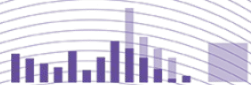


CSNOG 2022
2022-06-20

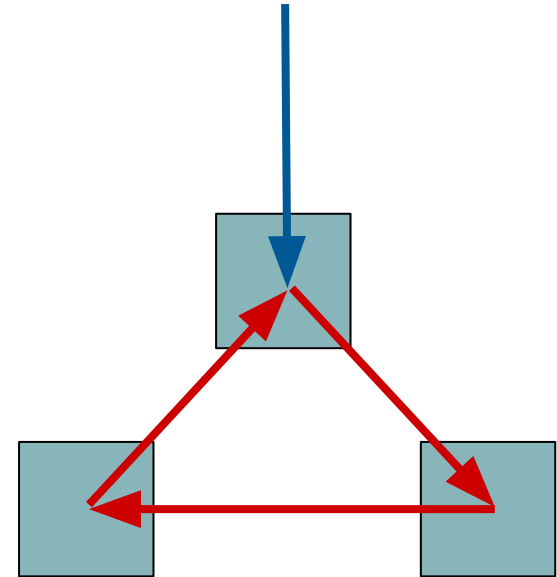
QRATOR
LABS

Routing loops

Alexander Zubkov
green@qrator.net



- packet is routed infinitely in a loop
- “protection” in IP protocol: TTL
 - maximum TTL 255
 - TTL can be modified
- TTL exceeded
 - real long path
 - loop
- traceroute/mtr



- packet is routed infinitely in a loop
- “protection” in IP protocol: TTL
 - maximum TTL 255
 - TTL can be modified
- TTL exceeded
 - real long path
 - loop
- traceroute/mtr



```

15:  xoxoxoxoxoxo .Ho .Ho .Ho .xoxoxoxoxoxo
16:  ooooxooooooxooo .V .ooooooxooooxoooo
17:  ooxooooooxooooo .MMM .ooooooxooxooxoo
18:  oooxooooooxooo .EEEE .oooxooooooxoooo
19:  ooooxooxooxoo .RRRRRRR .ooooooxooooox
20:  oxooooooxoo .RRRRRRRRR .oooxooooooxoo
21:  xoooxooooo .YYYYYYYYYYY .oooxooooxoo
22:  ooxooooooxooooo .CCC .ooooooxooxoo
23:  oooooxooo .HHHHHHHHHHHHH .oxoooxooo
24:  ooxooxoo .RRRRRRRRRRRRRRR .ooxooxoo
25:  oxoooxo .IIIIIIIIIIIIIIIIII .ooxooxo
26:  oooxoo .SSSSSSSSSSSSSSSSSS .ooxoooo
27:  oooxoooxoooooo .TTT .ooooooxoooooxoo
28:  ooxoo .MMMMMMMMMMMMMMMMMMMM .oooxo
29:  xxoo .AAAAAAAAAAAAAAAAAAAAAA .oxoo
30:  oxo .SSSSSSSSSSSSSSSSSSSSSSSSSS .ooo
31:  ooxooooooxoooo .XXX .ooooooxoooooxo
32:  oxooooooxoooo .XXX .ooooooxoooooxo
33:  Oh.the.weather.outside.is.frightful
34:  But.the.fire.is.so.delightful
35:  And.since.weve.no.place.to.go
36:  Let.It.Snow.Let.It.Snow.Let.It.Snow
...

```

```

6 Episode.IV
7 A.NEW.HOPE
8 It.is.a.period.of.civil.war
9 Rebel.spaceships
10 striking.from.a.hidden.base
11 have.won.their.first.victory
12 against.the.evil.Galactic.Empire
13 During.the.battle
14 Rebel.spies.managed
15 to.steal.secret.plans
16 to.the.Empires.ultimate.weapon
17 the.DEATH.STAR
18 an.armored.space.station
19 with.enough.power.to
20 destroy.an.entire.planet
21 Pursued.by.the.Empires
22 sinister.agents
23 Princess.Leia.races.home
24 aboard.her.starship
25 custodian.of.the.stolen.plans
26 that.can.save.her
27 people.and.restore
28 freedom.to.the.galaxy

```

```

29 0-----0
30 0-----0
31 0-----0
32 0-----0
33 0-----0
34 0-----0
35 0-----0
36 0-----0
37 0-----0
38 0-----0
39 0-----0
40 0-----0
41 0-----0
42 0-----0
43 0----0
44 0---0
45 0--0
46 0-0
47 0-0
48 00
49 I
50 By.Ryan.Werber
51 When.CCIEs.Get.Bored
52 read.more.at.beaglenetworks.net

