toreator

Tor relay parser and UI

User guide

Libor Polčák

TARZAN project VI20172020062 document
Faculty of Information Technology, Brno University of Technology

Last change: April 24, 2020

toreator — User guide

Libor Polčák

Faculty of Information Technology, Brno University of Technology, e-mail:
polcak@fit.vutbr.cz

This software processes data about Tor network. The main source is Tor CollecTor[1], which is available for research purposes. The tool augment the data with data from MaxMind GeoLite2 data[2] and data fetched from public DNS.

CollecTor contains all regisered public nodes that were usable as entry nodes, exit nodes, and relays. Neither Tor bridges nor Tor users are stored in the database.

Conceptually, this tool was inspired by the service Exonerator[3] and Torstatus[4]. The database created by this tool differs from:

- *Exonerator* because it contains more details about the Tor nodes;
- *Torsatus* because it contains historical data.

The database is available as a REST API[5] and allows to fetch structured data as JSON (for machine processing) and HTML (for human processing). Data encoding is controlled by the `Accept` header.

You should set up the program (see install guide) so that the database is automatically updated. CollecTor updates once per hour. MaxMind updates once per week (ASN database) or month (location database).

1 Tool use cases

This section describes the anticipated use cases for the Toreator API and complementary web service.

1.1 Law enforcement agencies

A detective learns an IP address that was revealed during an investigation to be connected with a crime. The detective can consult the Toreator database. If the IP address was not a part of the Tor network, the detective can continue with the investigation according to regular schedule.

[1] https://metrics.torproject.org/collector.html
[2] http://www.maxmind.com/
[3] https://exonerator.torproject.org/
[4] http://torstatus.blutmagie.de/
[5] You can try http://toreator.fit.vutbr.cz/ for an example instance of the API. The FIT Toreator database was created as part of the *Integrated platform for analysis of digital data from security incidents* project. This database contains data about Tor network since January 2008.

However, for IP addresses that are part of the Tor network, the investigator will need to take additional steps. The detective should learn if the relay was an exit or not. If so, learning the real identity of the offender might be impossible.

we anticipate that such investigation will involve an IP address and a time. See Fig. 1 to see an example how such a search might look like.

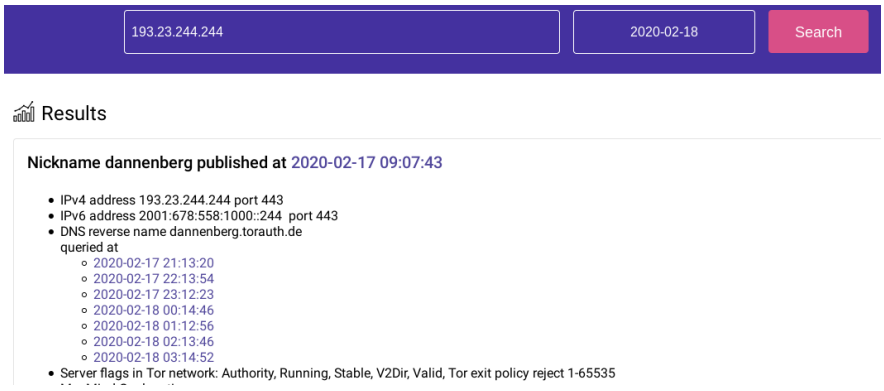

Fig. 1. A detective inserts an IP address and time into the tool to learn that the IP address was a part of the Tor network.

The REST API allows law enforcement to develop own tools that use data from Toreator database.

1.2 Network administrator

A network administrator might be interested if there are any Tor relays in the maintained network. This information can be useful, for example, to detect nodes that violate network policy as the Tor relay inevitably consume some bandwidth, electricity and computing power. It is also possible that a botnet installs Tor software. Such relays can be revealed by Toreator.

As Fig. 2 shows, the network administrator is interested only in the maintained network. The REST API can be used to develop custom tools that automatically obtains the updated information.

1.3 Tor relay operator

A Tor relay operator can be interested about the state of their node, or what information about the node is publicly available. Hence the operator performs a search for the IP address of the relay, as already depicted in Fig. 1.

