Rusty Clusters? Dusting an IPv6 Research Foundation

Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, Georg Carle

Wednesday 26th October, 2022

ACM Internet Measurement Conference 2022
Nice, France

Dusting an IPv6 Research Foundation - The IPv6 Hitlist

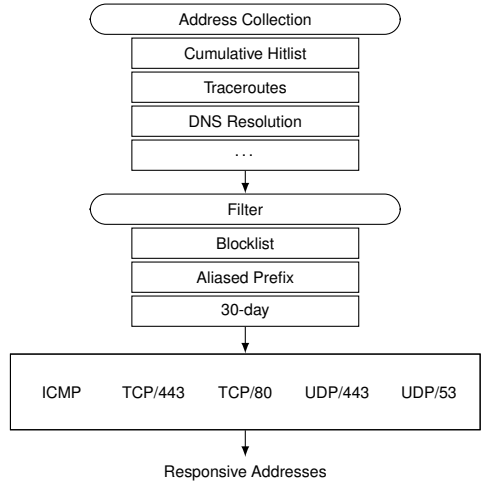
The large and sparsely used IPv6 address space is infeasible to scan.

Dusting an IPv6 Research Foundation - The IPv6 Hitlist

The large and sparsely used IPv6 address space is infeasible to scan.

Gasser et al.[1] established an ongoing IPv6 Hitlist in 2018, that:

- collects address candidates from multiple sources,
- applies different filters,
- and tests addresses for their responsiveness.



[1] O. Gasser et al. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In Proc. ACM Int. Measurement Conference [1]

Dusting an IPv6 Research Foundation - The IPv6 Hitlist

The large and sparsely used IPv6 address space is infeasible to scan.

Gasser et al.[1] established an ongoing IPv6 Hitlist in 2018, that:

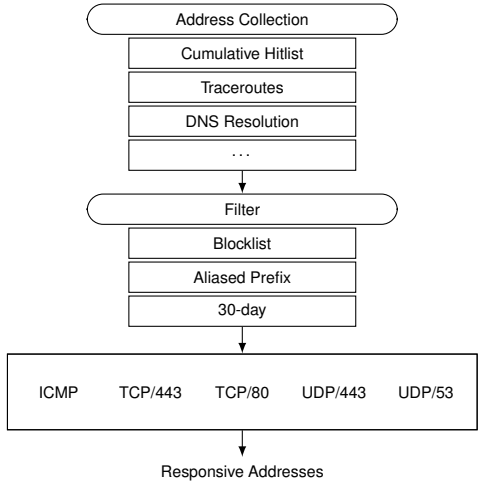
- collects address candidates from multiple sources,
- applies different filters,
- and tests addresses for their responsiveness.

Research questions:

RQ1 How did the IPv6 Hitlist develop?

RQ2 Should addresses from aliased prefix be strictly excluded?

RQ3 Can we improve the IPv6 Hitlist with new sources?



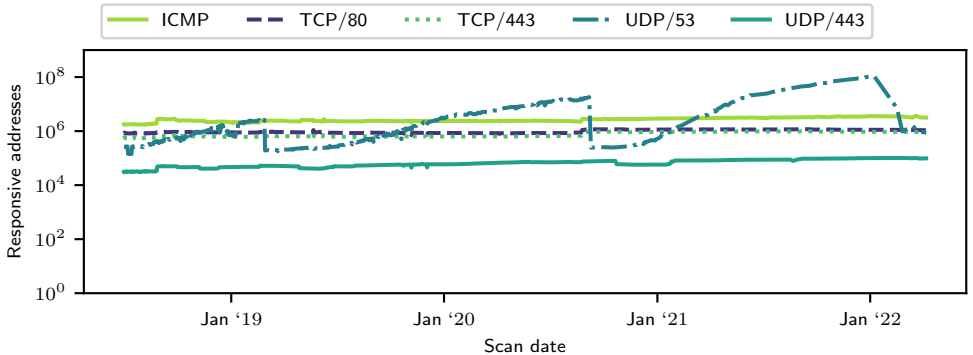
[1] O. Gasser et al. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In Proc. ACM Int. Measurement Conference [1]

How did the IPv6 Hitlist develop? (1/3)

- Input increased from 90 M to 790 M addresses.
- 250 M addresses are within aliased prefixes.
- 405 M addresses are unresponsive for at least 30 d.

