateway Protocol). Outro problema que o EGP não resolve — e é abordado pelo BGP — é o do roteamento baseado em política (policy-based routing), um roteamento com base em um conjunto de regras não técnicas, definidas pelos Sistemas Autônomos.

A última versão do BGP, o BGP4, foi projetada para suportar os problemas causados pelo grande crescimento da internet.

Para pensar

Parafraseando Douglas Comer, o BGP-4 atualmente é “a cola que mantém a internet unida e permite a interconexão universal”.



O BGP-4 possibilita o intercâmbio de informações de roteamento entre os diversos sistemas autônomos, ou ASs (Autonomous Systems), que em conjunto formam a internet. Explicando de uma forma simplificada: ele permite que os dados trafeguem entre os ASs até chegar ao AS de destino, e dentro dele siga até o seu destino final (host). A seguir vamos rever um breve histórico da criação da internet, com o objetivo de justificar a necessidade do protocolo BGP-4.

Há alguns anos, quando o principal backbone da internet era a Arpanet, as instituições de pesquisa conectadas à rede precisavam gerenciar manualmente as tabelas de rotas para todos os possíveis destinos, ou seja, todas as outras redes conectadas. Com o crescimento da internet, verificou-se que era impraticável manter todas as tabelas atualizadas dessa forma, e que eram necessários mecanismos de atualização automática. Os pesquisadores da internet optaram, então, por usar uma arquitetura que consistia de um reduzido e centralizado grupo de roteadores (core routers) que tinham, em suas tabelas, as rotas para todos os possíveis destinos da internet; e um outro grupo maior de roteadores que possuíam em suas tabelas apenas informações (rotas) parciais, e não para toda a internet. Os core routers eram administrados pelo Internet Network Operations Center (INOC), e o grupo maior de roteadores externos ficou conhecido pelo termo noncore routers (roteadores fora do núcleo), que conectavam as redes locais das instituições de pesquisa ao backbone da Arpanet. Foi desenvolvido, então, o protocolo Gateway-to-Gateway Protocol (GGP), que foi usado nos core routers para atualização automática das tabelas de rotas entre eles.

O GGP era um protocolo baseado no algoritmo de vetor de distância (Vector-Distance, também conhecido como Bellman-Ford).

Essa arquitetura tem, tecnicamente, graves pontos fracos, principalmente com relação à sua capacidade de expansão, e a internet acabou crescendo muito, indo além de um único backbone gerenciado de forma centralizada. Verificou-se, portanto, não ser possível expandir esse backbone arbitrariamente, pelas diversas limitações técnicas.

Como o backbone de cada site pode ter uma estrutura complexa, o esquema de core routers não iria conseguir conectar todas as redes diretamente. Era necessário um novo esquema que permitisse aos noncore routers passar informações aos core routers sobre as redes que estavam “atrás” deles, além de oferecer autonomia de gerenciamento aos sites. Até o momento, estava sendo usado o conceito de interconexão, que levava em conta apenas a arquitetura do roteamento em uma internet e não contemplava as questões administrativas envolvidas.

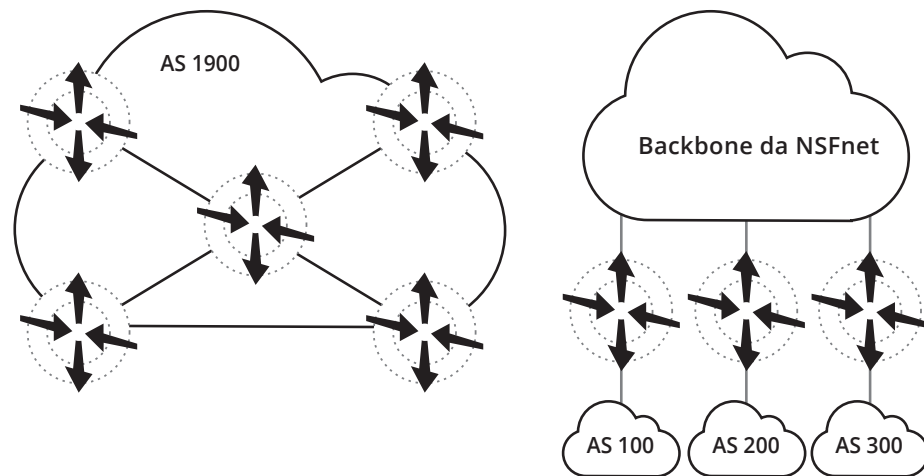


Figura 4.2
Exemplos
de Sistemas
Autônomos (AS).

Os projetistas notaram que as interconexões de um backbone com arquitetura complexa não devem ser encaradas como várias redes independentes conectadas a uma internet, mas como uma organização que controla várias redes e garante que as informações sobre as rotas internas são consistentes, e que pode escolher um de seus roteadores para fazer a ponte de comunicação para o “mundo exterior”.

Entra em cena o conceito do Sistema Autônomo (Autonomous Systems – AS), no qual as redes e roteadores estão sob o controle de uma mesma entidade administrativa. Esse conceito substituiu a ideia das redes locais conectadas ao backbone central. Cada AS tem a liberdade de escolher o esquema e a arquitetura mais adequados para si para descobrir, propagar, validar e verificar a consistência das suas rotas internas e a responsabilidade de anunciar para os outros ASs as rotas para suas redes internas invisíveis. A figura acima ilustra o conceito de Sistema Autônomo (AS 1900).

Para anunciar as rotas para suas redes internas entre si, os ASs precisavam concordar em usar um esquema único, como um mesmo idioma por toda a internet. Para permitir que um algoritmo de roteamento automatizado pudesse distinguir entre um AS e outro, foi designado a cada AS um número (Autonomous System Number – ASN) por uma autoridade central, a Internet Assigned Numbers Authority – IANA (<http://www.iana.org/>) – encarregada de atribuir todos os endereços identificadores das redes conectadas à internet. A figura do backbone da National System Foundation Network (NSFnet) ilustra a interconexão de ASs.

Exterior Gateway Protocol (EGP):

- Vizinhos internos.
- Vizinhos externos.

Problemas:

- Loops de roteamento.
- Pouca flexibilidade para políticas de roteamento.

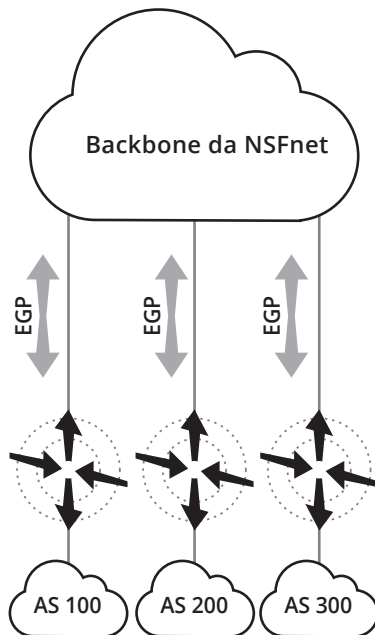


Figura 4.3
Backbone da
NSFnet que
interliga ASs.

Dois roteadores que pertençam ao mesmo AS são considerados “vizinhos internos” (interior neighbors). Se ambos pertencerem a ASs diferentes e trocarem informações de roteamento entre si, são considerados “vizinhos externos” (exterior neighbors). O protocolo de roteamento usado pelos exterior neighbors é o Exterior Gateway Protocol ou simplesmente EGP (RFC 904). É ele que permite o anúncio das rotas para as redes internas do AS para o núcleo (core) da internet, como mostra a figura anterior.

Com o tempo, o EGP apresentou diversas limitações técnicas e potenciais problemas ao ser usado na internet. Apesar das tentativas para produzir novas versões (EGP2 e EGP3) do protocolo, os projetistas não obtiveram sucesso, por haver a necessidade de muitas alterações fundamentais na estrutura do protocolo.

O EGP apresentou deficiências insustentáveis, como restrições em topologia, incapacidade de evitar loops de roteamento e pouca flexibilidade para a configuração de políticas de roteamento.

Um grande desafio para os projetistas era transformar uma arquitetura internet que não dependesse de um sistema centralizado (core routers), deixando uma topologia organizada hierarquicamente e iniciando outra com estrutura diferente. Além disso, tinha o desafio de fazer uma arquitetura internet suportar uma forma de colaboração mais próxima entre certos ASs do que entre outros. Isso levou os engenheiros do IETF a desenvolverem uma solução para esses problemas através de um novo, mais moderno e mais robusto protocolo de roteamento externo, como veremos.

Border Gateway Protocol (BGP-4)



- Sucessor do EGP.
- Roteamento entre ASs.
- Suporta CIDR.
- Interage com IGPs: RIP, OSPF etc.
- Usa TCP porta 179.
- Estabelecimento de uma sessão BGP.
 - ▣ Estabelecimento da conexão TCP entre os roteadores.
 - ▣ Envio da tabela de rotas completa só uma vez.
 - ▣ Atualização parcial da tabela (incremental).
 - ▣ Mensagens de Keep-Alive para manter a sessão aberta.

O BGP é um protocolo de roteamento para ser usado entre múltiplos sistemas autônomos em redes baseadas no protocolo TCP/IP. O BGP-4 (RFCs 4271, 1772) tornou-se o sucessor natural do EGP, efetivamente atacando suas deficiências mais sérias, ou seja, evitando loops de roteamento e permitindo o uso de políticas de roteamento entre ASs baseadas em regras arbitrárias por eles definidas. Além disso, o BGP-4 foi a primeira versão do BGP a suportar endereços agregados (Classless Interdomain Routing ou simplesmente CIDR) e o conceito de supernets.

O protocolo BGP-4 assume que o roteamento interno do AS é feito através de um sistema IGP (Interior Gateway Protocol) de roteamento interno. Esse pode ser um protocolo de roteamento como o RIP, OSPF, IGRP ou EIGRP; ou até mesmo através de rotas estáticas. O BGP constrói um gráfico dos ASs, usando as informações trocadas pelos “vizinhos BGP” (BGP neighbors), que são compostas dos números identificadores dos ASs, os ASN. A conexão entre ASs forma um “caminho” (path), e a coleção desses caminhos acaba formando uma rota composta pelos números dos ASs que devem ser percorridos até se chegar a um determinado AS de destino.

O BGP faz uso do TCP (porta 179) para o transporte das informações de roteamento, de modo que ele próprio não precisa preocupar-se com a correta transmissão das informações.

Outra característica do BGP-4 é a atualização das tabelas de rotas feitas de forma incremental, como nos algoritmos de estado de enlace. A atualização completa da tabela de rotas é feita somente uma vez, quando se estabelece a sessão entre os neighbors (vizinhos) ou peers (pares).

Para o estabelecimento de uma sessão BGP entre neighbors ou peers, basicamente, os seguintes passos são executados:

- É estabelecida a conexão TCP entre os dois roteadores que trocam mensagens de abertura da sessão e negociam os parâmetros de operação.
- O primeiro fluxo de dados transmitido é a tabela completa de rotas BGP. Atualizações posteriores são feitas nessa tabela, incrementalmente, à medida que as mudanças ocorrem.
- Como não há a atualização completa da tabela após a primeira atualização, o roteador mantém a informação da versão da tabela que todos os seus peers possuem, enquanto durar a sessão entre eles. Se esta for interrompida por qualquer motivo, o processo é iniciado novamente a partir do primeiro passo.
- Mensagens de Keep-Alive são enviadas periodicamente para manter a sessão aberta.
- Mensagens de aviso são enviadas quando ocorrem erros ou outras situações especiais.



- Caso uma conexão verifique um erro, uma mensagem é enviada e a conexão fechada, encerrando a sessão.

Basicamente, o BGP serve para informar às redes externas a um AS quais são as rotas para redes possíveis de se atingir dentro de sua rede. Falando de outra forma: o propósito do BGP-4 é anunciar rotas para outras redes externas ou sistemas autônomos. Esses anúncios são como “garantias” de que os dados serão transportados para o espaço IP representado pela rota anunciada.

Se, por exemplo, um AS anunciar uma rota para 200.130.26.0/24 (na sintaxe anterior ao CIDR, esse endereço é a classe C, que começa em 200.130.26.0 e termina em 200.130.26.255), e alguém enviar dados destinados a qualquer endereço dentro dessa faixa, esse AS estará “garantindo” que sabe enviar os dados até o destino.

Routing Information Base (RIB)

O protocolo BGP assume que esteja rodando um protocolo interior no AS. Ele não tem algoritmos próprios de cálculo da tabela de rotas, apenas importa as rotas de muitas fontes de informação: rotas diretamente conectadas, estáticas, agregadas, protocolos IGP e outros protocolos EGP. A Routing Information Base armazena todas essas rotas das diversas fontes, submetendo-as, quando for o caso, às políticas de importação de rotas e aos mapas de rotas para estabelecer as rotas aceitáveis dentro dos critérios estabelecidos. A Figura 9.42 mostra como a RIB é montada a partir de todas as informações recebidas.

Note que a partir das informações armazenadas na RIB e de acordo com o processo de seleção de rotas, é montada a tabela das melhores rotas, que será usada nas decisões de roteamento pelo BGP, chamada de Forwarding Information Base (FIB). As políticas de exportação de rotas e os mapas de rotas determinam que rotas serão informadas aos protocolos interiores e aos pares BGP.

- BGP assume a existência de um protocolo IGP.
- BGP não calcula tabela de rotas.
- Importa rotas de muitas fontes de informação.
 - Diretamente conectadas.
 - Estáticas.
 - Agregadas.
 - Protocolos IGP.
 - Outros protocolos EGP.
- A RIB armazena todas essas rotas.



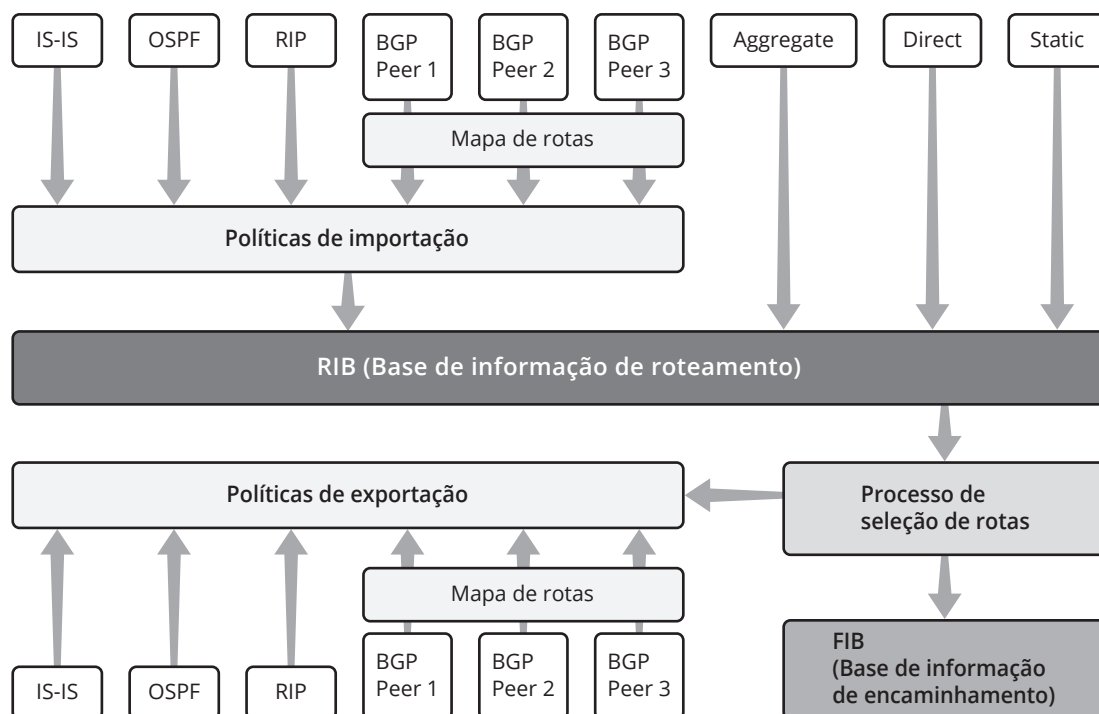


Figura 4.4
Routing Information
Base (RIB).

Observe que as políticas de importação de rotas filtram as rotas que serão armazenadas na RIB, sejam elas oriundas dos protocolos IGP (RIP, OSPF, IS-IS etc.), sejam de pares BGP. As rotas oriundas de pares BGP são filtradas adicionalmente pelos mapas de rotas, que serão descritos mais adiante. As demais rotas são armazenadas sem qualquer filtragem. Na divulgação das rotas atuam as políticas de exportação de rotas e também os mapas de rotas.

- Políticas de importação de rotas filtram as rotas que serão armazenadas na RIB.
- Políticas de exportação filtram as rotas que serão anunciadas aos BGPs pares (peer BGP).

O RFC 4271 define três RIBs para armazenamento de informações de rotas:

- **Adj-RIBs-In:** todas as rotas recebidas de um BGP peer.
- **Loc-RIB:** todas as rotas BGP válidas.
- **Adj-RIBs-Out:** as rotas anunciadas a um BGP peer.

Vizinhos e pares BGP

Sistemas (roteadores) que são “vizinhos BGP” (BGP neighbors) comunicam-se através de sessões TCP estabelecidas entre eles. Os roteadores de borda (border routers) de ASs vizinhos são considerados peers (pares). Esses pares são as fronteiras políticas dos ASs, que trocam tráfego de acordo com as regras definidas pelos ASs participantes.

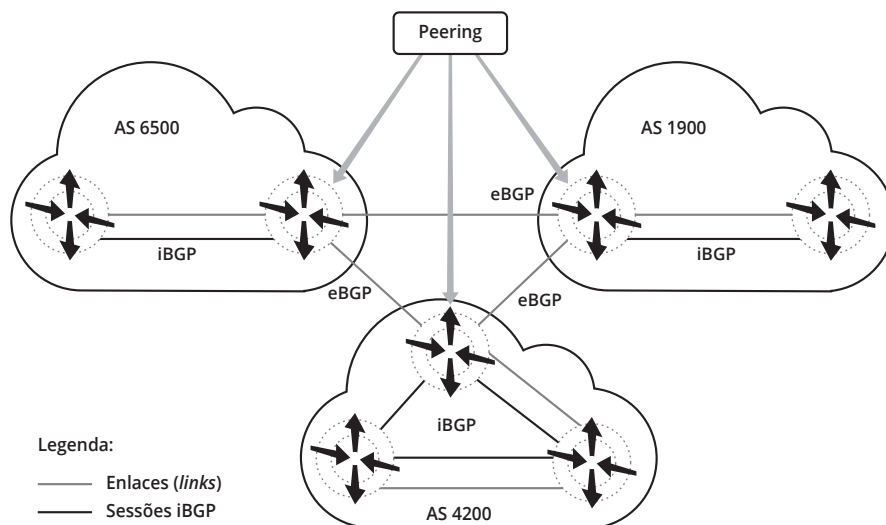


Figura 4.5
Vizinhos e pares BGP.

São chamados neighbors os sistemas BGP (roteadores) que possuem sessões BGP estabelecidas entre eles. Então, os roteadores de borda são neighbors? Sim. Porém, quando uma importância política é atribuída a eles, a forma correta de chamá-los é de peers, enquanto que os neighbors são quaisquer vizinhos BGP.

Existem outras situações em que os vizinhos BGP não são, obrigatoriamente, os roteadores entre ASs, e sim roteadores do mesmo AS. Nesse caso, as sessões estabelecidas entre eles acontecem internamente ao AS. O que permite isso é o iBGP ou internal BGP, que permite a troca de rotas no mesmo AS. De forma análoga, a troca de rotas entre ASs é feita pelo eBGP (exterior BGP). Um importante conceito do iBGP é que os neighbors não têm a obrigação de estarem diretamente conectados (como na figura acima) através de uma linha serial ou via interface Ethernet, por exemplo. Os peers, por outro lado, não podem estar conectados de outra forma que não seja a direta, seja link serial ou interface Ethernet.

O algoritmo do eBGP trabalha, basicamente, anunciando todas as rotas que conhece, enquanto o do iBGP faz o possível para não anunciar rotas. Assim, para fazer o iBGP funcionar adequadamente dentro de um AS, é necessário estabelecer sessões BGP entre todos os roteadores que “falam” iBGP (AS 4200), formando uma “malha completa” (full mesh) de sessões iBGP dentro do AS.

Atributos do BGP

- Controlam informações relativas a rotas.
- Usados pelo algoritmo para tomada de decisões.
- Representados no formato TLV.
 - ▣ Type.
 - ▣ Length.
 - ▣ Value.
- Enviados nas mensagens de Update na informação de NLRI associada.

A seguir são descritos alguns dos atributos do BGP. Para conhecer todos os atributos, ver RFC 4271.

Atributos de caminho (path attributes) são um conjunto de parâmetros que descrevem as várias características de um caminho (path) para um determinado prefixo IP de destino. Eles são muito usados no processo de seleção de rotas para a escolha da melhor rota entre várias disponíveis e também para construção das políticas de roteamento através da comparação e definição de atributos.

Quatro categorias de atributos BGP.

- **Mandatário bem conhecido:** *tem* de existir e *tem* de ser implementado por todos os protocolos BGP.
- **Discrecionário bem conhecido:** *pode* existir e *tem* de ser implementado por todos os protocolos BGP, embora não faça sentido em certas situações.
- **Transitivo opcional:** *pode* existir e *pode* ser implementado por todos os protocolos BGP; encaminhado para os outros protocolos.
- **Não transitivo opcional:** *pode* existir e *pode* ser implementado por todos os protocolos BGP; ignorado e não encaminhado para os outros protocolos.

Cada atributo BGP é representado no formato Type Length Value (TLV – Tipo Comprimento Valor) e enviado com a informação Network Layer Reachability Information (NLRI) associada, nas mensagens de Update (o formato das mensagens BGP será visto mais adiante). Há quatro categorias de atributos BGP que descrevem como processar e distribuir a informação de roteamento para os outros protocolos BGP (BGP speakers):

- **Mandatário bem conhecido** (well-known mandatory): *tem* de estar presente na mensagem de Update e *tem* de ser implementado por todos os protocolos BGP para garantir que todas as implementações BGP suportem um conjunto padrão de atributos.
- **Discrecionário bem conhecido** (well-known discretionary): *pode* estar presente na mensagem de Update e *tem* de ser implementado por todos os protocolos BGP, embora algumas vezes um particular atributo de caminho não faça sentido estar presente.
- **Transitivo opcional** (optional transitive): *pode* estar presente na mensagem de Update e *pode* ser implementado por todos os protocolos BGP; encaminhado para os outros protocolos BGP, mesmo que não seja entendido pelo protocolo BGP local.
- **Não transitivo opcional** (optional non-transitive): *pode* estar presente na mensagem de Update e *pode* ser implementado por todos os protocolos BGP; ignorado e não enviado para os outros protocolos BGP, mesmo que não seja entendido pelo protocolo BGP local.

Formato do Atributo de Caminho

Enviado no campo *Atributos de Caminho* na mensagem de Update:

- Primeiros bits descrevem a categoria.
- Demais bits contém o TLV do atributo.

Figura 4.6
Formato do Atributo
de Caminho
(Path Attribute).

W/O	N/T	C/P	E/L	Não usado (4 bits)	Código do tipo de atributo (1 byte)
Tamanho do atributo (1 ou 2 bytes)					Valor do atributo (variável)

A Figura 4.6 mostra o formato de cada atributo de caminho e como ele é enviado no campo de *Atributo de Caminho* (Path Attributes) da mensagem Update. Os primeiros 8 bits descrevem a categoria do atributo, conforme detalhado a seguir, e os demais campos contém o formato TLV do atributo.

Descrição dos campos:

- Bit 0: (W/O) Well-known (valor 0) ou Optional (valor 1).
- Bit 1: (N/T) Non-transitive (valor 0) ou Transitive (valor 1).
- Bit 2: (C/P) Complete (valor 0) ou Partial (valor 1).
 - ▣ Complete: o atributo foi enviado ao longo de todo o caminho.
 - ▣ Partial: um roteador no caminho não implementou o atributo ou a informação de roteamento pode ter sido perdida.
- Bit 3: (EL) Extended Length, define se o campo comprimento do atributo tem 1 byte (valor 0) ou 2 bytes (valor 1).
- Bits 4-7: não usados (devem estar com zeros).
- Código do tipo de atributo (Attribute Type Code).
- Tamanho do atributo (Attribute Length).
- Valor do atributo (Attribute Value).

