

FACULDADE DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA
LICENCIATURA EM ENGENHARIA INFORMÁTICA
REDES DE COMPUTADORES II

TEMA: Border Gateway Protocol

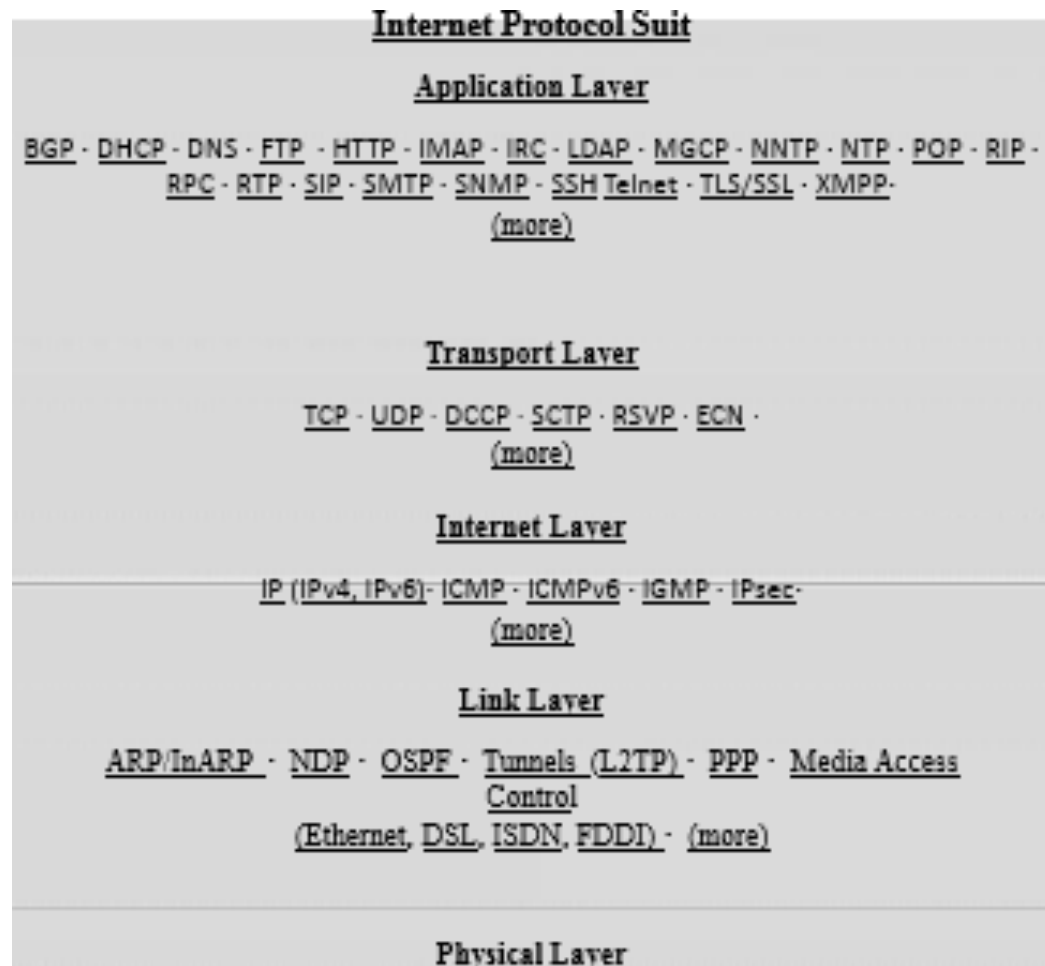
Grupo Docente:

- **Regente:** Eng^o. Felizardo Munguambe
- **Assistente:** Eng^o. Délcio Chadreca

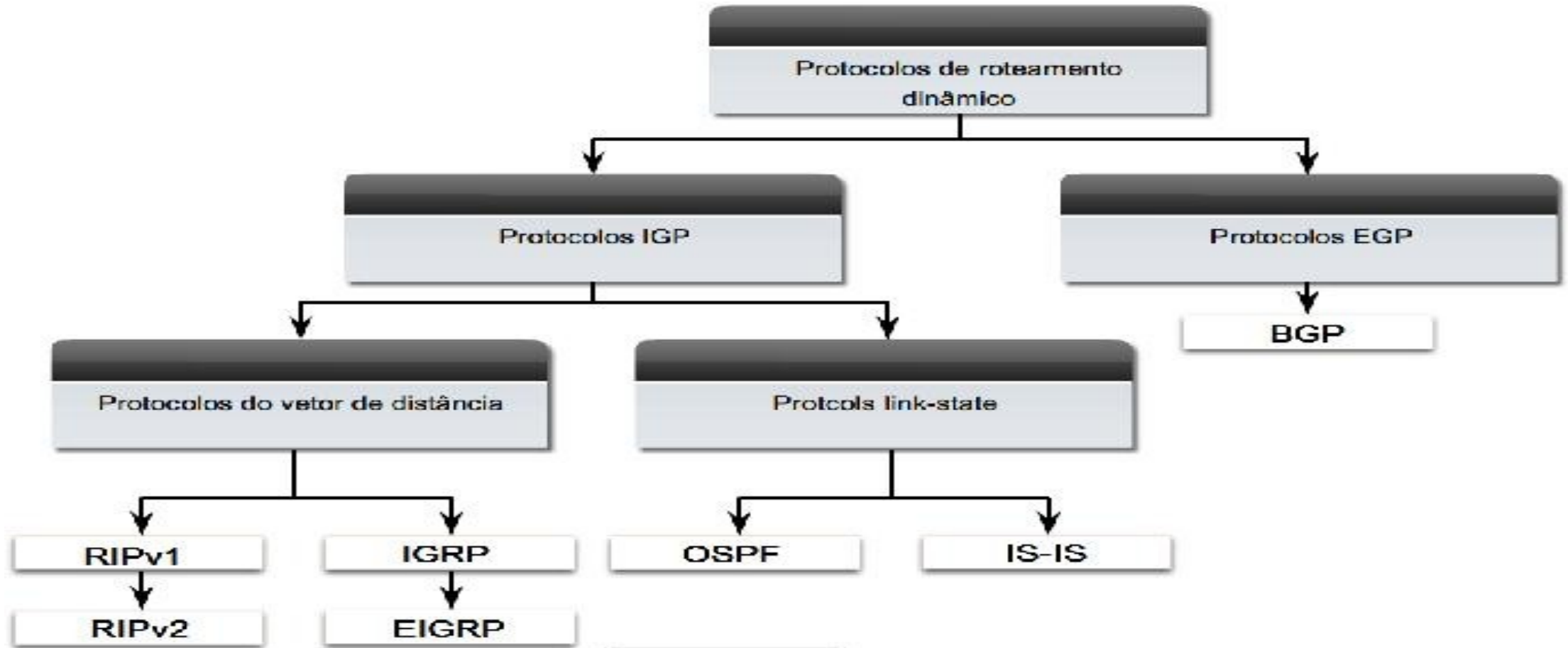
Tópicos da Aula

- 1. Introdução
- 2. Sistemas Autonomos (AS)
- 3. Protocolo BGP (*Border Gateway Protocol*) e Atributos
- 4. Formato das mensagens trocadas pelo protocolo BGP
- 5. Estados de Sessoes BGP

