Vol. 41 (2010) *ACTA PHYSICA POLONICA B* No 5

COOPERATION IN PEER-TO-PEER NETWORKS*

Tomasz Gradowski[a], Maciej J. Mrowinski[a]
Robert A. Kosiński[a,b]

[a]Faculty of Physics, Warsaw University of Technology
Koszykowa 75, 00-662 Warszawa, Poland
[b]Central Institute of Labor Protection — National Research Institute
Czerniakowska 16, 00-701 Warszawa, Poland

(Received February 17, 2010)

This paper presents results of a research conducted on a simple model of a peer-to-peer network (a network in which users exchange files directly, without any central server involved). The conditions necessary for the file exchange process to be efficient and stable are investigated through numerical simulations and analytical calculations based on the master equation. Ways of preventing free-riding (selfish behavior, when users download files without sharing them) are also discussed.

PACS numbers: 87.23.Ge, 89.20.Hh

1. Introduction

The Internet has become one of the most important means of communication of our century. Every second it is being used worldwide to deliver an unimaginable amount of content. A substantial part of this traffic can be attributed to the so-called peer-to-peer networks, for example BitTorrent, in which users can exchange data directly, without the assistance of any intermediate server (there may be some servers involved in establishing the connection — BitTorrent is a hybrid system [1]). While used mostly for illegal file sharing, peer-to-peer networks are an effective way of distributing files to a large group of users at virtually no cost for the publisher. Unlike centralized server-based solutions, peer-to-peer networks are as reliable (in terms of a file availability) as their users and therefore cooperation between the file sharers plays a crucial role in spreading files over the network. In peer-to-peer networks files are divided into segments. Users who want to

---

* Presented at the XXII Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland, September 12–17, 2009.

download a specific file connect to a server (a tracker). The server provides them with a list of other users who already own at least one segment of the requested file. The list allows users to connect with each other directly and exchange available segments. Users who share an entire file (the ones who are in possession of all the segments) are called seeders. Other users are called peers.

One of the biggest issues in peer-to-peer networks is the so-called free-riding — many users stop sharing files after acquiring all the segments. Such a behavior has a significant impact on the overall dynamics of the network. Users compete for available network resources and selfish behavior causes a loss of efficiency. Leech (or leacher), in other fields referred to as a free-rider, is a person who downloads far more amount of data than he uploads. Free-riding behavior problem (originally describing the practice of using public transportation without paying the fare) is well described in economics, psychology and political science. Free-riding and punishing strategies on evolutionary games was extensively investigated by many authors (for review see [2]). Hence peer-to-peer networks can be considered as shared resources system, where users play the role of resource contributors. In such systems so-called altruistic punishment is essential to gain a high level of cooperation.

Free-riding was recognized as a major problem in the very early days of peer-to-peer networks. Studies of real networks indicate that most users tend to behave selfishly and do not share files. According to [3] "almost 70% of Gnutella users share no files, and nearly 50% of all responses are returned by the top 1% of sharing hosts". Various mechanisms that provide an incentive to share have been proposed ([4–6]) and analyzed using the game theory framework. Experimental studies on socio-psychological motivating methods have also been conducted [7].

We use simple models to study the cooperative and dynamical phenomena in peer-to-peer networks. Our goal is to look for critical parameters that determine the dynamics of such systems and discuss ways of preventing the free-riding behavior.

2. Numerical model

To simulate numerically the process of file exchange in peer-to-peer networks we introduce an agent-based model. The agents in the model are irrational and behave according to probabilistic rules. The dynamics step is as follows: each user tries to download each file with probability α. Then the user deletes his files with probability β. After that, the file exchanging process starts. Each agent sends to other users a requests for files which he wants to download. Then the queue of users, who have one of the requested files, is created and randomly permuted. Each agent, a peer-to-peer network user, has constrains in download/upload capacity for one step of the dynamics, so the number of concurrent connections per user is limited. The requests are executed in turn, bit by bit, until the download limit is exceeded or the queue vanishes. Senders are removed from the queue if they exceed their upload limit. If the sender has only some part of the requested file, he seeds it with a probability proportional to the size of the completed part. New downloads are randomly selected from a pool of files, number of files is constant, and their sizes are uniformly distributed.