Muitos atributos de caminho foram definidos em RFCs, mas nem todos foram implementados ou estão sendo usados atualmente. Serão discutidos aqui somente os atributos de caminho que são relevantes na internet atual.

<http://web.archive.org/web/20070323205145/http://www.riverstonenet.com/support/bgp/scalability/index.htm>

Atributo Origin

Type Code 1, Well-known Mandatory – RFC 1771.

- Indica a origem do anúncio de rota.
- 0 IGP (i) – origem interna ao AS.
- 1 EGP (e) – origem AS externo, vindo de um EGP.
- Incomplete (?) – NLRI desconhecida.
- Origin: indica a origem do anúncio de rota ou NLRI (que indica o prefixo e a máscara de bits), com relação ao AS que o originou. Pode conter um dos valores:
 1. 0 IGP: a origem é interna ao AS originário da mensagem (indicado por um “i” na tabela de rotas), seja ela recebida através da redistribuição das rotas do IGP para o BGP (daquele AS) ou pela simples configuração do BGP naquele roteador.
 2. 1 EGP: a origem é de um AS externo e foi recebida por um anúncio de um EGP. É identificada por um “e” na tabela de rotas. Esse tipo de entrada dificilmente será vista nas tabelas de rotas atualmente.
 3. Incomplete: a NLRI é desconhecida ou aprendida por outros meios (além dos citados acima). Geralmente acontece quando uma rota estática (configurada manualmente por um operador) é redistribuída no BGP e a origem da rota fica incompleta. É indicada pelo caractere “?” na tabela de rotas.



Saiba mais

Para conhecer mais alguns atributos que são usados em certas situações, veja o documento “IBGP Scalability”:



Atributo As_Path

Type Code 2, Well-known Mandatory – RFC 1771.

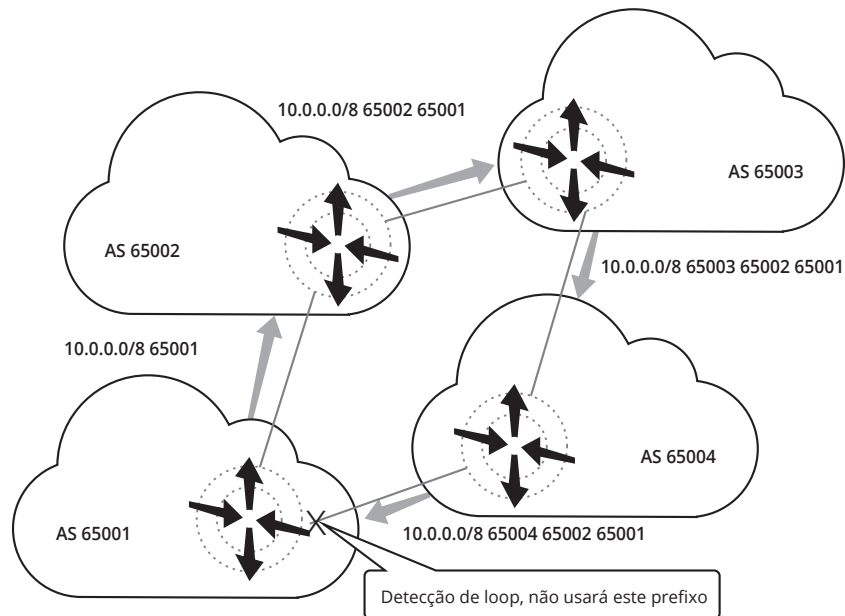


Figura 4.7
Exemplo do atributo
As_Path.

As_path é uma sequência de ASNs que uma rota cruza para alcançar uma determinada rede de destino. O AS que origina uma rota acrescenta seu ASN ao anunciar uma rota sua para seus vizinhos BGP externos. Daí para frente, cada AS que receber a rota acrescenta seu próprio ASN no início da sequência de ASNs e repassa a rota para outros peers seus, que farão o mesmo. A lista final vai representar todos os ASNs que uma rota atravessou com o ASN do AS de origem da rota no final da sequência, também conhecida como AS_Sequence. Caso um AS receba um anúncio de rota que contenha seu próprio ASN na sequência inclusa no As_Path, esse anúncio será rejeitado e descartado, garantindo que não haverá loop de roteamento na tabela BGP desse AS. Caso o As_Path seja anunciado para um vizinho do mesmo AS, a informação contida no As_Path não é alterada. A informação contida no As_Path é uma das informações usadas no processo de seleção da melhor rota para determinado destino. Ao comparar duas rotas para um mesmo destino (considerando que os outros atributos sejam idênticos), o BGP vai preferir a que possuir o As_Path menor. Caso o caminho (path) seja do mesmo tamanho, o BGP vai usar outros atributos para fazer a sua escolha da melhor rota.

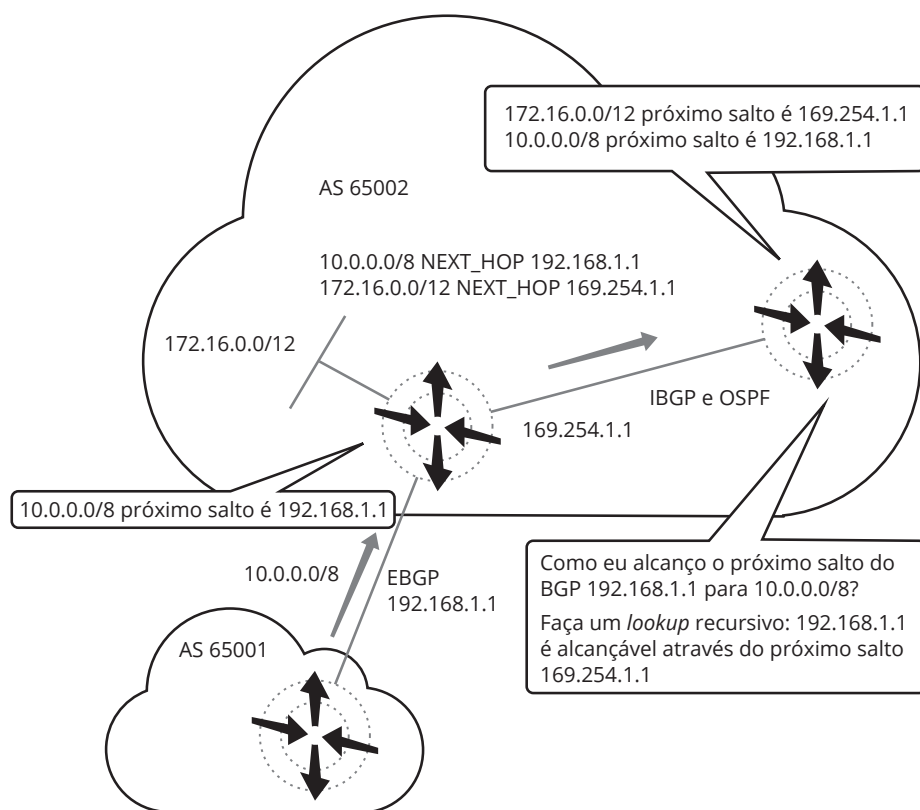
A Figura 4.7 ilustra o uso desse atributo para o prefixo de rede 10.0.0.0/8 anunciado pelo AS65001 e propagado para os ASs 65002, 65003 e 65004. Cada roteador acrescenta no início da lista de AS seu próprio número, quando faz o anúncio do prefixo para seu EBGp par. Quando o roteador que originou o anúncio detecta seu AS na lista, descarta o prefixo.

Atributo Next_Hop

Type Code 3, Well-known Mandatory – RFC 1771.



Basicamente, o atributo Next Hop recebe o endereço IP da interface do próximo roteador — próximo salto (next hop) a ser dado — para alcançar determinado destino.



Existem três situações diferentes que determinam o Next_Hop:

1. Em sessões eBGP, o Next_Hop será sempre o IP de um roteador de borda (peer BGP) de um AS vizinho que originou a rota.
2. Em sessões iBGP, onde a rota foi originada dentro do AS, o Next_Hop será o endereço IP do vizinho que anunciou a rota originalmente.
3. O Next_Hop aprendido pelo eBGP não é alterado pelo iBGP, permanecendo o endereço IP do peer eBGP que originou o anúncio da rota. Quando a rota é anunciada em mídias de multiacesso (como Ethernet e Frame Relay), o Next_Hop geralmente é o endereço IP da interface do roteador conectada à mídia que originou a rota. Existem ainda outras regras definidas pelo RFC 4271.

A Figura 4.8 mostra um exemplo desse atributo. Cada roteador precisa resolver o próximo passo BGP (BGP next hop) antes que a rota possa ser usada. Não confunda os dois next hops. O BGP next hop (o atributo Next_Hop) é transportado pelo BGP, enquanto o endereço do próximo passo (next hop address) é usado para resolver o próximo passo do BGP da FIB.

Atributo Multi_Exit_Disc

Type Code 4, Optional Non-transitive – RFC 1771.



O atributo Multi-Exit Discriminator (MED) tem como finalidade informar para os vizinhos BGP externos (peers) o melhor caminho (path) para uma determinada rota do próprio AS, influenciando-os, assim, em relação ao caminho que deve ser seguido no caso do AS possuir diversos pontos de entrada. Esse atributo influencia o tráfego entrante no AS. O MED é anunciado somente entre ASs. Porém, só o AS de origem pode fazer anúncios com valores nesse atributo, enquanto um AS vizinho que receba o atributo via mensagem Update não pode repassar o valor desse atributo a outros ASs, fazendo uso destes apenas para tomadas de decisão internas do AS.

A Figura 4.9 mostra um exemplo como MEDs podem ser usados para obter uma distribuição de carga e caminhos de backup de uma forma bem simples. Essa é uma das aplicações mais comuns dos MEDs. O usuário (AS 65002) anuncia prefixos com MEDs de valores opostos (alto e baixo) para o seu provedor (ISP), com o objetivo de direcionar o tráfego entrante para uma rota particular. Quando ambos os enlaces estiverem operando (up), a carga é compartilhada entre os enlaces. Se um enlace falhar, o roteador do provedor ainda tem um caminho backup através do roteador que está anunciando o MED com o valor mais alto.

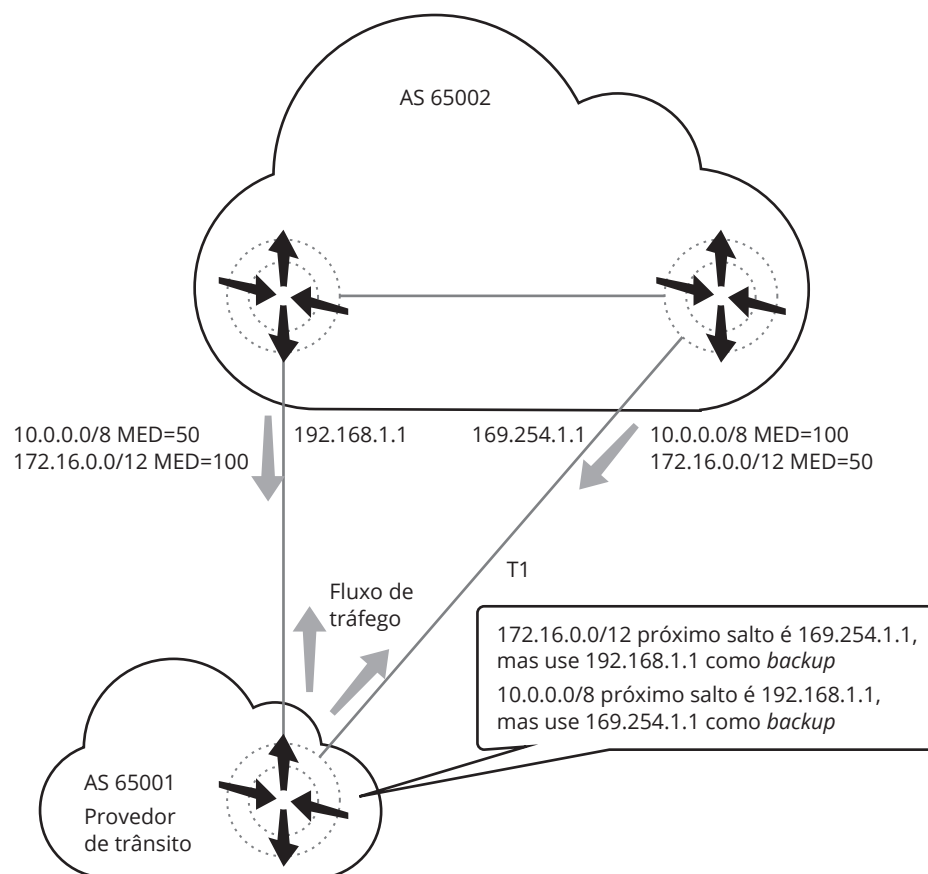


Figura 4.9
Exemplo do atributo
Multi_Exit_Disc.

A métrica do IGP também pode ser usada como MED para permitir ao IGP dinamicamente direcionar o tráfego entrante para o melhor roteador de borda através do BGP.

Atributo Local_Pref

Type Code 5, Well-known Discretionary – RFC 1771.

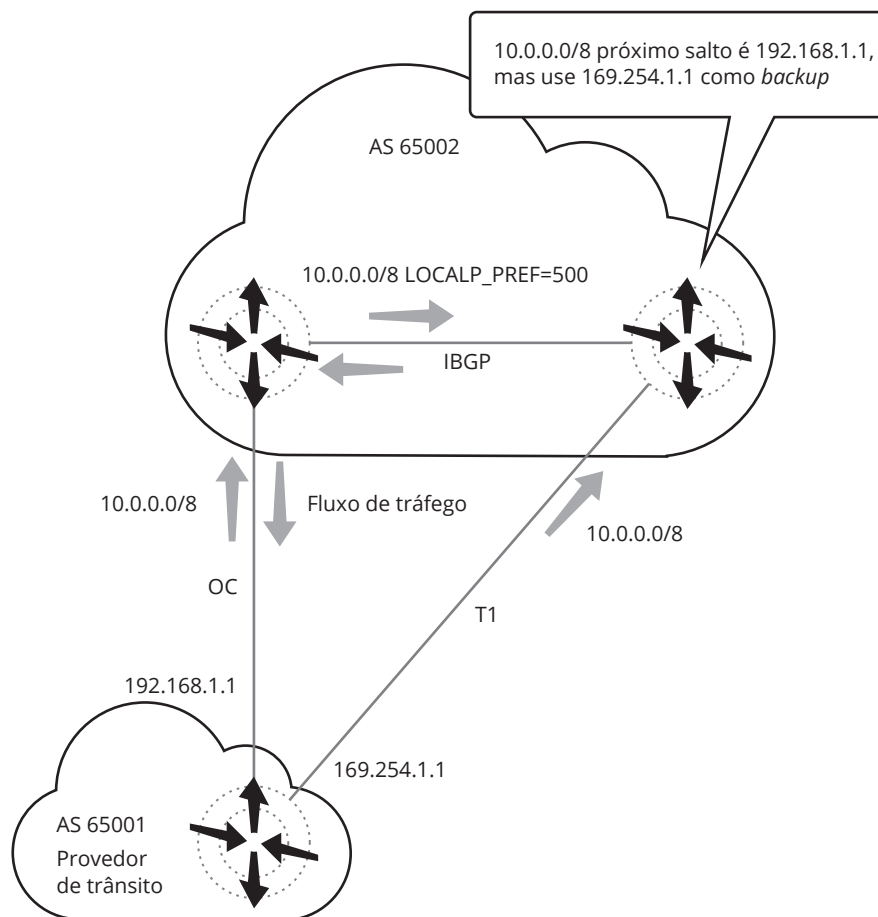


Figura 4.10
Exemplo do atributo
Local_Pref.

O atributo Local_Pref serve para anunciar o caminho preferencial de saída (de pacotes) para uma determinada rota, destinada a uma rede externa ao AS. Esse atributo influencia o tráfego que sai do AS. Como o próprio nome do atributo sugere, o Local_Pref somente é anunciado (repasado) entre os roteadores vizinhos BGP (iBGP) do mesmo AS, e não é repasado aos roteadores vizinhos externos (eBGP). Caminhos (paths) que possuem o Local_Pref com maior valor são preferidos pelo BGP. O valor padrão do Local_Pref é 100.

A Figura 4.10 mostra um exemplo como Local_Prefs podem ser usados para obter uma distribuição de carga e caminhos de backup de uma forma bem simples. Local preference é normalmente usado para escolher uma saída particular de um AS. No exemplo, o usuário (AS 65002) está conectado ao provedor por um enlace OC3 (155 MBPS) e um enlace T1 (1.5 MBPS) de backup. Todo o tráfego que sai da rede do usuário deveria usar o enlace OC3, a menos que o enlace falhe. O usuário define um alto valor de local preference nas rotas recebidas pelo enlace OC3 e mantém o valor default de 100 nas rotas recebidas pelo enlace T1. Quando ambos os enlaces estão operacionais, o tráfego é enviado pelo enlace OC3, de acordo com o maior valor do atributo local preference. Se o enlace OC3 falhar, então o tráfego será enviado pelo enlace T1 de backup.

Atributo Atomic_Aggregate

Type Code 6, Well-known Discretionary – RFC 1771.



O atributo Atomic Aggregate é usado por um roteador que, ao ter de selecionar uma rota dentre outras — recebidas de seu peer — que se sobrepõem, escolhe uma, ignorando a mais específica. Então, ele deve incluir o atributo Atomic_Aggregate à rota quando for propagá-la a seus vizinhos (caso o atributo ainda não esteja presente na rota menos específica recebida). Um roteador que receba uma rota com o atributo Atomic_Aggregate não deve removê-lo e não deve fazer nenhum NLRI da rota mais específica quando for propagar a rota aos vizinhos BGP. Ele precisa também reconhecer que o caminho atual para os destinos (como especificado no campo NLRI da rota), respeitando a ausência de loops de roteamento, pode cruzar ASs que não estejam listados no As_Path. Outra observação importante: não é possível agregar um endereço sem ter uma rota mais específica daquele endereço na tabela de roteamento.

Por exemplo: um roteador não pode gerar uma rota agregada para 160.0.0.0 sem possuir previamente uma rota de 160.0.0.0 em sua tabela de roteamento.

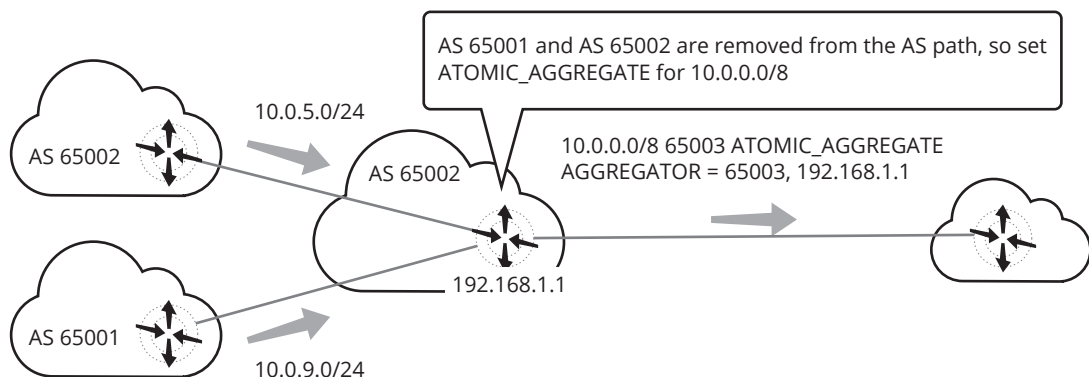


Figura 4.11
Exemplo dos
atributos Atomic_
Aggregate
e Aggregator.

Atributo Aggregator

Type Code 7, Optional Transitive – RFC 1771.



O atributo Aggregator pode ser incluído em mensagens Update formadas por agregação. O atributo Aggregator contém o ASN do último roteador que formou uma rota agregada, seguido de seu próprio ASN e endereço IP.

A Figura 4.11 ilustra o uso em conjunto dos atributos Atomic_Aggregate e Aggregator. O prefixo 10.0.0.0/8 é atribuído ao AS 65003, o qual por sua vez atribui sub-redes dessa rede a seus usuários. Cada usuário anuncia uma sub-rede /24 com BGP. Quando AS65003 anuncia seu agregado /8, ele provoca perda de informação de AS path porque os ASs 65001 e 65002 serão removidos do AS path agregado. Adicionalmente, o atributo Aggregator é incluído para indicar qual roteador originou o agregado.

Atributo Communities

Type Code 8, Optional Transitive – RFC 1991.



Communities é um atributo usado para representar um agrupamento de destinos que compartilhem uma ou mais características, não sendo essas restritas a um mesmo AS, rede ou conjunto de redes. As delimitações do agrupamento são políticas, podendo envolver mais de um AS, inclusive. As comunidades (communities) podem ser compostas por diversas



redes pertencentes a qualquer AS, usadas para simplificar políticas de roteamento, identificando rotas por algum parâmetro lógico em vez de por prefixos CIDR ou ASNs. Usando esses atributos, um roteador pode combiná-los com outros para determinar, para cada comunidade, quais rotas devem ser aceitas, descartadas, preferidas ou repassadas para outros vizinhos. O valor desse atributo pode estar entre 0 (zero) e 4.294.967.200, e consiste de conjuntos de valores de 4 bytes.

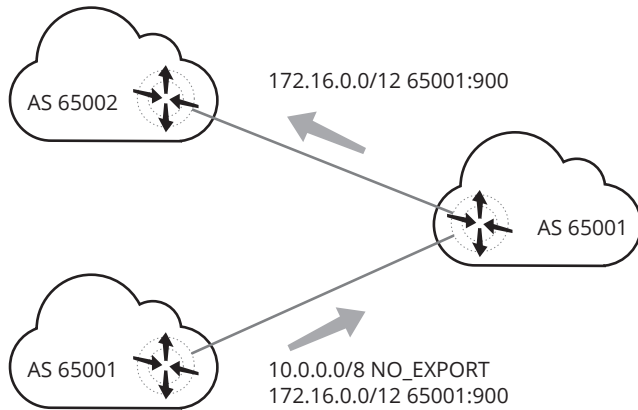


Figura 4.12
Exemplo do atributo Communities.

Esse atributo fornece uma maneira de classificar de forma lógica um prefixo para uso em políticas de roteamento, anexando um identificador que é significativo dentro de uma rede. Por exemplo, ele pode ser usado para informar sobre onde um prefixo entrou na rede ou como um prefixo deveria ser anunciado para diferentes pares. Prefixos podem ter múltiplas comunidades e políticas de roteamento podem ser baseadas em uma ou mais comunidades.

Na linha de comando (CLI), comunidades são representadas com dois números separados por um ":", por exemplo, "65001:500" ou "65000:750". Cada número pode estar na faixa 0 – 65535. A convenção usada é definir o primeiro número para o AS local e o segundo número como um valor arbitrário que é definido pelas políticas administrativas de rede. Isso torna possível rastrear qual AS definiu uma comunidade particular. As comunidades "0:0" a "0:65535" e "65535:0" a "65535:65535" são reservadas e não podem ser usadas, de forma que a linha de comando (CLI) não aceitará comunidades nesta faixa.

Algumas comunidades bem conhecidas dessa faixa reservada foram definidas por conveniência e são entendidas por todos os roteadores. São elas:

- No_Export (65535:65281): não deveria ser anunciado ao par EBGp; é útil para filtrar anúncios nas bordas da rede; marcado como no-export na linha de comando (CLI).
- No_Advertise (65535:65282): não deveria ser anunciado a *nenhum* par; marcado como no-advertise na linha de comando (CLI).

Weight

Definido pela Cisco Systems, o Weight não é propriamente um atributo BGP. Ele influencia no processo de seleção da melhor rota do roteador onde for definido e, como é um atributo local ao roteador, não é repassado e nem propagado aos seus vizinhos nas mensagens de Update. O Weight é um valor decimal situado entre 0 e 65535, sendo o valor padrão igual a 32768, assumido para rotas originadas pelo roteador. Outras rotas possuem o Weight igual a 0 (zero), por padrão. Havendo mais de uma rota possível para um mesmo destino, o BGP-4 seleciona a que possuir o atributo Weight com maior valor. Esse atributo é comumente usado pelos operadores de redes para influenciar o processo de escolha de rotas do BGP.

