BGP Add-Paths
Caminhos Adicionais

GTER 35

23 Maio 2013

Eduardo Ascenço Reis
<eascenco@nic.br>
<eduardo@intron.com.br>

Equipe PTT.br
<eng@ptt.br>

Resumo

Na atual versão do protocolo de roteamento utilizado entre os Sistemas Autônomos, o BGP-4, cada roteador calcula o melhor caminho para um determinado prefixo IP (NLRI), dentre os disponíveis, e apenas este é repassado aos demais roteadores com os quais possui vizinhança (sessões estabelecidas).

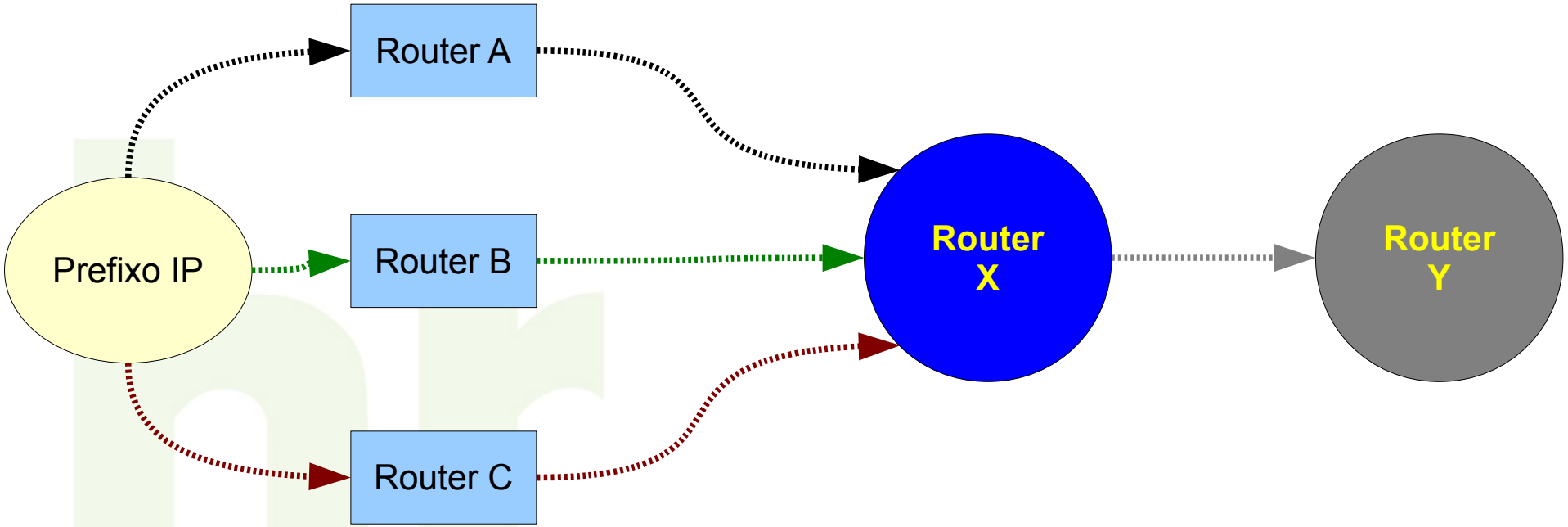
Neste trabalho serão apresentadas algumas propostas que estão em andamento no IETF para permitir que caminhos adicionais sejam anunciados via BGP, e algumas possíveis decorrências destas modificações.

Agradecimentos

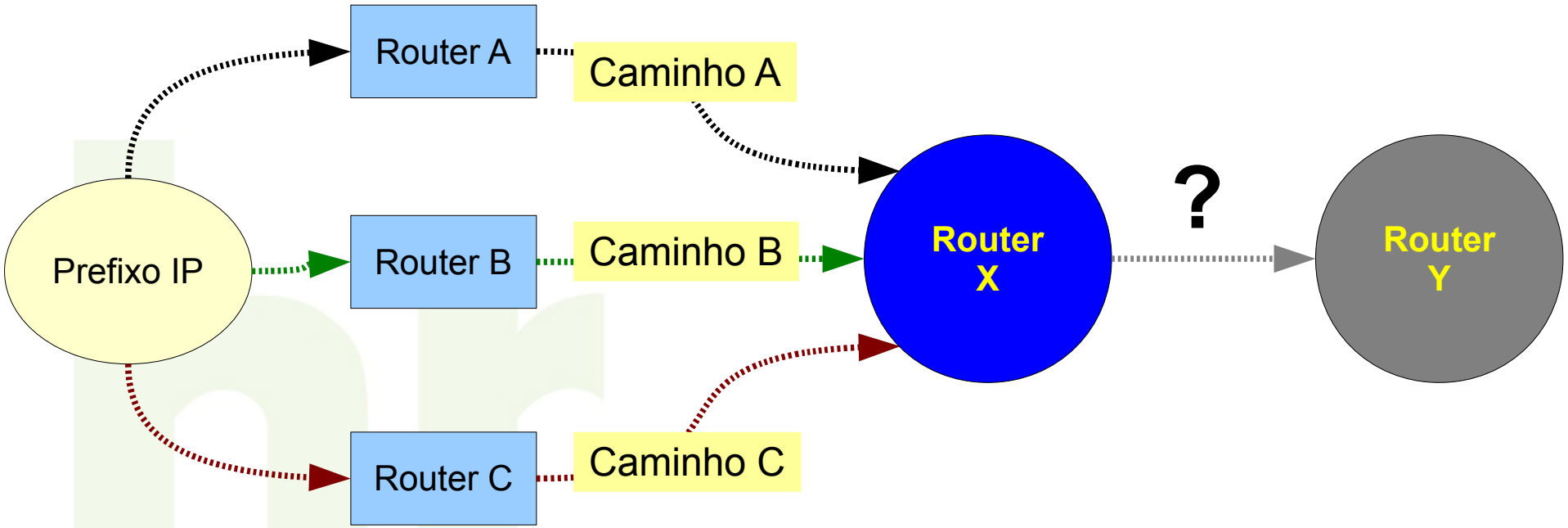
Ao NIC.br/PTT.br por ceder a janela de uso do laboratório de roteamento para a implantação da rede de São Paulo (VPLS).

A equipe do PTT.br pelo apoio na preparação do laboratório para o Experimento desta apresentação, em especial ao:
Ademar Francisco de Almeida
Ailton Soares da Rocha

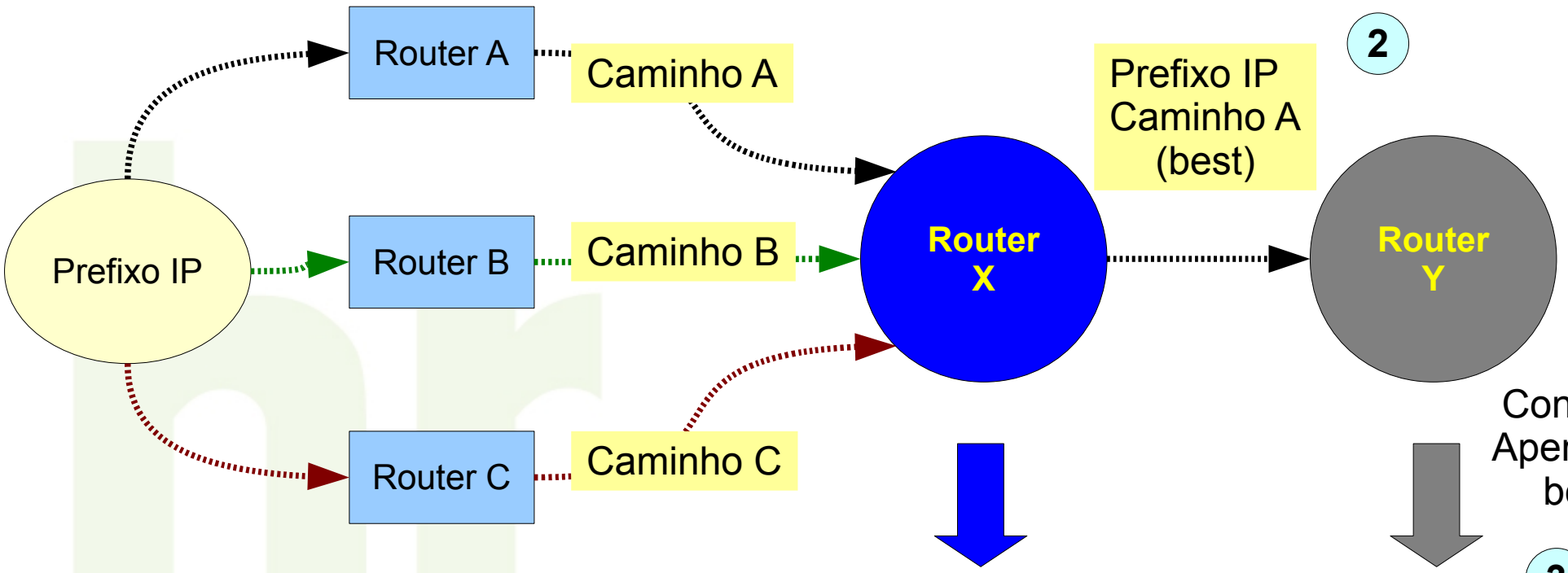
Modelo de Referência – Anúncio Somente Best Path