Introdução



Introdução



Introdução

- ***Interior Gateway Routing Protocols (IGP)***: Usados para roteamento dentro de um sistema autónomo e entre as próprias redes individuais.
 - RIP (*Routing Information Protocol*),
 - IGRP (*Interior Gateway Routing Protocol*),
 - EIGRP (*Enhanced IGRP*),
 - OSPF (*Open Shortest Path First*)
- ***Exterior Routing Protocols (EGP)***: Usados para roteamento entre sistemas autónomos
 - BGPv4 (*Border Gateway Protocol Versao 4*)

Introdução

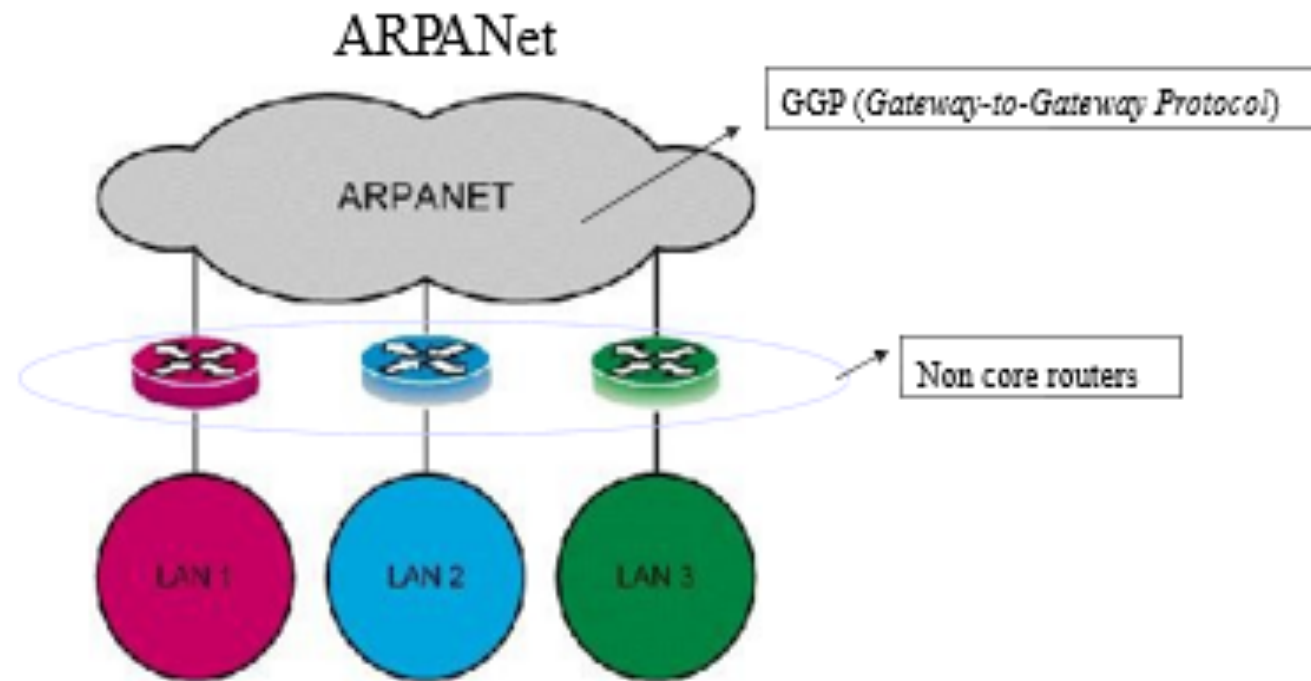
- Roteamento é o mecanismo através do qual duas máquinas em comunicação “acham” e usam um caminho óptimo (o melhor) através de uma rede. O processo envolve:
 - *Determinar que caminhos estão disponíveis;*
 - *Selecionar o “melhor” caminho para uma finalidade particular;*
 - *Usar o caminho para chegar aos outros sistemas;*
 - *Ajustar o formato dos dados (datagramas) às tecnologias de transporte disponíveis (MTU, MSS, etc.).*
- Na arquitetura TCP/IP, o roteamento é baseado no endereçamento IP, particularmente, na parte de identificação de rede de um endereço IP. Toda a tarefa é desenvolvida na camada Inter-rede da pilha de protocolos TCP/IP.
- A ARPANET (*Advanced Research Projects Agency Network*) foi a rede que iniciou o backbone da Internet.
- Nessa época cada participante administrava suas tabelas de roteamento e suas atualizações eram feitas manualmente. Com o rápido crescimento da rede, verificou-se que essa estratégia adotada tornava-se impraticável e Adotou-se um roteamento centralizado no núcleo (core).

