

# Boas práticas para peering no PTTMetro

**Luís Balbinot**

[lbabinot@commcorp.com.br](mailto:lbabinot@commcorp.com.br)

Commcorp Telecom

GTER 30 - São Leopoldo, RS

26/11/2010

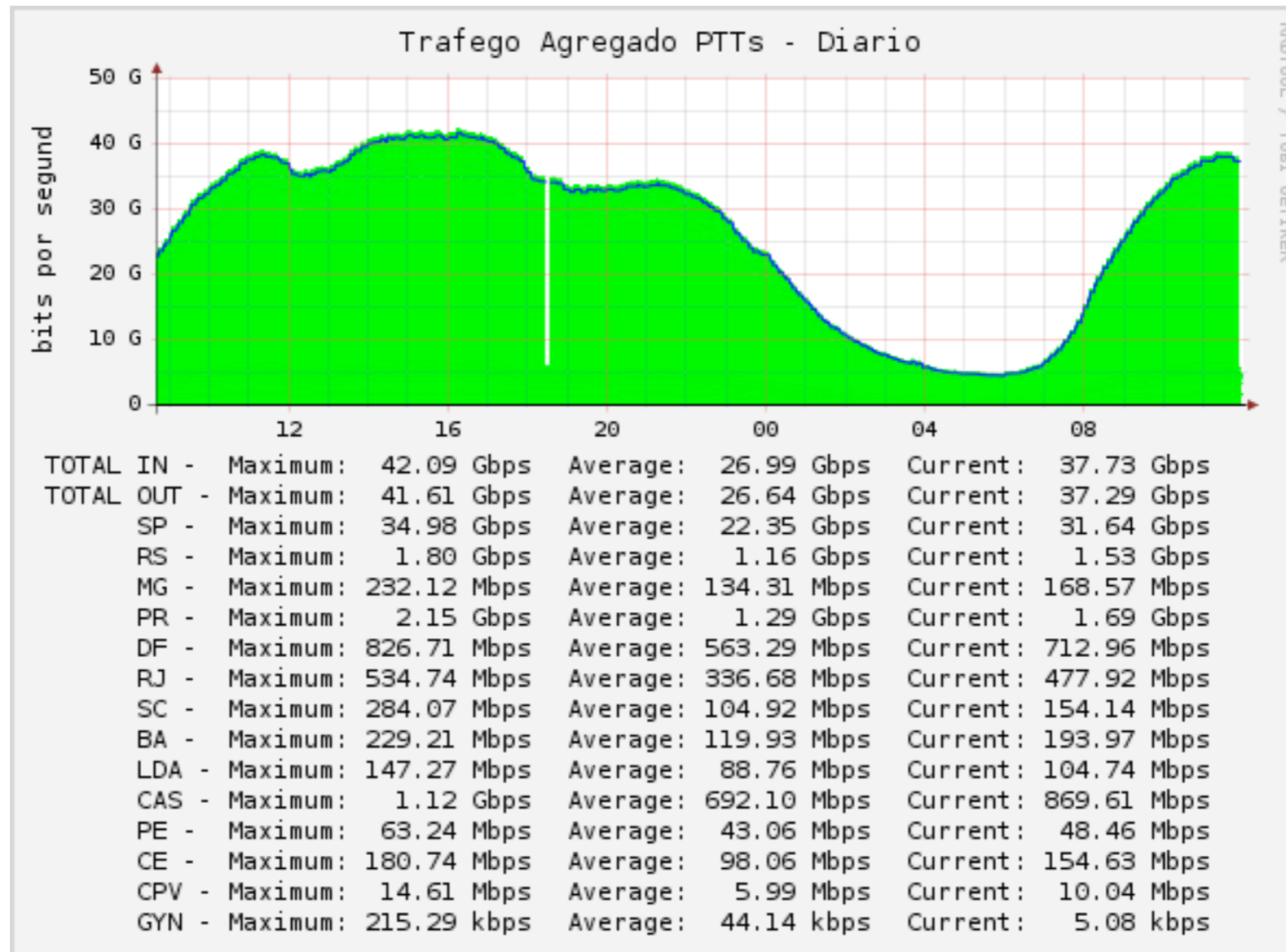
# Objetivo

Mostrar boas práticas de peering no PTTMetro de forma que se obtenha o maior proveito possível e, ao mesmo tempo, se colabore com a manutenção do projeto e organização da malha de troca de tráfego

# PTTMetro

- Projeto do NIC.br (Ceptro) - ptt.br
  - Sem fins lucrativos
  - Conjunto de pontos de interconexão (PIX) interligados por fibras óticas apagadas iluminadas com a tecnologia Ethernet
  - Sistema metropolitano (80km), colaborativo, distribuído, neutro, centralmente administrado e aberto
  - Suporte a serviços avançados (IPv6 e IP Multicast)
  - Atualmente presente em 14 localidades e crescendo rápido
    - Belo Horizonte, Brasília, Campina Grande, Campinas, Curitiba, Florianópolis, Fortaleza, Goiânia, Londrina, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo
- Qualquer AS pode participar!

