

14° GTER



“Graceful Restart” - Mais um passo em direção a redes mais estáveis

caio@juniper.net

14° GTER

Sumário

- ◆ **Objetivos e Motivações**
- ◆ **Modificações propostas para o BGP (draft-ietf-idr-restart-03.txt)**
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ **Procedimentos para execução do "Graceful Restart"**
- ◆ **Exemplo prático**
- ◆ **Considerações**

Sumário

- ◆ **Objetivos e Motivações**
- ◆ Modificações propostas para o BGP (draft-ietf-idr-restart-03.txt)
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ Procedimentos para execução do "Graceful Restart"
- ◆ Exemplo prático
- ◆ Considerações

Efeitos negativos da reinicialização

- ◆ Normalmente quando o BGP reinicializa em um roteador :
 - ❖ Todos os peers detectam que a conexão caiu
 - ❖ A transição “down/up” resulta em um “routing-flap” com recálculo das rotas e possível “flap” da “forwarding-table”.
 - ❖ Esse comportamento pode se espalhar para múltiplos domínios de roteamento.
 - ❖ Pode criar “buracos” e “loops” temporários
 - ❖ Consomem recursos do plano de controle dos roteadores afetados pelo “flap”.

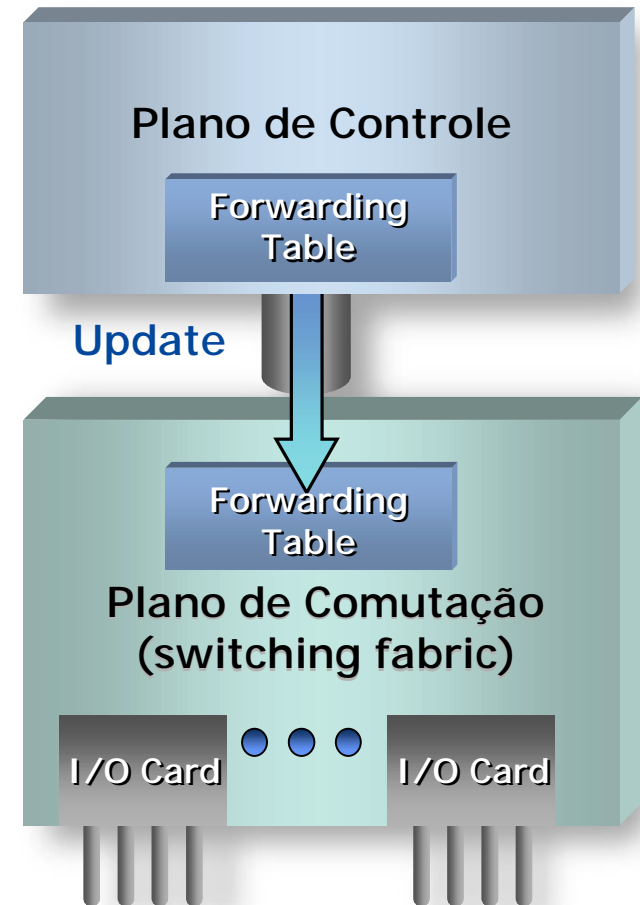
Plano de Controle e Plano de Comutação de um Roteador

◆ Plano de Controle

- ❖ Manipula pacotes de roteamento, OSPF, IS-IS, BGP,...
- ❖ Roda os protocolos de roteamento
- ❖ Gera a tabela de roteamento e tabela de "forwarding"

◆ Plano de Comutação

- ❖ Manipula todos os pacotes que atravessam o roteador
- ❖ Executa operações sobre pacotes ex: "route-lookup"

