



HEXA

NETWORKS

Routing and Internet **Experts.**

Chega de rede hambúrguer!
Tenha escalabilidade com BGP
Free Core e 6PE!

André Dias

- Fundador e CEO da **Hexa Networks** e **NetConfig**;
- Mais de **14** anos atuando na área de **redes**;
- Mais de **22** anos atuando na área de **tecnologia** (iniciando aos 8 anos de idade);
- Aos **16** anos iniciou em um provedor de VoIP (Telefonia sobre IP);
- Aos **17** ingressou em um ISP (Provedor de Acesso a Internet);
- Certificado **MTCNA, MTCRE, MTCINE, MTCIPv6E, JNCIA-JUNOS, JNCIA-DC, JNCIS-SP, NRS-I** ;
- Autor de artigos para o "**Brasil Peering Fórum**";
- Fanático assíduo por **IPv6**;
- Já fez parte da maior orquestra do planeta;
- Sou chato e teimoso;
- Colaborador para a maior obra coletiva da humanidade, a Internet.



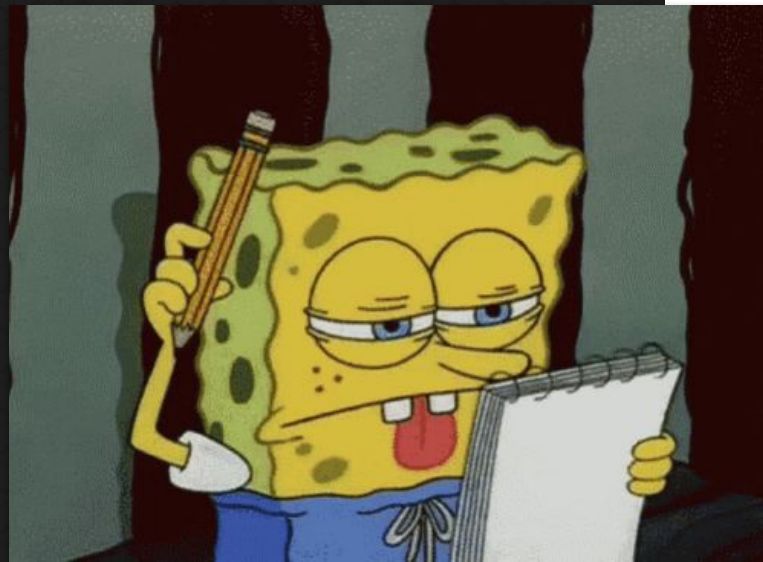
Agenda:

- Motivação;
- Contexto do caso;
- O problema;
- A pseudo-solução (rede hambúrguer);
- Solução proposta
- BGP FREECORE;
- 6PE.



Motivação Para a **Palestra**

- Problema de Escalabilidade;
- Erro no uso generalizado de L2VPN;
- Mostrar que L2VPN é só a ponta do iceberg do MPLS.





As aventuras do Bob Esponja na EsponjaNET

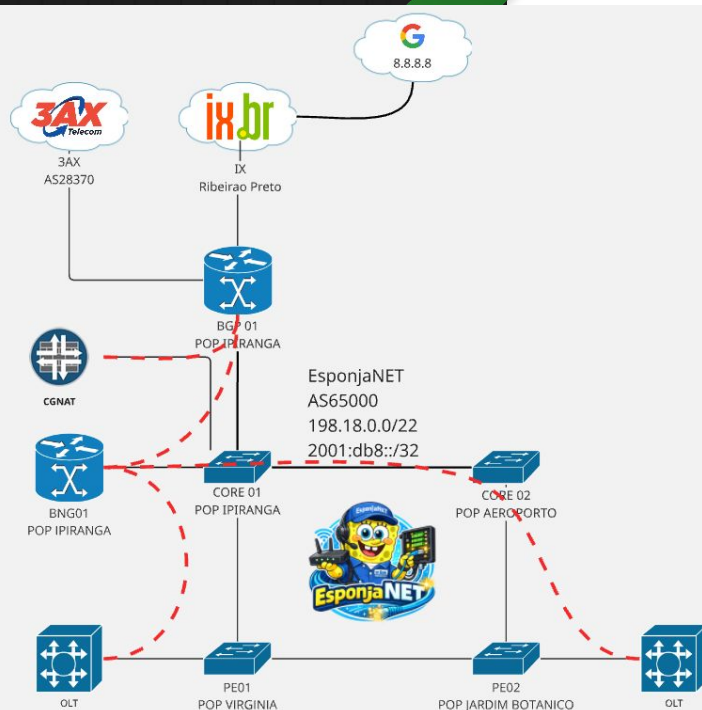
- Bob Esponja tem um provedor de internet!
- Provedor fictício de Ribeirão Preto (sim, mais um provedor em Ribeirão);
- Roteadores de Borda Huawei NE8000;
- BNG/BRAS (Concentrador) Huawei NE8000;
- Malha MPLS mista (Datacom, Huawei e Mikrotik).