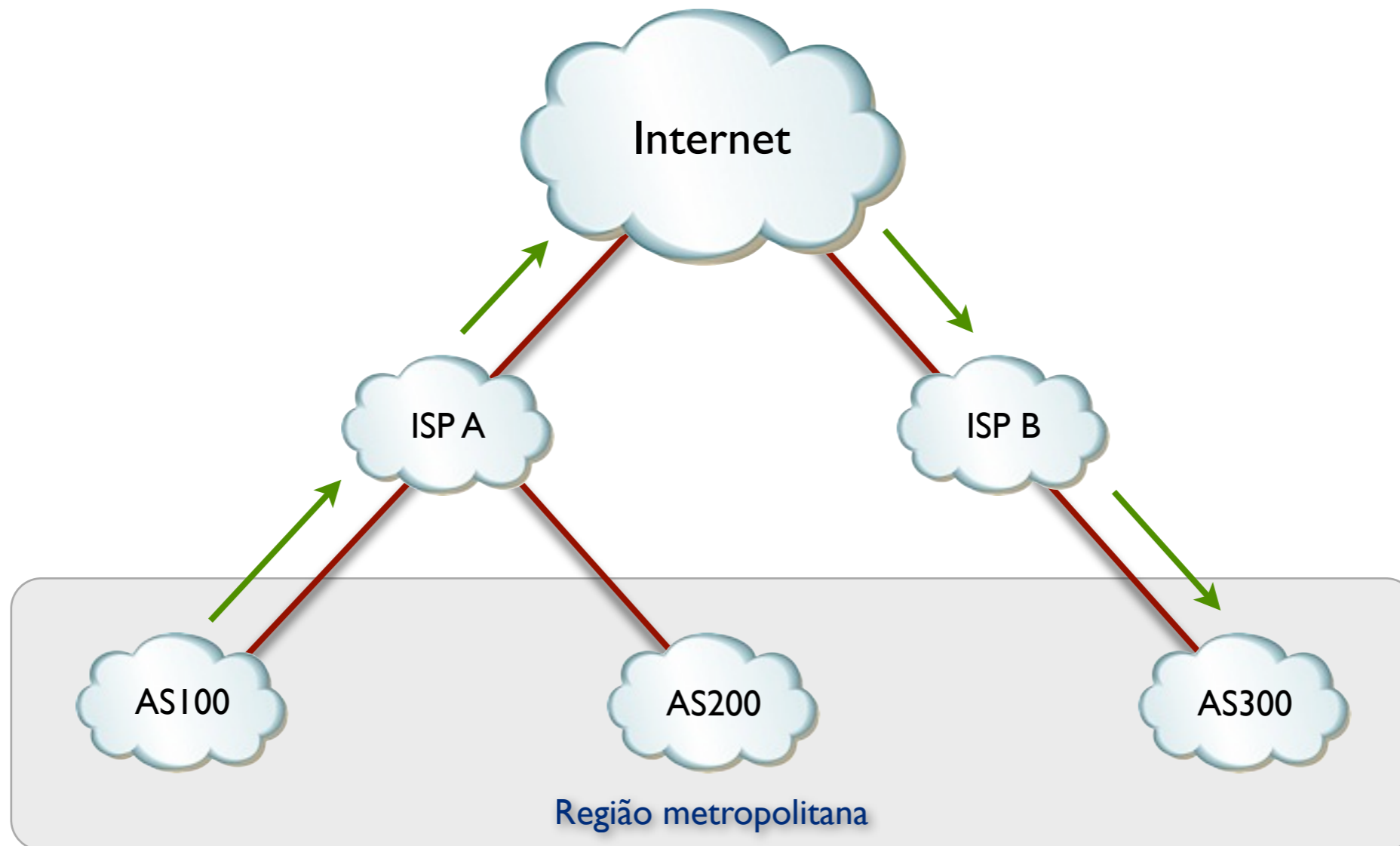
# Tráfego total