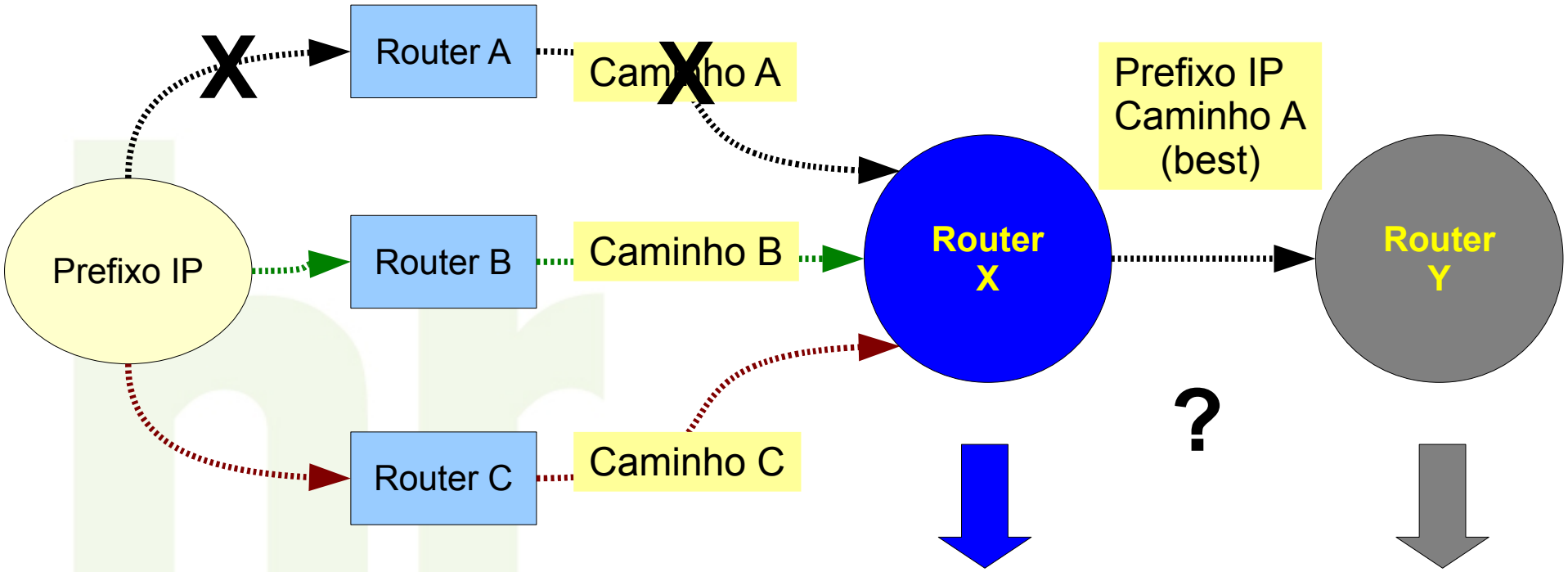
Modelo de Referência – Anúncio Somente Best Path



Modelo de Referência – Anúncio Somente Best Path



Evento no qual o caminho selecionado (best) torna-se inacessível



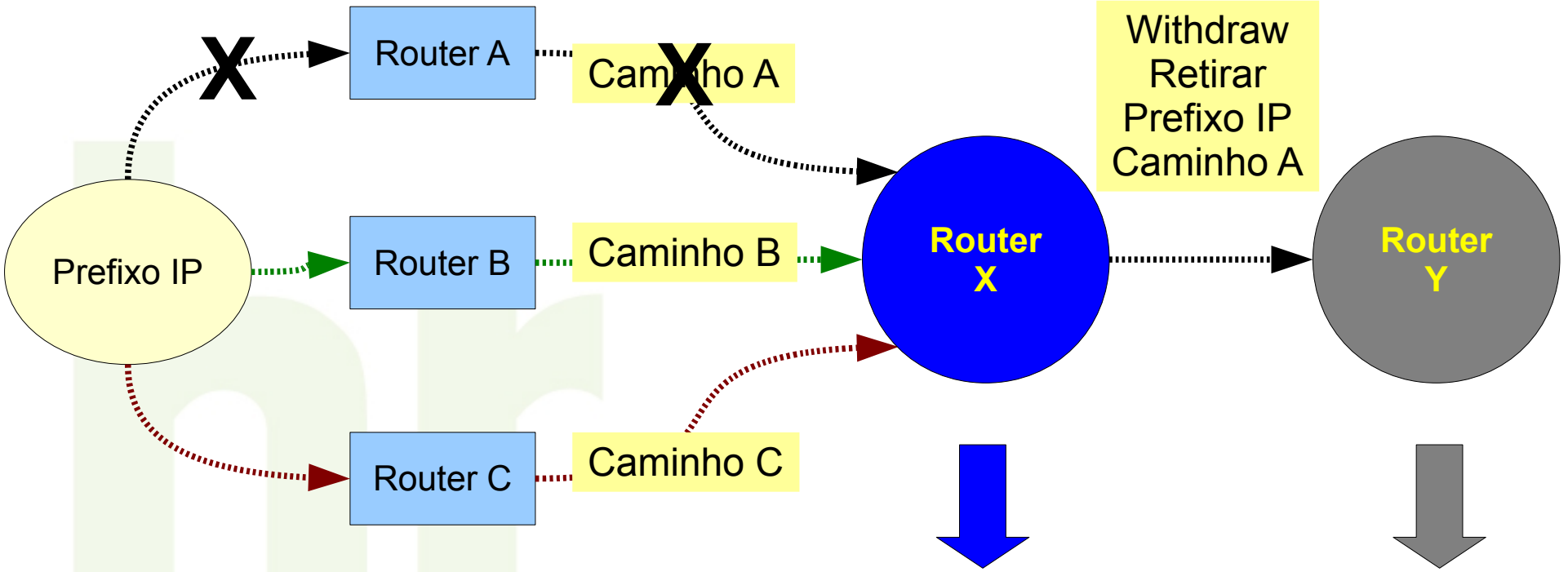
Router X

Prefixo IP
- **Caminho A (best)**
- Caminho B
- Caminho C

Router Y

Prefixo IP
- **Caminho A (best)**

Modelo de Referência – Anúncio Somente Best Path



** implicit withdraw*
De acordo com a implementação de BGP
As mensagens de retirada e novo entrada
De prefixo podem ser somadas, o que também
Reduz o tempo de convergência

