

Pridobivanje podatkov iz omrežja DHT

Analiza pretočnega prometa skozi vozlišča protokola BitTorrent in prenos metapodatkov

Anton Luka Šijanec

Mentor: Andrej Šuštaršič

Gimnazija Bežigrad
Ljubljana

15. maja 2023

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

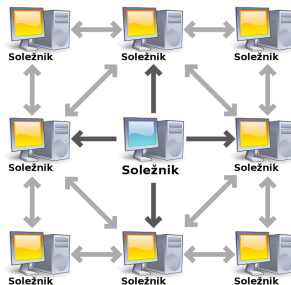
Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Opis protokolov

Protokol BitTorrent

- P2P izmenjava podatkov brez centralnega strežnika
- soležnik, roj, torrent, metapodatki

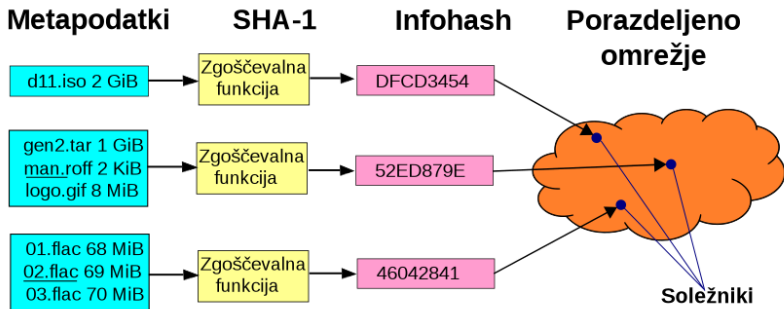


Slika: Shematski prikaz povezav med soležniki

Opis protokolov

Protokol DHT

- Kademia — veliko vozlišč, povezanih v graf

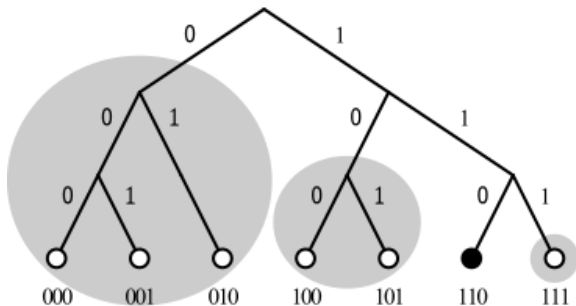


Slika: Shematski prikaz DHT

Opis protokolov

Protokol DHT — proces iskanja

- Kompleksnost iskanja je $O(\log n)$



Slika: Drevesno iskanje soležnikov

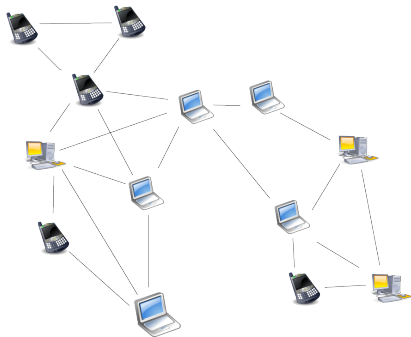
Opis protokolov

Protokol DHT — proces iskanja

- Vsako vozlišče
 - Ima naključno generiran ID
 - Hrani seznam soležnikov
 - Hrani usmerjevalno tabelo (seznam) drugih vozlišč (usmerjene povezave na grafu)
- Poizvedba po soležnikih
 - Zahteva: UDP paket z zgoščeno vrednostjo
 - Odgovor:
 - Soležniki (če jih vozlišče ima)
 - $K = 8$ vozlišč iz usmerjevalne tabele, ki imajo ID blizu zgoščene vrednosti v poizvedbi

Osnovna predstavitev koncepta

- Vozlišče pridobiva seznam obstoječih torrentov s shranjevanjem zgoščenih vrednosti iz poizvedb in nato za vsako zgoščeno vrednost poizkusi prenesti metapodatke.