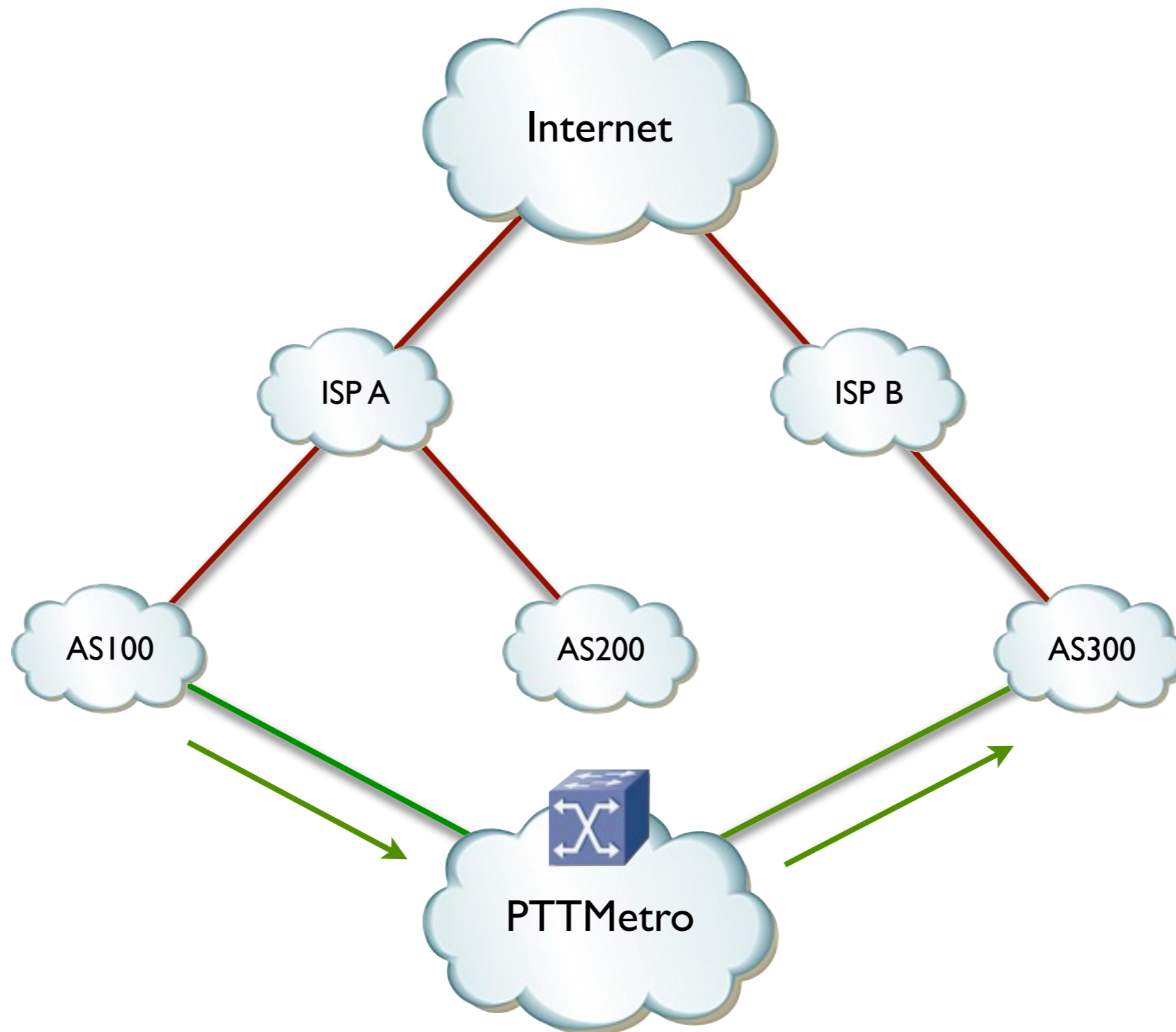
Fonte: site do pt.br (25/11/2010)

