DHT-crawler
BitTorrent Distributed Hash Table Monitor

User guide

Libor Polčák

DHT-crawler — User guide

Libor Polčák

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

BitTorrent is a major file sharing peer-to-peer network. BitTorrent is decentralised and the anyone can share content through the network. Although BitTorrent network is used to distribute content in accordance to copyright (for example, by GNU/Linux distributions[1]), the network is often misused for file sharing without necessary rights[2].

The aim of this document is to introduce the tool DHT-crawler[3] developed by *Integrated platform for analysis of digital data from security incidents* project. For the terminology and introduction to the Distributed Hash Table (DHT) see, for example, the *Integrated platform for analysis of digital data from security incidents* technical report[1].

1 Tool use case — Who is sharing my data?

In the case of stolen data, such as the database leaked from the Mall.cz or for the prevention of violation of copyright, it is beneficial to monitor BitTorrent network. Nevertheless, such monitoring is beneficial also in case of legal sharing. For example, GNU/Linux distribution can maintain statistics of the number of users participating in file sharing.

The DHT-crawler focuses on DHT monitoring based on a known infohash. It is up to the user to select the infohash to be monitored.

2 An example of DHT-crawler usage

This document anticipates that DHT-crawler is already installed according to the installation manual. See the installation manual for more details.

You should know the infohash that you want to search for. If you want to try the program, you can try an arbitrary infohash, for example:

The Pirate Bay - UNIX list: `https://thepiratebay.org/rss/top100/303`:

```
wget https://thepiratebay.org/rss/top100/303 -O - | less
```

For example, Tails v3.6 the amnesic incognito live system is distributed with infohash of 17CAD44F3C43C98F46813A28859C78BEB2D37D0F.

To search for current peers distributing Tails, run:

---

[1] `https://thepiratebay.org/rss/top100/303`, `https://torrent.fedoraproject.org/`

[2] For example, `https://thepiratebay.org/rss/top100/201`

[3] `https://github.com/polcak/DHT-crawler`

```
$ dht_crawler/exec.py --hash 17CAD44F3C43C98F46813A28859C78BEB2D37D0F \
--db_format --duration 60
```

The DHT-crawler outputs a JSON like:

```
{"<the searched infohash>":
{"peers": [
{"timestamp":<time when the peer was detected>,
   "addr": ["<peer IP address>", <peer TCP port>]
  },
  ... list of detected peers ...
]
"nodes": [
{"timestamp": <time when the node was detected>,
   "nodeID": "<the DHT node ID>",
   "nodeAddr": ["<node IP address>", <node TCP port>]},
... list of peers maintaining the DHT for the infohash ...
]
}
}
```

It is up to the user to process and use the information. The JSON format
helps into using the program in a pipeline with other automated tools.

Bibliography

[1] Polčák, L.: Co skrývá síť BitTorrent? Technická zpráva, 2018, only available in Czech.
URL https://www.fit.vut.cz/research/publication/11885