3. Attempts of analytical description

In order to understand the dynamics of peer-to-peer networks we can create a mathematical model of such a system. While it is not possible to recreate the complexity of real networks, we can try to isolate the key components that determine their behavior. Let p be the probability that a user in the peer-to-peer network is downloading a specific file (we do not make a distinction between sharing and downloading files; we also assume that the download process is instantaneous). During each time step users already sharing the file delete it with probability β and user who do not share the file try to download it with probability α. The time evolution of probability p, consistent with these two simple actions, can be described by the following master equation:

$$\frac{dp}{dt} = (1-p)\alpha - p\beta \,. \tag{1}$$

This equation has a single stable fixed point

$$p^* = \frac{1}{1+\xi} \,, \tag{2}$$

where $\xi = \beta/\alpha$. If neither files nor users in the system are distinguishable, then the probability that a file is being downloaded by a user is also equal p^*. Using this fact and equation (2) we can derive the probability that the file is being downloaded by at least one user:

$$p_{\mathrm{av}} = 1 - (1-p^*)^N = 1 - \left(\frac{1}{1+\frac{1}{\xi}}\right)^N \,, \tag{3}$$

where N is the total number of users in the system. Both (2) and (3) are in agreement with the results of our simulations. Probability p_{av} can be used as a measure of file availability and, when treated as such, it gives us information about efficiency of the network and cooperation between its users. As we can see in Fig. 1, p_{av} nearly equals 1 for small ξ and starts to decrease rapidly when ξ reaches certain value ξ_{o}. Networks with ξ lower than ξ_{o} are optimal in the sense that all files are always available. We can approximate ξ_{o} by checking for which ξ the tangent to p_{av} at the inflection point equals 1:

$$\xi_{\mathrm{o}} = \frac{1}{4}\frac{(N-1)^2}{N} \approx \frac{1}{4}N\,. \tag{4}$$

Equation (4) specifies the number of files (denoted by f) a user must download at each time step for the network to remain optimal. If the total number of files in the network is equal F then (from (2) and (4))

$$f \approx \frac{4F\beta}{N} \tag{5}$$

and the total number of files in the user's download queue (shared files) is $\frac{4F}{N}$.

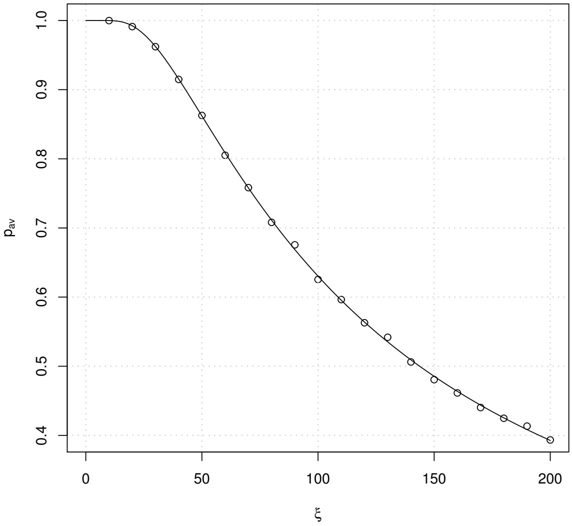

Fig. 1. Probability p_{av} that file is being downloaded by at least one user (for 100 users) compared with analytical results (3).

4. Results

In this section we present the results of numerical simulations for the model introduced in Section 2 and compare them with the results obtained analytically in Section 3. In the simulations performed for different values of the control parameters we examine how the β/α ratio affects the efficiency of the network. Following parameters were used for simulations: number of users $N = 1000$, number of files $F = 100$, upload/download bandwidth 16/64 units, file size 100–150 units, number of dynamic steps $t = 1000$.

In order to determine the efficiency and stability of the peer-to-peer network, we calculate the number of files available in the network (files seeded by at least one user). This number varies from the total number of files in the pool (when every file from the pool is owned by at least one user) to zero. The number of files present in the system as a function of the delete/download probabilities ratio ξ for a few values of α is shown in Fig. 2. We can distinguish two phases separated by a critical value of ξ. In the first phase the cooperation between users occurs and almost every file is present in the system. In the second phase there are no files in the system.

