



**WTR**2015  
II WORKSHOP DE TECNOLOGIA DE REDES

# DNS Anycast com ExaBGP

Luiz Gustavo Barros

Universidade Estadual de Ponta Grossa

gustavo {at} uepg.br

# Intro



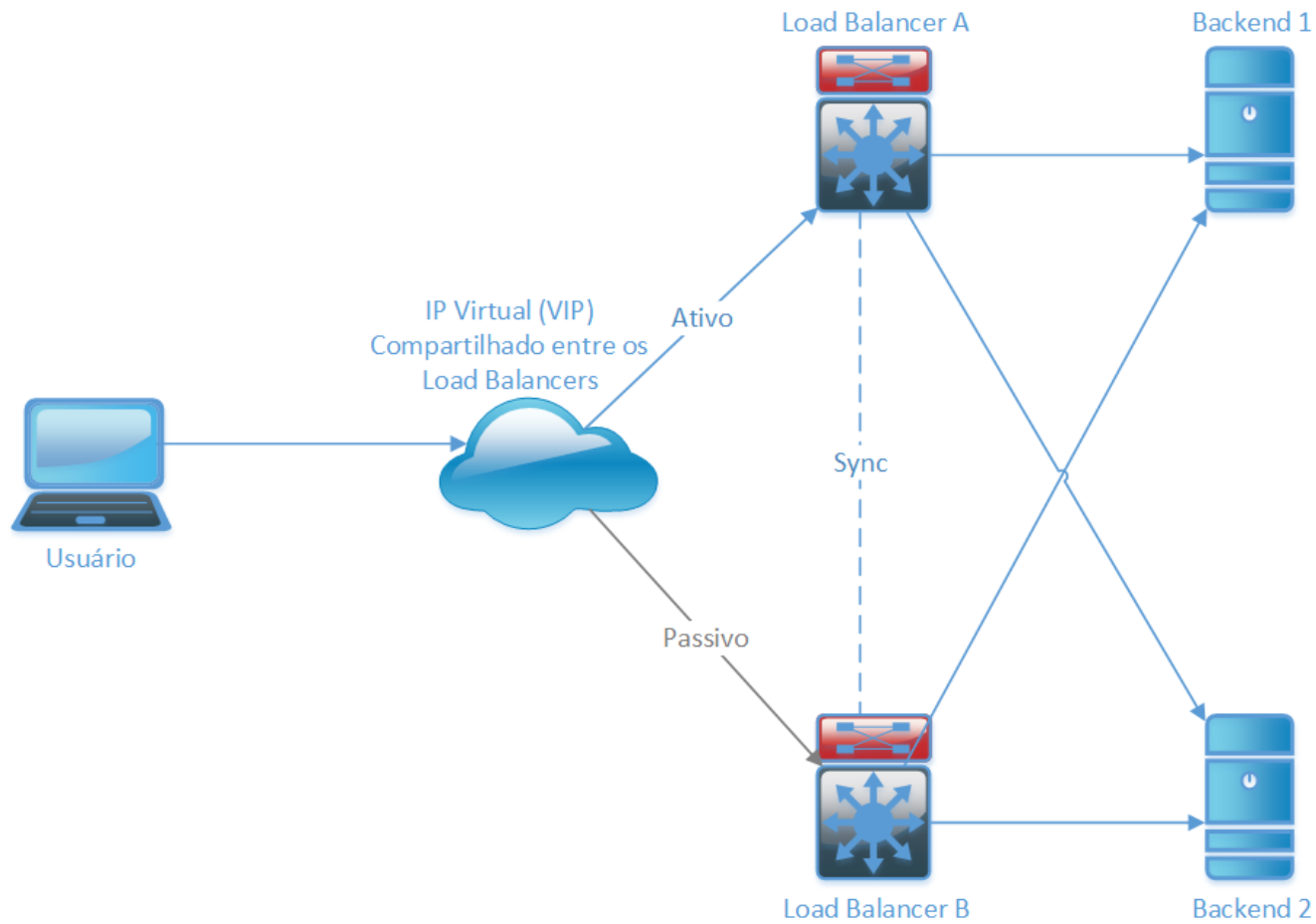
- DNS
- Anycast
  - Encaminhamento ao melhor (mais próximo) destino definido pelo roteamento
  - Mesmo IP anunciado em pontos distintos
    - Unicast: one to one
    - Broadcast: one to all
    - Multicast: one to many
    - **Anycast: one to nearest**

