

Objective 2: Generate a specification for statistical assessment of aggregate data.

- **Activity 2.1: Document where Tor users are from and how they connect to Tor.**
- **Activity 2.2: How many relays and bridges are online and what we know about them.**
- **Activity 2.3: How much traffic the Tor network can handle and how much traffic there is.**
- **Activity 2.4: Write specification for assessing how fast and reliable the Tor network is.**
- **Activity 2.5: How many onion services there are and how much traffic do they pull.**
- **Activity 2.6: How many applications like Tor Browser have been downloaded/updated.**

We published a documentation page on Tor Metrics that we're calling "Reproducible Metrics". In short, this page specifies how one can reproduce the data behind graphs on Tor Metrics by using original Tor network data as input.

The page is here:

<https://metrics.torproject.org/reproducible-metrics.html>

Sources – Tor Metrics

https://metrics.torproject.org/reproducible-metrics.html

News Sources Services Development Research About

Home Users Servers Traffic Performance Onion Services Applications

Home » Sources » Reproducible Metrics

Work in progress notice

As of July 2018, this page is still a work in progress. Handle with care!

Reproducible Metrics

The graphs and tables on Tor Metrics are the result of aggregating data obtained from several points in the Tor network. Some of these aggregations are straightforward, but some are not.

We want to make the graphs and tables on this site easier to access and reproduce, so on this page, we specify how you can reproduce the data behind them to create your own. We also provide background for some of the design decisions behind our aggregations and link to [technical reports](#) and other additional information.

This page is a living document that reflects the latest changes to graphs and tables on Tor Metrics. Whenever we create new aggregations or visualizations, we may write down our thoughts in technical reports; but, if we later expand or change a statistic, we don't update the original technical reports. Instead, we update the specification here.

While we may refer to technical reports for additional details, we do not assume their knowledge in order to make sense of the specifications here. Knowledge of our source code is not needed, either.

Users

The number of Tor users is one of our most important statistics. It is vital for us to know how many people use the Tor network on a daily basis, whether they connect via [relays](#) or [bridges](#), from which countries they connect, what [transports](#) they use, and whether they connect via

This page is marked as work in progress, because we still want to show it to experts in our field and ask them to review and maybe even try out these instructions as a way to measure the quality of our work. However, other than suggested revisions coming from those experts, we're not planning to make any further changes to this document. It's like a feature that we built and that we're going to maintain, in case there are bugs.

In addition to the "Reproducible Metrics" page, we extended another web page on Tor Metrics with "Statistics files used on this website":

<https://metrics.torproject.org/stats.html>

The screenshot shows a web browser window with the URL `https://metrics.torproject.org/stats.html`. The page has a purple header with the Tor logo and the word "Metrics". The navigation menu includes links for News, Sources, Services, Development, Research, and About. Below the header is a secondary navigation bar with icons and labels for Home, Users, Servers, Traffic, Performance, Onion Services, and Applications. The breadcrumb trail reads "Home » Sources » Statistics".

[Home](#) » [Sources](#) » [Statistics](#)

Statistics files used on this website

This page contains specifications of statistics files used on this website.

Parameters

All per-graph statistics files (excluding the now deprecated pre-aggregated files) are available for download via an URL of the form:

```
https://metrics.torproject.org/identifier.csv
```

These URLs all support a set of *optional* parameters that can be used to further customize their content. Typically, these are **start** and **end** plus additional parameters as specified below. Including a parameter in a URL typically filters the resulting statistics file by the given parameter value. In reverse, omitting a parameter produces a larger statistics file that is not filtered by that parameter.

Columns

Each per-graph statistics file (again, excluding the deprecated pre-aggregated files) starts with a comment section, followed by a header line and then the actual data lines. Columns are pre-defined and specified further down below. The rule of thumb for columns is that neither the choice of parameters nor availability of data should affect the set of columns, but that only a code change can add, change, or remove a column. This rule of thumb is not yet implemented for all per-graph statistics files.

Applications must not rely on the order of columns, as this order may change when columns are removed. Instead, applications should refer to columns by their name. Applications should be able to handle newly added columns and fail gracefully in case of removed columns.

Changes

The background is that we added new data files containing exactly the data as found in our graphs earlier this year. This page is the written documentation for requesting those data files, including possible customizations to contain more or less data. It is also a data format specification for the contents of these data files.

We have scheduled minor changes to these data files, also as a result of writing this specification page. But we need to give our users a heads-up before simply going ahead and making changes. It's also something we'd consider as maintenance, not as part of providing the documentation in the first place.

Both documents described above complete objective 2 with activities 2.1 to 2.6.