Easy7

opensource CGNAT implementation

Pascal Gloor @ SwiNOG-38

Init7

CGNAT on Fiber7? Seriously?

NO, NEVER !

Thanks for listening

THE END

Init7 - highspeed internet on fiber optics

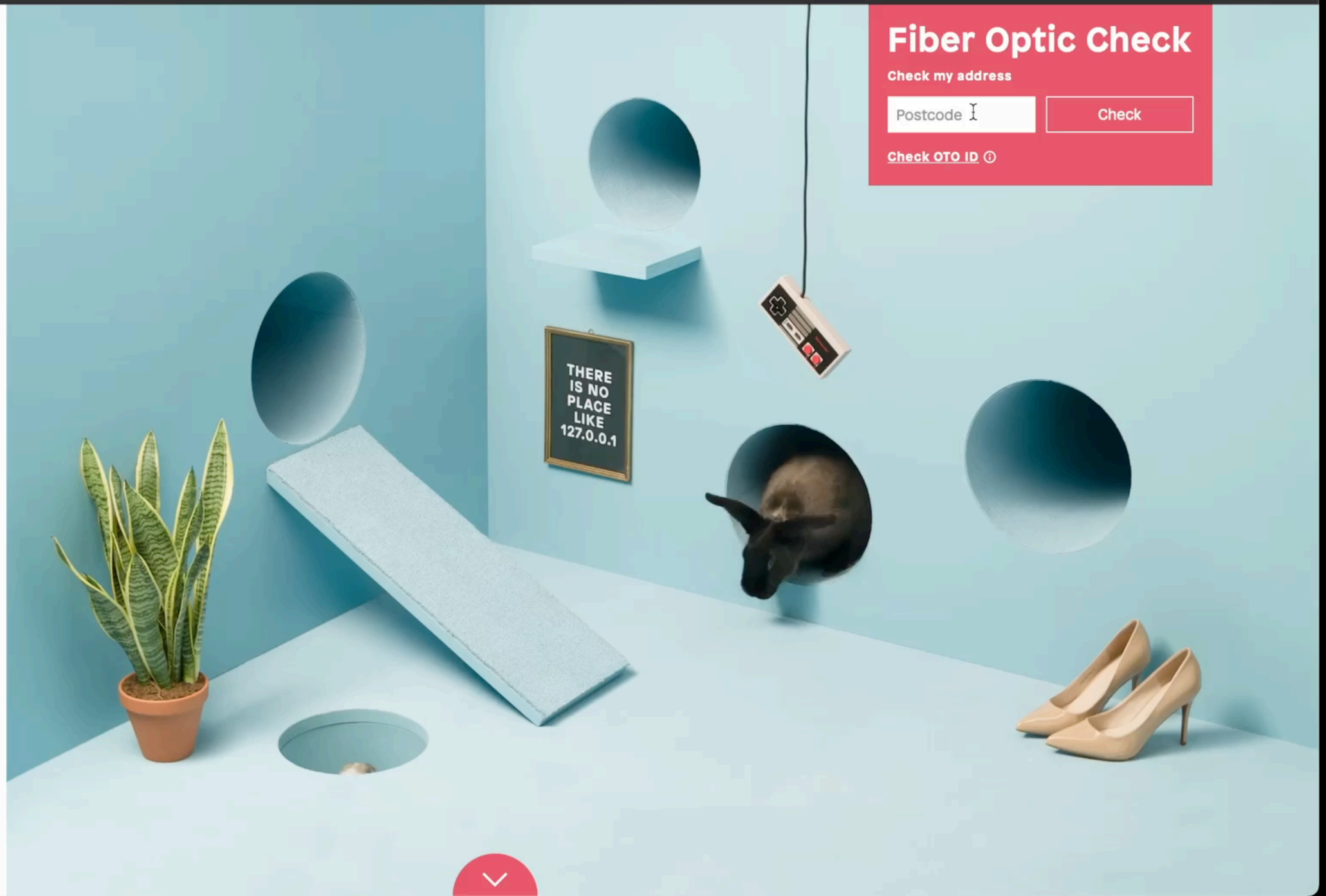
Init7

EN ▲

Private | Business

- Recommend Init7
- Internet
- TV
- Support / FAQ
- Infrastructure
- About Init7

nerd mode



Fiber Optic Check

Check my address

Postcode | Check

Check OTO ID

Easy7 is not Fiber7

- Designed as a « classic » Internet product
- 1Gbps only (10/25Gbps for Fiber7)
- IPv6 /56 « only »
- Includes a Managed CPE
- **No public IPv4 → CGNAT**