```

```
28. | -- bad.horse
29. | -- bad.horse
30. | -- bad.horse
31. | -- bad.horse
32. | -- he.rides.across.the.nation
33. | -- the.thoroughbred.of.sin
34. | -- he.got.the.application
35. | -- that.you.just.sent.in
36. | -- it.needs.evaluation
37. | -- so.let.the.games.begin
38. | -- a.heinous.crime
39. | -- a.show.of.force
40. | -- a.murder.would.be.nice.of.course
41. | -- bad.horse
```

```
42. | -- bad.horse
43. | -- bad.horse
44. | -- he-s.bad
45. | -- the.evil.league.of.evil
46. | -- is.watching.so.beware
47. | -- the.grade.that.you.receive
48. | -- will.be.your.last.we.swear
49. | -- so.make.the.bad.horse.gleeful
50. | -- or.he-ll.make.you.his.mare
51. | -- o_o
52. | -- you-re.saddled.up
53. | -- there-s.no.recourse
54. | -- it-s.hi-ho.silver
55. | -- signed.bad.horse
```

1.	-- ge-9-1.ce26.ams-01.nl.leaseweb.net	0.0%
2.	-- xe-7-2-1.br01.ams-01.nl.leaseweb.net	0.0%
3.	-- be-101.bb03.ams-01.leaseweb.net	0.0%
4.	-- ge-1-1-1-501.edge00.nik.nl.hso-group.net	0.0%
5.	-- xe-0-1-2-0.pe-r-01.thn.uk.hso-group.net	0.0%
6.	-- et-0-2-0-0.pe-r-00.thn.uk.hso-group.net	0.0%
7.	-- xe-0-1-2-0.pe-r-00.gs1.uk.hso-group.net	0.0%
8.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
9.	-- ???	100.0
10.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
11.	-- ???	100.0
12.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
13.	-- ???	100.0
14.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
15.	-- ???	100.0
16.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
17.	-- ???	100.0
18.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
19.	-- ???	100.0
20.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%
21.	-- ???	100.0
22.	-- ae0-1203.edge00.sov.uk.hso-group.net	0.0%

- BGP, OSFP, ...
- loop protection
- BGP can converge for minutes
- mostly short-term loops
- but:
 - stuck routes
 - different masks: aggregation, partial table, ...

- unused IPs / NAT-pools
 - provider: routes 192.0.2.0/24 to the client
 - client:
 - 192.0.2.0/25 — is routed inside client's network
 - 192.0.2.128/25 — via default, goes to the provider
- long-term loops

- client side
 - null route assigned IP ranges
 - even when fully used
 - foolproof
 - vlan down → no route
- provider
 - filter spoof (BCP38) if possible

I: you have a routing loop, here is a traceroute

...

...

P: is not a routing loop “by definition”, address pool is not assigned to any specific interface, in your traceroute x.x.x.x is the final hop, it is a cosmetic problem, not worth the effort

The loop persisted for almost 5 years since then.



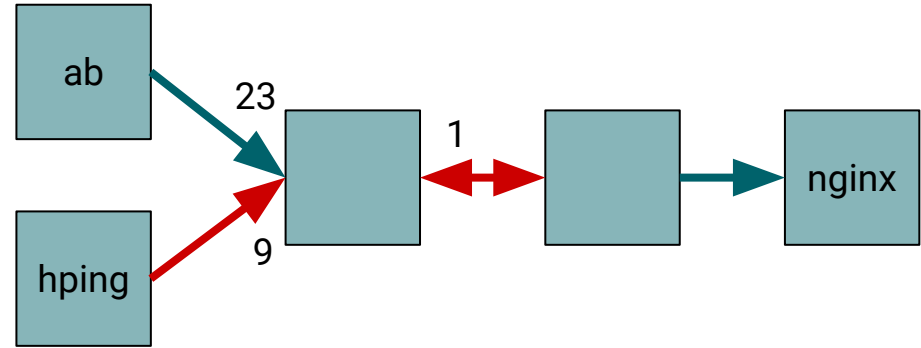
- link utilization
 - TTL >200, 2 hop loop → 100 times amplification
 - DoS target
 - charged traffic
- other problems
 - means of DoS-attacks
 - probe possibility of spoofing

- [Flooding Attacks by Exploiting Persistent Forwarding Loops](#) (2005), Jianhong Xia, Lixin Gao, Teng Fei
- [Weaponizing Middleboxes for TCP Reflected Amplification](#) (2021), Kevin Bock, Abdulrahman Alaraj, Yair Fax, Yair Fax, Eric Wustrow, Dave Levin
- [Using Loops Observed in Traceroute to Infer the Ability to Spoof](#) (2017), Qasim Lone, Matthew Luckie, Maciej Korczyński, Michel van Eeten
- [Hunting down the stuck BGP routes](#) (2021), Ben Cox
- [The Risks and Dangers of Amplified Routing Loops](#) (2021), Andree Toonk

From “Weaponizing Middleboxes for TCP Reflected Amplification”:

“We were unable to terminate the barrage of packets sent to us ... the traffic stopped after approximately six days ... We believe the reason they finally stopped was because the routing loop eventually dropped a packet.”

