BitTorrent – from swarms to collectives

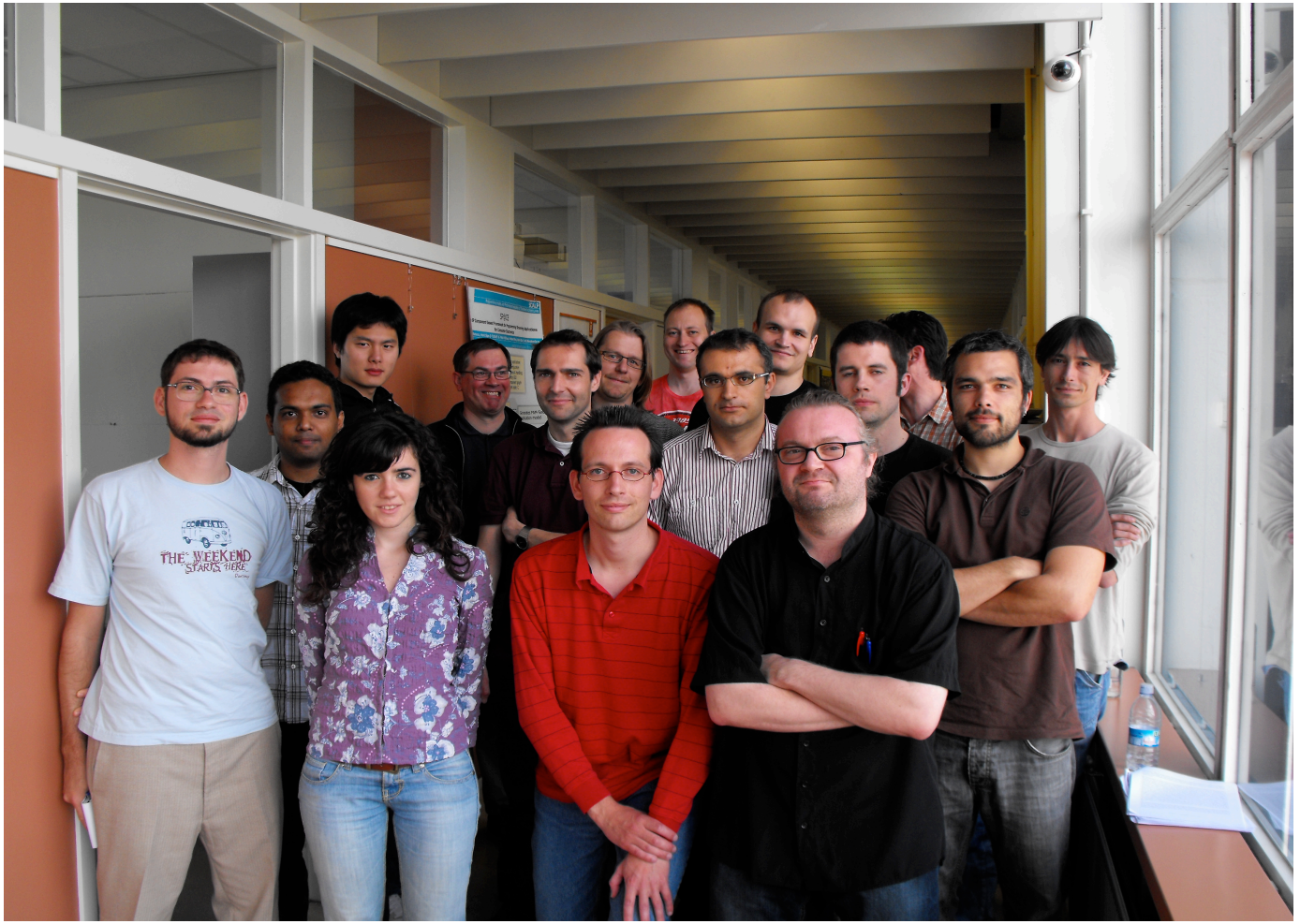
Some on-going research in the tribler team at TU Delft

