

# Tor Metrics Ecosystem

Data Collection, Archive, Analysis and Visualisation

Iain R. Learmonth (irl)

September 17, 2018

Tor Project



Tor Metrics Team Member

Background in Internet  
Measurement

@iainlearmonth  
@irl@mastodon.technology

Contributing to Tor Project since  
2015

# Tor Metrics

## Introduction

*The Metrics Team is a group of people who care about measuring and analyzing things in the public Tor network.*

# Tor Metrics

## Philosophy

We only use **public, non-sensitive data**. Each analysis goes through a rigorous review and discussion process before publication.

We **never** publish statistics—*even aggregate statistics*—of sensitive data, such as unencrypted contents of traffic.

# Tor Metrics

Research Safety Board

The goals of a privacy and anonymity network like Tor are not easily combined with extensive data gathering, but at the same time data is needed for monitoring, understanding, and improving the network.

Safety and privacy concerns regarding data collection by Tor Metrics are guided by the Tor Research Safety Board's guidelines.

<https://research.torproject.org/safetyboard.html>

<http://wcgqzqyfi7a6iu62.onion/safetyboard.html>

# Tor Metrics

## Key Safety Principals

1. Data minimalization
2. Source aggregation
3. Transparency

# Tor Metrics

## Data minimalization

The first and most important guideline is that only the **minimum amount** of statistical data should be gathered to solve a given problem. The **level of detail** of measured data should be as **small as possible**.

# Tor Metrics

## Source aggregation

Possibly sensitive data should exist for **as short a time as possible**. Data should be aggregated at its source, including **categorizing** single events and memorizing category counts only, **summing** up event counts over large time frames, and being **imprecise** regarding exact event counts.



# Tor Metrics

## Transparency

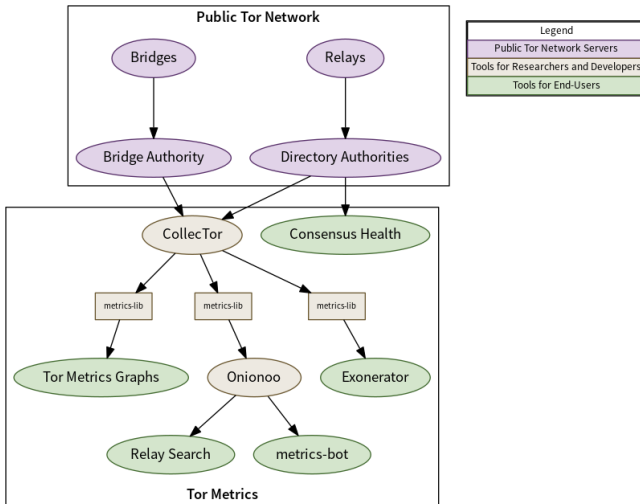
All algorithms to gather statistical data need to be **discussed publicly** before deploying them. All measured statistical data should be made **publicly available** as a **safeguard** to *not gather data that is too sensitive*.

Data and analysis can be used to:

- detect possible censorship events
- detect attacks against the network
- evaluate effects on performance of software changes
- evaluate how the network scales
- argue for a more private and secure Internet from a *position of data, rather than just dogma or perspective*

# Tor Metrics

Ecosystem



# Collecto**r**

## Introduction







Collecto**r** **fetches data** from various nodes and services in the public Tor network and **makes it available** to the world.

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

<http://rougmnvswfsmd4dq.onion/collector.html>

- Tor Relay Descriptors
  - Relay Server Descriptors
  - Relay Extra-info Descriptors
  - Network Status Consensuses
  - Network Status Votes
  - Directory Key Certificates
  - Microdescriptor Consensuses
  - Microdescriptors
- Tor Hidden Service Descriptors
- Tor Bridge Descriptors
  - Bridge Network Statuses
  - Bridge Server Descriptors
  - Bridge Extra-info Descriptors
- TorDNSSEL's Exit Lists
- Torperf's and OnionPerf's Performance Data
- Tor web server logs

