Combating DDoS and why peering is important in Asia

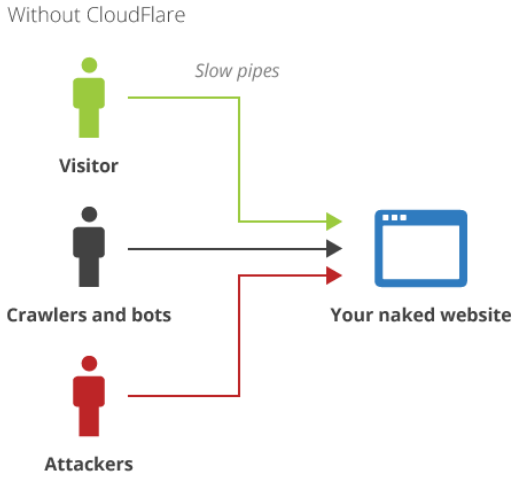
Marty Strong

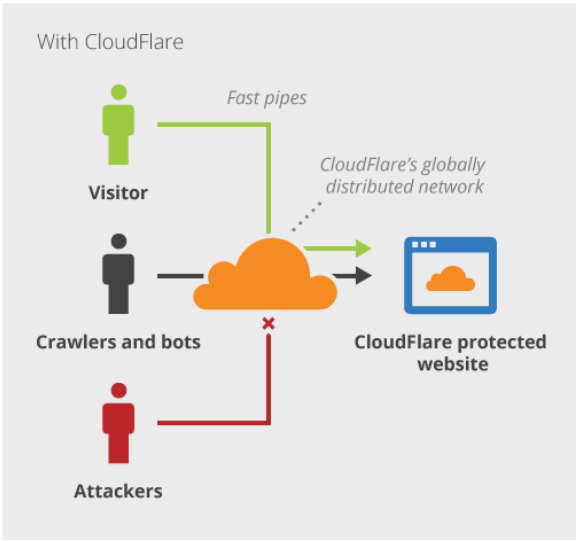
HKNOG 2.0 - Hong Kong
14th September 2015

What is CloudFlare?

CloudFlare makes websites faster and safer using our globally distributed network to deliver essential services to any website

- Performance
- Content
- Optimisation
- Security
- 3rd party services
- Analytics





How does CloudFlare work?

CloudFlare works at the network level

- Once a website is part of the CloudFlare community, its web traffic is routed through our global network of 30+ data centres.
- At each edge node, CloudFlare manages DNS, caching, bot filtering, web content optimisation and third party app installations.

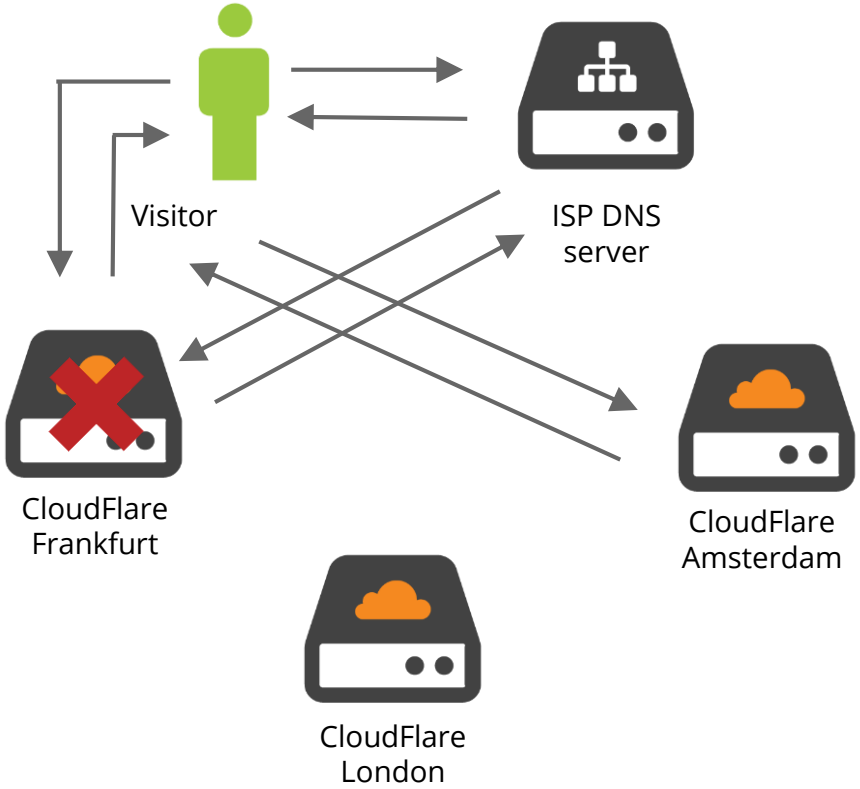
How does CloudFlare work?