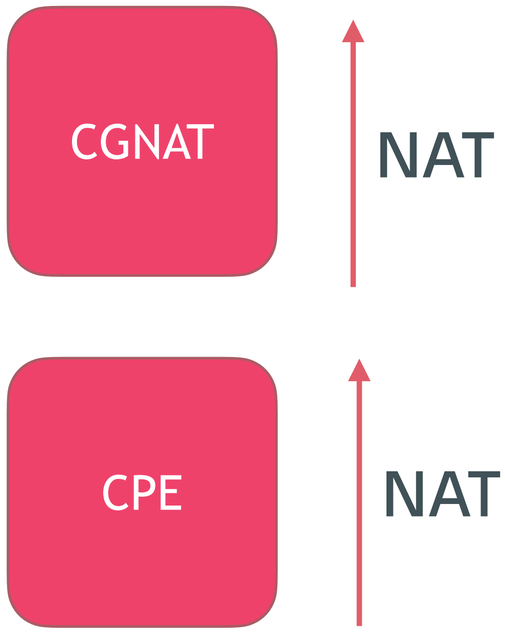
Init7

Carrier Grade NAT / NAT444

Shared Public IPv4

RFC6598 - IPv4 Prefix for Shared Address Space
100.64.0.0/10

RFC1918 - Address Allocation for Private Internets
10.0.0.0/8 - 172.16.0.0/12 - 192.168.0.0/16



Why CGNAT?

Because IPv4...

- is over.... right?
- is scarce
- is expensive
- is now a premium service

Init7

What does CGNAT break?

According to purevpn.com

- « CGNAT is a barrier usually employed by ISPs to **restrict bandwidth.** »
- « CGNAT is deployed by the ISPs to conserve bandwidth or **comply with a fair usage policy.** »
- « CGNAT also has some drawbacks, including making it **difficult for customers to communicate with devices on the public internet** and reducing the security of the ISP's network. »
- « CGNAT can cause jitter because it can cause packets to be dropped or **delivered out of order.** »

Init7

What does CGNAT break?

According to purevpn.com

- « CGNAT is a barrier usually employed by ISPs to **restrict bandwidth**. »
- « CGNAT is deployed by the ISPs to conserve bandwidth or **comply with a**
- « CGNAT also has some drawbacks, including making it **difficult for customers to communicate with devices on the public internet** and reducing the security of the ISP's network. »
- « CGNAT can cause jitter because it can cause packets to be dropped or **delivered out of order**. »

CGNAT cons?

- ❌ Port forwarding
- ❌ IP Blacklisting (IPs are shared, so it bad behaviour)
- ❌ Bittorrent seeding (leeching is fine)

Init7

CGNAT considerations

- Lawful Interception requirements aka BÜPF / LSCPT
- do 50-100Gbps
- support 50-100k users
- be georedundant (breaking tcp sessions on failover tolerated)
- be load-balanced / distributed / scalable
- be affordable

Init7

CGNAT considerations

- **Lawful Interception requirements aka BÜPF / LSCPT**
- do 50-100Gbps
- support 50-100k users
- be georedundant (breaking tcp sessions on failover tolerated)
- be load-balanced / distributed / scalable
- be affordable

Init7

Lawful Interception requirements aka BÜPF / LSCPT

VÜPF/OSCPT Art. 39 → Information Request Type IR_9_NAT:

A. the **source IP** address after or before the NAT translation procedure
B. the **source port** number after or before the NAT translation procedure
C. if required for identification, the public **destination IP** address;
D. if required for identification, the **destination port** number;
E. the type of the **transport protocol**
F. the **date and time** of the NAT translation procedure.