name: David Hales
date: Sept. 21, 2009
event: ECCS 2009

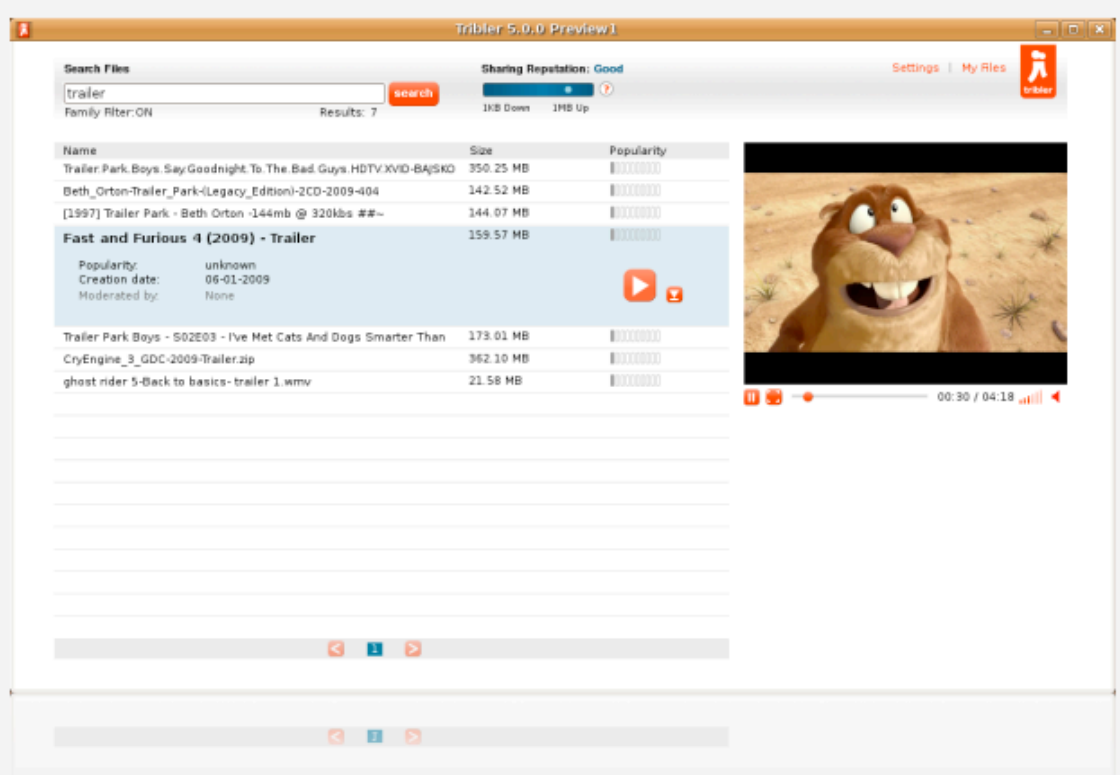
TU Delft

Tribler team: 35 + collaborators

tribler.org



Tribler BitTorrent Client



Tribler 5.1

We aim to show just how easy file-sharing can be. Due to our simplified search box you can find your files without the need for a website.

- Most easy to use
- Integrated search box
- Video-on-demand support
- Fully distributed

Download page

[Windows | Mac]

Talk Overview

- What is BitTorrent and specifically and Peer-to-Peer (P2P) systems in general?
- Axelrod and tit-for-tat experiment
- From swarms to communities: BitCrunch

What is BitTorrent

- BitTorrent is a Peer-to-Peer (P2P) file sharing protocol
- It lets users connect together to share any data they want to
- It self-organises and scales allowing millions of users to share files easily
- Current powerful economic actors don’t like it due to piracy
- But what is a P2P system in general?

What are Peer-to-Peer Systems?