How does it work?

- DNS Query - to anycast DNS address
- DNS result returned with Anycast IP
- Client makes connection to returned IP
- CloudFlare replies, session established

What happens in the event of an outage?

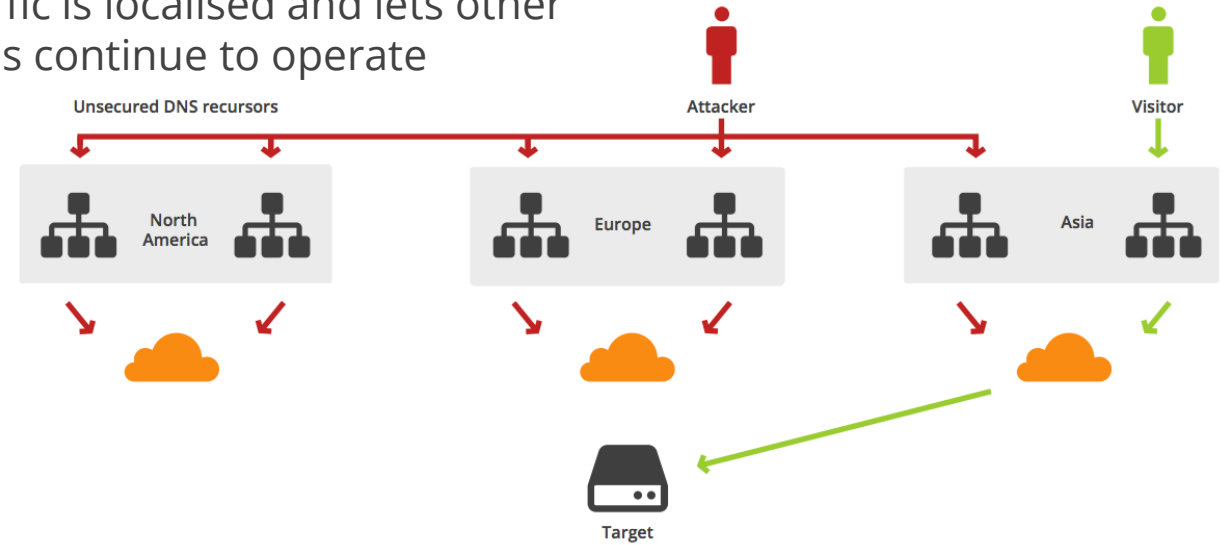
- Anycast prefixes are withdrawn from problematic PoP
- Traffic re-routes to next closest PoP
  - TCP session resets at this point



CloudFlare works globally

CloudFlare protects globally

- DDoS attack traffic is localised and lets other geographic areas continue to operate



Why do we peer?

Why do we peer?

"In computer networking, peering is a voluntary interconnection of administratively separate Internet networks for the purpose of exchanging traffic between the users of each network."

- To improve performance (reduce hop count, reduce latency etc.)
- To reduce costs
- To ensure anycast traffic lands locally
- To gain more control over routing
- To gain more control of DDoS traffic

Where do we peer?

Where do we peer?