Router X

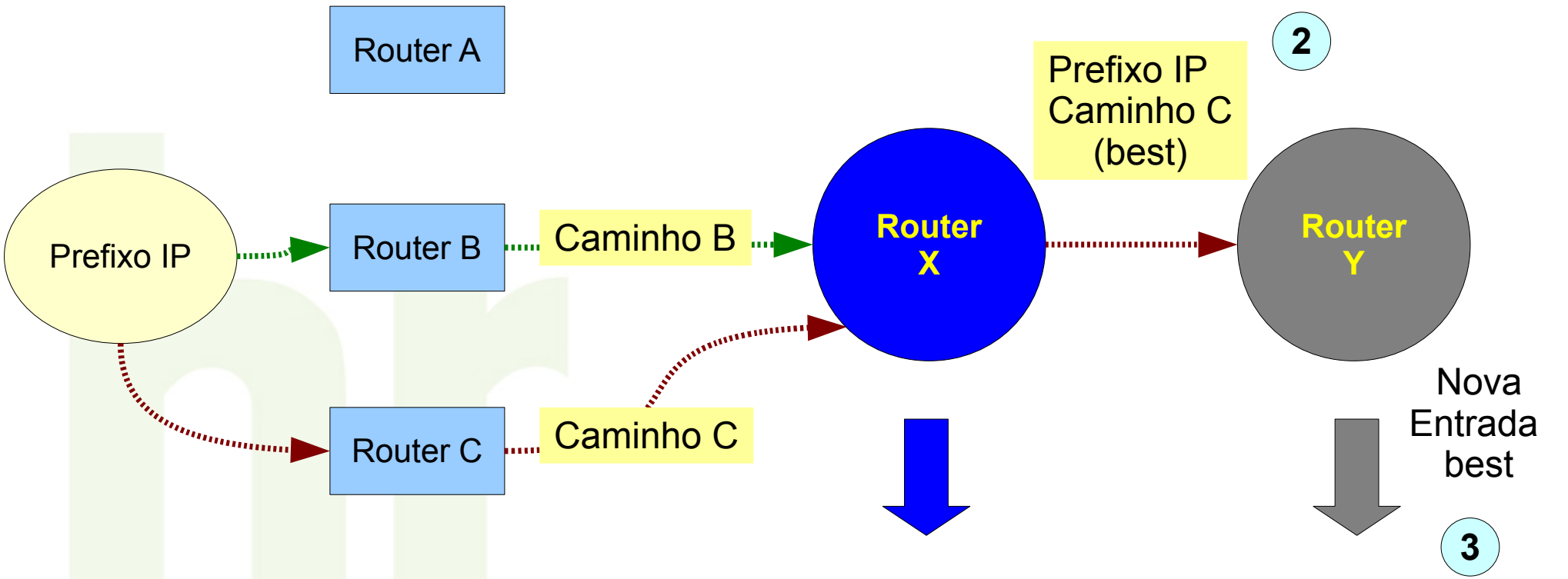
Prefixo IP

- Caminho B
- Caminho C

Router Y

Prefixo IP
unreachable
inacessível

Modelo de Referência – Anúncio Somente Best Path



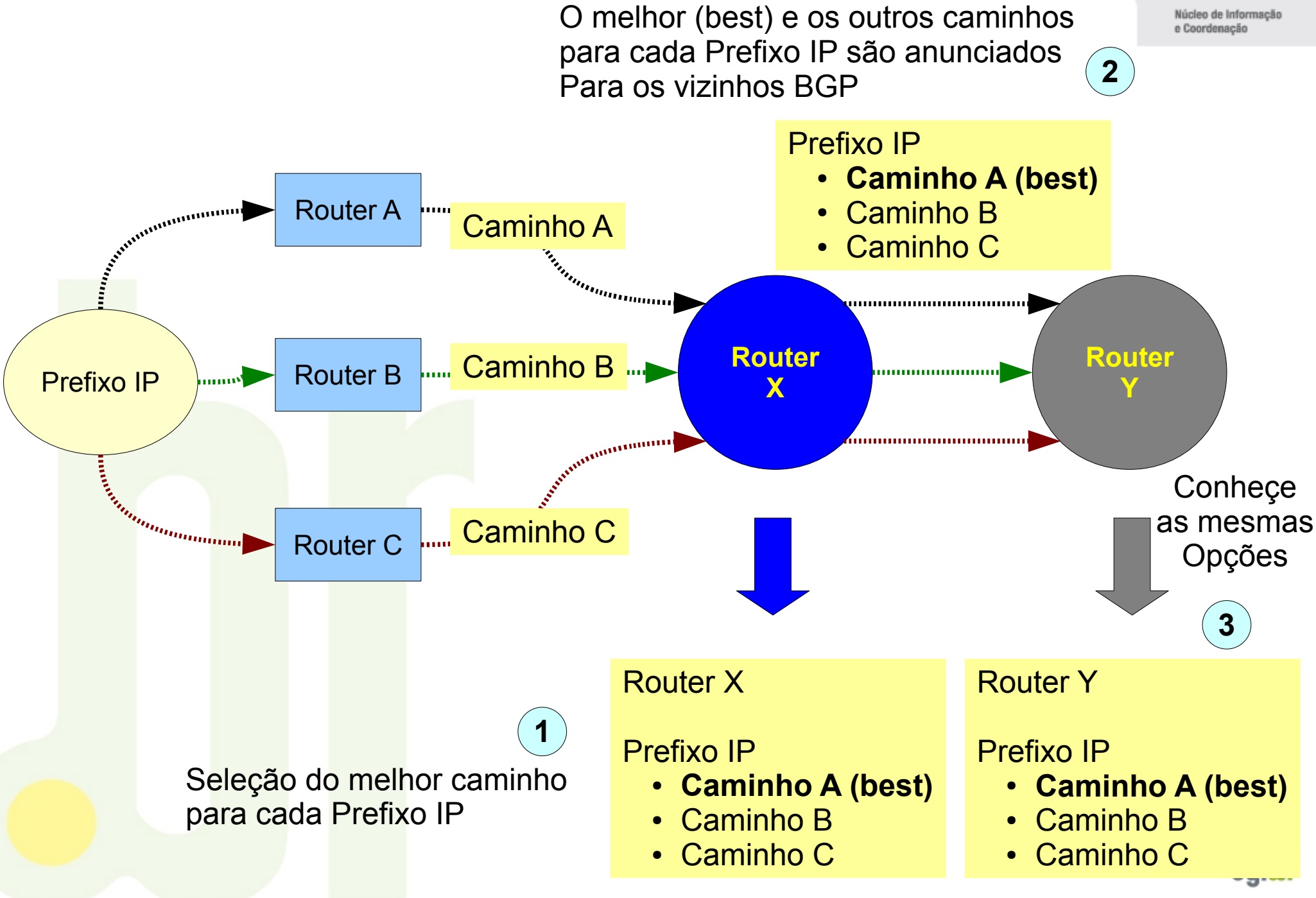
Proposta de Caminhos Adicionais (Add-Paths)

Ao invés de enviar apenas a melhor opção de caminho (best)

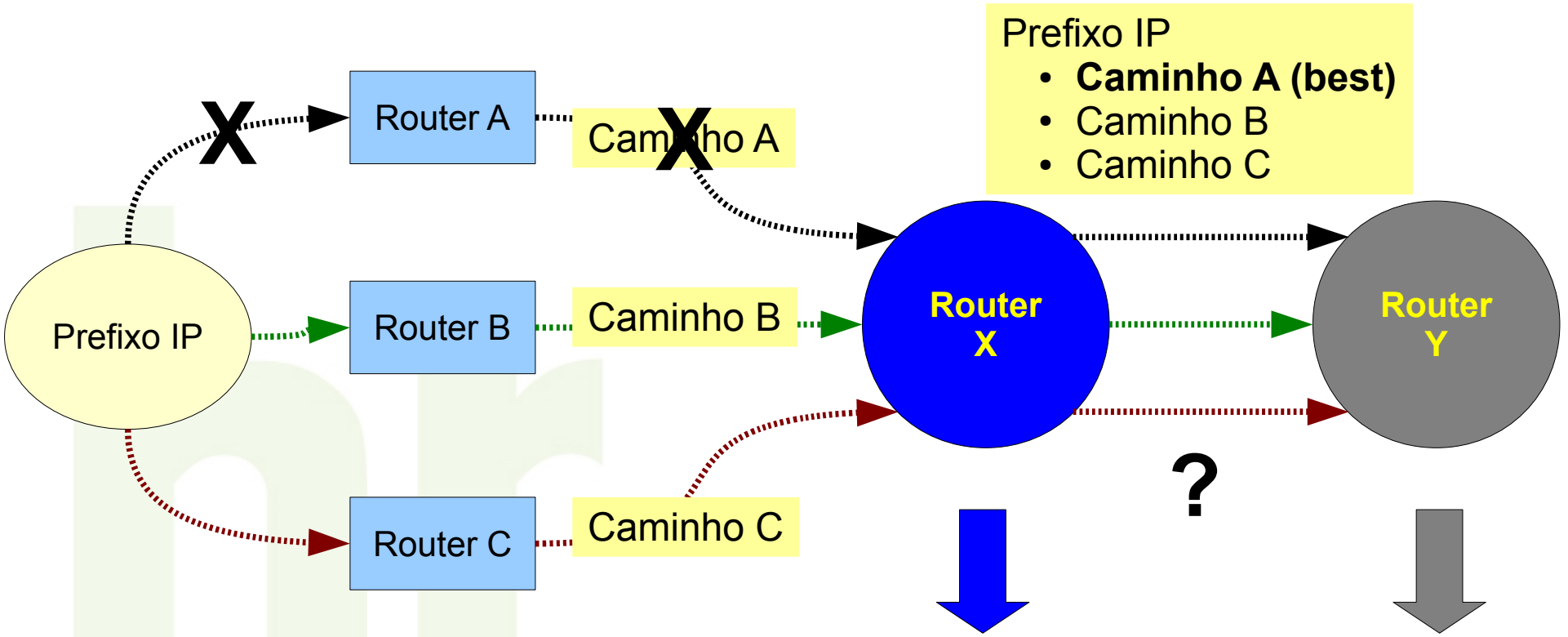
Enviar também a segunda melhor opção de caminho (backup) ou mesmo