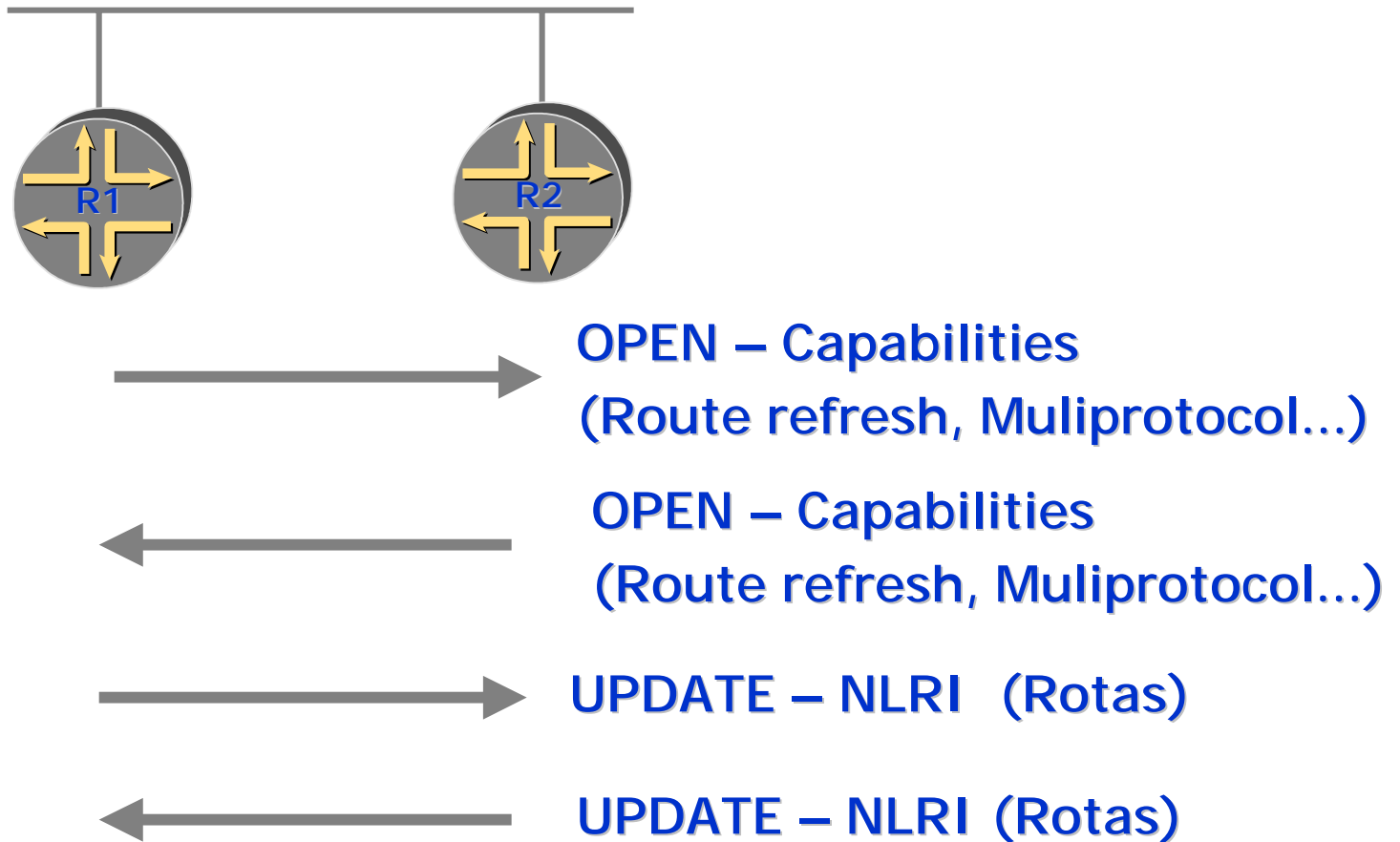


Após dispor de uma "forwarding table", o Plano de Comutação pode executar suas funções sem o Plano de Controle

Sumário

- ◆ **Objetivos e Motivações**
- ◆ **Modificações propostas para o BGP (draft-ietf-idr-restart-03.txt)**
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ **Procedimentos para execução do "Graceful Restart"**
- ◆ **Exemplo prático**
- ◆ **Considerações**

Abertura de Sessão BGP Convencional



Não existe sinalização explícita de "fim-de-update"

End-of-RIB

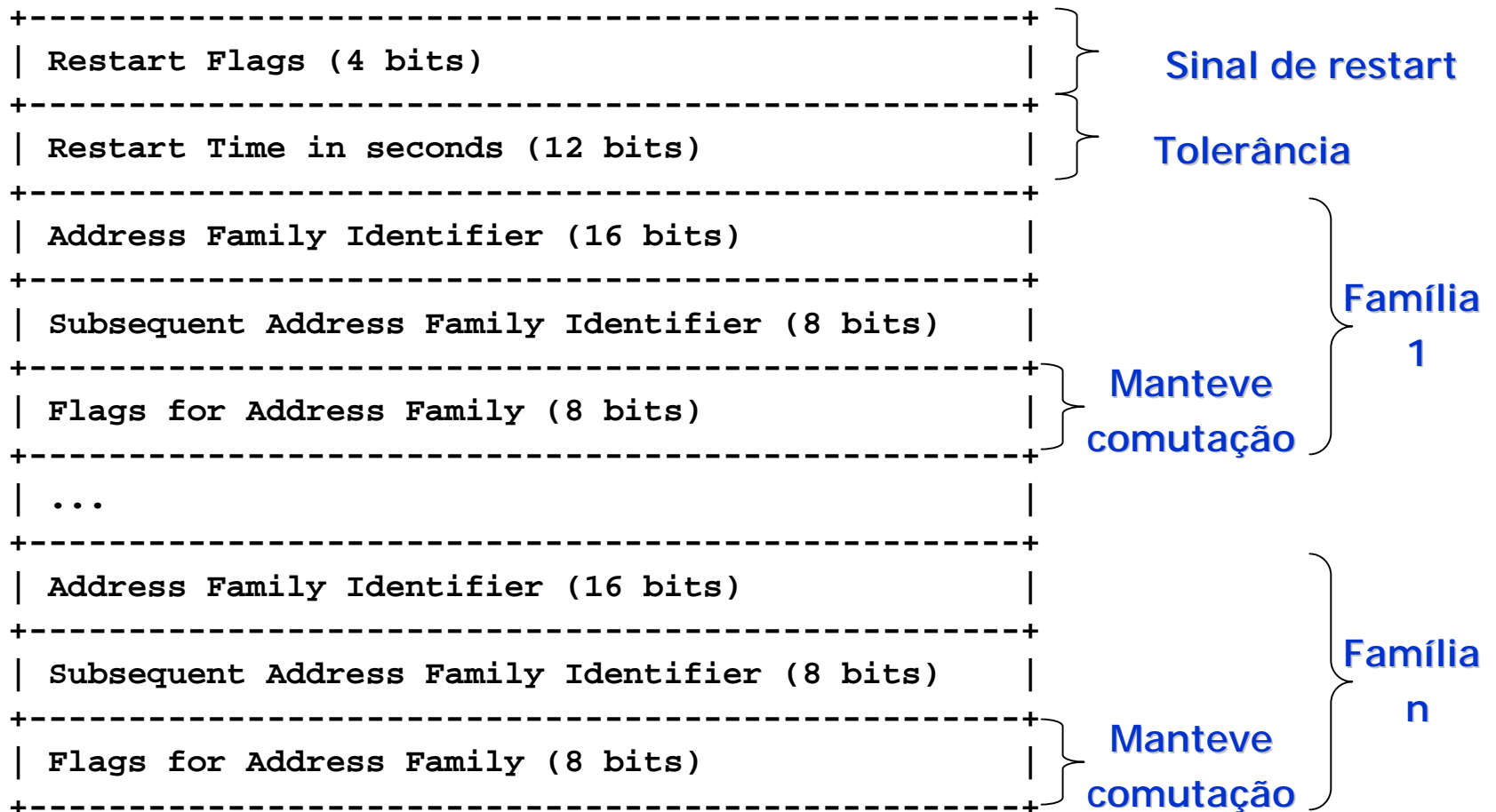
- ◆ Uma mensagem UPDATE com “withdrawn NLRI” vazio é especificada como “End-Of-RIB Marker”. Indica ao peer que o “update” de roteamento foi finalizado.
- ◆ Embora o “End-of-RIB Marker” seja especificado para o graceful restart, é recomendado o seu uso em todas as conexões pois pode ser útil na convergência do protocolo de roteamento.