Init7

Lawful Interception requirements aka BÜPF / LSCPT

VÜPF/OSCPT Art. 18 → Obligations for the supply of information:

TSPs, **with the exception of those with reduced surveillance duties** ... must be able to provide the information specified in Articles 38 and 39...

Init7

Lawful Interception requirements aka BÜPF / LSCPT

VÜPF/OSCPT Art. 51 → TSP with reduced surveillance duties:

A. < 10 different surveillance targets in the last 12 months
B. Annual turnover (in CH) of < 100mio CHF

Init7

CGNAT considerations

- Lawful Interception requirements aka BÜPF / LSCPT
- **do 50-100Gbps**
- **support 50-100k users**
- be georedundant (breaking tcp sessions on failover tolerated)
- be load-balanced / distributed / scalable
- **be affordable**

Init7

CGNAT / Linux - netfilter it is!

- Software has « No costs »
- Well known and used
- Known behaviour
- Flexible
- Standard hardware / can do >50Gbps
- Specialised Hardware is completely overpriced!

Init7

Hardware

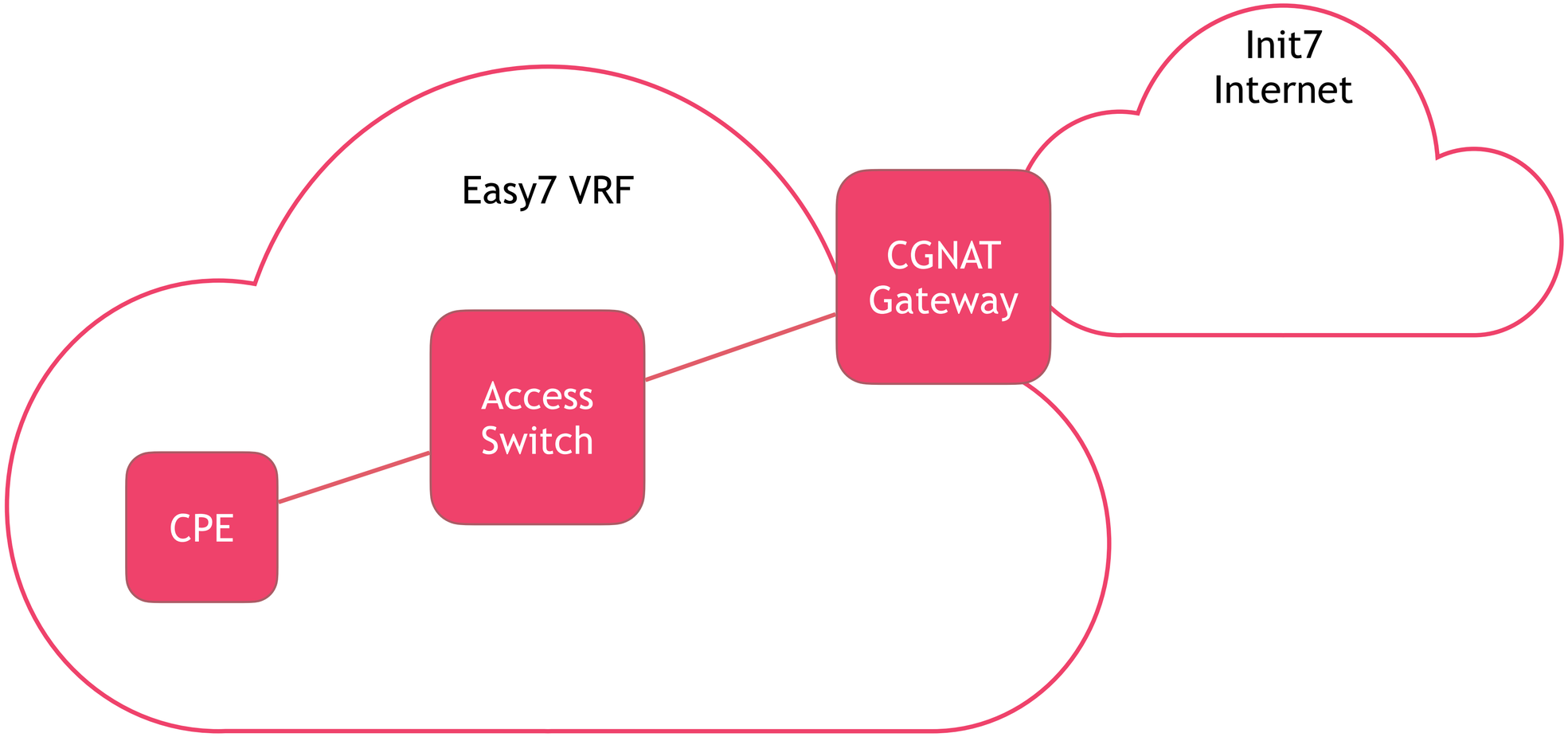
- AMD EPYC 7313 / 64GB RAM / 500GB redundant NVMe
- 100Gbps PCIe Mellanox adapter