As aventuras do Bob Esponja na EsponjaNET

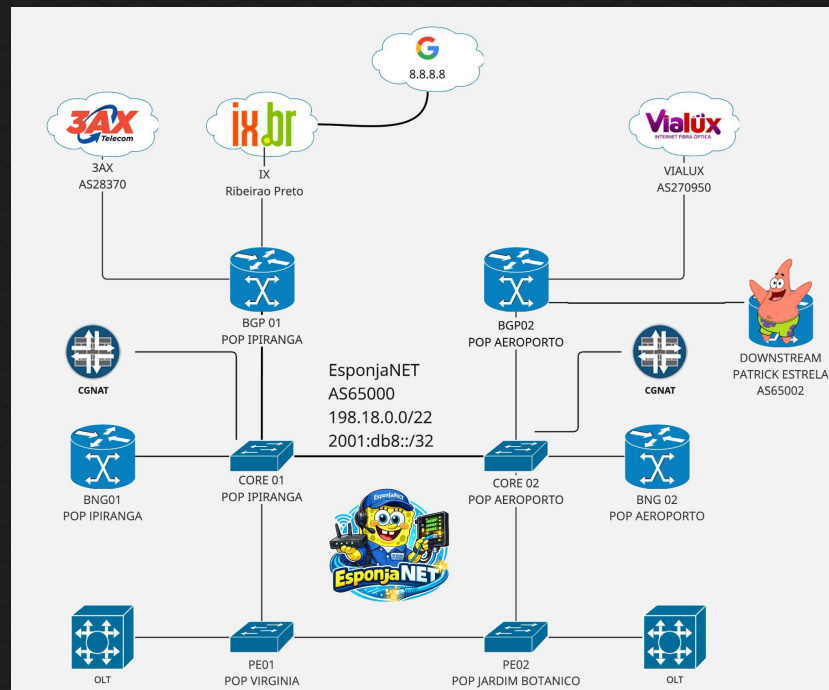
- Provedor começou a crescer, e precisou de escalabilidade, principalmente de redundância ágil no anel;
- Bob Esponja foi apresentado ao mundo do MPLS;
- Aprendeu a transportar clientes PPPoE para os BNG utilizando L2VPN (VPLS), e amou!
- Seus problemas de redundância estavam resolvidos....por enquanto.



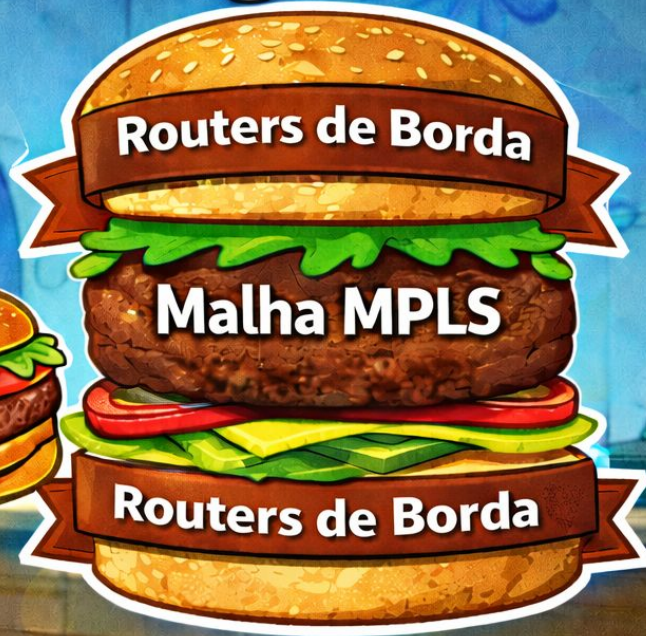
As aventuras do Bob Esponja na **EsponjaNET**

- Bob Esponja resolveu adicionar novo roteador de borda e novo BNG no Aeroporto;
- Contrata um novo Upstream;
- Passa ter um cliente de trânsito!
- E aí que o problema começa...
- Como comunicar esses roteadores?
- Replicar a solução de L2VPN para essa finalidade também?

Senhoras e Senhores...Sejam apresentados ao que ~~nao tao~~ carinhosamente apelidamos de....