Slika: Shematski prikaz grafa vozlišč DHT

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - **Namen raziskave**
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Namen raziskave

- Predstaviti metodo učinkovitega odkrivanja podatkov
- Podatki se lahko uporabljajo za
 - analizo vsebin, ki se prenaša po omrežju BitTorrent,
 - analizo reprezentativnosti programske opreme,
 - določanje izvora uporabnikov omrežja,
 - predvidevanje preferenc uporabnikov omrežja (korelacija tipov vsebine), ...
- V nalogi je podroben opis implementacije

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - **Podobni obstoječi projekti**
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Podobni obstoječi projekti

- Btdigg — odkrivanje vsebin
- I know what you download — določanje popularnosti vsebin skozi čas, shranjevanje identitete prenašalcev
- Izpopolnitev v tej nalogi:
 - Odprtokodna implementacija
 - Začetek analiziranja
 - Opis koncepta za preprostejše nadgrajevanje

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 **Praktično delo**
 - **Program za iskanje**
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Program za iskanje

- Travnik v C za POSIX
- Povezava v omrežje, prenos vsake datoteke z metapodatki
- Brez stanja: možno iskati na več strežnikih hkrati
- Neinvazivno/nemoteče delovanje za omrežje

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 **Praktično delo**
 - Program za iskanje
 - **Poganjanje programa — raziskovalno okolje**
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Poganjanje programa — raziskovalno okolje

- Domači optični priključek: en torrent na 29 sekund
 - Težava: veliko UDP paketov ohromi linijo
- Dva VPS GrNet v Grčiji: en torrent na 3,5 sekunde
- rsync

Analiza podatkov

Pregled

- Razčlenjevalnik torrentov v Pythonu
- Obdelava v Jupyter Notebook
- 285000 torrentov do 22. marca 2023, prejeti od 174000 različnih IP naslovov
- Metapodatki o 15 milijonih datotek v skupni velikosti 1,55 petabajtov
- Do 14. maja 2023 pa 773000 torrentov

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Pogonjanje programa — raziskovalno okolje
- 3 Rezultati
 - **Primer strukture torrent datoteke z metapodatki**
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Analiza podatkov

Primer strukture torrent datoteke z metapodatki

- Zgoščena vrednost:

696802a16728636cd72617e4cd7b64e3ca314e71

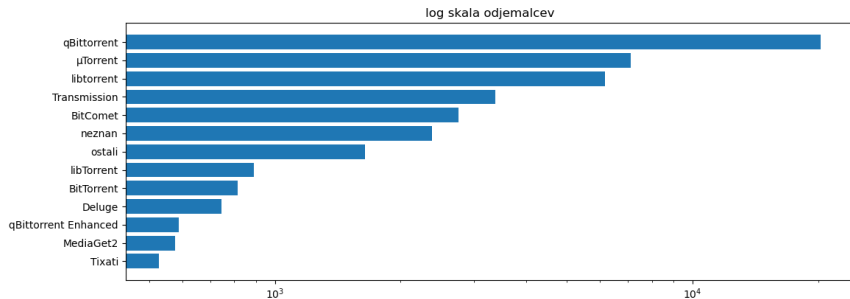
```
1 {
2   "creation date": 1676645016,
3   "encoding": "UTF-8",
4   "info": {
5     "files": [
6       { "length": 5903711328,
7         "path": [ "John.Wick.Chapter.2.2017.2160p.4K.BluRay.x265.10bit.AAC5.1-[YTS.MX].mkv" ] },
8       { "length": 358, "path": [ "YTSPproxies.com.txt" ] },
9       { "length": 53226, "path": [ "www.YTS.MX.jpg" ] }
10    ],
11    "name": "John Wick Chapter 2 (2017) [2160p] [4K] [BluRay] [5.1] [YTS.MX]",
12    "piece length": 4194304
13  },
14  "source": { "ip": "::ffff:185.242.21.95/9090", "v": "Transmission 3.00" }
15 }
```

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Pogonjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - **Programska oprema odjemalcev**
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Analiza podatkov

