

ASINVS ASINVM FRICAT

– ou –

**Configuração de um acordo bilateral
num PTT para backup mútuo**

Danton Nunes

danton.nunes@inexo.com.br

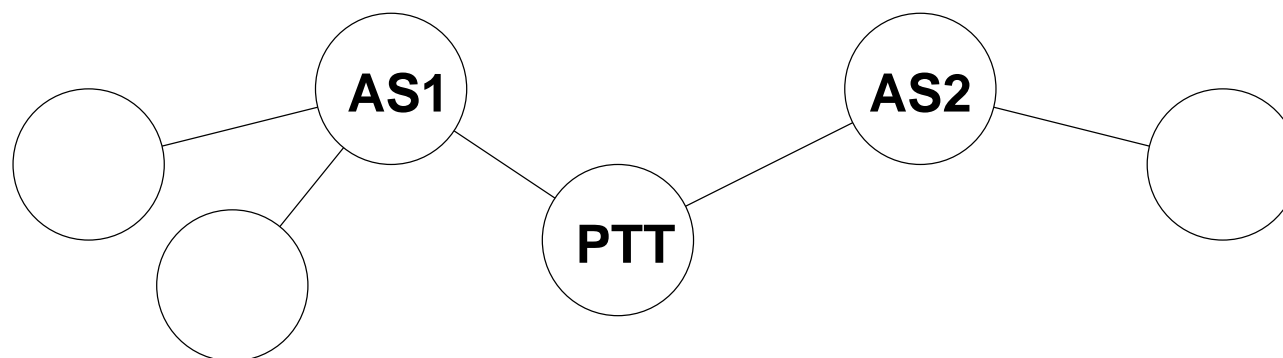
O problema

Duas empresas tem suas redes ligadas a um PTT e tem provedores diferentes.

Um dia seus diretores se encontram no café e tem uma ideia luminosa: "Ei, através do PTT eu poderia te prover trânsito caso teu acesso regular falhe e vice-versa!"



"Redundância a custo zero!" logo entendeu o outro.



Porém há algumas condições...

O arranjo é só para emergências, o trânsito entre os dois ASs deve ser minimizado em condições normais de funcionamento.

Os dois participantes recebem tabelas diferentes de seus 'upstreams', um recebe tabela completa e outro parcial com rota default.

O esquema só servirá IPv4 pois um dos participantes ainda não tem ou não ativou seu bloco de IPv6, mas deve ser facilmente estendido para IPv6.

O tráfego deste arranjo será segregado da troca regular feita dentro do ATM do PTT, ou seja, será estabelecido um acordo bilateral com VLAN própria para conter este tráfego.

Para complicar um pouco

Os participantes não devem saber a priori quais prefixos serão anunciados pelo outro.

Os participantes não devem anunciar prefixos recebidos de seus 'upstreams' ou dos servidores de rotas do PTT, somente de seus clientes.

Para facilitar

Os dois lados usam quagga!



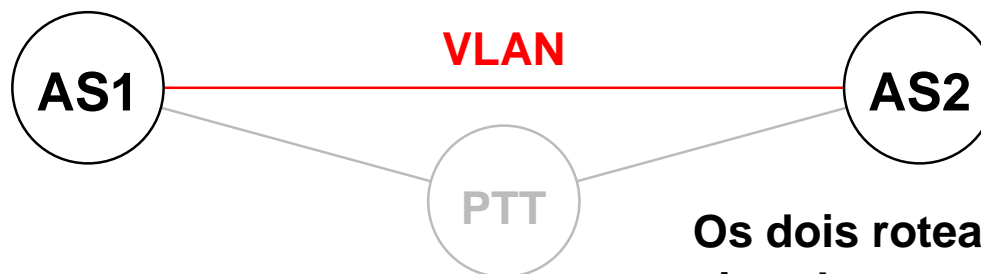
O arranjo

VLAN realizada via PTT entre os dois roteadores.

Uso de 'community' para marcar as rotas recebidas do outro lado.

Uso de 'prepend' para fazer com que o enlace de backup pareça ser distante para terceiros.

Monitoramento do tráfego na VLAN por meio do bom e velho MRTG.



Os dois roteadores são virtualmente adjacentes.

Exemplo

Um Lado

ASN 64496

bloco IPv4: 198.51.100.0/24

Outro Lado

ASN 64511

bloco: 203.0.113.0/24

VLAN

Os endereços da VPN vem de um bloco /30 do ASN 64511.
(poderiam ser usados endereços BCP-5 ou locais)

203.0.113.253/30

203.0.113.254/30

VLAN id: 1001

NOTA: estes endereços não são reais!