Rede Hambúrguer



PÃES = Routers de Borda

HAMBÚRGUER = Malha MPLS



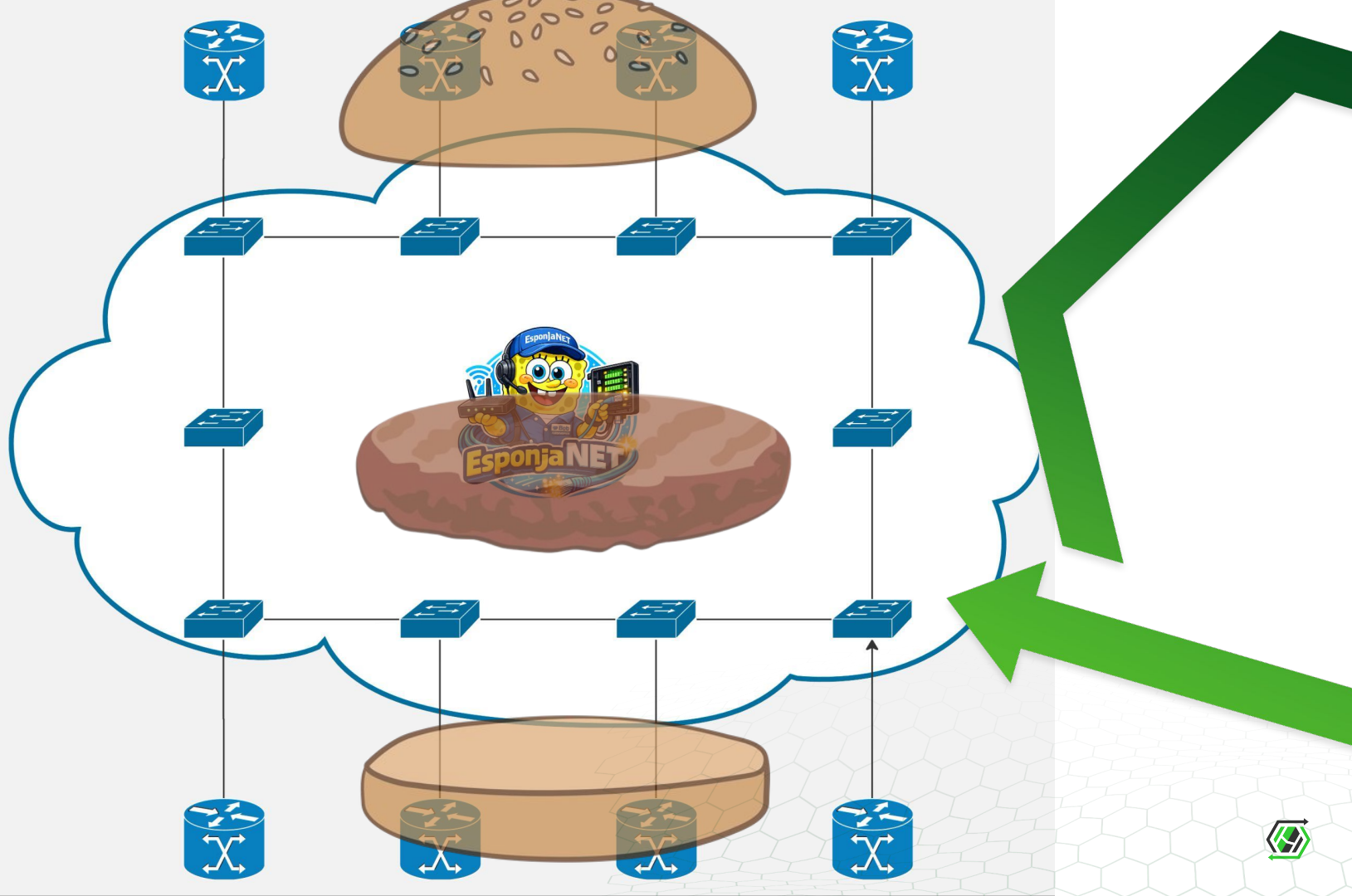
Rede Hambúrguer

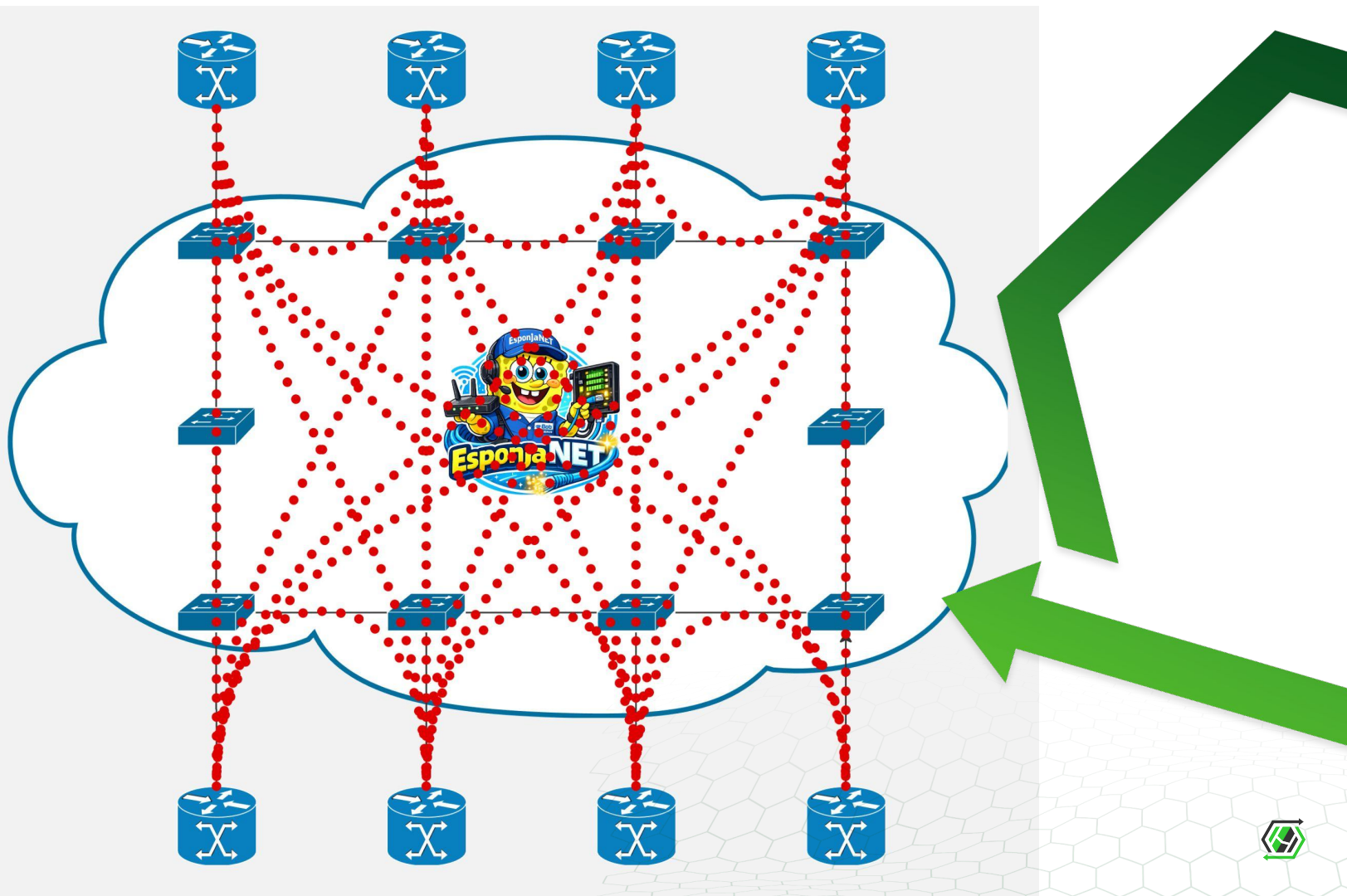


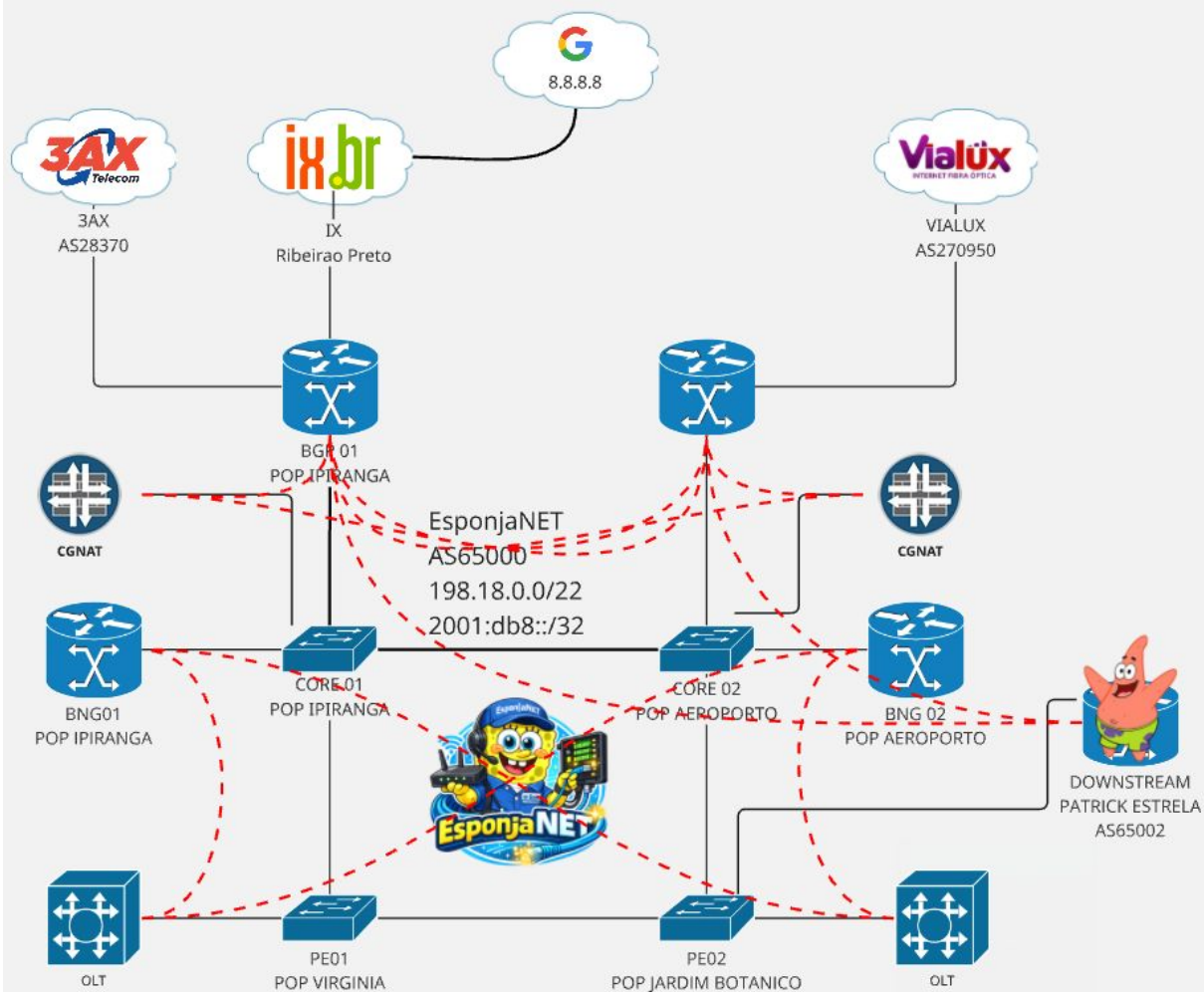
PÃES = Routers de Borda

HAMBÚRGUER = Malha MPLS





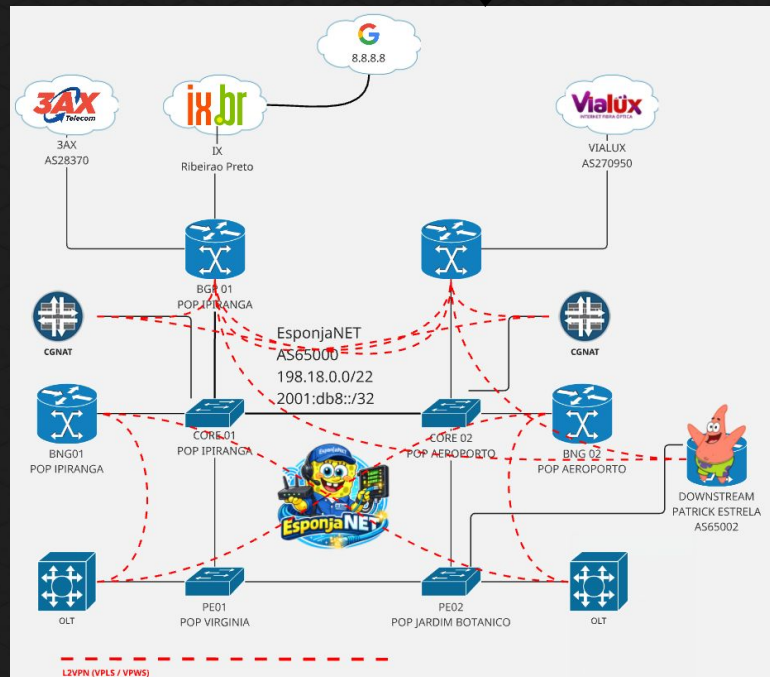






Problemas dessa pseudo-solução

- Utilização generalizada de L2VPN;
- Aprendizado de MACs descontrolado;
- Mais domínios de broadcast (BUUM);
 - Broadcast, Unknown Unicast, Multicast
- Alta possibilidade de erros humanos;
- Complexidade totalmente desnecessária;
- ZERO ESCALABILIDADE.



A Solução: BGP Free Core

Alguns passos simples são necessários para que o BGP Free Core funcione, sendo eles:

- ✓ IGP rodando (OSPF ou IS-IS);
- ✓ Algum protocolo de distribuição de label atuando (usaremos LDP);
- ✓ iBGP entre as bordas;
- **Habilitar suporte a roteamento recursivo nas bordas.**
 - Algumas caixas o fazem por padrão, outras não! Roteadores Nokia e Huawei precisam ter a feature habilitada manualmente, enquanto roteadores Cisco têm isso configurado implicitamente.



