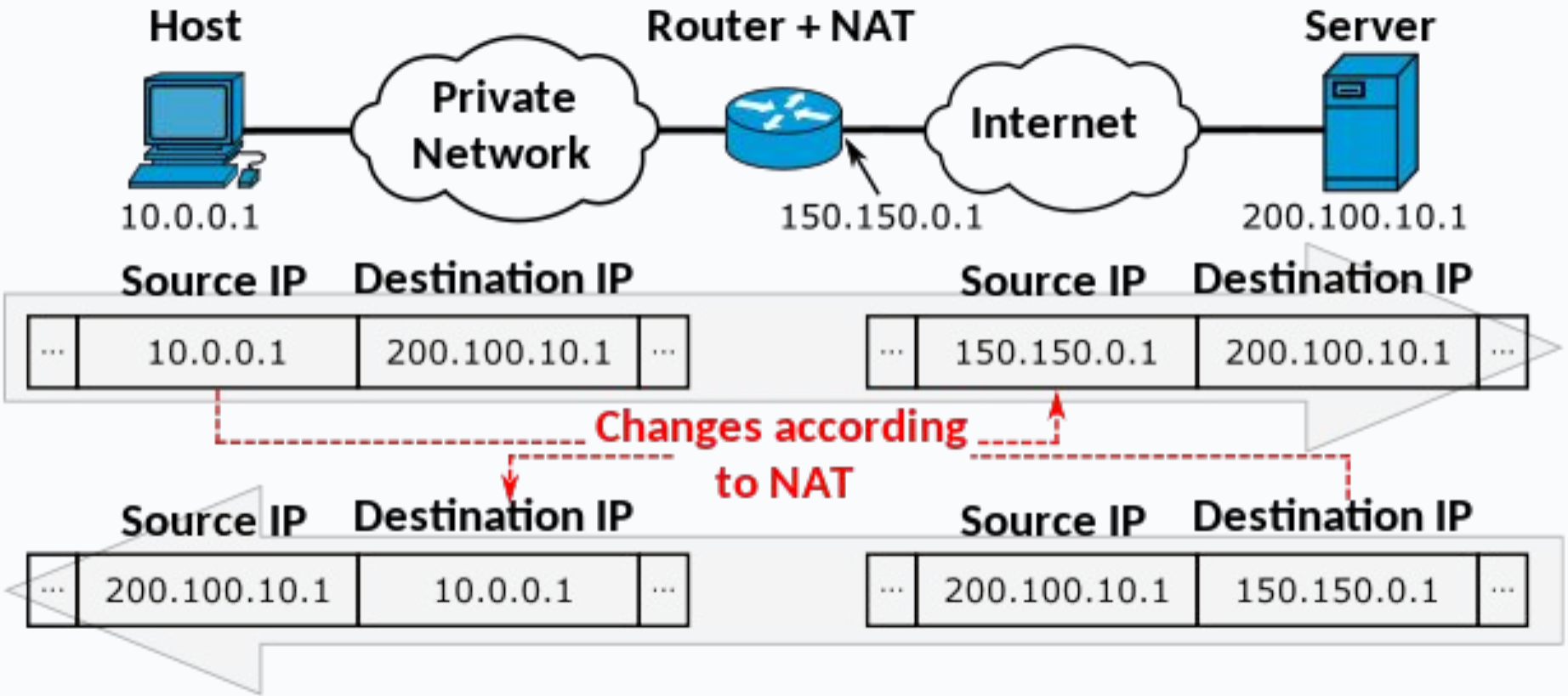
CGNAT no Debian com nftables

DebianDay Maceió 2024 - Henrique Silva

debian 12

NAT

- Network address translation(NAT) é um método de mapeamento de um espaço de endereço IP em outro, modificando as informações de endereço de rede no cabeçalho IP dos pacotes que estão em trânsito em um roteador.


Host
Router + NAT
Server
Private Network
Internet
10.0.0.1
150.150.0.1
200.100.10.1
Source IP
Destination IP
10.0.0.1
200.100.10.1
Source IP
Destination IP
150.150.0.1
200.100.10.1
Changes according to NAT
Source IP
Destination IP
200.100.10.1
10.0.0.1
Source IP
Destination IP
200.100.10.1
150.150.0.1

CGNAT

- Carrier-grade NAT (CGN ou CGNAT), também conhecido como large-scale NAT(LSN), é um tipo de tradução de endereço de rede (NAT) usado por ISPs em redes IPv4.

- Vantagens
  - Maximiza o uso de espaço limitado de endereço IPv4 público.
  - Mapeia dispositivos na rede para interface externa.
- Desvantagens
  - Pode criar um gargalo de desempenho que limita a escalabilidade.
  - Não resolve o problema de exaustão de endereço IPv4.
  - ...