Graceful Restart Capability

- ◆ A “Graceful Restart Capability” é uma capacidade do BGP [BGP-CAP] que pode ser usada por um “BGP speaker” para indicar a sua habilidade de preservar o seu estado de “forwarding” (comutação) durante o restart do BGP. Também pode ser utilizado para sinalizar aos seus peers a intenção de gerar a “End-Of-RIB marker” ao final das atualizações iniciais de roteamento.

Graceful Restart Capability

- ◆ A capacidade (capability) é definida abaixo:
- ◆ Capability code: 64
- ◆ Capability length: variable



Graceful Restart Capability - Campos

- ◆ **Restart Flags:** O bit mais significativo é definido como Restart State bit. Quando "1" indica que o "BGP speaker" reinicializou.
- ◆ **Restart Time:** Tempo estimado (s) para reestabelecimento da sessão BGP após o "restart". É usado como "timeout" para acelerar a convergência caso o roteador não retorne após o "restart".

Graceful Restart Capability - Campos

- ◆ **Address Family Identifier (AFI):** Identifica o protocolo suportado pela facilidade de graceful restart
- ◆ **Subsequent Address Family Identifier (Sub-AFI):** Informação adicional sobre NLRI.
- ◆ **Flags for Address Family:** Contém flags para o protocolo em questão. O bit mais significativo é definido como "Forwarding State bit" que pode ser usado para indicar se o estado de comutação para a família <AFI, Sub-AFI> foi preservado no último restart. Quando "1" indica que a comutação foi preservada.

Sumário

- ◆ Objetivos e Motivações
- ◆ Modificações propostas para o BGP (draft-ietf-idr-restart-02.txt)
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ **Procedimentos para execução do "Graceful Restart"**
- ◆ Exemplo prático
- ◆ Considerações

Abertura de Sessão BGP



OPEN – Graceful Restart Capability



OPEN – Graceful Restart Capability

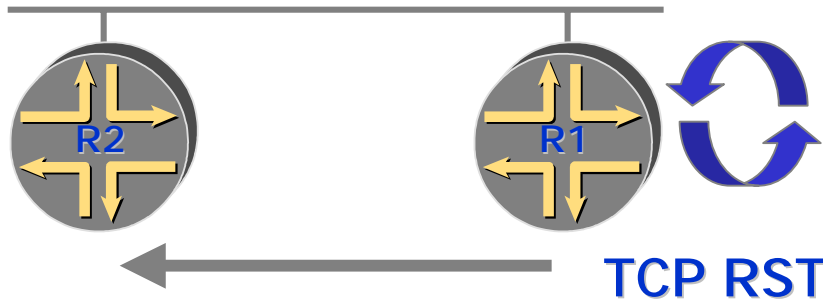


UPDATE com EOR



UPDATE com EOR

Procedimentos



- Mantém o "forwarding state" das rotas BGP na Loc-RIB.
- Marca rotas como "stale"
- Mantém comutação normalmente

Marca rotas como "stale" e mantém tabela de comutação
Verifica restart time



OPEN – GRCapability (Restart State = 1 e Forwarding State = 1)

OPEN – GRCapability (FS = 1)