- AKL-IX (Auckland)
- APE (Auckland)
- BBIX (Tokyo, Osaka, Singapore)
- Equinix (Hong Kong, Osaka, Singapore, Sydney, Tokyo)
- HKIX (Hong Kong)
- IX Australia (Melbourne, Sydney)
- JPIX (Tokyo, Osaka)
- JPNAP (Tokyo, Osaka)
- Megaport (Auckland, Singapore, Sydney)
- MyIX (Kuala Lumpur)
- PIPE (Melbourne, Sydney)

Plus many more @
http://as13335.peeringdb.com

What is a DDoS attack?

What is a DDoS attack?

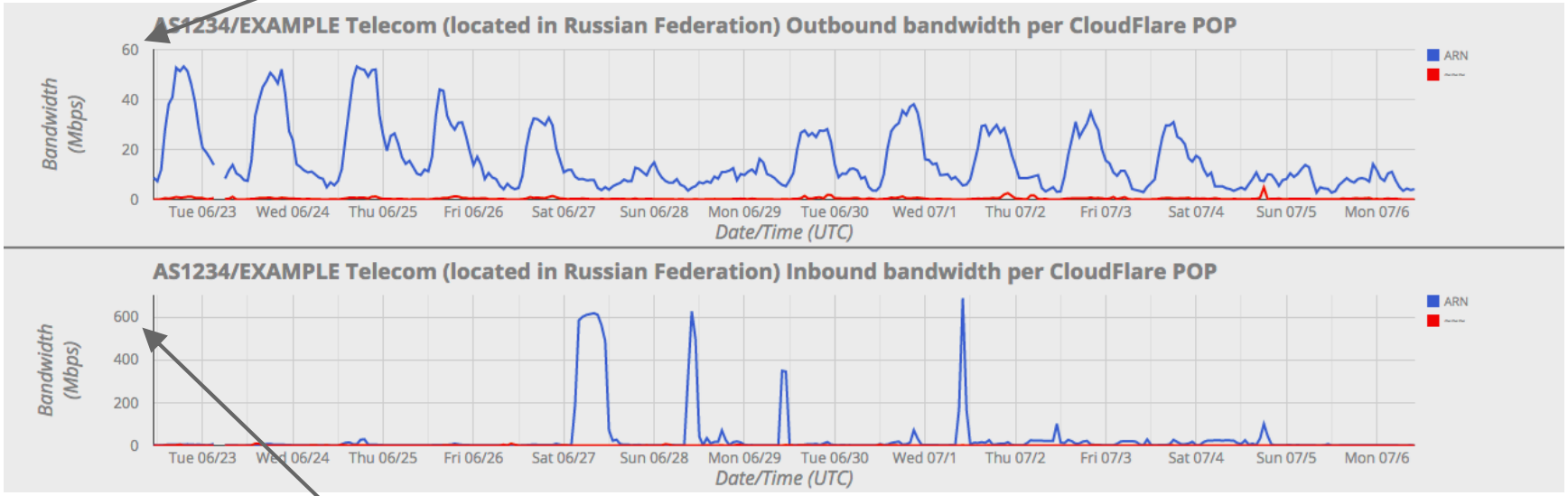
According to WikiPedia:

“In computing, a **denial-of-service** (**DoS**) **attack** is an attempt to make a machine or network resource unavailable to its intended users. This could be CPU resources, but often involves efforts to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. A **distributed denial-of-service** (**DDoS**) is where incoming traffic comes from more than one - and often thousands - of unique IPs, either from botnets or via various types of reflection attack.”
https://en.wikipedia.org/wiki/Denial-of-service_attack

Learn more here: https://www.cloudflare.com/ddos

DDoS network

60 Mbps peak



600 Mbps peak

DDoS network

- Our usual traffic ratio to eyeball ISPs is around 1:20 inbound:outbound
- However the ratio from the previous slide was 10:1 inbound:outbound
- The attacks shown on the graph are highly likely part of a much bigger global DDoS

How do we connect to this ISP?

