

Tor Metrics Ecosystem

Data Collection, Archive, Analysis, and Visualisation

Iain R. Learmonth (irl)

October 3, 2018

Tor Project



Tor Metrics Team Member

Background in Internet
Measurement

irl@torproject.org
@iainlearmonth
@irl@chaos.social

Contributing to Tor Project since
2015

A8F7 BA50 41E1 3333 9CBA 1696 76D5 8093 F540 ABCD

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.

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

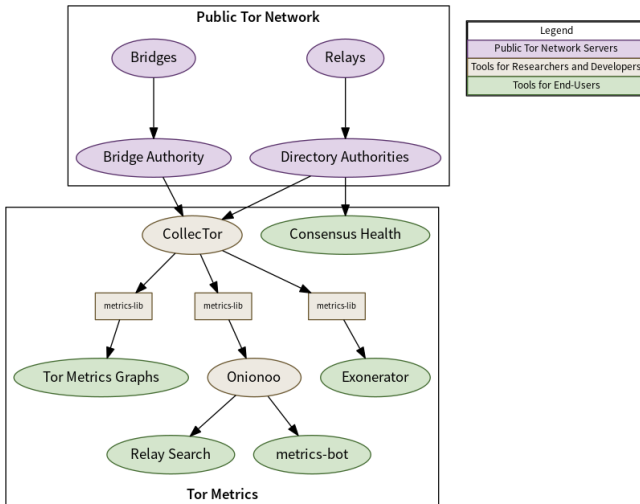
All algorithms to gather statistical data need to be **discussed publicly** before deploying them. All measured statistical data should be made **publicly available** as a **safeguard** to *not gather data that is too sensitive*.

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



CollecTor

Introduction







CollecTor **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 Bridge Descriptors
 - Bridge Network Statuses
 - Bridge Server Descriptors
 - Bridge Extra-info Descriptors
- TorDNSEL's Exit Lists
- Torperf's and OnionPerf's Performance Data
- Tor web server logs

Index of /recent

Name	Last modified	Size	Description
 Parent Directory		-	
 bridge-descriptors/	2016-09-18 19:09	-	
 exit-lists/	2018-07-14 21:02	-	
 relay-descriptors/	2015-10-28 09:37	-	
 torperf/	2018-07-14 06:01	-	
 webstats/	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.

metrics-lib

Introduction

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 4ucDsJwPHxC8K99hdgZFXHd4fDy5zpEBg2uBhb9zygk
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+2C4RS0MMyzyPSuYks=
reject *:*
tunnelled-dir-server
router-sig-ed25519 IA3YlX7tL88eKSo0GLmbYiEA0zAa2NQ5M3jDeQ9sqa0/
    IE32sVvfWQUM+Pd20ZP3oUlJJJa5f40ozBPz63nZMCA
```

metrics-lib

Parsing Relay Descriptors

```
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
```

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

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**

metrics-lib is 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 version 3 bandwidth lists.

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

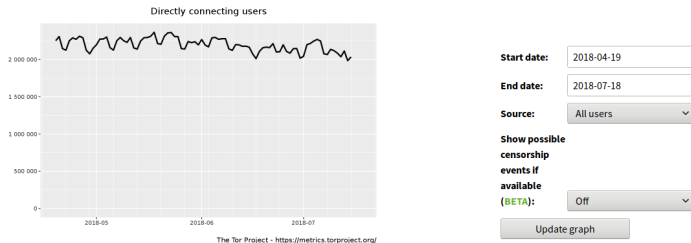
Bridge users by IP version

Top-10 countries by relay users

Top-10 countries by possible censorship events

Top-10 countries by bridge users

"The anonymous Internet"



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](#).



<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, , , ,
```

Tor Metrics Statistics

Web Server Tweaks

Project idea alert!

Idea: **Disabling session cookie**

Tor Metrics uses Jetty as a web server which currently sets a global session cookie. We have no use for this so would like to disable it.

