Análise dos Prefixos IPv4 BR na tabela BGP e dos impactos decorrentes das soluções para redução do seu tamanho

Eduardo Ascenço Reis
<eduardo@intron.com.br>
<eascenco@ctbc.com.br>

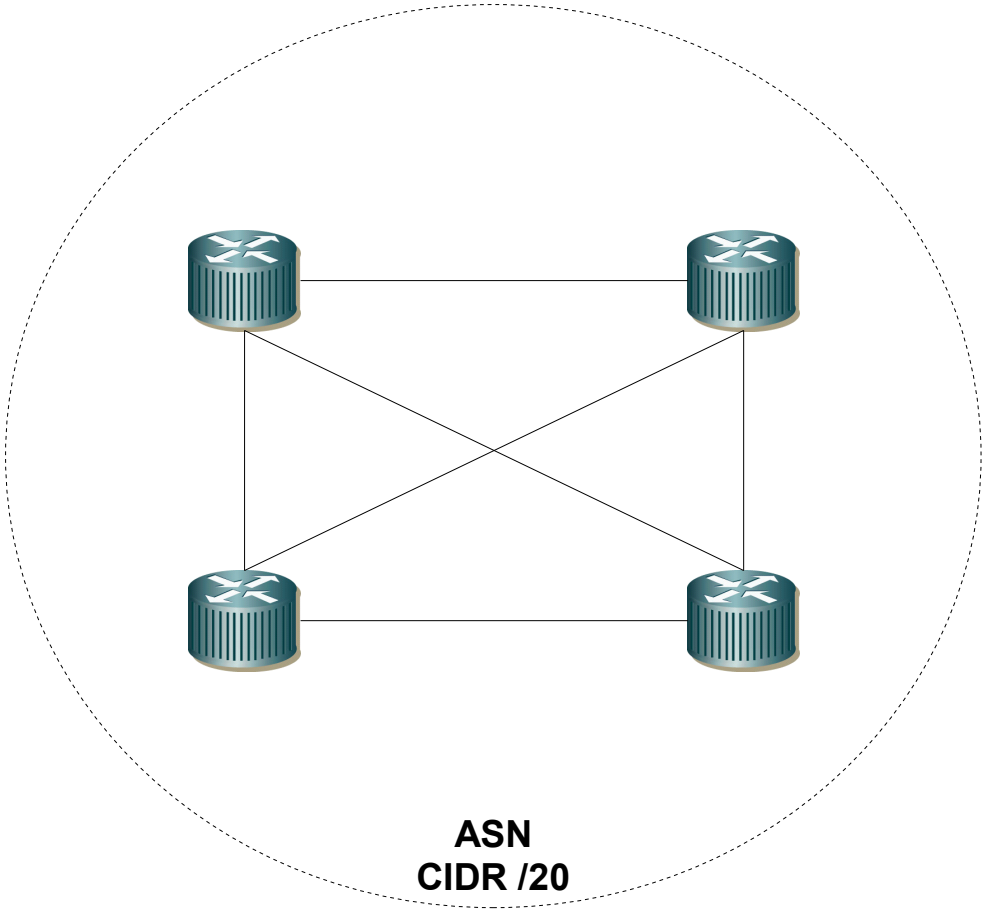
Resumo

O crescente aumento da tabela BGP IPv4 tem ocasionado problemas operacionais para os Sistemas Autônomos (AS) em toda a Internet, devido a necessidade de proporcional crescimento dos recursos de hardware (memória e processamento) dos roteadores que utilizam o protocolo BGP, em especial com tabela completa.

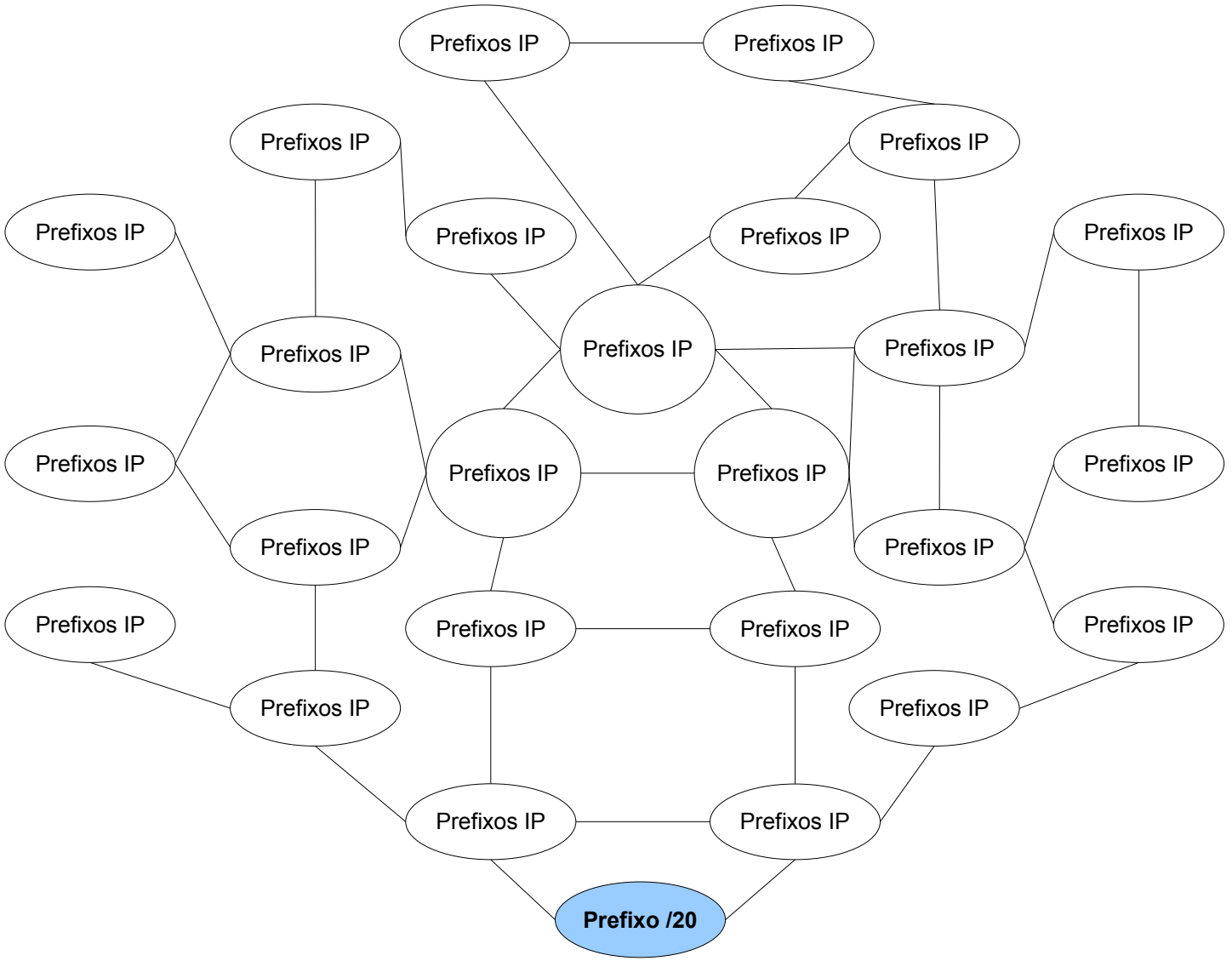
Algumas soluções técnicas para reduzir o tamanho da tabela BGP completa estão em plena discussão (e.g. NANOG), das quais destacam-se os filtros de prefixos: pelo limite das menores alocações por bloco CIDR de cada RIR (Regional Internet Registry) ou pela agregação baseada em prefixos com o mesmo AS-PATH, como o algorítmo utilizado pelo CIDR Report (http://www.cidr-report.org).