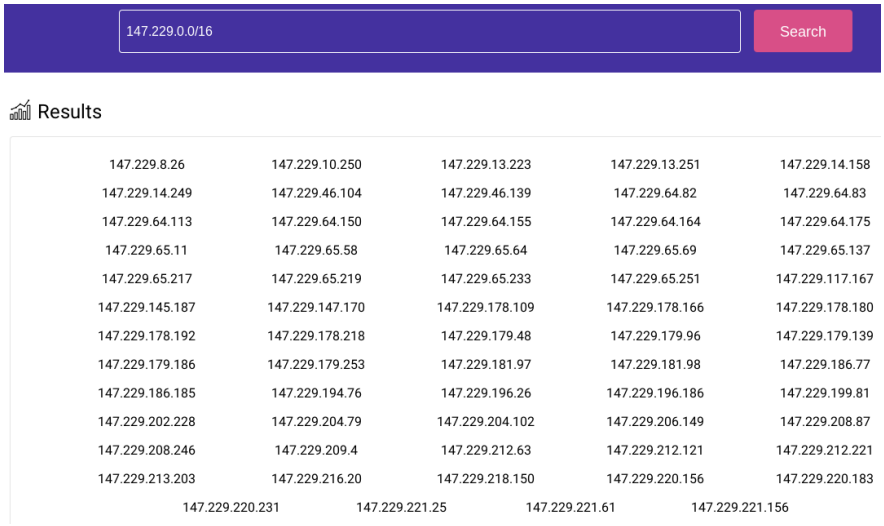

Fig. 2. A network administrator inserts an IP address range to learn Tor relays in the maintained network.

1.4 Tor network user

Tor users are often interested about the network state. Where are the relays located? Who operates them. Toreator can shed more light for these users. As Fig. 3 shows, the users can perform different actions depending on their interests. Besides searching (a), they have the opportunity to browse the networks and look for the relays (b). Once they find the relay, the users can be interested in the domain name of the relay (c) and geographic and network location of the node (d).

2 Personal data in the database

Generally speaking, the database created by this tool does not contain personal data as defined by EU Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) as the Tor nodes typically operate independently on their operator, see for example The Tor Relay Guide[6]. However, some volunteers provide their own computers and in such case, the IP addresses and DNS names identifies their computers. In such cases, the database contains the personal data based on the legitimate interests for scientific research. Hence, by running the package, you become a controller

[6] https://trac.torproject.org/projects/tor/wiki/TorRelayGuide

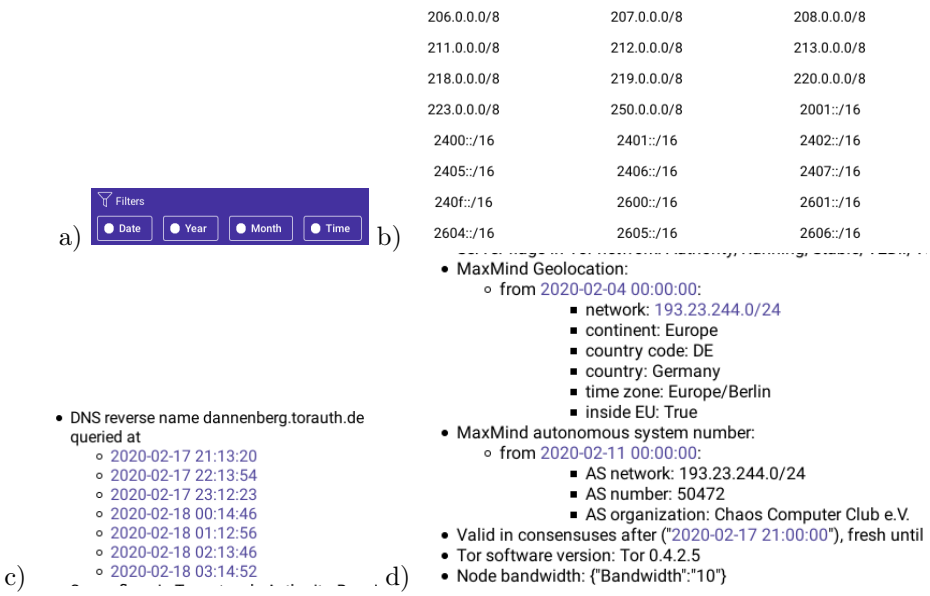

Fig. 3. A user might be interested in many attributes that our tool provides.

and need to asses the legal basis for processing the data, possibly perfom inga balance test if you can process the data based on legitimate interests.

FIT hoste the database based on legitimate interests, as we believe, our legitimate interests are not overriden by the rights of individuals for the following reasons:

1. We process only information that is already public (CollecTor, DNS) or available after simple registration (MaxMind).
2. Tools like Exonerator, Torsatus, and CollecTor already contain very similar data and are available for public use.
3. The web page https://metrics.torproject.org/collector.html explicitly states: *If you're doing research on the Tor network, or if you're developing an application that uses Tor network data, this is your place to start.*.
4. Web pages describing Tor relay setup warn operators that all relays will be listed in the public list of Tor relays. Operators that do not want their data to be processed can operate a bridge.
5. The use cases described in this document shows that the tool is beneficial for a wide range of use cases and actors, including for the security purposes.