Todas as opções de caminhos conhecidos

Modelo de Referência – BGP Add Paths



Evento no qual o caminho selecionado (best) torna-se inacessível



Prefixo IP
- **Caminho A (best)**
- Caminho B
- Caminho C

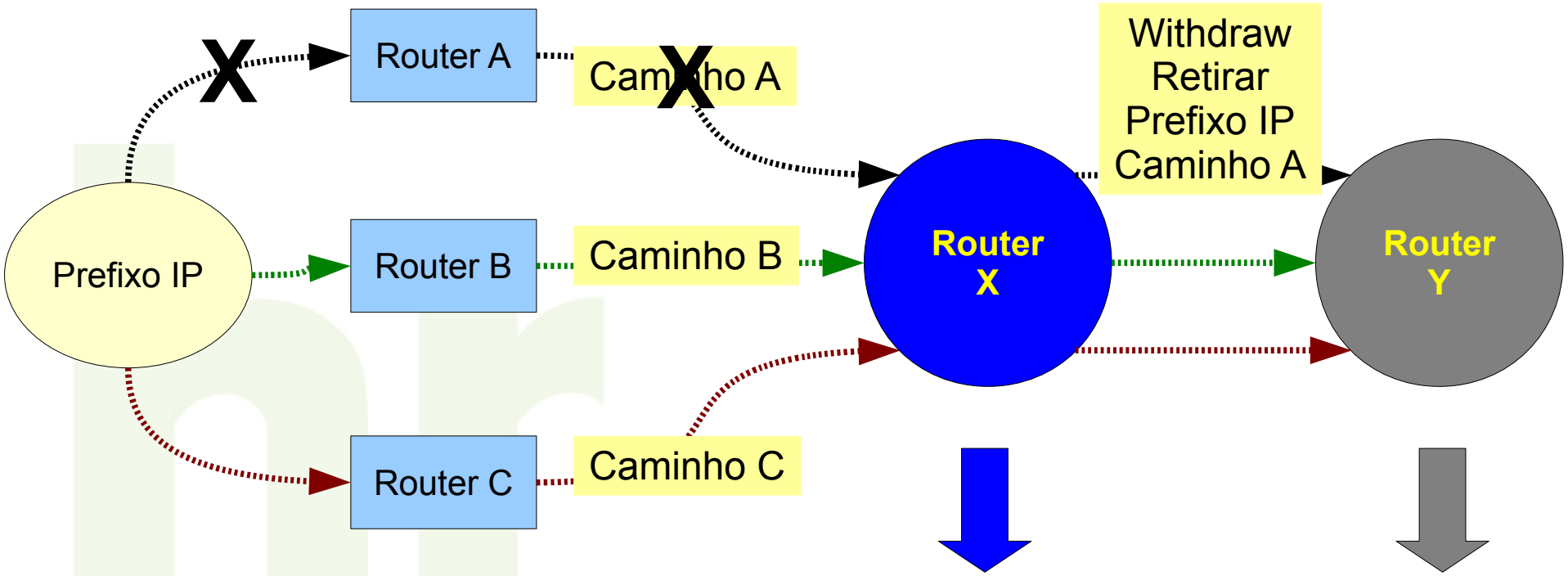
Router X

Prefixo IP
- **Caminho A (best)**
- Caminho B
- Caminho C

Router Y

Prefixo IP
- **Caminho A (best)**
- Caminho B
- Caminho C

Evento que o caminho selecionado (best) torna-se inacessível



Router X

Prefixo IP

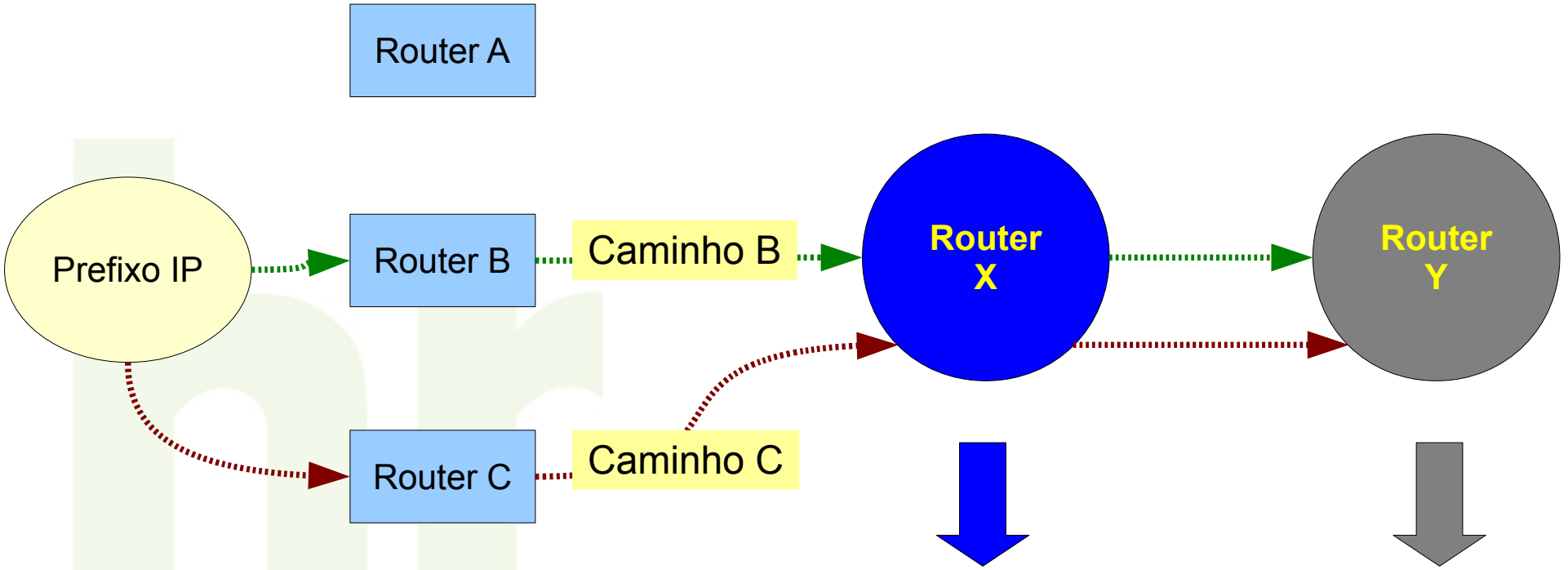
- Caminho B
- Caminho C

Router Y

Prefixo IP

- Caminho B
- Caminho C

Modelo de Referência – BGP Add Paths



1 Nova Seleção do melhor caminho para o Prefixo IP que oscilou

Router X 1

Prefixo IP

- Caminho B
- **Caminho C (best)**

Router Y 1

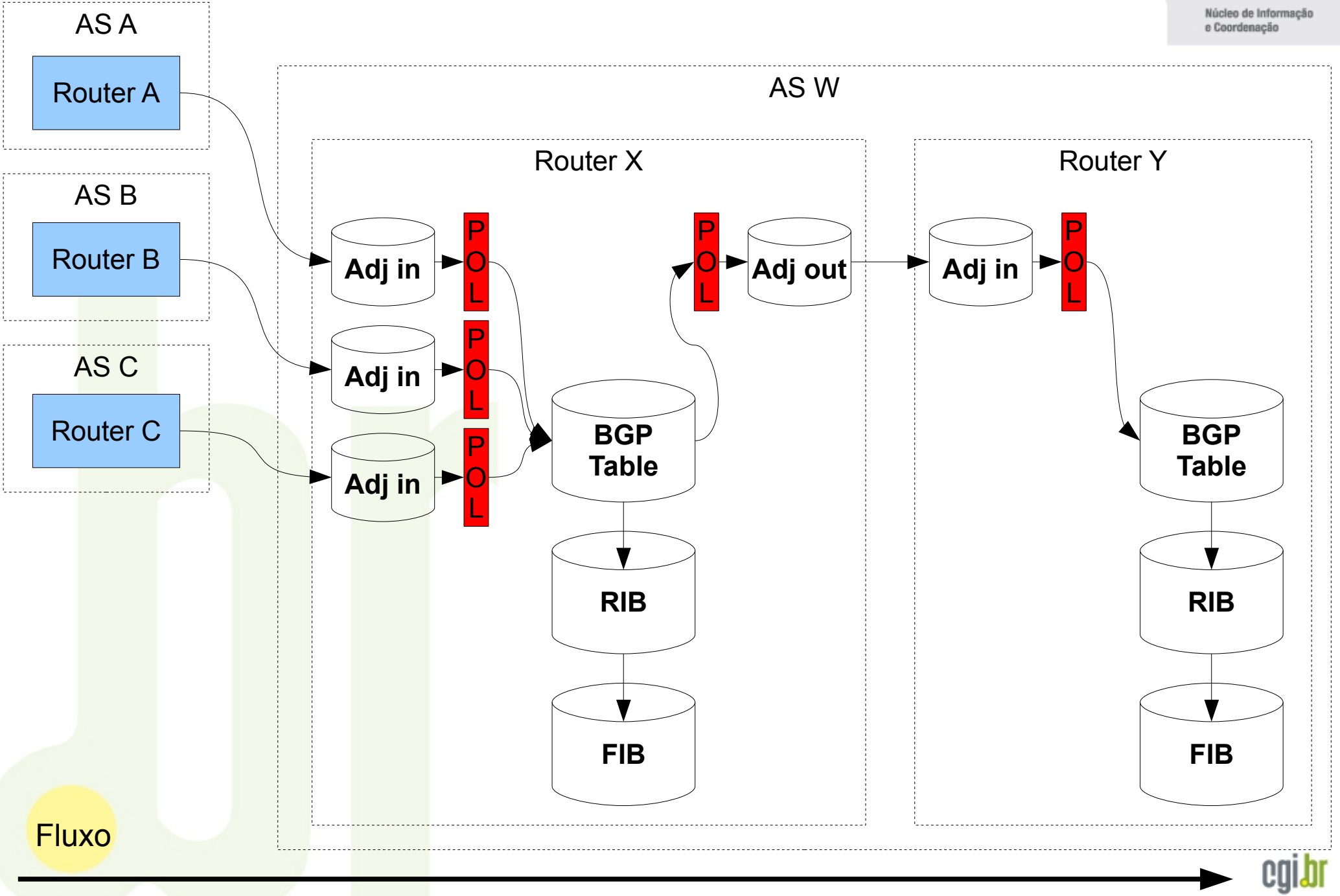
Prefixo IP

- Caminho B
- **Caminho C (best)**

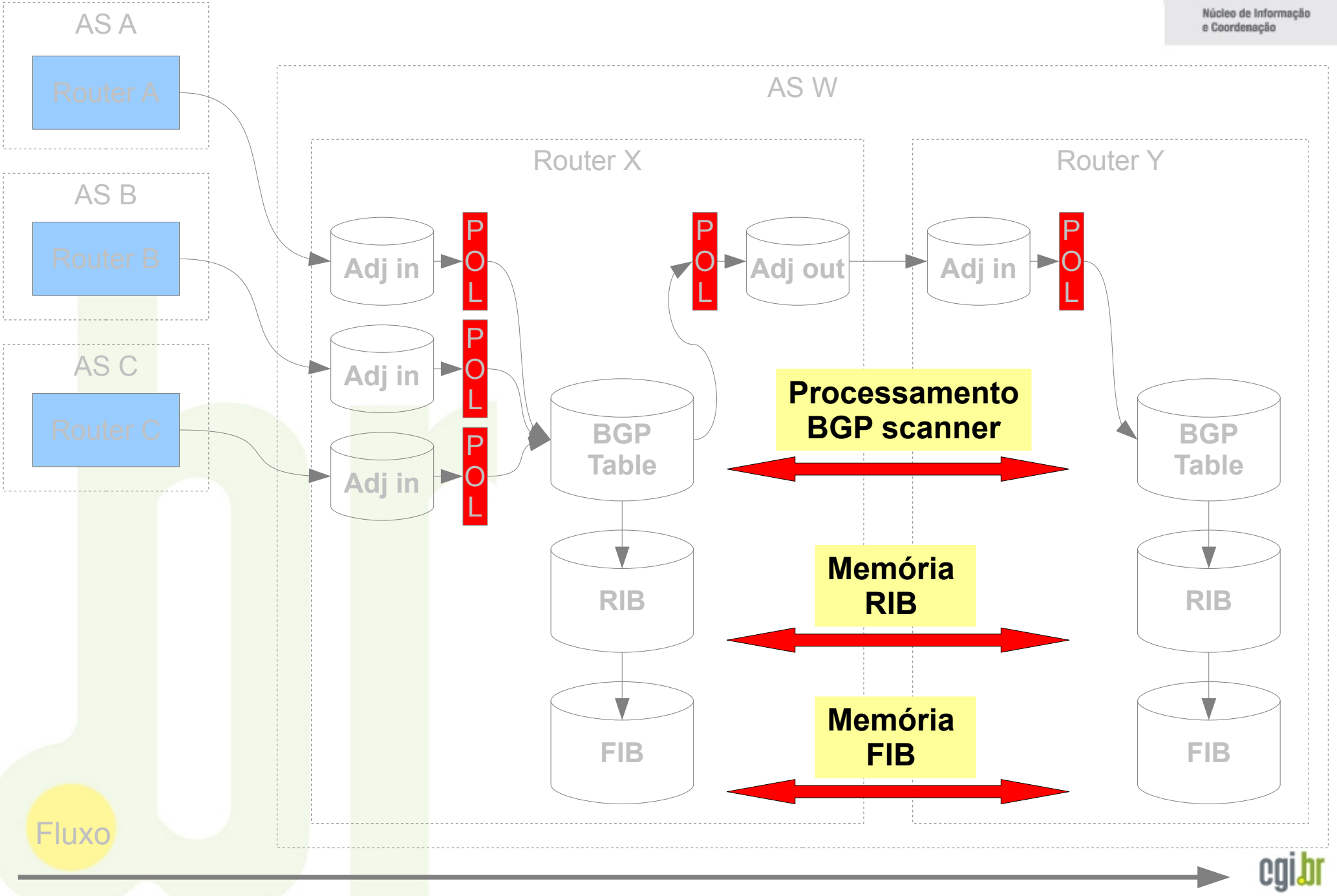
Possível resultado imediato do uso de Caminhos Adicionais:

Menor Tempo de Convergência da Rede após Interferência.

Modelo de Referência – Roteador com BGP



Modelo de Referência – Roteador com BGP



http://datatracker.ietf.org/doc/draft-ietf-idr-add-paths/

Advertisement of Multiple Paths in BGP
Draft-ietf-idr-add-paths-08

Abstract

In this document we propose a BGP extension that allows the advertisement of multiple paths for the same address prefix without the new paths implicitly replacing any previous ones. The essence of the extension is that each path is identified by a path identifier in addition to the address prefix.

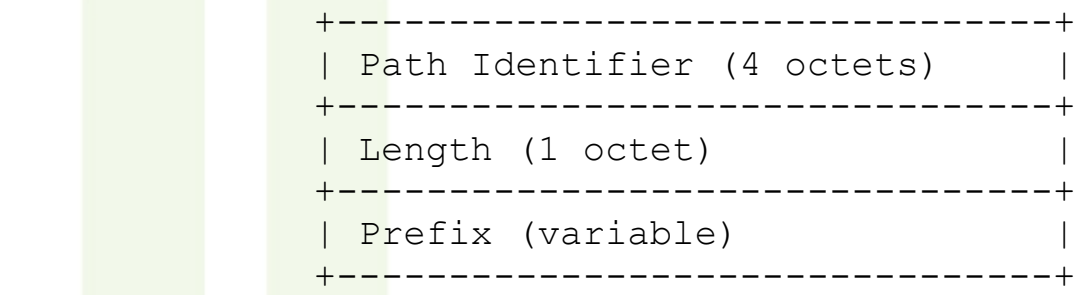
Advertisement of Multiple Paths in BGP

Draft-ietf-idr-add-paths-08