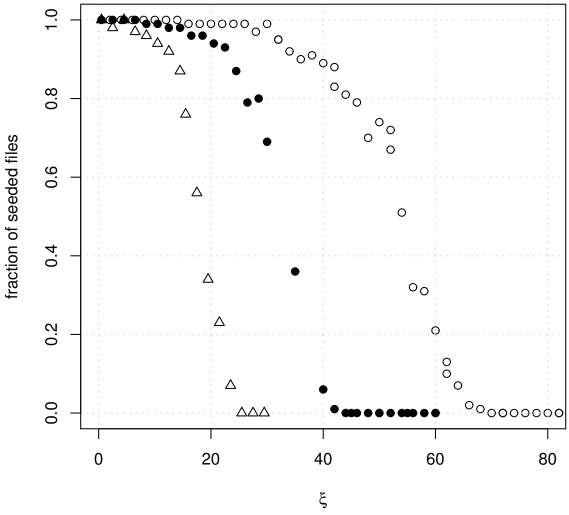

Fig. 2. Fraction of seeded files as a function of ξ for different values of α: 0.005 (circles), 0.01 (filled circles) and 0.02 (triangles).

Let n_i denote the number of users who are downloading or already downloaded file i. Then $p_i = n_i/N$, where N is the number of users, is the probability that file i is downloaded. Thus the average probability is $p = \frac{1}{F}\sum_i^F p_i$, where F is the number of files. This probability, shown in Fig. 3 for different values of the control parameters, is in agreement with the theoretical results (2).

We have calculated the total number of files seeded or downloaded in the system for different $\xi = \beta/\alpha$ ratios. It is shown in Fig. 1 with theoretical curve obtained in (3).

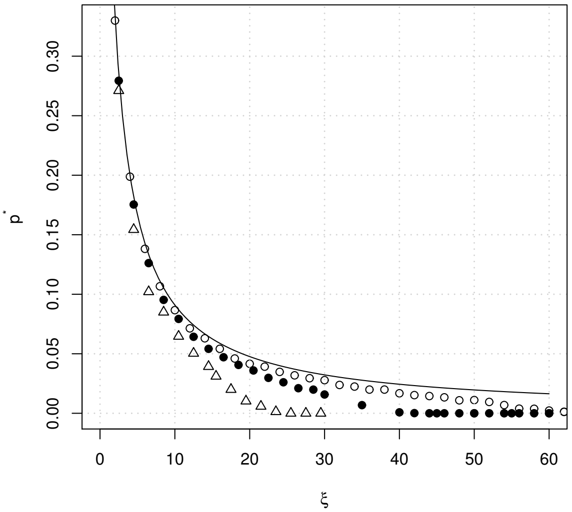

Fig. 3. Probability p that file is being downloaded averaged over all files. Solid curve is for analytical result obtained in (2). Numerical results for $\alpha = 0.005$ (circles), 0.01 (filled circles) and 0.02 (triangles).

5. Conclusions

In our work we have studied a simple model of a peer-to-peer network, both analytically and numerically. Our results show how users' behavior determines the efficiency and stability of the file exchange process. We have calculated parameters for which the network works in an optimal way — namely, the number of files users must share and download during each time step. However, the model itself does not take into account any incentive for the users to actually share enough files.

In real peer-to-peer networks there is no 100% working method of preventing free-riding. The most important one is spreading the etiquette which demands that every user should keep his share ratio equal at least 1.0. That is, users should upload at least the same amount of resources as they download. Of course, the peer-to-peer etiquette, like any other etiquette is not obligatory. It would be possible to implement such a requirement directly in peer-to-peer clients (computer programs used to exchange files in peer-to-peer networks), but effectiveness of such a solution would depend on popularity of this client. Peer-to-peer clients and servers implement various incentives to share files. Clients often impose download limits on users who limit their upload bandwidth. Some trackers, in order to achieve the desired share ratio, require registration and track users activity. This method is the most effective one but not widely used.

This work was supported by the European Cooperation in the field of Scientific and Technical Research — COST Action MP0801 "Physics of competition and conflicts".

REFERENCES

[1] F. Benayoune, L. Lancier, *Lect. Notes Comput. Sci.* **3262**, 327 (2004).

[2] K. Sigmund, *TRENDS in Ecology and Evolution* **22**, 593 (2007).

[3] E. Adar, B.A. Huberman, *First Monday* **5**(10) (2000).

[4] N. Andrade, M. Mowbray, W. Cirne, F. Brasileiro, CCGRID 04: Proceedings of the 2004 IEEE International Symposium on Cluster Computing and the Grid, IEEE Computer Society, p. 440, 2004.

[5] C. Buragohain, D. Agrawal, S. Suri, Proceedings of the 3rd International Conference on Peer-to-Peer Computing, 2003.

[6] D. Qiu, R. Srikant, *ACM SIGCOMM*, Portland, OR, USA, 2004.

[7] H. Bretzke, J. Vassileva, *Lect. Notes Comput. Sci.* **2702**, 148 (2003).