Configuração BGP – roteadores Cisco



Configuração básica:

- Enable BGP Routing (obrigatório).
- BGP Neighbors (obrigatório).
- BGP Soft Reconfiguration.
- Reset BGP Connections.
- BGP Interactions with IGPs.
- BGP Route Filtering by Neighbor.
- BGP Path Filtering by Neighbor.
- Disable Next-Hop Processing on BGP Updates.
- BGP Version.
- Multi Exit Discriminator Metric (MED).

A configuração do protocolo BGP pode ser dividida em configuração básica e avançada. As duas primeiras tarefas da configuração básica são obrigatórias e as demais, bem como as tarefas da configuração avançada, são opcionais.

A seguir descrevemos cada tarefa da configuração básica e sua implementação em roteadores Cisco:

- **Enable BGP Routing:** para habilitar o roteamento BGP, use os seguintes comandos no modo de configuração global no console do roteador:
 - `router bgp <AS>`, onde <AS> é o número do Sistema Autônomo.
 - `network <network number> mask <network mask> route-map <route-map-name>`.

O primeiro comando estabelece um processo de roteamento BGP no AS indicado. O segundo comando indica que essa rede é local a esse AS e a coloca na tabela BGP. A máscara de rede permite subnetting e supernetting.

- **BGP Neighbors:** é necessário configurar os vizinhos BGP manualmente. BGP suporta dois tipos de vizinhos: 1) vizinhos externos que residem em diferentes ASs e, normalmente, são adjacentes e compartilham uma sub-rede; 2) vizinhos internos que residem em qualquer lugar do mesmo AS, não sendo necessariamente adjacentes. Para isso, use o seguinte comando no modo de configuração de roteador: `neighbor {ip-address|peer-group-name} remote-as <number>`, onde o primeiro parâmetro especifica o endereço IP do vizinho ou o nome do grupo par (peer-group-name) ao qual ele pertence, e o segundo parâmetro especifica o número do AS remoto (se for um vizinho remoto).
- **BGP Soft Reconfiguration:** sempre que houver uma modificação na política de roteamento, a sessão BGP tem de ser encerrada e reiniciada para que as alterações tenham efeito. Isso provoca tremendo impacto na operação das redes. Para permitir que as políticas possam ser modificadas e ativadas sem encerrar as sessões BGP, usa-se essa opção, que deve ser configurada para cada vizinho. Comando: `neighbor {ip-address|peer-group-name} soft-reconfiguration inbound`.
- **Reset BGP Connections:** sempre que dois roteadores forem definidos como vizinhos, eles estabelecerão uma conexão BGP e trocarão informações de roteamento. Se, posteriormente, forem feitas alterações de filtro BGP, peso (weight), distância, versão ou outras alterações similares, a conexão BGP deve sofrer um *reset* para que as alterações tenham efeito. Qualquer um dos dois comandos a seguir pode ser usado:



▣ *Clear ip bgp <address>* dá *reset* numa conexão BGP específica.

▣ *Clear ip bgp ** dá *reset* em todas as conexões BGP.

- ▣ **BGP Interactions with IGP:** se o seu AS estiver conduzindo tráfego de um AS para outro AS, é importante que o protocolo BGP esteja sincronizado com o protocolo IGP do seu AS, para que as rotas anunciadas sejam consistentes. A sincronização BGP/IGP é habilitada por default; porém, nos casos em que o seu AS não conduz tráfego de um AS para outro, não há necessidade dessa sincronização. Para desabilitá-la, use o comando *no synchronization*. Também é preciso dar *reset* nas conexões BGP.
- ▣ **BGP Route Filtering by neighbor:** é possível filtrar os anúncios de rotas BGP de duas maneiras: 1) através de filtros de trajetória de AS (AS-path) usando os comandos *ip as-path access-list* e *neighbor filter-list*; 2) usando listas de acesso ou de prefixo com o comando *neighbor distribute-list*, cujo formato é: *neighbor {ip-address|peer-group-name} distribute-list {access-list-number|name} {in|out}*.
- ▣ **BGP Path Filtering by neighbor:** além da filtragem das atualizações de roteamento baseado nos números de redes, é possível especificar um filtro de lista de acesso em ambas as atualizações (de entrada e saída) baseado nas trajetórias BGP do AS. Cada filtro é uma lista de acesso. Na configuração de filtragem BGP são usados os comandos (iniciando no modo de configuração global):
 - ▣ *ip as-path access-list access-list-number {permit|deny} as-regular-expression*, onde o parâmetro *as-regular-expression* permite complexas manipulações de filtros (ver bibliografia para exemplos);
 - ▣ *router bgp <AS>; 3 neighbor {ip-address|peer-group-name} filter-list access-list-number|name {in|out}*.
- ▣ **Disable Next-Hop Processing on BGP Updates:** o IOS Cisco pode ser configurado para desabilitar o processamento do próximo hop (Next-Hop Processing) nas atualizações BGP (BGP Updates). Isso é útil em redes Frame Relay e X.25 que possuem topologia em malha parcialmente ligada, onde os vizinhos BGP não têm acesso direto a todos os outros vizinhos na mesma sub-rede. Há duas maneiras de fazer isso:
 - ▣ *neighbor {ip-address|peer-group-name} next-hop-self* faz com que o roteador corrente anuncie a si mesmo como o próximo hop para o vizinho especificado; todos os demais roteadores enviarão para esse roteador os pacotes em vez de enviar para o vizinho especificado, e esse roteador se encarregará de encaminhá-los.
 - ▣ *set ip next-hop ip-address [...ip-address] [peer-address]* especifica que o próximo hop é o endereço IP do par remoto (remote peer).
- ▣ **BGP Version:** por default, a versão do BGP é a 4. Se necessário, o BGP negocia a operação em versões anteriores. Para impedir a negociação e forçar o uso específico de uma versão, use o comando *neighbor {ip-address|peer-group-name} version value*.
- ▣ **Multi Exit Discriminator Metric (MED):** BGP usa essa métrica para indicar aos vizinhos externos as trajetórias preferidas. O comando é *default-metric number*.

Configuração BGP – simulador Zebra

- ▣ BGP router.
- ▣ BGP route.
- ▣ Route Aggregation.
- ▣ Redistribute to BGP.





- Defining Peer Group.
- Defining Peer.
- BGP Peer neighbor.
- Show IP BGP.

Os comandos a seguir se referem ao simulador Zebra, usado pelo Imunes:

- **BGP router:** é preciso primeiro configurar o roteador BGP definindo o número do AS onde ele reside. O número do AS é usado pelo protocolo BGP para detectar se a conexão BGP é interna ou externa. Para habilitar o protocolo BGP num determinado AS, use o comando *router bgp asn*, onde *asn* é o número do AS. Depois podem ser digitados quaisquer comandos BGP. Para especificar a identificação do roteador (router-ID), use o comando *bgp router-id ip-address*, onde o endereço IP deve ser o da interface de *loopback*, porque essa interface nunca fica fora (down); se não for especificada uma interface, o BGP usará como identificação do roteador a interface de maior endereço IP; se por algum problema no enlace a interface cair (down), o protocolo BGP ficará instável.
- **BGP route:** é preciso adicionar redes ao AS com o comando *network A.B.C.D/M*, onde *A.B.C.D* é o endereço de rede e */M* é a máscara de sub-rede (notação CIDR); essa rede será anunciada pelo BGP aos seus pares. Exemplo: *router bgp 1; network 10.0.0.0/8*. A rede 10.0.0.0/8 será anunciada a todos os vizinhos. Alguns fabricantes de roteadores não permitem que os roteadores anunciem redes que não estejam nas tabelas de roteamento do protocolo IGP (OSPF, RIP etc.). Nessa implementação o BGP não leva em consideração as rotas IGP.
- **Route Aggregation:** é possível reduzir as tabelas de roteamento agregando rotas (ou supernets), usando a facilidade do Classless Inter-Domain Routing (CIDR). O comando é *aggregate-address A.B.C.D/M {as-set}*. Para que as rotas agregadas não sejam anunciadas no AS, use o comando *aggregate-address A.B.C.D/M summary-only*.
- **Redistribute to BGP:** o protocolo BGP pode aprender rotas internas ao AS, sejam elas do kernel, estáticas, de redes diretamente conectadas ou de protocolos IGP (RIP, OSPF). O comando é *redistribute {kernel | static | connected | rip | ospf}*. Apenas um deles pode ser informado de cada vez. Se necessário, redistribua várias rotas, digitando vários comandos.
- **Defining Peer Group:** para simplificar a configuração de vizinhos, pode-se criar um ou mais peer group e definir os vizinhos dentro dos grupos; são necessários três passos:
 1. Criar o peer group com o comando *neighbor peer-group-name peer group*, onde *peer-group-name* é o nome do grupo.
 2. Configurar opções para o grupo, entre elas:
 - 2.1. *neighbor peer-group-name remote-as asn*, para especificar um vizinho BGP.
 - 2.2. *neighbor peer-group-name update-source interface*, para permitir que as sessões BGP internas possam usar qualquer interface operacional para as conexões TCP.
 - 2.3. *neighbor peer-group-name description*, para associar uma descrição a um vizinho BGP.
 3. Definir o peer group com o comando *neighbor ip-address remote-as asn*, onde o endereço IP informado é o do vizinho no AS remoto de número *asn*. Exemplo: *router bgp 1; neighbor 10.0.0.1 remote-as 2*. O roteador do AS 1 está tentando o acesso ao par (peer) 10.0.0.1 no AS 2. Esse comando tem de ser o primeiro a ser usado na configuração de vizinho.

- **BGP Peer neighbor:** o protocolo BGP requer configurações específicas de vizinhos.

Vejamos algumas delas:

- ▣ *neighbor ip-address shutdown:* esse comando tira do ar o vizinho, mas preserva a configuração dele; para excluir também a configuração do vizinho, use o comando *no neighbor ip-address remote-as asn*.
 - ▣ *neighbor ip-address description:* faz descrição do vizinho.
 - ▣ *neighbor ip-address next-hop-self:* faz com que o roteador corrente anuncie a si mesmo como o próximo hop para o vizinho especificado; todos os demais roteadores enviarão para esse roteador os pacotes, em vez de enviar para o vizinho especificado, e esse roteador se encarregará de encaminhá-los.
 - ▣ *neighbor ip-address default-originate:* por default, o BGP não anuncia a rota padrão (0.0.0.0/0), mesmo que ela esteja na tabela de roteamento; esse comando faz com que a rota padrão seja anunciada para o vizinho.
- **Show IP BGP:** lista as rotas BGP; se for especificado um endereço, lista as rotas relacionadas; se não, lista todas as rotas. O comando é *show ip bgp A.B.C.D*.



Roteiro de Atividades 4

Atividade 4.1 – Configuração do protocolo BGP

- AS6500 – roteadores ROT1, ROT2, ROTA, ROTB e ROTC.
- AS1900 – roteadores router0, router1 e router2.

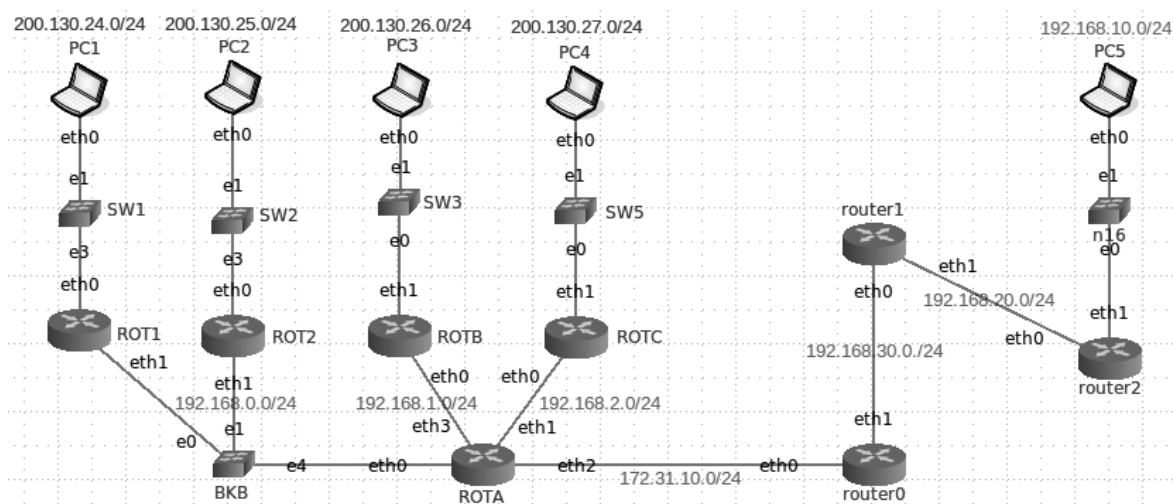


Figura 4.13
Rede Rede1_
Sessao4_ADR8.

Descrição da rede

Na rede da Figura 4.13 temos dois ASs: 6500 e 1900.

- O AS 6500 é composto por cinco roteadores: ROT1 ROT2, ROTA, ROTB e ROTC, e pelas redes 200.130.24.0/24, 200.130.25.0/24, 200.130.26.0/24 e 200.130.27.0/24 (bloco 200.130.24.0/22).
- Os roteadores ROTB e ROTC usam somente o protocolo OSPF. Os demais roteadores usam OSPF e BGP.
- O ROTA mantém sessão iBGP com os roteadores ROT1 e ROT2 e sessão eBGP com o router0.
- O AS 1900 é composto por três roteadores: router0, router1 e router2, e pelas redes 192.168.10.0/24, 192.168.20.0/24 e 192.168.30.0/24 (bloco 192.168.0.0/16).
- O router2 usa somente o protocolo OSPF. Os demais roteadores usam OSPF e BGP.
- O router0 mantém sessão iBGP com o router1 e sessão eBGP com o ROTA.

Os hosts pc1 e pc5 têm endereços IP: 200.130.24.2/24 e 192.168.10.2/24, respectivamente.

Atividade 4.2 – Configuração do protocolo BGP

- Carregar a rede *Rede1_sessao4_ADR8* no simulador Core.
- Verificar as tabelas de rotas OSPF dos roteadores.
- Configurar o protocolo BGP nos roteadores.
- Verificar as tabelas de rotas BGP dos roteadores.
- Verificar a continuidade entre os PCs.



Conclusão

Nestas atividades práticas aprendemos a:

- Configurar protocolo BGP;
- Analisar anúncios de rotas BGP;
- Verificar a conectividade da rede.

É assim que os ASs se comunicam na internet. Esta configuração é adequada para interligar dois ASs privados, mas não é adequada pra interligar um AS privado a um AS público como, por exemplo, de uma Telemar, Embratel ou Brasil Telecom. A razão disso é que as rotas BGP são redistribuídas para as rotas OSPF intra-domínio. Se o AS vizinho for público, os roteadores internos do AS privado que estão rodando OSPF receberão todas as rotas do AS público, quantidade essa que pode chegar facilmente a 200 mil rotas, sobrecarregando a tabela de rotas sem necessidade.

No próximo capítulo veremos como conectar um AS privado relativamente pequeno com um AS público, sem redistribuir as rotas BGP para o OSPF.

5

Protocolo de roteamento BGP4 – Parte 2

objetivos

Estudar os conceitos avançados do protocolo Border Gateway Protocol versão 4 (BGP-4), bem como o funcionamento dos Pontos de Troca de Tráfego entre os ASs.

Sessão BGP, Mensagens BGP, Mapas de rotas e Pontos de Troca de Tráfego (PTT).

conceitos

Sessão BGP

- Entre vizinhos BGP (BGP neighbors).
- Usa o protocolo TCP (porta 179).
- Negocia diversos parâmetros.
- Envia as “melhores rotas” (best paths) conhecidas.
- Depois as atualizações são incrementais.
 - ▣ Mensagens Update.
 - ▣ Somente quando houver alteração.
 - ▣ Controle do número de versão da atualização.
 - ▣ Mensagens Keep-Alive (“Estou Vivo”).



Antes do estabelecimento de uma sessão BGP, os roteadores “vizinhos BGP” trocam mensagens entre si para entrar em acordo sobre quais serão os parâmetros (exemplo: tempo máximo de espera entre mensagens – hold time) da sessão. Não havendo discordância nem erros durante a negociação dos parâmetros entre as partes, a sessão BGP é estabelecida. Caso contrário, serão enviadas mensagens de erro e a sessão não será aberta.

Quando a sessão é estabelecida entre os roteadores, são trocadas mensagens contendo todas as informações de roteamento, ou seja, todos os “melhores caminhos” (best paths) previamente selecionados por eles, para os destinos conhecidos. Posteriormente, eles trocarão somente mensagens de atualização das informações de roteamento (mensagens do tipo Update) de forma incremental. Essa técnica mostrou-se um avanço no que se refere à diminuição de carga nas CPUs dos roteadores e na economia da largura de banda dos enlaces, quando comparada a outros protocolos que, ao comunicarem suas atualizações, enviam periodicamente a totalidade das rotas existentes em suas tabelas.



Nesse sentido, o BGP é bem econômico, somente enviando mensagens de atualização quando ocorrem mudanças nas rotas (exemplo: uma rota se tornou inválida) e informando novas rotas. Caso não existam atualizações a serem informadas, os roteadores trocam apenas mensagens do tipo Keep-Alive para se certificarem de que a comunicação entre eles está “viva”, ou seja, ainda está ativa. Essas mensagens são pequenas (apenas 19 bytes), não sobrecarregando a CPU dos roteadores e nem o enlace entre eles.

Uma característica das tabelas de rotas BGP é a existência de um número de versão, que é incrementado a cada atualização feita (através das mensagens do tipo Update), permitindo assim a verificação de inconsistências nas informações de roteamento.

Se ocorrer rápido aumento no número da versão das tabelas, isso pode ser um indicativo de instabilidade na rede.

Mensagens BGP

- Tamanho mínimo de 19 e máximo de 4.096 bytes.
- Cabeçalho + Dados.
- Campos do cabeçalho.
 - ▣ Marcador.
 - ▣ Comprimento.
 - ▣ Tipo.
 - ▣ Dados (opcional).

Marcador (Marker) 16 bytes	Comprimento (Length) 2 bytes	Tipo (Type) 1 byte
-------------------------------	---------------------------------	-----------------------



Figura 5.1
Cabeçalho das mensagens BGP.

As mensagens trocadas em sessões BGP têm o comprimento mínimo de 19 bytes e máximo de 4.096 bytes. Todas as mensagens são compostas de, no mínimo, um cabeçalho e, opcionalmente, de dados. O formato do cabeçalho das mensagens BGP está descrito na figura 5.1. É opcional uma sequência de dados após o cabeçalho.

Campos do cabeçalho:

- **Marcador** (Marker): serve para verificar a autenticidade da mensagem recebida e se houve perda de sincronização entre os roteadores vizinhos BGP. Pode ter dois formatos: caso a mensagem seja do tipo Open (abrir), ou se a mensagem do tipo Open não possuir informação de autenticação, o campo deve estar todo preenchido com números um (1); se não, o campo *marker* terá o seu conteúdo baseado em parte do mecanismo de autenticação usado.
- **Comprimento** (Length): deve conter um número que representa o comprimento total da mensagem, incluindo o cabeçalho. Como pode haver mensagens que não possuem dados após o cabeçalho, a menor mensagem BGP enviada é de 19 bytes (16 + 2 + 1 bytes).
- **Tipo** (Type): deve conter um número que representa o código de um tipo de mensagem.

Tipos de mensagens BGP

Os tipos de mensagens são:

- Open.
- Notification.
- Keep-Alive.
- Update.
- Route-Refresh.



Mensagem Open

- Versão.
- Número do AS (ASN).
- Tempo de espera.
- Identificador BGP.
- Comprimento dos parâmetros opcionais.
- Parâmetros opcionais.

A mensagem do tipo Open (tipo 1) é enviada para que seja iniciada a abertura de uma sessão BGP entre neighbors ou peers BGP. O formato dessa mensagem está mostrado na Figura 5.2.

Versão (Version) 1 byte	Nº do AS (ANS) 2 bytes	Tempo de espera (Hold Time) 2 bytes	Identificador de BGP (BGP ID) 4 bytes	C.P.O. (ANS) 1 byte
Parâmetros Opcionais (Tipo/Comprimento/Valor) Tamanho variável				

Figura 5.2
Mensagem Open.

Descrição dos campos:

- **Versão** (Version): identifica a versão do BGP (3 ou 4). Esse é um dos parâmetros negociados pelos roteadores que, normalmente, tentam entrar em acordo para usar a maior versão suportada. Não havendo possibilidade de consenso (exemplo: um dos roteadores não suporta o BGP-4), eles tentam usar versões anteriores, até que coincidam. Nos roteadores Cisco, há como configurar a versão a ser usada pelos roteadores (se previamente sabemos a versão que ambos suportam), eliminando essa fase de negociação do processo de abertura da sessão BGP, resultando em economia de tempo.
- **Número do AS** (AS Number): deve conter o número do AS ao qual o roteador (remetente da mensagem tipo Open) pertence.
- **Tempo de espera** (Hold Time): deve conter o valor, em segundos, do maior tempo de espera (hold time), permitido entre mensagens do tipo Update ou Keep-Alive. O BGP mantém um contador do hold time, que é reiniciado (zerado) a cada vez que uma mensagem do tipo Keep-Alive ou Update é recebida. Caso nenhuma das duas seja recebida no prazo máximo, o BGP considera que a comunicação com o outro roteador foi perdida, e a sessão é encerrada, tendo de ser reiniciada novamente. Os roteadores tentam usar o menor hold time entre os dois. Caso o valor seja estabelecido como zero, a sessão será considerada como estando sempre “viva” (ativa) e mensagens de Keep-Alive não serão transmitidas, pois os contadores do hold time e do Keep-Alive não serão zerados nunca. O valor mínimo recomendado para esse parâmetro é de três segundos.



- **Identificador BGP (BGP ID):** é a identificação BGP do roteador que enviou a mensagem Open. Contém o endereço IP definido no comando *bgp router-id*.
- **Comprimento dos Parâmetros Opcionais** (Optional Parameters Length): indica o comprimento total do campo de *Parâmetros Opcionais* (Optional Parameters). No caso de ausência de parâmetros opcionais, esse campo será preenchido com zero.
- **Parâmetros Opcionais** (Optional Parameters): pode conter vários parâmetros opcionais para a negociação de abertura de uma sessão BGP. Esse campo deve ser preenchido com conjuntos formados por três valores:
 1. Tipo do parâmetro (1 byte).
 2. Comprimento do parâmetro (1 byte).
 3. Valor do parâmetro (comprimento variável).

Um exemplo de parâmetro é o de informação de autenticação (tipo 1), usado para autenticar a sessão com o vizinho BGP.

Mensagem Notification

- Erro.
- Subcódigo de erro.
- Dados.



Esse tipo de mensagem (tipo 4) é enviada no caso de detecção de erros durante ou após o estabelecimento de uma sessão BGP. O formato da mensagem Notification está mostrado na Figura 5.3.

Erro (Error) 1 byte	Subcódigo de erro (Error Subcode) 2 bytes	Dados (Data) Tamanho variável
----------------------------------	--	--

Figura 5.3
Mensagem
Notification.

- **Erro** (Error): deve conter o tipo da notificação.
- **Subcódigo de Erro** (Error subcode): deve conter um valor que fornece mais informações sobre o erro.
- **Dados** (Data): pode conter dados referentes ao erro, como por exemplo um cabeçalho mal formado (inválido) ou um número de AS inválido.

Mensagem Keep-Alive

Somente o cabeçalho padrão BGP (19 bytes).



São mensagens (tipo 3) trocadas periodicamente com o propósito de verificar se a comunicação entre os vizinhos está ativa. A mensagem do tipo Keep-Alive é composta apenas pelo cabeçalho padrão das mensagens BGP, sem dados transmitidos após o cabeçalho. O tempo máximo permitido para o recebimento de mensagens Keep-Alive ou Update é definido pelo hold time, como foi visto na descrição do tipo de mensagem Open.

Para manter aberta a sessão, a mensagem de Keep-Alive deve ser enviada antes que expire o prazo definido no hold time; caso contrário, a sessão será encerrada. A recomendação é de que a mensagem seja enviada em até 1/3 do tempo definido no hold time. Se o seu valor for igual a zero, então as mensagens do tipo Keep-Alive não serão enviadas.

Mensagem Update

- Rotas removidas (unfeasible routes).
 - ▣ Comprimento do campo (length).
 - ▣ Rotas removidas (withdrawn routes).
- Atributos Caminho (Path Attributes).
 - ▣ Comprimento do campo (length).
 - ▣ Atributos Caminho (Path Attributes).
- Informações NLRI.