Introdução

- Arquitetura proposta na época
 - Um conjunto reduzido e centralizado de roteadores no núcleo da rede (*Core routers*)
 - Mantinham rotas para todos os possíveis destinos na Internet
 - Administrados pelo INOC (*Internet network Operation Center*)
 - Desenvolvimento do protocolo GGP (*Gateway-to-Gateway Protocol*) para atualização automática das tabelas de rotas
 - Baseado no algoritmo vetor-distância (*Bellman Ford*)
 - Um conjunto maior de roteadores (*Non cores routers*) com rotas parciais
 - Administrados pelas instituições de pesquisa

Histórico

- Arquitetura proposta na época



Histórico

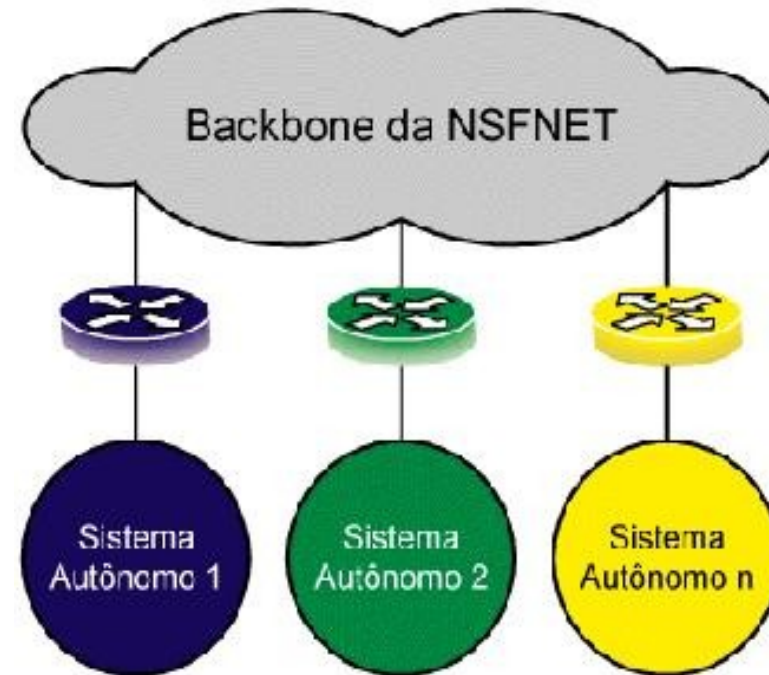
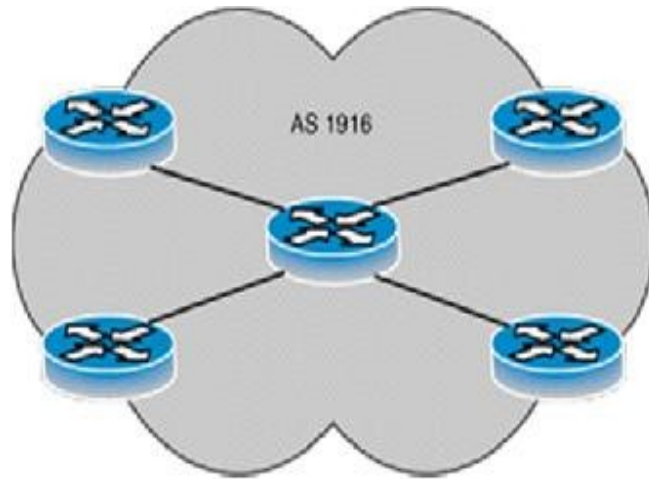
- Limitações da arquitetura proposta na época
 - *Backbone* complexo de cada site (instituição)
 - Impossibilidade de conectar diretamente as redes
 - Levava em conta apenas interligação com apenas uma Internet
 - Não contemplava questões administrativas
- Solução
 - Desenvolver um mecanismo para possibilitar a comunicação com o “mundo exterior”
 - Nasce e firma-se o conceito de AS
 - Desenvolvimento do EGP (Que também mostrou-se limitado)
 - Desenvolvimento do BGP em substituição ao EGP

Sistema Autônomo (*Autonomous System*) - AS

- Um grupo de roteadores
 - Administrados com uma política comum de roteamento,
 - Operam sob a mesma administração técnica,
 - Percebidos externamente como um único domínio de roteamento.
- **Inteiro de 16 bits (1-65535)**
- **64512-65535 AS - Privados**
- Segundo a 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 **outros ASs**.
 - Requisito básico: uma política de roteamento única
 - A política de roteamento define como são tomadas as decisões de roteamento na Internet.

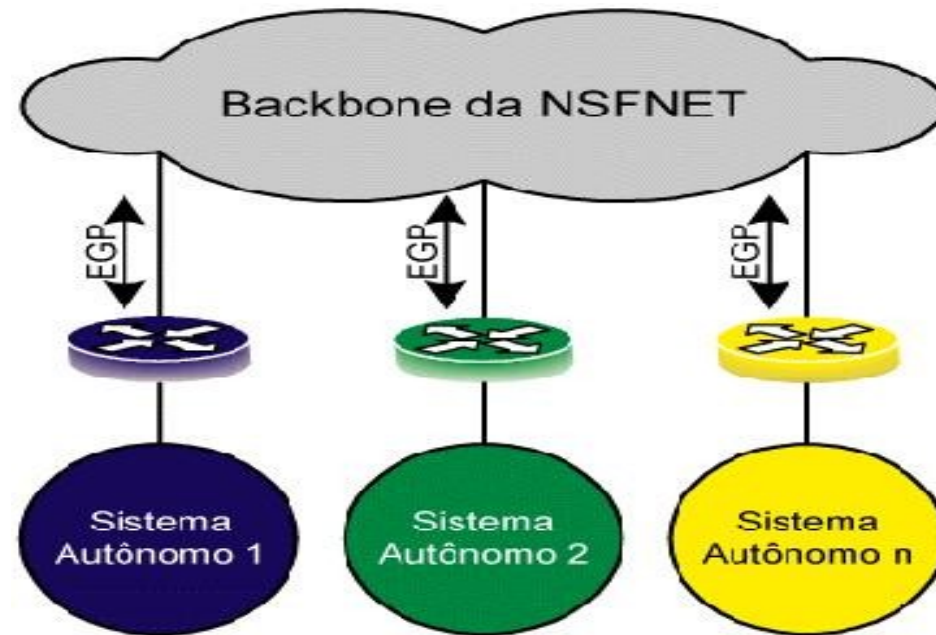
Sistema Autônomo (*Autonomous System*) - AS

- Conjunto de redes compartilhando a mesma política
- Utilizam um único protocolo de roteamento
- Estão sob a mesma administração técnica
 - Exemplos de ASs



Sistema Autônomo (*Autonomous System*) - AS

- Desafio:
 - como transformar uma arquitetura Internet para não depender de um sistema centralizado (*core routers*) - deixando uma topologia organizada hierarquicamente e iniciando outra, com diferente estrutura