Criando as interfaces (Linux)

```
# /sbin/ip link set link eth1 name gambi id 1001
# /sbin/ip link set gambi up
# /sbin/ip addr add 203.0.113.253/30 dev gambi

# /sbin/ip link set link eth0 name bkp id 1001
# /sbin/ip link set bkp up
# /sbin/ip addr add 203.0.113.254/30 dev bkp
```

os dois
devem usar
o mesmo n^o!

O outro lado está vivo?

```
# ping -c 3 203.0.113.254
PING 203.0.113.254 (203.0.113.254): 56 data bytes
64 bytes from 203.0.113.254: icmp=0 loss=(0%, 0 pkts) ttl=64 time=0.3 ms
64 bytes from 203.0.113.254: icmp=1 loss=(0%, 0 pkts) ttl=64 time=0.5 ms
64 bytes from 203.0.113.254: icmp=2 loss=(0%, 0 pkts) ttl=64 time=0.3 ms

--- 203.0.113.254 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
```

indica adjacência

```
# ping -c 3 203.0.113.253
.....
```

bgpd.conf (só um lado, o outro fica como exercício para o leitor)

```
router bgp 64496
!....
neighbor 203.0.113.254 remote-as 64511
neighbor 203.0.113.254 description exemplo para mostrar no GTER-32
neighbor 203.0.113.254 soft-reconfiguration inbound
neighbor 203.0.113.254 route-map tobkpeer out
neighbor 203.0.113.254 route-map frombkpeer in
!....
```

Marca as rotas recebidas para poder exportá-las

Limita quais prefixos serão anunciados

bgpd.conf (continuação)

```
neighbor 192.0.2.79 remote-as 64500  
neighbor 192.0.2.79 description Provedor genérico a montante  
neighbor 192.0.2.79 soft-reconfiguration inbound  
neighbor 192.0.2.79 route-map toUpstream out
```

!.....

Aqui controlamos quais prefixos anunciamos para montante, os nossos mais os recebidos do parceiro.

```
route-map frombkeeper permit 10  
set community 64511:100
```

A ideia aqui é marcar todos os prefixos recebidos do parceiro para que possam ser exportados para montante.

```
ip community list standard partner permit 64511:100  
!...
```

```
route-map toUpstream permit 50  
match community partner  
set as-path prepend 64496 64496 ..
```

prepends para fazer parecer distante

funcionou?

```
aqui> show ip bgp summary  
BGP router identifier 203.0.113.253, local AS number 64496  
6988 BGP AS-PATH entries  
166 BGP community entries
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
...									
203.0.113.254	4	64511	21635	50985	0	0	0	01w5d20h	1
...									

olha nosso vizinho aqui!

e está funcionando!

e nos mandou um prefixo!

```
aqui> show ip bgp neighbor 203.0.113.254 routes
```

blá blá blá

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 203.0.113.0/24	203.0.113.254	0	100	0	64511 i

```
Total number of prefixes 1
```

funcionou?

```
aqui> show ip bgp neighbor 203.0.113.254 advertised-routes
```

```
... blá blá blá ... patati-patatá ...
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 198.51.100.0/24	203.0.113.254	0		32768	i

```
Total number of prefixes 1
```

oba! estamos anunciando nosso prefixo para o parceiro.

```
aqui> show ip bgp neighbor 192.0.2.79 advertised-routes
```

```
... blá blá blá ... patati-patatá ...
```