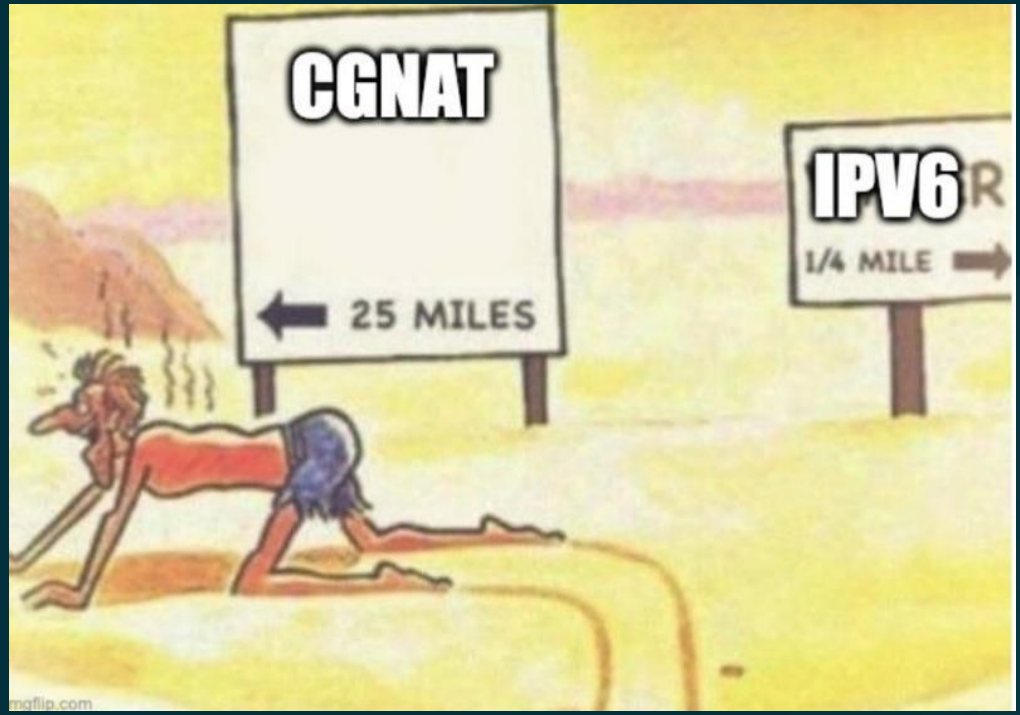
nftables

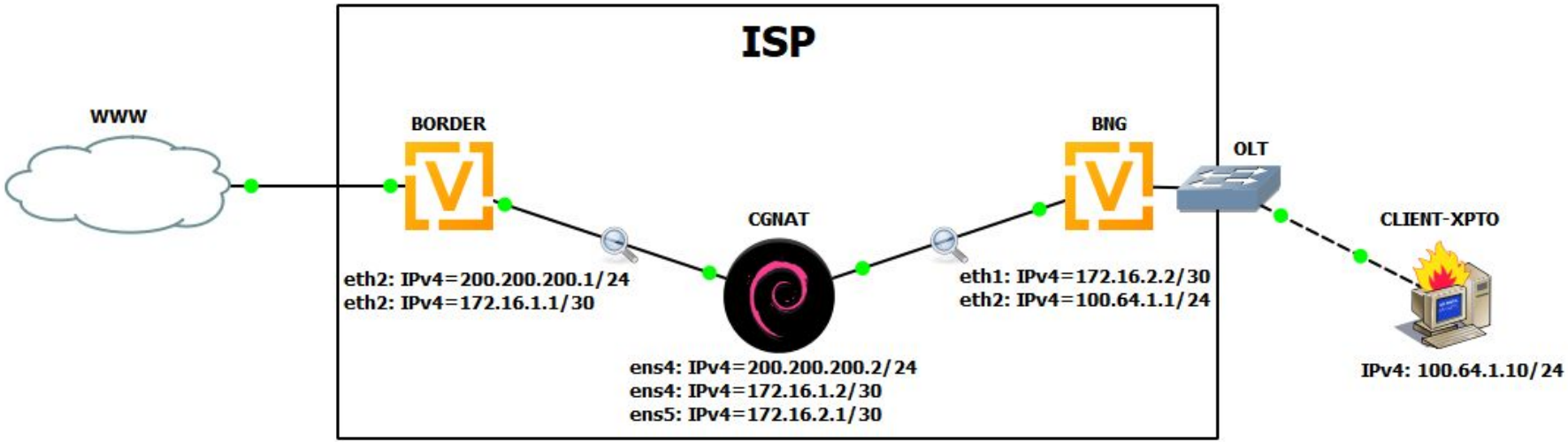
- nftables substitui o {ip,ip6,arp,eb}tables.

- Vantagens
  - Execução mais eficiente.
  - Sintaxe mais intuitiva.
  - Manipulação avançada de pacotes.
  - …

Casos de uso

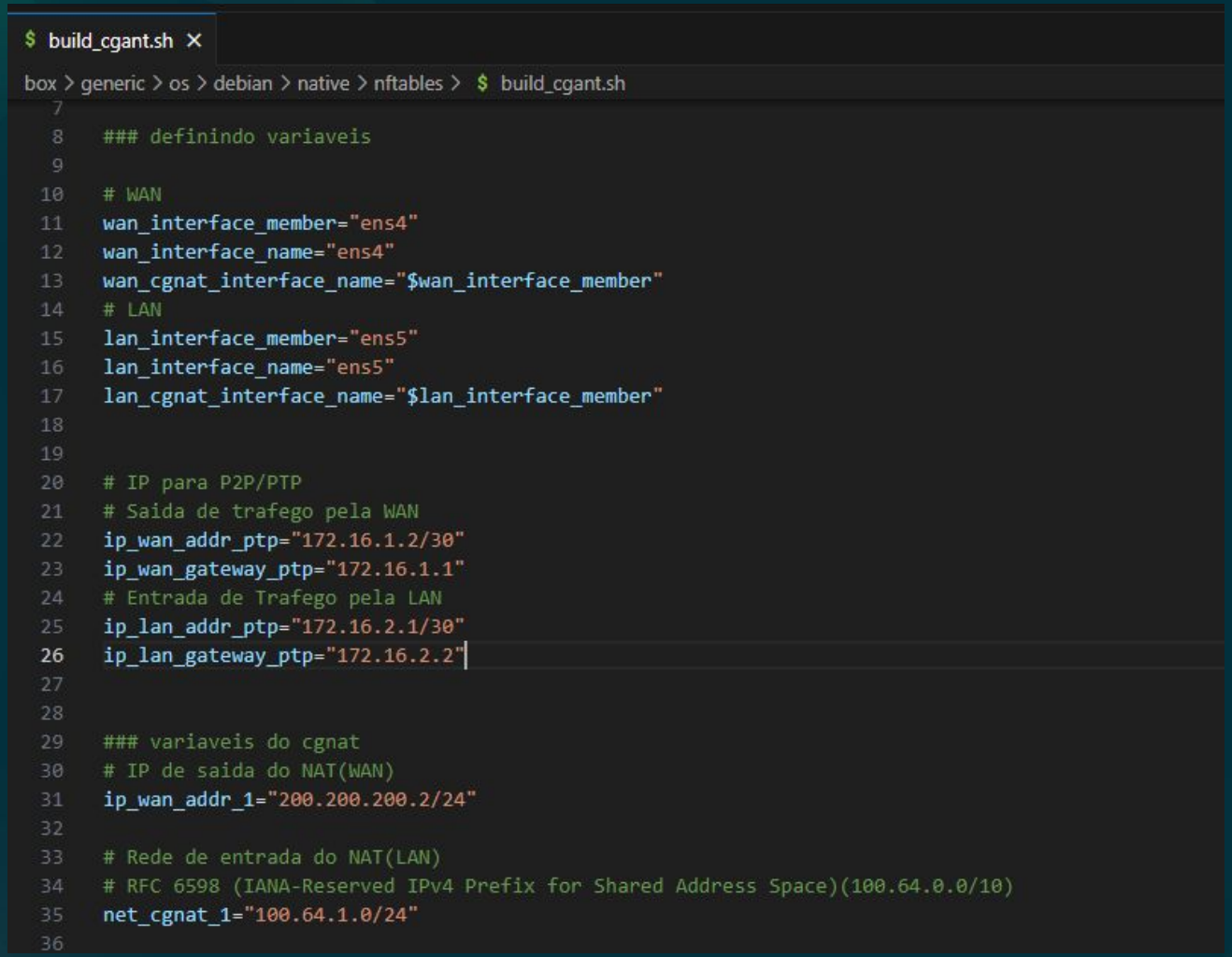
- Internet service provider (ISP).
- Manter redes IPv4 funcionando.