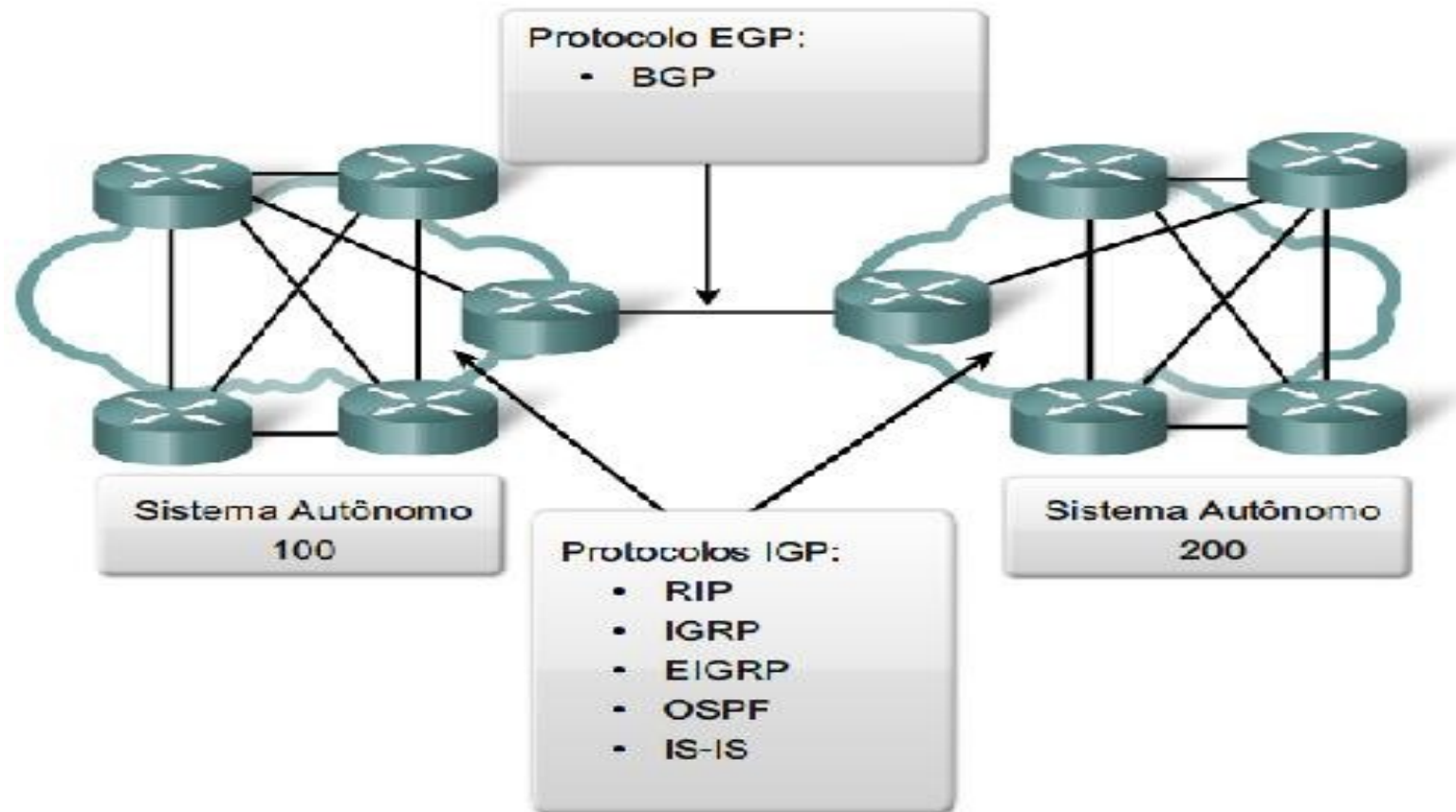


Sistema Autônomo (*Autonomous System*) - AS

- Solução:
 - Roteadores utilizados para trocar informações dentro de um AS
 - Roteadores interiores (*Internal Routers*)
 - Utilizam algum protocolo IGP (*Interior Gateway Protocol*)
 - » RIP, OSPF, IS-IS, IGRP, EIGRP
 - Roteadores utilizados para trocar informações entre **ASs**
 - Roteadores exteriores (*External Routers*)
 - Utilizam algum protocolo EGP (*External Gateway Protocol*)
 - » BGP-4 (RFC 4271)
 - Consideram blocos CIDR (Super Redes)

IGP Vs EGP

Protocolos de roteamento EGP em comparação com IGP



Border Gateway Protocol – BGP-4

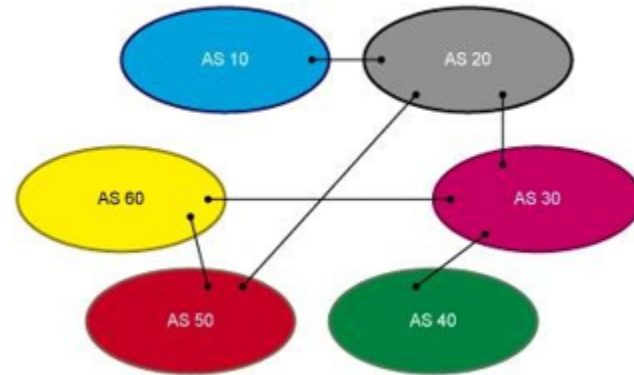
- É um protocolo de roteamento do tipo inter-domínio, criado transmite informações de prefixo;
- É um protocolo do tipo “ path vector”;
- Percebe a Internet como uma coleção de AS
- Projetado para evitar *loops* de roteamento em topologias
- O BGP é baseado em política (*Policy-based routing*)
 - Atende a um conjunto de regras não técnicas
 - As regras são definidas pelo administrador do AS

Border Gateway Protocol – BGP-4

- Sucessor do EGP (Substituiu o EGP)
- Roteamento entre ASs
- Suporte a Super-redes (CIDR – *Classless Interdomain Router*)
- Interage com IGP's: RIP, OSPF, etc.
- Usa TCP porta 179
- Estabelece sessões BGP
 - Estabelecimento de conexão TCP entre os roteadores
 - Envio de tabela de rotas completas apenas uma vez
 - Atualização parcial da tabela (Incremental)
 - Mensagens de *keepalive* para manter a sessão

Border Gateway Protocol – BGP-4

- 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.