3. Extended NLRI Encodings

In order to carry the Path Identifier in an UPDATE message, the existing NLRI encodings are extended by prepending the Path Identifier field, which is of four-octets.

For example, the NLRI encodings specified in [RFC4271, RFC4760] are extended as the following:

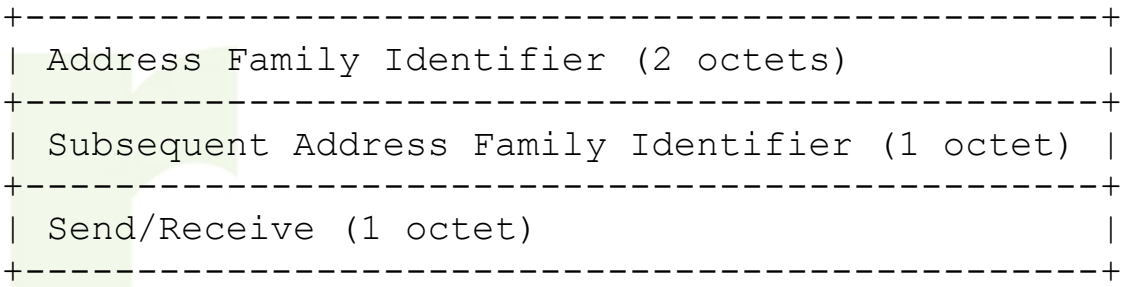


Advertisement of Multiple Paths in BGP

Draft-ietf-idr-add-paths-08

4. ADD-PATH Capability

The ADD-PATH Capability is a new BGP capability [RFC5492]. The Capability Code for this capability is specified in the IANA Considerations section of this document. The Capability Length field of this capability is variable. The Capability Value field consists of one or more of the following tuples:



The meaning and use of the fields are as follows:

Address Family Identifier (AFI):
This field is the same as the one used in [RFC4760].

Subsequent Address Family Identifier (SAFI):
This field is the same as the one used in [RFC4760].

Send/Receive:
This field indicates whether the sender is (a) willing to receive multiple paths from its peer (value 1), (b) would like to send multiple paths to its peer (value 2), or (c) both (value 3) for the <AFI, SAFI>.

Advertisement of Multiple Paths in BGP

Draft-ietf-idr-add-paths-08

6. Applications

The BGP extension specified in this document can be used by a BGP speaker to advertise multiple paths in certain applications. The availability of the additional paths can help reduce or eliminate persistent route oscillations [RFC3345]. It can also help with optimal routing and routing convergence in a network. The applications are detailed in separate documents.