O objetivo deste trabalho é apresentar a participação atual dos AS Brasileiros, e respectivos prefixos, no tamanho da tabela BGP IPv4, e um estudo dos impactos previstos nesses mesmos AS com a aplicação dos filtros de prefixos que estão em discussão.

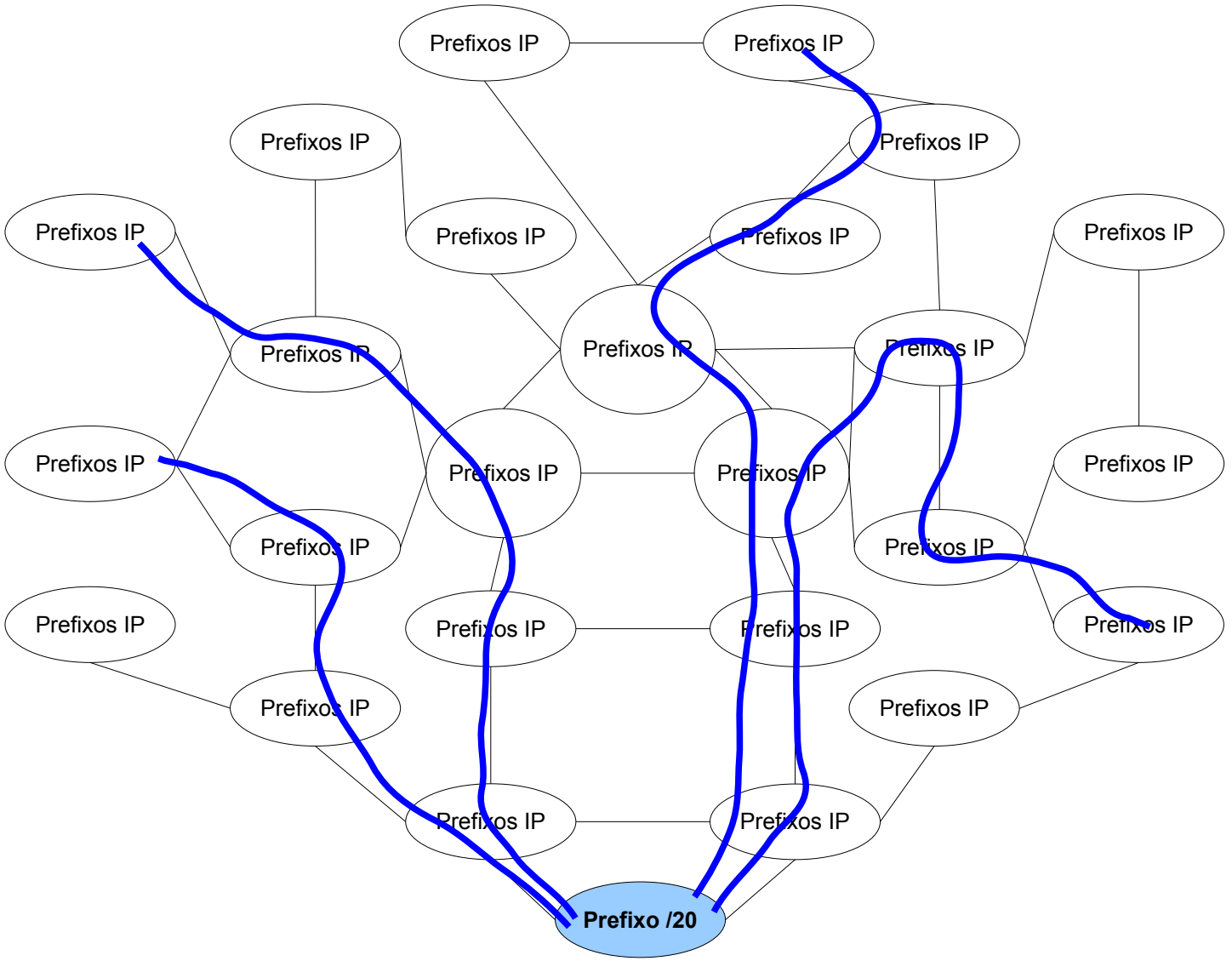
Modelo: Novo Sistema Autônomo (AS) Brasileiro



Internet: Rede de AS Interconectados



Internet: Opções de Caminho – AS-PATH (Vetor de AS até o Destino)