P2P systems *are* socio-economic systems

- Peers cooperate collectively to achieve their goals
- No peer in the system controls everything
- Performance results from interactions
- At the end-of-day users (people) are in control
- P2P designers confront socio-economic issues

Individualism v. Collectivism

In socio-economic systems individual interests may conflict with collective interests:

- e.g. over exploitation of a common resource (a river, a field, the atmosphere etc.)
- e.g. banks - lending (to those who they know can not repay) to gain a commission by selling on the debt to other banks
- e.g. P2P file sharing system - downloading more than uploading

Individualism v. Collectivism

Consider a P2P file sharing system:

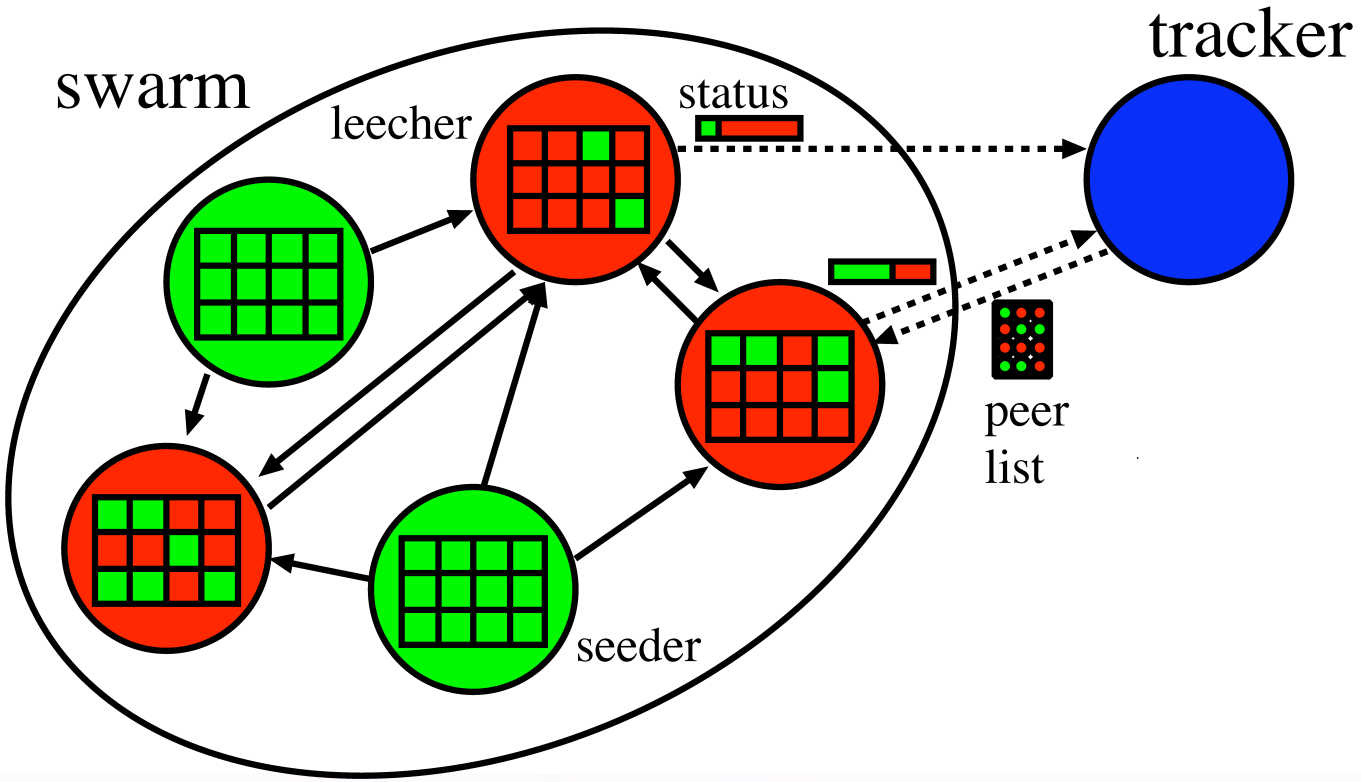
- It is in the *collective interest* for all to upload to others so everyone gets the file quickly
- But it is in the *individual interest* to save bandwidth by only downloading and hence free-riding on others
- Free-riding (or free-loading) is a perennial problem in P2P file-sharing systems
- Any efficient system needs to tackle it in some way