Comprimento (Length), Prefixo (Prefix)...

As mensagens Update (tipo 2) trocadas entre os peers ou neighbors BGP são de extrema importância, pois são elas que levam as informações para a atualização da tabela de rotas mantida pelo BGP. O formato da mensagem do tipo Update está mostrado na Figura 5.4.

Unfeasible Routes Length 2 bytes	Withdrawn Routes <Length, Prefix>
Total Path Attributes Length 2 bytes	Path Attributes Length
NLRL Information 2 bytes	

Figura 5.4
Mensagem Update.

A estrutura básica das mensagens do tipo Update é composta por três itens:

- Rotas removidas (unfeasible routes).
- Atributos de caminhos (path attributes).
- Informação de alcance da camada de rede (Network Layer Reachability Information – NLRI).

O formato dos campos é:

- Rotas removidas:
 - ▣ Comprimento do campo (2 bytes);
 - ▣ Rotas removidas <Tamanho,Prefixo>.
- Atributos do Caminho:
 - ▣ Comprimento do campo (2 bytes);
 - ▣ Atributos do Caminho.
- Informação NLRI: <Tamanho,Prefixo>
- **Comprimento das rotas removidas ou inalcançáveis** (Unfeasible Routes Length): nesse campo é indicado o comprimento total, em bytes, do total de rotas removidas. Um comprimento igual a 0 (zero) indica que não há rotas a serem removidas nessa mensagem Update, portanto, não existe o campo de rotas removidas.
- **Rotas removidas** (Withdrawn routes): este campo inclui uma lista de prefixos de endereços para rotas que devem ser removidas da tabela de rotas BGP. É composto de endereços IP (prefixos) e do tamanho do prefixo (notação CIDR). Um exemplo de entrada seria: <16,143.54.0.0>, que representa a rede 143.54.0.0, dizendo que o endereço de rede é até o décimo sexto bit, ou seja, em notação CIDR representa a rede 143.54.0.0/16.

- **Comprimento Total do Atributo Caminho** (Total Path Attribute Length): deve indicar o comprimento total, em bytes, do campo *Atributo Caminho*. O valor contido nesse campo deve permitir determinação do comprimento do campo NLRI (ver RFC1771). Se o valor desse campo for 0 (zero), significa que não há informação NLRI presente na mensagem Update.

Atributos do Caminho (Path Attributes): são um conjunto de parâmetros associados a uma determinada rota que influenciam no processo de decisão feito pelo BGP para escolha da melhor rota, como, por exemplo, Local_Pref, Next_Hop, Origin etc.

- **Informações NLRI** (NLRI Information): são prefixos de endereços IP de informações no formato igual ao do campo de *rotas removidas* (*withdrawn routes*). Esse campo é preenchido por várias entradas. Um exemplo de entrada seria: <18,192.213.134.0>, que indica uma rota para a rede 192.213.14.0/18, onde 18 representa o tamanho do prefixo (número de bits 1 na máscara do bloco).

Mensagem Route-Refresh

- Definida no RFC2918.
 - Address Family Identifier (AFI).
 - Reservado (valor 0).
 - Subsequent Address Family Identifier (SAFI).
- Serve para solicitar a retransmissão de todas as informações de roteamento de um vizinho BGP.
- Não precisa fechar e reiniciar a sessão BGP.
- Independente do protocolo (IPv4 ou IPv6).



A mensagem Route-Refresh (tipo 5) não está definida no RFC4271, mas sim no RFC2918, como uma capacidade do BGP. O leiaute dela está mostrado na Figura 5.5.



Figura 5.5
Mensagem
Route-Refresh.

Ela é usada para solicitar a completa retransmissão de todas as informações de roteamento de um vizinho, sem necessidade de encerrar e reabrir uma sessão BGP com o vizinho. Dessa forma, mudanças nas políticas de roteamento podem ser feitas dinamicamente, economizando recursos do roteador. Essa mensagem foi projetada para ser independente de protocolo; assim pode ser solicitada a retransmissão das rotas IPv4, mas não das rotas IPv6.

Um exemplo: o campo AFI pode ser IPv4 ou IPv6, enquanto o campo SAFI pode ser unicast ou multicast.

ERROR CODE		ERROR SUBCODE
1	Erro no cabeçalho da mensagem	1. Conexão não sincronizada 2. Comprimento da mensagem inválido 3. Tipo de mensagem inválido
2	Erro na mensagem OPEN	1. Número de versão não suportado 2. Número de AS vizinho inválido 3. Identificador BGP inválido 4. Parâmetro opcional não suportado 5. Falha na autenticação 6. Tempo de espera inaceitável
3	Erro na mensagem UPDATE	1. Lista de atributos mal formada 2. Atributo well-know desconhecido 3. Atributo well-know faltando 4. Erro nas flags de atributos 5. Erro no comprimento do atributo 6. Atributo de origem inválido 7. Loop de roteamento em AS 8. Atributo NEXT HOP inválido 9. Erro no atributo opcional 10. Campo de rede inválido 11. AS_PATH mal formado
4	Hold time expired	No subcodes
5	Finite state error	No subcodes
6	Cease	No subcodes

Figura 5.6
Mensagens de erro.

As mensagens de erro BGP sinalizam as diversas condições de erro que podem ocasionar a perda de informações ou até o encerramento da conexão TCP e, portanto, da sessão BGP.

Mapas de rotas

Redistribuição:

- Cada protocolo de roteamento mantém sua própria tabela.
- O roteador mantém uma tabela de todas as rotas.
- A redistribuição é o repasse de rotas entre os protocolos.

Mapas de rotas (route maps).

- Controlam e modificam informações de roteamento.
- Formato: `route-map nome permit|deny seq`.
- Comandos `match` e `set`.

Listas de prefixo (prefix lists):

- `ip prefix-list nome seq número permit|deny prefix [le len] [ge len]`

Num roteador, cada protocolo de roteamento mantém a sua tabela de rotas individual na memória, enquanto o próprio roteador mantém outra tabela montada com rotas fornecidas por todos os protocolos de roteamento que estiverem sendo executados nele. Essa é a tabela utilizada pelo roteador para executar sua função de rotear pacotes de dados.



A redistribuição acontece quando, em um roteador, um protocolo de roteamento repassa as rotas de sua tabela para outro protocolo de roteamento. O outro protocolo pode aceitar (ou não) todas ou apenas algumas e incluí-las em sua tabela de rotas. Posteriormente, essas rotas serão anunciadas por esse outro protocolo para os roteadores vizinhos que “falam” esse mesmo protocolo.

O comando *network* é uma das formas de anunciar as redes de um AS no protocolo BGP. Outra forma é redistribuir as rotas conhecidas pelo IGP para o BGP. Isso pode ser muito perigoso, pois pode-se injetar todas as rotas internas do AS no BGP desnecessariamente. Se, por exemplo, uma das rotas foi aprendida através do próprio BGP, então não há necessidade de repassá-la novamente. Uma filtragem cuidadosa deve ser aplicada para garantir que só serão anunciadas para a internet rotas que realmente desejamos anunciar, e não anunciar todas indiscriminadamente.

No capítulo anterior vimos o exemplo oposto: as rotas BGP sendo anunciadas para o protocolo OSPF, que é um IGP. Também existem riscos naquele caso, conforme foi mostrado.

Route maps servem para o BGP controlar e modificar informações de roteamento e também definir as condições da propagação de rotas entre domínios de roteamento. Os route maps possibilitam a definição de condições para, por exemplo, redistribuição de rotas entre protocolos de roteamento (BGP e algum IGP) ou para o controle das rotas injetadas (ou removidas) no BGP.

Sintaxe de um route map:

- **route-map** *nome* [[**permit** | **deny**] | [*seq*]]: o *nome* identifica o route map. O *seq* indica a posição que a instância do route map deve ter em relação a outras instâncias do mesmo route map, sendo as instâncias ordenadas sequencialmente. Exemplos de route maps:
 - *route-map* *TESTE* *permit* 10: primeiro conjunto de condições.
 - *route-map* *TESTE* *permit* 20: segundo conjunto de condições.

Quando o BGP aplica o route map *TESTE* nas atualizações de roteamento (route updates), primeiro é aplicada a instância que possuir o menor número sequencial (no exemplo acima, a instância 10) e depois as subsequentes, se houver. Se o primeiro conjunto de condições não for satisfeito, o segundo será aplicado e assim por diante, até que algum conjunto de condições seja satisfeito ou quando não houver mais condições a serem aplicadas.

Os comandos *match* e *set* são usados para definir as condições no route map. O comando *match* define a condição a ser satisfeita e o comando *set* especifica a ação a ser tomada caso o update corresponda à condição. Abaixo, um exemplo de route map simples:

```
route-map TESTE permit 10
match ip address 10.10.8.1
set metric 10
```

Quando uma rota corresponder ao endereço IP 10.10.8.1, o BGP vai configurar a métrica do update para 10, propagá-lo (pelo uso da palavra-chave *permit*) e sair da lista de instâncias de route maps. Caso o update não corresponda ao critério de uma instância definida, o BGP vai comparar com a instância seguinte, até que uma ação seja tomada ou até que a última instância seja comparada. Se o update não satisfizer nenhuma das condições, o update não será propagado. Caso seja usada a palavra-chave *deny* na configuração do route map e o update corresponder ao critério definido, o BGP vai interromper a comparação com a lista de instâncias e o update não será propagado.

Uma restrição que deve ser observada no uso de route maps é que eles podem ser usados para filtrar anúncios (redistribuição) de updates baseados em endereços IP, mas não para filtrar o recebimento de updates baseados nos endereços IP.

Listas de Prefixo (Prefix Lists) podem ser usadas como alternativa para as listas de acesso (Access Control Lists – ACLs) em muitos comandos de filtragem de rotas BGP. Tais listas fornecem o mais poderoso mecanismo de filtragem baseado em prefixos, com as seguintes vantagens sobre as listas de acesso:

- Significativa melhoria de performance na carga e pesquisa de rotas em grandes listas.
- Suporte para atualizações incrementais; a filtragem que usa listas de acesso estendidas não suporta atualizações incrementais.
- Interface de linha de comando mais amigável; a interface de linha de comando para uso de listas de acesso na filtragem de atualizações BGP é difícil de compreender e de utilizar, porque utiliza o formato de filtragem de pacotes.
- Maior flexibilidade; antes de usar uma lista de prefixo num comando é necessário preparar a aplicação da lista num mapa de rotas, como veremos mais adiante.

O formato é: `ip prefix-list nome seq número permit|deny prefix [le len] [ge len]`, onde:

- **nome**: nome da lista de prefixo.
- **número**: número sequencial que determina a ordem dentro da lista; pode ser numerado manualmente ou automaticamente; nesse último caso a numeração será de 5 em 5.
- **le len**: esse comando especifica o comprimento do *prefix* (prefix length); as condições da lista de prefixo serão aplicadas se o comprimento do prefixo for menor ou igual ao valor *len*.
- **ge len**: esse comando especifica o comprimento do *prefix* (prefix length); as condições da lista de prefixo serão aplicadas se o comprimento do prefixo for maior ou igual ao valor *len*.

Esses dois últimos comandos podem ser usados sozinhos ou em conjunto, não importando a ordem.

A filtragem por lista de prefixo envolve a comparação dos prefixos das rotas com aquelas relacionadas na lista de prefixo. Quando ocorre uma igualdade (match), a rota é usada. A comparação é similar àquela usada nas listas de acesso.

Mais especificamente, a permissão ou a negação de um prefixo é baseada nas seguintes regras:

- Uma lista de prefixo vazia permite (permit) todos os prefixos.
- Uma negação (deny) implícita é assumida se um determinado prefixo não existe (doesn't match) na lista de prefixo.
- Quando um dado prefixo aparece várias vezes na lista de prefixo (multiple entries), a ocorrência com o menor número de sequência será a escolhida (match).

O roteador inicia a pesquisa no início (top) da lista de prefixo, pelas ocorrências de menor número sequencial. Quando ocorrer uma igualdade (match), o roteador interrompe a pesquisa e ignora o resto da lista, executando as ações definidas na ocorrência em que ocorreu a igualdade. Para maior eficiência na pesquisa da lista, é recomendável colocar as ocorrências mais comuns no topo da lista.

Uso de mapas de rotas

- `route-map meumapa permit 10`
 - `match ip address 10.0.0.1`
 - `set metric 5`
- `ip prefix-list abc permit 192.0.0.0/8 le 24`
- `ip prefix-list abc deny 192.0.0.0/8 ge 25`
- `ip prefix-list abc permit 0.0.0.0/0 le 32`
- `route-map rotaspadrao permit 10`
 - `match ip address prefix-list BOGONS`



Embora existam muitos métodos para filtragem de rotas em BGP, vamos exemplificar aqui o uso de listas de prefixo e mapas de rotas. O primeiro exemplo verifica as atualizações que têm o endereço IP 10.0.0.1, alterando a métrica para 5. Como o mapa foi declarado com a condição `permit`, a rota será propagada. Se a condição fosse `deny`, a rota não seria propagada. Nota: é inútil declarar a cláusula `set` quando a condição for `deny`, porque a rota não será propagada, e a alteração não será feita.

Os três exemplos seguintes são de lista de prefixo. O primeiro deles permite rotas 192.0.0.0/8 de tamanho de prefixo de até 24. O segundo deles nega as rotas 192.0.0.0/8 de tamanho de prefixo maior ou igual a 25. O terceiro deles permite todas as rotas.

O último exemplo define um mapa de rotas chamado *rotaspadrao*, que usa a lista de prefixo chamada BOGONS, que define as rotas que não vamos aceitar porque são as rotas óbvias: rota padrão, endereços privados (RFC1918), endereço multicast e outras específicas do domínio local.



A palavra-chave BOGONS é usada para definir esse tipo de lista de prefixo.

A lista aceita como referência padrão é:

- `ip prefix-list BOGONS description (redes “óbvias” que não serão aceitas)`
- `ip prefix-list BOGONS seq 5 deny 0.0.0.0/8 le 32`
- `ip prefix-list BOGONS seq 10 deny 10.0.0.0/8 le 32`
- `ip prefix-list BOGONS seq 15 deny 127.0.0.0/8 le 32`
- `ip prefix-list BOGONS seq 20 deny 172.16.0.0/12 le 32`
- `ip prefix-list BOGONS seq 25 deny 169.254.0.0/16 le 32`
- `ip prefix-list BOGONS seq 30 deny 192.168.0.0/16 le 32`
- `ip prefix-list BOGONS seq 35 deny 192.0.2.0/24 le 32`
- `ip prefix-list BOGONS seq 40 deny 224.0.0.0/3 le 32`
- `ip prefix-list BOGONS seq 45 permit 0.0.0.0/0 le 32`

Route Reflector

- Número de sessões BGP: $n(n-1)/2$.
- Full-mesh BGP.
- Solução: Route Reflector



Como vimos anteriormente, não há detecção de loop de roteamento em sessões iBGP. Dessa forma, o anúncio de rotas para um vizinho iBGP pode causar um loop de roteamento que não será detectado. É por essa razão que os roteadores iBGP não anunciam rotas para seus vizinhos. Em outras palavras: cada roteador iBGP *tem* de manter uma sessão BGP com todos os outros roteadores iBGP, mesmo não tendo uma conexão física com eles. É o que nós chamamos de full mesh BGP.

Na Figura 5.7, com apenas 6 roteadores teríamos $n(n-1)/2$ sessões: $6(6-1)/2 = 15$ sessões, o que torna quase inviável a comunicação iBGP. Por outro lado, não configurar todos os vizinhos (peers) dentro do AS pode fazer com que alguns roteadores desconheçam algumas rotas.

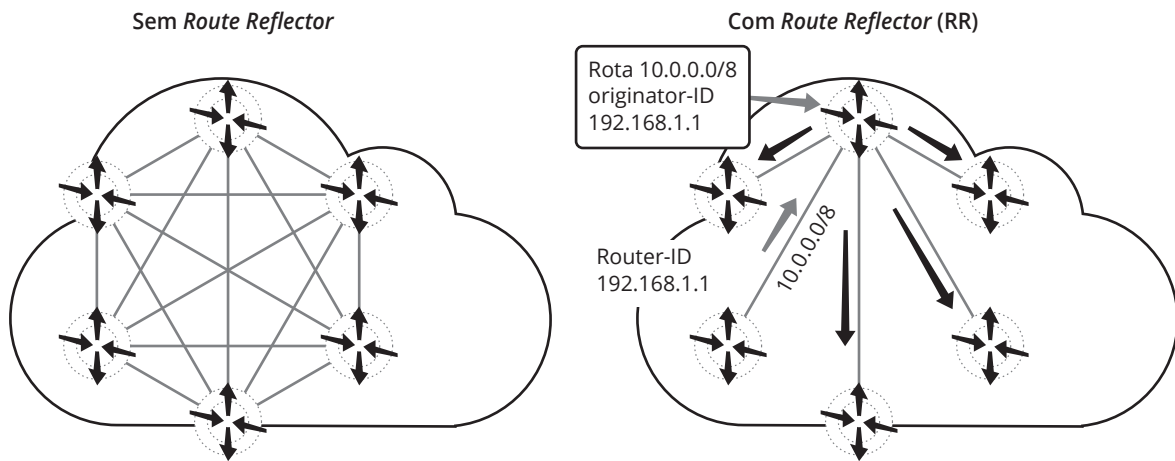


Figura 5.7
Sem Route
Reflector/Com
Route Reflector (RR).

Para resolver esse problema, duas soluções podem ser adotadas: route reflection (reflexão de rotas) e confederations (confederações). Ambas são amplamente utilizadas e não são mutuamente exclusivas, podendo ser utilizadas dentro do mesmo AS.

Route reflection é definida no RFC2796 e introduz um enfoque de hierarquia para resolver o problema do full-mesh BGP. Em vez de definir vizinhança (peering) com todos os roteadores do AS, apenas define-se vizinhança com um roteador escolhido para ser o route reflector, conforme mostrado na Figura 5.7. Com 6 roteadores, reduzimos 15 sessões para apenas 5 sessões iBGP.

Os vizinhos do route reflector são chamados clientes (clients). O roteador escolhido para ser o route reflector pode redistribuir rotas iBGP para seus clientes. Cada grupo de clientes com seu respectivo route reflector é chamado de cluster (concentração). Cada cluster recebe uma identificação única (cluster ID).

Todos os atributos de caminho (path attributes) são passados para os clientes sem modificação, especialmente o endereço do próximo salto (next hop address), se não todo o tráfego teria de passar pelo route reflector (RR).

Já que agora o route reflector pode anunciar rotas iBGP para seus vizinhos, podem ocorrer loops de roteamento. Para evitar isso foram definidos dois novos atributos de caminho: Originator-ID e Cluster List, ambos definidos no RFC2796.

Suponha, na figura anterior, que o roteador 192.168.1.1 anunciou a rota 10.0.0.0/8 para o RR que, por sua vez, a repassou para os seus clientes. O RR anunciou a rota com o atributo Originator-ID = 192.168.1.1, para indicar o roteador que anunciou aquela rota. Esse anúncio nunca será feito para o roteador que originou essa rota (o próprio 192.168.1.1), se não poderíamos ter um loop de roteamento. Os clientes *nunca* usam esse atributo.

Cluster list

A lista de cluster (cluster list) registra os clusters que um anúncio de prefixo atravessou. Os RRs acrescentam o cluster-ID à lista de clusters quando anunciam a rota para outro cluster. Se um RR detectar seu próprio cluster-ID no anúncio feito por um vizinho, esse RR não aceitará o anúncio do prefixo, pois isso poderia provocar um loop de roteamento. Os clientes nunca modificam o atributo Cluster-List.

Na Figura 5.8, a rota 10.0.0.0/8 foi anunciada pelo cluster 192.168.1.1 e, depois de passar pelos clusters 192.168.1.2 e 192.168.1.3, será anunciada para o cluster 192.168.1.1, que originou essa rota; por isso ela é rejeitada.

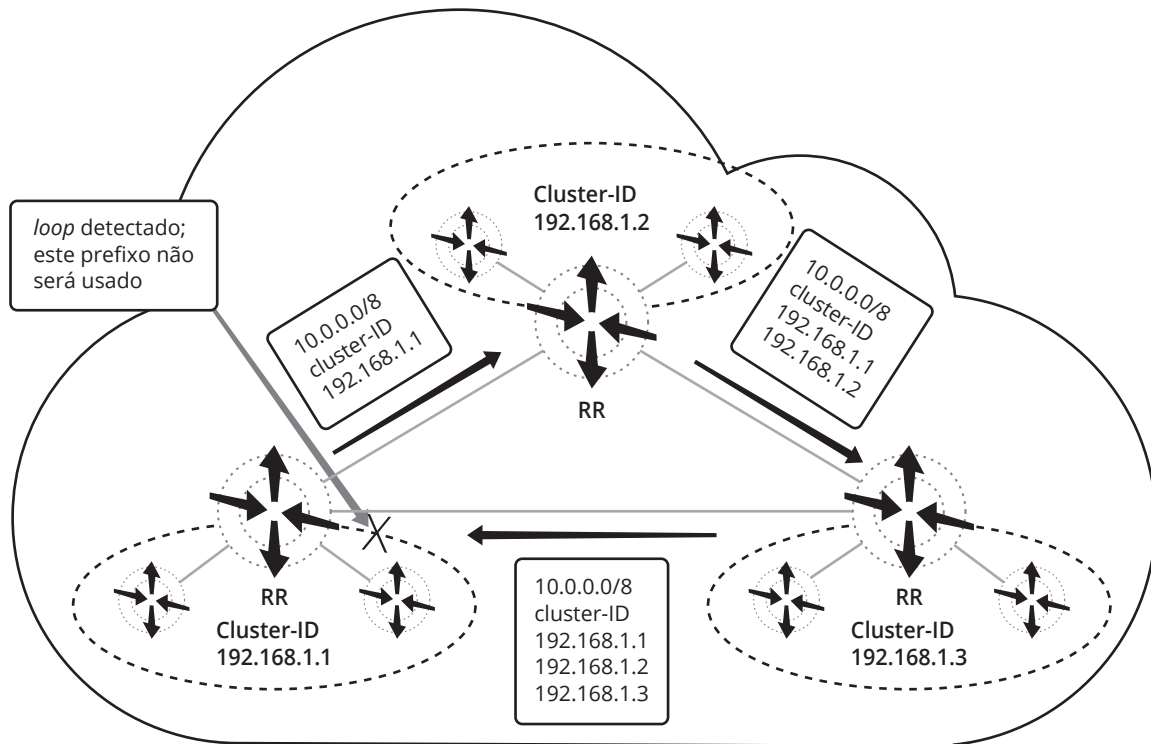


Figura 5.8
Exemplo de
Cluster list.

Pontos de Troca de Tráfego (PTT)

- A internet é uma “nuvem”.
- Compra de trânsito via provedor de acesso.

É comum imaginar a internet como uma “nuvem” a qual se está, de alguma forma, conectado; e a forma de conexão mais comum é uma relação comercial de compra de trânsito com um provedor, como representado na Figura 5.9.

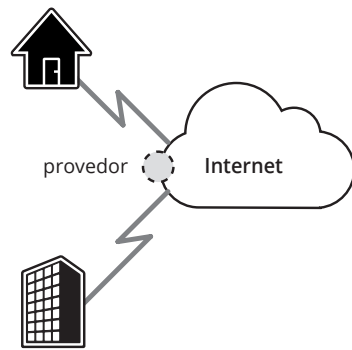


Figura 5.9
Provedor de acesso
à internet.

Normalmente uma empresa ou usuário doméstico realizam sua conexão à internet através de um provedor. Esse desenho reflete uma forma comum de pensar e não ajuda a perceber a realidade de que a empresa, os computadores domésticos e o próprio provedor, todos fazem parte da internet.

Frequentemente, costuma-se pensar nessa forma de conexão à internet como a única possível. Não é hábito refletir sobre a natureza da “nuvem” nem pensar na possibilidade da conexão direta entre redes diferentes como uma realidade prática. Contudo, uma vez que se esteja ligado à internet, passa-se a fazer parte dessa “nuvem” imaginária, como ilustra a Figura 5.10. A “nuvem” internet é formada pelos seus participantes, interligados por meios físicos e falando o Protocolo Internet (IP).