Domínios de Informação – Diferença de Números de Prefixos

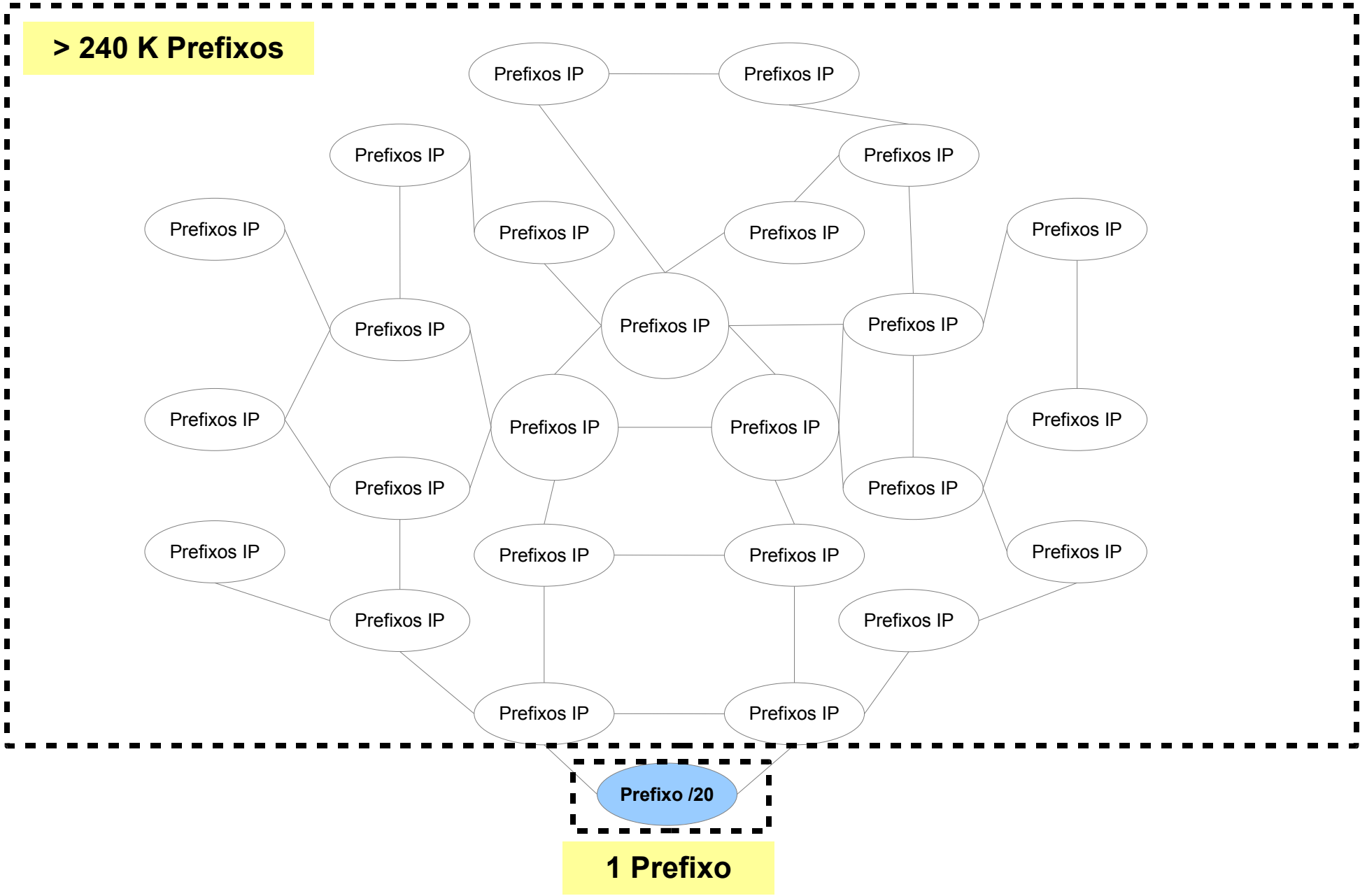


Tabela BGP IPv4 Completa # Parcial

Informação é Poder

Política de Roteamento AS modelo (as-out)

Controle dos Anúncios feitos => Interfere com tráfego de entrada.

Grande interesse para provedores de acesso.

Exemplo de anúncios mais específicos além do /20:
2x /21
4x /22
8x /23
16x /24

Soma: 30 prefixos - sob definição de cada AS em toda a Internet

Política de Roteamento AS modelo (as-in)

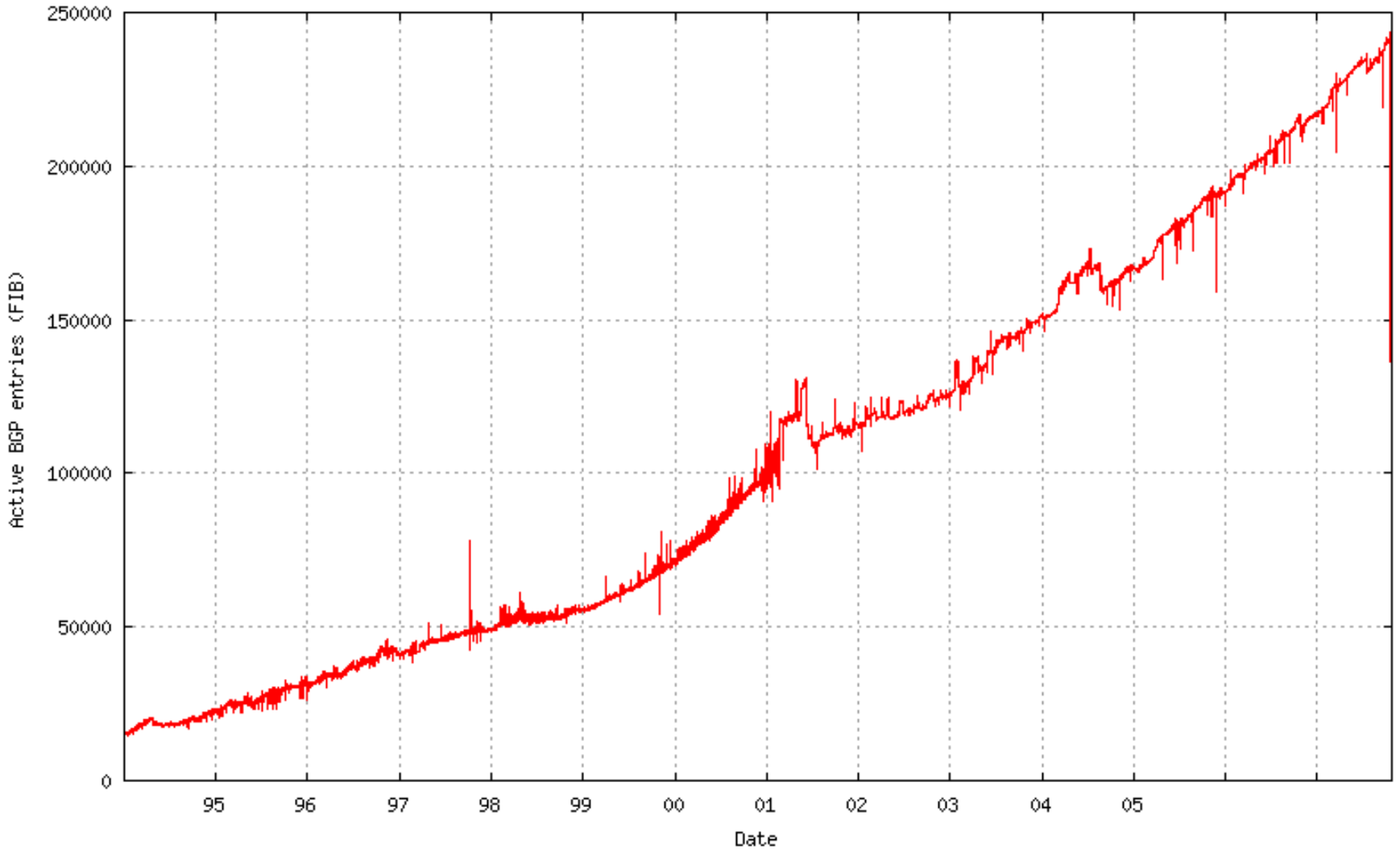
Controle dos Anúncios Recebidos => Interfere com tráfego de saída.