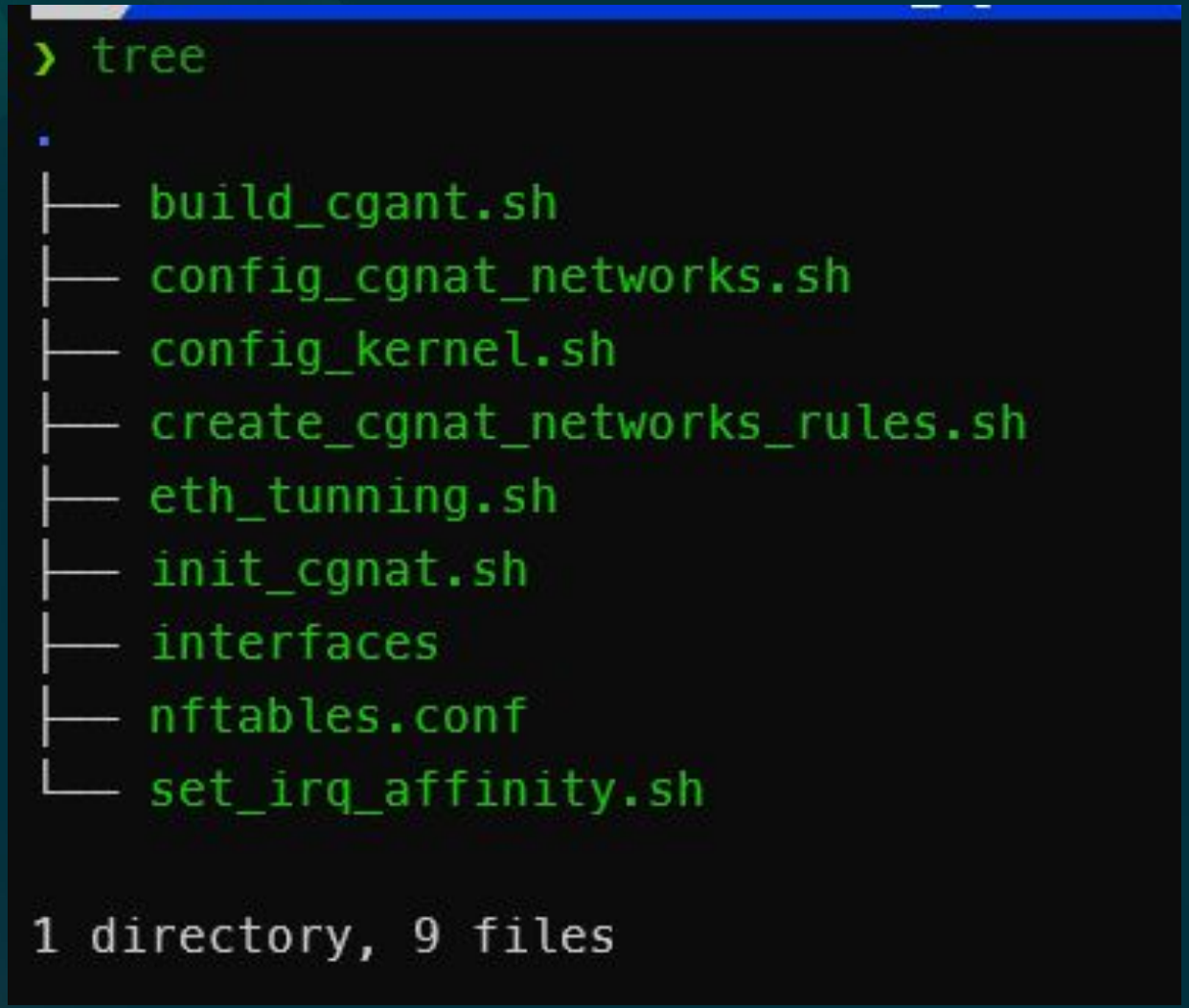

ISP
WWW
BORDER
BNG
OLT
CGNAT
CLIENT-XPTO
eth2: IPv4=200.200.200.1/24
eth2: IPv4=172.16.1.1/30
eth1: IPv4=172.16.2.2/30
eth2: IPv4=100.64.1.1/24
ens4: IPv4=200.200.200.2/24
ens4: IPv4=172.16.1.2/30
ens5: IPv4=172.16.2.1/30
IPv4: 100.64.1.10/24