# Cenário Anterior



- Redundância com corosync + ldirectord
- 8 IPs no balanceamento (VIPs)
- Todos os VIPs ficam nos Load Balancers (LB)
  - Corosync mantém somente um LB com os VIPs
  - Ativo/passivo
- Balanceamento em si é feito com ldirectord

# Cenário Anterior



# Motivação



- Não depender de uma única implementação de balanceamento
  - Pode ocorrer o mesmo bug ao mesmo tempo em todo o ambiente
- Permitir implantar em local físico distinto

# Requisitos



- Obrigatórios
  - Nova metodologia
  - IPv6
- Desejáveis
  - Ativo/Ativo
  - Operação em camada 3
- Solução adotada: Anycast

# Anycast



- Pode ser feito com qualquer protocolo de roteamento
- OSPF
  - OSPFv3 do Quagga ainda é experimental
    - Processo para sem razão alguma
  - Poderia usar outro daemon (BIRD?)
- BGP
  - Injeção de rotas mais simples
  - Permite posicionar os nós em outro local sem a necessidade de um enlace L2

# ECMP



- Equal-cost multi-path routing
- Múltiplas rotas ativas para encaminhamento
  - Faz balanceamento entre as rotas disponíveis
  - Mais de uma rota para o mesmo prefixo é inserida na FIB
- Necessário habilitar e configurar nos roteadores para o balanceamento funcionar corretamente



# ExaBGP



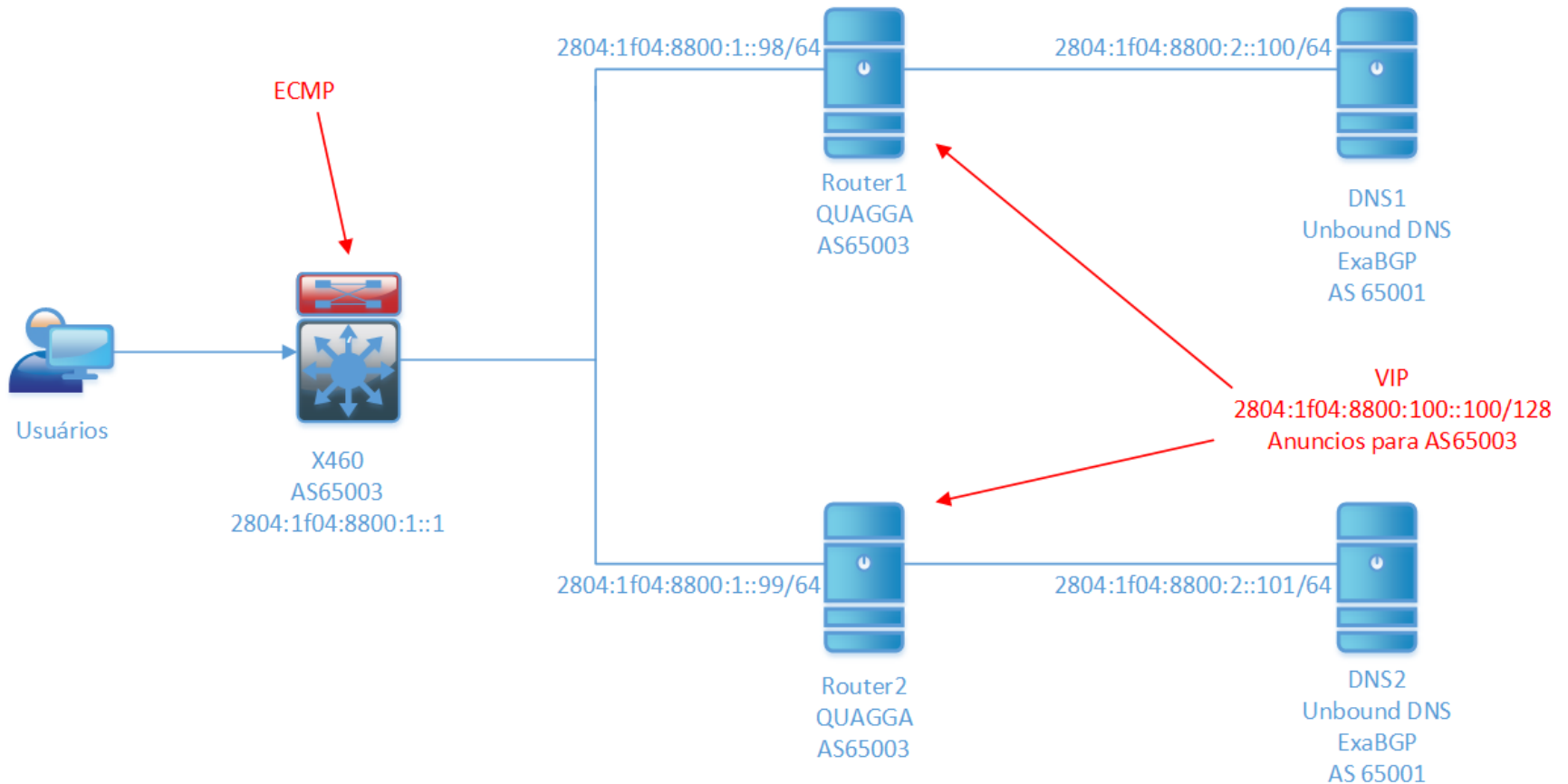
- É um daemon BGP
  - Porém, não manipula a FIB
- Permite injeção de rotas no BGP
  - Transforma mensagens de texto em anuncios de rotas
  - Transforma anuncios recebidos em texto/JSON
- Feito para rodar com scripts