Some BitTorrent Terminology

- Swarm: set of peers interested in a file
  - file is split in smaller chunks called pieces
  - seeder: holds a full copy of the data
  - leecher: holds only a part of the data (initially nothing)
- Tracker: centralized manager
  - keep track of all peers in the swarm
  - return list of current peers in swarm
- Torrent file: meta-data
  - contains pointer to tracker hosting the swarm
  - details about the file - hash, no. of pieces, size etc.

BitTorrent Protocol

- Get a list of other peers in the swarm from the tracker
- Ask peers their list of pieces and tell them what is yours
- Exchange pieces with appropriate peers



How to avoid the commons tragedy?

Central enforcement of correct behaviour

- require centralised agencies and policing
- ability to identify and track individuals centrally
- not suitable for pure P2P (but see: private trackers)

Decentralised methods

- self-policing producing incentives for cooperation
- do not require centralised coordination
- more suitable for pure P2P
- can apply ideas from early complexity science results (Robert Axelrod: Iterated Prisoner's Dilemma Tournament)

Robert Axelrod's Tournament - programs as strategies

Axelrod organised an open IPD tournament:

- Academics were asked to submit programs (BASIC or FORTRAN) that would play the IPD against each other
- Nobody knew competitors code
- The only input would be the on-going past history of the game (a string of C's and D's)
- The aim was to get the highest score (utility) based on round-robin playoffs between all pairs of programs
- Axelrod's aim was to see which programs did best against all the others and understand why
- He wrote-up his results in the famous book "the evolution of cooperation"

Axlerod's Tournament - what happened?

Basic results were:

- many strategies were submitted (complex and simple)
- the one with the highest overall score turned out to be simple: *tit-for-tat* (TFT) or "look back"
- starts playing C, then "looked back" at the last move made by opponent and copied that move
- submitted by Psychologist Anatol Rapoport
- didn't "win" against each strategy but did better overall on average against all strategies
- TFT mechanism an example of "*reciprocal altruism*" (Robert Trivers)

What has this got to do with BitTorrent?

In the *BitTorrent protocol*:

- TFT-like method used for sharing files
- nodes form groups interested in a particular file (swarms) and swap or “barter” pieces with each other
- if a node does not upload data then this can be compared to playing defection
- it is punished in the future by being “choked” - not getting upload from others
- even if you hack your client to be selfish the chances are the standard TFT-like protocol will do better overall
- *Bram Cohen* - original BT designer - inspired by Axelrod’s tournaments

The Global Ecology of BitTorrent Clients

Many *bittorrent clients* exist in "the wild"

- Bittorrent 6 (from Bittorrent.com, formally utorrent)
- Others: Azureus, ABC, Transmission, many others...
- Tribler (of course)
- bad guy clients: BitThief, BitTyrant

Hence:

- The current bittorrent ecosystem is a ***global on-going experiment***, like Axelrod's, but with huge user base and rich interactions (not just TFT) incredible strategy sophistication
- This is unprecedented and could lead to new economic theory - in general!