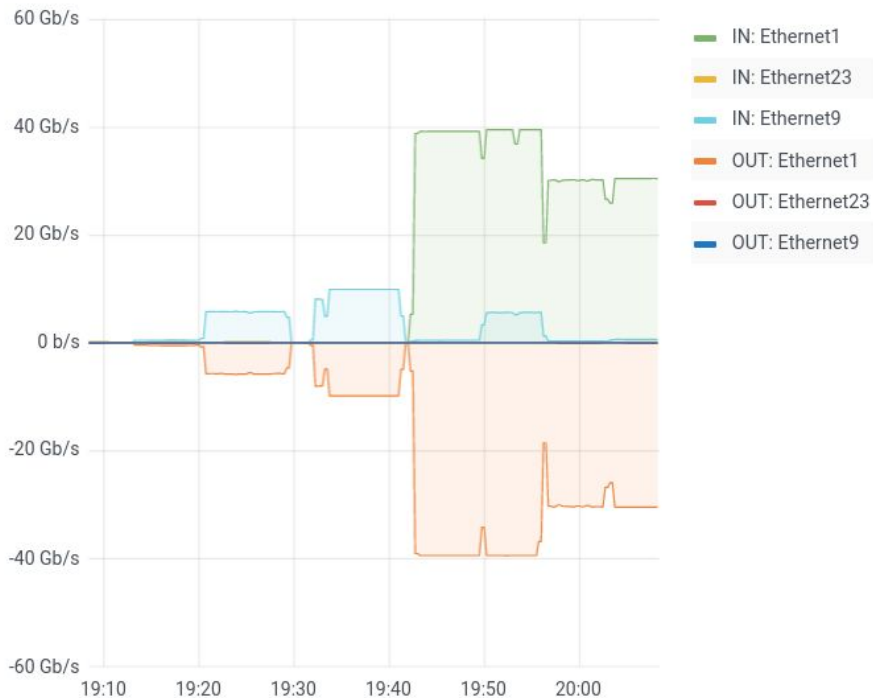
- 3 servers
 - nginx (server)
 - ab (client)
 - hping (flood)
- 2 L3-switches
- links
 - server-switch: 10G
 - switch-switch: 40G



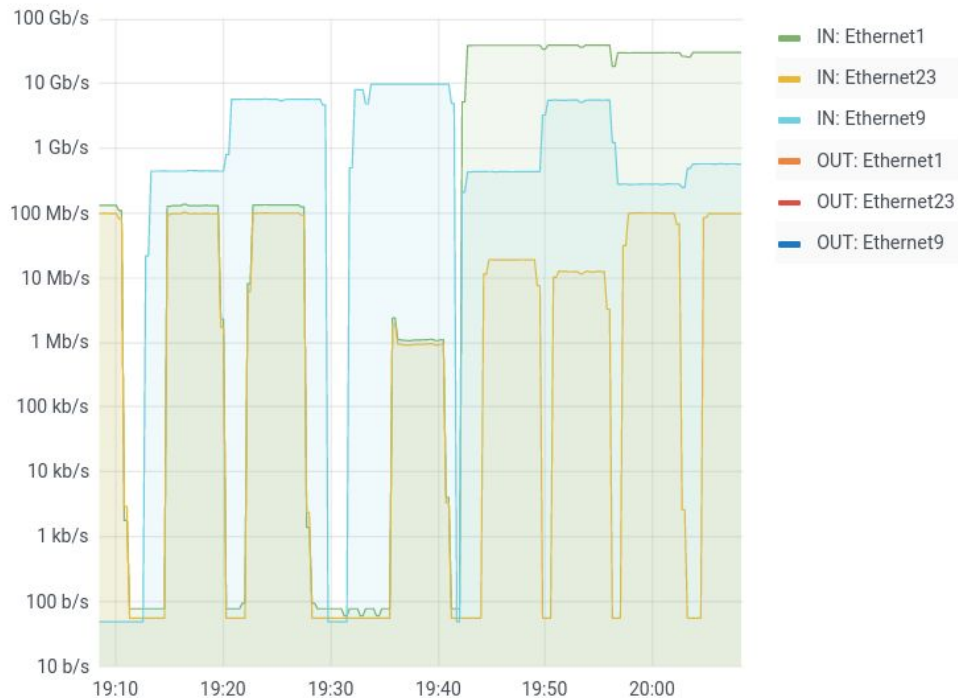
- client
 - `ab -r -t 300 -n 100000000 -c 100 ...`
 - 100 threads
 - for 5 minutes or up to 10M requests
- flood
 - `hping -2 -p 80 -t 220 [-d 1400] [-i u20|--flood] ...`
 - destination: server or loop
 - vary speed, packet size