Grande interesse para provedores de conteúdo (grandes sites, IDC, etc)

Definição de Prefixos Recebidos

- Tabela Completa
- Tabela Parcial (e.g. somente prefixos nacionais) + Rota Default

Curva de Crescimento da Tabela BGP IPv4 (AS6447)



http://www.cidr-report.org/
http://bgp.potaroo.net/

Tabela BGP IPv4 Completa – Demanda de Hardware

Tabela BGP IPv4 Completa impõem demanda de recursos de hardware dos participantes (provedor e cliente):

1. Capacidade de Processamento
2. Memória de Trabalho

- Recursos normais em equipamentos computadorizados.
- Recursos finitos, que portanto se esgotam!
- Roteador é um computador dedicado !

Situação de Esgotamento de Recursos: Memória e/ou Processamento

Em situações de falta desses recursos (limite) há duas soluções básicas:

1. Adição de mais recurso.
Normalmente: Exige Investimento

2. Restrição do consumo.
Normalmente: Resulta em Perda de Capacidade

Esgotamento de Recursos – AS Afetados

Problema afeta todos os AS na Internet!!!

Soluções de Redução da Tabela BGP IPv4 - Uso de filtros

* bloqueio de prefixos >= /24 != BR
* permite somente prefixos BR
* permite somente prefixos de PTTs
* bloqueio de prefixos mais espefícos do que as alocações mínimas dos RIR
* agregação de prefixos com AS-PATH idêntico

=> Normalmente essas soluções precisão estar associadas ao uso de rota default !!!

Cuidados com soluções de redução da tabela

- Perigo de gerar inconsistência entre
  - RIB e FIB
  - Roteadores do mesmo AS
- Condição Básica de Boa Operação de AS
  - Política de Roteamento Externo Única
  - Consistência das FIB de todos os roteadores que utilizam BGP no AS

Restrição da Tabela BGP IPv4 pelo Limite das Alocações Mínimas dos RIR

Proposta de utilização de filtro de prefixos na regra de entrada, bloqueando prefixos IPv4 mais específicos do que a alocação mínima do bloco correspondente feito pelo Registro Regional.

09/07/07 Route table growth and hardware limits.. Jon Lewis
http://www.merit.edu/mail.archives/nanog/msg02822.html

Barry Greene - bgreene@cisco.com
ftp://ftp-eng.cisco.com/cons/isp/security/Ingress-Prefix-Filter-Templates/
T-ip-prefix-filter-ingress-strict-check-v18.txt

Autoridades Internacionais para Alocação de Endereços IPv4

ICANN (Internet Corporation for Assigned Names and Numbers)
http://www.icann.org/

IANA (Internet Assigned Numbers Authority)
http://www.iana.org/

No ínicio de operação da Internet a IANA realizava diretamente as alocações de recursos (IP e ASN) para os AS.

Essa antigas alocações (blocos /8 e outros) são hoje identificados como alocações legadas.

Atualmente as alocações são feitas pelos Registros Regionais (RIR), de acordo com as suas áreas de atuação.

Autoridades Regionais para Alocação de Endereços IPv4

Regional Internet Registries (RIRs)

AfriNIC (African Network Information Centre)
http://www.afrinic.net/

APNIC (Asia Pacific Network Information Centre)
http://www.apnic.net/

ARIN (American Registry for Internet Numbers)
http://www.arin.net/

LACNIC (Regional Latin-American and Caribbean IP Address Registry)
http://lacnic.net/

RIPE NCC (Réseaux IP Européens Network Coordination Centre)
http://www.ripe.net/

Regional Internet Registries – Alocações Mínimas de Blocos IPv4

AfriNIC
http://www.afrinic.net/documents.htm#templates

APNIC
http://www.apnic.net/db/min-alloc.html

ARIN
http://www.arin.net/reference/ip_blocks.html#ipv4

LACNIC
http://lacnic.net/pt/registro/index.html

RIPE
https://www.ripe.net/ripe/docs/ripe-ncc-managed-address-space.html

Alocações Mínimas de Blocos IPv4 – exemplo LACNIC

Bloco IPv4	alocação mínima
186.0.0.0/8	/20
187.0.0.0/8	/20
189.0.0.0/8	/20
190.0.0.0/8	/20
200.0.0.0/8	/24
201.0.0.0/8	/20

Autoridade para Alocação de Endereços IPv4 - Brasil

No caso Brasil, o LACNIC delegou o controle de alocações de recursos Internet (ASN, IPv4 e IPv6) para o Registro.br (http://registro.br/)

Atualmente o Registro.br adota como padrão de alocação mínima de blocos CIDR IPv4 um /20.

Exemplos de alocações BR

USP
143.107.0.0/16 - Alocação: Legada (IANA)

CTBC
Bloco CIDR: 200.160.112.0/20 - Alocação: Registro.br

Polêmica sobre Tamanho da Alocação Mínima de Blocos IPv4

Há demanda de pretendentes a AS para alocações mínimas mais específicas (e.g. /24).

1 bloco /20 atenderia 16 AS com /24

✗ Consequente aumento da Tabela BGP !!!