Take home message

- Previous complexity work (Axelrod's IPD) has provided a basis for protocol design in a P2P system
- Deployed variants of the protocol are creating a massive global economic experiment
- Measurements can be made and these could inform new theory and new protocols

From swarms to collectives

Communities have formed around BitTorrent Trackers

Quite a few of these:

ONE BIG TORRENT .ORG

DutchReverge

jamendo
open your ears

bt.etree.org

Demonoid.com

ISOHUNT

TVTORRENTS.COM

The Pirate Bay

BITSOUP

Linux tracker

muslim torrents

TorrentPortal

indietorrents

Public Trackers (e.g. PirateBay)

- BitTorrent uses Trackers to index swarms
- Public trackers let anyone join or create a swarm
- Sharing within a swarm is incentivised via a form of tit-for-tat (as we have seen)
- However there is no incentive for:
  - Seeding (uploading after file is downloaded)
  - Capping (creating and injecting a new file)
  - Maintaining a Tracker in the first instance

Private Trackers (Many)

- Private Trackers have emerged more recently
- Only allow registered users to join swarms
- May track upload / download of each user
- Some keep centralised accounts for each user
  - When users download much more than upload they may be kicked out
  - Many different schemes: ratio, credits, points etc
- Some rely on users to just be nice with various “gentleman's club” methods

A little detail on credit systems

- We will give a little detail on credit systems in private BT communities
- Give a flavour of how economic / collective issues are becoming significant
- Present results from a simple (agent-based) model and some measurements of a real private tracker

Private Trackers - Credit

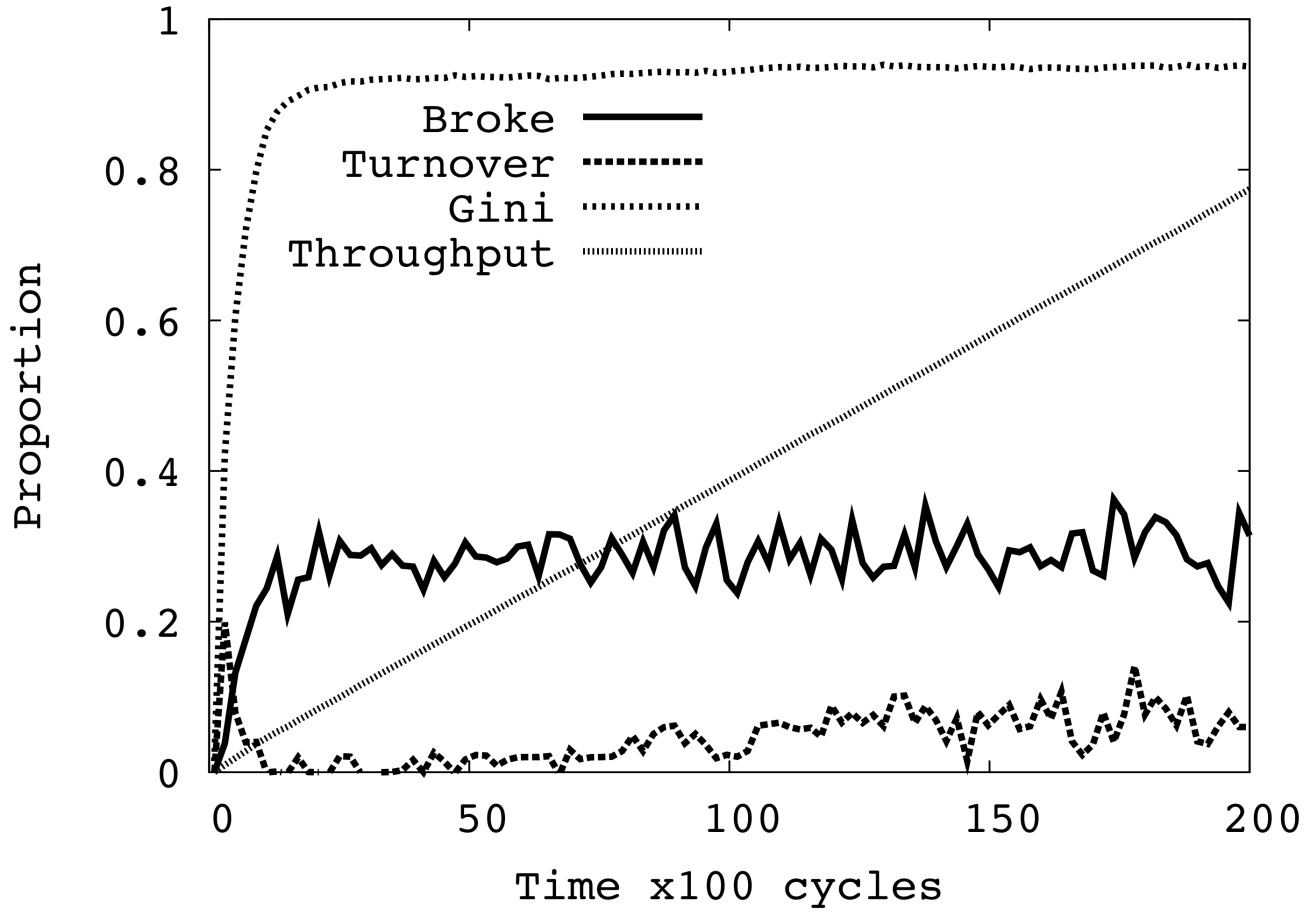
- Consider a scheme based on credits
  - Uploading 1MB earns one credit
  - Downloading 1MB costs one credit
  - A user with no credits can’t download
- Users must be given some initial credit
- In fixed size pop. total credit remains constant
- Similar to a fixed supply of money in an economy (loose analogy!)