Aguarda RIB completa de todos os peers



UPDATE com EOR

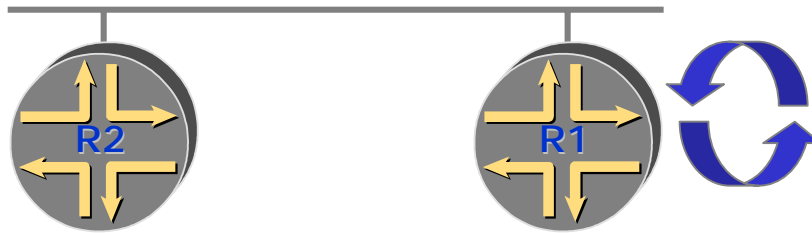
Seleciona rotas, remove "stale" e atualiza tabela

Remove "stale" e atualiza tabela

UPDATE com EOR

Procedimentos

R2 não percebe inicialmente o restart



- Mantém o "forwarding state" das rotas BGP na Loc-RIB.
- Marca rotas como "stale"
- Mantém comutação normalmente

OPEN ??? - Assume restart, marca rotas como "stale" e mantém tabela de comutação

OPEN – GRCapability (Restart State = 1 e Forwarding State = 1)

OPEN – GRCapability (FS = 1)

Aguarda RIB completa de todos os peers



UPDATE com EOR

Seleciona rotas, remove "stale" e atualiza tabela

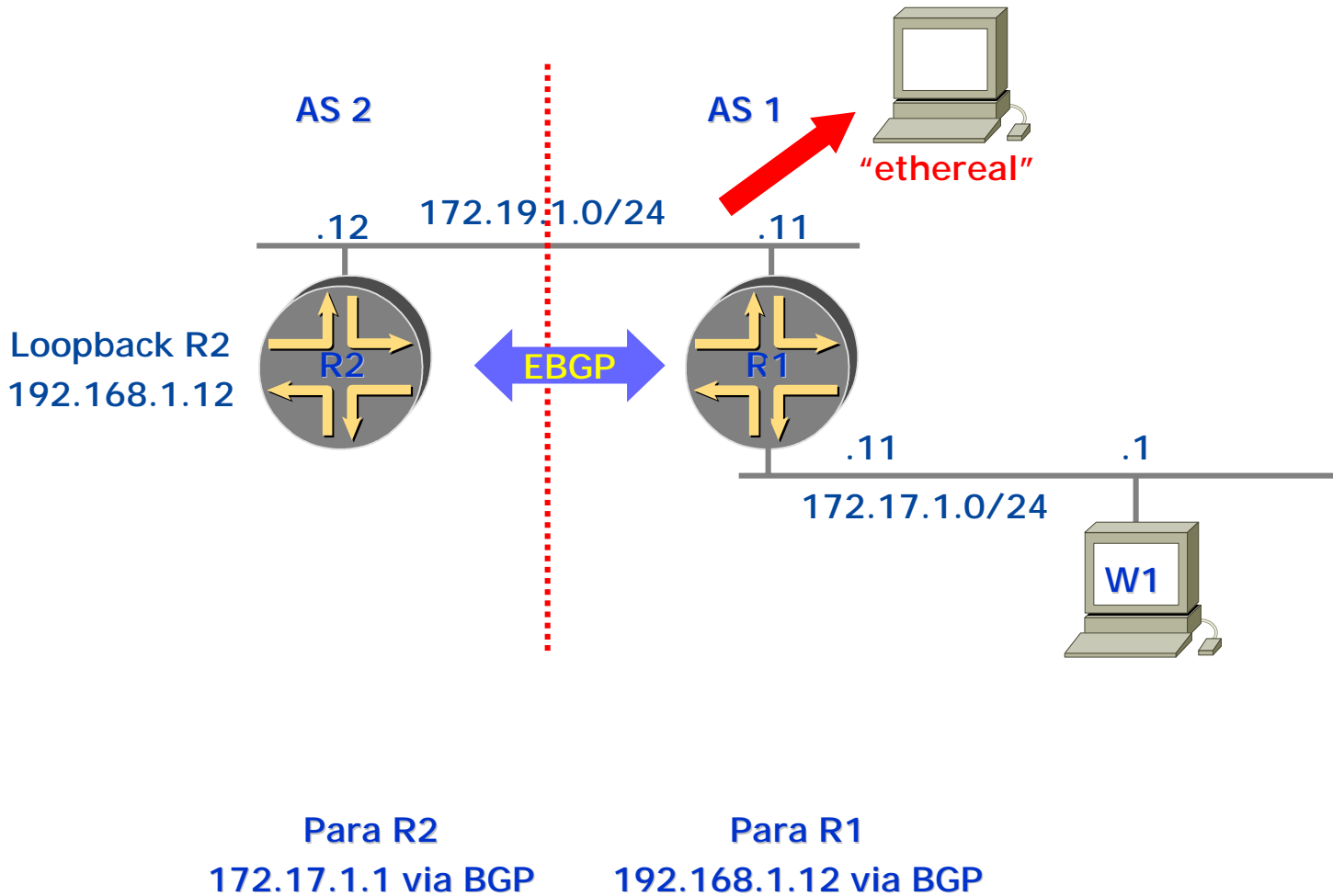
remove "stale" e atualiza tabela

UPDATE com EOR

Sumário

- ◆ Objetivos e Motivações
- ◆ Modificações propostas para o BGP (draft-ietf-idr-restart-02.txt)
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ Procedimentos para execução do "Graceful Restart"
- ◆ **Exemplo prático**
- ◆ Considerações