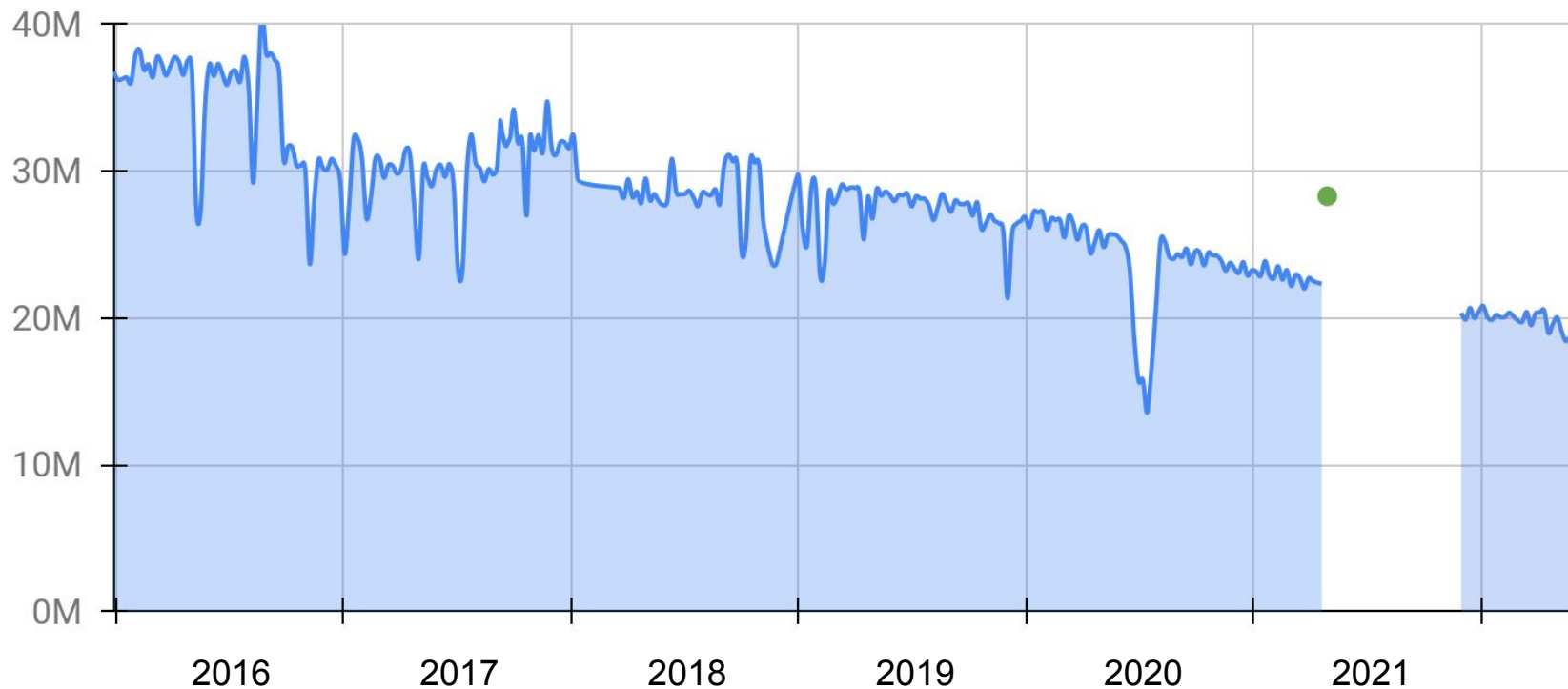
	flood dst	flood rate	link utilization	req/s	slowdown
	no flood	0	0	27.8k	
→	server	450M	450M	27.6k	
	server	5.8G	5.8G	28.1k	
	server	9.9G	9.9G	257	108
→	loop	440M	39G	5.3k	5.2
	loop	5.6G	39G	3.8k	8
	loop (small)	280M	30G	28.1k	
	loop (small)	570M	30G	27.5k	