# Funcionamento

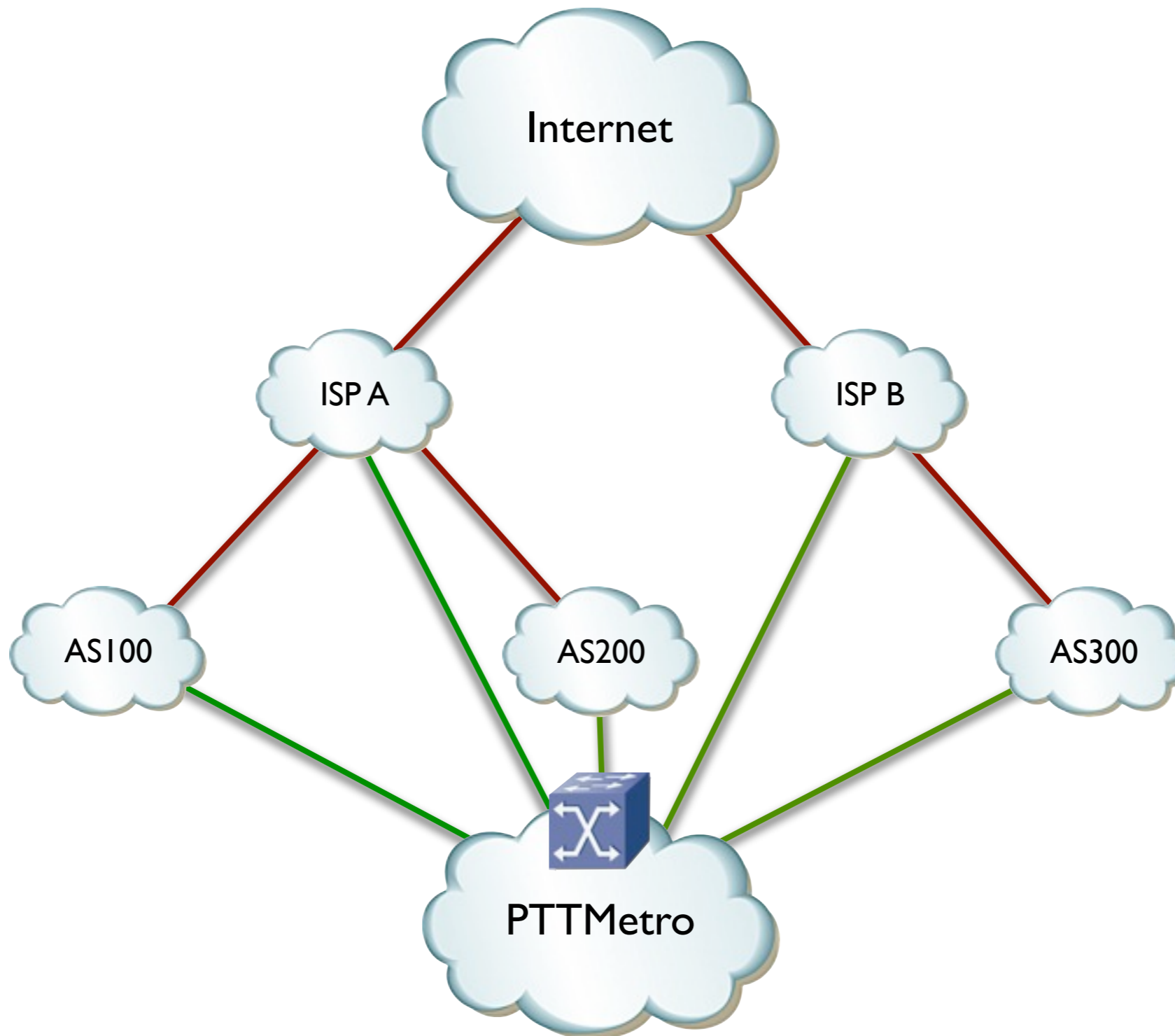
- Matriz de comutação non-blocking
- Isolamento lógico entre ASs através de VLANs
- Acordo de Troca Multilateral (ATM)
  - Todos trocam com todos, sem restrições
  - Utiliza um par de route-servers e um looking-glass
- Acordo de Troca Bilateral (ATB)
  - Entre dois ASs através de VLAN dedicada ou compartilhada
  - Feito normalmente para venda de trânsito ou Interconexão Classe V (Anatel)



Problema: AS100 precisa fazer trânsito pela Internet para chegar no AS300 e vice-versa: maior latência, perdas e custo



Solução: AS100 e AS300 entram no PTT local e fazem troca de tráfego direta, sem intermediários



Cenário ideal: todos, incluindo os ISPs de trânsito, trocam tráfego no ATM do PTT local



