DNS recursivo estável e escalável

Marcelo Gardini
Registro.br

marcelo@registro.br

GTER23 – 29/jun/2007

Agenda

- Objetivo
- A solução deve oferecer
- Exemplos
- Solução
- Topologia
- Anycast dentro da sua rede
- ECMP
- Detalhes do BIND
- Controle do cluster
- Recursos
- Monitoração

Objetivo

- Apresentar uma solução estável e escalável de um servidor DNS recursivo

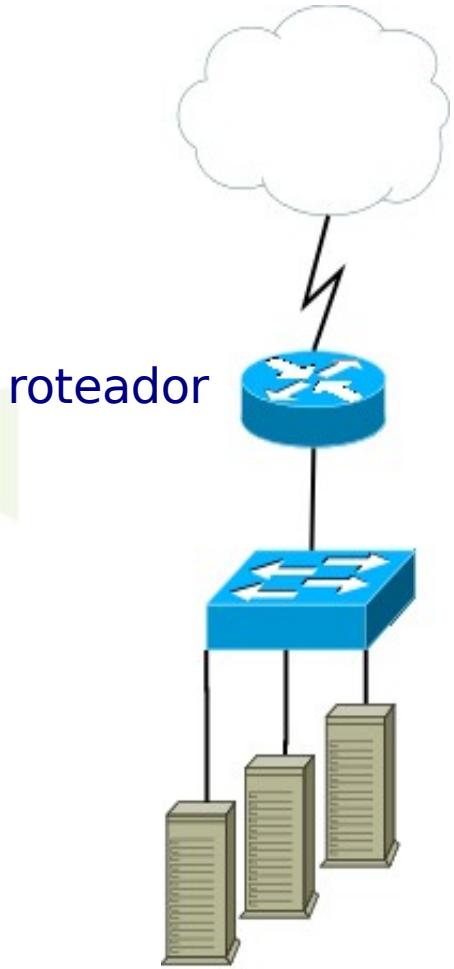
- Motivação
  - DNS recursivo instável é foco de reclamações constantes

A solução deve oferecer

- Escalabilidade e Estabilidade:
  - Suporte a alto tráfego
  - Suporte a muitos usuários
  - Fácil upgrade caso a demanda aumente
  - Manutenção sem parada do sistema
  - Alta disponibilidade
  - Balanceamento de carga