## Index of /recent

| <a href="#">Name</a>  | <a href="#">Last modified</a> | <a href="#">Size</a> | <a href="#">Description</a> |
|---|-------------------------------|----------------------|-----------------------------|
|  <a href="#">Parent Directory</a>    |                               | -                    |                             |
|  <a href="#">bridge-descriptors/</a> | 2016-09-18 19:09              | -                    |                             |
|  <a href="#">exit-lists/</a>         | 2018-07-14 21:02              | -                    |                             |
|  <a href="#">relay-descriptors/</a>  | 2015-10-28 09:37              | -                    |                             |
|  <a href="#">torperf/</a>            | 2018-07-14 06:01              | -                    |                             |
|  <a href="#">webstats/</a>           | 2018-07-14 10:50              | -                    |                             |

*Apache Server at collector.torproject.org Port 443*

<https://collector.torproject.org/>

<http://qigcb4g4xxbh5ho6.onion/>

# Collector

## Accessing the data

```
#!/bin/sh
wget --recursive \                # turn on recursive retrieving
     --reject "index.html*" \     # don't retrieve indexes
     --no-parent \               # don't ascend to parent directory
     https://collector.torproject.org/recent/relay-descriptors/microdescs/
```

Another automated way to download descriptors is to develop a tool that uses the provided `index.json` file (or one of its compressed versions `index.json.gz`, `index.json.bz2`, or `index.json.xz`).

These files contain a machine-readable representation of all descriptor files available on this site.



Project idea alert!

Idea: **CollectorFS**

Write a **FUSE filesystem** that utilises the `index.json` file provided by collector to present files from Collector as if they were a local filesystem. Files should be downloaded and cached on demand.

Tor Metrics Library API (a.k.a. metrics-lib) is a **Java library** to **obtain and process descriptors** containing Tor network data.

<https://metrics.torproject.org/metrics-lib/>

<http://rougmnvswfsmd4dq.onion/>

# metrics-lib

## Example Descriptor

```
router milliways 83.68.131.4 9042 0 9030
master-key-ed25519 4ucDsjsjPHxC8K99hdgZFXHd4fDy5zpEBg2uBHB9zygk
or-address [2a01:190:1501:9050::1]:9042
platform Tor 0.3.3.8 on Linux
proto Cons=1-2 Desc=1-2 DirCache=1-2 HSDir=1-2 HSIntro=3-4 HSRender=1-2
  Link=1-5 LinkAuth=1,3 Microdesc=1-2 Relay=1-2
published 2018-07-14 17:28:37
fingerprint E59C C006 0074 E14C A8E9 4699 99B8 62C5 E1CE 49E9
uptime 194521
bandwidth 819200 1638400 702464
extra-info-digest 3306B53F8969F3B82903E5F22B40B5F2067453DF
  kHyXz1yPrw7kn98dnHqVwCDkQySBZ26Ptyu9SjK6thw
family $CF0CC69DE1E7E75A2D995FD8D9FA7D20983531DA
hidden-service-dir
contact 0xF540ABCD Iain R. Learmonth <irl@fsfe.org>
ntor-onion-key rFSc06l+7ByBC5huXeEX/FTdC+2C4RSoMMyzyPSuYks=
reject *:*
tunnelled-dir-server
router-sig-ed25519 IA3YlX7tL88eKSo0GLmbYiEA0zAa2NQ5M3jDeQ9sqa0/
  IE32sVvfwQUM+Pd20ZP3oUljJa5f40ozBPz63nZMCA
```

```
public interface RelayServerDescriptor  
extends ServerDescriptor
```

Contains a relay server descriptor.

Relay server descriptors share many contents with sanitized bridge server descriptors (**BridgeServerDescriptor**), which is why they share a common superinterface (**ServerDescriptor**). The main purpose of having two subinterfaces is being able to distinguish descriptor types more easily.

