

# Internship proposal:

## Domain popularity ranking based on passive DNS traffic

**Location:** The internship will be located at Grenoble Informatics Laboratory (LIG) (<https://www.liglab.fr>), Bâtiment IMAG, 700 Avenue Centrale, 38401 Saint-Martin-d'Hères

**Team:** Drakkar group (<http://drakkar.imag.fr>)

**Advisor:** Maciej Korczyński ([maciej.korczynski@univ-grenoble-alpes.fr](mailto:maciej.korczynski@univ-grenoble-alpes.fr))

**Dates:** February to July 2018 (6 months)

**Keywords:** Algorithms, Domain Name System, Data Analytics, Security

**Prerequisites:**

- Good knowledge in algorithms
- Good knowledge in programming (e.g. Python)
- Knowledge in networking would be a plus

**Remarks:** This internship will be performed in close collaboration with industry partners. It can be followed by a PhD thesis.

## 1 Context

The most popular techniques for measuring the popularity of a website depend on search hits or the deployment of client-side measurement agents that may make them susceptible to various manipulation methods and therefore may undermine their results. The most prominent example is Alexa Traffic Ranking [1] that is based on aggregated historical traffic from millions of Alexa Toolbar users [2]. Therefore, it depends on the adoption of the Alexa Browser extension. Another important limitation is that websites with low measured traffic will not be accurately ranked [3].

## 2 Internship Program and Objectives:

The objective of this internship is to create an alternative ranking based on the analysis of Domain Name System (DNS) traffic rather than on HTTP traffic that would potentially overcome the above-mentioned limitations. While there have been some recent efforts using various active [4] and passive measurements [5, 6, 7] of DNS traffic, which reflect the popularity of Internet activity, we argue that ranking based on passively observed

DNS traffic coming from thousands of sensors located around the world would be the most accurate.

The project would consist on the analysis of DNS queries and developing a popularity algorithm that takes into account the number of queries observed in the passive dataset in defined time windows, a Time to Live (TTL) value of each domain indicating how long the domain “stays” in DNS caching servers, and more.

To validate this work, the intern would be able to compare the obtained results with ground truth data, more specifically with the DNS traffic collected at authoritative domain name servers of one of the leading Top Level Domain (TLD) registry operators.

The research would include a comparison with the Alexa 1M list and Cisco Umbrella popularity rankings. A list of 1 Million most popular domains along with the algorithm could be available for Internet users free of charge.

## References

- [1] “Alexa Top Sites,” <https://aws.amazon.com/alexa-top-sites>.
- [2] “The Alexa Extension,” <http://www.alexa.com/toolbar>.
- [3] “How are Alexa’s traffic rankings determined,” <https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined->.
- [4] M. A. Rajab, F. Monroe, A. Terzis, and N. Provos, “Peeking through the cloud: Dns-based estimation and its applications,” in *Proceedings of the 6th International Conference on Applied Cryptography and Network Security*, ser. ACNS’08, 2008, pp. 21–38. [Online]. Available: <https://www.cs.jhu.edu/~moheeb/webpage.files/DNS.pdf>
- [5] L. Deri, S. Mainardi, M. Martinelli, and E. Gregori, “Exploiting dns traffic to rank internet domains,” in *2013 IEEE International Conference on Communications Workshops (ICC)*, June 2013, pp. 1325–1329. [Online]. Available: <http://luca.ntop.org/tricans2013.pdf>
- [6] A. Mayrhofer, “How popular is this Domain? Yet another (DNS based) approach,” [http://sched.ws/hosted\\_files/icann58copenhagen2017/7a/DNS%20Magnitude%20-%20Mayrhofer.pdf](http://sched.ws/hosted_files/icann58copenhagen2017/7a/DNS%20Magnitude%20-%20Mayrhofer.pdf).
- [7] “Cisco Umbrella 1 Million,” <https://umbrella.cisco.com/blog/blog/2016/12/14/cisco-umbrella-1-million/>.