# BGP Anycast



- Equipamentos para testes
  - Backend: 2x Unbound DNS + ExaBGP
  - Routers: 2x Quagga
  - Core: 1x Extreme X460

# Topologia de testes



# Testes



- Healthcheck - ExaBGP

```
run /usr/bin/python
/opt/exabgp/etc/exabgp/processes/healthcheck.py
--cmd "/opt/exabgp-conf/check_dns.sh" --start-ip 0
```

- Adiciona e remove rotas de acordo com a disponibilidade do serviço monitorado

```
Wed, 02 Sep 2015 11:14:39 | INFO      | 5715   | reactor      | Route added to
neighbor 2804:1f04:8800:1::1 local-ip 2804:1f04:8800:1::100 local-as 65001 pe
er-as 65003 router-id 177.101.22.100 family-allowed in-open : 2804:1f04:8800:1
00::100/128 next-hop 2804:1f04:8800:1::100 med 103
Wed, 02 Sep 2015 11:14:40 | INFO      | 5715   | reactor      | Performing dyn
amic route update
Wed, 02 Sep 2015 11:14:40 | INFO      | 5715   | reactor      | Updated peers
dynamic routes successfully
```

# Testes



- dig hostname.bind

```
cpdv0025:~# dig @2804:1f04:8800:100::100 chaos txt hostname.bind +short
"cpdv0089"
cpdv0025:~# dig @2804:1f04:8800:100::100 chaos txt hostname.bind +short
"cpdv0090"
```