**Since:**

1.1.0

### ***Method Summary***

#### **Methods inherited from interface org.torproject.descriptor.**ServerDescriptor****

```
getAddress, getAllowSingleHopExits, getBandwidthBurst, getBandwidthObserved, getBandwidthRate,  
getCachesExtraInfo, getCircuitProtocolVersions, getContact, getDigestShalHex, getDigestSha256Base64,  
getDirPort, getExitPolicyLines, getExtraInfoDigestShalHex, getExtraInfoDigestSha256Base64,  
getFamilyEntries, getFingerprint, getHiddenServiceDirVersions, getIdentityEd25519, getIpv6DefaultPolicy,  
getIpv6PortList, getLinkProtocolVersions, getMasterKeyEd25519, getNickname, getNtorOnionKey,  
getNtorOnionKeyCrosscert, getNtorOnionKeyCrosscertSign, getOnionKey, getOnionKeyCrosscert, getOrAddresses,  
getOrPort, getPlatform, getProtocols, getPublishedMillis, getReadHistory, getRouterSignature,  
getRouterSignatureEd25519, getSigningKey, getSocksPort, getTunnelledDirServer, getUptime,  
getUsesEnhancedDnsLogic, getWriteHistory, isHibernating, isHiddenServiceDir
```

#### **Methods inherited from interface org.torproject.descriptor.**Descriptor****

```
getAnnotations, getDescriptorFile, getRawDescriptorBytes, getRawDescriptorLength, getUnrecognizedLines
```

# metrics-lib

Alternative: stem

stem is a Python library that includes parsers for various Tor descriptors. One notable feature of stem is that it can use a tor process to fetch descriptors live from the network. It also is able to check signatures on descriptors.

[https://stem.torproject.org/tutorials/mirror\\_mirror\\_on\\_the\\_wall.html](https://stem.torproject.org/tutorials/mirror_mirror_on_the_wall.html)

# metrics-lib

Alternative: zoossh

zoossh is a Go library that includes parsers for various Tor descriptors.  
zoossh is fast, but doesn't support as many descriptor formats as stem.

<https://gitweb.torproject.org/user/phw/zoossh.git/>

Project idea alert!

### Idea: **Extend a library**

Each of metrics-lib, stem and zoosh are incomplete when it comes to parsing every kind of descriptor currently in use in the wider Tor ecosystem. You could extend one of these libraries to add support for a descriptor that currently is not understood.

# Tor Metrics Statistics

## Introduction

### Analysis

View visualizations of statistics collected from the public Tor network and from Tor Project infrastructure.



#### Users

Where Tor users are from and how they connect to Tor.



#### Servers

How many relays and bridges are online and what we know about them.



#### Traffic

How much traffic the Tor network can handle and how much traffic there is.



#### Performance

How fast and reliable the Tor network is.



#### Onion Services

How many onion services there are and how much traffic they pull.



#### Applications

How many Tor applications, like Tor Browser, have been downloaded or updated.



<https://metrics.torproject.org/>  
<http://rougmnvswfsmd4dq.onion/>



# Tor Metrics Statistics

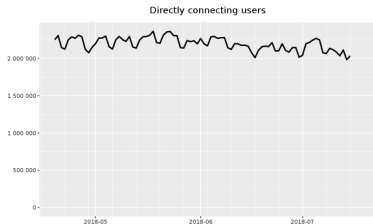
## Example Analysis

### Users

We estimate the number of users by analyzing the requests induced by clients to relays and bridges.

Relay users   Bridge users by country   Bridge users by transport   Bridge users by country and transport   Bridge users by IP version

Top-10 countries by relay users   Top-10 countries by possible censorship events   Top-10 countries by bridge users   "The anonymous Internet"



The Tor Project - <https://metrics.torproject.org/>

This graph shows the estimated number of directly-connecting [clients](#); that is, it excludes clients connecting via [bridges](#). These estimates are derived from the number of directory requests counted on [directory authorities](#) and [mirrors](#). Relays resolve client IP addresses to country codes, so that graphs are available for most countries. Furthermore, it is possible to display indications of censorship events as obtained from an anomaly-based censorship-detection system (for more details, see [this technical report](#)). For further details see these [questions and answers about user statistics](#).