Filtro de Prefixos IPv4 por Limite de RIR (Cisco prefix-list) – Parte 1/2

```
ip prefix-list ISP-Ingress-In-Strict: 51 entries
   seq 4000 deny 58.0.0.0/8 ge 22
   seq 4001 deny 59.0.0.0/8 ge 22
   seq 4002 deny 60.0.0.0/7 ge 22
   seq 4004 deny 116.0.0.0/6 ge 22
   seq 4008 deny 120.0.0.0/6 ge 22
   seq 4011 deny 124.0.0.0/7 ge 22
   seq 4013 deny 126.0.0.0/8 ge 22
   seq 4014 deny 202.0.0.0/7 ge 25
   seq 4016 deny 210.0.0.0/7 ge 22
   seq 4018 permit 218.100.0.0/16 ge 17 le 24
   seq 4019 deny 218.0.0.0/7 ge 22
   seq 4021 deny 220.0.0.0/7 ge 22
   seq 4023 deny 222.0.0.0/8 ge 22
   seq 5000 deny 24.0.0.0/8 ge 21
   seq 5001 deny 63.0.0.0/8 ge 21
   seq 5002 deny 64.0.0.0/6 ge 21
   seq 5006 deny 68.0.0.0/7 ge 21
   seq 5008 deny 70.0.0.0/7 ge 21
   seq 5010 deny 72.0.0.0/6 ge 21
   seq 5014 deny 76.0.0.0/8 ge 21
   seq 5015 deny 96.0.0.0/6 ge 21
   seq 5020 deny 198.0.0.0/7 ge 25
   seq 5022 deny 204.0.0.0/7 ge 25
   seq 5023 deny 206.0.0.0/7 ge 25
   seq 5032 deny 208.0.0.0/8 ge 23
   seq 5033 deny 209.0.0.0/8 ge 21
   seq 5034 deny 216.0.0.0/8 ge 21
```

Filtro de Prefixos IPv4 por Limite de RIR (Cisco prefix-list) – Parte 2/2

```
seq 6000 deny 62.0.0.0/8 ge 20
seq 6001 deny 77.0.0.0/8 ge 22
seq 6002 deny 78.0.0.0/7 ge 22
seq 6004 deny 80.0.0.0/7 ge 21
seq 6006 deny 82.0.0.0/8 ge 21
seq 6007 deny 83.0.0.0/8 ge 22
seq 6008 deny 84.0.0.0/6 ge 22
seq 6012 deny 88.0.0.0/7 ge 22
seq 6014 deny 90.0.0.0/8 ge 22
seq 6015 deny 91.0.0.0/8 ge 25
seq 6016 deny 92.0.0.0/6 ge 22
seq 6020 deny 193.0.0.0/8 ge 25
seq 6021 deny 194.0.0.0/7 ge 25
seq 6023 deny 212.0.0.0/7 ge 20
seq 6025 deny 217.0.0.0/8 ge 21
seq 7010 deny 186.0.0.0/8 ge 21
seq 7020 deny 187.0.0.0/8 ge 21
seq 7100 deny 189.0.0.0/8 ge 21
seq 7201 deny 190.0.0.0/8 ge 21
seq 7302 deny 200.0.0.0/8 ge 25
seq 7403 deny 201.0.0.0/8 ge 21
seq 8000 deny 41.0.0.0/8 ge 23
seq 8001 deny 196.0.0.0/8 ge 23
seq 10200 permit 0.0.0.0/0 le 24
```

Laboratório para Testes de Soluções de Restrição da Tabela BGP

Objetivos

- Analisar a eficiência da solução
  Validar os métodos propostos.

- Análisar os impactos resultantes
  Estimar os efeitos negativos decorrentes da sua adoção.

Laboratório – Recursos

Análise

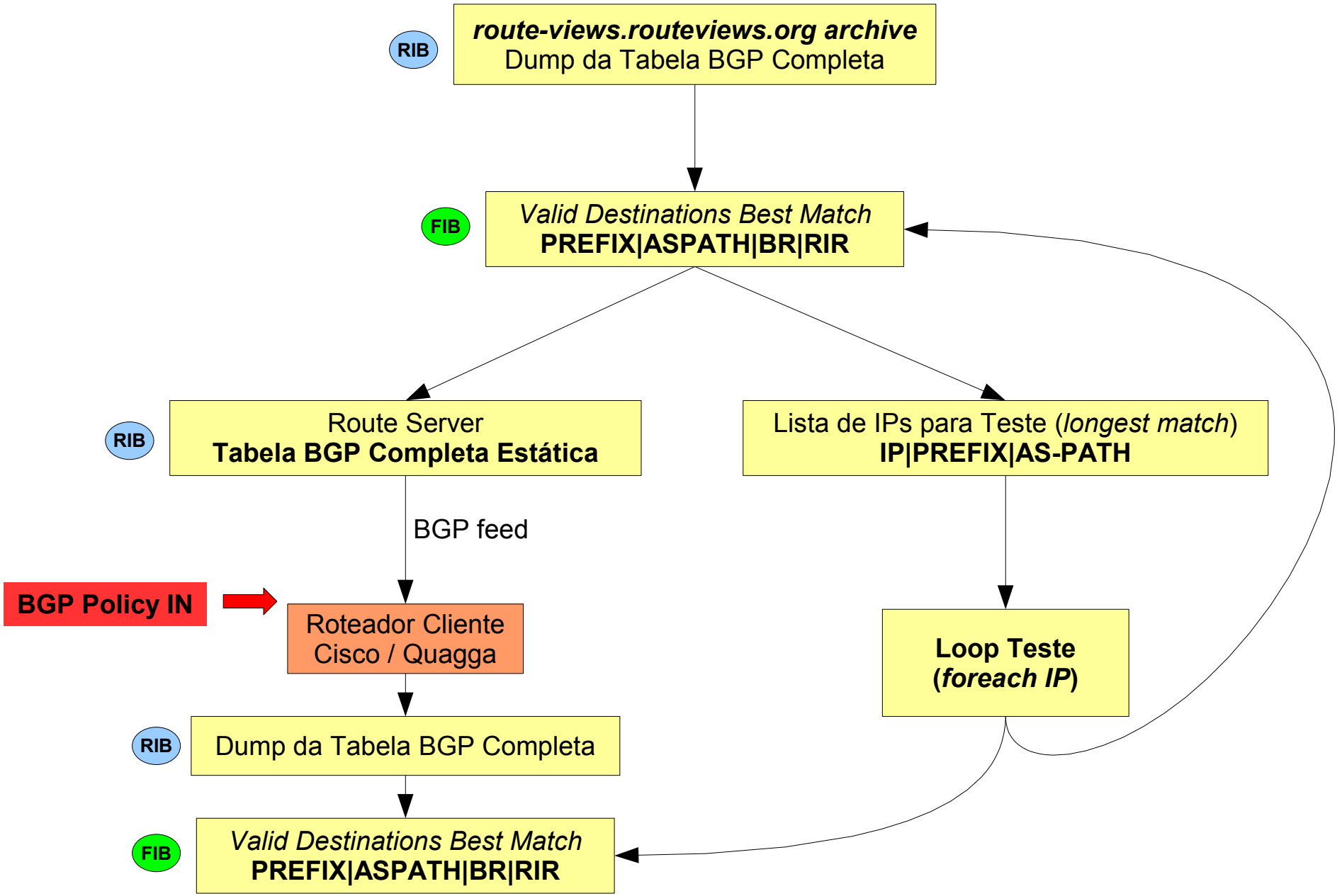
Algorítmo macro de análise é composto por uma série de algorítmos menores, este feitos em Perl e Bash + Aplicativos GNU (awk, grep, etc).