A actual arquitetura da Internet, onde ASs comunicam-se via BGP-4

Atributos BGP

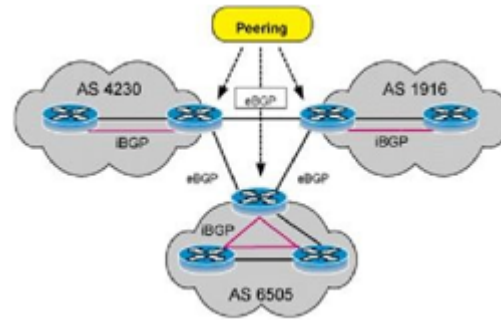
- AS-Path
 - Sequência de ASs que a rota atravessou; usado para detecção de loop; Aplicação de políticas.
- BGP Nexthop
 - é loopback do roteador iBGP
 - eBGP – endereço do neighbor externo
 - iBGP – Next-hop do eBGP
 - BGP speaker deve conhecer caminho para o next-hop
- Local-preference
 - Determina melhor caminho para tráfego saínte
 - Caminho com maior local-preference vence
 - Local-preference default 100 (JUNOS)

Atributos BGP

- MED (*Multi-Exit Discriminator*)
 - Inter-AS não transitivo; determina melhor caminho para tráfego entrante
 - Seu uso deve ser acordado entre ASs
- Origin
 - Informa a origem do prefixo
 - Influencia seleção do melhor caminho
 - Três tipos:
 - IGP – configurada de forma explícita no BGP (agregado, policy)
 - EGP – gerada pelo EGP
 - Incomplete – redistribuída por outro protocolo de roteamento

Border Gateway Protocol – BGP-4

- Neighbors, Peers, eBGP e iBGP
 - Sistemas (roteadores) que são "vizinhos BGP" (BGP *neighbors*) comunicam-se através de secções estabelecidas entre eles
 - Os roteadores de "borda" (*border routers*) de ASs vizinhos são considerados peers
- *Esses peers são as "fronteiras políticas" dos ASs, que trocam tráfego de acordo com as regras definidas pela Ass participantes*



Exemplo de Peers, Neighbors, eBGP e iBGP

Exterior BGP (eBGP)

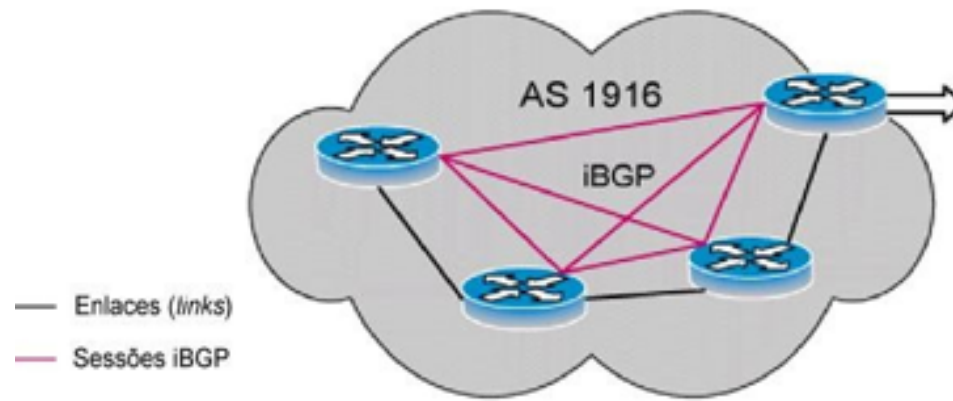
- Utilizado para passar rotas entre ASs
- Características
 - BGP nexthop é modificado
 - AS-Path é adicionado
 - Peer geralmente entre endereços de interfaces físicas
- AS-Path é utilizado como mecanismo de prevenção de loop de roteamento

Interior BGP (iBGP)

- Utilizado no interior de um AS
- Next-hop BGP não é modificado
- AS-Path não é adicionado
- É implementado tipicamente com peers totalmente interconectados (full mesh)
 - Análise de AS-PATH não é aplicável para prevenir loops internos
 - Roteador não pode repassar via iBGP rotas aprendidas através de outros peers iBGP

Border Gateway Protocol – BGP-4

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



Border Gateway Protocol – BGP-4 (MENSAGENS)

- Sessão BGP
 - Antes do estabelecimento de uma sessão BGP, os roteadores "vizinhos BGP" trocam mensagens entre si para entrar em acordo sobre quais serão os parâmetros
 - Ex.: tempo máximo de espera entre mensagens - *hold time*)
 - Quando a sessão é estabelecida entre os roteadores, são trocadas mensagens contendo todas as informações de roteamento, ou seja, todos os "melhores caminhos" (*best path*) previamente selecionados por cada um

Tipos de Mensagens

- Para comunicação entre roteadores BGP existem alguns tipos de mensagens onde cada um deles tem um papel importante na comunicação BGP, nomeadamente:
- **OPEN:** são utilizadas para o estabelecimento de uma conexão BGP;
- **NOTIFICATION:** reportam erros e serve para representar possíveis problemas nas conexões BGP.
- **UPDATE:** são utilizadas para os anúncios propriamente ditos, incluindo rotas que devem ser incluídas na tabela e também rotas que devem ser removidos da tabela BGP.
- **KEEPALIVE:** são utilizadas para manter a conexão entre roteadores BGP caso não existam atualizações através de mensagens **UPDATE**.