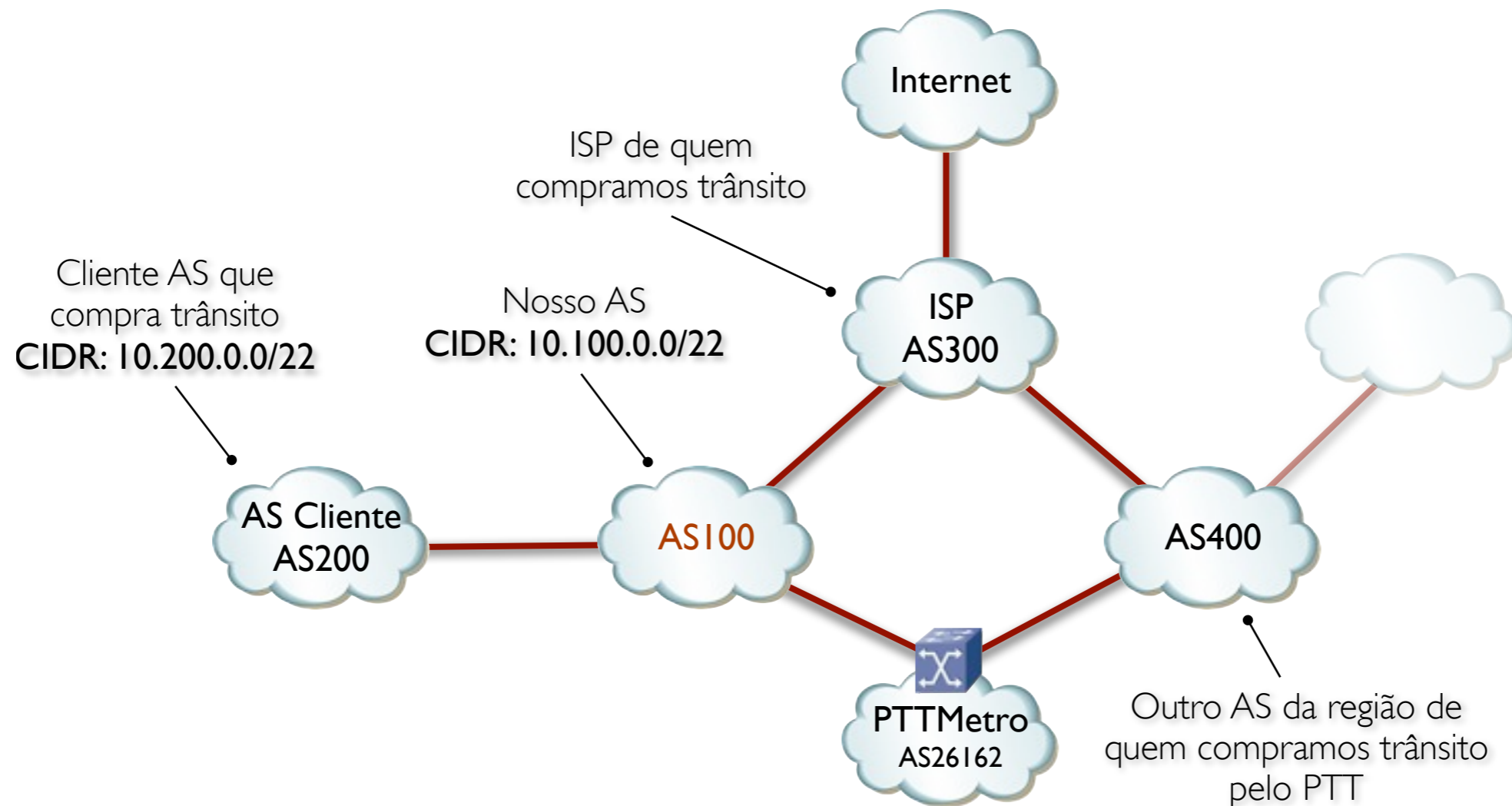
# Troca de tráfego no RS

- Começou com o RSiX em 2001 (7 ASs)
- Em 2005 vira PTTMetro com fundos do CGI.br
  - É criado um PIX Central (UFRGS) e um PIX na Procempa
- Em 2006 a SIM Telecom (Commcorp) vira PIX
- Em 2008 a Brasil Telecom (Oi) vira PIX
- Em 2010 a Stech via PIX
- Em 2010 é instalada uma cópia do Root Server “I”
- Hoje conta com 30 participantes e quase 2 Gbps de tráfego trocado no agregado total

# Boas práticas na interconexão

- PTT filtra um único MAC por participante
  - Qualquer outro MAC em uso gera alertas e sobrecarrega os logs
- Evite que pacotes indesejados sejam enviados para o PTT desligando ou filtrando:
  - BDPU's (STP e variantes)
  - Multicast (OSPF, IS-IS, IGMP, etc)
  - Recursos de rádios em bridge no acesso
  - Protocolos proprietários (CDP da Cisco, recursos obscuros do Mikrotik)
  - Etc...
- Filtros L3 de entrada e saída (não esqueça do IPv6)
- O PTT agradece!

# Topologia utilizada



Consiste de um AS (AS100) com um ISP de trânsito (AS300), uma conexão ao PTT (AS26162) um AS cliente (AS200) e um outro AS (AS400) que troca tráfego e vende trânsito pelo PTT

# Política de tráfego

- Prioridade de tráfego, em ordem decrescente:
  - Clientes AS (localpref 400)
  - ATM no PTTMetro (localpref 300)
  - Trânsito comprado dentro do PTTMetro (localpref 200)
  - Trânsito Internet (localpref 100)
- Tradução de políticas para a engenharia de tráfego via BGP em duas etapas:
  - Filtros de entrada: controlam a saída do tráfego
  - Filtros de saída: controlam a entrada do tráfego