Sistemas

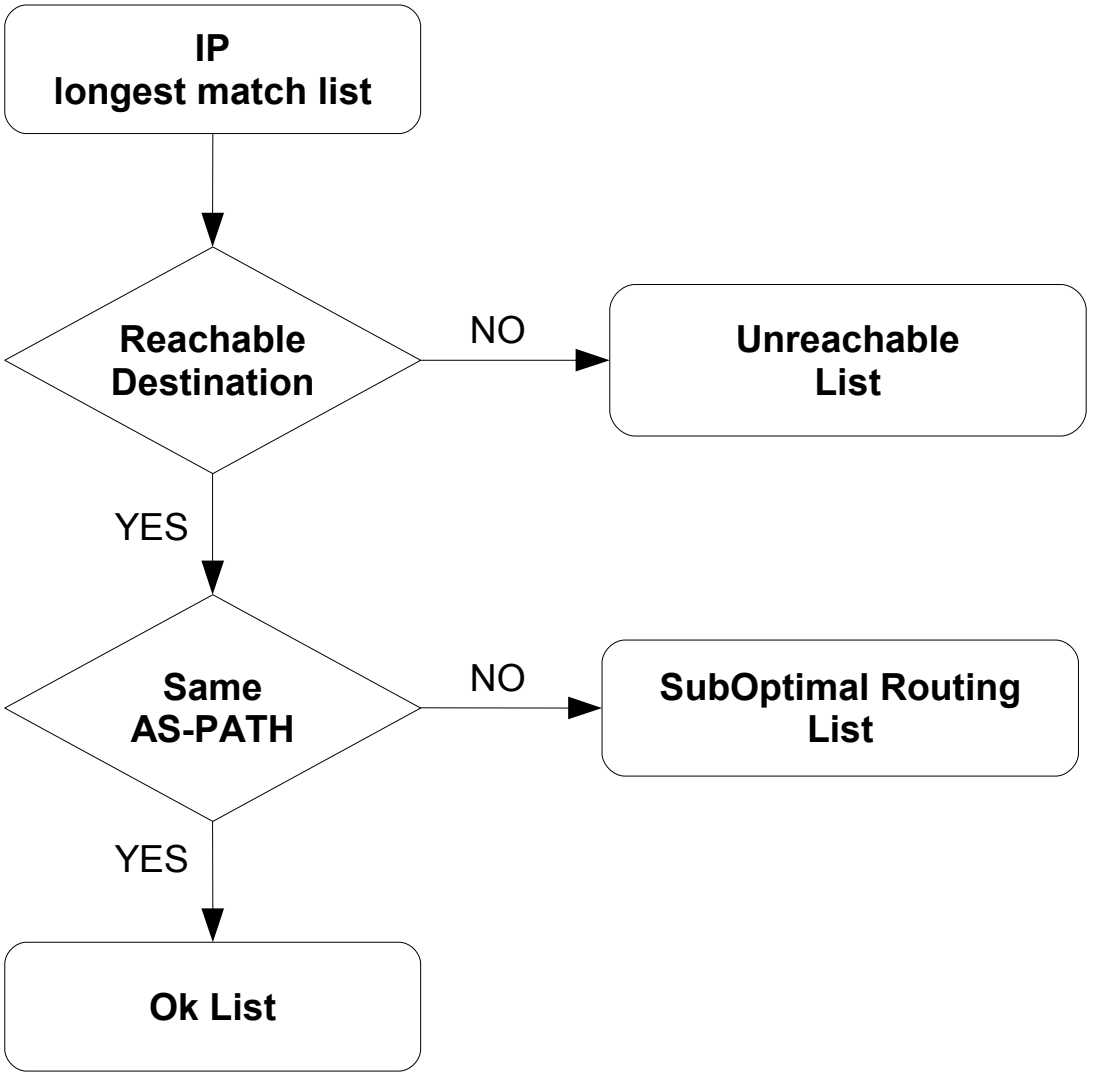
Route Servers - Debian GNU Linux e FreeBSD + Quagga

Roteador Cisco

Laboratório – Diagrama Lógico



Laboratório – Fluxograma de Teste



Laboratório – Informações de Roteamento (Tabela BGP IPv4 Completa)

University of Oregon Route Views Project

http://www.routeviews.org/

University of Oregon Route Views Archive Project

David Meyer

http://archive.routeviews.org/

RIB

```
$ wc -l oix-full-snapshot-2007-10-23-2000.dat
9010444 oix-full-snapshot-2007-10-23-2000.dat
$
```

Laboratório – Informações de Alocações por RIR

IANA

http://www.iana.org/assignments/ipv4-address-space

Observação

Exceto pelas alocações para RIR, as demais foram identificadas como IANA (blocos legados, reservados, etc)

Laboratório – Informações de Encaminhamento (FIB)

FIB

Criada a partir da RIB apenas com a melhor opção de caminho para os destinos válidos

```
$ head oix-full-snapshot-2007-10-23-2000.dat.prefixes-aspath.txt.v9
3.0.0.0/8|3356 701 703 80|NOBR|IANA
4.0.0.0/8|3356|NOBR|IANA
4.0.0.0/9|3356|NOBR|IANA
4.23.112.0/22|6079 174 21889|NOBR|IANA
4.23.112.0/24|3561 174 21889|NOBR|IANA
4.23.113.0/24|3561 174 21889|NOBR|IANA
4.23.114.0/24|3561 174 21889|NOBR|IANA
4.36.116.0/23|3561 174 21889|NOBR|IANA
4.36.116.0/24|3561 174 21889|NOBR|IANA
4.36.117.0/24|3561 174 21889|NOBR|IANA
$

$ wc -l oix-full-snapshot-2007-10-23-2000.dat.prefixes-aspath.txt.v9
242151 oix-full-snapshot-2007-10-23-2000.dat.prefixes-aspath.txt.v9
$
```

Laboratório – Validação do Teste

O Loop de teste foi executado contra a FIB original de OIX e contra a FIB do roteador Cliente.

Ambos os testes resultaram em 100% dos endereços IP na Lista Ok (reachable and same AS-PATH).

Laboratório – Teste pelo Limite das Alocações Mínimas dos RIR

Teste de restrição pelo limite das alocações mínimas dos RIR.

Utilização do prefix-list ISP-Ingress-In-Strict como filtro na regra de entrada (policy in) do roteador cliente.

Cliente utilizando tabela BGP parcial SEM rota default.