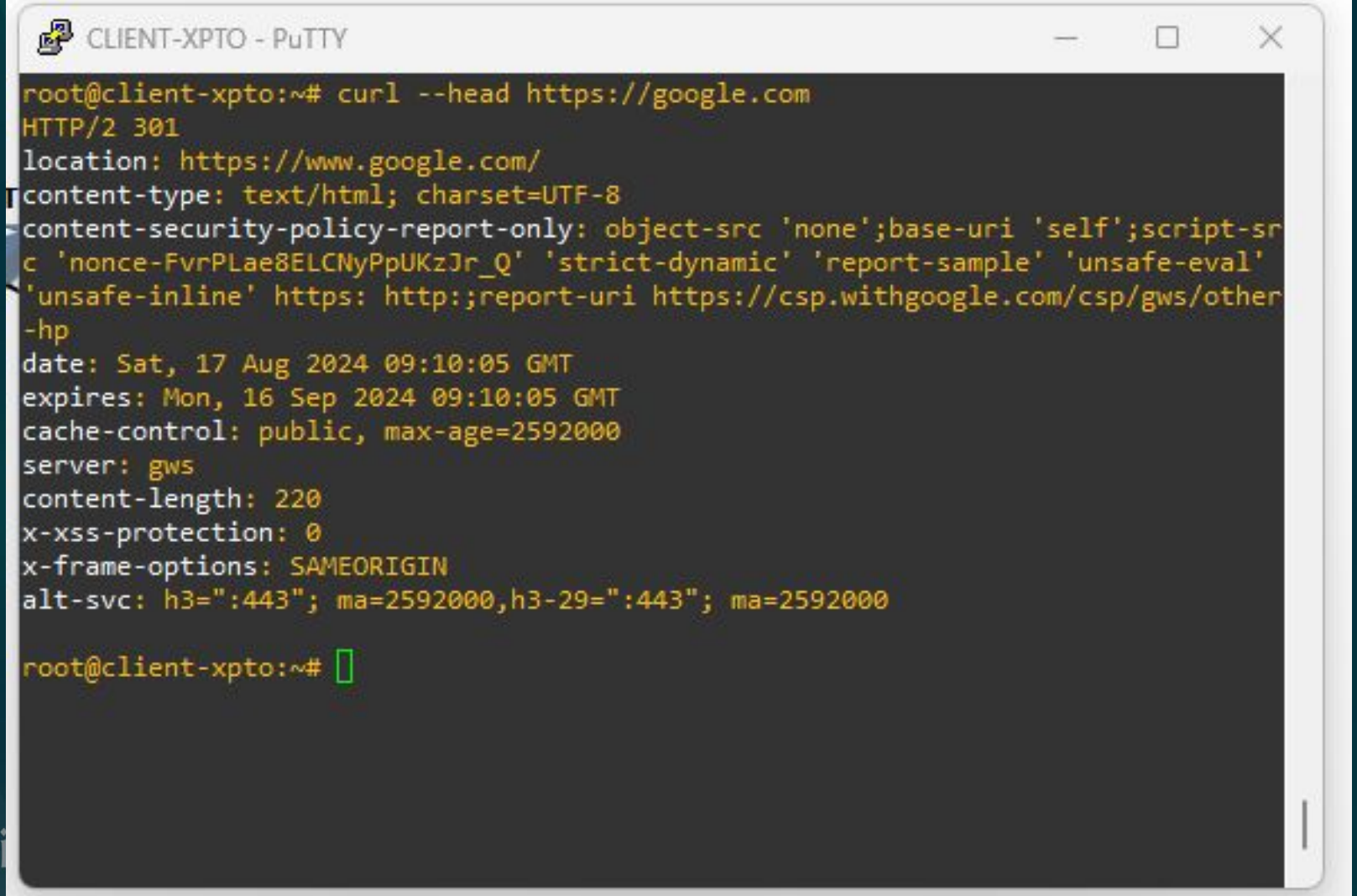
Criando o CGNAT

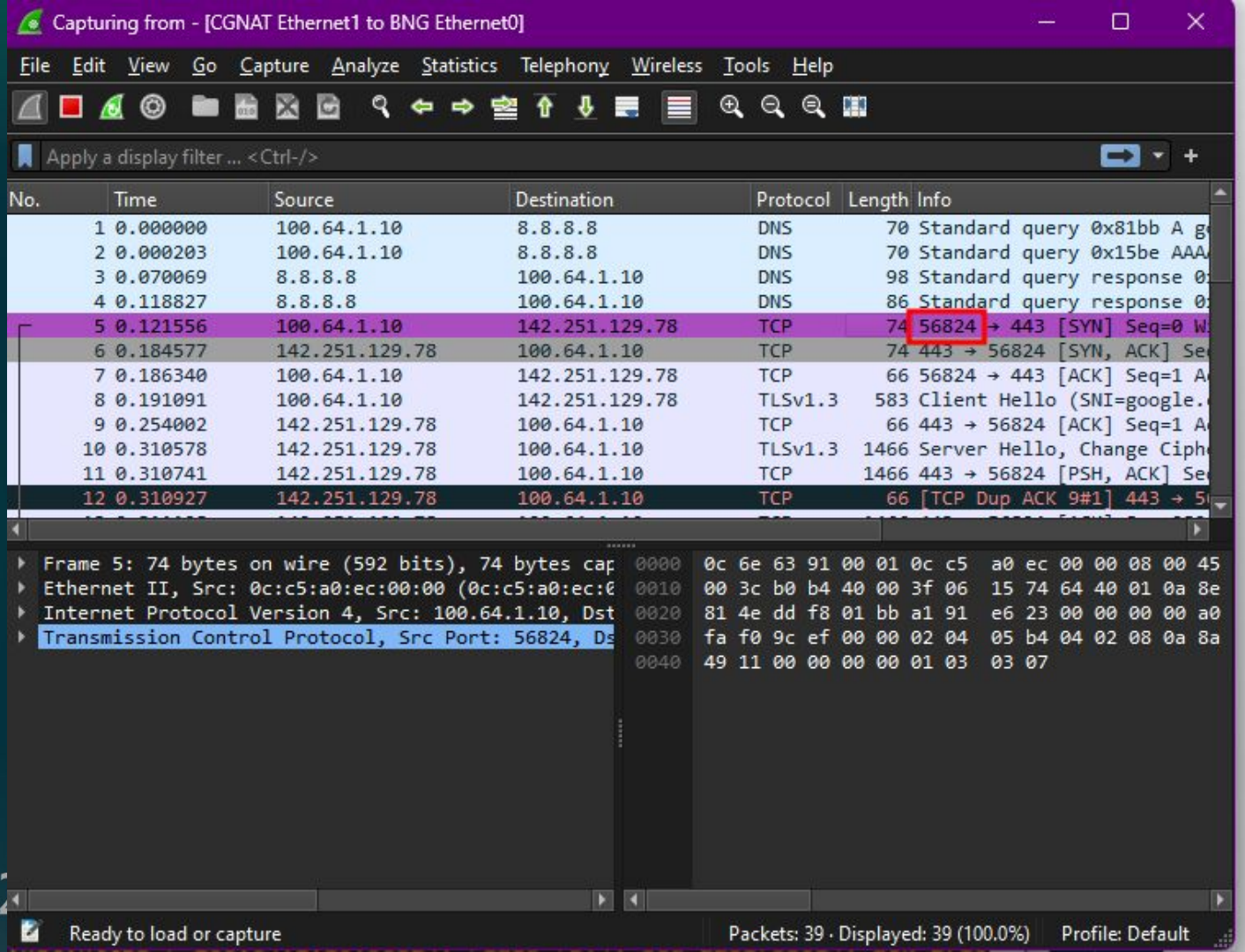
```
# git clone https://github.com/rick0x00/net_cgnat.git

# cd net_cgnat/box/generic/os/debian/native/nftables/

# vim build_cgant.sh

# bash build_cgant.sh
```