AS1234 / EXAMPLE Telecom

IXP	DESCRIPTION	ASN	ROUTER	IF		Gbps	IPv4		IPv6	
DE-CIX Frankfurt	Example,RU	AS1234	edge01.fra02	ae2	.0	:	✔ 80.81.194.	8/126 6d00h	✔ 2001:7f8::	17/2 16d01h
AMS-IX	Example,RU	AS1234	edge01.ams01	ae3	.0	:	✔ 80.249.209.	1,401/125 9d23h	✔ 2001:7f8:	9d23h
Netnod Stockholm	Example,RU	AS1234	edge01.arn01	irb	.15	0.00 : 0.00	✔ 194.68.128.	1/160 20d01h	2001:7f8:	
LINX Juniper LAN	Example,RU	AS1234	edge01.lhr01	ae3	.0	0.00 : 0.00	✔ 195.66.225.	1,391/164 17d23h	✔ 2001:7f8:	29/4 1m09d

DDoS look-and-feel

DNS Attacks look different

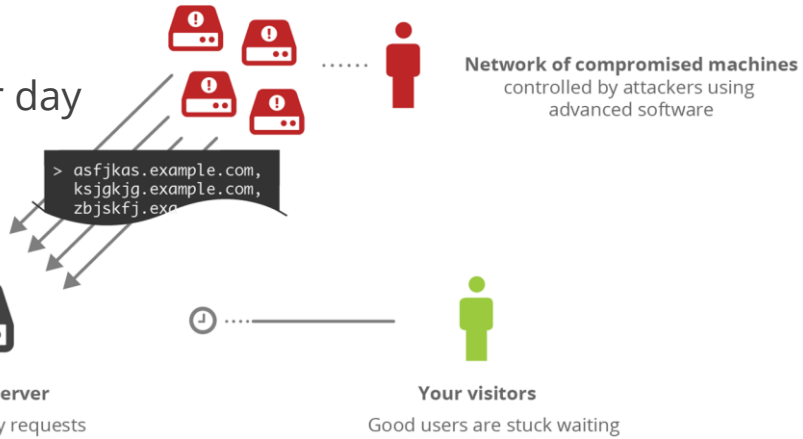
- Layer-7 attacks (hitting the application layer)
- Purpose: exhaust the CPU (vs. bandwidth)

Malicious payload

- Request sent to exploit vulnerability on server
- Purpose: gain control or release sensitive data
- CloudFlare WAF blocks ~1.2 billion request per day

Volumetric attack

- Send as many small packets as possible
- Purpose: overwhelm the router ports

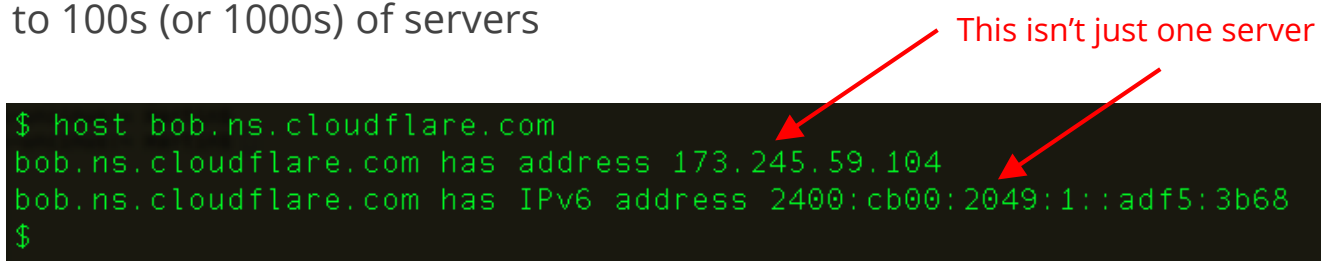


Why run 1,000s and 1,000s of servers?

Geography
- Spread the load for both content delivery and DDoS processing
- Allows us to distribute the attack more effectively
- Allow specific attack sources to be isolated

In-PoP load balancing
- Allows us to ensure no one server bears the entire brunt of an attack

Externally presented IP addresses
- One IP can map to 100s (or 1000s) of servers



Anycast routing

- You can’t guarantee which path ISPs will take
- Routing is down to the eyeball ISP
- There are a small number of ways to influence it
  - Use BGP communities to adjust announcements (e.g. do not announce to ASN X)
  - Use AS-Path prepending
  - Peer with ISPs

What if there was no peering?