Traffic



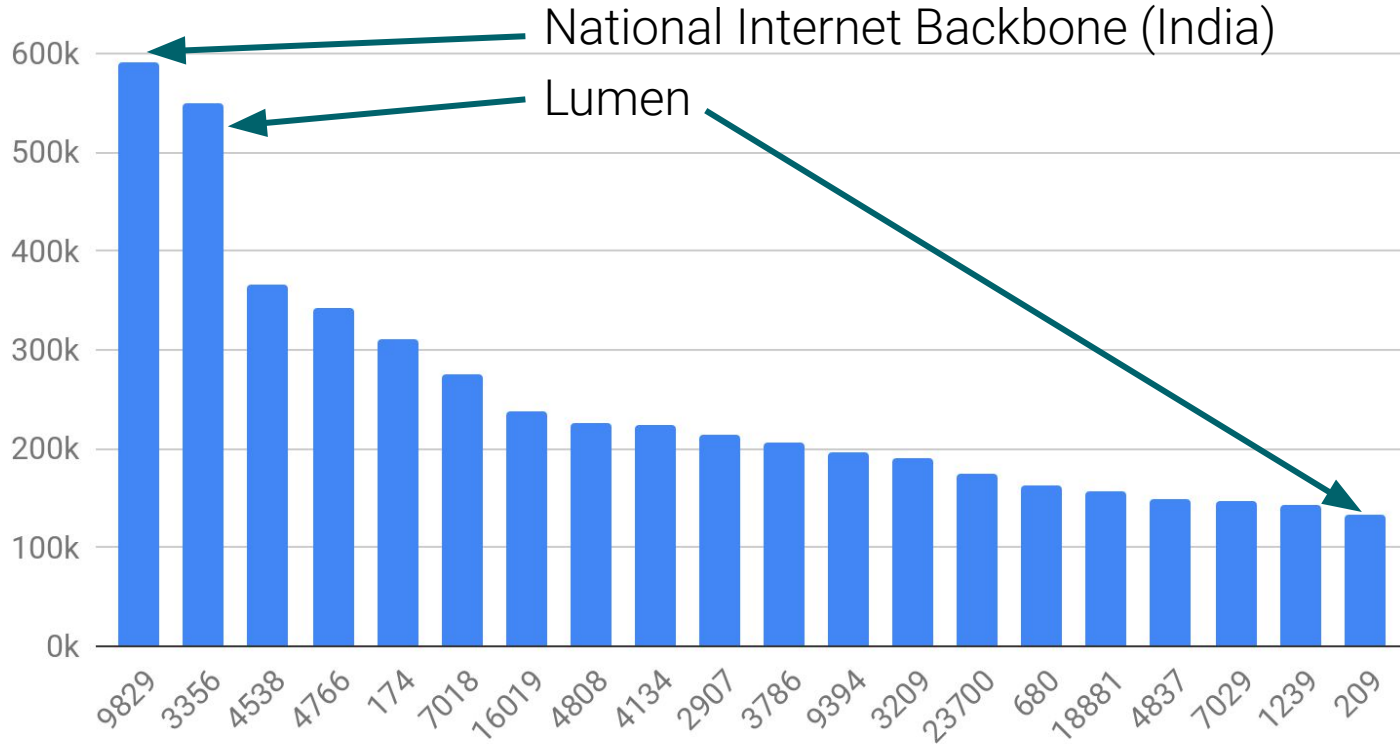
Traffic



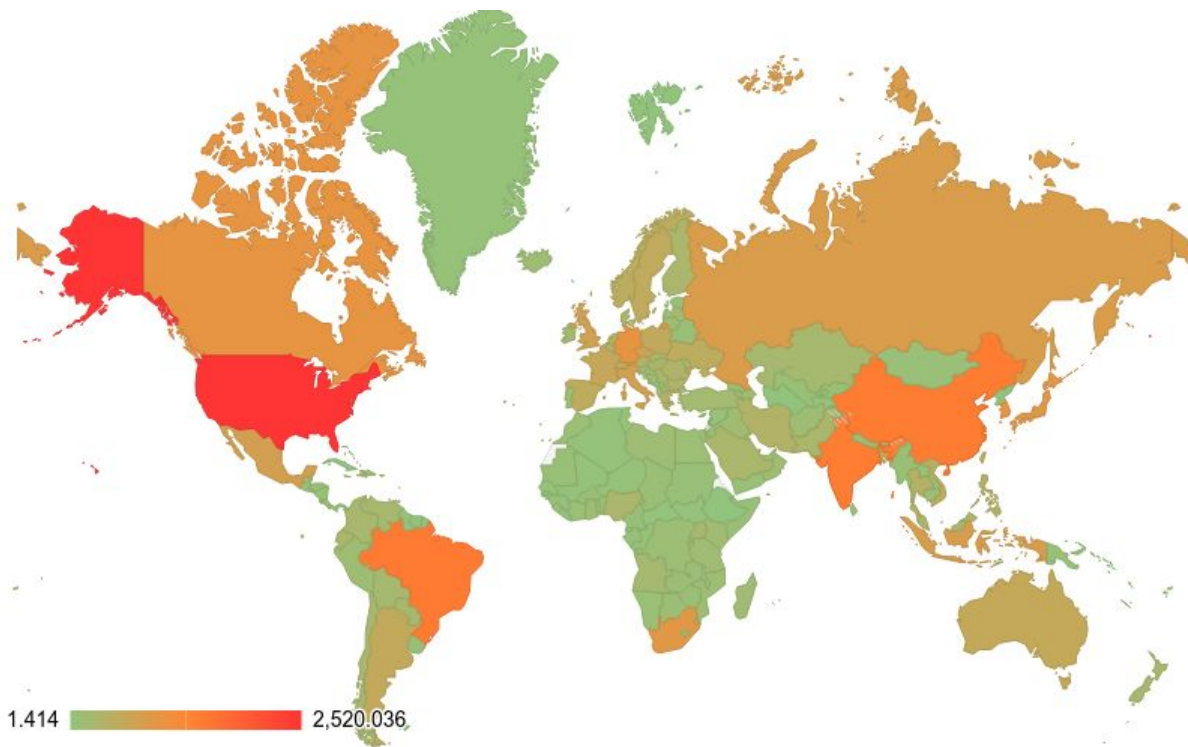


- scan Internet (IPv4 only)
 - ICMP, 4 probes TTL 125–128
 - TTL exceeded → loop
- more scans
- traceroutes
- tcpdumps