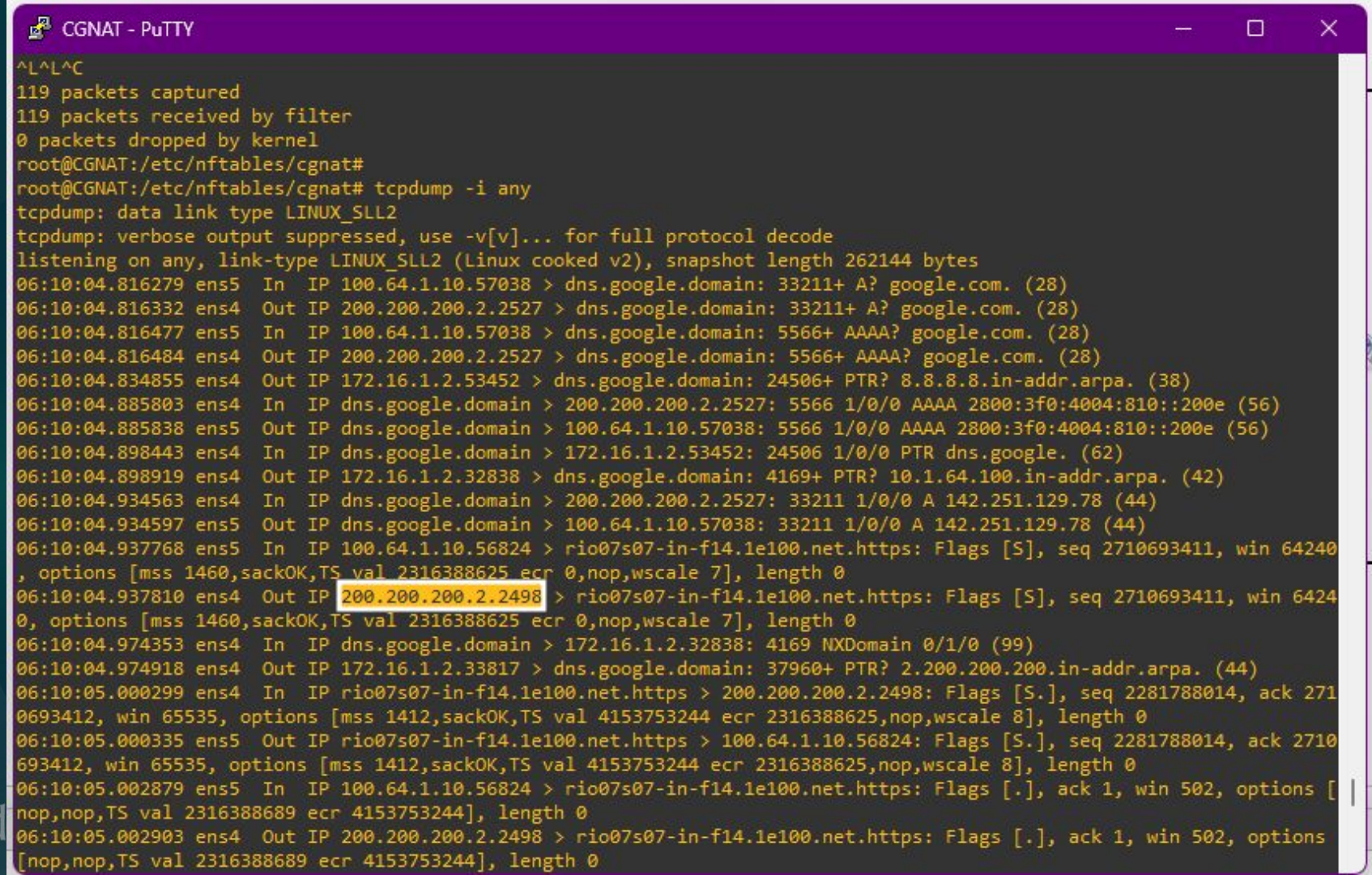

$ build_cgant.sh
box > generic > os > debian > native > nftables > $ build_cgant.sh
definindo variaveis
# WAN
wan_interface_member="ens4"
wan_interface_name="ens4"
wan_cgnat_interface_name="$wan_interface_member"
# LAN
lan_interface_member="ens5"
lan_interface_name="ens5"
lan_cgnat_interface_name="$lan_interface_member"
# IP para P2P/PTP
# Saida de trafego pela WAN
ip_wan_addr_ptp="172.16.1.2/30"
ip_wan_gateway_ptp="172.16.1.1"
# Entrada de Trafego pela LAN
ip_lan_addr_ptp="172.16.2.1/30"
ip_lan_gateway_ptp="172.16.2.2"
variaveis do cgnat
# IP de saida do NAT(WAN)
ip_wan_addr_1="200.200.200.2/24"
# Rede de entrada do NAT(LAN)
# RFC 6598 (IANA-Reserved IPv4 Prefix for Shared Address Space)(100.64.0.0/10)
net_cgnat_1="100.64.1.0/24"