# Política de communities

- Sintaxe:
  - **ASN:Código** (padrão de facto)
- Controle de prefixos aprendidos:
  - **100:10** - Prefixos aprendidos de peerings de trânsito
  - **100:20** - Prefixos aprendidos do PTTMetro
  - **100:30** - Prefixos aprendidos de clientes AS
- Controle de anúncios:
  - **100:100** - Anuncia para todos os peerings
  - **100:200** - Anuncia apenas em peerings de trânsito
  - **100:300** - Anuncia apenas no PTTMetro

# Balanciamento de tráfego



- Anúncio do CIDR completo (/22) em todos peerings
- Anúncio da primeira metade (10.100.0.0/23) para o ISP de trânsito (AS300)
- Anúncio da segunda metade (10.100.0.2/23) para o AS400 de trânsito via PTTMetro
- Anúncio de todos os /24 que compõem o CIDR somente para o PTTMetro

# Injetando prefixos

```
router bgp 100
  network 10.100.0.0 mask 255.255.252.0 route-map BGP-ORIGINATE
  network 10.100.0.0 mask 255.255.254.0 route-map BGP-ORIGINATE
  network 10.100.0.2 mask 255.255.254.0 route-map BGP-ORIGINATE
  network 10.100.0.0 mask 255.255.255.0 route-map BGP-ORIGINATE-SO-PTT
  network 10.100.0.1 mask 255.255.255.0 route-map BGP-ORIGINATE-SO-PTT
  network 10.100.0.2 mask 255.255.255.0 route-map BGP-ORIGINATE-SO-PTT
  network 10.100.0.3 mask 255.255.255.0 route-map BGP-ORIGINATE-SO-PTT
!
route-map BGP-ORIGINATE
  set local-preference 500
  set community 100:100
!
route-map BGP-ORIGINATE-SO-PTT
  set local-preference 500
  set community 100:300
!
```

The diagram illustrates the mapping between route-map names and prefix lengths. Blue arrows point from the route-map names in the configuration to their corresponding prefix lengths on the right:

- `BGP-ORIGINATE` (first instance) points to `/22`.
- `BGP-ORIGINATE` (second instance) points to `/23`.
- `BGP-ORIGINATE-SO-PTT` (all three instances) points to `/24`.

# Peering

## PTTMetro RS

```
router bgp 100
  neighbor ptt-rs peer-group
  neighbor ptt-rs remote-as 26162
  neighbor 200.219.143.252 remote-as 20121
  neighbor 200.219.143.253 peer-group ptt-rs
  neighbor 200.219.143.254 peer-group ptt-rs
  address-family ipv4
    neighbor ptt-rs route-map PTT-RS-ATM-IN in
    neighbor ptt-rs route-map PTT-RS-ATM-OUT out
    neighbor ptt-rs maximum-prefix 10000 restart 5
    neighbor 200.219.143.252 send-community
    neighbor 200.219.143.252 allowas-in
    neighbor 200.219.143.252 route-map NO-ROUTES in
  !
  route-map NO-ROUTES deny 10
  !
```



# Filtros

## ATM no PTTMetro RS

- Rejeitar:
  - Rota default
  - Bogons (<http://www.team-cymru.org/Services/Bogons/bogon-bn-nonagg.txt>)
  - Prefixos do nosso próprio AS
- Marcar prefixos recebidos na community 100:20
- Evitar ser usado como trânsito! Pacotes não destinados aos nossos blocos (ou clientes AS):
  - Filtro na interface (ACLs)
  - Unicast RPF (loose mode)  
<http://www.cisco.com/web/about/security/intelligence/unicast-rpf.html>
  - Caixa “default-free” somente para o PTT

# Filtros

## Entrada do PTTMetro RS

```
route-map PTT-RS-ATM-IN deny 10
  description Não aceita rota default pelo PTT-RS
  match ip address prefix-list DEFAULT-ROUTE
```

!

```
route-map PTT-RS-ATM-IN deny 20
  description Não aceita bogons (CYMRU)
  match ip address prefix-list CYMRU-BOGONS
```

!

```
route-map PTT-RS-ATM-IN deny 30
  description Não aceita nossos prefixos
  match ip address prefix-list AS100-CIDR
```

!

```
route-map PTT-RS-ATM-IN permit 40
  description Localpref 300 e marca com 100:20
  set local-preference 300
  set community 100:20
```

!

ip prefix-list DEFAULT-ROUTE permit 0.0.0.0/0

ip prefix-list CYMRU-BOGONS permit 0.0.0.0/8 1e 32  
ip prefix-list CYMRU-BOGONS permit 5.0.0.0/8 1e 32  
ip prefix-list CYMRU-BOGONS permit 10.0.0.0/8 1e 32  
ip prefix-list CYMRU-BOGONS permit 23.0.0.0/8 1e 32  
...