Estados de Sessão BGP

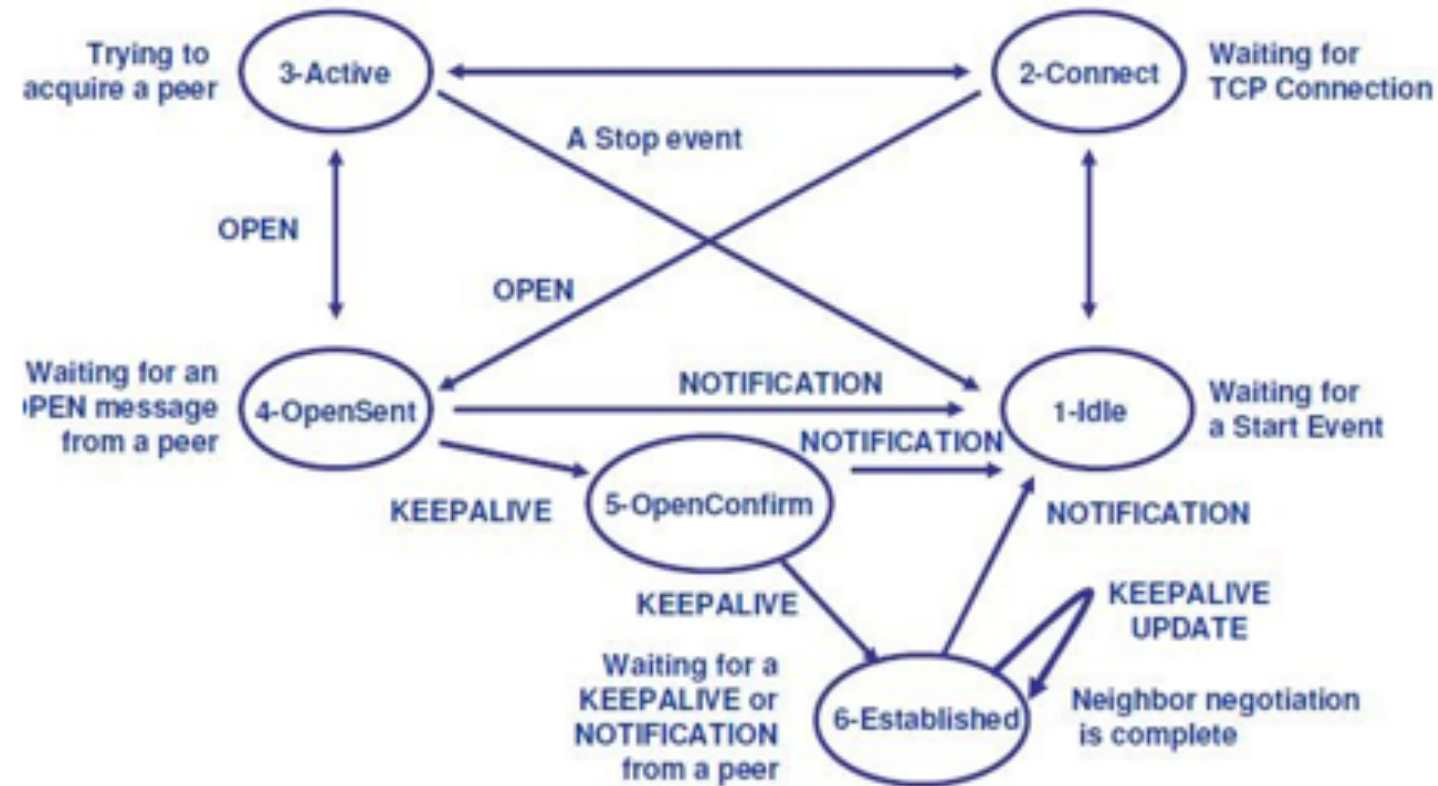


Figura: Máquina de estados finitos para sessões BGP

Estados de Sessão BGP

- **IDLE:** identifica o primeiro estágio de uma conexão BGP, onde o protocolo está aguardando por uma conexão de um *peer* remoto. O próximo estado é o de CONNECT e no caso da tentativa ser mal sucedida, volta ao estado IDLE.
- **CONNECT:** BGP aguarda pela conexão no nível de transporte, com destino na porta 179. Quando a conexão a este nível estiver estabelecida, ou seja, com o recebimento da mensagem de OPEN, passa-se ao estado de **OPENSENT**. Se a conexão nível de transporte não for bem sucedida, o estado vai para ACTIVE. No caso do tempo de espera ter sido ultrapassado, o estado volta para CONNECT. Em qualquer outro evento, é retorna-se para IDLE.
- **ACTIVE:** O BGP tenta estabelecer comunicação com um *peer* inicializando uma conexão no nível de transporte. Caso esta seja bem sucedida, passa-se ao estado OPENSENT. Se esta tentativa não for bem sucedida, pelo motivo de expiração do tempo, por exemplo, o estado passa para CONNECT. Em cada de interrupção pelo sistema ou pelo administrador, volta ao estado IDLE.

Estados de Sessão BGP

- **OPENSENT:** o BGP aguarda pela mensagem de OPEN e faz uma checagem de seu conteúdo. Caso seja encontrado algum erro como número de AS incoerente ao esperado ou a própria versão do BGP, envia-se uma mensagem tipo NOTIFICATION e volta ao estado de IDLE. Caso não ocorram erros na checagem, inicia-se o envio de mensagens KEEPALIVE. Em seguida, acerta-se o tempo de *Hold Time*, sendo optado o menor tempo entre os dois *peers*. Depois deste acerto, compara-se o número AS local e o número AS enviado pelo peer, com o intuito de detectar se trata-se de uma conexão iBGP (números de AS iguais) ou eBGP (números de AS diferentes). Em caso de desconexão a nível de protocolo de transporte, o estado passa para ACTIVE. Para as demais situações de erro, como expiração do Hold Time, envia-se uma mensagem de NOTIFICATION com o código de erro correspondente e retorna-se ao estado de IDLE. No caso de intervenção do administrador ou o próprio sistema, também retorna-se o estado IDLE.
- **OPENCONFIRM:** Neste estado o BGP aguarda pela mensagem de KEEPALIVE e quando esta for recebida, o estado segue para ESTABLISHED e a negociação do peer é finalmente completa. Com o recebimento da mensagem de KEEPALIVE, é acertado o valor negociado de Hold Time entre os peers. Se o sistema receber uma mensagem tipo NOTIFICATION, retorna-se ao estado de IDLE. O sistema também envia periodicamente, segundo o tempo negociado, mensagens de KEEPALIVE. No caso da ocorrência de eventos como desconexão ou intervenção do operador, retorna-se ao estado de IDLE também. Por fim, na ocorrência de eventos diferentes aos citados, envia-se uma mensagem NOTIFICATION, retornando ao estado de IDLE.

- **ESTABLISHED:** Neste estado, o BGP inicia a troca de mensagens de UPDATE ou KEEPALIVE, de acordo com o Hold Time negociado. Caso seja recebida alguma mensagem tipo NOTIFICATION, retorna-se ao estado IDLE. No recebimento de cada mensagem tipo UPDATE, aplica-se uma checagem por atributos incorretos ou em falta, atributos duplicados e caso algum erro seja detectado, envia-se uma mensagem de NOTIFICATION, retornando ao estado IDLE. Por fim, se o Hold Time expirar ou for detectada desconexão ou intervenção do administrador, também retorna-se ao estado de IDLE.
- A partir da máquina de estados apresentada anteriormente, é possível saber qual o *status* de uma sessão BGP entre dois roteadores, podendo também iniciar uma investigação sobre qual problema pode estar ocorrendo em alguma sessão. O objetivo esperado é que todas as sessões BGP de um roteador mantenham-se no estado ESTABLISHED, visto que somente neste estado ocorre a troca de anúncios com o roteador vizinho.