BGP FREE CORE

BGP Free Core não é “uma feature” — é um **técnica de roteamento: BGP nas borda + MPLS/IGP** na rede para possibilitar o encaminhamento baseado em labels.

Bob Esponja resolveu testar essa solução! Bora acompanhar ele?



Implementando BGP Free Core

Bora fazer o deploy na **EsponjaNET?**
O Bob é esperto e não é **emocionado** de trocar e implantar na rede toda sem antes testar e validar. Então decide implementar apenas para o **cliente Patrick Estrela**.

-  **IGP** rodando (OSPF ou IS-IS)
-  **Algun protocolo de distribuição de label MPLS** atuando (LDP por exemplo)

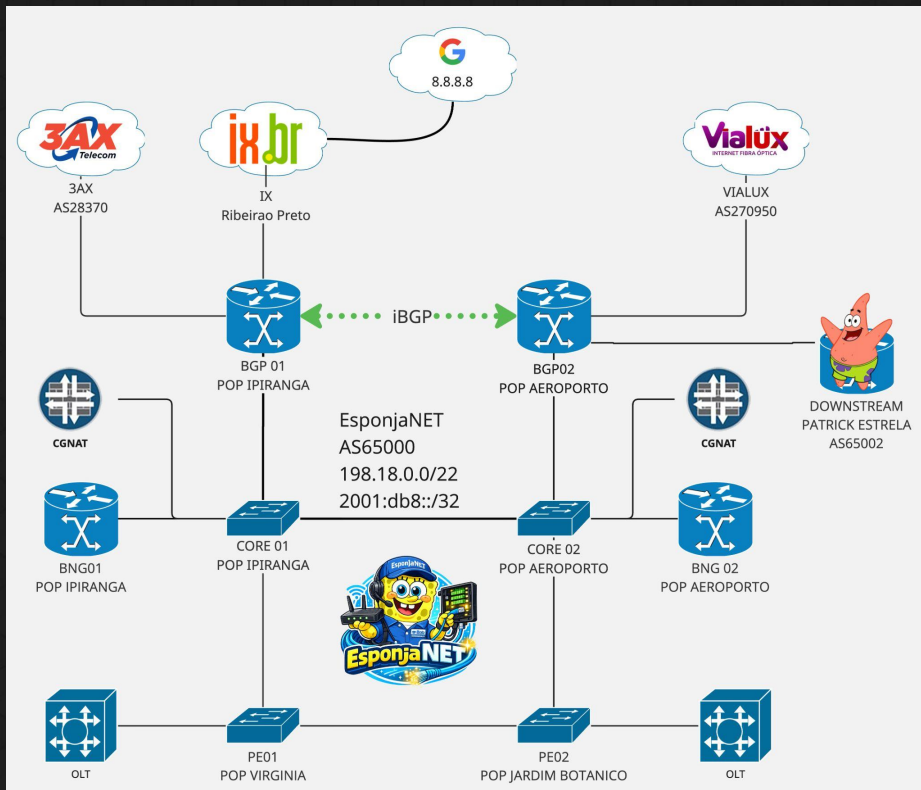
Falta apenas:

- **iBGP entre as Routers;**
- **Habilitar suporte a roteamento recursivo nos PEs/Routers.**



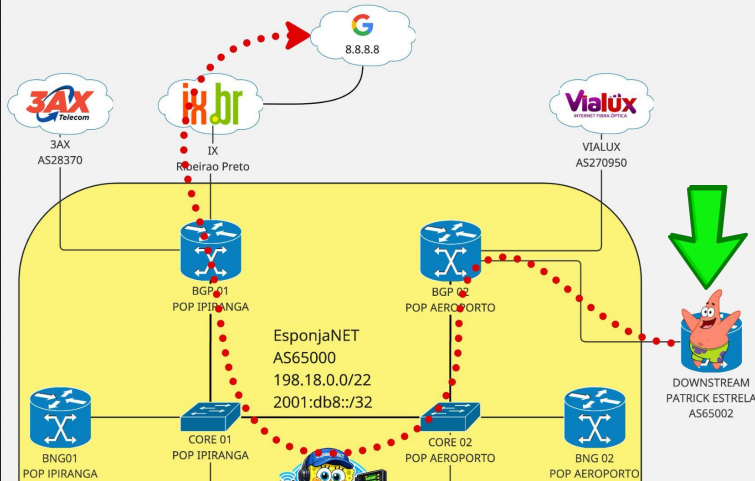
Como comunicar os Routers (BNG/Border)?

- Vamos conectar nossos routers via BGP FREE CORE
- Apenas subir sessão iBGP entre eles resolverá....certo?
- É... Nope. Não vai funcionar. Vamos entender o motivo?



Troubleshoot do Bob:

✗ • Não pinga;



```
[admin@PATRICK ESTRELA] > /ping 8.8.8.8
```

```
SEQ HOST
0 8.8.8.8
1 8.8.8.8
```

```
sent=2 received=0 packet-loss=100%
```

```
SIZE TTL TIME STATUS
timeout
timeout
```

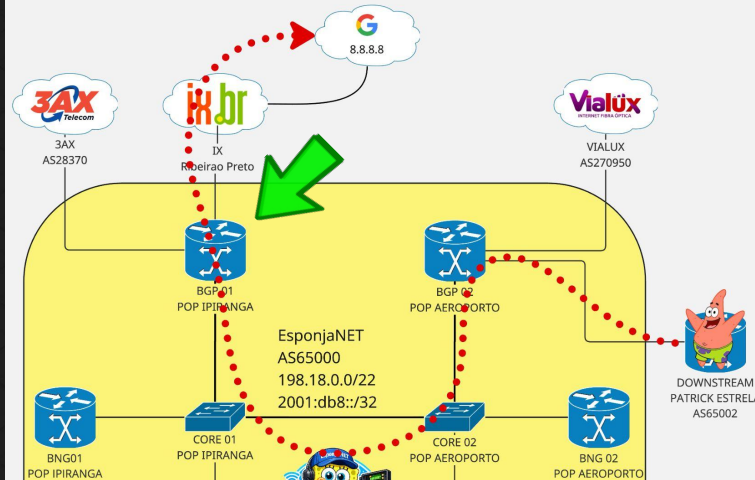
```
[admin@PATRICK ESTRELA] > /tool traceroute 8.8.8.8
```