Demonstração - Topologia



Início - OPEN

The screenshot shows the Wireshark interface for a capture named 'inicio-com-gr - Ethereal'. The packet list pane displays the following BGP messages:

No.	Time	Source	Destination	Protocol	Info
15	14.592255	172.19.1.12	172.19.1.11	BGP	OPEN Message
16	14.592989	172.19.1.11	172.19.1.12	BGP	OPEN Message
17	14.593082	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
18	14.593152	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
19	14.593476	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
20	14.593548	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
21	14.593721	172.19.1.12	172.19.1.11	BGP	UPDATE Message
22	14.593778	172.19.1.11	172.19.1.12	BGP	UPDATE Message, UPDATE Me
23	14.593829	172.19.1.12	172.19.1.11	BGP	UPDATE Message

The packet details pane shows the structure of the selected packet (No. 15):

- Unknown capability (10 bytes)
 - Parameter type: Capabilities (2)
 - Parameter length: 8 bytes
 - Capability code: Unknown (64)
 - Capability length: 6 bytes
 - Capability value: Unknown

The packet bytes pane shows the raw data for the selected packet:

```
0060 06 01 04 00 01 00 01 02 02 80 00 02 02 02 00 02 .....  
0070 08 40 06 00 78 00 01 01 80 .@.x...
```

Restart State = 0 (início normal)

Restart Time = 120 (segundos)

AFI = 01 (IPv4)

Sub-AFI = 01

Forwarding State = 1 (manteve forwarding anterior)

Início - UPDATE

The screenshot shows the Wireshark interface with a capture of BGP messages. The packet list pane shows a series of BGP messages, with the selected packet (No. 23) being a BGP UPDATE Message from 172.19.1.12 to 172.19.1.11. The packet details pane shows the structure of the UPDATE message, including an unfeasible routes length of 0 bytes. The packet bytes pane shows the raw data of the message, with the end-of-RIB marker (00 00) highlighted.

No.	Time	Source	Destination	Protocol	Info
15	14.592255	172.19.1.12	172.19.1.11	BGP	OPEN Message
16	14.592989	172.19.1.11	172.19.1.12	BGP	OPEN Message
17	14.593082	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
18	14.593152	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
19	14.593476	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
20	14.593548	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
21	14.593721	172.19.1.12	172.19.1.11	BGP	UPDATE Message
22	14.593778	172.19.1.11	172.19.1.12	BGP	UPDATE Message, UPDATE Me
23	14.593829	172.19.1.12	172.19.1.11	BGP	UPDATE Message

Border Gateway Protocol

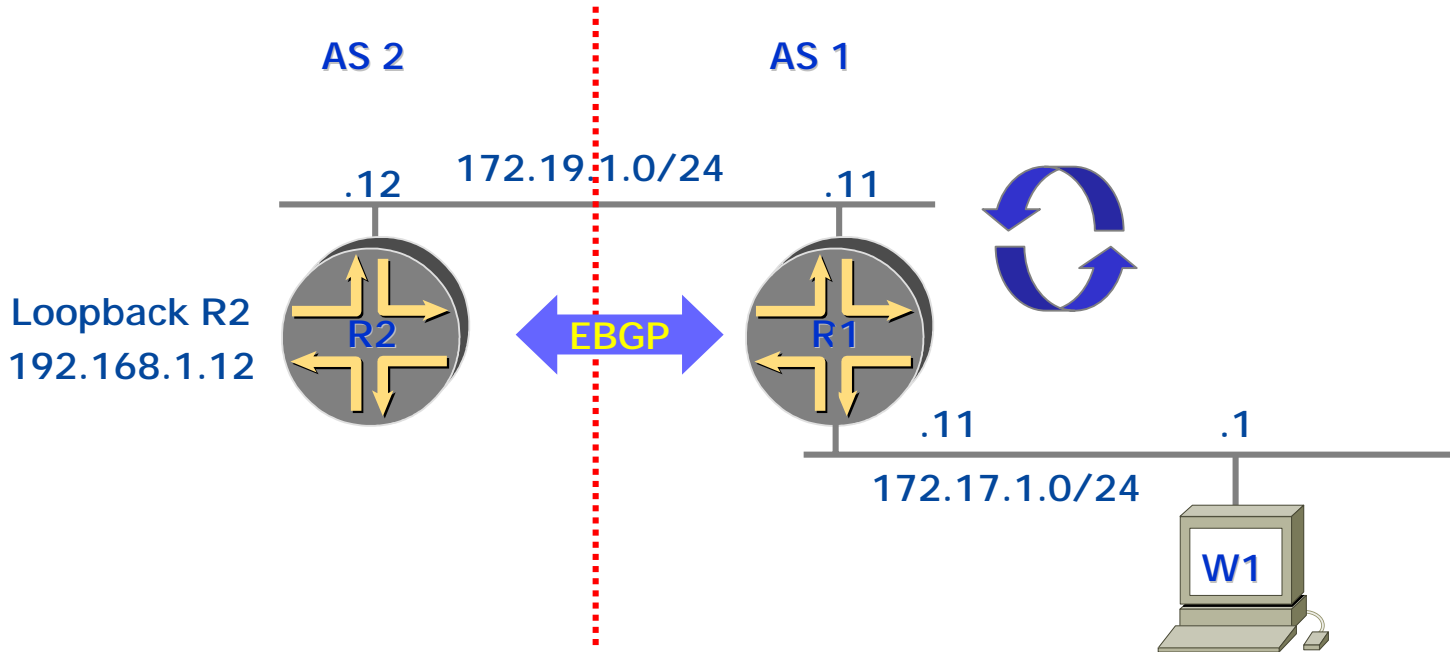
- UPDATE Message
 - Marker: 16 bytes
 - Length: 23 bytes
 - Type: UPDATE Message (2)
 - unfeasible routes length: 0 bytes
 - Total path attribute length: 0 bytes