Private Trackers - Credit

- How much credit should be put into the system?
- How would it effect the efficiency of the system?
- When do credit squeezes occur?
- How can they be avoided?

We define a credit squeeze as a situation in which, due to lack of credit, the efficiency of the system is significantly reduced.

Time series of simple model



Some aggregate results

C	T	β	G	φ
1	0.56	0.39	0.90	0.82
10	0.71	0.32	0.93	0.44
100	0.77	0.29	0.94	0.06
100++	0.97	0.01	0.71	0.00

C = initial credit
T = total throughput = total number of units uploaded as proportion of maximum possible (infinite credit)
B = proportion of nodes that are “broke” (zero credit)
G = Gini measure (simple measure of inequality of credit)
Phi = turnover of top 10% of peers ranked by credit (credit mobility)
100++ indicates initial credit of 100 with 1.5 credit seeding bonus

Observations from simulations

- Even when all peers are equal and are good guys performance of the system may be inhibited by credit shortages
- Adding extra capacity to the system, in the form of upload and download, can *actually reduce* the performance
- By injecting new credit into the system in the form of a “seeding bonus” a credit squeeze can be ameliorated

Statistics from a Private Tracker

Day	T	Δ	Δ_0	δ	S/L
1	48	24	17	0.23	26
2	40	20	15	0.25	26
3	50	25	12	0.16	25
4	67	33.5	17	0.17	25
5	52	26	19	0.24	25
6	46	23	15	0.21	25
7	87	43.5	17	0.13	25
Ave.	56	28	16	0.19	25

Approx. 50,000 peers per day, 10,000 swarms,
access to credit balances of top 10%

T = throughput in TB over all swarms
Delta = total credit increase that day in the entire system
Delta0 = total credit increase for top 10% of peers
Delta = minimum fraction of credit increase that goes to top 10% of peers
S/L = seeder to leecher ratio over all swarms

Statistics from a Private Tracker

- Indicates “rich getting richer” since top 10% are getting a lot of the new credit
- High Seeder / Leecher ratio suggestive that a credit squeeze is happening for many
- But need more information to verify this
- Would be interesting to see what happened to throughput if there was a “free day” or seeding bonus was increased

