Identifying Aggressive DNS Resolver Behaviors using Unsupervised Machine Learning (abstract)

Natália G. Knob[1], Ricardo de O. Schmidt[1], and Marco A. S. Trentin[1]

[1]University of Passo Fundo, Brazil

Recursive resolvers operate making requests to domain name servers and caching retries to reduce latency and improve resilience [9]. Because of their role, the DNS resolvers are an essential component of the DNS infrastructure, and their misbehavior can affect DNS clients and authoritative servers. Besides that, misbehaving resolvers can be misused in attacks [11], or cause delays in responses to clients [1, 3].

Despite previous research about DNS environment [10, 6, 8, 13, 2], implementations and behaviors of recursive resolvers are not well known. However, it is a consensus that an excessive amount of queries sent by recursives to the DNS infrastructure [5, 4, 12] misuse precious resources of authoritative nameservers, probably without any useful purpose.

In prior analysis [7], where we looked at 2019 DITL[1] dataset, we observed that 99% of all recursives send moderate amount of queries. However, the 1% that send aggressively were responsible for sending up to 87% of all traffic. Moreover, the 1% of abusive recursives correspond to more than 8.5k resolvers which indicates a high amount of them making a massive amount of queries, misusing the DNS infrastructure.

In order to identify, quantify and characterize aggressive resolvers, we now analyze DNS-OARC DITL datasets from 2016 to 2020 using an unsupervised clustering algorithm known as GMM (Gaussian Mixture Models). We chose to use the GMM, given that it allows the classification of non-spherical data groups. In preliminary results, it is possible to verify the existence of four distinct groups of resolvers, grouped similarly, despite the year observed.

In figure 1 it is possible to identify DITL 2019 resolvers[2] grouped according to their behavior: non-aggressive or with low, medium, or high aggressiveness. This categorization considers as attributes the number of requests made, the number of minutes a resolver has been active (where there has been at least one request), and the quantity of peaks across the queries distribution.

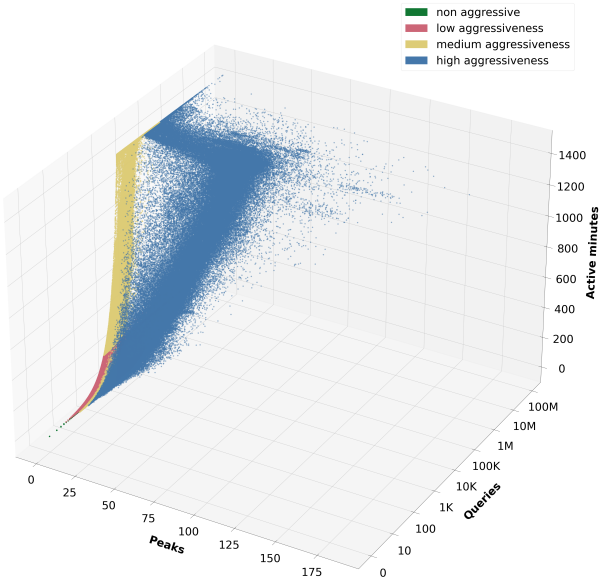

Figure 1: DITL2019 resolvers grouped according to their behavior

	Qty. Resolvers	% of all	Qty. Requests	% of all
Non aggr.	5793290	65,39	15091425	0,19
Low aggr.	2508122	28,31	290867213	3,75
Medium aggr.	342548	3,87	808205837	10,43
High aggr.	215139	2,43	6632439789	85,62
All	8859099	100	7746604264	100

Table 1: Amount and percentage of queries and resolvers according to the groups formed by the clustering method

According to table 1, it should be noted that non-aggressive resolvers account for 65% of the total number of resolvers and only 0.2% of the total queries sent, on the other hand, the most aggressive recursives account for 2.4% of the total number of them and 85.6% of total queries sent. It is also observed that the non-aggressive recursives, although they are in greater quantity, are superimposed in a few points of the graph, with little varied behavior regarding the analyzed attributes.

Finally, it is also possible to observe that resolvers with more abusive behavior are mainly related to a greater number of peaks and a greater number of queries performed, with the active minutes factor being less relevant.

Future work intends to identify ranges of values for the selected attributes that allow easy and without clusterization help us recognize recursives that potentially have aggressive behavior.

[1]`https://www.dns-oarc.net/oarc/data/ditl`

[2]Results already obtained from other years can be seen through address `https://drive.google.com/drive/folders/1_ShHdSGXvTkYkEupcbNmVhcwoOGI-Rza?usp=sharing`

References

[1] Rami Al-Dalky, Michael Rabinovich, and Kyle Schomp. A Look at the ECS Behavior of DNS Resolvers. In *Internet Measurement Conference*, 2019.

[2] Mark Allman. Comments on dns robustness. In *Proceedings of the Internet Measurement Conference 2018*, IMC '18, pages 84–90, New York, NY, USA, 2018. ACM.

[3] Tom Callahan, Mark Allman, and Michael Rabinovich. On Modern DNS Behavior and Properties. *ACM Computer Communication Review*, 43(3), July 2013.

[4] Sebastian Castro, Duane Wessels, Marina Fomenkov, and Kimberly C. Claffy. A day at the root of the internet. *Computer Communication Review*, 38(5):41–46, 2008.

[5] Sebastian Castro, Min Zhang, Wolfgang John, Duane Wessels, and Kimberly C. Claffy. Understanding and preparing for DNS evolution. In *Traffic Monitoring and Analysis, Second International Workshop, TMA 2010, Zurich, Switzerland, April 7, 2010, Proceedings*, pages 1–16, 2010.

[6] Andrew Kalafut, Minaxi Gupta, Pairoj Rattadilok, and Pragneshkumar Patel. Surveying dns wildcard usage among the good, the bad, and the ugly. In Sushil Jajodia and Jianying Zhou, editors, *Security and Privacy in Communication Networks*, pages 448–465, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[7] Natália G Knob, Ricardo de O Schmidt, Marco AS Trentin, Jelena Mirkovic, Wes Hardaker, and John Heidemann. Understanding and quantifying aggressive resolver behaviors. *DNS and Internet Naming Research Directions 2020*.

[8] Daiping Liu, Shuai Hao, and Haining Wang. All your dns records point to us: Understanding the security threats of dangling dns records. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1414–1425, New York, NY, USA, 2016. ACM.

[9] P. Mockapetris. Domain names - implementation and specification. RFC 1035, Internet Engineering Task Force, November 1987.

[10] Giovane C. M. Moura, John Heidemann, Ricardo de O. Schmidt, and Wes Hardaker. Cache me if you can: Effects of DNS Time-to-Live (extended). Technical Report ISI-TR-734b, USC/Information Sciences Institute, July 2019. Released May 2019, updated Sept. 2019.

[11] Christian Rossow. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. In *In Proceedings of the 2014 Network and Distributed System Security Symposium, NDSS*, 2014.

[12] Duane Wessels and Marina Fomenkov. Wow, that's a lot of packets. In *Proc. of Passive and Active Measurement Workshop*, 2003.

[13] Ming Zhang, Yaoping Ruan, Vivek Pai, and Jennifer Rexford. How dns misnaming distorts internet topology mapping. In *Proceedings of the Annual Conference on USENIX '06 Annual Technical Conference*, ATEC '06, pages 34–34, Berkeley, CA, USA, 2006. USENIX Association.