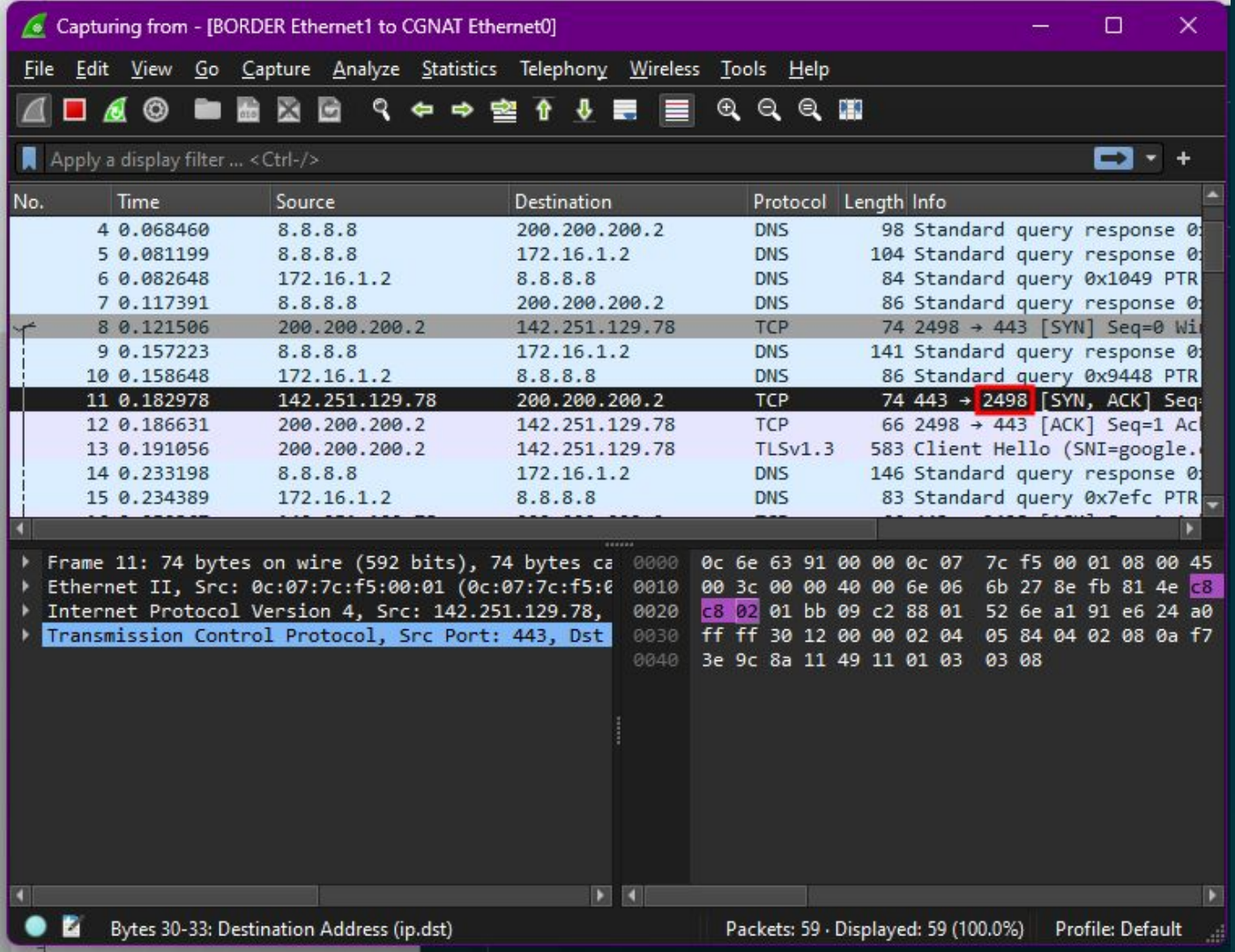
debian 12

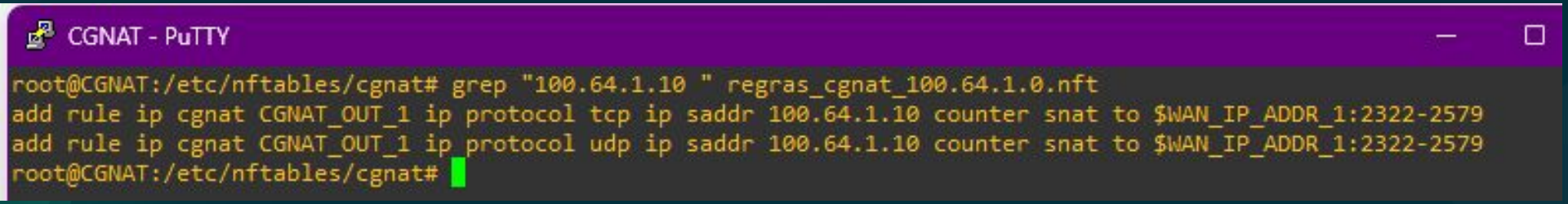

❯ tree
.
├── build_cgant.sh
├── config_cgnat_networks.sh
├── config_kernel.sh
├── create_cgnat_networks_rules.sh
├── eth_tunning.sh
├── init_cgnat.sh
├── interfaces
├── nftables.conf
└── set_irq_affinity.sh
1 directory, 9 files


CLIENT-XPTO - PuTTY
root@client-xpto:~# curl --head https://google.com
HTTP/2 301
location: https://www.google.com/
content-type: text/html; charset=UTF-8
content-security-policy-report-only: object-src 'none';base-uri 'self';script-src 'nonce-FvrPLae8ELCNyPpUKzJr_Q' 'strict-dynamic' 'report-sample' 'unsafe-eval' 'unsafe-inline' https: http:;report-uri https://csp.withgoogle.com/csp/gws/other-hp
date: Sat, 17 Aug 2024 09:10:05 GMT
expires: Mon, 16 Sep 2024 09:10:05 GMT
cache-control: public, max-age=2592000
server: gws
content-length: 220
x-xss-protection: 0
x-frame-options: SAMEORIGIN
alt-svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000
root@client-xpto:~#