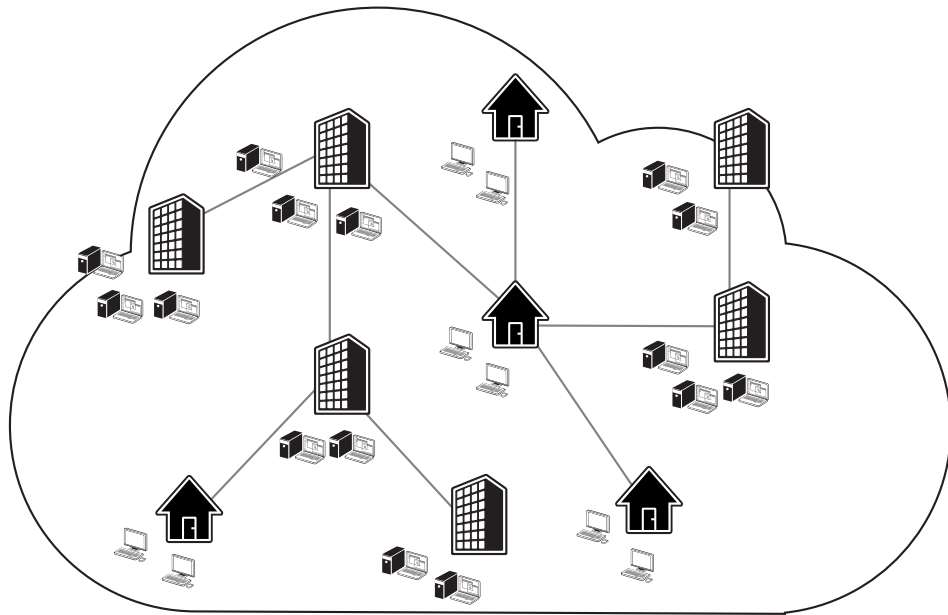


Figura 5.10
A internet como
uma ‘nuvem’.

Essa “nuvem” é, de fato, uma abstração que representa todas as redes interligadas, incluindo os usuários domésticos, empresas de todos os tamanhos, redes acadêmicas e governamentais etc. Ligações diretas entre essas redes, sem a participação obrigatória de provedores, são possíveis, reais, e parte integrante da internet, e não apenas uma abstração teórica na sua definição.

Troca de tráfego

Mesmo técnicos experientes em redes, por vezes, fazem confusão quando se trata desse assunto. Não se está afirmando aqui que os provedores sejam dispensáveis.

Para a maioria das situações eles devem ser o principal meio de conexão à internet. Contudo, pode-se identificar outros participantes da internet com quem a comunicação através da rede seja relevante e estabelecer um enlace físico direto, trocando através dele o tráfego que antes passava pelo provedor.

Isso se chama troca de tráfego, e continua sendo parte da internet. Essa substituição de uma relação de compra de trânsito pela troca de tráfego está ilustrada na Figura 5.11, onde a primeira situação está representada do lado esquerdo e a segunda do lado direito.

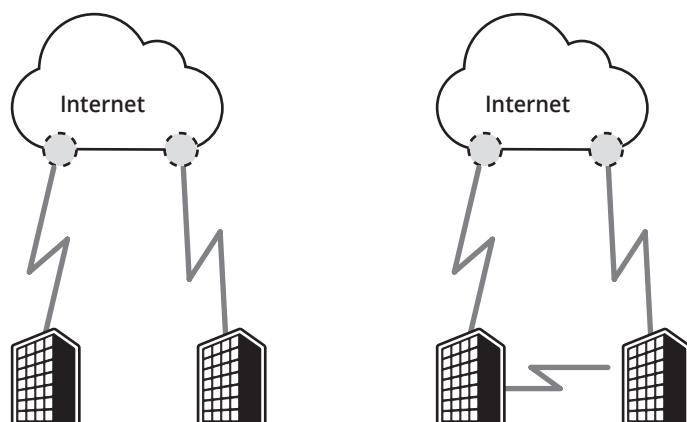


Figura 5.11
Compra de trânsito
x troca de tráfego.

Pode-se identificar, então, dois tipos básicos de relação entre participantes da internet: a compra de trânsito, bem conhecida, onde um provedor fornece acesso a parte ou à totalidade das demais redes interligadas, em troca de dinheiro; e a troca de tráfego (em inglês: peering), onde redes conectam-se diretamente, fornecendo acesso umas às outras mutuamente.

A troca de tráfego traz economia, porque se deixa de pagar ao provedor pelo tráfego que é trocado diretamente com as outras redes. Traz também melhoria de qualidade, porque conexões diretas são mais rápidas e confiáveis.

No entanto, há despesas envolvidas. Enlaces devem ser estabelecidos, o que costuma custar muito caro, especialmente no Brasil. Equipamentos, como roteadores, podem ter de ser trocados. E deve-se dispor de mão de obra especializada, capaz de lidar com as configurações necessárias. Além disso, como muitas vezes a troca de tráfego implica numa relação não comercial, nem sempre há acordos de nível de serviço estabelecidos; em caso de problemas, conta-se com a boa vontade do parceiro para resolvê-los, sem garantias contratuais.

Os Pontos de Troca de Tráfego existem para ajudar os participantes da internet a estabelecer relações de troca de tráfego, mantendo as vantagens já apresentadas, mas reduzindo as despesas e problemas envolvidos. O conceito em que se baseiam é extremamente simples: consistem numa estrutura centralizada, onde várias redes podem se interligar.

Dessa forma, não são necessários vários enlaces distintos para estabelecer relações de troca de tráfego com diferentes redes, mas apenas um enlace, para o PTT. Esse conceito está ilustrado na Figura 5.12.

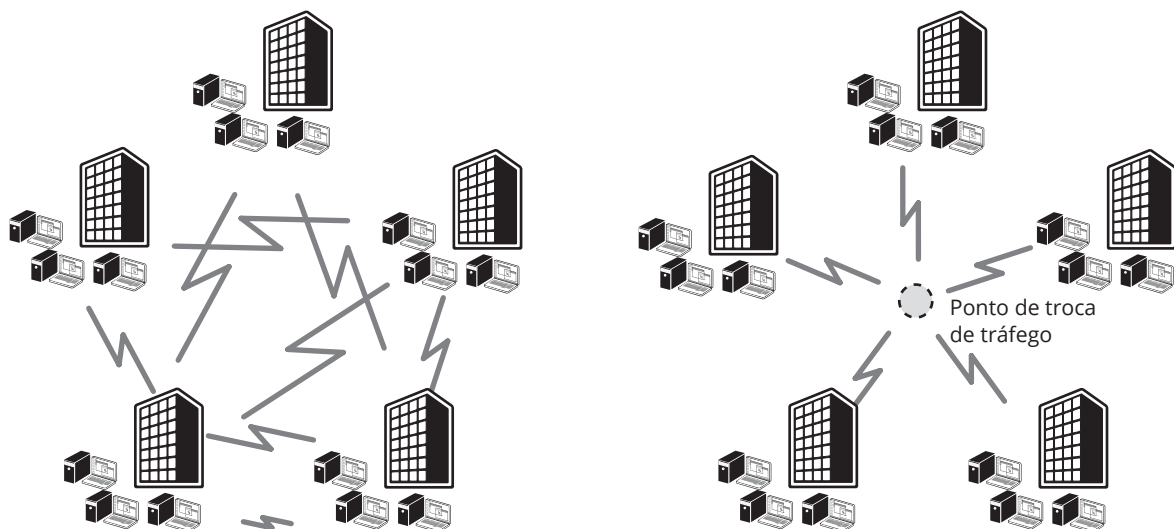


Figura 5.12
Pontos de Troca de Tráfego (PTT).

Uma vez conectadas, as empresas e instituições podem fazer acordos bilaterais ou multilaterais para troca de tráfego, de caráter comercial ou não. Mesmo relações de compra de trânsito podem também ser estabelecidas através dos PTTs, com um ou mais provedores e em conjunto ou não com relações de troca de tráfego, embora isso não seja o objetivo principal de sua existência.

Costuma-se fazer uma analogia, comparando um Ponto de Troca de Tráfego a uma mesa de bar. Várias pessoas podem estar presentes. A cerveja está disponível. O bar oferece um ponto de encontro e toda a infraestrutura necessária. Isso não quer dizer que todos estejam bebendo, e menos ainda que todos bebam ou conversem juntos. Essa possibilidade existe e pode até ser bem interessante, mas as circunstâncias podem levar um pequeno grupo a se reunir para beber e conversar num canto, outro no canto oposto, etc. Há vários tipos de situações que podem levar determinadas redes a terem, ou não, interesse em trocar tráfego com outras, mesmo participando do PTT.

Estar em um PTT não significa a obrigatoriedade em se trocar tráfego com todos os outros participantes, mas traz, isso sim, essa possibilidade. Não se deve confundir os Pontos de Troca de Tráfego com backbones. Os PTTs são regionais, normalmente de caráter metropolitano. Sua função não é carregar o tráfego das redes a longas distâncias, mas sim, melhorar os custos e a qualidade das conexões das redes de uma mesma localidade. O ideal é que haja um PTT por região.

Estrutura da internet

Estrutura hierárquica:

- Provedores de nível 1 – Sprint, Genuity/BBN, AT&T...
- Provedores de nível 2 – Embratel, Telefônica, Telemar...
- Demais provedores.

Troca de tráfego regional via PTTMetro (CGIbr).

A estrutura da internet pode ser considerada, de forma aproximada, como hierárquica. Em seu centro estão os provedores de nível 1, que são aqueles que têm acesso a toda a internet sem necessidade de pagar a ninguém. São exemplos de provedores nível 1 a Sprint, a Genuity/BBN, a AT&T, a UUNet, entre outros. Eles possuem grandes backbones e trocam tráfego entre si diretamente e através de PTTs.



Os provedores que não conseguem acesso a toda a internet através da troca de tráfego devem se tornar clientes dos provedores de nível 1, pagando a eles pela conexão à internet. Eles são chamados de provedores nível 2 e nessa categoria incluem-se nossos principais provedores nacionais, como Embratel, Telefônica, Telemar, Brasil Telecom etc.

A troca de tráfego regional entre os provedores nível 2 brasileiros, e mesmo entre provedores menores e usuários finais, é recomendada, pois traz as vantagens já mencionadas anteriormente: custos menores, com a redução do valor pago aos provedores estrangeiros, e melhoria de qualidade, com diminuição da latência e da taxa de erros nas conexões.

No Brasil, o Comitê Gestor da Internet (CGI.br) lançou mão do projeto PTTMetro, como forma de incentivar e apoiar a troca de tráfego regional. O projeto PTTMetro foi criado em meados de 2004, tendo o escopo inicial de construir cinco PTTs em importantes capitais brasileiras. No final do mesmo ano entrou em operação o PTT de São Paulo.

Em maio de 2008 eram oito os Pontos de Troca de Tráfego do PTTMetro: São Paulo, Porto Alegre, Belo Horizonte, Curitiba, Brasília, Rio de Janeiro, Santa Catarina e Salvador. Juntos, eles são responsáveis por lidar com um tráfego médio de 3,3 Gb/s de dados, e que apresenta picos de 6,4Gb/s.

Em abril de 2012 havia 20 PTTs (URL: www.ptt.br/particip.php) cujo tráfego agregado está mostrado na Figura 5.13 (atualizado em 17/04/2012 15:00hs).



Figura 5.13
Tráfego agregado diário PTTs.

Pontos de acesso

É importante notar que, quando se diz que os PTTs devem ser regionais, ou ainda que há um único PTT numa determinada cidade, não significa que possa haver apenas um único ponto de conexão físico ao PTT. Um Ponto de Troca de Tráfego pode ter vários Pontos de Acesso, chamados também de PIXes. No PTTMetro, empresas particulares, como datacenters, podem ser PIXes. Para isso elas devem estabelecer uma conexão com o PIX Central da região através de uma fibra óptica apagada (que permite grande escalabilidade no tocante ao volume de dados) e arcar com os custos do equipamento local (switch).

Elas podem, então, estabelecer condições e valores para a conexão dos participantes do PTT ao seu PIX. Os equipamentos, no entanto, são administrados pelo NIC.br, representando o Comitê Gestor da Internet, que define também a política de troca de tráfego e do uso em geral do sistema de interconexão. Essa infraestrutura é considerada pelo Comitê Gestor como de uso público e, portanto, seu uso é, hoje, gratuito. Em outras palavras, pode-se pagar ao administrador de um determinado PIX para se conectar, e o preço e condições podem variar de um PIX para outro; no entanto, todo o uso da infraestrutura do PTT, seja para troca de tráfego ou compra de trânsito de outros participantes, é gratuito. Não se paga pelo volume de tráfego trocado.

A Figura 5.14 mostra a estrutura do PTT de São Paulo, para permitir a melhor compreensão do conceito de PIX. Há um PIX Central, no NIC.br; um PIX acadêmico, na USP; e diversos PIX comerciais, na Brasil Telecom, Locaweb etc. Não importa onde um determinado participante se conecte, a comunicação com os demais é transparente. Um participante conectado, por exemplo, no PIX Locaweb, pode ter um acordo de troca de tráfego com outro, digamos, ligado ao PIX Tivit.

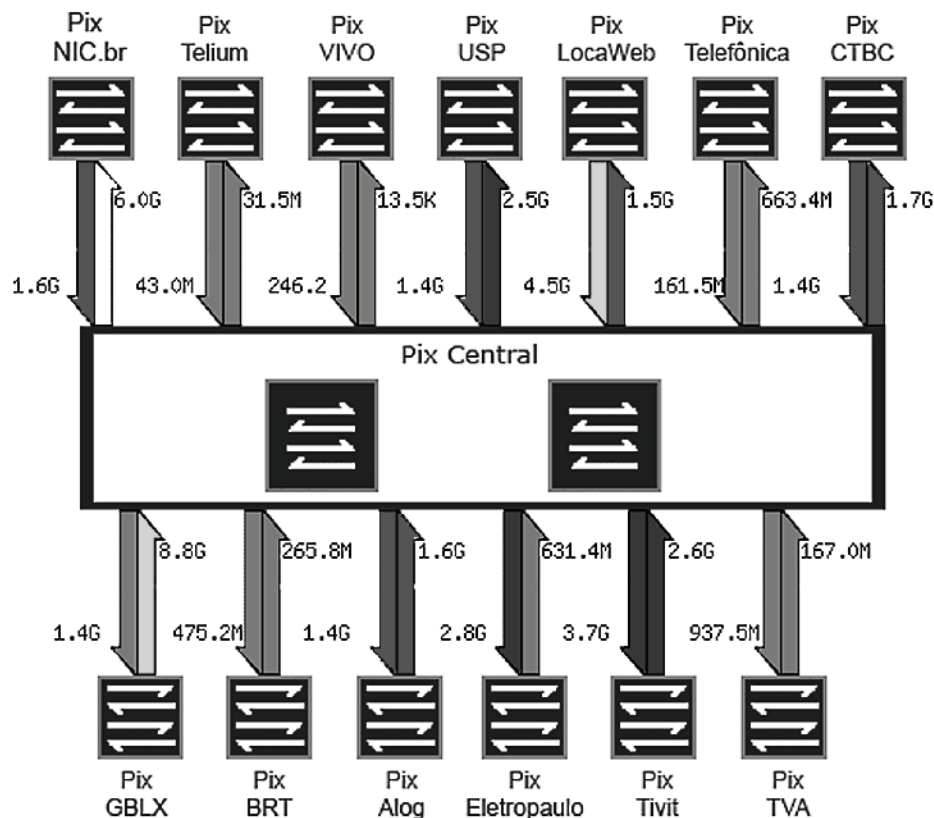


Figura 5.14
Estrutura do PTT de
São Paulo do NIC.br.

Como no Brasil os custos dos enlaces locais são muito altos, essa diversidade de PIXes colabora de forma importante para o sucesso do projeto. Um participante pode escolher conectar-se ao PIX que implicará num custo de enlace menor. Como muitos dos PIXes são datacenters comerciais, pode também existir o caso em que toda a rede ou parte importante da rede da instituição participante esteja dentro do próprio datacenter, levando o custo de conexão para próximo de zero.

O PTTMetro tem hoje participantes importantes, como os principais provedores de banda larga: Brasil Telecom, Oi, Telefônica, Embratel, CTBC Telecom, GVT e Net. Conta também com a participação da RNP, que conecta as principais universidades e centros de pesquisa do país. Tem ainda os principais conteúdos da internet brasileira, através da participação da Locaweb, Terra, Yahoo! e UOL. São cerca de 80 participantes no total, alguns dos quais presentes em mais de um dos PTTs.

Conexão do AS ao PTT

Cada AS deve ser registrado no IANA, obtendo assim seu número de AS (ASN – Autonomous System Number). Esse número é usado na configuração do protocolo BGP, como já vimos na sessão anterior. Veja os RFCs: 1930, 4893 e 5398.

Para se conectar à internet, a entidade interessada deve fazê-lo inicialmente através de um provedor no modelo conhecido como Modelo Acesso IP – Entidade Cliente ISP, mostrado na Figura 5.15 a seguir. Nesse modelo a entidade não tem opções de conexão direta com redes externas (internet). Para a internet a entidade faz parte do ISP A, que provê seu acesso.

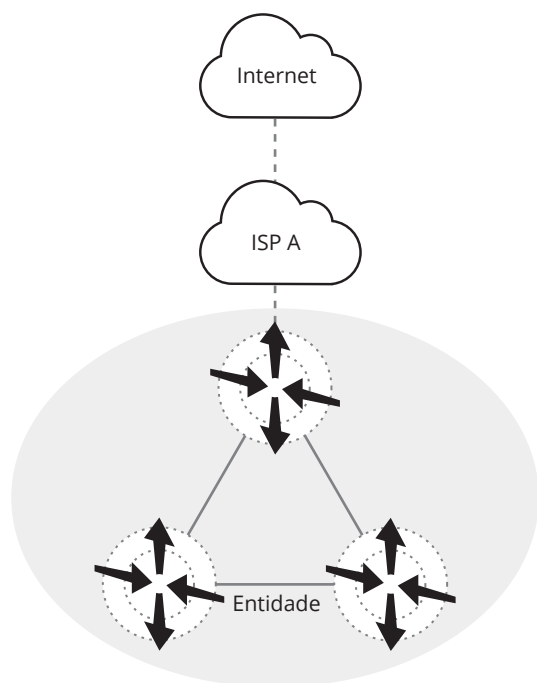


Figura 5.15
Modelo Acesso IP –
Entidade Cliente ISP.

Se a entidade tiver um ASN, ela pode se conectar ao PTT, além de se conectar ao provedor, conforme mostrado na Figura 5.16.

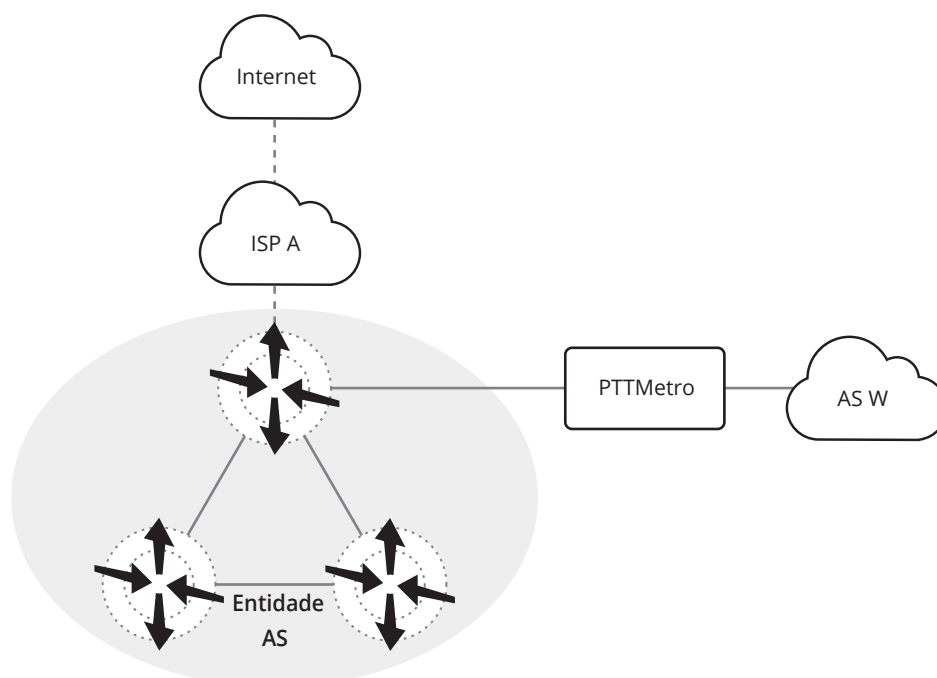


Figura 5.16
Modelo Acesso IP –
Entidade AS.

Nesse modelo a entidade AS tem opção de conexão com o AS W via internet (ou com outros ASs) ou via PTT, conforme mostra a Figura 5.17.

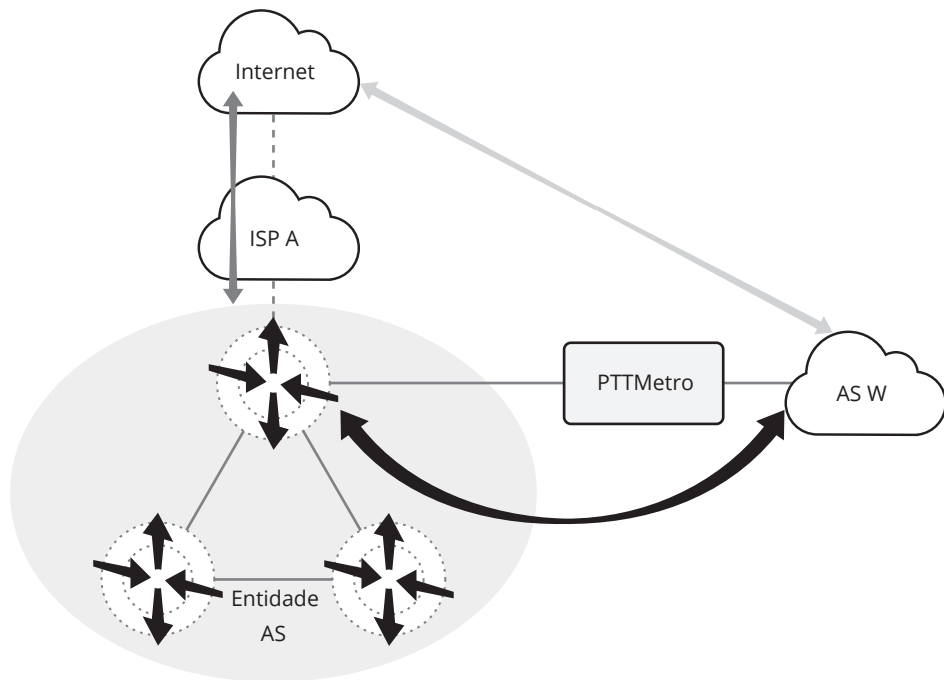


Figura 5.17
Opções de conexão
da Entidade AS.

Podemos então visualizar a inter-rede internet não apenas como uma interligação de redes físicas, mas sim como um conjunto de sistemas autônomos (AS) interconectados, conforme esquematizado na Figura 5.18.

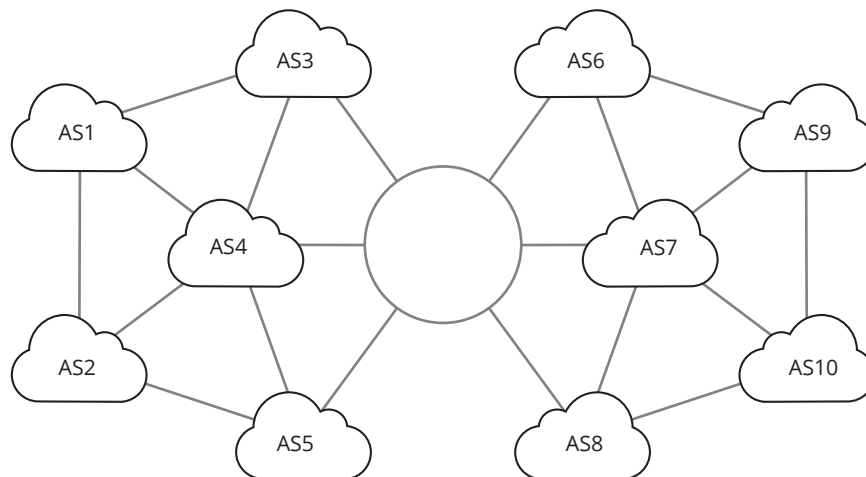


Figura 5.18
A internet como
um conjunto de sis-
temas autônomos
interconectados.