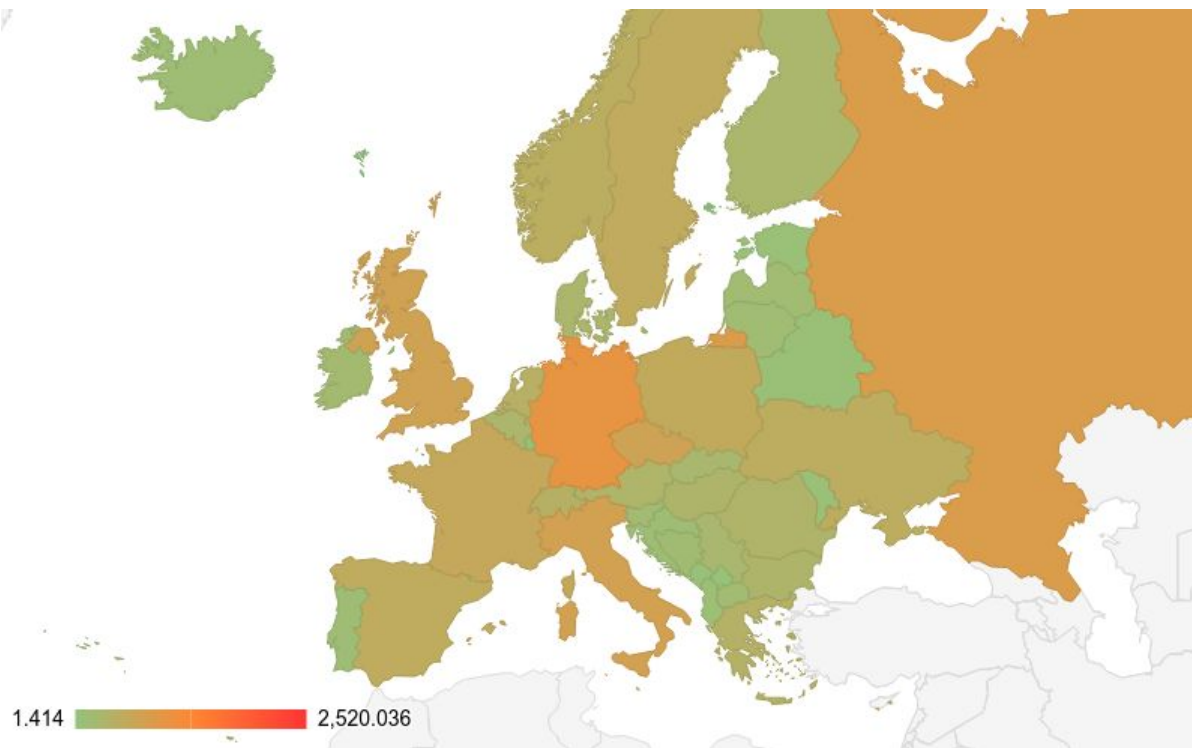
- 28.3M destination IPs (1.1% of all announced)
 - there are more actually
 - silent routers, random drops
- 25.5k destination AS numbers (35% of all announced)
 - top 20 AS numbers (<0.1%) – 17.6% of all loops
 - various ASes
 - CDN networks, DDoS mitigation services
 - well-known names



1	22.4%	USA
2	7.6%	Brazil
3	7.2%	China
4	7.0%	India
5	3.6%	South Korea
6	3.5%	Germany
7	3.4%	Canada
8	3.0%	Japan
9	2.9%	South Africa
10	2.4%	Russia

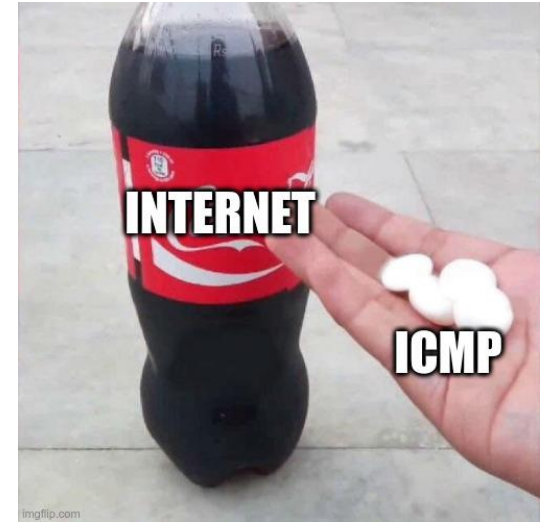


1	3.5%	Germany
2	2.4%	Russia
3	1.9%	Italy
4	1.8%	UK
5	1.7%	Czechia
6	1.2%	France
7	1.1%	Poland
8	1.0%	Sweden
9	0.9%	Spain
10	0.8%	Ukraine
...	...	...
22	0.24%	Slovakia



- 4 probes → 4.23 replies on average
- >100k replies for some probes
- most active router IP — 6.4% replies
- various patterns:
 - single IP
 - multiple IPs
 - multiple IPs, another finishing the series

- 4 probes → 4.23 replies on average
- >100k replies for some probes
- most active router IP — 6.4% replies
- various patterns:
 - single IP
 - multiple IPs
 - multiple IPs, another finishing the series