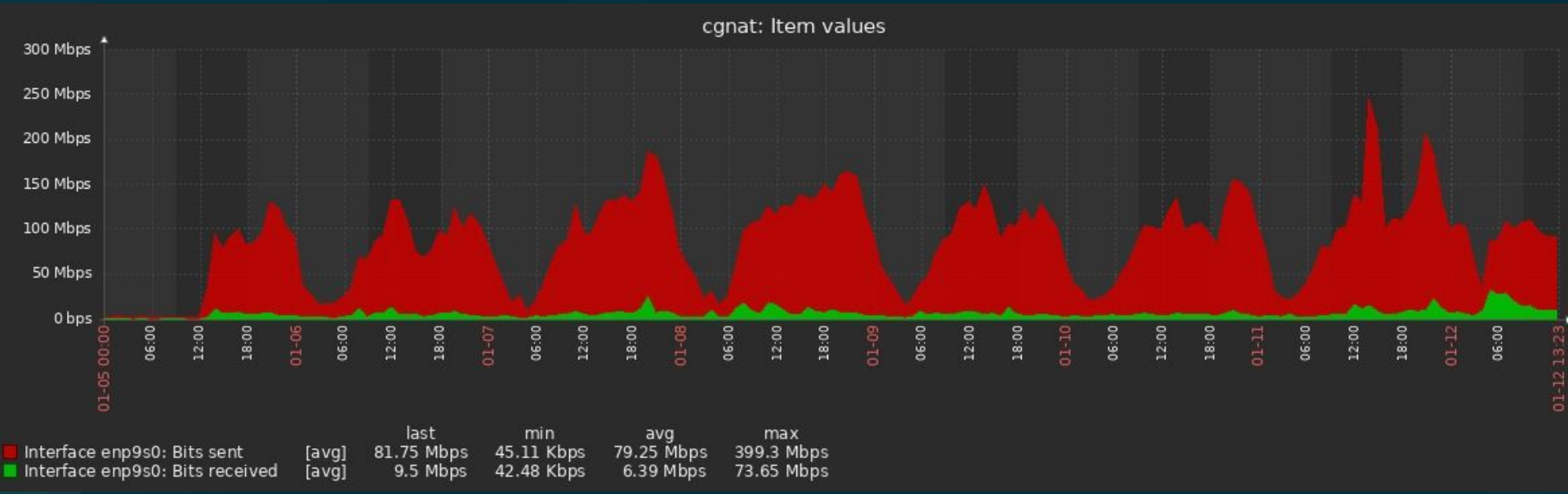

Capturing from - [CGNAT Ethernet1 to BNG Ethernet0]
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
No. Time Source Destination Protocol Length Info
1 0.000000 100.64.1.10 8.8.8.8 DNS 70 Standard query 0x81bb A g
2 0.000203 100.64.1.10 8.8.8.8 DNS 70 Standard query 0x15be AAA
3 0.070069 8.8.8.8 100.64.1.10 DNS 98 Standard query response 0
4 0.118827 8.8.8.8 100.64.1.10 DNS 86 Standard query response 0
5 0.121556 100.64.1.10 142.251.129.78 TCP 74 56824 → 443 [SYN] Seq=0 W
6 0.184577 142.251.129.78 100.64.1.10 TCP 74 443 → 56824 [SYN, ACK] Se
7 0.186340 100.64.1.10 142.251.129.78 TCP 66 56824 → 443 [ACK] Seq=1 A
8 0.191091 100.64.1.10 142.251.129.78 TLSv1.3 583 Client Hello (SNI=google.
9 0.254002 142.251.129.78 100.64.1.10 TCP 66 443 → 56824 [ACK] Seq=1 A
10 0.310578 142.251.129.78 100.64.1.10 TLSv1.3 1466 Server Hello, Change Ciph
11 0.310741 142.251.129.78 100.64.1.10 TCP 1466 443 → 56824 [PSH, ACK] Se
12 0.310927 142.251.129.78 100.64.1.10 TCP 66 [TCP Dup ACK 9#1] 443 → 5
Frame 5: 74 bytes on wire (592 bits), 74 bytes cap
Ethernet II, Src: 0c:c5:a0:ec:00:00 (0c:c5:a0:ec:0
Internet Protocol Version 4, Src: 100.64.1.10, Dst
Transmission Control Protocol, Src Port: 56824, Ds
0000 0c 6e 63 91 00 01 0c c5 a0 ec 00 00 08 00 45
0010 00 3c b0 b4 40 00 3f 06 15 74 64 40 01 0a 8e
0020 81 4e dd f8 01 bb a1 91 e6 23 00 00 00 00 a0
0030 fa f0 9c ef 00 00 02 04 05 b4 04 02 08 0a 8a
0040 49 11 00 00 00 00 01 03 03 07
Ready to load or capture
Packets: 39 · Displayed: 39 (100.0%)
Profile: Default


CGNAT - PuTTY
^L^L^C
119 packets captured
119 packets received by filter
0 packets dropped by kernel
root@CGNAT:/etc/nftables/cgnat#
root@CGNAT:/etc/nftables/cgnat# tcpdump -i any
tcpdump: data link type LINUX_SLL2
tcpdump: verbose output suppressed, use -v[v]... for full protocol decode
listening on any, link-type LINUX_SLL2 (Linux cooked v2), snapshot length 262144 bytes
06:10:04.816279 ens5 In IP 100.64.1.10.57038 > dns.google.domain: 33211+ A? google.com. (28)
06:10:04.816332 ens4 Out IP 200.200.200.2.2527 > dns.google.domain: 33211+ A? google.com. (28)
06:10:04.816477 ens5 In IP 100.64.1.10.57038 > dns.google.domain: 5566+ AAAA? google.com. (28)
06:10:04.816484 ens4 Out IP 200.200.200.2.2527 > dns.google.domain: 5566+ AAAA? google.com. (28)
06:10:04.834855 ens4 Out IP 172.16.1.2.53452 > dns.google.domain: 24506+ PTR? 8.8.8.8.in-addr.arpa. (38)
06:10:04.885803 ens4 In IP dns.google.domain > 200.200.200.2.2527: 5566 1/0/0 AAAA 2800:3f0:4004:810::200e (56)
06:10:04.885838 ens5 Out IP dns.google.domain > 100.64.1.10.57038: 5566 1/0/0 AAAA 2800:3f0:4004:810::200e (56)
06:10:04.898443 ens4 In IP dns.google.domain > 172.16.1.2.53452: 24506 1/0/0 PTR dns.google. (62)
06:10:04.898919 ens4 Out IP 172.16.1.2.32838 > dns.google.domain: 4169+ PTR? 10.1.64.100.in-addr.arpa. (42)
06:10:04.934563 ens4 In IP dns.google.domain > 200.200.200.2.2527: 33211 1/0/0 A 142.251.129.78 (44)
06:10:04.934597 ens5 Out IP dns.google.domain > 100.64.1.10.57038: 33211 1/0/0 A 142.251.129.78 (44)
06:10:04.937768 ens5 In IP 100.64.1.10.56824 > rio07s07-in-f14.1e100.net.https: Flags [S], seq 2710693411, win 64240
, options [mss 1460,sackOK,TS val 2316388625 ecr 0,nop,wscale 7], length 0
06:10:04.937810 ens4 Out IP 200.200.200.2.2498 > rio07s07-in-f14.1e100.net.https: Flags [S], seq 2710693411, win 6424
0, options [mss 1460,sackOK,TS val 2316388625 ecr 0,nop,wscale 7], length 0
06:10:04.974353 ens4 In IP dns.google.domain > 172.16.1.2.32838: 4169 NXDomain 0/1/0 (99)
06:10:04.974918 ens4 Out IP 172.16.1.2.33817 > dns.google.domain: 37960+ PTR? 2.200.200.200.in-addr.arpa. (44)
06:10:05.000299 ens4 In IP rio07s07-in-f14.1e100.net.https > 200.200.200.2.2498: Flags [S.], seq 2281788014, ack 271
0693412, win 65535, options [mss 1412,sackOK,TS val 4153753244 ecr 2316388625,nop,wscale 8], length 0
06:10:05.000335 ens5 Out IP rio07s07-in-f14.1e100.net.https > 100.64.1.10.56824: Flags [S.], seq 2281788014, ack 2710
693412, win 65535, options [mss 1412,sackOK,TS val 4153753244 ecr 2316388625,nop,wscale 8], length 0
06:10:05.002879 ens5 In IP 100.64.1.10.56824 > rio07s07-in-f14.1e100.net.https: Flags [.], ack 1, win 502, options [
nop,nop,TS val 2316388689 ecr 4153753244], length 0
06:10:05.002903 ens4 Out IP 200.200.200.2.2498 > rio07s07-in-f14.1e100.net.https: Flags [.], ack 1, win 502, options
[nop,nop,TS val 2316388689 ecr 4153753244], length 0


Capturing from - [BORDER Ethernet1 to CGNAT Ethernet0]
File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help
Apply a display filter ... <Ctrl-/>
No. Time Source Destination Protocol Length Info
4 0.068460 8.8.8.8 200.200.200.2 DNS 98 Standard query response 0
5 0.081199 8.8.8.8 172.16.1.2 DNS 104 Standard query response 0
6 0.082648 172.16.1.2 8.8.8.8 DNS 84 Standard query 0x1049 PTR
7 0.117391 8.8.8.8 200.200.200.2 DNS 86 Standard query response 0
8 0.121506 200.200.200.2 142.251.129.78 TCP 74 2498 → 443 [SYN] Seq=0 Wi
9 0.157223 8.8.8.8 172.16.1.2 DNS 141 Standard query response 0
10 0.158648 172.16.1.2 8.8.8.8 DNS 86 Standard query 0x9448 PTR
11 0.182978 142.251.129.78 200.200.200.2 TCP 74 443 → 2498 [SYN, ACK] Seq
12 0.186631 200.200.200.2 142.251.129.78 TCP 66 2498 → 443 [ACK] Seq=1 Ac
13 0.191056 200.200.200.2 142.251.129.78 TLSv1.3 583 Client Hello (SNI=google.
14 0.233198 8.8.8.8 172.16.1.2 DNS 146 Standard query response 0
15 0.234389 172.16.1.2 8.8.8.8 DNS 83 Standard query 0x7efc PTR
Frame 11: 74 bytes on wire (592 bits), 74 bytes ca
Ethernet II, Src: 0c:07:7c:f5:00:01 (0c:07:7c:f5:0
Internet Protocol Version 4, Src: 142.251.129.78,
Transmission Control Protocol, Src Port: 443, Dst
0000 0c 6e 63 91 00 00 0c 07 7c f5 00 01 08 00 45
0010 00 3c 00 00 40 00 6e 06 6b 27 8e fb 81 4e c8
0020 c8 02 01 bb 09 c2 88 01 52 6e a1 91 e6 24 a0
0030 ff ff 30 12 00 00 02 04 05 84 04 02 08 0a f7
0040 3e 9c 8a 11 49 11 01 03 03 08
Bytes 30-33: Destination Address (ip.dst)
Packets: 59 · Displayed: 59 (100.0%)
Profile: Default