CGNAT considerations

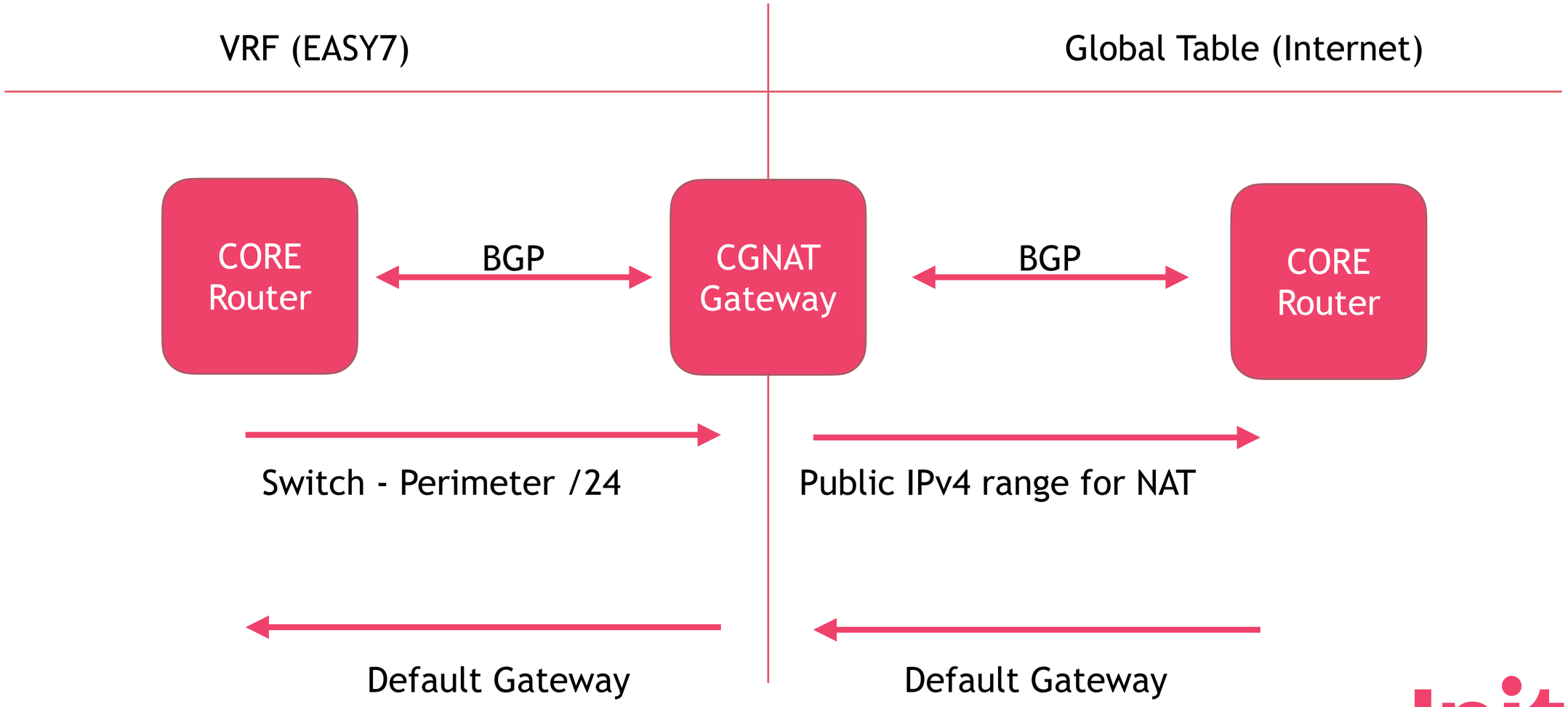
- Lawful Interception requirements aka BÜPF / LSCPT
- do 50-100Gbps
- support 50-100k users
- **be georedundant (breaking tcp sessions on failover tolerated)**
- **be load-balanced / distributed / scalable**
- be affordable