- unique loops
 - hard to match
 - missing replies
 - repeating IPs
 - multiple IPs
 - >100k unique loops
- 585k unique router IPs in 20.0k ASes

- from 1 to 34 hops
 - 2 hops – more than a half
- up to 7 ASes, up to 8 countries (rough)
 - border router IPs
 - geolocation issues
- loop time
 - 1 hop – up to 18 seconds
 - 2 hops – up to 8 seconds

~500 replies to a single ICMP request

```
02:44:44.628392 IP 130.88.118.250 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.628392 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.628552 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.628944 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.645634 IP 130.88.118.250 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.645872 IP 130.88.118.250 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.646357 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.647456 IP 130.88.118.250 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.647594 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.647921 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.651247 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.656137 IP 130.88.118.250 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:44.656902 IP 130.88.116.126 > 192.168.0.2: ICMP time exceeded in-transit, length 36
...
02:44:45.293832 IP 130.88.116.61 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:45.338794 IP 130.88.116.61 > 192.168.0.2: ICMP time exceeded in-transit, length 36
02:44:45.343614 IP 130.88.116.61 > 192.168.0.2: ICMP time exceeded in-transit, length 36
```

37 different IPs reply to a single ICMP request

```
12:39:35.137148 IP 91.250.182.149 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.137148 IP 91.250.182.112 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.137148 IP 91.250.186.48 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.137452 IP 91.250.187.6 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.137452 IP 91.250.178.85 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.137452 IP 91.250.181.59 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.138024 IP 91.250.184.116 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.138040 IP 91.250.186.241 > 192.168.0.2: ICMP time exceeded in-transit, length 36
...
12:39:35.138846 IP 91.250.187.170 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.138856 IP 91.250.176.57 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.138859 IP 91.250.176.59 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139178 IP 91.250.176.132 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139185 IP 91.250.185.108 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139187 IP 91.250.183.58 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139190 IP 91.250.180.243 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139192 IP 91.250.184.245 > 192.168.0.2: ICMP time exceeded in-transit, length 36
12:39:35.139384 IP 91.250.186.206 > 192.168.0.2: ICMP time exceeded in-transit, length 36
```

17-hops loop, 15 silent in a row

```
42. dis4-torontoxn_ae1.net.bell.ca
43. 67.69.163.242
44. ???
45. ???
46. ???
47. ???
48. ???
49. ???
50. ???
51. ???
52. ???
53. ???
54. ???
55. ???
56. ???
57. ???
58. ???
```


16 silent in a row, 33-hops loop

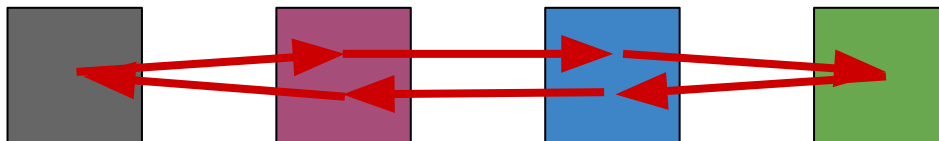
48. be-1412-cr12.pittsburgh.pa.ibone.comcast.net (96.110.38.158)	65. ???
49. be-301-cr14.350ecermak.il.ibone.comcast.net (96.110.39.157)	66. ???
50. be-1214-cs02.350ecermak.il.ibone.comcast.net (96.110.35.53)	67. ???
51. be-1211-cr11.350ecermak.il.ibone.comcast.net (96.110.35.6)	68. ???
52. be-302-cr11.1601milehigh.co.ibone.comcast.net (96.110.37.150)	69. ???
53. be-1311-cs03.1601milehigh.co.ibone.comcast.net (96.110.39.73)	70. ???
54. be-1314-cr14.1601milehigh.co.ibone.comcast.net (96.110.39.122)	71. ???
55. be-304-cr14.champa.co.ibone.comcast.net (96.110.39.13)	72. ???
56. be-1214-cs02.champa.co.ibone.comcast.net (96.110.37.245)	73. ???
57. be-1211-cr11.champa.co.ibone.comcast.net (96.110.37.198)	74. ???
58. be-302-cr01.seattle.wa.ibone.comcast.net (96.110.36.214)	75. ???
59. be-10846-pe01.seattle.wa.ibone.comcast.net (68.86.86.90)	76. ???
60. 96-87-8-90-static.hfc.comcastbusiness.net (96.87.8.90)	77. ???
61. border1.ae2-bbnet2.sef.pnap.net (63.251.160.68)	78. ???
62. usd-30.edge1.sef.pnap.net (64.94.137.194)	79. ???
63. core.sea.dedicated.com (167.160.89.2)	80. ???
64. 167.160.89.18	