Laboratório – Prefixos Filtrados

```
For address family: IPv4 Unicast
 BGP table version 150833, neighbor version 150833/0
 Output queue size: 0
 Index 1, Offset 0, Mask 0x2
 1 update-group member
 Inbound soft reconfiguration allowed
 Incoming update prefix filter list is ISP-Ingress-In-Strict
                                Sent       Rcvd
 Prefix activity:               ----       ----
   Prefixes Current:               0     150832 (Consumes 12591852 bytes)
   Prefixes Total:                 0     150832
   Implicit Withdraw:              0          0
   Explicit Withdraw:              0          0
   Used as bestpath:             n/a     150832
   Used as multipath:            n/a          0
   Saved (soft-reconfig):        n/a      91319 (Consumes 4748588 bytes)

                                 Outbound    Inbound
 Local Policy Denied Prefixes:   --------    -------
   prefix-list                          0      91319
   Bestpath from this peer:        150832        n/a
   Total:                          150832      91319
 Number of NLRIs in the update sent: max 0, min 0
```

Laboratório – Resultado Total – Análise de Eficiência

FIB Cliente	Tabela BGP	Prefixos IPv4
Original OIX	Completa	242.151
Cliente	Parcial	150.832

Diferença	-91.319
Diferença Relativa (%)	-37,71

Laboratório – Resultado Total – Análise de Impactos

Lista	Prefixos	% TOTAL
Ok	159.520	69,07
SubOptimal Routing	43.557	18,86
Unreachable	27.885	12,07
TOTAL	230.962	100

Laboratório – Resultado Total – Distribuição Prefixos por RIR

OIX

RIR	Prefixos	%
AfriNIC	2396	0,99
APNIC	55065	22,74
ARIN	83214	34,36
IANA	38804	16,02
LACNIC	15438	6,38
RIPE	47234	19,51
TOTAL	242151	100

Laboratório – Resultado Total – Análise de Impactos por RIR

Teste Ips

	Ok List		SubOptimal List		Unreachable List	
RIR	**Prefixes**	**%**	**Prefixes**	**%**	**Prefixes**	**%**
AfriNIC	2270	0,98	465	0,2	925	0,4
APNIC	50997	22,08	11206	4,85	3275	1,42
ARIN	80207	34,73	23969	10,38	12857	5,57
IANA	37676	16,31	549	0,24	20	0,01
LACNIC	14450	6,26	1028	0,45	2537	1,1
RIPE	45362	19,64	6340	2,75	8271	3,58
TOTAL	230962	100	43557	18,86	27885	12,07

Laboratório – Resultado Total – Prefixos BR

	Prefixes	% BR
OIX	242.151	1
LACNIC	15438	19
BR	2948	100

Laboratório – Resultado Total – Análise de Impactos BR

Teste Ips – BR

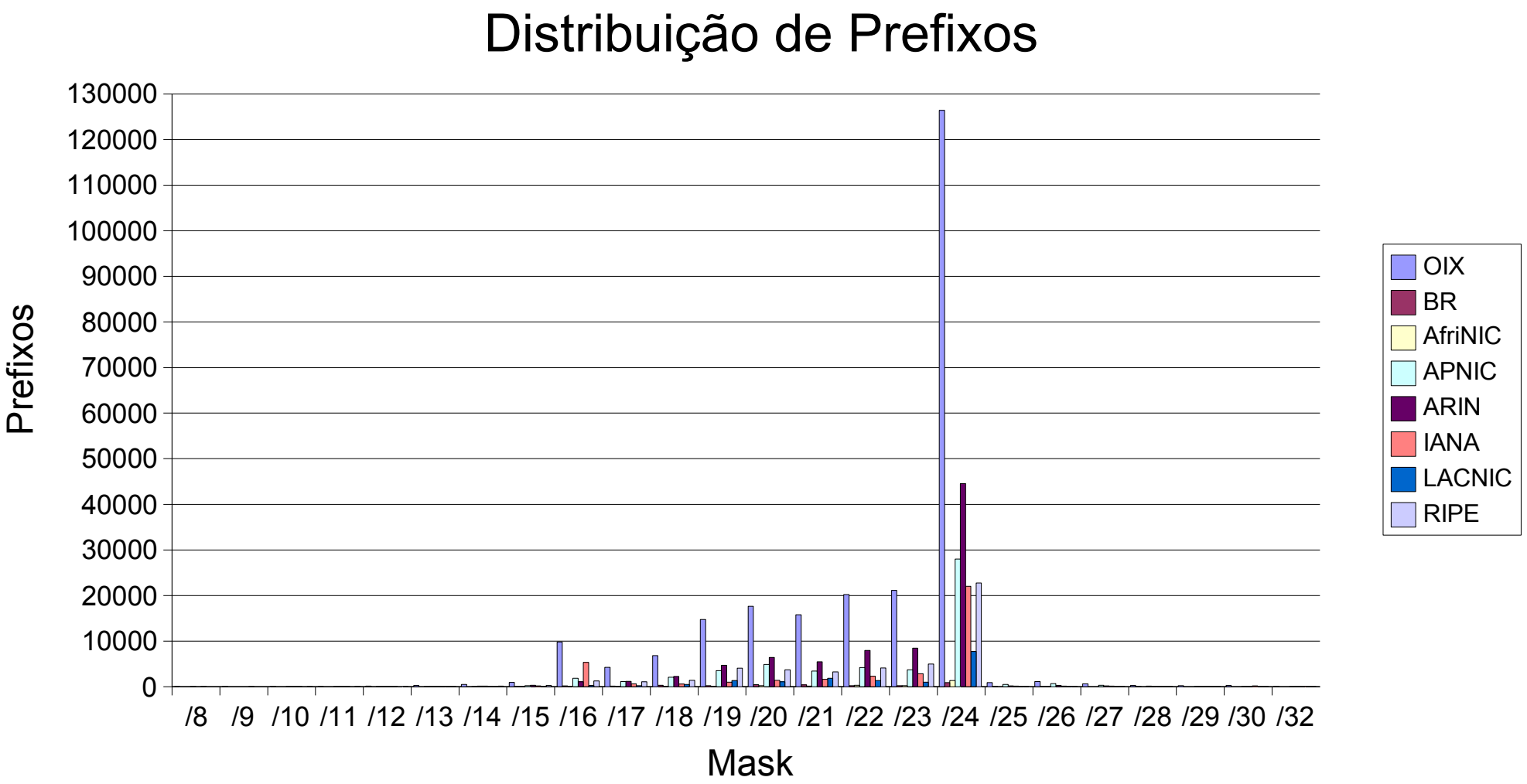
List	Prefixes	%
Ok	2087	79,9
SubOptimal	129	4,94
Unreachable	396	15,16
TOTAL	2612	100