0040 11 b3 ff ff ff ff ff ff ff ff ff ff ff ff ff .C.....
0050 ff ff 00 17 02 00 00 00 00

Filter: [] Reset

End-Of-RIB Marker

Restart em R1



Graceful Restart - OPEN

The screenshot shows a Wireshark capture of BGP messages. The packet list pane shows the following entries:

No.	Time	Source	Destination	Protocol	Info
29	12.241120	172.19.1.11	172.19.1.12	BGP	OPEN Message
30	12.241746	172.19.1.12	172.19.1.11	BGP	OPEN Message
31	12.241863	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
32	12.241931	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
33	12.242313	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
34	12.242388	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
35	12.340159	172.19.1.11	172.19.1.12	TCP	1026 > 179 [ACK] Seq=3546
36	12.340221	172.19.1.12	172.19.1.11	BGP	UPDATE Message, UPDATE Me
37	12.340743	172.19.1.11	172.19.1.12	BGP	UPDATE Message

The details pane for the selected packet (No. 29) shows the following structure:

- Capability length: 0 bytes
- Unknown capability (10 bytes)
 - Parameter type: Capabilities (2)
 - Parameter length: 8 bytes
 - Capability code: Unknown (64)
 - Capability length: 6 bytes
 - Capability value: Unknown

The hex dump pane shows the following data:

```
0060 06 01 04 00 01 00 01 02 02 80 00 02 02 02 00 02 .....  
0070 08 40 06 80 78 00 01 01 80 .@.x...
```

Restart State = 1 (graceful restart)

Restart Time = 120 (segundos)

AFI = 01 (IPv4)

Sub-AFI = 01

Forwarding State = 1 (manteve forwarding anterior)

Graceful Restart - UPDATE

The screenshot shows the Wireshark interface for a capture named 'ping-e-gr'. The main pane displays a list of captured packets. Packet 36 is highlighted in blue. Below the list, the packet details pane shows the structure of the BGP UPDATE message. The hex dump pane shows the raw bytes of the message.

No.	Time	Source	Destination	Protocol	Info
29	12.241120	172.19.1.11	172.19.1.12	BGP	OPEN Message
30	12.241746	172.19.1.12	172.19.1.11	BGP	OPEN Message
31	12.241863	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
32	12.241931	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
33	12.242313	172.19.1.11	172.19.1.12	BGP	KEEPALIVE Message
34	12.242388	172.19.1.12	172.19.1.11	BGP	KEEPALIVE Message
35	12.340159	172.19.1.11	172.19.1.12	TCP	1026 > 179 [ACK] Seq=3546
36	12.340221	172.19.1.12	172.19.1.11	BGP	UPDATE Message, UPDATE Me
37	12.340743	172.19.1.11	172.19.1.12	BGP	UPDATE Message