Init7

Easy7 - Architecture



Redundancy


VRF (EASY7)
Global Table (Internet)
CORE Router
BGP
CGNAT Gateway
BGP
CORE Router
Switch - Perimeter /24
Public IPv4 range for NAT
Default Gateway
Default Gateway

How to do NAT?

- Trivial! We've been NATting for over 20 years! Or.. is it?
- Even if only « on demand », identification must be possible
- But how many public IPs?
- How much « IP over subscribing » can really you do?
- Can you still identify your customers?
- Should a customer be « load-balanced » over multiple IPs?
- Customers won't be evenly balanced across the « inside » IP ranges

Init7

Not all NAT are equal

- netfilter NAT needs one source IP-port pair for each destination IP-port pair
- Not all customers have the same needs for « ports »
- Who needs so many ports?

NAT - the « tech » friendly way

```
iptables -t nat \
  -A POSTROUTING -s 100.64.0.0/10 -j MASQUERADE
OR
  -A POSTROUTING -s 100.64.0.0/10 -j SNAT --to 192.0.2.0-192.0.2.255
```

✅ Will evenly use source ports → HUGE BUCKET

✅ Doesn’t matter which switch has more/less customers

❌ Must log every session

Init7

NAT - the « Lawful Intercept » friendly way

```
iptables -t nat \
-A POSTROUTING -s 100.64.0.1 -p tcp -j SNAT --to 192.0.2.0:1024-2047
```

✅ Will uniquely identify customer, no logs needed!

❌ Creates a lot of unused buckets!

❌ Treats all customers equally, but they aren't

❌ Uses too many public IPv4 (or too few ports)

❌ thousands of rules... (3 x $pools_size)