Laboratório – Contabilidade de Prefixos

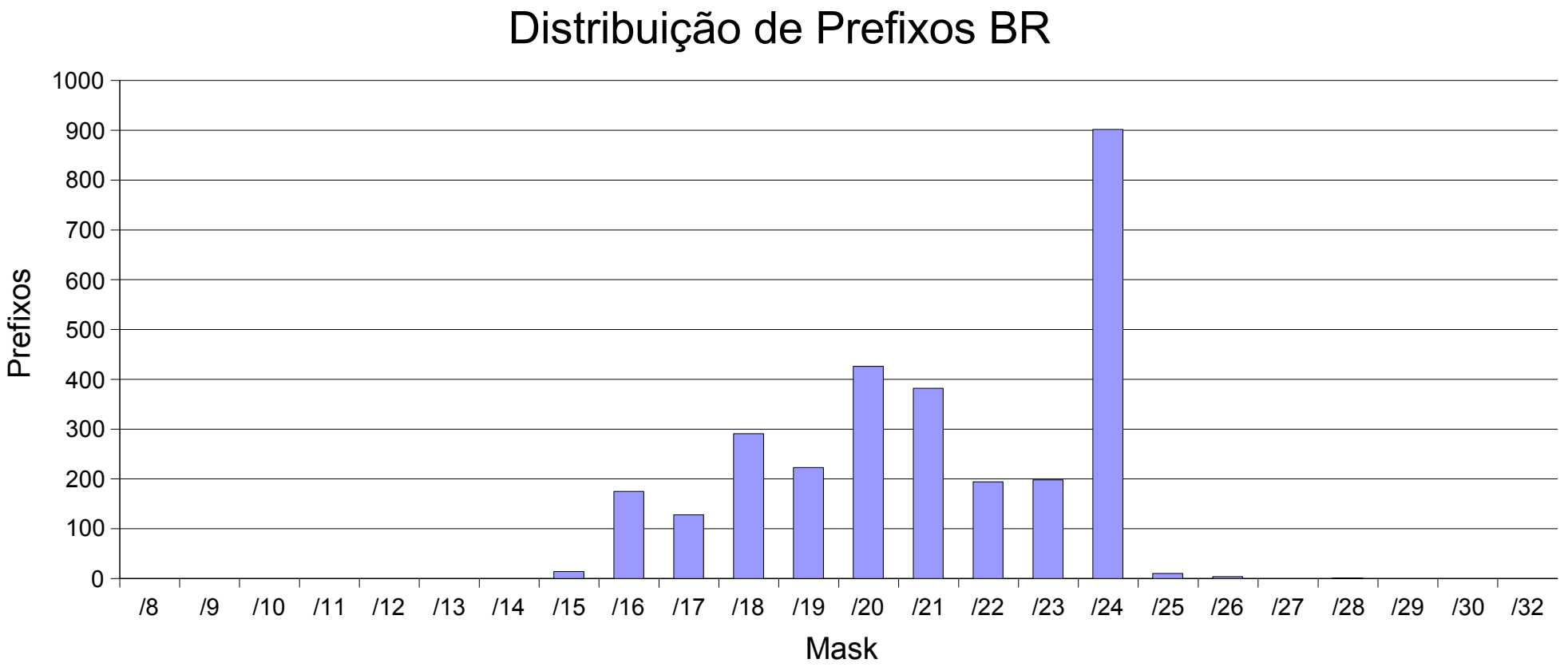
	Prefixes							
Mask	**OIX**	**BR**	**AfriNIC**	**APNIC**	**ARIN**	**IANA**	**LACNIC**	**RIPE**
/8	19			1		18		
/9	9					9		
/10	16			4	2	4		6
/11	38			8	15	2		13
/12	135		1	38	42	24		30
/13	273		4	95	72	43	1	58
/14	484		1	139	128	87	5	124
/15	952	14	4	223	283	169	22	251
/16	9799	175	31	1833	1091	5339	266	1239
/17	4242	128	18	1134	1171	622	212	1085
/18	6811	291	39	2078	2261	581	478	1374
/19	14727	223	77	3522	4679	1009	1368	4072
/20	17650	426	211	4877	6434	1380	1062	3686
/21	15766	382	139	3443	5446	1625	1861	3252
/22	20221	194	297	4210	7967	2299	1340	4108
/23	21115	198	230	3688	8426	2840	971	4960
/24	126446	902	1343	28017	44539	22042	7739	22766
/25	899	10		504	166	133	32	64
/26	1126	4	1	658	255	133	25	54
/27	624			315	156	115	11	27
/28	270	1		120	19	54	40	37
/29	194			88	18	76	3	9
/30	264			68	21	164	1	10
/32	71			2	23	36	1	9

TOTAL	242151	2948	2396	55065	83214	38804	15438	47234

Laboratório – Gráfico de Contabilidade de Prefixos



Laboratório – Gráfico de Contabilidade de Prefixos BR



Sistemas Autônomos BR # Tabela BGP IPv4

Alertas

- Evitar anúncios específicos desnecessários.
- Sempre que possível fazer o anúncio do prefixo correspondente ao bloco total, ou baseado no limite da alocação mínima (/20) para casos de AS com CIDR menos específicos.
- Cuidado com as soluções de restrição de anúncios para evitar inconsistências

Provável Consequência dos Efeitos do Aumento de Tabela BGP IPv4

Adoção de IPv6 é possívelmente afetada de modo negativo devido as restrições de hardware para as atuais necessidades de IPv4 (está em operação e gera receita).

Apêndice

Rota Default é solução para grande maioria (99,9% dos AS) conforme lista ???

Consequências

situações de multi-homming

* tráfego sub-ótimo
* Complexidade operacional (configs) para distribuição de tráfego de saída
  pode gerar custo com equipamentos e pessoal.

* tráfego basal de lixo na Internet decorrente de computadores comprometidos
  varrendo faixas de IP.
  => teste para novos AS
  No início da operação, direcionar todo o tráfego do /20 para um PC
  de teste (e.g. com Quagga fechando BGP com o provedor).
  coletar e medir banda com destino para o /20 recém anúnciado (antes não existia !!!)

  => com rota default o tráfego dos clientes do AS com destino para qualquer IP,
  mesmo que não roteáveis na Internet (não há prefixos correspondentes da tabela BGP)
  serão enviados para o provedor => consumo desnessário de link (Opex)

Dificulta ou inviabiliza mecanismos de segurança para validação (e.g. uRPF) e contabilidade (Netflow) de tráfego
perda de recursos de segurança => necessidade de investimento em recursos extras
para resolver o mesmo problema

Obrigado !