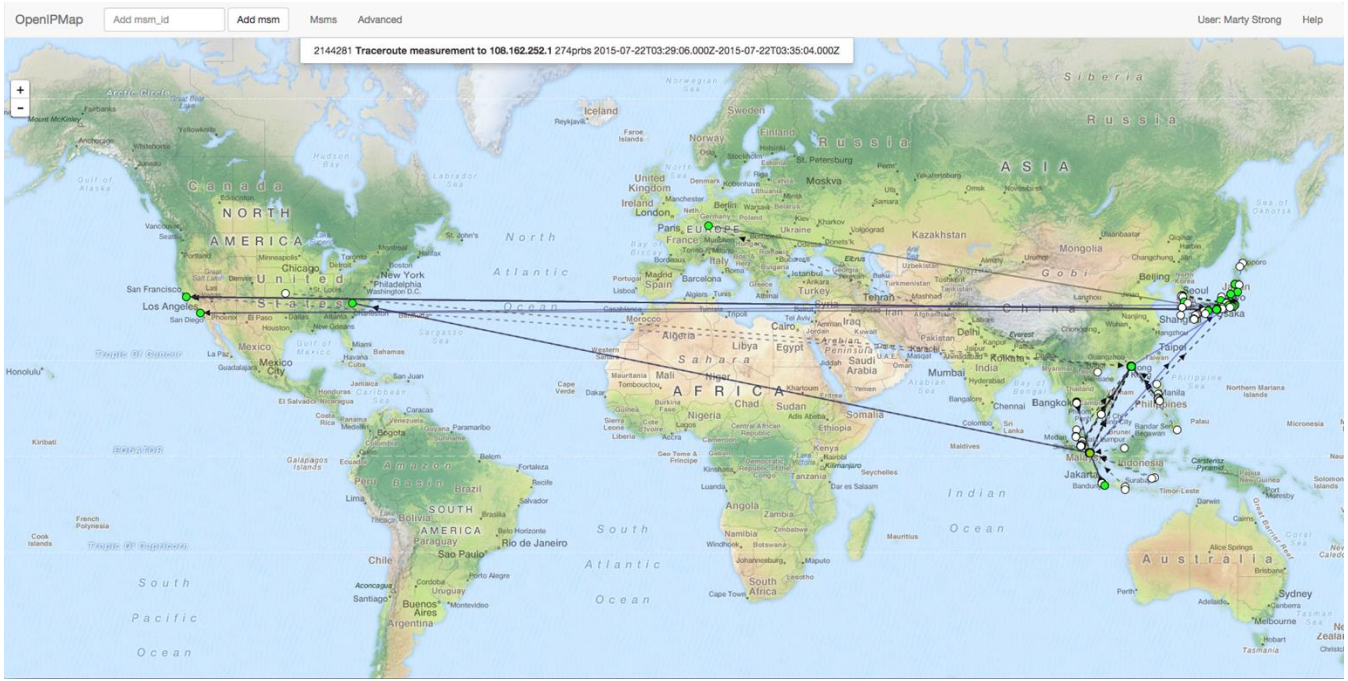
- You are reliant on your transit carriers' routing and interconnection with other providers
- Performance could be affected (long path, more hops etc.)
- Higher likelihood of sporadic changes

Why is this so important in Asia?