How did the IPv6 Hitlist develop? (1/3)

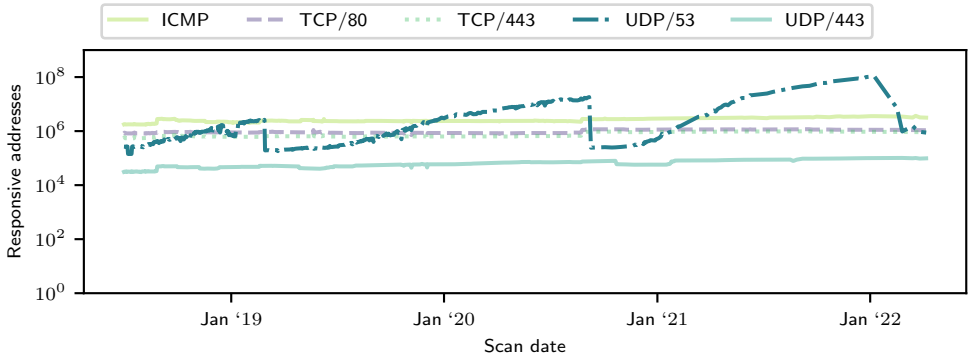
- Input increased from 90 M to 790 M addresses.
- 250 M addresses are within aliased prefixes.
- 405 M addresses are unresponsive for at least 30 d.
- Up to 100 M addresses are responsive to at least one protocol.



How did the IPv6 Hitlist develop? (1/3)

- Input increased from 90 M to 790 M addresses.
- 250 M addresses are within aliased prefixes.
- 405 M addresses are unresponsive for at least 30 d.
- Up to 100 M addresses are responsive to at least one protocol.

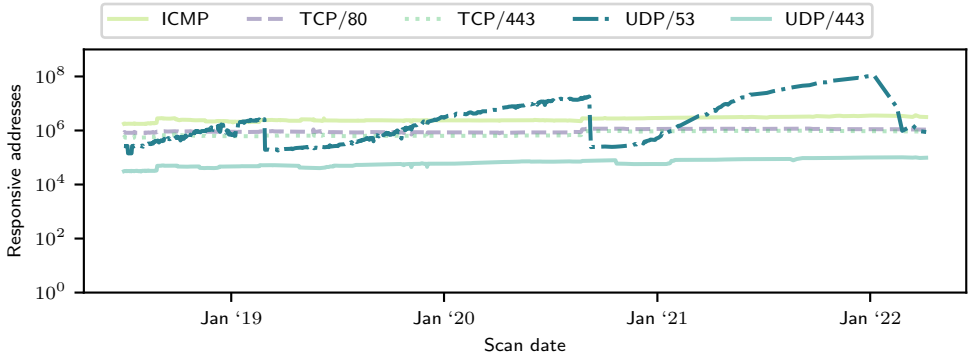
→ However, large spikes in addresses responsive to UDP/53 are visible.



How did the IPv6 Hitlist develop? (2/3)

- ZMapv6 is configured to send DNS queries for `www.google.com`
- Most responses
  - contain invalid addresses (Teredo),
  - are received multiple times,
  - and responsive addresses are mostly related to Chinese ASes.

→ Prevalence of DNS responses due to injected DNS responses by Chinese censorship mechanisms.



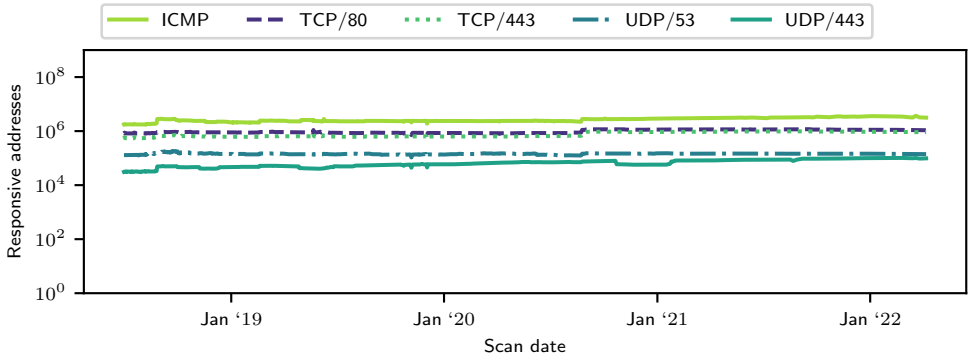
How did the IPv6 Hitlist develop? (3/3)

- We accumulated a blocklist of 134M unresponsive hosts receiving injections.
- We filter injected packets directly after our scans.