192.168.1.12/32

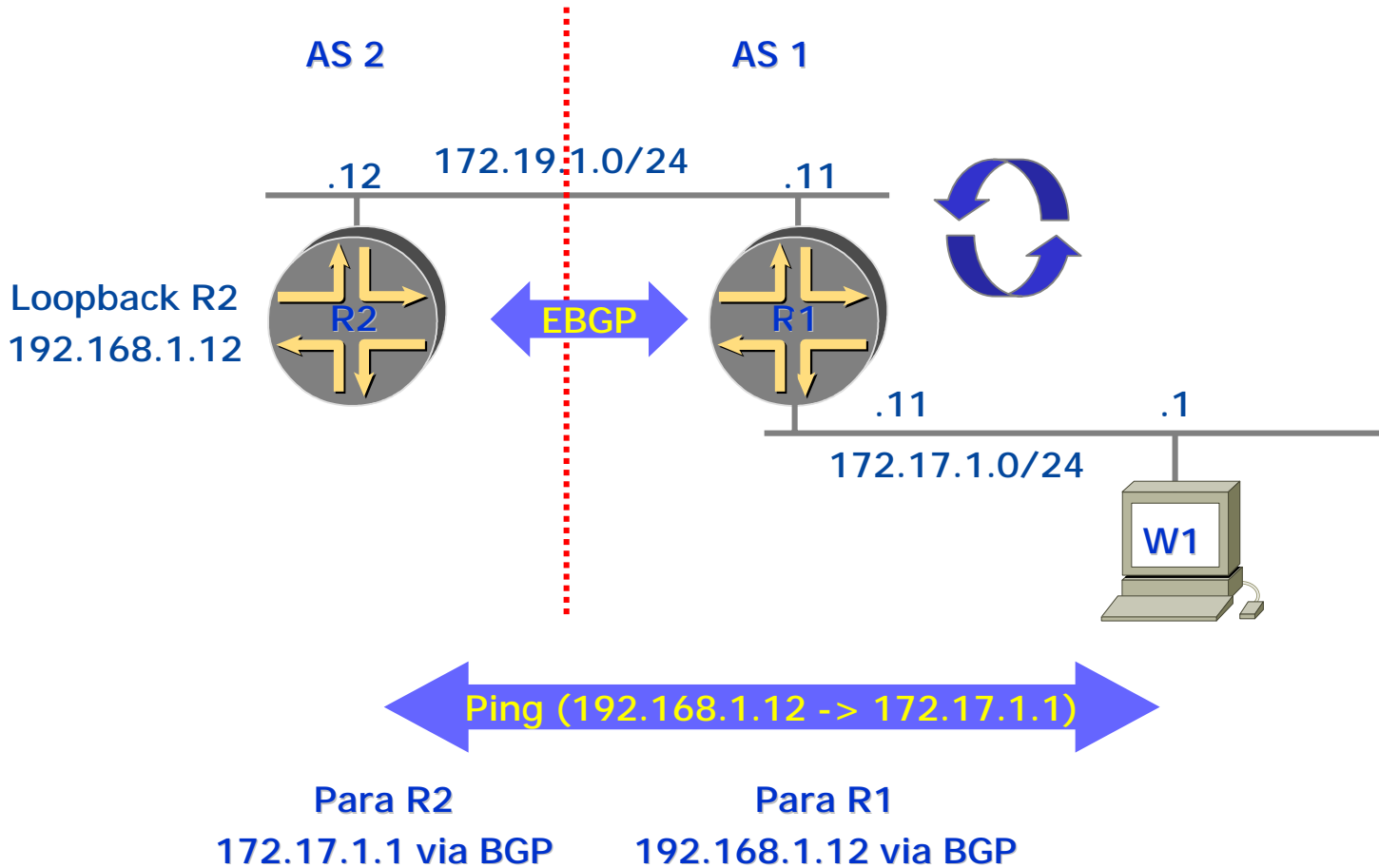
- UPDATE Message
 - Marker: 16 bytes
 - Length: 23 bytes
 - Type: UPDATE Message (2)
 - unfeasible routes length: 0 bytes
 - Total path attribute length: 0 bytes

00a0 CU a8 01 UC FF FF FF FF FF FF FF FF FF FF FF FF
00b0 ff ff ff ff 00 17 02 00 00 00 00

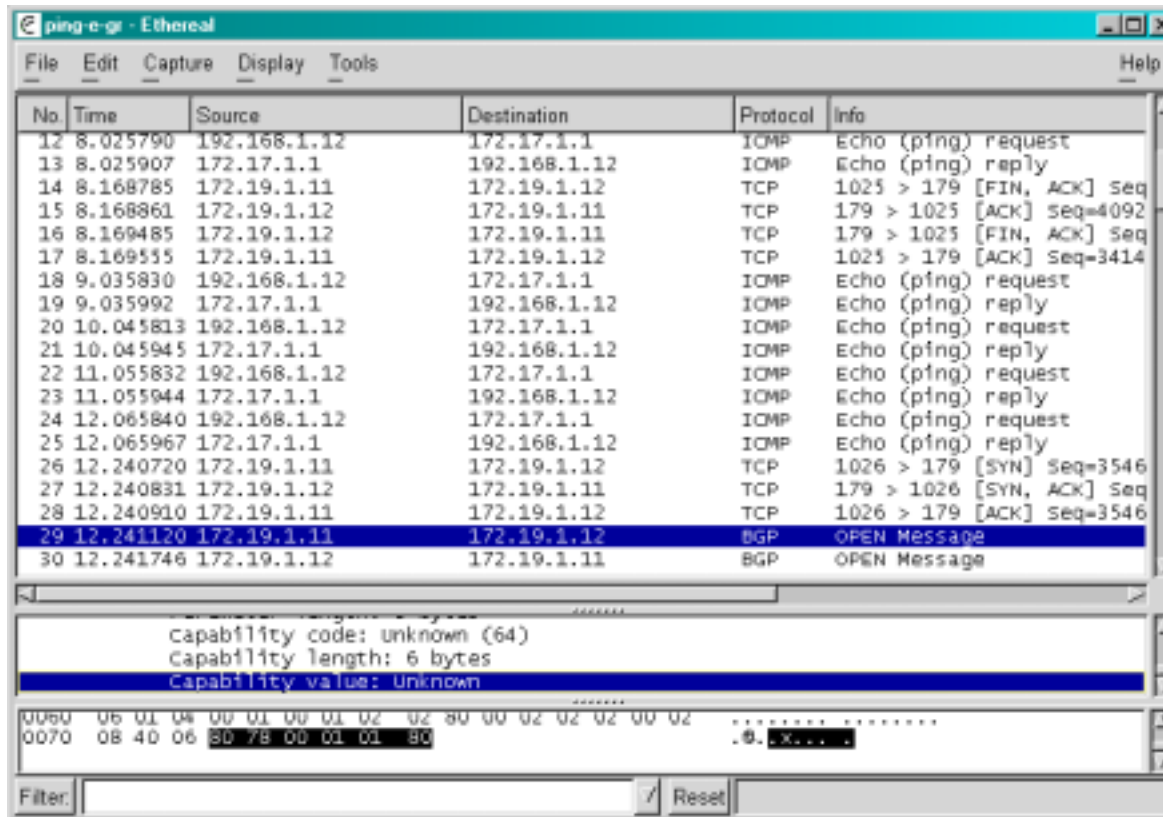
36 - R2 (receiving speaker) envia RIB e EOR