7. Deployment Considerations

The extension proposed in this document provides a mechanism for a BGP speaker to advertise multiple paths over a BGP session. Care needs to be taken in its deployment to ensure consistent routing and forwarding in a network, the details of which will be described in separate application documents.

8. IANA Considerations

IANA has assigned capability number 69 for the ADD-PATH Capability described in this document. This registration is in the BGP Capability Codes registry.

http://datatracker.ietf.org/doc/draft-ietf-idr-add-paths-guidelines/

Best Practices for Advertisement of Multiple Paths in IBGP
draft-ietf-idr-add-paths-guidelines-04

Abstract

Add-Paths is a BGP enhancement that allows a BGP router to advertise multiple distinct paths for the same prefix/NLRI. This provides a number of potential benefits, including reduced routing churn, faster convergence and better loadsharing.

This document provides recommendations to implementers of Add-Paths so that network operators have the tools needed to address their specific applications and to manage the scalability impact of Add-Paths. A router implementing Add-Paths may learn many paths for a prefix and must decide which of these to advertise to peers. This document analyses different algorithms for making this selection and provides recommendations based on the target application.

Best Practices for Advertisement of Multiple Paths in IBGP

draft-ietf-idr-add-paths-guidelines-04

Table of Contents

Best Practices for Advertisement of Multiple Paths in IBGP

draft-ietf-idr-add-paths-guidelines-04

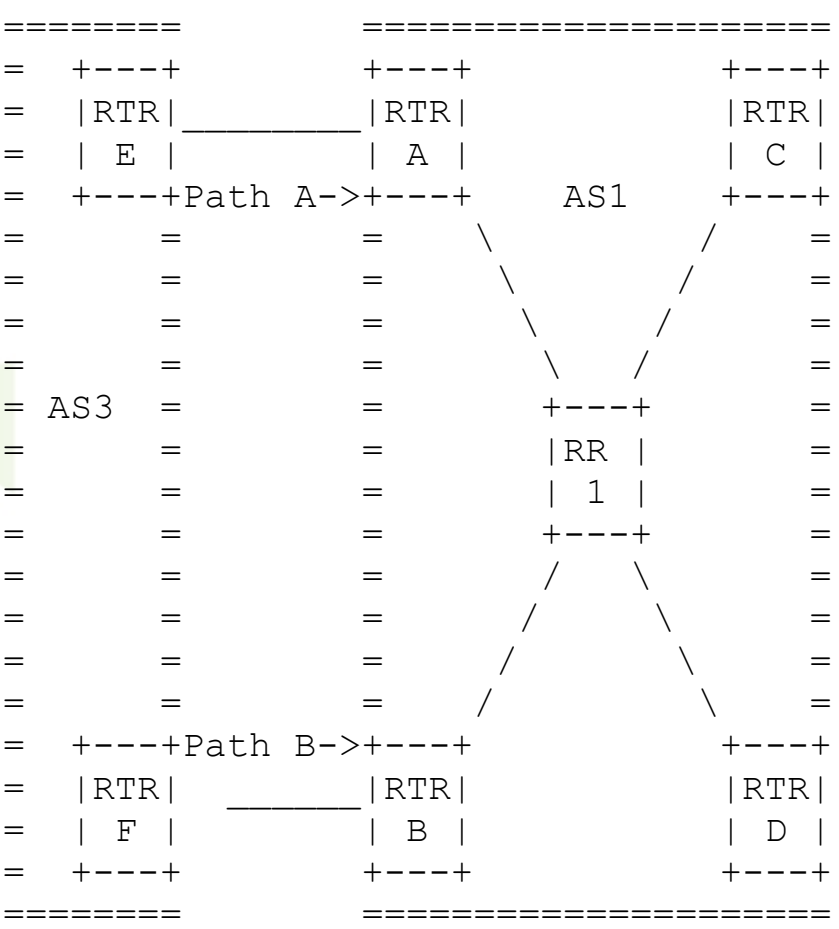


Figure 1: Example Topology

http://datatracker.ietf.org/doc/draft-pmohapat-idr-fast-conn-restore/

Fast Connectivity Restoration Using BGP Add-path
Draft-pmohapat-idr-fast-conn-restore-03

Abstract

A BGP route defines an association of an address prefix with an "exit point" from the current Autonomous System (AS). If the exit point becomes unreachable due to a failure, the route becomes invalid. This usually triggers an exchange of BGP control messages after which a new BGP route for the given prefix is installed. However, connectivity can be restored more quickly if the router maintains precomputed BGP backup routes. It can then switch to a backup route immediately upon learning that an exit point is unreachable, without needing to wait for the BGP control messages exchange. This document specifies the procedures to be used by BGP to maintain and distribute the precomputed backup routes. Maintaining these additional routes is also useful in promoting load balancing, performing maintenance without causing traffic loss, and in reducing churn in the BGP control plane.

Cisco - Referência BGP Add-Paths

Cisco ASR 9000 Series Aggregation Services Router Routing Configuration Guide, Release 4.2.x

Implementing BGP on Cisco ASR 9000 Series Router

http://www.cisco.com/en/US/docs/routers/asr9000/software/asr9k_r4.2/routing/configuration/guide/b_routing_cg42asr9k_chapter_00.html#task_6510DCF00831458692D0AC4285E08FD3

Advanced Routing Resiliency
http://www.cisco.com/web/SK/expo2012/pdf/advanced_routing_resilliency_josef_ungerman_cisco.pdf

Configuring BGP Additional Paths: Example - TAC engineer IOS XR group

This is a sample configuration for enabling BGP Additional Paths send, receive, and selection capabilities:

```
!
route-policy add_path_policy
  if community matches-any (*) then
    set path-selection all advertise
  else
    pass
  endif
end-policy
!
router bgp xxxxx
!
 address-family ipv4 unicast
  additional-paths receive
  additional-paths send
  additional-paths selection route-policy add_path_policy

!
```

Configuring BGP Additional Paths: Example - TAC engineer IOS XR group

Enable BGP peers to advertise best-external

```
!
router bgp xxxxx
 address-family ipv4 unicast
  advertise best-external
 !
!
```

Configuring BGP Additional Paths: Example - TAC engineer IOS XR group

Tell BGP to install a backup path in the FIB table

```
!
route-policy backup
  set path-selection backup 1 install multipath-protect advertise
end-policy
!
router bgp xxxxxx
 address-family ipv4 unicast
  additional-paths selection route-policy backup
 !
!
end
```