Tor Metrics Statistics

Tor Browser Update Pings by Locale and Platform

Project idea alert!

Idea: **Which locales and platforms are popular? (#27931, #27932)**

We have plots on Tor Metrics to show initial downloads by locale and platform but for update pings we only have a global aggregate. We already have the data in our database but need to turn it into plots.

Tor Metrics Statistics

Modernizing R Code

Project idea alert!

Idea: **Using dplyr and tidyr (#22423)**

We use R to produce the plots you will find on Tor Metrics and have started to update our codebases using the dplyr and tidyr packages but have not yet updated all of the code.

Tor Metrics Statistics

Helping Data Journalism

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

Introduction

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>

Onionoo

API Overview

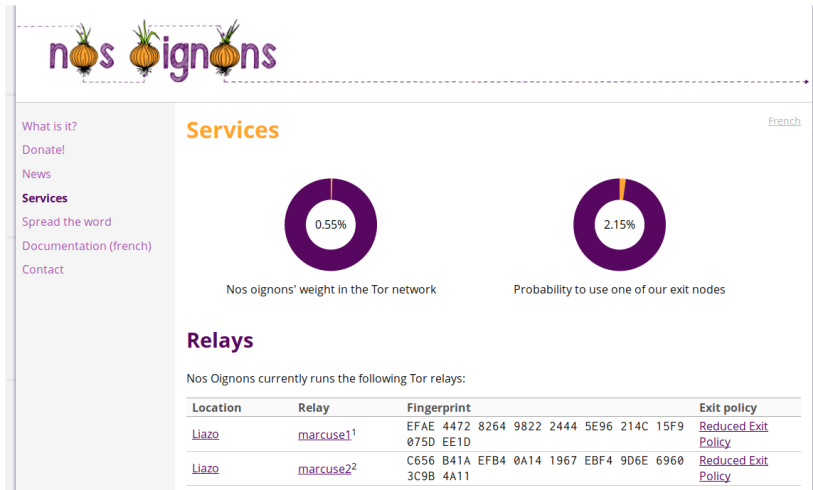
Method	URL	Description
GET	/summary	returns a summary document
GET	/details	returns a details document
GET	/bandwidth	returns a bandwidth document
GET	/weights	returns a weights document
GET	/clients	returns a clients document
GET	/uptime	returns an uptime document