Take home message

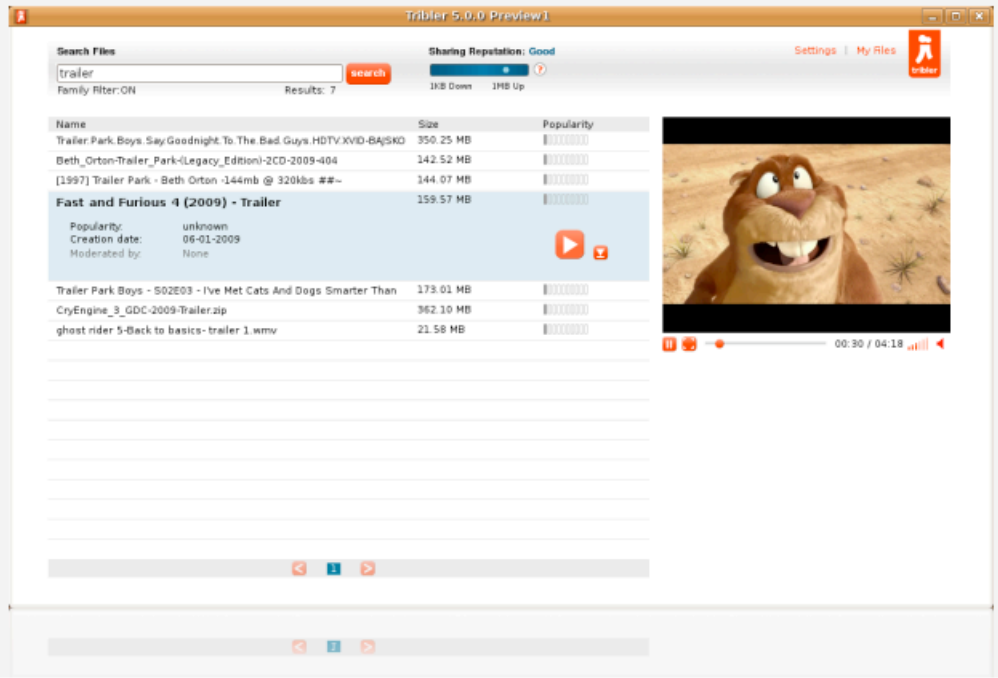
- Communities formed around trackers provide an on-going global socio-economic experiment
- Self-organisation of socio-economic structures in measurable forms
- Ideas, models and theories from complexity science may inform and learn from this
  - Cultural group selection, Indirect reciprocity, Altruistic Punishment etc.
- Such communities so strong don't be surprised if they start influencing the "real world" (e.g. the PirateParty)

On-going work in Tribler team

- Community measurements and studies, including qualitative - Nazareno Andrade et al
- Alternative economic incentive models, incl. Participatory Econ.- Rameez Rahman et al
- Distributed credit systems, including currency-type approaches – Michel Meulpolder et al
- Self-organising locality for increased performance - Maciej Wojciechowski et al
- Many many others…

Advert / plug / shameless promotion!

Download tribler 5.1 at: www.tribler.org



Tribler 5.1

We aim to show just how easy file-sharing can be. Due to our simplified search box you can find your files without the need for a website.

- Most easy to use
- Integrated search box
- Video-on-demand support
- Fully distributed

Download page

[Windows | Mac]

Comments / discussion and suggestions on:

forum.tribler.org

References

Nash, John (1950) Equilibrium points in n-person games. Proceedings of the National Academy of Sciences 36(1):48-49.

John von Neumann and Oskar Morgenstern: Theory of Games and Economic Behavior, Princeton University Press (1944)

Robert Axelrod (1984) The Evolution of Cooperation, Basic Books

Nowak, M.A. and Sigmund, K. (1998) Evolution of indirect reciprocity by image scoring, Nature 393, 573.

Garrett Hardin (1968) The Tragedy of the Commons Science 162, 1243-1248.

Trivers, R.L. (1971). The evolution of reciprocal altruism. Quarterly Review of Biology. 46: 35-57

Maynard Smith, J. (1982) Evolution and the Theory of Games. Cambridge University Press