Diagrama Lógico - Topologia

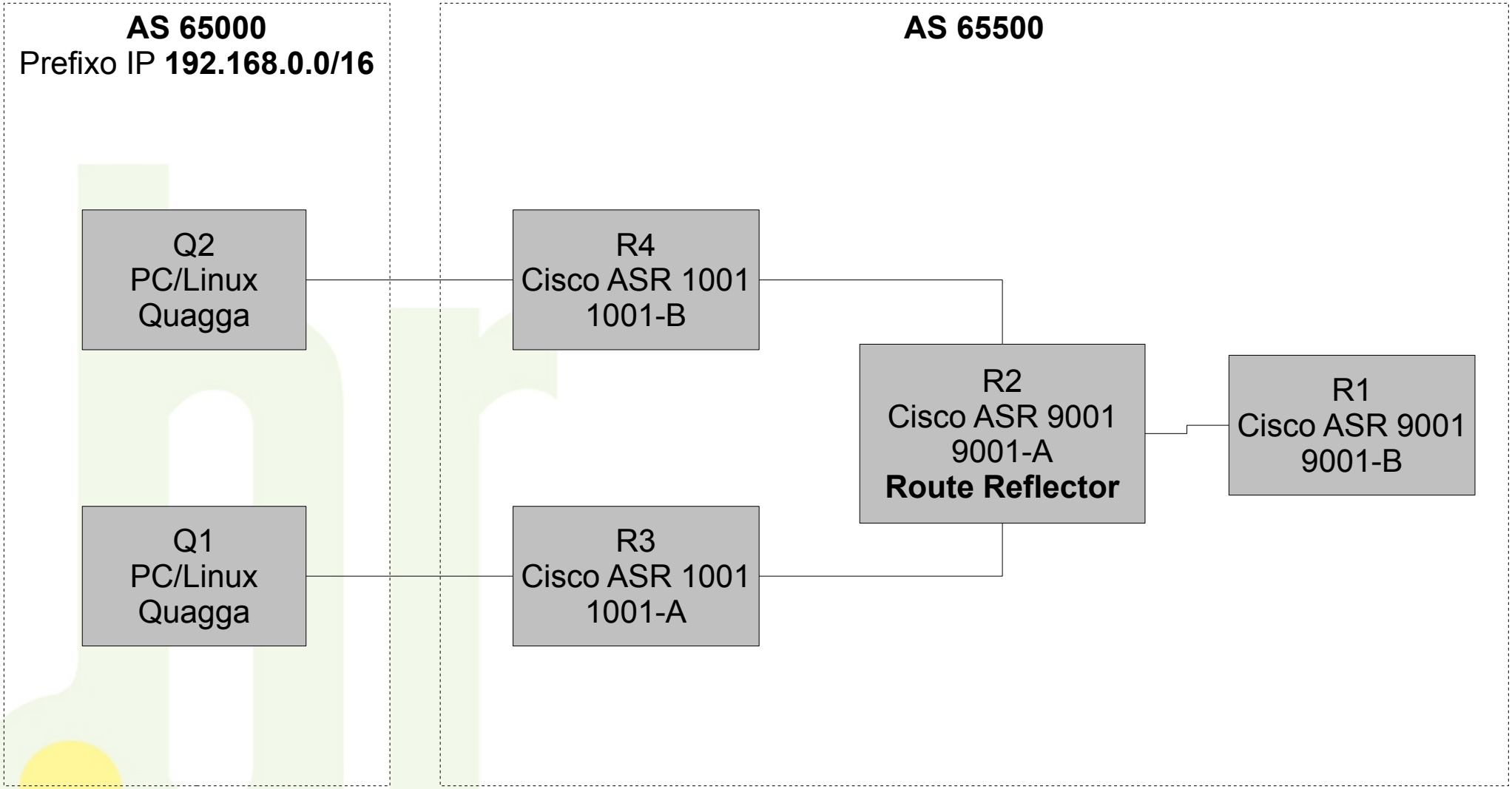
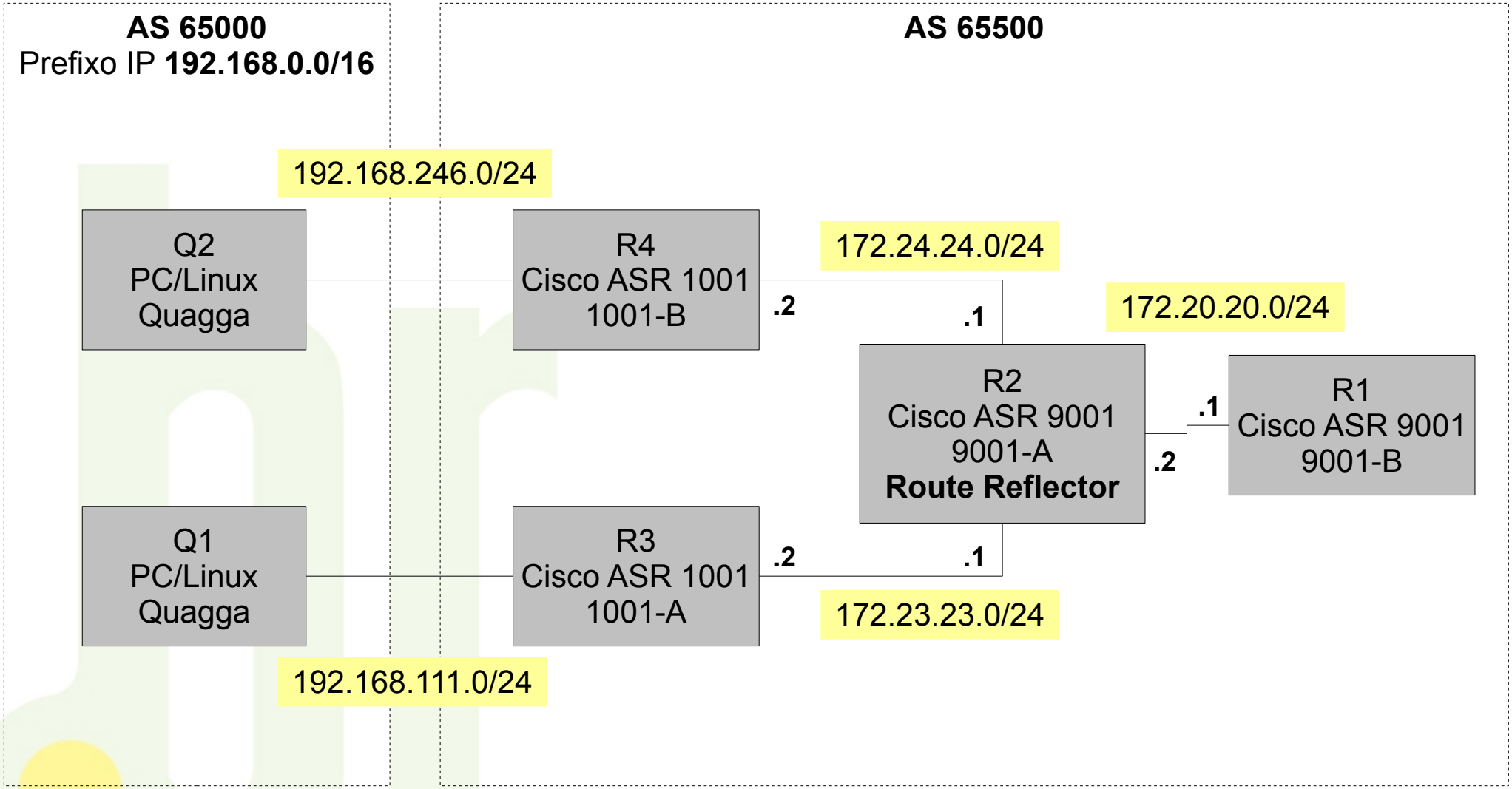


Diagrama Lógico – Endereçamento IP



```
RP/0/RSP0/CPU0:9001-A#show version
Thu May 23 16:09:55.157 UTC

Cisco IOS XR Software, Version 4.2.3[Default]
Copyright (c) 2012 by Cisco Systems, Inc.

ROM: System Bootstrap, Version 1.29(20120327:203546) [ASR9K ROMMON],

9001-A uptime is 5 days, 11 hours, 53 minutes
System image file is "bootflash:disk0/asr9k-os-mbi-4.2.3/0x100000/mbiasr9k-rp.vm"

cisco ASR9K Series (P4040) processor with 8388608K bytes of memory.
P4040 processor at 1500MHz, Revision 2.0
ASR-9001 Chassis

2 Management Ethernet
12 TenGigE
12 DWDM controller(s)
12 WANPHY controller(s)
219k bytes of non-volatile configuration memory.
2940M bytes of hard disk.
4014064k bytes of disk0: (Sector size 512 bytes).

Configuration register on node 0/RSP0/CPU0 is 0x2102
(...)
```

```
RP/0/RSP0/CPU0:9001-A#show running-config router bgp
Thu May 23 15:50:07.464 UTC
router bgp 65500
 bgp router-id 172.20.20.2
 bgp cluster-id 172.20.20.2
 address-family ipv4 unicast
  additional-paths receive
  additional-paths send
  advertise best-external
  additional-paths selection route-policy backup
 !
 neighbor 172.20.20.1
  remote-as 65500
  address-family ipv4 unicast
   route-policy pass-all in
   route-reflector-client
   route-policy pass-all out
  !
 !
(...)
```

```
RP/0/RSP0/CPU0:9001-A#show bgp summary
Thu May 23 15:53:15.447 UTC
BGP router identifier 172.20.20.2, local AS number 65500
BGP generic scan interval 60 secs
BGP table state: Active
Table ID: 0xe0000000   RD version: 39
BGP main routing table version 39
BGP scan interval 60 secs

BGP is operating in STANDALONE mode.


Process       RcvTblVer   bRIB/RIB   LabelVer  ImportVer  SendTblVer  StandbyVer
Speaker              39         39         39         39          39          39

Neighbor        Spk    AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  St/PfxRcd
172.20.20.1       0 65500    3439    3487       39    0    0 00:06:24           0
172.23.23.2       0 65500    3912    3547       39    0    0 00:06:48           1
172.24.24.2       0 65500    3901    3557       39    0    0 00:06:48           1

RP/0/RSP0/CPU0:9001-A#
```