Let’s test: Methodology

- Take an IP prefix it and announce it in multiple locations (anycast)
  - Singapore
  - Hong Kong
  - Toyko
  - Osaka
- Do this separately for each provider in use (NTT, Tata, Pacnet)
- Make RIPE Atlas measurement
  - Probes from HK, ID, JP, KR, MY, PH, SG, TH, VN

Let’s test: NTT (AS2914)



https://marmot.ripe.net/openipmap/tracemap?msm_ids=2144281&show_suggestions=1&max_probes=274

Let’s test: NTT (AS2914)

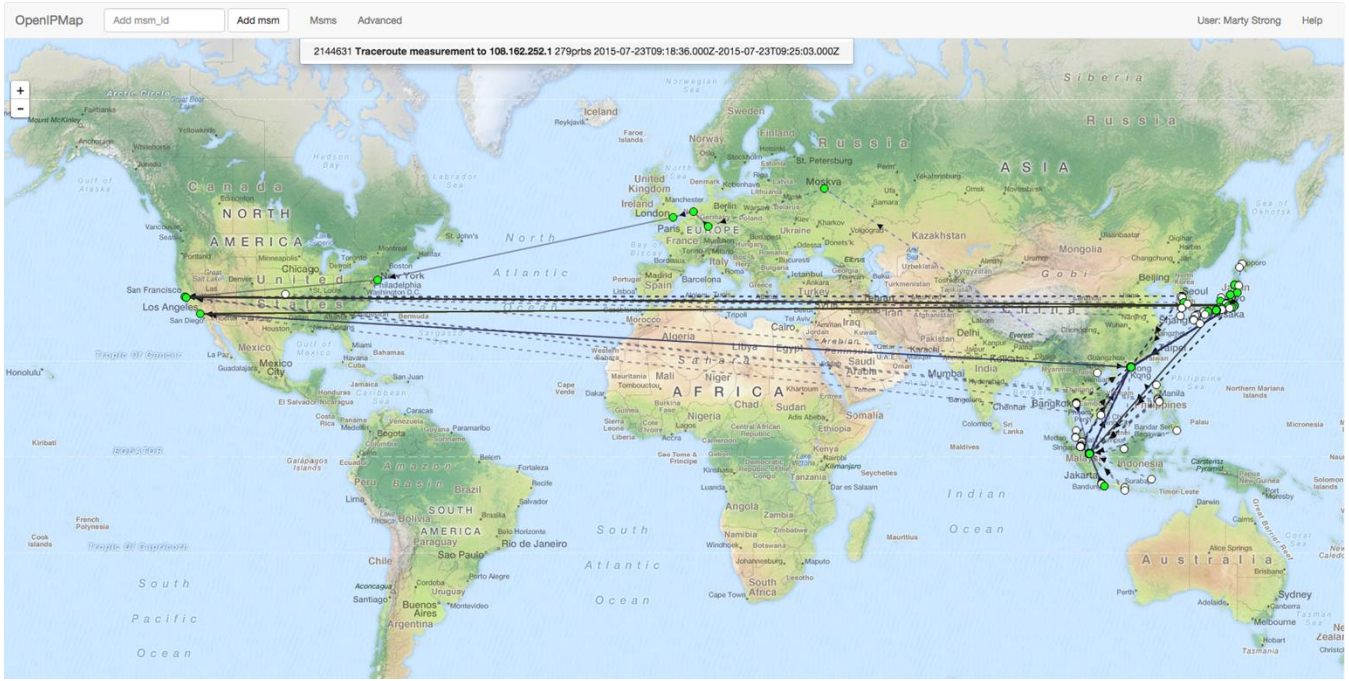
msm:2144281 prb:18166 ts:2015-07-22T03:29:12.000Z