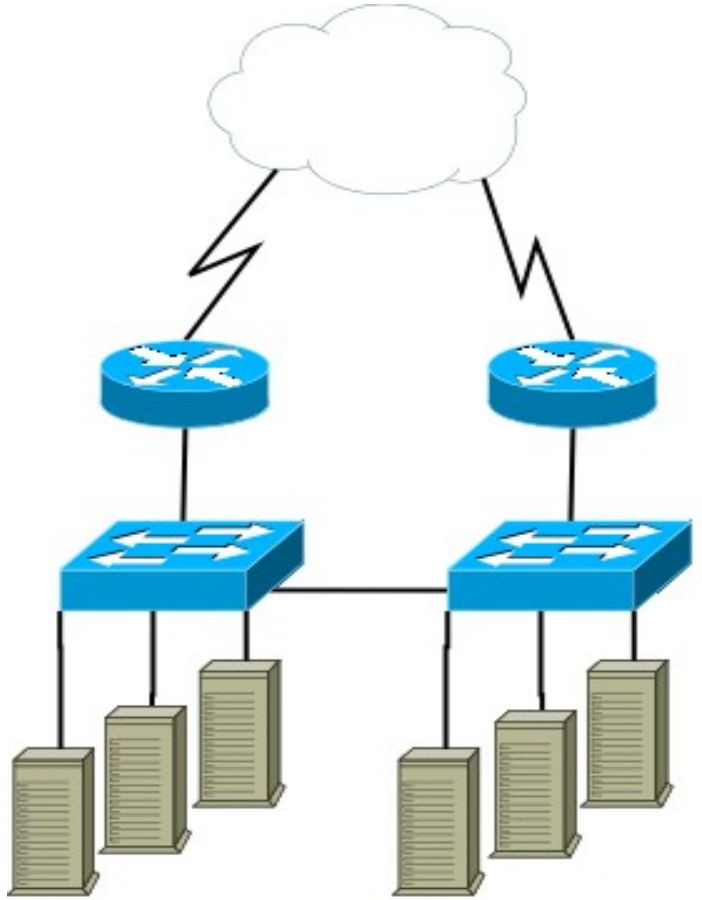
Exemplos

- Cluster



Exemplos

- Flexibilidade
  - Mais de um roteador
  - Mais de um switch



Solução

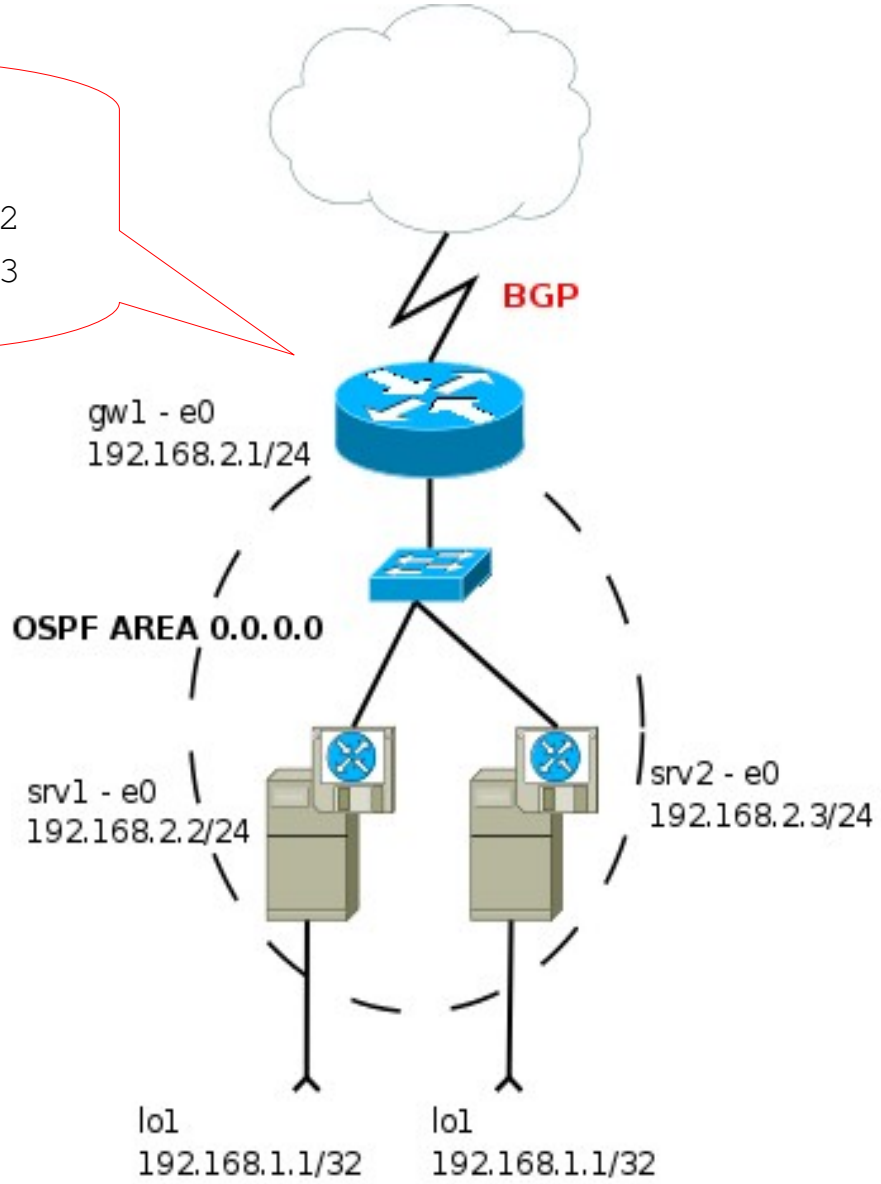
- Roteador + [UNIX + Quagga + BIND]