#	ADDRESS	LOSS	SENT	LAST	AVG	BEST
1	198.18.1.1	0%	1	7.9ms	7.9	7.9
2		100%	1	timeout		
3		100%	1	timeout		
4		100%	1	timeout		
5		100%	1	timeout		
6		0%	1	0ms		

```
[admin@PATRICK ESTRELA] > █
```

Troubleshoot do Bob:

- ✗ ~~Não pinga;~~
- ✓ Sessões eBGP UP;
- ✓ Recebe o 8.8.8.0/24 pelo IX;



```
[ESPONJANET-RT-NE8KM8-BGP01-IPIRANGA-bgp]dis bgp peer
```

```
BGP local router ID : 192.0.2.2
Local AS number : 65000
Total number of peers : 3
```

```
Peers in established state : 3
```

Peer	V	AS	MsgRcvd	MsgSent	OutQ	Up/Down	Sta
172.16.0.4	4	65000	196	240	0	02:42:40	Establish
192.0.2.1	4	270950	273	240	0	03:14:26	Establish
200.192.110.253	4	26162	20	24	0	00:11:07	Establish

```
[ESPONJANET-RT-NE8KM8-BGP01-IPIRANGA-bgp]dis ip routin
```

```
[ESPONJANET-RT-NE8KM8-BGP01-IPIRANGA-bgp]dis ip routing-table 8.8.8.8
```

```
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black
```

```
Routing Table : _public_
```

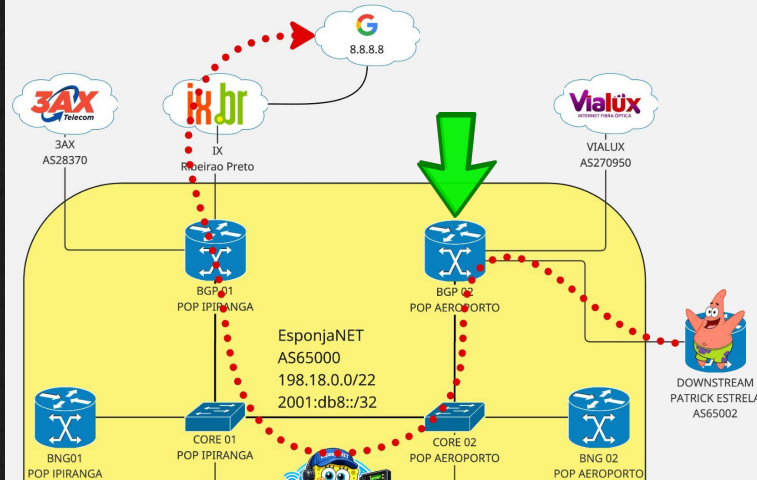
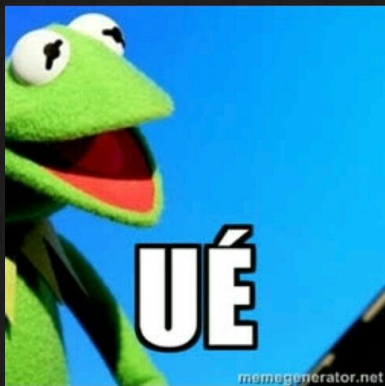
```
Summary Count : 1
```

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Interfa
8.8.8.0/24	EBGP	255	0	RD	200.192.110.88	Etherne

```
[ESPONJANET-RT-NE8KM8-BGP01-IPIRANGA-bgp]
```

Troubleshoot do Bob:

- ✗ ~~Não pinga;~~
- ✓ ~~Sessão eBGP UP;~~
- ✓ ~~Recebe o 8.8.8.0/24 do IX;~~
- ✓ Sessão iBGP UP;
- ✓ Recebe o 8.8.8.0/24 do BGP01;
- ✓ Troca de NEXT_HOP ok.



```
[ESPONJANET-RT-NE8KM8-BGP02-AEROPORTO]dis bgp peer
```

```
BGP local router ID : 172.16.0.4
Local AS number : 65000
Total number of peers : 2
```

```
Peers in established state :
```

Peer	V	AS	MsgRcvd	MsgSent	OutQ	Up/Down
172.16.0.1	4	65000	250	206	0	02:50:11 Establ
192.0.2.1	4	270950	8	13	0	00:04:20 Establ

```
[ESPONJANET-RT-NE8KM8-BGP02-AEROPORTO]dis ip routin
```

```
[ESPONJANET-RT-NE8KM8-BGP02-AEROPORTO]dis ip routing-table 8.8.8.8
```

```
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - b
```

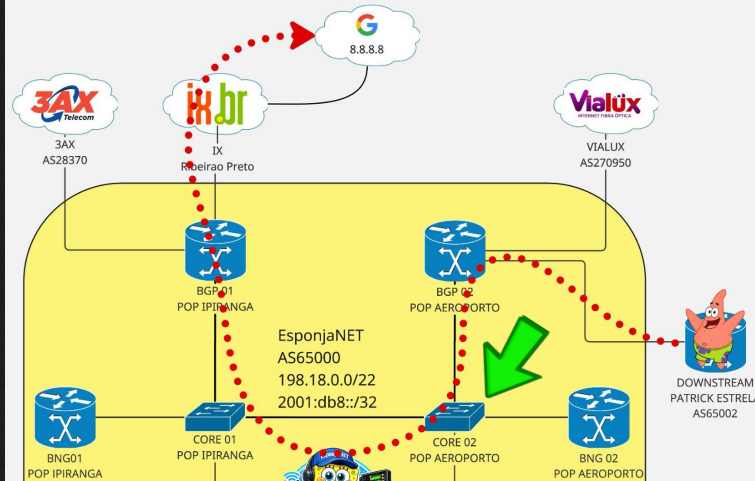
```
Routing Table : _public_
```

```
Summary Count : 1
```

