Internet Freedom Nonprofit Seeks Metrics Data Architect

February 4, 2021

The Tor Project, Inc., a 501(c)(3) nonprofit organization advancing human rights and freedoms by creating and deploying free and open source anonymity and privacy technologies, is seeking an experienced Data Architect to take our metrics work to the next level.

Tor is for everyone, and we are actively working to build a team that represents people from all over the world — people from diverse ethnic, national, and cultural backgrounds; people from all walks of life. Racial minorities, non-gender-binary people, women, and people from any group that is generally underrepresented in tech are encouraged to apply.

This is a full-time remote position.

Our Metrics Team has been collecting data since 2004 to help improve the tools we build and learn more about the Tor network. For example, we monitor the number of relays and clients in the network, their respective capabilities, the number of clients connecting via bridges, fluctuations in network speed, etc. Gathering this data results in huge data archives, so we are also working to develop tools to process this data and make it available to everyone.

How we achieve our goals:

- Robustness (We want to avoid bugs and/or bad design decisions that cause us to miss data)
- Timeliness (users need up-to-date network status information)
- Scalability (as the network grows, so does our data)
- Transparency (our community rightly wants to know what data we're collecting)

The most interesting challenge for metrics collection at Tor is how to gather data on an anonymity system without de-anonymizing users.

We are integrating the metrics team with a new network health team that will be also focusing on taking care of the health of the Tor network. Additionally, Tor has just begin a large network performance improvement project that will vastly increase network utilization and client performance. There will be close interaction between this role and our performance effort, in the form of ensuring data is collected properly during deployment of new performance features and performance tuning. This will also include producing custom visualizations, doing analysis, and comparisons between historical and current data.

The person in this position will work directly with helping us maintain existing systems, and design new systems for gathering and analyzing data. They will help the rest of the teams understand the data available to improve our tools as well as the Tor network's health.

The bulk of our code is written in Java, but smaller portions are written in R, Python, PostgreSQL, and JavaScript. Part of this job will be to analyze and fix bugs in our current code and review patches.

You will be working with the existing teams in Tor with support and advice on data collection and measurements for their needs in their respective projects.

Our main five codebases:

- Collector
- metrics-lib
- Onionoo
- Exonerator
- metrics-web

Requirements:

Technical abilities/experience:

- Have experience finding your way into existing Java, R, and PostgreSQL-based code bases and the ability to review patches and make changes to fix bugs/smaller enhancements.
- Have experience with web application development in the Python programming language. Experience with known frameworks is a plus.
- Able to identify shortcomings in our data pipeline and suggest improvements to reduce complexity and future maintenance efforts.
- Experience working with Git and Gitlab or similar issue tracking tools.
- Ability to learn quickly and can adapt to our current processes; being able to improve future processes for releasing software and operating services.
- Understanding of the inherent privacy implications of gathering data in an anonymity system, the security implications of gathering metrics data from semi-trusted relays in the Tor network, and the challenges of processing large amounts of data per day (specifically performance and scalability challenges).
- Data analysis: Ability to make sense of data sets and use data analysis tools to find and visualize interesting patterns.
- Mathematics: Knowledge of basic statistics.

Collaborative requirements:

- Ability to work remotely 100% of the time, as synchronization happens via email and/or IRC.
- Participation in weekly IRC meetings and monthly team video chats.
- Language: write and speak fluent English.
- Comfortable posting to a public mailing list or speaking up in a public IRC channel to ask questions, even when you think the question might be obvious or silly.

Bonus skills:

- Open source experience: You have contributed to an open source project before, and you're accustomed to a pattern of early and frequent releases without attempting to finalize things on your own.
- Scientific writing: Experience writing technical reports about data findings.
- Networking background: Experience working with networks and measurements in the past.
- You support Internet Freedom!

To apply, submit a cover letter, your CV/resume (including three professional references), and a link to a code sample or some non-trivial software project you have significantly contributed to. In your cover letter, please include the reason you want to work at the Tor Project.

IMPORTANT: Please email application materials in plain text or PDF format to job-metrics at torproject dot org with “Metrics Data Architect” in the subject line.

The Tor Project's workforce is smart, committed, and hard working. We currently have a paid and contract staff of around 24 developers and operational support people, plus many thousands of volunteers who contribute to our work. The Tor Project is funded in part by government research and development grants, and in part by individual, foundation, and corporate donations.

Salary for this position depends on experience and there is voluntary opt-in salary transparency for employees and contractors.

The Tor Project has a competitive benefits package, including a generous PTO policy, 16 paid holidays per year (including the week between Christmas and New Year's, when the office is closed), and flexible work schedule. Insurance benefits vary by employment status and country of residence.

The Tor Project, Inc., is an equal opportunity, affirmative action employer.