- Rotas

```
cpdv0083-bgpd# sh ipv6 bgp
BGP table version is 0, local router ID is 177.101.31.31
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 2804:1f04:8800:100::100/128	2804:1f04:8800:2::100	100		0	65001 i
*	2804:1f04:8800:2::101	101		0	65001 i

Total number of prefixes 1

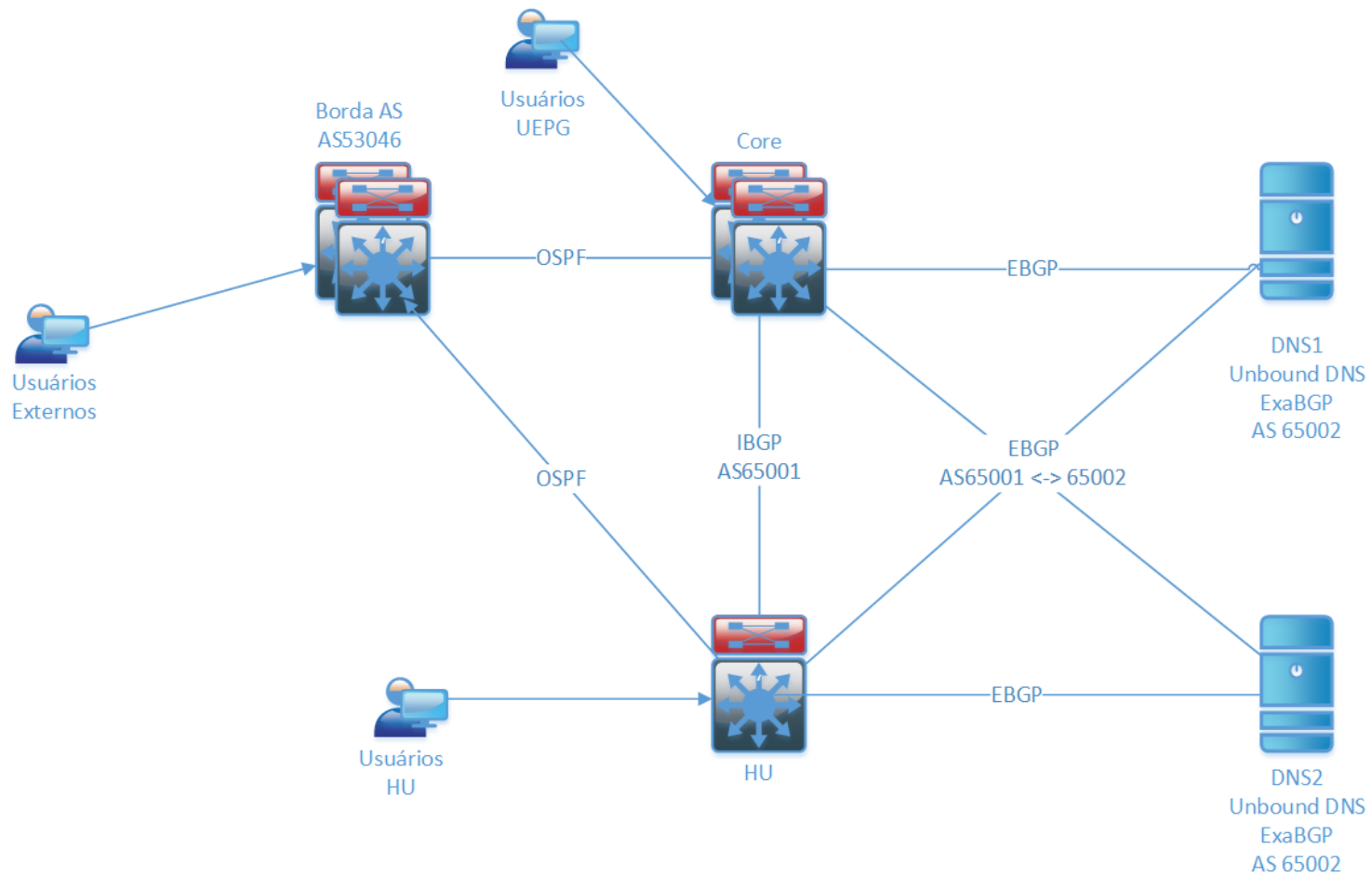
```
cpds0011.57 # sh iproute ipv6 origin bgp
Ori Destination
Gateway
#bi 2804:1f04:8800:100::100/128
2804:1f04:8800:1::99
#bi 2804:1f04:8800:100::100/128
2804:1f04:8800:1::98
```

Quagga

Extreme XOS

Mtr	Flags	Duration
0	UGHD---um--f-	0d:0h:0m:56s
	v0277-dnsAny	
0	UGHD---um--f-	0d:0h:0m:39s
	v0277-dnsAny	

# Topologia Final



OBS: A topologia do DNS recursivo e autoritativo é idêntica

# Conclusões



- Funciona!
- Pode escalar para dezenas de Mpps com routers dedicados na função de load balancer

# Trabalhos futuros



- Implantar em mais serviços
  - LDAP
  - HTTP
- Adicionar Route Reflector ou Route Server caso a quantidade de nós aumente



# Funcionalidades ExaBGP



- BGP Looking Glass
  - Análise da tabela de roteamento (MySQL)
  - Em produção: <http://lg.uepg.br>
- Manipulação de rotas
  - Injeção de rotas nos routers do AS
  - Útil para bloquear prefixos (blackhole)
    - Coordenar com upstreams RTBH via communities BGP

# Referências



- ExaBGP
  - <https://github.com/Exa-Networks/exabgp>
  - <http://vincent.bernat.im/en/blog/2013-exabgp-highavailability.html>
- Anycast
  - <http://ddiguru.com/blog/118-introduction-to-anycast-dns>
  - <http://bits.shutterstock.com/2014/05/22/stop-buying-load-balancers-and-start-controlling-your-traffic-flow-with-software/>
  - <https://www.nanog.org/meetings/nanog37/presentations/matt.levine.pdf>