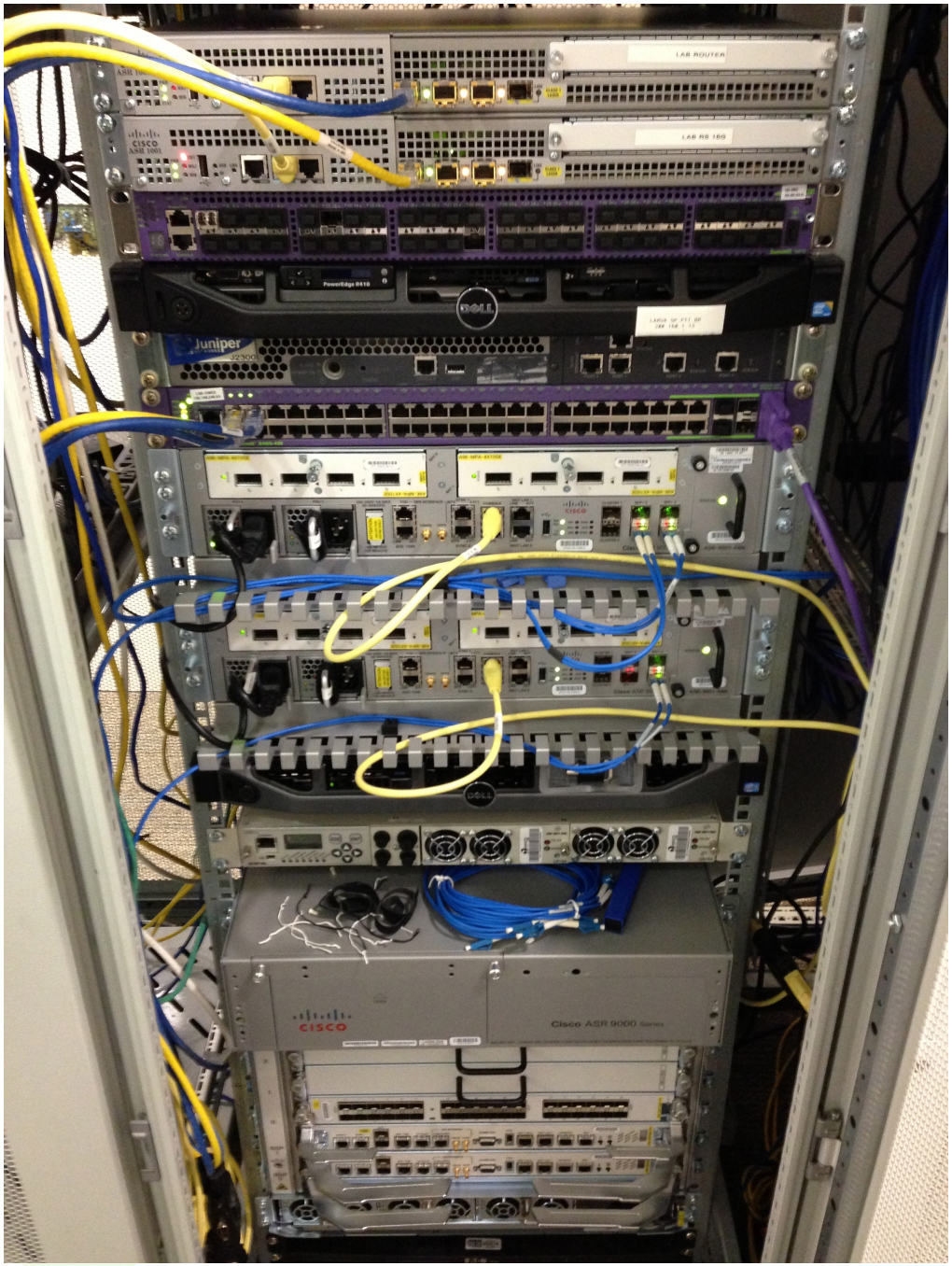
```
RP/0/RSP0/CPU0:9001-A#show bgp
Thu May 23 15:54:48.432 UTC
BGP router identifier 172.20.20.2, local AS number 65500
BGP generic scan interval 60 secs
BGP table state: Active
Table ID: 0xe0000000   RD version: 39
BGP main routing table version 39
BGP scan interval 60 secs

Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network          Next Hop            Metric LocPrf Weight Path
*>i192.168.0.0/16    172.23.23.2              0    100      0 65000 i
* i                  172.24.24.2              0    100      0 65000 i

Processed 1 prefixes, 2 paths
RP/0/RSP0/CPU0:9001-A#
```

```
RP/0/RSP0/CPU0:9001-A#show bgp 192.168.0.0
Thu May 23 15:55:48.222 UTC
BGP routing table entry for 192.168.0.0/16
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 39          39
Last Modified: May 23 15:46:27.550 for 00:09:20
Paths: (2 available, best #1)
  Advertised to update-groups (with more than one peer):
    0.3
  Advertised to peers (in unique update groups):
    172.20.20.1
  Path #1: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.3
  Advertised to peers (in unique update groups):
    172.20.20.1
  65000, (Received from a RR-client)
    172.23.23.2 from 172.23.23.2 (172.23.23.2)
      Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
      Received Path ID 0, Local Path ID 1, version 39
  Path #2: Received by speaker 0
  Advertised to peers (in unique update groups):
    172.20.20.1
  65000, (Received from a RR-client)
    172.24.24.2 from 172.24.24.2 (172.24.24.2)
      Origin IGP, metric 0, localpref 100, valid, internal, backup, add-path
      Received Path ID 0, Local Path ID 2, version 39
RP/0/RSP0/CPU0:9001-A#
```

```
RP/0/RSP0/CPU0:9001-A#show bgp neighbor 172.20.20.1
Thu May 23 15:59:04.137 UTC

BGP neighbor is 172.20.20.1
 Remote AS 65500, local AS 65500, internal link
 Remote router ID 172.20.20.1
 Cluster ID 172.20.20.2
(…)
 For Address Family: IPv4 Unicast
  BGP neighbor version 39
  Update group: 0.4 Filter-group: 0.4  No Refresh request being processed
  Route-Reflector Client
  AF-dependent capabilities:
    Additional-paths Send: advertised and received
    Additional-paths Receive: advertised and received
  Route refresh request: received 0, sent 0
  Policy for incoming advertisements is pass-all
  Policy for outgoing advertisements is pass-all
  0 accepted prefixes, 0 are bestpaths
  Cumulative no. of prefixes denied: 0.
  Prefix advertised 2, suppressed 0, withdrawn 0
  Maximum prefixes allowed 1048576
  Threshold for warning message 75%, restart interval 0 min
  AIGP is enabled
  An EoR was not received during read-only mode
  Last ack version 39, Last synced ack version 0
  Outstanding version objects: current 0, max 4
  Additional-paths operation: Send and Receive
(...)
```

```
RP/0/RSP0/CPU0:9001-B#show bgp summary
Thu May 23 16:02:46.114 UTC
BGP router identifier 172.20.20.1, local AS number 65500
BGP generic scan interval 60 secs
BGP table state: Active
Table ID: 0xe0000000   RD version: 29
BGP main routing table version 29
BGP scan interval 60 secs

BGP is operating in STANDALONE mode.


Process       RcvTblVer   bRIB/RIB   LabelVer  ImportVer  SendTblVer  StandbyVer
Speaker              29         29         29         29          29          29

Neighbor        Spk    AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  St/PfxRcd
172.20.20.2       0 65500    2444    2420       29    0    0 00:16:08           2

RP/0/RSP0/CPU0:9001-B#
```

```
RP/0/RSP0/CPU0:9001-B#show bgp 192.168.0.0
Thu May 23 16:03:34.957 UTC
BGP routing table entry for 192.168.0.0/16
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 29          29
Last Modified: May 23 15:46:42.540 for 00:16:52
Paths: (2 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  65000
    172.23.23.2 (metric 2) from 172.20.20.2 (172.23.23.2)
      Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
      Received Path ID 1, Local Path ID 1, version 29
      Originator: 172.23.23.2, Cluster list: 172.20.20.2
  Path #2: Received by speaker 0
  Not advertised to any peer
  65000
    172.24.24.2 (metric 2) from 172.20.20.2 (172.24.24.2)
      Origin IGP, metric 0, localpref 100, valid, internal
      Received Path ID 2, Local Path ID 0, version 0
      Originator: 172.24.24.2, Cluster list: 172.20.20.2
RP/0/RSP0/CPU0:9001-B#
```


LAB ROUTER
LAB RS 1GS
Juniper
CISCO
Cisco ASR 9000 Series

http://datatracker.ietf.org/wg/idr/

e.g. RFC 4271

A Border Gateway Protocol 4 (BGP-4)

http://datatracker.ietf.org/doc/rfc4271/

http://www.ietf.org/

The goal of the IETF is to make the Internet work better.

The mission of the IETF is to make the Internet work better by producing high quality, relevant technical documents that influence the way people design, use, and manage the Internet.

http://datatracker.ietf.org/wg/

Active IETF Working Groups

IETF Working Groups (WGs) are the primary mechanism for development of IETF specifications and guidelines, many of which are intended to be standards or recommendations.

http://www.internetsociety.org/

Home » What We Do » Leadership Programmes »
IETF and OIS Programmes » Fellowship: IETF

The Internet Society Fellowship to the Internet Engineering Task Force (IETF) Programme

The Internet Engineering Task Force (IETF) is the world's premier open Internet standards-development body. The Internet Society Fellowship to the IETF is available to technology professionals, advanced IT students, and other qualified individuals from developing and emerging economies.

Fellows to the IETF attend an IETF meeting where they are paired with an experienced mentor and are expected to make a positive contribution to IETF work.

Obrigado & Contatos

Eduardo Ascenço Reis

<eascenco@nic.br>

<eduardo@intron.com.br>

Equipe PTT.br

<eng@ptt.br>

http://ptt.br/