ip prefix-list AS100-CIDR permit 10.100.0.0/22 1e 32

# Filtros

## Saída do PTTMetro RS

```
!  
route-map PTT-RS-ATM-OUT deny 10  
  description Não anuncia prefixos nessas communities  
  match community 100:200  
!  
route-map PTT-RS-ATM-OUT permit 50  
  description Anuncia tudo com community 100:100 e/ou 100:300  
  match community 100:100 100:300  
!
```

# Filtros

## Trânsito pelo PTTMetro via AS400

- Rejeitar:
  - Rota default (opcional)
    - Se aceitar é recomendado marcar como local-AS
    - Mesmo com full-route ajuda nos tempos de convergência
    - Em partial-route pode ser melhor que default estática
  - Bogons
  - Prefixos do nosso próprio AS
- Marcar prefixos recebidos na community 100:10

# Filtros


## Entrada do AS400 via PTT

```
route-map PTT-RS-AS400-IN deny 10
  match ip address prefix-list DEFAULT-ROUTE
!
route-map PTT-RS-AS400-IN deny 20
  match ip address prefix-list CYMRU-BOGONS
!
route-map PTT-RS-AS400-IN deny 30
  match ip address prefix-list AS100-CIDR
!
route-map PTT-RS-AS400-IN permit 40
  description Localpref 200 e marca com 100:10
  set local-preference 200
  set community 100:10
!
```

# Filtros

## Saída do AS400 via PTT

```
route-map PTT-RS-AS400-OUT deny 10
  description Não anuncia prefixos nessas communities
  match community 100:300
!
route-map PTT-RS-AS400-OUT permit 20
  description Anuncia tudo com community 100:100 e/ou 100:200,
  description todo CIDR e a segunda metade mais específica
  match ip address prefix-list AS400-OUT
  match community 100:100 100:200
!
```



```
ip prefix-list AS400-OUT permit 10.100.0.0/22
ip prefix-list AS400-OUT permit 10.100.0.2/23
```

# Filtros

## Trânsito para Cliente AS

- Aceitar somente:
  - Prefixos do CIDR do cliente ( $\geq /24$ )
  - Prefixos com o AS\_path do cliente
- Marcar prefixos recebidos na community 100:30 e 100:100 para permitir anúncio em todos os peerings
- Política de communities pode ser estendida aos clientes:
  - Menos trabalho para o NOC!
  - Maior autonomia dos clientes (blackholes, etc)

# Filtros

## Entrada do Cliente AS

```
route-map CLIENTE-AS-IN permit 10
description Localpref 400, marca como cliente (100:30) e
description libera para trânsito por todos os peerings
set local-preference 400
set community 100:30 100:100
```

!

```
router bgp 100
neighbor xxx.xxx.xxx.xxx remote-as 200
address-family ipv4
neighbor xxx.xxx.xxx.xxx prefix-list AS200-CIDR in
neighbor xxx.xxx.xxx.xxx filter-list 10 in
neighbor xxx.xxx.xxx.xxx route-map CLIENTE-AS-IN in
neighbor xxx.xxx.xxx.xxx route-map CLIENTE-AS-OUT out
```

!

```
ip prefix-list AS200-CIDR permit 10.200.0.0/22 le 24
```

```
ip as-path access-list 10 permit ^(200_)+$
```



# Filtros

## Saída do Cliente AS

```
route-map CLIENTE-AS-OUT permit 10
description Manda prefixos de trânsito (100:10), do PTT (100:20),
description de outros clientes AS (100:30) e do nosso AS (100:100)
match community 100:10 100:20 100:30 100:100
!
```

# Filtros

## Trânsito pelo o ISP

- Rejeitar:
  - Rota default (opcional)
  - Bogons
  - Prefixos do nosso próprio AS
- Marcar prefixos recebidos na community 100:10

# Filtros


## Entrada do ISP

```
route-map ISP-IN deny 10
  match ip address prefix-list DEFAULT-ROUTE
!
route-map ISP-IN deny 20
  match ip address prefix-list CYMRU-BOGONS
!
route-map ISP-IN deny 30
  match ip address prefix-list AS100-CIDR
!
route-map ISP-IN permit 40
  description Localpref 100 e marca com 100:10
  set local-preference 100
  set community 100:10
!
```