remove

hop	IP	ASN	hostname	location		RTTs
1	203.30.39.254	23855	(none)			0.7 7.9 0.5
2	202.147.33.137	10026	Gi2-0-403.gw2.sin1.asianetcom.net	Singapore,,SG	ok	0.5 0.4 0.4
3	61.14.157.113	10026	te0-0-2-0.wr2.sin0.asianetcom.net	Singapore,,SG	ok	4.7 1.4 1.7
4	61.14.157.62	10026	te0-0-2-0.wr2.osa0.asianetcom.net	Osaka,Ōsaka,JP	ok	176.6 177.3 177.7
5	202.147.50.197	10026	te0-0-0-4.gw3.sjc1.asianetcom.net	San Jose,California,US	ok	177.9 177.9 177.8
6	213.248.83.113	1299	sjo-b21-link.telia.net	San Jose,California,US	ok	178.1 178.0 179.2
7	62.115.44.46	1299	ntt-ic-306350-sjo-b21.c.telia.net	San Jose,California,US	ok	186.3 186.4 186.4
8	129.250.4.25	2914	ae-4.r23.snjsca04.us.bb.gin.ntt.net	San Jose,California,US	ok	194.1 186.2 186.0
9	129.250.2.131	2914	ae-6.r21.osakjp02.jp.bb.gin.ntt.net	Osaka,Ōsaka,JP	ok	361.8 361.8 361.8
10	129.250.6.144	2914	ae-5.r23.osakjp02.jp.bb.gin.ntt.net	Osaka,Ōsaka,JP	ok	368.2 370.2 396.1
11	129.250.3.199	2914	ae-2.r01.osakjp02.jp.bb.gin.ntt.net	Osaka,Ōsaka,JP	ok	370.0 371.1 370.1
12	108.162.252.1	13335	(none)			199.1 199.1 199.0

https://marmot.ripe.net/openipmap/tracemap?msm_ids=2144281&show_suggestions=1&max_probes=274

Let’s test: Tata (AS6453)



https://marmot.ripe.net/openipmap/tracemap?msm_ids=2144631&show_suggestions=1&max_probes=274

Let’s test: Tata (AS6453)

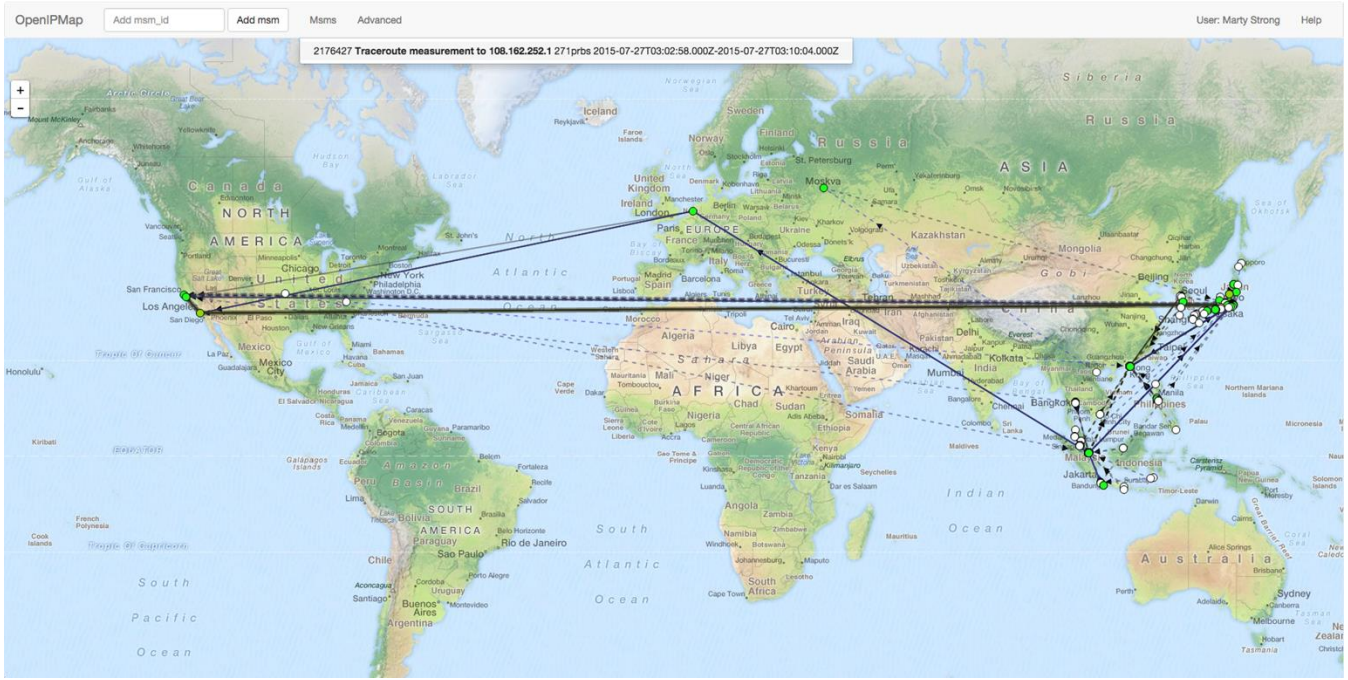
msm:2144631 prb:18497 ts:2015-07-23T09:18:41.000Z