**Start date:**

**End date:**

**Source:**

**Show possible censorship events if available (BETA):**

[Download graph as PNG or PDF.](#)

[Download data as CSV.](#)



<https://metrics.torproject.org/userstats-relay-country.html>

<http://rougmnvswfsm4dq.onion/userstats-relay-country.html>

# Tor Metrics Statistics

## Query Features

- Date Ranges
- Country
- Pluggable Transport
- IP Version

# Tor Metrics Statistics

## Export Formats

- PNG
- PDF
- CSV

# Tor Metrics Statistics

## Example CSV

```
1 #
2 # The Tor Project
3 #
4 # URL: https://metrics.torproject.org/userstats-
      relay-country.csv?start=2018-04-19&end=2018-07-
      18&country=all&events=off
5 #
6 date , country , users , downturns , upturns , lower , upper
7 2018-04-19 , , 2253583 , , , ,
8 2018-04-20 , , 2308749 , , , ,
9 2018-04-21 , , 2147036 , , , ,
10 2018-04-22 , , 2126204 , , , ,
11 2018-04-23 , , 2251922 , , , ,
12 2018-04-24 , , 2292202 , , , ,
13 2018-04-25 , , 2272599 , , , ,
14 2018-04-26 , , 2313660 , , , ,
15 2018-04-27 , , 2292282 , , , ,
16 2018-04-28 , , 2125045 , , , ,
17 2018-04-29 , , 2077537 , , , ,
18 2018-04-30 , , 2151478 , , , ,
```

Project idea alert!

Idea: **Tools for data journalists using Tor Metrics CSV files**

Create tools that make it easier for data journalists to create visualisations using Tor Metrics CSV files. This might include mash-ups with other data sources such as the CIA World Factbook or DBpedia.

<https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism>

Onionoo is a **web-based protocol** to learn about currently running Tor relays and bridges. Onionoo itself was not designed as a service for human beings—at least not directly. Onionoo **provides the data for other applications and websites** which in turn present Tor network status information to humans.

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

<http://rougmnvswfsmd4dq.onion/onionoo.html>

| Method | URL        | Description                  |
|--------|------------|------------------------------|
| GET    | /summary   | returns a summary document   |
| GET    | /details   | returns a details document   |
| GET    | /bandwidth | returns a bandwidth document |
| GET    | /weights   | returns a weights document   |
| GET    | /clients   | returns a clients document   |
| GET    | /uptime    | returns an uptime document   |

# Onionoo

## Example Summary Document

---

```
1 {"version":"6.1",
2  "build_revision":"eee9cf8",
3  "relays_published":"2018-07-16 20:00:00",
4  "relays":[
5    {"n":"seele","f":"000A10D43011EA4928A35F610405F92B4433B4
        DC","a":["67.161.31.147"],"r":true},
6    {"n":"CalyxInstitute14","f":"0011BD2485AD45D984EC4159C88
        FC066E5E3300E","a":["162.247.74.201"],"r":true},
7    {"n":"Neldoreth","f":"001524DD403D729F08F7E5D77813EF1275
        6CFA8D","a":["185.13.39.197"],"r":false}
8  ],
9  "relays_truncated":8109,
10 "bridges_published":"2018-07-16 19:51:42",
11 "bridges":[
12 ]}
```

---

<https://onionoo.torproject.org/summary?limit=3&type=relay>





What is it?

Donate!

News

**Services**

Spread the word

Documentation (french)

Contact

## Services

French



Nos oignons' weight in the Tor network




Probability to use one of our exit nodes

## Relays


Nos Oignons currently runs the following Tor relays:

| Location              | Relay                                | Fingerprint  | Exit policy                         |
|-----------------------|--------------------------------------|--|-------------------------------------|
| <a href="#">Liazo</a> | <a href="#">marcuse1<sup>1</sup></a> | EFAE 4472 8264 9822 2444 5E96 214C 15F9<br>075D EE1D | <a href="#">Reduced Exit Policy</a> |
| <a href="#">Liazo</a> | <a href="#">marcuse2<sup>2</sup></a> | C656 B41A EFB4 0A14 1967 EBF4 9D6E 6960<br>3C9B 4A11 | <a href="#">Reduced Exit Policy</a> |
|                       |                                      | 0D18 4500 0000 3070 5000 7107 0001 0500              | <a href="#">Reduced Exit Policy</a> |

<https://nos-oignons.net/Services/index.en.html>



**ORNETSTATS**  
Stats about the Tor Network



## OrNetStats

OrNetStats shows you statistics about the Tor network.