# Filtros

## Saída do ISP

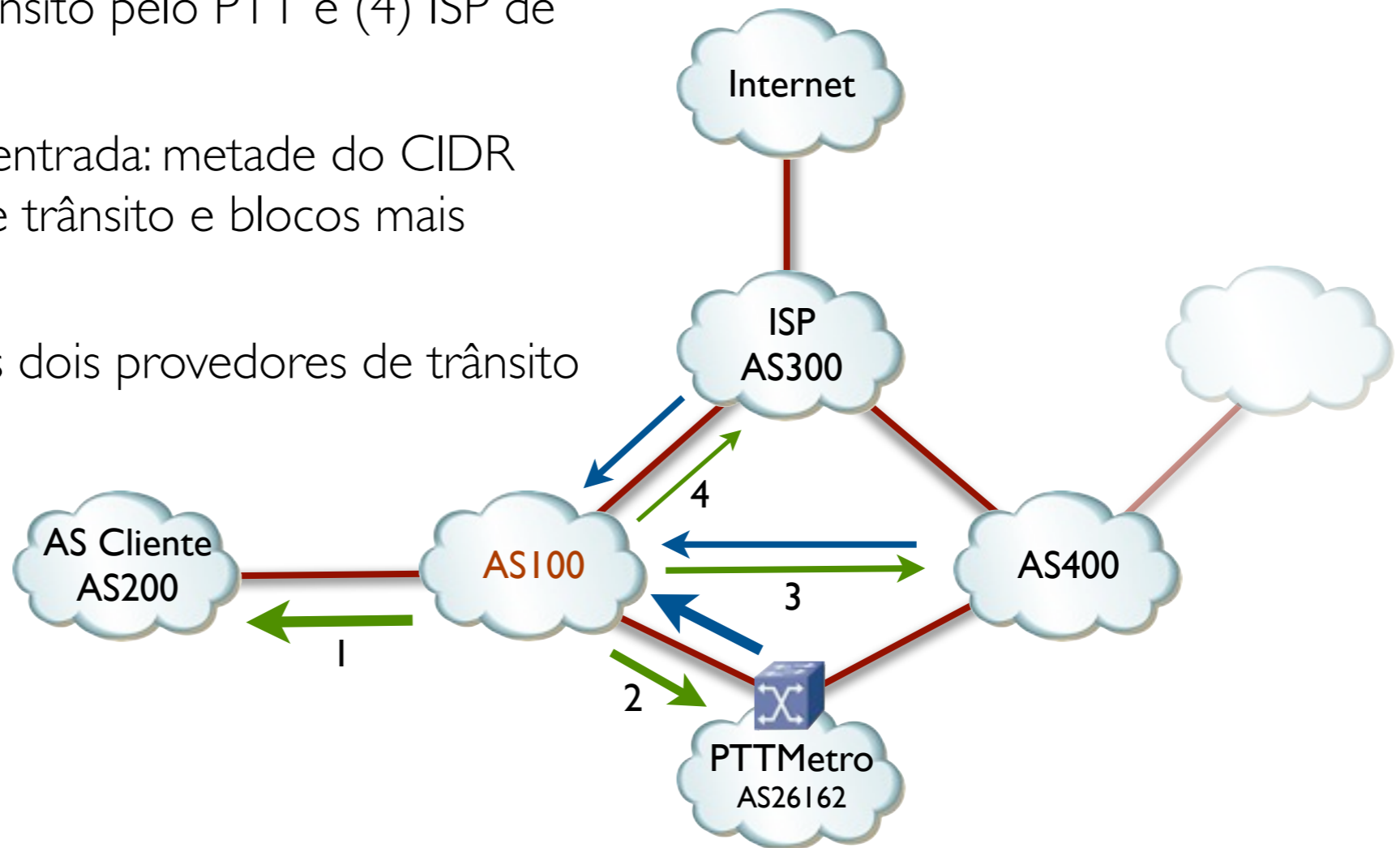
```
route-map ISP-OUT deny 10
  description Não anuncia prefixos nessas communities
  match community 100:300
!
route-map ISP-OUT permit 20
  description Anuncia tudo com community 100:100 e/ou 100:200,
  description todo CIDR e a primeira metade mais específica
  match ip address prefix-list ISP-OUT
  match community 100:100 100:200
!
```



```
ip prefix-list ISP-OUT permit 10.100.0.0/22
ip prefix-list ISP-OUT permit 10.100.0.0/23
```

# Política final

- Ordem de preferência de saída: (1) Clientes AS, (2) PTTMetro, (3) Trânsito pelo PTT e (4) ISP de trânsito
- “Balanceamento” de entrada: metade do CIDR por cada provedor de trânsito e blocos mais específicos pelo PTT
- Redundância entre os dois provedores de trânsito



# Considerações finais

- Casos especiais
  - Clientes AS que também fazem parte do PTT
  - Peering em múltiplos PTTs
  - Peering IPv6
  - ASN 32 bits
  - Granularidade de anúncios para novos ASs
- Participe do looking-glass!
- Peering DB  
<https://www.peeringdb.com/>
- Participe do PTT Fórum!  
<http://pttforum.nic.br/>

# Perguntas?

[lbalbinot@commcorp.com.br](mailto:lbalbinot@commcorp.com.br)