Utilização de Endereçamento IP Portável (PI).

- Espaço de endereçamento IP Próprio.
- Redundância.
- Acordos de troca de tráfego.
- Critérios para se tornar um AS:
 - ▣ Estrutura complexa de rede.
 - ▣ Duas conexões para acesso à internet.
 - ▣ Equipe técnica capacitada.
 - ▣ Equipamentos com suporte a BGP.
 - ▣ Condições financeiras.



A motivação para tornar-se um AS pode ser resumida nos seguintes pontos positivos:

- Utilização de Endereçamento IP Portável (PI): quando a entidade torna-se independente de provedor, o processo de troca do provedor de acesso à internet passa a ser mais simples, pois não envolve mudanças de configuração interna.
- Espaço de endereçamento IP Próprio: para alocação de endereços IP públicos diretamente para clientes, o que melhora a utilização de algumas aplicações, facilita o processo de rastreabilidade de clientes (segurança) etc.
- Redundância: possibilita a implementação de redundância do acesso à internet, pela conexão com dois ou mais provedores, e aumento da disponibilidade dos serviços prestados.
- Acordos de troca de tráfego: possibilita a conexão da entidade com pontos de troca de tráfego (e.g. PTTMetro) e o estabelecimento de acordos multilaterais e bilaterais, o que pode resultar em economia de recursos com a contratação de banda e melhor qualidade de interconexão.

Critérios para se tornar um AS:

- Estrutura com complexidade mínima de rede (hoje medida pela necessidade de endereços IPv4).
- Duas conexões para acesso à internet ou uma conexão com a internet e um link de transporte L2 até um Ponto de Troca de Tráfego (e.g. PTTMetro).
- Equipe técnica capacitada para criar, implementar e operar a política de roteamento da entidade, pela utilização do protocolo BGP.
- Equipamentos com suporte (recursos de hardware e software) para utilizar o protocolo BGP.
- Condições financeiras para implantação e operação/administração da nova estrutura.

A entidade pode se conectar à internet através de um PTT, cujo objetivo é viabilizar a conexão direta entre as entidades que compõe a internet, os ASs. A interconexão entre os ASs é otimizada pelo PTT, pois possibilita melhor qualidade (menor latência) de tráfego entre os ASs, menor custo e maior organização da estrutura de rede regional (pontos concentradores).

O PTT é estruturado segundo o modelo de matriz de comutação, esquematizado na Figura 5.19.

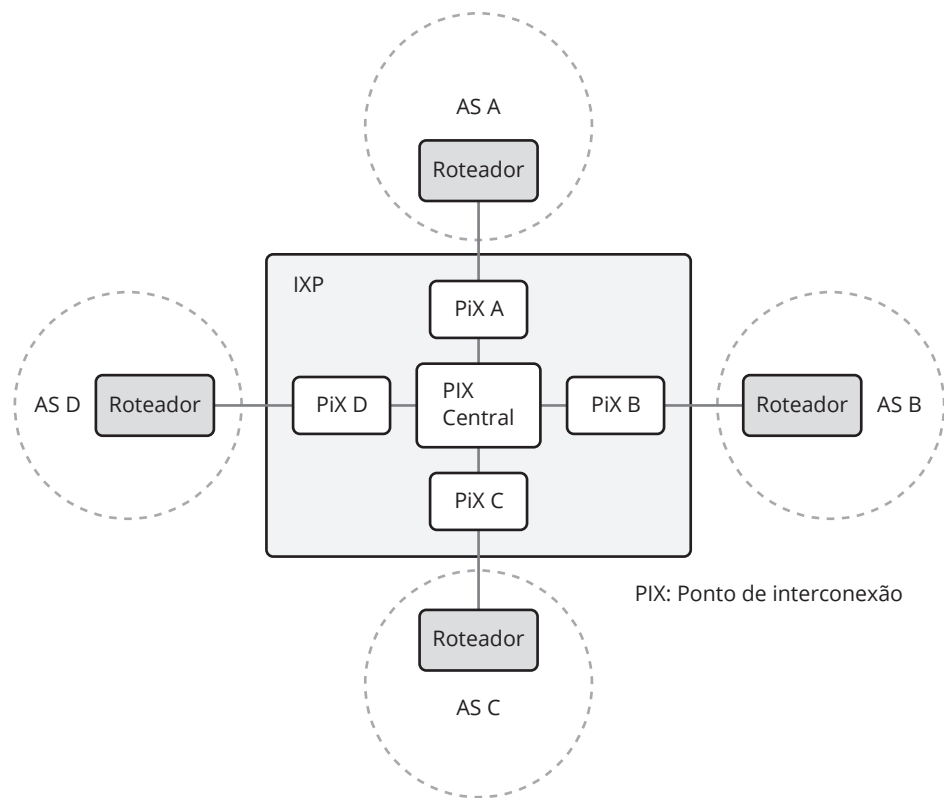


Figura 5.19
Matriz de comu-
tação do PTTMetro.

PTTMetro é o nome dado ao projeto do Comitê Gestor da Internet no Brasil (CGI.br), que promove e cria a infraestrutura necessária (Ponto de Troca de Tráfego – PTT) para a interconexão direta entre as redes (Autonomous Systems – ASs) que compõem a internet brasileira. A atuação do PTTMetro é voltado às regiões metropolitanas no país que apresentam grande interesse de troca de tráfego internet.

O projeto arca com os equipamentos ativos (hardware), responsáveis pela transmissão intra e inter PiXes e pelas interfaces de conexão dos participantes. Não há repasse de custo para os participantes, sobre as suas interfaces de conexão, independente da capacidade (Fast Ethernet, Gigabit Ethernet ou 10 Gigabit Ethernet), e mesmo considerando eventual redundância.

Pontos de Interconexão (PIX) proveem ao projeto recursos de infraestrutura: espaço, alimentação elétrica, refrigeração, segurança física e um ou dois (preferência) pares de fibras ópticas apagadas até o PIX central. O CGI.br não tem planos de interconectar as localidades do PTTMetro e competir com as operadoras de telecomunicações.

Numa determinada região, ASs com localização física relativamente próxima, podem trocar tráfego entre si através de seus respectivos provedores de trânsito (ISP), junto com o restante do tráfego internet. A Figura 5.20 mostra essa situação.

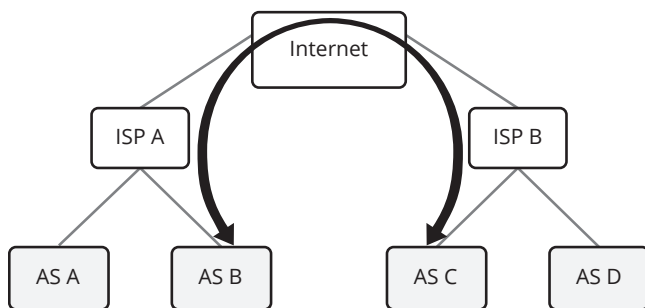


Figura 5.20
Troca de tráfego entre ASs relativamente próximos.

Com um PTT na região, os AS participantes podem trocar tráfego entre si pelo PTT (menor custo e latência) e deixar os seus links de trânsito para acesso aos outros AS da internet. A Figura 5.21 mostra essa situação.

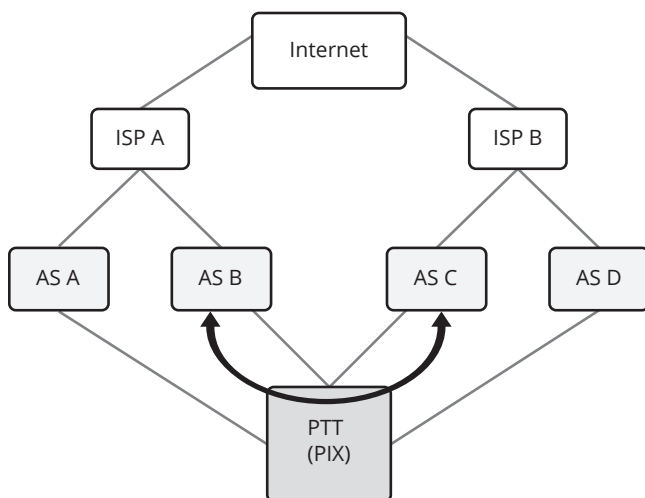
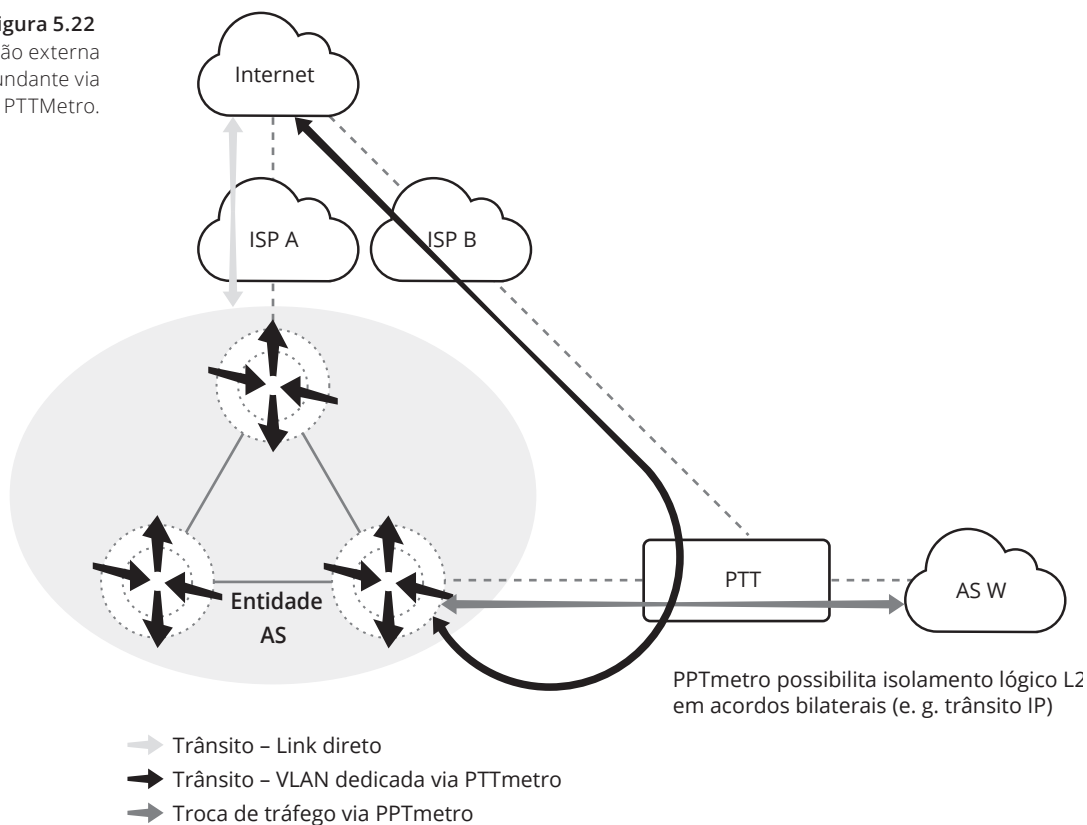


Figura 5.21
Troca de tráfego entre ASs através de um PTT regional.

A interconexão com um PTT permite também várias alternativas de conexão externa redundante através de acordos bilaterais entre os ASs participantes, conforme mostrado na Figura 5.22.

Figura 5.22
Conexão externa
redundante via
PTTMetro.



Looking Glass é um servidor (ou roteador) de acesso público que tem uma cópia da tabela de roteamento BGP: <http://www.bgp4.com.br/lg>. Para mais informações sobre um determinado AS, por exemplo, o AS1916 da RNP: <http://bgp.potaroo.net/cgi-bin/as-report?as=AS1916>



No acordo de troca de tráfego multilateral (ATM), os participantes estabelecem sessões BGP apenas com os dois servidores de rotas do PTT e não com todos os demais participantes. Os participantes do ATM ainda podem optar por se conectar ao servidor de Looking Glass, para alimentá-lo com seus prefixos, e assim contribuir com o projeto. A figura 5.23 exemplifica essa possibilidade.

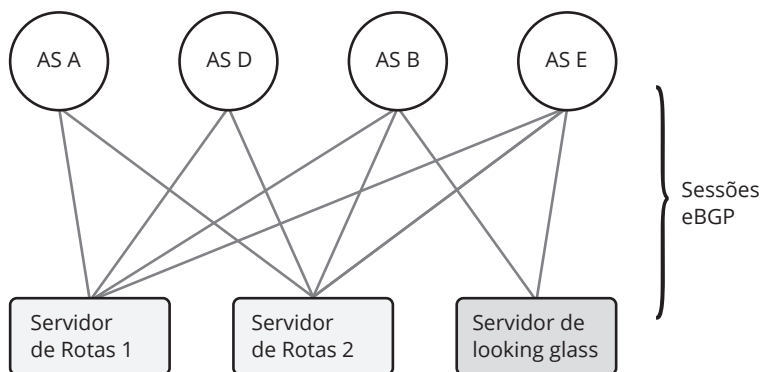


Figura 5.23
Acordo de troca de
tráfego multilateral
(ATM).

Anúncios de rotas

Os anúncios de rotas devem ser feitos com muito cuidado, porque você está anunciando para a internet (o mundo todo), que tem rota para determinadas redes. Se houver algum engano nesses anúncios, na prática você está contando uma mentira para o mundo todo. Vamos analisar um caso que ocorreu entre o Paquistão e o YouTube, que está bem documentado e ilustra bem esse ponto.

Estudo de caso:

- 24/02/2008 Pakistan Telecom (AS 17557) fez anúncio não autorizado: 208.65.153.0/24.
- PCCW Global (AS 3491) propagou este anúncio para toda a internet.
- Sequestro de tráfego do YouTube.



Em 24 de fevereiro de 2008, domingo, a empresa Pakistan Telecom (AS 17557) iniciou um anúncio não autorizado do prefixo 208.65.153.0/24, que pertence ao YouTube. Um dos provedores da Pakistan Telecom, a PCCW Global (AS 3491), encaminhou esse anúncio para o resto da internet, o que resultou no sequestro do tráfego do YouTube numa escala global.

Linha do tempo

- **Antes, durante e depois de domingo, 24/2/2008:** AS 36561 (YouTube) anuncia o prefixo 208.65.152.0/22, entre outros (que não foram envolvidos nesse evento).
- **Domingo, 24/2/2008, 18:47 (UTC):** AS 17557 (Pakistan Telecom) inicia o anúncio não autorizado do prefixo 208.65.153.0/24. Esse prefixo pertence ao bloco anunciado pelo YouTube, que abrange as redes 208.65.152.0/24, 208.65.153.0/24, 208.65.154.0/24 e 208.65.155.0/24. O provedor AS 3491 (PCCW Global) propaga esse anúncio. Roteadores ao redor do mundo recebem esse anúncio e, como consequência, o tráfego do YouTube para esse prefixo é redirecionado para o Paquistão.
- **Domingo, 24/2/2008, 20:07 (UTC):** AS 36561 (YouTube) percebe o que está acontecendo e inicia o anúncio do prefixo 208.65.153.0/24. Note que esse prefixo é uma rota mais específica do que o anúncio do bloco 208.65.152.0/22, mas ainda não foi suficiente para recuperar o tráfego, porque com dois prefixos idênticos no sistema de roteamento, as regras de política do BGP, tais como a preferência pelo caminho mais curto (shortest AS path), determinam qual a rota a ser escolhida. Isso significa que o AS 17557 (Pakistan Telecom) continua a atrair algum tráfego do YouTube.
- **Domingo, 24/2/2008, 20:18 (UTC):** AS 36561 (YouTube) percebe que não deu certo e inicia o anúncio dos prefixos 208.65.153.0/25 e 208.65.153.128/25, que são rotas mais específicas do que a rota para o prefixo 208.65.253.0/24. Devido à regra da rota mais específica (prefixo maior: /25 é maior do que /24), cada roteador que recebe esses anúncios envia o tráfego para o YouTube.
- **Domingo, 24/2/2008, 20:51 (UTC):** todos os anúncios de prefixos, incluindo o sequestrado /24 que foram originados pelo AS 17557 (Pakistan Telecom) via AS 3491 (PCCW Global), são vistos por todos como tendo o prefixo 17557, parecendo que o AS path é maior do que é realmente, fazendo com que os roteadores prefiram o anúncio originado pelo YouTube.
- **Domingo, 24/2/2008, 21:01 (UTC):** AS 3491 (PCCW Global) remove todos os prefixos originados pelo AS 17557 (Pakistan Telecom), interrompendo o sequestro do prefixo 208.65.153.0/24. Note que o AS 17557 não foi completamente desconectado pelo AS 3491. Prefixos originados por outros ASs do Paquistão ainda continuam sendo anunciados pelo AS 17557 através do AS 3491.



Este estudo de caso está documentado no vídeo "YouTube Hijacking: A RIPE NCC RIS case study".

Observe que esses eventos ocorreram num período de tempo de pouco mais de duas horas. Os prefixos envolvidos nesse sequestro e as contramedidas tomadas pelo YouTube foram relatadas pelo RIPE NCC's Routing Information Service (RIS). O que o Paquistão pretendia fazer era bloquear o website do YouTube, especificamente três endereços IP: 208.65.153.238, 208.65.153.251 e 208.65.153.253. Acontece que esse anúncio, que devia ter efeito internamente no Paquistão, foi repassado para a internet pelo provedor PCCW Global, resultando num efeito global indesejado.

A Figura 5.24 resume o início e o fim desse evento.

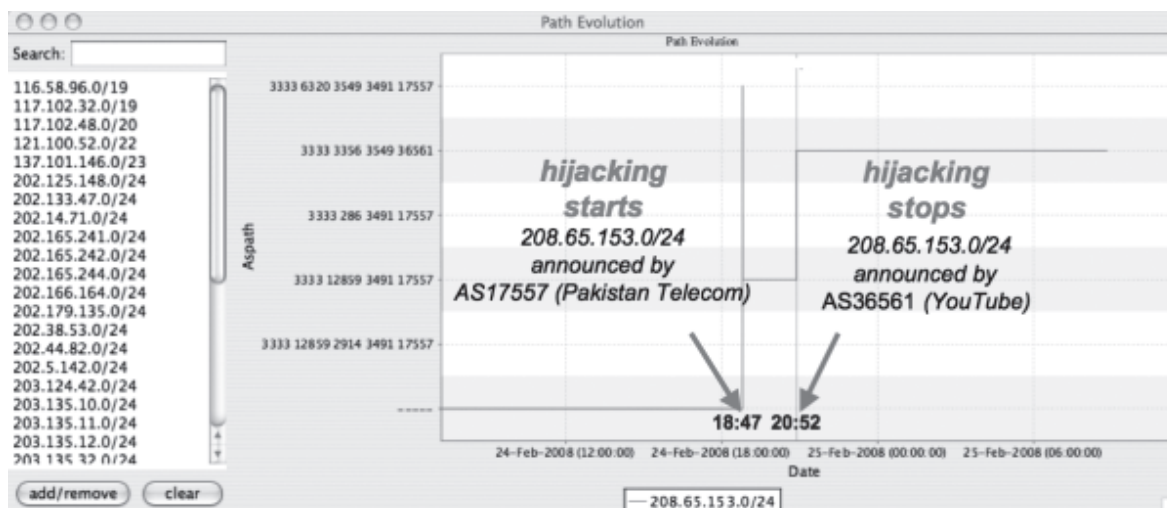


Figura 5.24
Início e fim do sequestro do tráfego do YouTube pelo Paquistão.

A Figura 5.25 mostra a situação de tráfego durante o sequestro e a Figura 5.26 mostra situação de tráfego depois do final do sequestro.

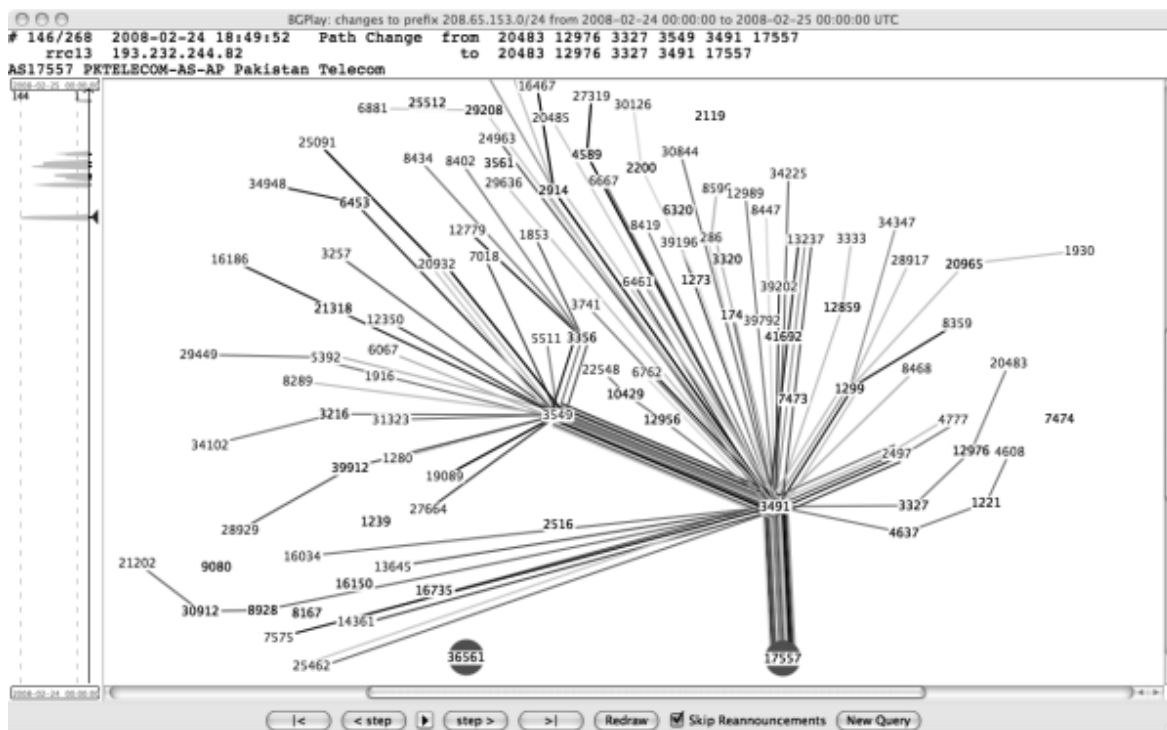


Figura 5.25
Situação de tráfego durante o sequestro.

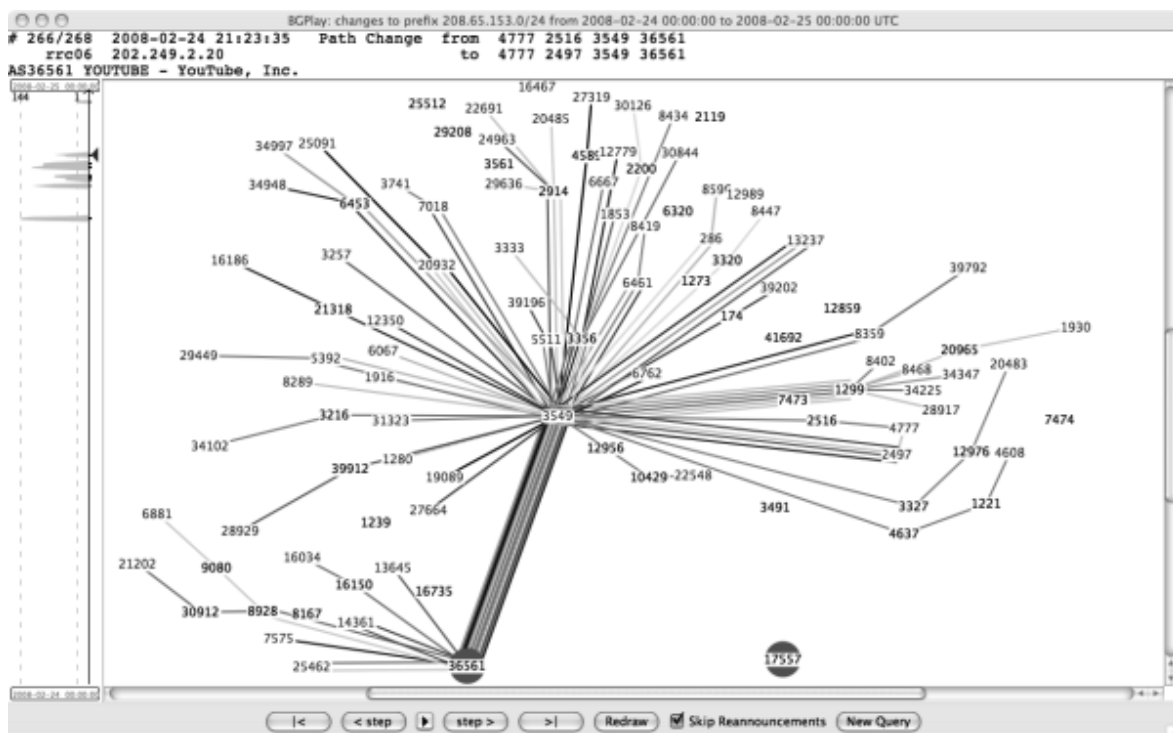


Figura 5.26
 Situação de tráfego
 depois do final do
 sequestro.