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Inte
8.8.8.0/24	IBGP	255	0	RD	172.16.0.1	Ethe

Troubleshoot do Bob:

- ✗ ~~Não pinga;~~
- ✓ ~~Sessão eBGP UP;~~
- ✓ ~~Recebe o 8.8.8.0/24 do IX;~~
- ✓ ~~Sessão iBGP UP;~~
- ✓ ~~Recebe o 8.8.8.0/24 do BGP01;~~
- ✓ ~~Troca de NEXT_HOP ok.~~
- ✗ Backbone nao tem o 8.8.8.0/24!
- Pacote está sendo descartado!



[ESPONJANET-SW-S6730-CORE02-AEROPORTO-Ethernet1/0/0]dis ip routing-table
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole route

Routing Table : public_
Destinations : 15 Routes : 15

Destination/Mask	Proto	Pre	Cost	Flags	NextHop	Interface
10.1.2.0/24	OSPF	10	2	D	10.2.3.2	Ethernet1/0/1
10.2.3.0/24	Direct	0	0	D	10.2.3.3	Ethernet1/0/1
10.2.3.3/32	Direct	0	0	D	127.0.0.1	Ethernet1/0/1
10.2.3.255/32	Direct	0	0	D	127.0.0.1	Ethernet1/0/1
10.3.4.0/24	Direct	0	0	D	10.3.4.3	Ethernet1/0/0
10.3.4.3/32	Direct	0	0	D	127.0.0.1	Ethernet1/0/0
10.3.4.255/32	Direct	0	0	D	127.0.0.1	Ethernet1/0/0
127.0.0.0/8	Direct	0	0	D	127.0.0.1	InLoopBack0
127.0.0.1/32	Direct	0	0	D	127.0.0.1	InLoopBack0
127.255.255.255/32	Direct	0	0	D	127.0.0.1	InLoopBack0
172.16.0.1/32	OSPF	10	2	D	10.2.3.2	Ethernet1/0/1
172.16.0.2/32	OSPF	10	1	D	10.2.3.2	Ethernet1/0/1
172.16.0.3/32	Direct	0	0	D	127.0.0.1	LoopBack0
172.16.0.4/32	OSPF	10	1	D	10.3.4.4	Ethernet1/0/0
255.255.255.255/32	Direct	0	0	D	127.0.0.1	InLoopBack0

[ESPONJANET-SW-S6730-CORE02-AEROPORTO-Ethernet1/0/0]dis ip routing-table 8.8.8.8

[ESPONJANET-SW-S6730-CORE02-AEROPORTO-Ethernet1/0/0]

Recursão de LSP

- Temos IGP;
- Temos Labels;
- Mas...Nada de ping! O que falta?

```
[ESPONJANET-RT-NE8KM8-BGP02-AEROPORTO]dis ip routing-table 8.8.8.0 24 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole route
-----
Routing Table : _public_
Summary Count : 1

Destination: 8.8.8.0/24
  Protocol: IBGP                                Process ID: 0
  Preference: 255                               Cost: 0
  NextHop: 172.16.0.1                           Neighbour: 172.16.0.1
    State: Active Adv Relied                     Age: 01h00m28s
    Tag: 0                                       Priority: low
    Label: NULL                                QoSInfo: 0x0
  IndirectID: 0x100007B                         Instance:
  RelayNextHop: 10.3.4.3                       Interface: Ethernet1/0/0
    TunnelID: 0x0                               Flags: RD
[ESPONJANET-RT-NE8KM8-BGP02-AEROPORTO]
```

Recursão de LSP

- Recursão!
- **route recursive-lookup tunnel**
- Permite que o pacote consiga recursar por um LSP (Label Switched Path) no backbone - encaminhamento baseado em labels
- Pode colocar a vontade nas caixas, zero downtime!

```
[ESPONJANET RT NE8KM8 BGP02 AEROPORTO]dis ip routing table 8.8.8.0 24 verbose
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole route
-----
Routing Table : _public_
Summary Count : 1

Destination: 8.8.8.0/24
  Protocol: IBGP                Process ID: 0
  Preference: 255                Cost: 0
  NextHop: 172.16.0.1            Neighbour: 172.16.0.1
  State: Active Adv Relied       Age: 00h00m02s
  Tag: 0                          Priority: low
  Label: NULL                     QoSInfo: 0x0
  IndirectID: 0x100007F           Instance:
  RelayNextHop: 10.3.4.3          Interface: Ethernet1/0/0
  TunnelID: 0x0000000001004c4b42 Flags: RD
[ESPONJANET RT NE8KM8 BGP02 AEROPORTO]
```





Isso é **mágica**?

O ping passou a ser possível devido ao conceito de encaminhamento baseado em labels!

Sem labels, o PE não conseguiria se comunicar diretamente com o DNS do Google!