→ We cleaned the hitlist and published data.

→ The result is more stable for all protocols.

→ 3.2 M addresses are responsive, covering 15.7 k ASes.



Should addresses from aliased prefix be strictly excluded?

The IPv6 Hitlist treats aliased prefixes as:

- a complete prefix
- with each IPv6 address used as alias
- by a single host.

→ A single such prefix could bias the hitlist and make scans infeasible

Should addresses from aliased prefix be strictly excluded?

The IPv6 Hitlist treats aliased prefixes as:

- a complete prefix
- with each IPv6 address used as alias
- by a single host.

→ A single such prefix could bias the hitlist and make scans infeasible

A prefix is labeled as aliased and excluded if

- 16 randomly generated addresses are responsive

Should addresses from aliased prefix be strictly excluded?

The IPv6 Hitlist treats aliased prefixes as:

- a complete prefix
- with each IPv6 address used as alias
- by a single host.

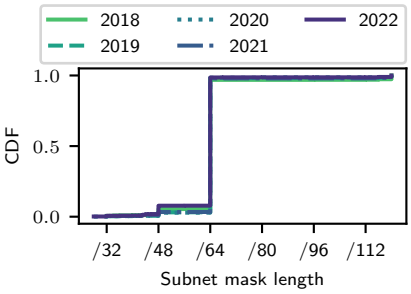
→ A single such prefix could bias the hitlist and make scans infeasible

A prefix is labeled as aliased and excluded if

- 16 randomly generated addresses are responsive

The presence of aliased prefixes is increasing:

- The amount of aliased prefixes increased from 12 k to 111 k.
- They cover more than 250 M addresses from the cumulative input.
- Most detected aliased prefixes are of size /64.



Should addresses from aliased prefix be strictly excluded?

However, many of these are not necessarily aliased but fully responsive:

- Aliased prefixes are often announced by CDNs.
  - For Fastly, more than 98 % of announced IPv6 addresses are aliased.

Should addresses from aliased prefix be strictly excluded?

However, many of these are not necessarily aliased but fully responsive:

- Aliased prefixes are often announced by CDNs.
  - For Fastly, more than 98 % of announced IPv6 addresses are aliased.
- Many domains resolve to IPv6 addresses in aliased prefixes.
  - For Cloudflare, aliased prefixes host more than 10 M domains.

Should addresses from aliased prefix be strictly excluded?

However, many of these are not necessarily aliased but fully responsive:

- Aliased prefixes are often announced by CDNs.
  - For Fastly, more than 98 % of announced IPv6 addresses are aliased.
- Many domains resolve to IPv6 addresses in aliased prefixes.
  - For Cloudflare, aliased prefixes host more than 10 M domains.
- Fingerprinting reveals different behavior between hosts within the same aliased prefix.

Should addresses from aliased prefix be strictly excluded?

However, many of these are not necessarily aliased but fully responsive:

- Aliased prefixes are often announced by CDNs.
  - For Fastly, more than 98 % of announced IPv6 addresses are aliased.
- Many domains resolve to IPv6 addresses in aliased prefixes.
  - For Cloudflare, aliased prefixes host more than 10 M domains.
- Fingerprinting reveals different behavior between hosts within the same aliased prefix.

We tested the responsiveness to other protocols for one address out of each aliased prefix:

- 32.3 k are responsive to TCP/443
- 28.8 k are responsive to UDP/443
- 172 are respnsive to UDP/53

Should addresses from aliased prefix be strictly excluded?

However, many of these are not necessarily aliased but fully responsive:

- Aliased prefixes are often announced by CDNs.
  - For Fastly, more than 98 % of announced IPv6 addresses are aliased.
- Many domains resolve to IPv6 addresses in aliased prefixes.
  - For Cloudflare, aliased prefixes host more than 10 M domains.
- Fingerprinting reveals different behavior between hosts within the same aliased prefix.