Onionoo

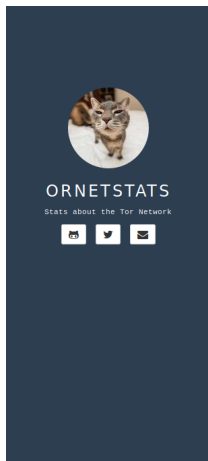
Example Summary Document

```
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>



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



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)
pm@dpjp.ru - 1Hr5ALwotveTsEJpxuyox2en6d62ZVedfs	0.19	0.19	3	3	2018-06-22	2
tor at releasing dot fun	0.02	0.16	4	4	2018-07-04	1
Total	0.21	0.35	7			

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>

Project idea alert!

Idea: **Get Onionoo ready for changes to the GeoIP database**

The format of the GeoIP database will soon be changing and we need to be ready for this change. Replace the current GeoIP lookup functions with functions that will work with the new format.

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., "contact:arma"), 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`.



Exonerator

Introduction

The ExoneraTor service maintains a database of IP addresses that have been part of the Tor network. It answers the question whether there was a Tor relay running on a given IP address on a given date.

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

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

Exonerator

Looking up an IP address

IP address

171.25.193.77

Date

2018-09-10

Search

Summary

Result is positive

We found one or more Tor relays on IP address 171.25.193.77 on or within a day of 2018-09-10 that Tor clients were likely to know.

Technical details

Looking up IP address 171.25.193.77 on or within one day of 2018-09-10. Tor clients could have selected this or these Tor relays to build circuits.

Timestamp (UTC)	IP address(es)	Identity fingerprint	Nickname	Exit relay
2018-09-09 00:00:00	171.25.193.77, [2001:67c:289c:3::77]	A10C4F666D27364036B562823E5830BC448E046A	DFRI1	Yes
2018-09-09 01:00:00	171.25.193.77, [2001:67c:289c:3::77]	A10C4F666D27364036B562823E5830BC448E046A	DFRI1	Yes
2018-09-09 02:00:00	171.25.193.77, [2001:67c:289c:3::77]	A10C4F666D27364036B562823E5830BC448E046A	DFRI1	Yes
2018-09-09 03:00:00	171.25.193.77, [2001:67c:289c:3::77]	A10C4F666D27364036B562823E5830BC448E046A	DFRI1	Yes





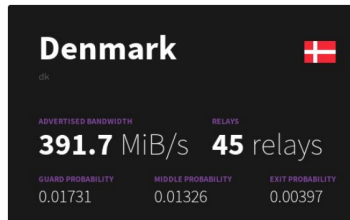
Tor Atlas

@TorAtlas

Follow



45 relays in Denmark are contributing 391.7 MiB/s bandwidth to the #Tor network.
metrics.torproject.org/rs.html#search...



2:33 am - 25 Sep 2018

7 Likes





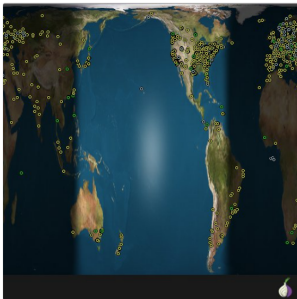
Tor Atlas

@TorAtlas

Follow



There are currently 6529 #Tor relays running providing 32.5 GiB/s total bandwidth.
[metrics.torproject.org/networksize.ht...](https://metrics.torproject.org/networksize.html)



2:33 pm · 2 Oct 2018

Consensus Health

Consensus Health shows statistics about the current consensus and votes to facilitate debugging of the directory consensus process.

<https://consensus-health.torproject.org/>
<http://tgnv2pssfumdedyw.onion/>

Consensus Health

Most well known for the detailed votes table:

Fingerprint	Nickname	maatu.	tor26	longc.	dizum	bastet	gabel.	morla
000A10D4	seele Relay Search ↵	Running Stable V2Dir Valid bw=12	Running Stable V2Dir Valid	Running Stable V2Dir Valid	Running Stable V2Dir Valid	Running Stable V2Dir Valid bw=40	Running Stable V2Dir Valid bw=11	Running Stable V2Dir Valid bw=20
000C1F7C	PutoElQueLee293884 Relay Search ↵	Fast Guard Running Stable V2Dir Valid bw=10200	Fast Guard Running Stable V2Dir Valid	Fast Guard Running Stable V2Dir Valid	Fast Guard Running Stable V2Dir Valid	Fast Guard Running Stable V2Dir Valid bw=21200	Fast Guard Running Stable V2Dir Valid bw=4460	Fast Guard Running Stable V2Dir Valid bw=60
0011BD24	CalyxInstitute14 Relay Search ↵	Exit Fast Guard HSDir Running Stable V2Dir Valid bw=17400	Exit Fast Guard HSDir Running Stable V2Dir Valid	Exit Fast Guard HSDir Running Stable V2Dir Valid	Exit Fast Guard HSDir Running Stable V2Dir Valid	Exit Fast Guard HSDir Running Stable V2Dir Valid bw=20700	Exit Fast Guard HSDir Running Stable V2Dir Valid bw=9690	Exit Fast Guard HSDir Running Stable V2Dir Valid bw=18
001524DD	Neldoreth Relay Search ↵	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir	FallbackDir Fast Guard Running Stable V2Dir

Votes for each relay from every directory authority are shown side-by-side for easy comparison.