Quando temos labels no backbone, os equipamentos não mais encaminham com base na FIB e sim com base na LFIB. (tabela de labels)!

```
[admin@PATRICK ESTRELA] > ping 8.8.8.8
```

SEQ	HOST	SIZE	TTL	TIME	STA
0	8.8.8.8	56	60	14ms	
1	8.8.8.8	56	60	8ms	

```
sent=2 received=2 packet-loss=0% min-rtt=8ms avg-rtt=11ms max
```

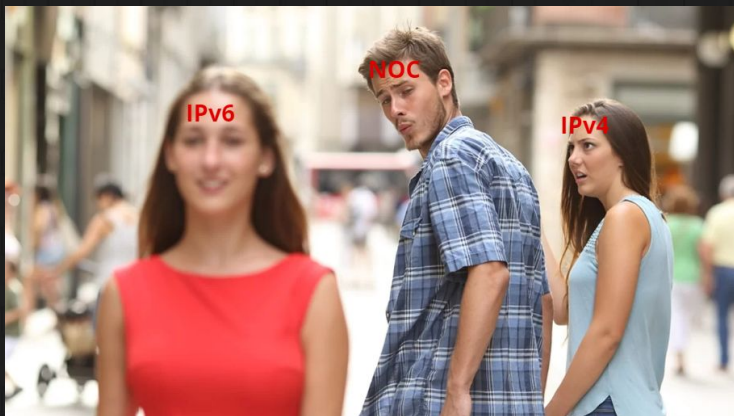
```
[admin@PATRICK ESTRELA] > /tool traceroute 8.8.8.8
```

#	ADDRESS	LOSS	SENT	LAST	AVG
1	198.18.1.1	0%	3	1.7ms	4
2		100%	3	timeout	
3		100%	2	timeout	
4	10.1.2.1	0%	2	6.5ms	11.9
5	8.8.8.8	0%	2	6.7ms	6.5

Legal esse IPv4, mas e o IPv6?

O LDP só possui suporte para gerar labels para rotas IPv4, não possuindo suporte para gerar labels para IPv6 (LDPv6 não teve sucesso).

Por esse motivo, precisamos usar outras técnicas para que possamos transportar rotas IPv6 por dentro de uma malha MPLS!



O famigerado 6PE!

O 6PE consiste em usar o MP-BGP (AFI 2, SAFI 4 - IPv6 Labeled Unicast) para transportar as rotas IPv6 através do backbone IPv4-only. Ou seja, apesar de usar AFI 2 (IPv6), a **sessão BGPv6 é estabelecida através do IPv4!**

Para viabilizar tal implementação é necessário **ativar o suporte à labels MPLS no peer BGP**, permitindo a negociação e interpretação entre eles. Obviamente, **o equipamento precisa ter suporte à esta feature!**

Como a sessão é estabelecida através do IPv4, porém usando AFI 2, o NEXT_HOP dessas rotas será bem diferente do convencional (IPv4-mapped IPv6 - RFC 4291):

- ::FFFF:192.0.2.54
- ::FFFF:203.0.113.34



Config de 6PE

Habilitar sessões iBGP IPv6 com os refletos utilizando o backbone IPv4:

```
bgp 65000
#
ipv6-family unicast
peer 198.18.0.255 enable
peer 198.18.0.255 label-route-capability
peer 198.18.2.255 enable
peer 198.18.2.255 label-route-capability
```

Config de 6PE

Verificar se a rota DEFAULT está sendo aprendida no PE02:

```
display bgp ipv6 routing-table 2001:4860:4860::8888
```

Repare no atributo **NEXT_HOP** da rota! Reconhecem a parte final do endereço?!

```
BGP local router ID : 172.16.0.3
Local AS number : 65000
Paths: 2 available, 1 best, 1 select, 0 best-external, 0 add-path
BGP routing table entry information of ::/0:
Label information (Received/Applied): 48030/NULL
From: 198.18.0.255 (198.18.0.255)
Route Duration: 0d00h03m42s
Relay IP Nexthop: A03:603::
Relay IP Out-Interface: Ethernet1/0/1
Relay Tunnel Out-Interface: Ethernet1/0/1
Original nexthop: ::FFFF:198.18.0.255
Qos information : 0x0
AS-path Nil, origin igp, MED 0, localpref 100, pref-val 0, valid, internal, best, select, pre 255, IGP cost 2000
Not advertised to any peer yet
```

Config de 6PE

ping ipv6 -a 2001:db8:faca:: 2001:4860:4860::8888

```
PING 2001:4860:4860::8888 : 56 data bytes, press CTRL_C to break
  Reply from 2001:4860:4860::8888
    bytes=56 Sequence=1 hop limit=60 time=15 ms
  Reply from 2001:4860:4860::8888
    bytes=56 Sequence=2 hop limit=60 time=10 ms
  Reply from 2001:4860:4860::8888
    bytes=56 Sequence=3 hop limit=60 time=9 ms
  Reply from 2001:4860:4860::8888
    bytes=56 Sequence=4 hop limit=60 time=13 ms
  Reply from 2001:4860:4860::8888
    bytes=56 Sequence=5 hop limit=60 time=10 ms

--- 2001:4860:4860::8888 ping statistics---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max=9/11/15 ms
```

tracert ipv6 -a 2001:db8:faca:: 2001:4860:4860::8888

```
tracert to 2001:4860:4860::8888 30 hops max,60 bytes packet
 1 ::FFFF:10.3.6.3 7 ms 3 ms 4 ms
 2 2804:248:CAFE:1::2 9 ms 3 ms 3 ms
 3 2804:248:CAFE:1::1 5 ms 5 ms 4 ms
 4 2804:5330:FACA::1 7 ms 5 ms 7 ms
 5 2600:1900:CAFE::1 16 ms 6 ms 8 ms
 6 2001:4860:4860::8888 7 ms 8 ms 8 ms
```



Conclusão do Bob!

- Escalabilidade - Próximas ativações **não causam impacto**
- Menos complexidade, gambiarras e riscos
- Maior **resiliência**
- Troubleshooting **facilitado**
- Deixe os hambúrgueres no Siri Cascudo, e utilize as soluções do MPLS das formas adequadas.



Melhorias possíveis!

- **Route Reflectors** (Inline e Outline)
- VPNs controladas por **Communities**
- **Serviços dedicados** com uso de LSPs (RSVP-TE)
- **Ocultar o backbone** através de não propagar ttl do pacote
- Engenharia de tráfego com maior **eficiência**

Meus **Contatos e links:**

E-mail: andredias@hexanetworks.com.br

Linkedin:

<https://www.linkedin.com/in/andreIrdias/>

Telefone: +55 (17) 99670-0482



Dúvidas ou Sugestões?