Border Gateway Protocol – BGP-4 (MENSAGENS)

- O BGP é bem económico nas mensagens de actualizações
 - Só envia quando ocorrem mudanças nas rotas
 - Ex.: uma rota se tornou inválida, informa novas rotas.
 - Caso não existam atualizações
 - Apenas mensagens de *KEEPALIVE* são trocadas
 - Certificar que a comunicação entre eles está "viva", ou seja, ainda está ativa

Border Gateway Protocol – BGP-4 (MENSAGENS)

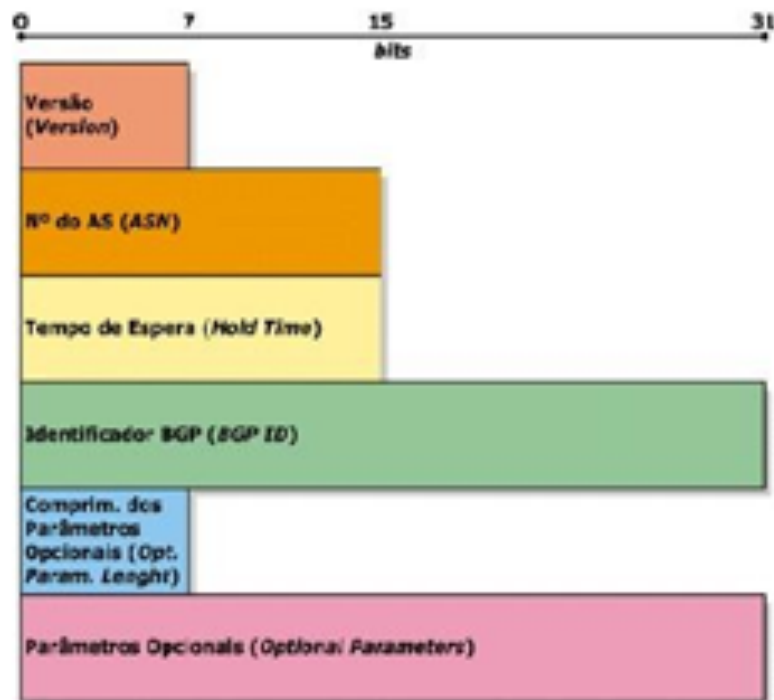
- Mensagens trocadas em sessões BGP
 - Cabeçalho 19 bytes
 - Campos do cabeçalho
 - Marker: verifica a autenticidade da mensagem recebida e a perda de sincronização entre os roteadores vizinhos BGP
 - Length: tamanho total (incluindo cabeçalho).
- Menor mensagem BGP enviada é de 19 bytes (16 + 2 + 1 bytes)
- Type: tipo da mensagem (OPEN, UPDATE, NOTIFICATION, KEEPALIVE)
 - Tamanho: 19 a 4096 bytes



Border Gateway Protocol – BGP-4 (MENSAGENS)

- Mensagem do tipo **OPEN**

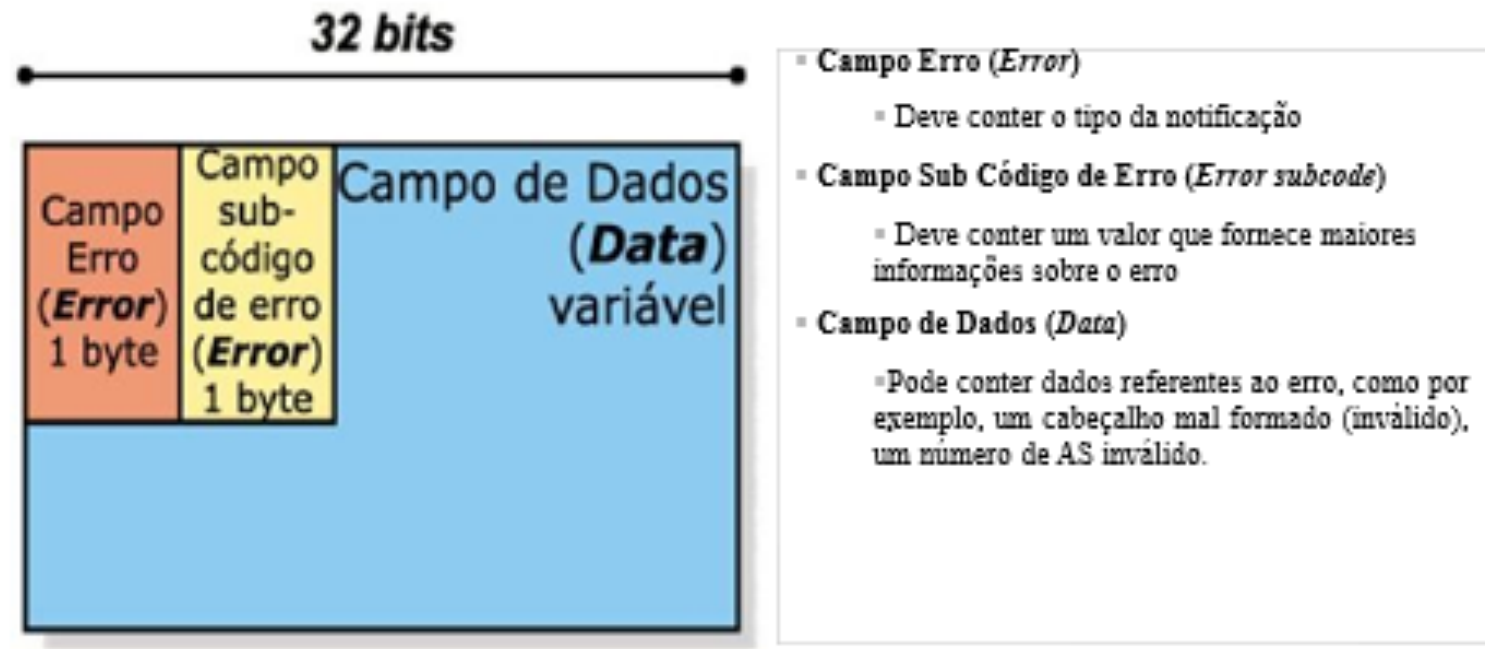
- A mensagem do tipo OPEN é enviada para se iniciar a abertura de uma sessão BGP entre neighbors ou peers BGP



- Versão (*Version*) – 3 ou 4
- Número do AS (*AS Number*)
 - Deve conter o número do AS a qual o roteador 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 KEEPALIVE
- Identificador BGP
 - Normalmente o Endereço IP do roteador
- Comprimento dos Parâmetros Opcionais (*Optional Parameters Length*)

Border Gateway Protocol – BGP-4 (MENSAGENS)

- Mensagem do tipo **NOTIFICATION**
 - Este tipo de mensagem é enviada no caso de detecção de erros durante ou após o estabelecimento de uma sessão BGP. Após o envio da mensagem “notification” as sessões BGP e a conexão TCP são encerradas.