Programska oprema odjemalcev



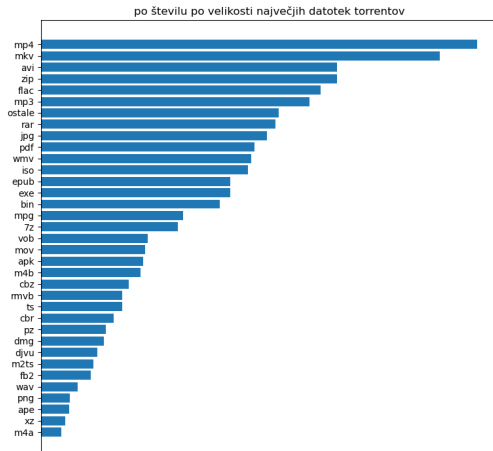
Slika: Reprezentativnost odjemalcev, ki predstavljajo vsaj odstotek populacije (*log skala*)

Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Analiza podatkov

Tipi datotek, ki se prenašajo v torrentih



Pregled vsebine

- 1 Uvod in teoretične osnove
 - Opis protokolov
 - Namen raziskave
 - Podobni obstoječi projekti
- 2 Praktično delo
 - Program za iskanje
 - Poganjanje programa — raziskovalno okolje
- 3 Rezultati
 - Primer strukture torrent datoteke z metapodatki
 - Programska oprema odjemalcev
 - Tipi datotek, ki se prenašajo v torrentih
 - Spletni vmesnik za iskanje datotek

Analiza podatkov

Spletni vmesnik za iskanje datotek

- Načini iskanja: regex z vsaj $O(n)$, iskalni indeks ključnih besed z $O(1)$

A search interface with a search bar containing 'Inkscape' and a 'Search' button. Below the search bar are two dropdown menus, both set to 'Normal'. The results section is titled 'Found 6 results' and lists four items:

- Inkscape-0a00cf5-x86_64.ApplImage**
size: 114.52 MB found: 3/18/2023 8:06:01 AM files: 1
- Inkscape 1.2.2 + Portable**
size: 414.6 MB found: 3/18/2023 4:31:38 PM files: 4
- Inkscape-1.1.2.dmg**
size: 127.1 MB found: 1/31/2023 6:43:59 AM files: 1
- Inkscape-1.2.1_x86_64.dmg**
size: 151.77 MB found: 3/20/2023 4:22:20 AM files: 1

Inkscape 1.2.2 + Portable

magnet	magnet
found	Sat Mar 18 2023 16:31:38 GMT+0100 (Central European Standard Time)
length	414.5961 MB
hash	8adabd7380a89fd25b117657848ba18605dc745
files	4
ip	89.28.165.253
found by	snf-59327

Files:

- Inkscape 1.2.2 + Portable/Inkscape-1.2.2_2022-12-09_732a01da63-x86.7z **106.05 MB**
- Inkscape 1.2.2 + Portable/Inkscape-1.2.2_2022-12-09_732a01da63-x86.7z **103.0 MB**
- Inkscape 1.2.2 + Portable/Inkscape-1.2.2_2022-12-09_732a01da63-x86.exe **103.47 MB**
- Inkscape 1.2.2 + Portable/Inkscape-1.2.2_2022-12-09_732a01da63-x86.exe **101.28 MB**

Slika: Primer uporabniškega vmesnika za iskanje

Konec predstavitve

- Hvala za pozornost!
- Vprašanja, prosim!
- Izvorna koda programov: <http://4a.si./prog> (git)
- Izvorna koda članka in predstavitve: <http://4a.si./clan> (git)
- 700000+ torrent datotek (~31 GiB) pošljem na zahtevo.
- anton@šijanec.eu.

Zahvala

- Oliverju Wagnerju (oliwerix.com.) in Adrianu Sebastianu Šiški (ass.si.) za pomoč pri obdelavi velike količine podatkov
- GrNet in njihovemu projektu ~okeanos (okeanos-global.grnet.gr.) za brezplačne VPS za raziskovanje

Viri slik in dovoljenje za uporabo

- Uporaba generiranega dokumenta je dovoljena v skladu s CC BY-SA (<http://4a.si./bysa>)



- Slika 1: Scott Martin: spremenjena, izvorna pod CC BY-SA
- Slika 3: Limaner: izvorna pod CC BY-SA