Init7

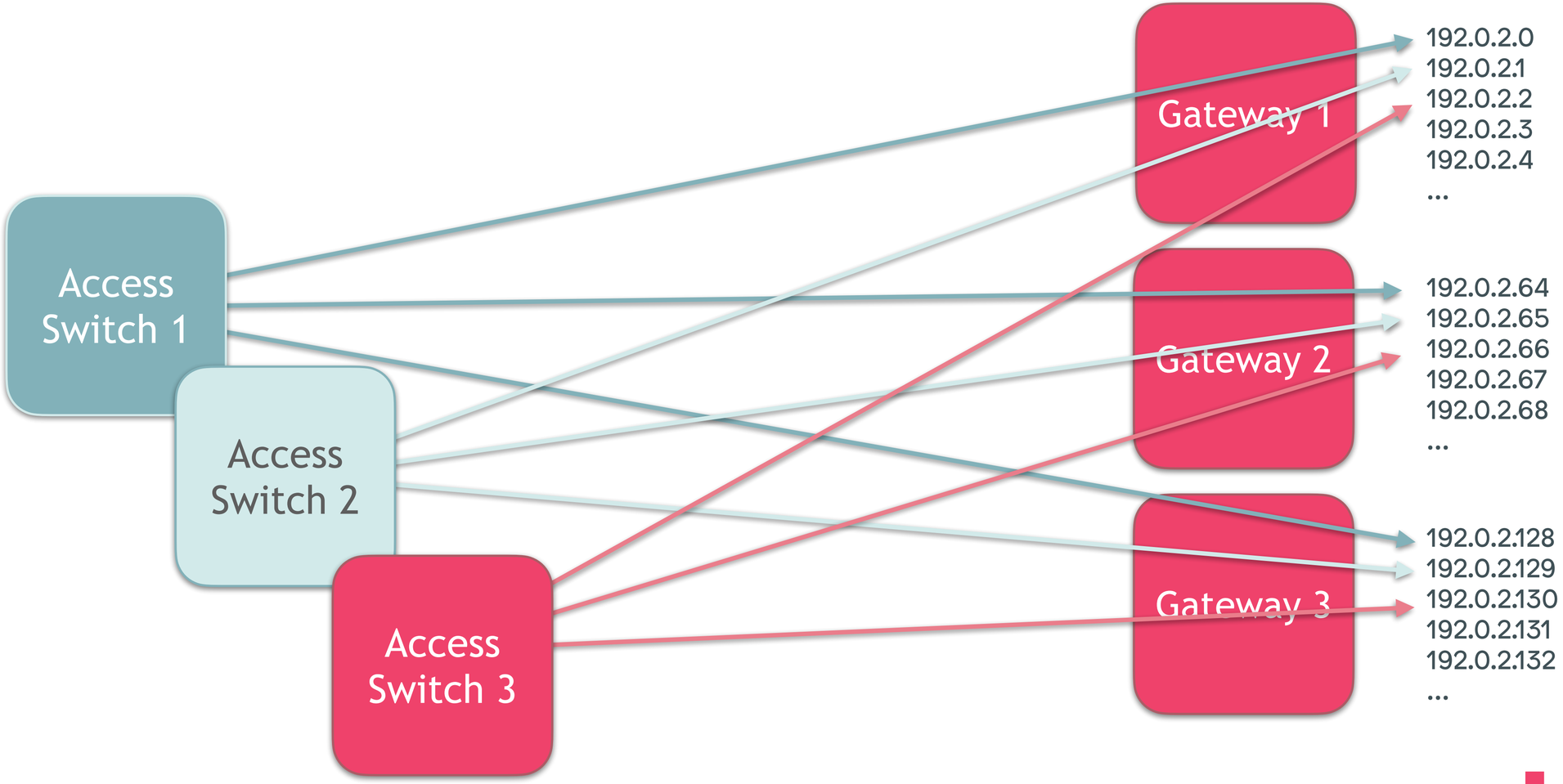
NAT - the Init7 way

- One 100.64.x.y/24 per switch
- One /27 per Gateway
- Mapping each switch to a public IP

NAT - the Init7 way

SWITCH Mgmt IP	SWITCH Perimeter	Gateway 1 IP	Gateway 2 IP
xx.yy.zz.0	100.64.0.0/24	192.0.2.0	192.0.2.128
xx.yy.zz.1	100.64.1.0/24	192.0.2.1	192.0.2.129
...			
xx.yy.zz.63	100.64.63.0/24	192.0.2.63	192.0.2.191
xx.yy.zz.64	100.64.64.0/24	192.0.2.0	192.0.2.128

NAT - the Init7 way



Netfilter tuning

```
net.netfilter.nf_conntrack_icmp_timeout=10
net.netfilter.nf_conntrack_max=1048576
net.netfilter.nf_conntrack_tcp_be_liberal=1
net.netfilter.nf_conntrack_tcp_loose=1
net.netfilter.nf_conntrack_tcp_timeout_established=7200
net.netfilter.nf_conntrack_udp_timeout=10
net.netfilter.nf_conntrack_udp_timeout_stream=300
```

Netfilter tuning

```
iptables -t filter -A FORWARD -p tcp
  -m connlimit
  --connlimit-above 8192
  --connlimit-mask 32
  --connlimit-saddr
  -j REJECT --reject-with tcp-reset
```


Init7

Conclusion

- It doesn’t have to cost so much
- It is professional
- Better control over your own infrastructure
- Keep it simple and stupid

Q&A

Thanks
and follow us

@init7
@kuenzler
@spale75