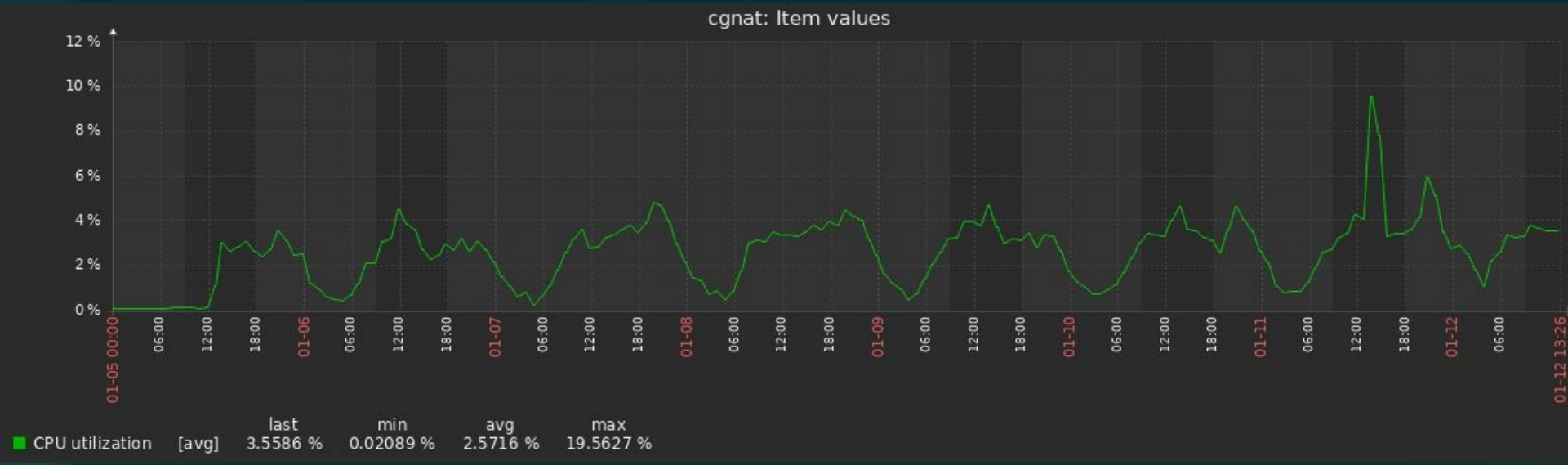

CGNAT - PuTTY
root@CGNAT:/etc/nftables/cgnat# grep "100.64.1.10 " regras_cgnat_100.64.1.0.nft
add rule ip cgnat CGNAT_OUT_1 ip protocol tcp ip saddr 100.64.1.10 counter snat to $WAN_IP_ADDR_1:2322-2579
add rule ip cgnat CGNAT_OUT_1 ip protocol udp ip saddr 100.64.1.10 counter snat to $WAN_IP_ADDR_1:2322-2579
root@CGNAT:/etc/nftables/cgnat#

Em produção: Uso de Rede

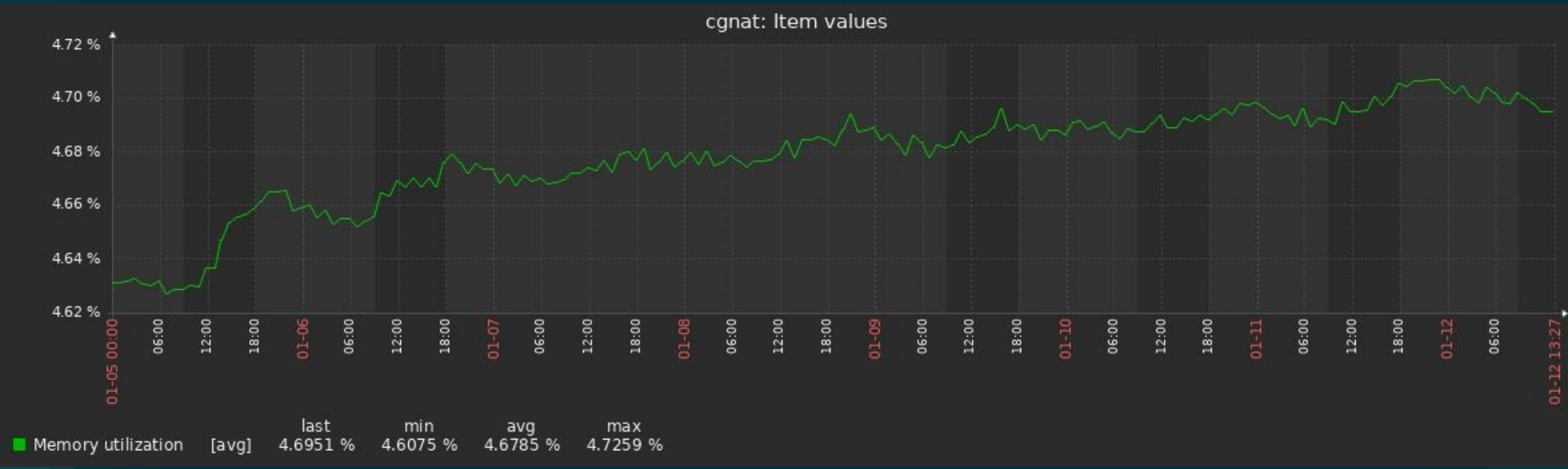


debian 12

Em produção: CPU utilization



Em produção: Memory utilization



Repositório

Referências

https://wiki.brasilpeeringforum.org/w/CGNAT_na_pratica

https://www.google.com/intl/en/ipv6/statistics.html#tab=ipv6-adoption

https://www.reddit.com/r/ipv6/comments/m0h38z/ipv6_illustrated_using_a_f1_meme/

https://wiki.nftables.org/wiki-nftables/index.php

https://en.wikipedia.org/wiki/Nftables