Roteiro de Atividades 5

Atividade 5.1 – Configuração do protocolo BGP

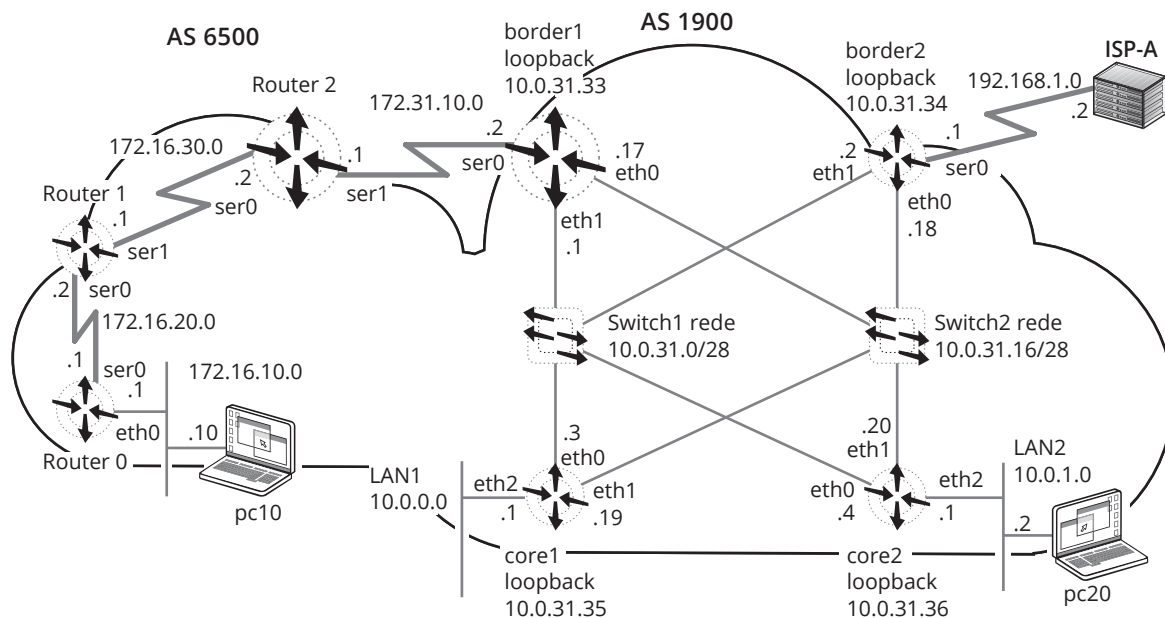


Figura 5.27
Rede1_Sessao5_
ADR8.

Descrição da rede

Na rede da Figura 5.27 temos dois ASs: 6500 e 1900 com Route Reflector.

- O AS 6500 é composto por três roteadores: router0, router1 e router2 e das redes 172.16.0.0/16.
- O router0 está configurado com duas interfaces Ethernet: eth0 (IP: 172.16.10.1/24) e eth1 (IP: 172.16.20.1/24).
- O router1 está configurado com duas interfaces Ethernet: eth0 (IP: 172.16.20.2/24) e eth1 (IP: 172.16.30.1/24).
- O router2 está configurado com duas interfaces Ethernet: eth0 (IP: 172.16.30.2/24) e eth1 (IP: 172.31.10.1/24).
- Todos os roteadores usam o protocolo OSPF (já está configurado) e são vizinhos BGP.
- O router2 é Route Reflector (RR) para os demais roteadores do AS 6500; portanto, mantém sessões iBGP com os roteadores router1 e router0. Os roteadores router0 e router1 não são vizinhos entre si e não mantêm sessões iBGP entre eles. Além disso, o router2 mantém uma sessão eBGP com o border1 e é a rota padrão para saída do AS 6500.
- O AS 1900 é composto por quatro roteadores: border1, border2, core1 e core2, e das redes 10.0.0.0/16 e 192.168.0.0/16. Somente a rede 10.0.0.0/16 está sendo anunciada pelo BGP para fora do AS 1900. Esse AS tem uma configuração dual, com caminhos redundantes entre os quatro roteadores. É uma configuração semelhante às usadas pelas empresas que proveem acesso à internet para seus clientes, inclusive para provedores de conteúdo (como o ISP-A). Observe que são usadas duas redes para permitir o estabelecimento dos caminhos duais: rede 10.0.31.0/28 e rede 10.0.31.16/28. As interfaces de loopback dos roteadores usam endereços IP da rede 10.0.31.32/28.

- Os quatro roteadores são identificados por suas interfaces de loopback: 10.0.31.33, 10.0.31.34, 10.0.31.35 e 10.0.31.36, respectivamente. Essa identificação tem a vantagem de garantir maior estabilidade nas conexões TCP do protocolo BGP, porque a interface de loopback, que é uma interface virtual, nunca sai do ar. Esse tipo de identificação é recomendado para uso local, somente dentro do AS. Para outros ASs, recomenda-se usar endereço IP de uma interface física do roteador de borda.
- O border1 é Route Reflector (RR) para os demais roteadores do AS 1900; portanto, mantém sessões iBGP com os roteadores border2, core1 e core2. Os roteadores border2, core1 e core2 não são vizinhos entre si e não mantêm sessões iBGP entre eles. Além disso, o border1 mantém uma sessão eBGP com o router2 e é a rota padrão para saída do AS 1900.
- O border1 está configurado com três interfaces Ethernet: eth0 (IP: 10.0.31.17/28), eth1 (IP: 10.0.31.1/28) e eth2 (IP: 172.31.10.2/24).
- O border2 está configurado com três interfaces Ethernet: eth0 (IP: 10.0.31.18/28), eth1 (IP: 10.0.31.2/28) e eth2 (IP: 192.168.1.1/28).
- O core1 está configurado com 3 interfaces Ethernet: eth0 (IP: 10.0.31.3/28), eth1 (IP: 10.0.31.19/28) e eth2 (IP: 10.0.0.1/24).
- O core2 está configurado com 3 interfaces Ethernet: eth0 (IP: 10.0.31.4/28), eth1 (IP: 10.0.31.20/28) e eth2 (IP: 10.0.1.1/24).
- Todos os roteadores usam o protocolo OSPF (já está configurado) e são vizinhos BGP.

Os hosts pc10 e pc20 têm endereços IP: 172.16.10.10/24 e 10.0.1.2/24, respectivamente.

Configuração do protocolo BGP

- Carregar a rede *Rede1_sessao5_ADR8* no simulador Core.
- Verificar as tabelas de rotas OSPF dos roteadores.
- Iniciar o Wireshark na interface eth1 do border1.
- Configurar o protocolo BGP nos roteadores.
- Verificar as tabelas de rotas BGP dos roteadores.
- Analisar o fluxo de pacotes.
- Verificar a continuidade entre os PCs.

Conclusão

Nesta atividade prática aprendemos a:

- Configurar o protocolo BGP;
- Utilizar recursos especiais do protocolo BGP;
- Analisar anúncios de rotas BGP;
- Verificar a conectividade da rede.

Esta rede simula a situação de interconexão de um AS privado (AS 6500) com um AS público (AS 1900). O recurso usado para que o OSPF saiba como rotear para fora do AS é definir uma rota padrão de saída do AS através do comando *default-configure*.

No capítulo anterior foi feita a configuração através da redistribuição das rotas BGP para o protocolo OSPF, que não é adequada para este caso, conforme já dissemos.

6

Resolução de problemas

objetivos

Estudar os procedimentos de testes e correção de erros.

Comandos *ping* e *traceroute*, configurações de roteadores e hosts, verificação das rotas aprendidas pelos roteadores, Wireshark para captura de pacotes.

conceitos

Orientações gerais

- Entender o problema.
 - ▣ Geralmente o problema é relatado pelo usuário e as informações são insuficientes.
- Fazer o diagnóstico correto.
- Pode haver mais de um erro de configuração.
- Usar ferramentas disponíveis.



Na resolução de problemas, muitas vezes as informações fornecidas pelos usuários são insuficientes e tecnicamente incorretas. O usuário se baseia no feeling dele, naquilo que ele está vendo e pensando que está ocorrendo, o que nem sempre é necessariamente a realidade.

É necessário um diagnóstico correto do problema antes de fazer qualquer modificação na configuração. Procure usar procedimentos que não alterem a “cena do crime”. Não é uma boa ideia partir para “tentativa e erro”. Em geral esse procedimento aleatório introduz mais erros e até mascara o problema real.

Lembre-se de que o problema pode ser ocasionado por uma série de erros e não apenas por um.

Procure usar todas as ferramentas disponíveis para auxílio no diagnóstico e verificação.

Formação de grupos de trabalho

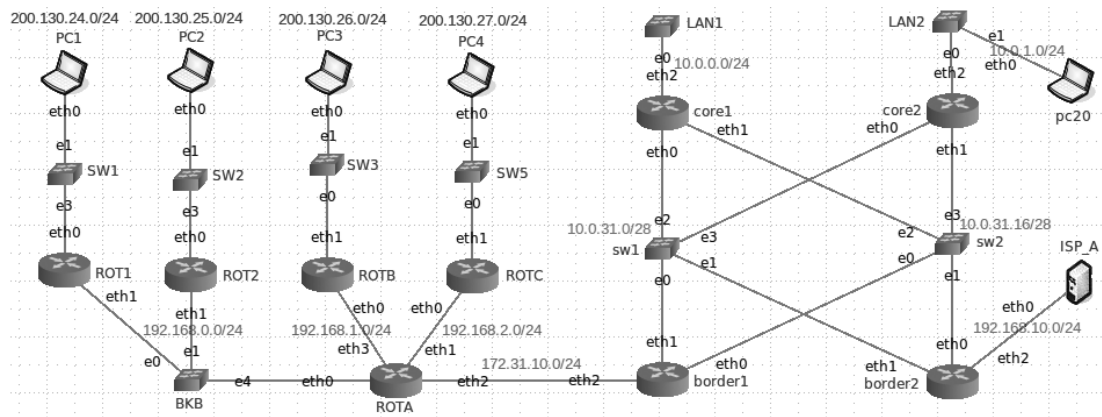
- Formar até seis grupos de trabalho.
- Cada grupo deve resolver um problema proposto.
- Tempo para resolução: 90 minutos.
- Cada grupo deve preparar uma apresentação da solução para os demais grupos.

O instrutor deve orientar a formação dos grupos e distribuir os problemas entre eles.

O tempo previsto para solução é de 90 minutos.



Problema 1



Essa rede apresenta uma configuração simétrica de roteadores com rotas alternadas no AS da direita. Em caso de falha de uma interface que está conectada ao sw1, por exemplo, há uma rota alternativa passando pelo sw2 e vice-versa.

Figura 6.1
Rede1_Sessao6_
ADR8.

Por exemplo, o PC1 pode atingir o PC20 passando pelo ROTA, border1, sw1 e core2. O caminho alternativo é: ROTA, border1, sw2 e core2. Ambas as rotas são iguais em termos de custo (número de hops).

Problema 2

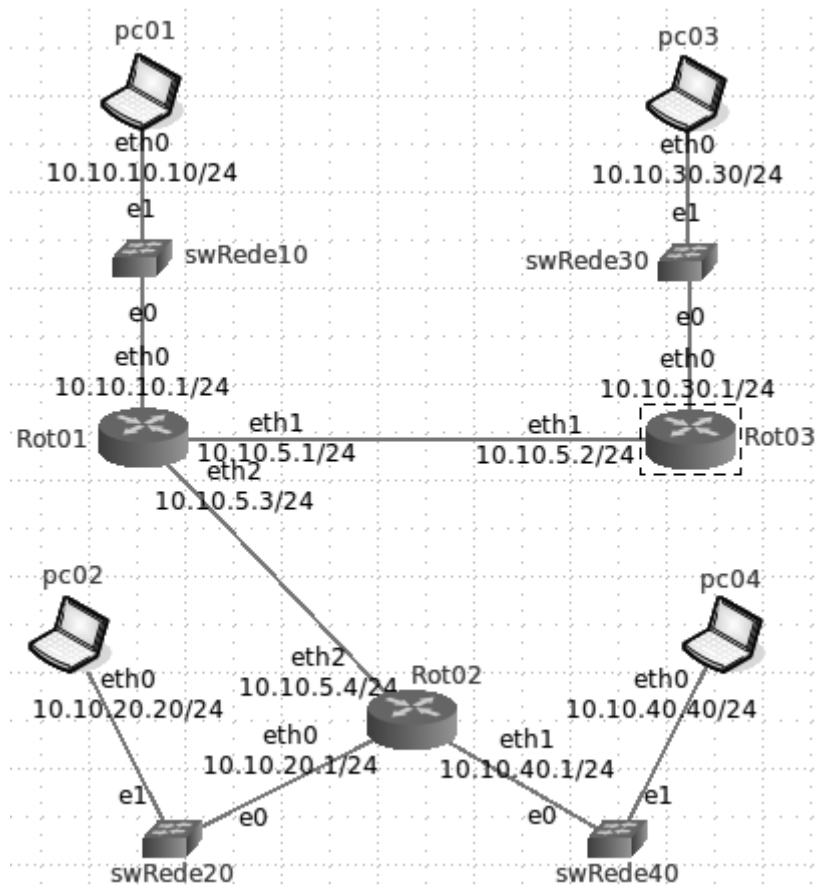


Figura 6.2
Rede2_Sessao6_
ADR8.

Essa rede apresenta quatro redes locais interligadas por três roteadores em diferentes localidades. Portanto, as ligações entre os roteadores constituem uma rede WAN que serve de “ponte” entre as redes LAN. O Rot02 tem duas redes locais, enquanto os outros têm apenas uma rede local cada um. Os enlaces entre os roteadores Rot01 e Rot03 e entre Rot01 e Rot02 são linhas de longa distância dedicadas (enlaces seriais).

Problema 3

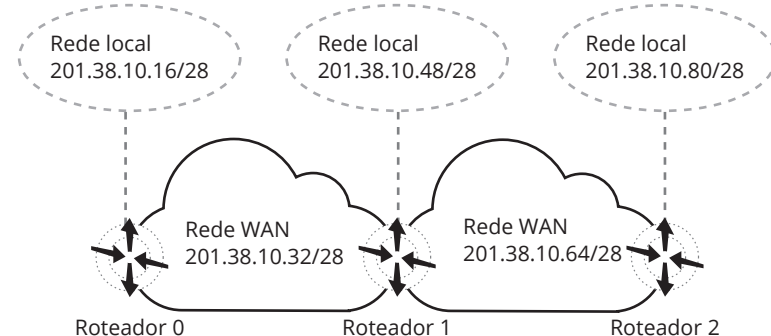


Figura 6.3
Rede3_Sessao6_
ADR8.

Essa rede apresenta três redes locais em localidades remotas interligadas por três roteadores. Foi utilizada uma subdivisão de uma rede Classe C que estava disponível: 201.38.10.0/24. A figura mostra o plano de endereçamento planejado pelo administrador da rede, usando o protocolo RIP.

Problema 4

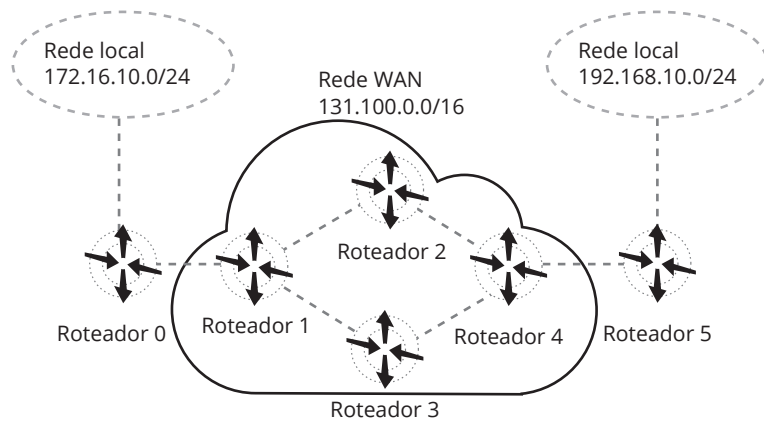


Figura 6.4
Rede4_Sessao6_
ADR8.

Essa rede apresenta duas redes locais interligadas pela rede WAN de um provedor. O provedor utiliza os endereços da rede Classe B (131.100.0.0/16). As redes locais do cliente usam endereços IP privados (RFC 1918), conforme mostrado na figura.

Problema 5

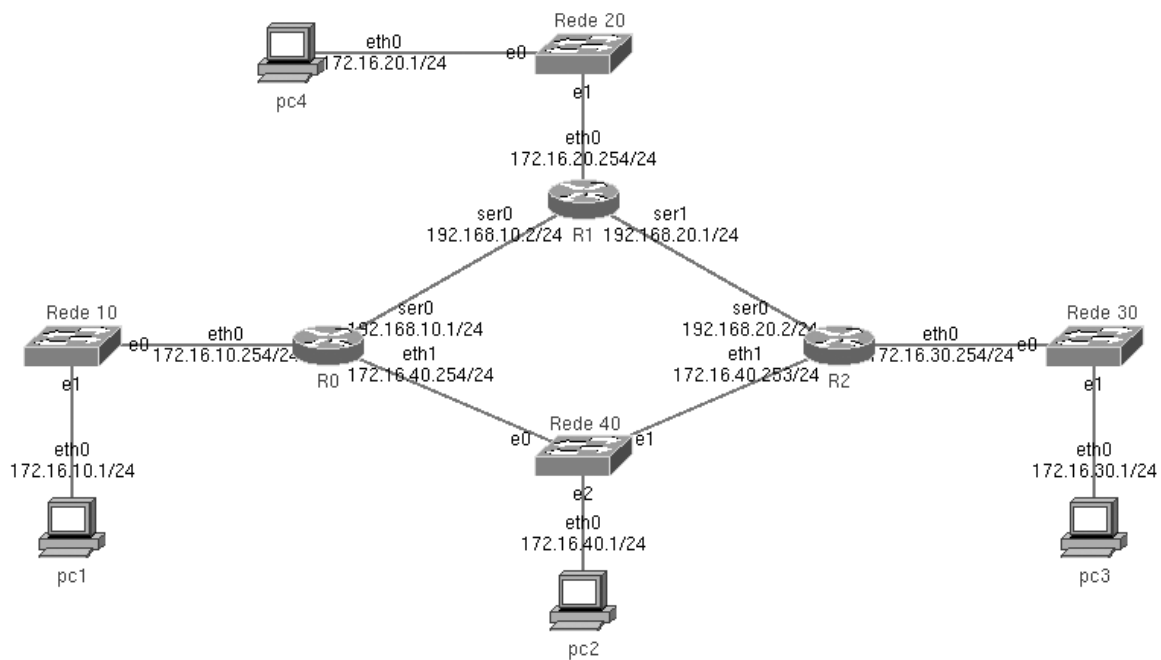


Figura 6.5
Rede5_Sessao6_
ADR8.

Essa rede apresenta quatro redes locais interligadas por três roteadores usando endereçamento privado.

Problema 6

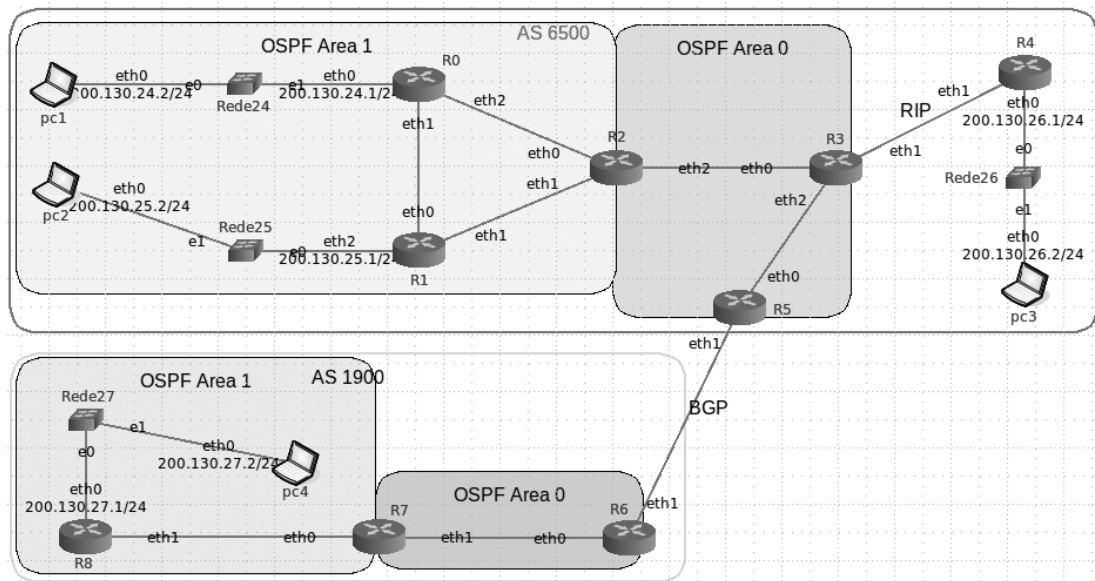


Figura 6.6
Rede6_Sessao6_
ADR8.

Essa rede apresenta dois ASs interligados por um enlace com protocolo BGP. Cada AS tem duas áreas OSPF, sendo que o AS 6500 tem um enlace com o protocolo RIP.

Apresentação das soluções

- Cada grupo apresenta para os demais.
- Tempo por grupo: 15 minutos + 5 minutos para perguntas.
- Tempo previsto total: 80 minutos.

O instrutor deve determinar a ordem de apresentação e controlar os tempos de cada grupo.

A seguir apresentamos a descrição detalhada dos problemas propostos.





Roteiro de Atividades 6

Atividade 6.1 – Configuração do protocolo BGP

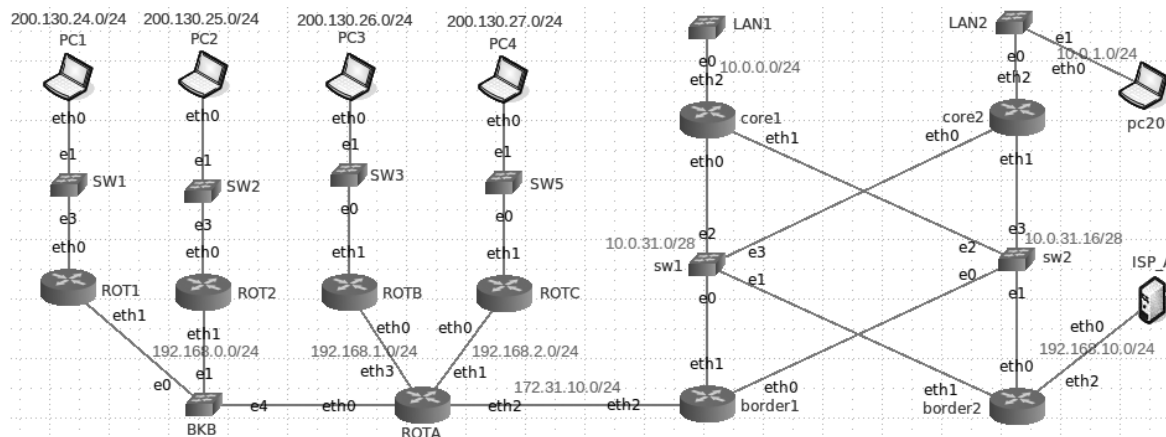


Figura 6.7
Rede1_Sessão6_
ADR8.