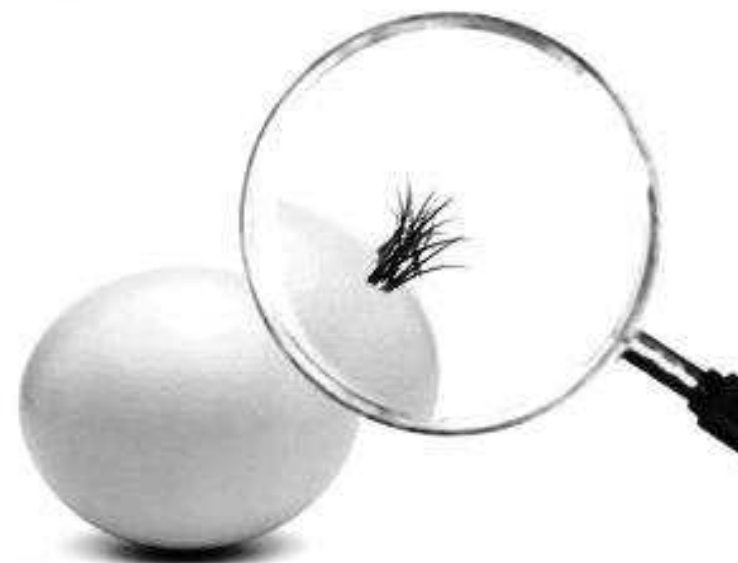
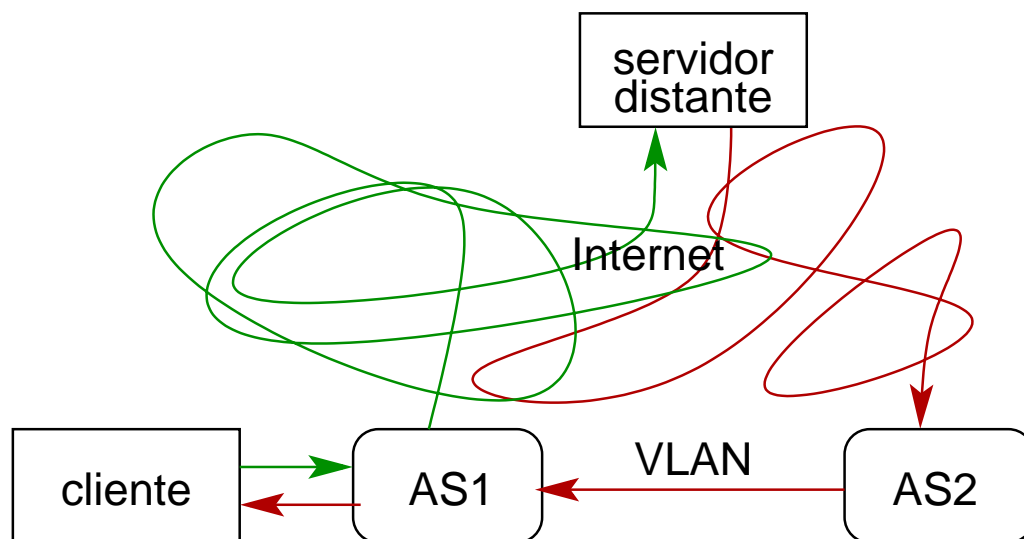
Network	Next Hop	Metric	LocPrf	Weight	Path
*> 198.51.100.0/24	192.0.2.79	0		32768	i
*> 203.0.113.254/24	192.0.2.79		100	0	64496 64496 64511 i

estamos anunciando nosso prefixo e também o do parceiro para o mundo!

mas o do parceiro sai com preprends!

Pequenas assimetrias, grandes problemas.

Caminhos diferentes para ida e volta



Caminho assimétrico pode ocorrer se o cliente parecer, para o servidor, mais próximo pelo caminho alternativo do que pelo regular (apesar dos preprends).

Isto pode ser problemático se o AS2 tem interfaces com MTU menor que as do AS1. O PMTU é estabelecido somente pelo caminho de ida, basta que o servidor distante ignore ou não leve em conta o ICMP gerado pelo AS2, o desastre estará armado.

Pequenas assimetrias, grandes problemas.

Tabela completa X parcial

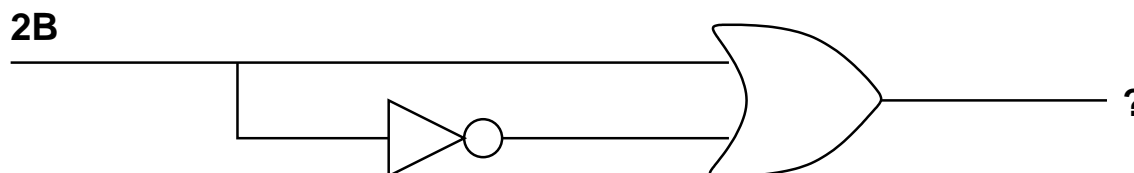
Um dos parceiros recebe tabela completa, o outro tabela parcial com rota default.

Anunciar a tabela completa para o outro lado faria com que todo o tráfego para alguns destinos fosse canalizado para a VLAN, o que não é desejado.

A abordagem adotada foi:

- Não anunciar para o outro as rotas recebidas de montante;
- Usar rotas estáticas de último recurso com métrica elevada na tabela default.

```
# /sbin/ip route add default via 203.0.113.254 dev gambi metric 100 table default
```



Conclusões

Um acordo bilateral no PTT pode ser usado para backup mútuo entre pequenos, não só para um grande vender trânsito para um pequeno.

A configuração do BGP é relativamente simples, tomando alguns cuidados com MTU, 'prepends' e preferência local.

O esquema proposto funciona para um acordo gratuito, mas poderia ser remunerado na base da diferença de fluxo, o que é facilmente mensurado com o SNMP (MRTG/RRDTool).

Se você tem um AS, este é (mais) um bom motivo para participar do PTT!

Agradecimentos

Fábio Aleixo Santos, DirectLink, SJC
PTT/Metro SJC

Perguntas?