- Endereço do serviço
  - Roda na loopback dos servidores
- Anycast dentro do cluster
  - Protocolo de roteamento dinâmico com suporte a ECMP - OSPF

Topologia

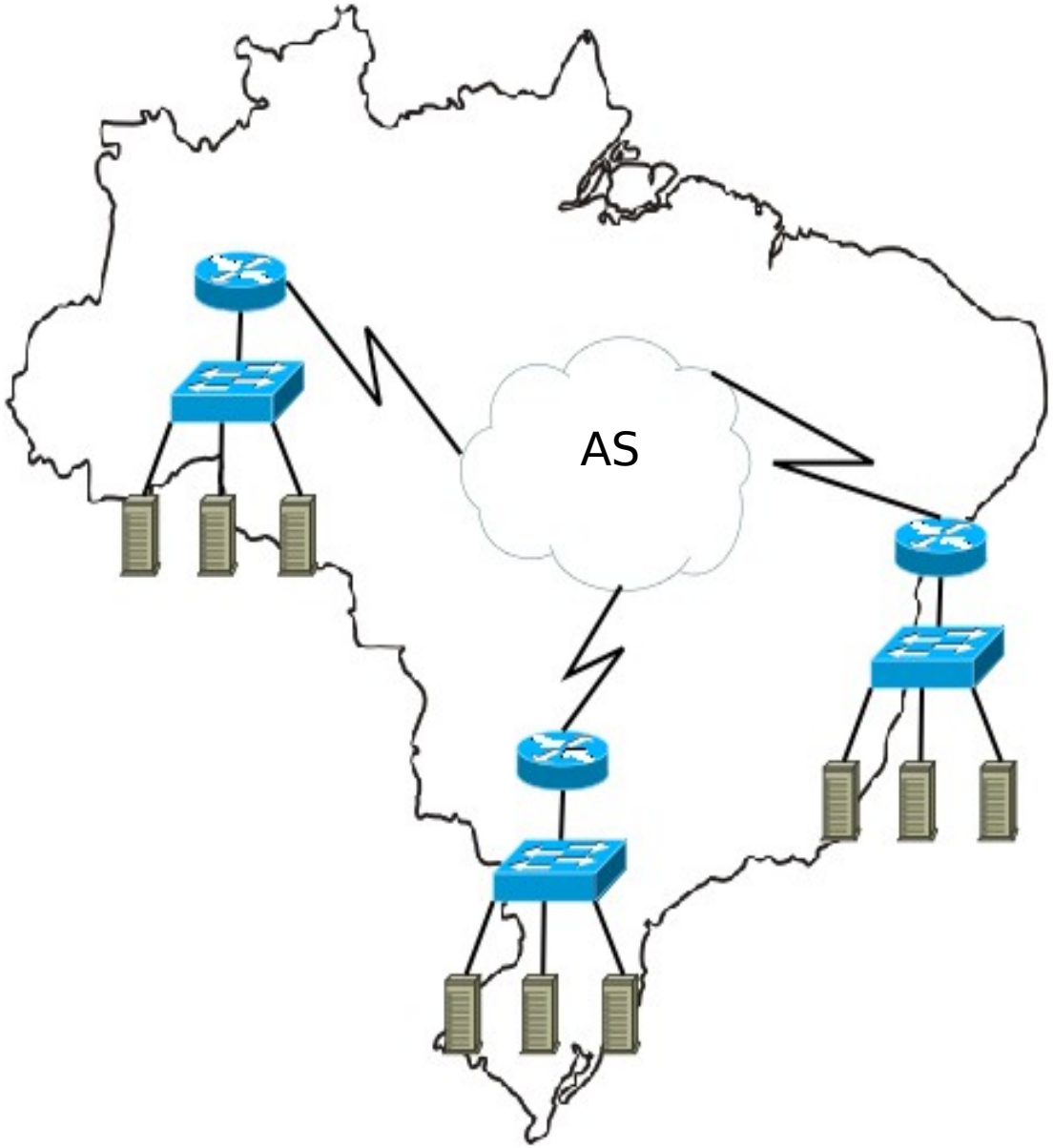
```
gw1# show ip ospf route
     192.168.1.0/24 is variably subnetted
O       192.168.1.1/32 [110/11] via 192.168.2.2
        192.168.1.1/32 [110/11] via 192.168.2.3
```