37 – Somente após receber EOR é que R1 (restarting speaker) anuncia

Trânsito de Pacotes



Graceful Restart – Comutação Ininterrupta



The screenshot shows a Wireshark capture of network traffic. The main pane displays a list of packets with columns for No., Time, Source, Destination, Protocol, and Info. The selected packet (No. 29) is a BGP OPEN Message from 172.19.1.11 to 172.19.1.12. Below the packet list, the packet details pane shows the BGP OPEN Message structure, including the Capability code (unknown (64)), Capability length (6 bytes), and Capability value (Unknown). The packet bytes pane shows the raw data of the BGP OPEN Message, including the BGP ID (807800010180).

No.	Time	Source	Destination	Protocol	Info
12	8.025790	192.168.1.12	172.17.1.1	ICMP	Echo (ping) request
13	8.025907	172.17.1.1	192.168.1.12	ICMP	Echo (ping) reply
14	8.168785	172.19.1.11	172.19.1.12	TCP	1025 > 179 [FIN, ACK] seq
15	8.168861	172.19.1.12	172.19.1.11	TCP	179 > 1025 [ACK] Seq=4092
16	8.169485	172.19.1.12	172.19.1.11	TCP	179 > 1025 [FIN, ACK] seq
17	8.169555	172.19.1.11	172.19.1.12	TCP	1025 > 179 [ACK] Seq=3414
18	9.035830	192.168.1.12	172.17.1.1	ICMP	Echo (ping) request
19	9.035992	172.17.1.1	192.168.1.12	ICMP	Echo (ping) reply
20	10.045813	192.168.1.12	172.17.1.1	ICMP	Echo (ping) request
21	10.045945	172.17.1.1	192.168.1.12	ICMP	Echo (ping) reply
22	11.055832	192.168.1.12	172.17.1.1	ICMP	Echo (ping) request
23	11.055944	172.17.1.1	192.168.1.12	ICMP	Echo (ping) reply
24	12.065840	192.168.1.12	172.17.1.1	ICMP	Echo (ping) request
25	12.065967	172.17.1.1	192.168.1.12	ICMP	Echo (ping) reply
26	12.240720	172.19.1.11	172.19.1.12	TCP	1026 > 179 [SYN] Seq=3546
27	12.240831	172.19.1.12	172.19.1.11	TCP	179 > 1026 [SYN, ACK] seq
28	12.240910	172.19.1.11	172.19.1.12	TCP	1026 > 179 [ACK] Seq=3546
29	12.241120	172.19.1.11	172.19.1.12	BGP	OPEN Message
30	12.241746	172.19.1.12	172.19.1.11	BGP	OPEN Message

- 12-13 Ping entre R2 e W1
- ... user@R1> restart routing
- 14-17 Término de conexão TCP
- 18-25 Comutação não interrompida
- 26-28 Reinício de conexão TCP/BGP
- 29 OPEN com restart flag = 1 e forwarding state = 1

Sumário

- ◆ Objetivos e Motivações
- ◆ Modificações propostas para o BGP (draft-ietf-idr-restart-02.txt)
 - ❖ End-of-RIB marker
 - ❖ "Graceful Restart Capability"
- ◆ Procedimentos para execução do "Graceful Restart"
- ◆ Exemplo prático
- ◆ **Considerações**

Considerações

- ◆ Existe a probabilidade de geração de loops ou “buracos” caso ocorra modificações nas informações de roteamento antes do roteador finalizar o “restart”
- ◆ Dependendo da topologia, caso alguns roteadores rodando IBGP não possuam Graceful Restart, há uma exposição maior à geração de loops e “buracos”.
- ◆ Restart Time ainda não foi bem definido. É necessário maior experiência.

Considerações, cont

- ◆ Há poucos benefícios em ativar o Graceful Restart para o BGP em ambientes onde existe uma interação forte entre BGP e IGP e o IGP não dispõe dessa facilidade.

Referências

- ◆ [draft-ietf-idr-restart-03.txt](#) Title: Graceful Restart Mechanism for BGP
- ◆ [draft-shand-isis-restart-01.txt](#) Title: Restart signaling for ISIS
- ◆ [draft-rekhter-bgp-mpls-restart-00.txt](#) Title: Graceful Restart Mechanism for BGP with MPLS
- ◆ [draft-ietf-ospf-hitless-restart-02.txt](#) Title: Hitless OSPF Restart

14° GTER



Obrigado

caio@juniper.net