34-hops loop in a single network

- | | |
|---|---|
| 12. sto-ste-dr1-ar1.sto-vas29-dr1.bahnhof.net | 29. sod-hfv-ar1.sod-sdt-dr1.bahnhof.net |
| 13. sto-vas29-dr1-ar1.sto-kn4-ar1.bahnhof.net | 30. sod-sdt-dr1.sod-sdt-dr2.bahnhof.net |
| 14. sto-kn4-dr1.sto-kn5-dr1.bahnhof.net | 31. sod-sdt-dr2.bot-tb-dr1.bahnhof.net |
| 15. sto-kn5-dr1.sto-kn5-dr2.bahnhof.net | 32. bot-tb-dr1.bot-tul-dr1.bahnhof.net |
| 16. sto-kn5-dr2.sto-sh-dr1.bahnhof.net | 33. bot-tul-dr1.hde-hud-dr1.bahnhof.net |
| 17. sto-sh-dr1.sto-ss-dr1.bahnhof.net | 34. hde-hud-dr1.sto-orby-dr1.bahnhof.net |
| 18. sto-ss-dr1.sto-ens-dr1.bahnhof.net | 35. sto-orby-dr1.sto-ars-dr1.bahnhof.net |
| 19. sto-ens-dr1.sto-ars-dr1.bahnhof.net | 36. sto-ars-dr1.sto-ens-dr1.bahnhof.net |
| 20. sto-ars-dr1.sto-orby-dr1.bahnhof.net | 37. sto-ens-dr1.sto-ss-dr1.bahnhof.net |
| 21. sto-orby-dr1.hde-hud-dr1.bahnhof.net | 38. sto-ss-dr1.sto-sh-dr1.bahnhof.net |
| 22. hde-hud-dr1.bot-tul-dr1.bahnhof.net | 39. sto-sh-dr1.sto-kn5-dr2.bahnhof.net |
| 23. bot-tul-dr1.bot-tb-dr1.bahnhof.net | 40. sto-kn5-ar1.sto-soder-dr1.bahnhof.net |
| 24. bot-tb-dr1.sod-sdt-dr2.bahnhof.net | 41. sto-soder-dr1.sto-soder-dr2.bahnhof.net |
| 25. sod-sdt-dr2.sod-sdt-dr1.bahnhof.net | 42. sto-soder-dr1.sto-pio-dr1.bahnhof.net |
| 26. sod-sdt-dr1.sod-hfv-ar1.bahnhof.net | 43. sto-pio-dr1.sto-ste-dr3.bahnhof.net |
| 27. bka-tgv1-ar1.sod-hfv-ar1.bahnhof.net | 44. sto-ste-dr2.sto-ste-dr3.bahnhof.net |
| 28. sod-hfv-ar1.bka-tgv1-ar1.bahnhof.net | 45. sto-ste-dr2.sto-ste-dr1.bahnhof.net |

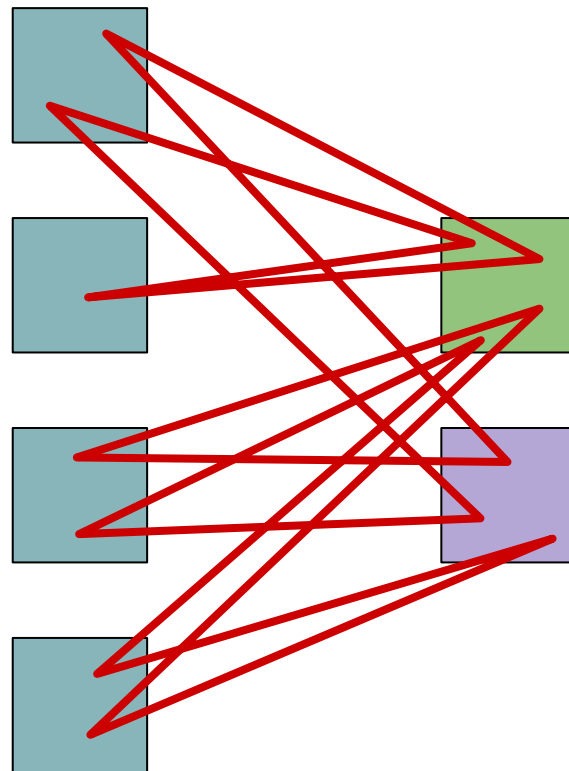
“flat” loops

```
15. 140.156.broadband18.io1.cz (109.81.156.140)
16. 194.228.159.255
17. 194.228.190.142
18. 194.228.190.141
19. 194.228.190.142
20. 194.228.159.255
```



“crazy” loop

37. AS23498	69.77.169.22
38. AS174	38.131.181.74
39. AS23498	158.106.103.38
40. AS174	38.131.181.74
41. AS23498	69.77.169.22
42. AS395965	69.194.36.66
43. AS23498	158.106.103.62
44. AS174	38.131.181.74
45. AS23498	69.77.169.42
46. AS395965	69.194.36.66
47. AS23498	69.77.169.42
48. AS174	38.131.181.74
49. AS23498	158.106.103.62
50. AS395965	69.194.36.66



- Qrator.Radar
 - <https://radar.qrator.net/>
 - loops
 - amplifiers
 - BGP incidents
- Alexander Zubkov
 - green@qrator.net