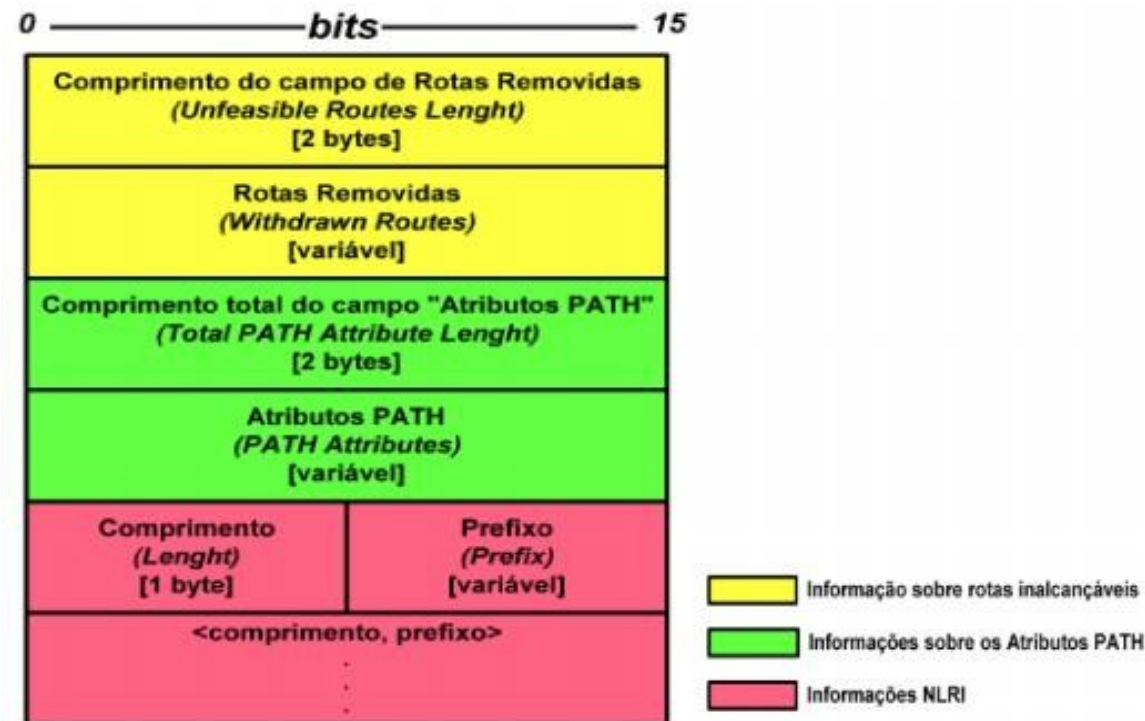
- Mensagem do tipo **NOTIFICATION**

- Tabela de códigos e sub-coódigos

Códigos de Erro	Sub códigos de Erro
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 <i>Well-Known</i> desconhecido 3 - Atributo <i>Well-Known</i> faltando 4 - Erro nas <i>flags</i> de atributos 5 - Erro no comprimento do atributo 6 - Atributo origem inválido 7 - <i>Loop</i> de roteamento em AS 8 - Atributo NEXT_HOP inválido 9 - Erro no atributo Opcional 10 - Campo de rede inválido 11 - <i>AS_path</i> mal formado

Border Gateway Protocol – BGP-4 (MENSAGENS)

- Mensagem to tipo **UPDATE**
 - As mensagens UPDATE, trocadas entre os peers ou neighbors BGP, são de extrema importância, pois são elas que levam as informações para a actualização da tabela de rotas mantida pelo BGP.



Border Gateway Protocol – BGP-4 (MENSAGENS)

- **Comprimento das Rotas Removidas ou Inalcançáveis (Unfeasible Routes Length):** Neste campo, é indicado o comprimento total, em bytes, do total 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 (CIDR)
- **Prefixo - (*Prefix*):** Contém prefixos de endereços IP seguidos de bits suficientes para fazer o final deste campo terminar "arredondado" em bytes completos. O valor dos bits complementares não têm importância
- **Comprimento (Length):** Deve indicar o comprimento total, em bits, do total de rotas removidas. Um comprimento igual a 0 (zero), indica que, nesta mensagem UPDATE, não há rotas a serem removidas.
- **Comprimento Total do Atributo PATH (*Total Path Attribute Length*):** Deve indicar o comprimento total, em bits, do campo Atributos PATH. O valor contido neste campo deve permitir a determinação do comprimento do campo NLRI.

Border Gateway Protocol – BGP-4 (MENSAGENS)

- Mensagem to tipo **KEEPALIVE**
 - São mensagens trocadas periodicamente com o propósito de verificar se a comunicação entre os vizinhos está activa
 - A mensagem do tipo KEEPALIVE é composta apenas do cabeçalho padrão das mensagens BGP, sem dados transmitidos após o cabeçalho (BGP header de 19 bytes)
 - O tempo máximo permitido para o recebimento de mensagens KEEPALIVE ou UPDATE é definido pelo *hold time*, das mensagens OPEN
 - Para manter aberta a sessão, a mensagem de KEEPALIVE deve ser enviada antes que o prazo definido no *hold time* expire

Bibliografia consultada

- ▶ Larry L. Peterson and Bruce S. Davie – Computer Network a system approach 5th Edition
- ▶ Tanenbaum A. S. and Wetherall D. J. - *Computer networks* 5th Edition.
- ▶ Mário Vestias Redes - Cisco para profissionais - 6ª Edição
- ▶ Adaptado do Professor Doutor Lourino Chemane

OBRIGADO !!!