Tor network data as of: **2018-07-16 22:00 UTC**



## Tor Relay Operators in End-to-End Correlation Position

The following table lists relay operators that are in a position to see a tor client's entry and exit connections. In the **worst-case a tor client would use these groups as entry (guard) and exit relay at the same time.**

Operators are only listed if they actually have a chance to do end-to-end correlation attacks, that is:

- their guard **and** exit probability is > 0%
- they did **not** properly configure [MyFamily](#)
- they run in **more** than a single /16 network block

This list might contain false-positives as [ContactInfo](#) is not authenticated.

The [ContactInfo](#) is truncated. Middle-only relays are not included in per-group relaycounts.

The table is sorted by guard probability.

| Contact  | Guard (%)   | Exit (%)    | #Relays /16 Netblocks | Newest Relay | Eff. Family Members (min) |
|--|-------------|-------------|-----------------------|--------------|---------------------------|
| <a href="#">pm@dppj.ru</a> -<br><a href="#">1Hr5ALwotveTsEJpxuyox2en6d62ZVedfs</a> | 0.19        | 0.19        | 3 3                   | 2018-06-22   | 2                         |
| <a href="#">tor at releasing dot fun</a>   | 0.02        | 0.16        | 4 4                   | 2018-07-04   | 1                         |
| <b>Total</b>   | <b>0.21</b> | <b>0.35</b> | <b>7</b>              |              |                           |

For a detailed list of (known) relays in end-to-end correlation position see [this page](#).

**NOTE:** There are many more relays with [MyFamily](#) configuration issues but most operate exit or guard relays exclusively or within a single /16 network block. Such operators can not become the first **and** last hop of your tor circuits, but they might be able to reveal your guard relay (when they act as the middle and exit relay in a single circuit).

<https://nusenu.github.io/OrNetStats/>

- OnionPy  
<https://github.com/duk3luk3/onion-py>
- onionoo-node-client  
<https://github.com/lukechilds/onionoo-node-client>
- tormetrics (PowerShell module)  
<https://github.com/lmillanta/tormetrics>
- konionoo<sup>1</sup> (Java CLI tool)  
<https://savannah.nongnu.org/projects/konionoo/>

---

<sup>1</sup>This is currently unmaintained

Project idea alert!

Idea: **New client library or command line tool**

Write a library or command-line tool using your favourite programming language for querying Onionoo. Queries should be cached.

# Relay Search

## Introduction

The relay search tool displays [data about relays and bridges](#) in the Tor network. It provides useful information on **how relays are configured** along with **graphs about their history**.

Relay Search is an [Onionoo client](#).

# Relay Search

## Introduction

## Relay Search

Simple Search

Aggregated Search

Advanced Search

The relay search tool displays data about single relays and bridges in the Tor network. It provides useful information on how relays are configured along with graphs about their past.

Query

Search

Top Relays

You can search for Tor relays and bridges by using keywords. In particular, this tool enables you to search for (partial) nicknames (e.g., "moria"), IP addresses (e.g., "128.31."), and fingerprints (e.g., "9695DFC3"). It is also possible to combine searches, e.g., "moria 128.31.". Finally, you can use qualifiers to search for relays in specific countries (e.g., "moria country.us"), with specific contact information (e.g., "contactarma"), or with specific flags (e.g., "flag:Authority").

If you are searching for a bridge, you will need to search by the hashed fingerprint. This prevents leaking the fingerprint of the bridge when searching. You can find this in the `hashed-fingerprint` file in the Tor data directory. On Debian systems, this is in `/var/lib/tor` but may be in another location on your system. The location is specified as `DataDirectory` in your `torrc`.



TODO

TODO



TODO