Trata-se de duas redes que foram interligadas via enlace serial entre os roteadores ROTA e border1 (rede 172.31.10.0/24). A rede do lado esquerdo (ROTA, ROTB, ROTC, ROT1 e ROT2) usa endereçamento público do bloco 200.130.24.0/22 para as estações dos usuários e endereçamento privativo da faixa 192.168.0.0 para interligação dos roteadores.

A rede do lado direito (border1, border2, core1 e core2) usa endereçamento privativo, sendo que na sua parte central (switches sw1 e sw2) usa rotas alternadas para maior confiabilidade.

Ambas as redes estão configuradas com o protocolo OSPF e a ideia do administrador de rede é interligá-las usando o protocolo BGP.

A tarefa a ser feita é a configuração do protocolo BGP e teste de conectividade entre as redes. Use o número de AS 64500 para a rede da esquerda e AS 64501 para a rede da direita.

Anote todos os procedimentos e aguarde a orientação do instrutor para fazer a apresentação do problema para os outros grupos.

Atividade 6.2 – Configuração de sub-redes

A implementação foi feita utilizando sub-redes da rede 10.10.0.0/16, conforme mostrado na figura.

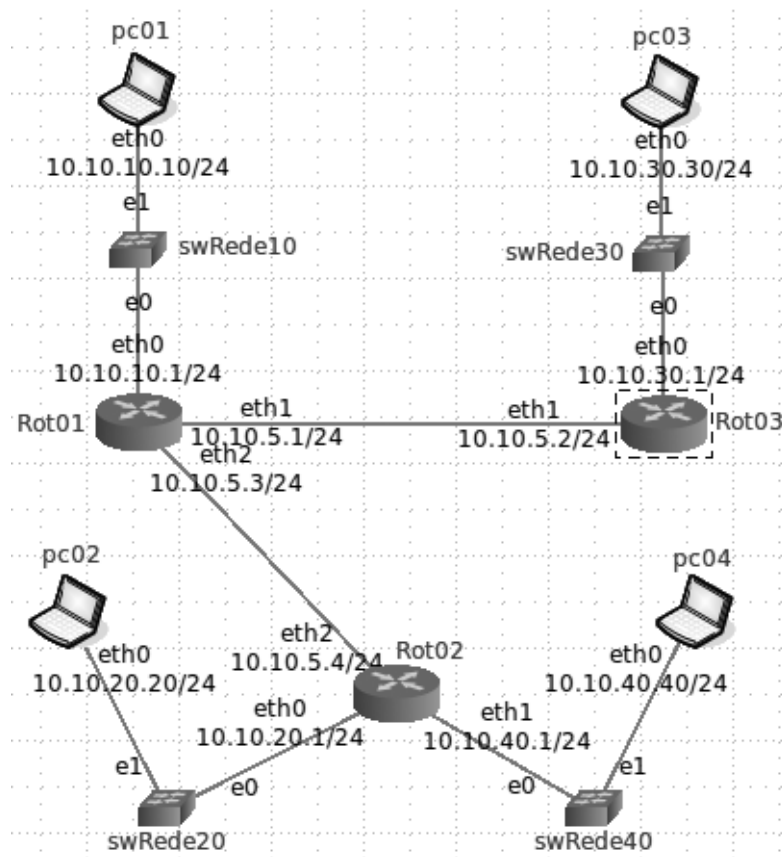


Figura 6.8
Rede2_Sessão6_
ADR8.

Infelizmente, alguns problemas surgiram:

1. O PC-01 não consegue enxergar a rede além do Rot01.
2. O PC-03 não consegue enxergar a rede além do Rot03.
3. O PC-02 não consegue enxergar a rede além do Rot02.
4. O PC-04 não consegue enxergar a rede além do Rot02.
5. Os roteadores Rot01 e Rot03 não enxergam as redes do Rot02.

Siga os seguintes procedimentos para fazer o diagnóstico dos problemas (há mais de um erro de configuração) e corrija-los:

1. Anote os passos do diagnóstico (ping, traceroute etc.) e os problemas encontrados.
2. Anote as correções efetuadas. Procure manter o esquema de endereçamento sempre que possível. Salve a configuração corrigida com outro nome.
3. Teste a continuidade entre todos os PCs.
4. Verifique se todos os roteadores aprenderam as rotas de todas as redes.

Anote todos os procedimentos e aguarde a orientação do instrutor para fazer a apresentação do problema para os outros grupos.

Atividade 6.3 – Configuração de sub-redes

A implementação foi feita utilizando sub-redes da rede 201.38.10.0/24, conforme mostrado na figura.

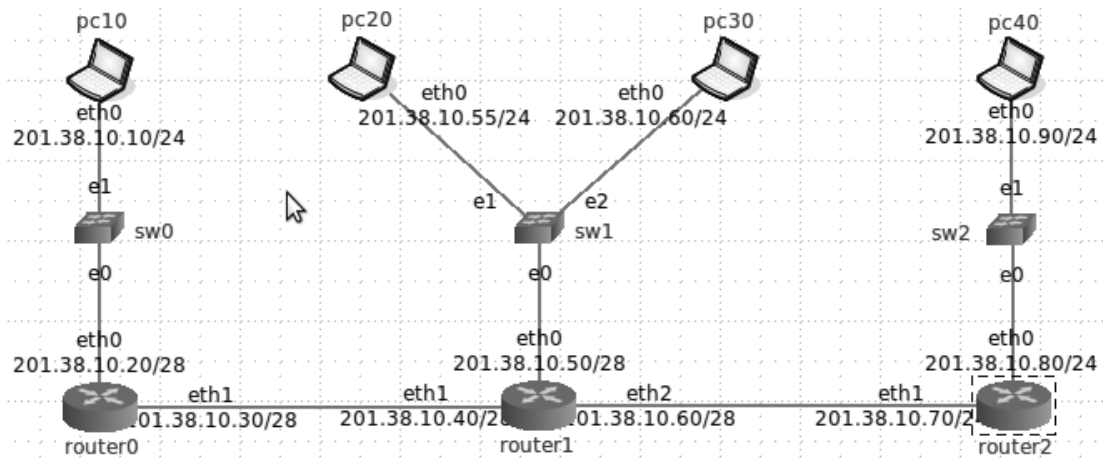


Figura 6.9
Rede3_Sessão6_
ADR8.

Infelizmente, alguns problemas surgiram:

1. Os PCs não conseguem enxergar todas as interfaces de seus respectivos roteadores.
2. A tabela de rotas dos roteadores não mostra todas as sub-redes.

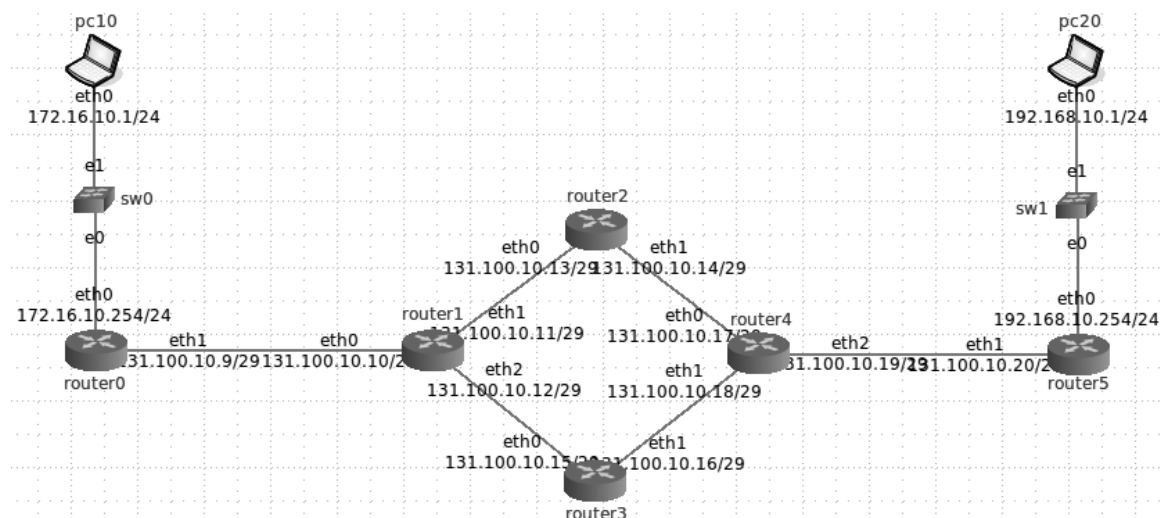
Siga os seguintes procedimentos para fazer o diagnóstico dos problemas (há mais de um erro de configuração) e corrigi-los:

1. Anote os passos do diagnóstico (ping, traceroute etc.) e os problemas encontrados.
2. Anote as correções efetuadas. Procure manter o esquema de endereçamento sempre que possível. Salve a configuração corrigida com outro nome.
3. Teste a continuidade entre todos os PCs.
4. Verifique se todos os roteadores aprenderam as rotas de todas as redes.

Anote todos os procedimentos e aguarde a orientação do instrutor para fazer a apresentação do problema para os outros grupos.

Atividade 6.4 – Projeto de endereçamento IP

A implementação foi feita utilizando sub-redes da rede 131.100.10.0/24 do provedor, conforme mostrado na figura.



Infelizmente, alguns problemas surgiram:

1. Os PCs não conseguem enxergar todas as interfaces de seus respectivos roteadores.
2. A tabela de rotas dos roteadores não mostra todas as sub-redes.
3. O provedor não aceitou a configuração proposta, alegando desperdício de endereços IP.
4. Questão especial:

Curiosamente, testando a continuidade no console do router1, quando se executa o comando `ping 131.100.10.15` (endereço IP da interface eth0/router3), o simulador pergunta: *Do you want to ping broadcast? Then -b*. Por quê?

Siga os seguintes procedimentos para fazer o diagnóstico dos problemas (há mais de um erro de configuração) e corrija-los:

1. Anotar os passos do diagnóstico (ping, traceroute etc.) e os problemas encontrados.
2. Anotar as correções efetuadas. Procure manter o esquema de endereçamento sempre que possível. Salve a configuração corrigida com outro nome.
3. Teste a continuidade entre os PCs.
4. Verifique se todos os roteadores aprenderam as rotas de todas as sub-redes.

Anote todos os procedimentos e aguarde a orientação do instrutor para fazer a apresentação do problema para os outros grupos.

Figura 6.10
Rede4_Sessão06_ ADR8.

Atividade 6.5 – Configuração de rotas estáticas

Essa rede apresenta problemas de roteamento. Os endereços estão corretos e não devem ser alterados.

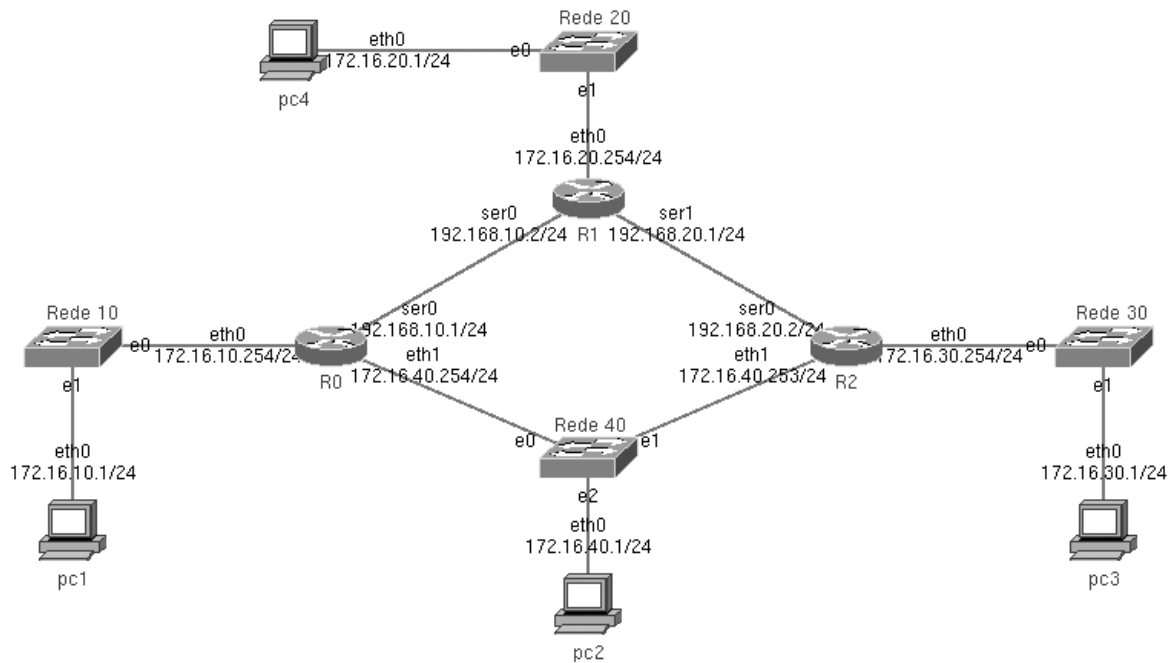


Figura 6.11
Rede5_Sessão6_
ADR8.

Problemas apresentados:

1. Quando o PC1 tenta alcançar as redes 172.16.20.0 e 172.16.30.0, o pacote cai num “loop” de roteamento.
2. Quando o PC2 tenta alcançar a rede 172.16.20.0, o pacote cai num “loop” de roteamento.
3. O PC3 não consegue enxergar as redes 172.16.10.0 e 172.16.20.0; quando tenta alcançar a rede 172.16.40.0, não enxerga a interface eth1 do roteador R0.
4. O PC4 não consegue enxergar as redes 172.16.10.0, 172.16.30.0 e 172.16.40.0.

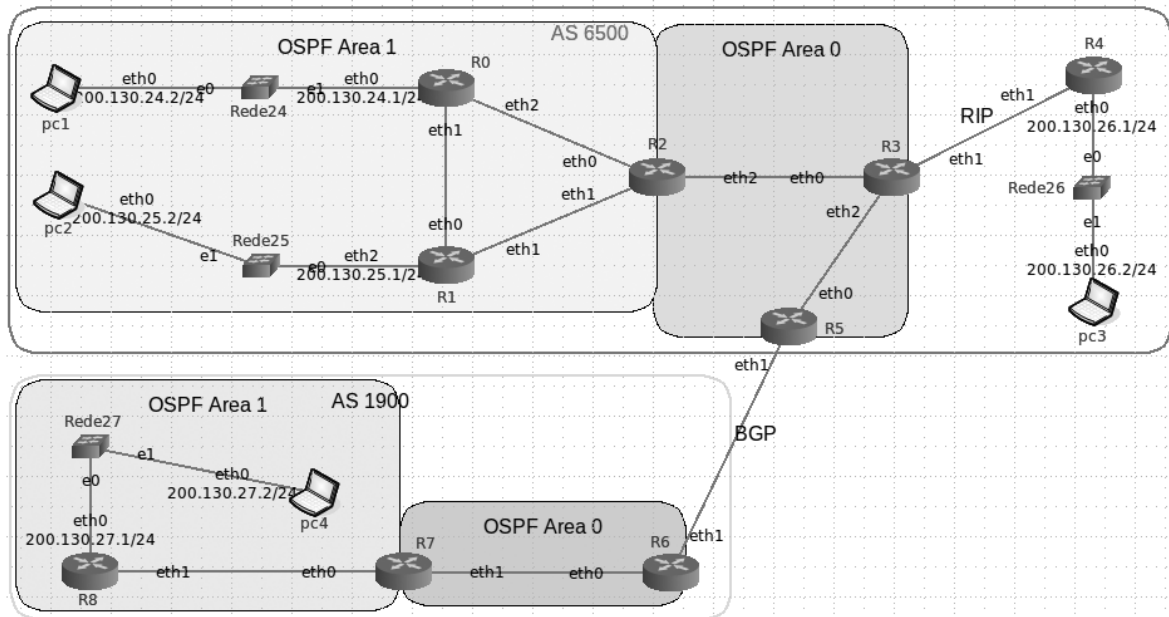
Siga os seguintes procedimentos para fazer o diagnóstico dos problemas (há mais de um erro de roteamento) e corrigi-los:

1. Anotar os passos do diagnóstico (ping, traceroute etc.) e os problemas encontrados.
2. Anotar as correções efetuadas. Mantenha o esquema de endereçamento. Salve a configuração corrigida com outro nome.
3. Testar a continuidade entre todos os PCs.
4. Verificar se todos os roteadores aprenderam as rotas de todas as redes.

Anote todos os procedimentos e aguarde a orientação do instrutor para fazer a apresentação do problema para os outros grupos.

Atividade 6.6 – Configuração de OSPF e BGP

Esta atividade consiste em configurar o protocolo OSPF em dois ASs (AS 6500 e AS 1900) e o protocolo BGP para interligá-los, além de configurar o protocolo RIP para uma rede isolada.



As áreas OSPF estão delimitadas na figura, bem como os enlaces BGP e RIP. As redes 200.130.24.0/24, 200.130.25.0/24, 200.130.26.0/24 e 200.130.27.0/24 já estão configuradas, mas os roteadores não estão. Os enlaces entre roteadores devem todos ser configurados com sub-redes da rede 192.168.10.0/24, que devem usar a máscara que permita a maior economia possível de endereços IP.

Figura 6.12
Rede6_Sessão6_
ADR8.

Conclusão

Nestas atividades práticas aprendemos a:

- Analisar problemas de configuração utilizando uma metodologia de solução de problemas;
- Configurar protocolos RIP, OSPF e BGP;
- Testar o funcionamento da rede.

Bibliografia

- ASSIS, ALEXANDRE FURTADO DE; ALVES JR., NILTON. Protocolos de Roteamento RIP e OSPF. Disponível em: <<http://mesonpi.cat.cbpf.br/naj/>>.
- COMER, DOUGLAS E. Interligação em rede com TCP/IP: princípios, protocolos e arquitetura. Rio de Janeiro: Campus; 1998.
- DE CASTRO, MARIA CRISTINA F. Planejamento de Redes Comutadas. Disponível em: <http://www.ee.pucrs.br/~decastro/pdf/Redes_Comutadas_Cap2_1.pdf>.
- HALABI, SAM. OSPF Design Guide. Cisco Systems, 1996. Disponível em: <<http://www.cisco.com/warp/public/104/1.html>>.
- LUCENA, SIDNEY CUNHA DE. Roteamento na RNP – uma visão geral. Disponível em: <http://www.rnp.br/_arquivo/sci/2002/roteamento.pdf>.
- MOURA, ALEX SOARES DE. O protocolo BGP4. Boletim trimestral sobre tecnologia de redes. RNP. 1999.
- _____. Dicas na Configuração do Protocolo BGP-4 – Parte 1. Boletim trimestral sobre tecnologia de redes, volume 5, nº 1. RNP.
- _____. Dicas na Configuração do Protocolo BGP-4 (final). Boletim trimestral sobre tecnologia de redes, volume 5, nº 5. RNP.
- _____. O Protocolo BGP4 – Parte 1. Boletim trimestral sobre tecnologia de redes, volume 3, nº 2. RNP.
- _____. O Protocolo BGP4 – Parte 2. Boletim trimestral sobre tecnologia de redes, volume 3, nº 3. RNP.
- _____. O Protocolo BGP4 – Parte 3 (final). Boletim trimestral sobre tecnologia de redes, volume 3, nº 4. RNP.
- OYAMA, CYBELLE SUEMI ODA; LUCENA, SIDNEY CUNHA DE. Considerações Acerca do Estabelecimento de QoS na RNP2. Boletim trimestral sobre tecnologia de redes, Volume 6, nº 3. RNP.
- RAMANATH, AVINASH. BGP OSPF Interaction Report. Disponível em: <<http://www.quagga.net/docs/>>.
- STEVENS, W. RICHARD. TCP/IP Illustrated, Volume 1: The Protocols. Addison-Wesley, 1994. ISBN 0-201-63346-9.

- ▣ Tutoriais de TCP/IP. Disponível em: <<http://www.juliobattisti.com.br/artigos>>.
- ▣ Protocolos de Roteamento RIP & OSPF. Disponível em:
<http://www.gta.ufrj.br/grad/98_2/aline/indice.html>.
- ▣ ZEBRA BGP commands. Disponível em:
<http://personals.ac.upc.edu/joseb/BGP_commands_zebra.pdf>.
- ▣ FRAIZER, John. Example zebra config. Disponível em:
<<http://tania.be.linux.org/zebra/msg00338.html>>.
- ▣ Network Protocols Configuration Guide, Part 1. Disponível em:
<http://www.cisco.com/univercd/cc/td/doc/product/software/ios113ed/113ed_cr/np1_c/1cbgp.pdf>.
- ▣ Using Regular Expressions in BGP. Disponível em:
<<http://www.cisco.com/warp/public/459/26.pdf>>.
- ▣ Route-Maps for IP Routing Protocol Redistribution Configuration. Disponível em: <http://www.cisco.com/warp/public/459/route-map_bestp.pdf>.
- ▣ Module 13 – Multihoming to Different ISPs. Disponível em:
<<http://www.pacnog.org/pacnog1/day5/module13.pdf>>.
- ▣ Module 12 – Multihoming to the Same ISP. Disponível em:
<<http://www.pacnog.org/pacnog1/day5/module12.pdf>>.
- ▣ Assigned Numbers. Disponível em:
<<http://www.ietf.org/rfc/rfc1700.txt>>. RFC 1700.
- ▣ Keywords for use in RFCs to Indicate Requirement Levels. Disponível em:
<<http://www.ietf.org/rfc/rfc2119.txt>>. RFC 2119.
- ▣ OSPF Version 2 . Disponível em: <<http://www.ietf.org/rfc/rfc2328.txt>>. RFC 2328.
- ▣ RIP Version 2 . Disponível em: <<http://www.ietf.org/rfc/rfc2453.txt>>. RFC 2453.
- ▣ Internet Official Protocol Standards. Disponível em: <<http://www.ietf.org/rfc/rfc3700.txt>>. RFC 3700.
- ▣ Protocolo RIP
<http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/rip.pdf>
- ▣ Pontos de troca de tráfego (NIC.br)
<http://www.ptt.br>
<http://www.equinix.com/pdf/whitepapers/PeeringWP.2.pdf>
<http://www.nanog.org/mtg-0405/pdf/norton.pdf>
<http://en.wikipedia.org/wiki/Peering>
<ftp://ftp.registro.br/pub/gter/gter20/06-migrando-as-intro.pdf>
<ftp://ftp.registro.br/pub/gter/gter18/08-pttmetro.pdf>
- ▣ Pontos de Troca de Tráfego na Internet
<http://penta2.ufrgs.br/comdex2002/ppt/PTTDemi/index.htm>
http://www.teleco.com.br/tutoriais/tutorialinter/pagina_4.asp
<http://www.rnp.br/ceo/peering.html>

- GTER29 - Estudo de Caso de Sistema Autônomo com Conexão a PTT Local, Remoto e Provedores de Trânsito:
<ftp://ftp.registro.br/pub/gter/gter29/10-PTTLocalRemotoTransito.pdf>
- Estudo de Caso sequestro tráfego Youtube
<http://www.ripe.net/internet-coordination/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study>
- OSPF Design Guide
<http://www.cisco.com/image/gif/paws/7039/1.pdf>
- Protocolo OSPF
<http://www.midiacom.uff.br/~debora/redes1/pdf/trab042/OSPF.pdf>
- Configuring OSPF
http://www.cisco.com/en/US/docs/ios/12_0/np1/configuration/guide/1cospf.pdf
- IBGP Scalability
<http://web.archive.org/web/20070323205145/http://www.riverstonenet.com/support/bgp/scalability/index.htm>
- BGP Fundamentals
<http://web.archive.org/web/20070317143705/http://www.riverstonenet.com/support/bgp/fundamentals/index.htm>
- Path Attributes
http://web.archive.org/web/20070320193903/http://www.riverstonenet.com/support/bgp/fundamentals/attributes.htm#_Path_Attributes



O curso prepara o aluno para projetar esquemas de roteamento para redes de diversos tamanhos, interligadas ou não a redes sob outra administração. O programa abrange roteamento estático e dinâmico, protocolos de roteamento interno RIP e OSPF e o protocolo de roteamento externo BGP.

Este livro inclui os roteiros das atividades práticas e o conteúdo dos slides apresentados em sala de aula, apoiando profissionais na disseminação deste conhecimento em suas organizações ou localidades de origem.

ISBN 978-85-63630-24-7



9 788563 630247