We tested the responsiveness to other protocols for one address out of each aliased prefix:

- 32.3 k are responsive to TCP/443
- 28.8 k are responsive to UDP/443
- 172 are respnsive to UDP/53

→ Including addresses from fully responsive prefixes should be considered in research relying on the IPv6 Hitlist.

Can we improve the IPv6 Hitlist with new sources?

While the existing IPv6 Hitlist sources regularly update the input, new sources have not been added.

We evaluated different approaches to extend our hitlist:

- Target Generation:
  - 6Tree, 6Graph, 6GAN, 6VecLM,
  - Distance Clustering (DC) (our own custom algorithm)
- 30-day unresponsive addresses

Method	Addr	Responsive Addr. ↓	Responsive ASes
6Graph	125.8 M		
6Tree	37.6 M		
DC	5.3 M		
6GAN	3.3 M		
6VecLM	70.3 k		
30-day Unresp.	405.0 M		

Can we improve the IPv6 Hitlist with new sources?

While the existing IPv6 Hitlist sources regularly update the input, new sources have not been added.

We evaluated different approaches to extend our hitlist:

- Target Generation:
  - 6Tree, 6Graph, 6GAN, 6VecLM,
  - Distance Clustering (DC) (our own custom algorithm)
- 30-day unresponsive addresses

Method	Addr	Responsive Addr. ↓	Responsive ASes
6Graph	125.8 M	3.8 M	10.7 k
6Tree	37.6 M	2.2 M	11.5 k
DC	5.3 M	651.9 k	5.5 k
6GAN	3.3 M	4.3 k	39
6VecLM	70.3 k	1.0 k	105
30-day Unresp.	405.0 M	1.3 M	9.0 k

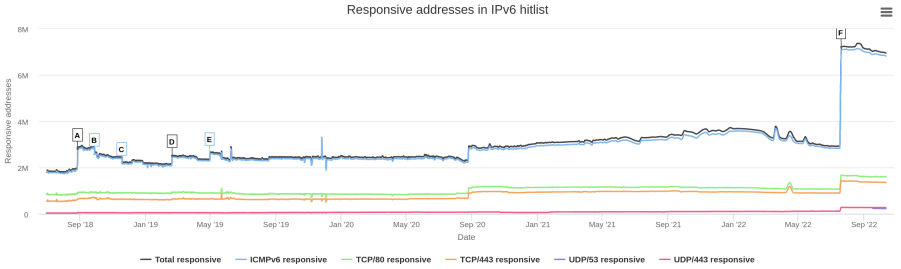
→ All sources contribute additional responsive addresses.
→ We identified 5.6 M new responsive IPv6 addresses from 14.6 k ASes.

Conclusion

We updated the IPv6 Hitlist service and

- removed the impact of DNS response injection;
- added new sources to the ongoing Hitlist service (F mark);
- recommend scanning targets from fully responsive prefixes in the future.

https://ipv6hitlist.github.io/



Conclusion

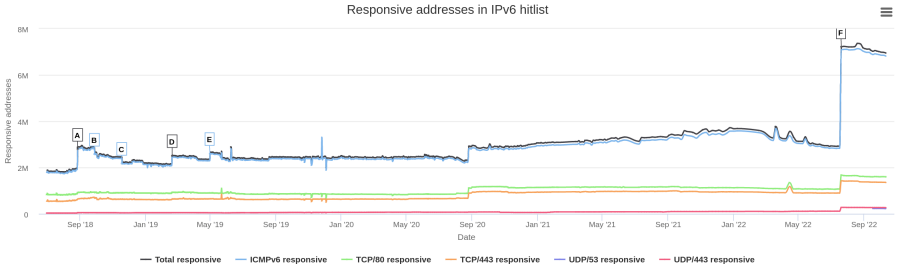
We updated the IPv6 Hitlist service and

- removed the impact of DNS response injection;
- added new sources to the ongoing Hitlist service (F mark);
- recommend scanning targets from fully responsive prefixes in the future.

We encourage everybody to share

- new address sources or target generation approaches,
- and new insights or ideas to update the ongoing service with us.

https://ipv6hitlist.github.io/



Bibliography

[1] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczynski, S. D. Strowes, L. Hendriks, and G. Carle.
Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists.
In Proc. ACM Int. Measurement Conference (IMC), 2018.