- ECMP
- Anycast no cluster



Anycast na rede

- Redundância
- iBGP entre os clusters



ECMP

- DNS: “stateless” ou de vida curta
- Deve-se garantir flow-hashing forwarding
  - CEF: Cisco Express Forwarding
  - Juniper: Load-Balance Per-Packet
  - Linux + quagga: patch no kernel

ECMP

- Assimetria de tráfego
  - O balanceamento de carga não é exato
- Incoerência de cache entre os servidores
  - O cache dos servidores não são idênticos, pois o tráfego que eles recebem também não é

Não há problema!

Detalhes do BIND

- BIND rodando nas duas interfaces
  - Queries chegam para a loopback
  - Monitoração pelo endereço individual
- query-source address
  - Forçar o endereço de origem para as consultas de cada servidor
- acl
- max-cache-size

named.conf

```
acl "lista" { 127.0.0.1; 10.0.0.0/8; };

options {
        directory       "/etc";
        pid-file        "/var/run/named/pid";
        dump-file       "/var/dump/named_dump.db";
        statistics-file "/var/stats/named.stats";
        allow-query { "lista"; };
        allow-query-cache { "lista"; };
        allow-recursion { "lista"; };
        query-source address 192.168.2.2 port 1050;
        max-cache-size 512M;
        cleaning-interval 60;
        clients-per-query 0;
        max-clients-per-query 0;
};
```

```
logging {
        channel all { file "/var/log/named.log"
                      versions 5 size 1M;
                      print-time yes; };
        category default { all; };
        category security { all; };
        category lame-servers { null; };
};

zone "." {
        type hint;
        file "named.root";
};

zone "0.0.127.in-addr.arpa" {
        type master;
        file "db.127.0.0";
};
```

named.conf

```
acl "lista" { 127.0.0.1; 10.0.0.0/8; };

options {
        ...
        allow-query { "lista"; };
        allow-query-cache { "lista"; };
        allow-recursion { "lista"; };
        query-source address 192.168.2.2 port 1050;
        max-cache-size 512M;
        cleaning-interval 60;
        clients-per-query 0;
        max-clients-per-query 0;
};
...
```

Controle

I. Iniciar processo do BIND
   - Para não perder consultas

II. Subir interface loopback
   - Quagga envia LSA para o roteador

```
ifconfig lo1 up
```

III. Tirar o servidor do ar
   - Enviar LSA removendo a rota

```
ifconfig lo1 down
```

Recursos

- Exemplo:
  - Rede com 1M de usuários simultâneos
  - Média de 50 q/h (por usuário)

50 x 1M = 50M q/h ~ 14k q/s

Um único servidor (com hardware robusto) é capaz de suportar esta carga

Sugestão para atingir alta disponibilidade: cluster com 3 ou 4 servidores

Monitoração

- Processo do BIND está no ar?
  - Monitorar endereço unicast de cada servidor
- Qual servidor que está respondendo?

```
dig @192.168.1.1 chaos txt hostname.bind +short
```

- Watchdog
  - Caso o BIND pare de responder

```
ifconfig lo1 down
```

Referências

- ISC-TN-2004-1
  A software aproach to distributing requests for DNS service
  http://www.isc.org/index.pl?/pubs/tn/index.pl?tn=isc-tn-2004-1.txt
- RFC 2328
  OSPF version 2
  ftp://ftp.registro.br/rfc/rfc2328.txt
- BIND 9.4 documentation
  http://www.isc.org/index.pl?/sw/bind/arm94/

Obrigado!