remove

hop	IP	ASN	hostname	location		RTTs
1	192.168.25.1	(RFC1918)	(none)			16.3 0.8 0.8
2	175.115.120.1	9318	(none)			3.8 3.7 3.8
3	211.58.43.121	9318	(none)			1.6 1.5 1.6
4	114.207.86.97	9318	(none)			1.3 1.3 1.3
5	1.255.23.53	9318	(none)			5.8 5.6 5.9
6	118.221.7.26	9318	(none)			3.2 5.5 5.8
7	58.229.15.226	9318	(none)			37.5 37.4 37.4
8	80.239.160.169	1299	hnk-b2-link.telia.net	Hong Kong,,HK	ok	159.6 159.8 159.5
9	62.115.138.118	1299	las-b21-link.telia.net	Los Angeles,California,U	ok	155.8 154.5 154.4
10	213.155.134.251	1299	las-b3-link.telia.net	Los Angeles,California,U	ok	154.1 154.0 154.0
11	66.110.59.65	6453	ix-23-0.tcore1.LVW-Los-Angeles.as6453.net	Los Angeles,California,U	ok	178.9 179.0 179.0
12	66.110.59.2	6453	if-2-2.tcore2.LVW-Los-Angeles.as6453.net	Los Angeles,California,U	ok	266.0 265.6 265.8
13	66.110.59.62	6453	if-12-2.tcore2.TV2-Tokyo.as6453.net	Tokyo,Tōkyō,JP	ok	265.5 265.5 265.3
14	108.162.252.1	13335	(none)			154.3 153.9 153.8

https://marmot.ripe.net/openipmap/tracemap?msm_ids=2144631&show_suggestions=1&max_probes=274

Let’s test: Pacnet (AS10026)



https://marmot.ripe.net/openipmap/tracemap?msm_ids=2176427&show_suggestions=1&max_probes=274

Let’s test: Pacnet (AS10026)

msm:2176427 prb:13531 ts:2015-07-27T03:03:04.000Z

remove

hop	IP	ASN	hostname	location		RTTs
1	208.69.37.1	36692	rtr1.nrt.opendns.com	Tokyo,Tōkyō,JP	ok	17.9 0.4 0.5
2	183.182.80.145	3257	ge-3-0-1.tyo10.ip4.gtt.net	Tokyo,Tōkyō,JP	ok	0.4 0.4 0.4
3	141.136.111.133	3257	xe-4-3-1.sjc12.ip4.gtt.net	San Jose,California,US	ok	109.1 108.8 109.8
4	77.67.68.234	3257	pacnet-gw.ip4.gtt.net		ok	109.3 109.2 109.4
5	202.147.50.134	10026	gi9-0-0.cr2.nrt1.asianetcom.net	Tokyo,Tōkyō,JP	ok	211.6 211.8 211.4
6	202.147.0.182	10026	ge-2-1-0-0.gw3.nrt5.asianetcom.net	Tokyo,Tōkyō,JP	ok	219.1 212.7 211.7
7	108.162.252.1	13335	(none)			106.7 109.9 106.7

https://marmot.ripe.net/openipmap/tracemap?msm_ids=2176427&show_suggestions=1&max_probes=274

How is this related to ingesting DDoS attacks?

- By utilising multiple transit carriers and peering extensively you have path diversity i.e. multiple ports that will ingest the attack
- You can geographically separate traffic
- There are less collateral issues caused to upstream backbones

Thank you!

Questions?

AS13335

http://as13335.peeringdb.com/

Marty Strong, Network Engineer
@martystronguk / @cloudflare
marty@cloudflare.com
https://www.cloudflare.com/