Social Networks Swarms in P2P File Sharing

Project Progress Report

Instructor: Prof. Stefan Saroiu

Group Members: Wael Aboelsaadat, wael@cs.toronto.edu, 991118170

                        Sadek Ali, sali@cs.toronto.edu, 994388989

Date: November 1, 2007

1   INTRODUCTION

On-line social networks (OSN) have been combined with peer-to-peer (P2P) file sharing systems to improve network performance, search, and user collaboration. A key research area that has not been widely explored is the critical role that social network incentives play in how content can be swarmed to pre-seed a network to improve content availability. P2P networks are largely designed for distributing popular content to a wide audience. Personal content that is shared within a semi-private social clique does not garner enough peers to make it available in protocols like BitTorrent because there may not be an available seeder to serve the content. This results in not only the file transfer being not much better than simple FTP or web download, but the file is potentially less available than posted web content because the seeder must be logged in to the P2P network in order for other users to find the content. An obvious solution to this problem is to create multiple seeders for the content. In P2P systems, connection speed, physical locality, bandwidth and the amount of uploaded/downloaded content are used as metrics for defining the preference relationships for file sharing between nodes in the network. Using these typical P2P incentives, such as bytes downloaded and uploaded, there is an inherent disincentive for peers to do a *favor* for another user, because the user who is downloading is consuming P2P resources for content that will be seldom requested, and the original publisher is rewarded. In addition, if we consider the heterogeneity of the hardware and disk space available across P2P peers, then there may be the case where a user who is socially connected may post desirable content, but the user is not able to participate in file sharing because they cannot guarantee consistent availability of the content. In this paper, we propose an extension to the BitTorrent protocol to include social network incentives based on an altruistic file-sharing algorithm where the benefit of multiple seeders for availability of content for smaller groups can be incorporated into P2P incentives for preferential benefits in peer selection for the altruistic peer(s). We consider the BitTorrent protocol (BTP/1.0), and show how it can be augmented to allow a peer with new group-oriented content to garner attention to their need for another seed,

and to negotiate a mechanism for altruistic peers that donate bandwidth, connections and disk space, until a new seed in the social network is created.

2 RELATED WORK

Incentive mechanisms are an important dynamic in how nodes interact in P2P networks. If there is no motivation to share information or resources, then the P2P network becomes subject to trends such as "free-riding" [18] or the "tragedy of the commons" [11] problem. Several incentive mechanisms have been proposed: auction based [15], service contribution [16], rank based incentives [10] and social network based incentives [8, 12, 4, 17]. The latter mechanism, social network incentives, is composed of social relationships, trust and reputation. These are strong facilitators to users sharing personal content with each other. Online social networks were initially for people who wanted to not just communicate with each other, but within the cultural milieu of a social group. One of the challenges today in OSN is the sharing of large content within a small group. This leads to the problem that the download of the content by group members is not necessarily fast, but also because of the small number of peer who want to share it, the content may often not be available for download when the P2P protocol requires that a seeder be always available to guarantee that the content can be downloaded at any time. The goal of our research is to explore how to best integrate the file-sharing efficiency of P2P with the compelling user experience of social networks. In the state-of-the-art in social-based P2P, the approach for content sharing has revolved around permissions, where resource access is given preferentially to members of the closest social network. Several existing commercial products [1, 9, 13] offer P2P using social networks. But, these commercial applications have largely been restricted to the realm of access controls for the purpose of restricting membership to a file-sharing network. Another approach to sharing content through the social network is seen in the BitTorrent-based system Azureus [3]. It does not directly integrate social relationships. Instead, it provides a feature to automatically link .torrent files into personal feeds of recommendations in OSN services such as FaceBook, Digg It! and del.ico.us [6, 7, 5]. The role of incentives such as social relationships, content ratings, identity ratings, network population size, social participation, real-world proximity and other social network-based factors help users to establish a following of social peers who may download content based on the recommendations. An outstanding key issue is that the recommendations in Azureus are typically made for commercial content, as opposed to personal content. For popular content, there will likely be multiple full copies of the content that exists on the P2P network. In contrast, we differentiate our approach to this problem by targeting the availability of content that is not popular, but its relevance to the social group may nonetheless be very high (for instance, sharing a wedding video to geographically distant relatives and parents). The social network provides a superb medium for communication, and by being able to improve the availability of less popular content, we believe that users of OSN's will more likely take the time to distribute large, media-rich content for sharing via P2P networks, while using the OSN to communicate to their social peers where in the P2P to access the content.

3   PROPOSED SOLUTION

We propose the development of an extension to the BTP/1.0 protocol [19] which we call Social Network Swarming. The goal of our extension is to provide an effective means to seed a given social network with one or more copies of any content that has been specially marked as group shareable content. The main reason for making multiple seeders is to help ensure the availability of the content.

The general concept for the solution is an altruistic tracker (alt-tracker) whose job is to track the seeding of content in social networks. It accepts requests from anonymous users to donate bandwidth to transfer pieces of anonymous content until a full copy has been completed between two socially-connect nodes.

The general flow of the protocol change involves three different peers, the publisher, the agents, and the new seeder. The publisher is the node that introduces new personal content to the P2P through a new torrent and requests for others to donate bandwidth to make one or more seeds in the network. The publisher shares the torrent with one or more candidate seeders through the social network. The candidate seeder joins the P2P network, contacts the alt-tracker, and begins downloading the content as per a social compact with the publisher. In order to make the process more reliable, any other peer in the P2P network can also visit the alt-tracker in order to donate bandwidth and hard drive space to download all of the pieces of the file from the publisher. Once enough candidate seeders have completed their downloads, so as to become seeders of the content, then the entire process is halted, all connections are freed, and all outstanding requests for data by donors and candidates are discarded.

We can summarize the main process flow as:

> **Publisher:**
> Publish a Torrent to New Seeder
>
> **Publisher:**
> ANNOUNCE the Swarm to "Altruistic" Tracker (Alt-Tracker)
>
> **Candidate Seeder (CS):**
> DONATING to Alt-Tracker with info_hash shared through OSN.
> Candidate Seeder is included in alt-tracker count in *numwanted*.
> Alt-Tracker responds with Swarm for CS to download content.
>
> **Anonymous Donor:**
> DONATING to Alt-Tracker with no info_hash
> Peer C donates bandwidth to download content

The publisher will use the following message to communicate the announcement to the alt-tracker of a request for a Donation to use a Social Network Swarm to transfer one or more copies of some new content (note, the dictionary is extended with a new value to state the size of the publisher's content):

ANNOUNCE (Request for Donation)	
info_hash	Required
peer_id	Required
port	Required
uploaded	Required
downloaded	Required
left	Required
ip	Required
numwant	Required – Number seeders requested
event	Null
size	Required - Bytes in content

The candidate seeder and anonymous donor use the following message to communicate with the alt-tracker to donate bandwidth and hard drive space (note, the dictionary is extended with a new value to state the size of the content that can be accepted):

DONATING (Request Intention to Donate)	
info_hash	(null | info_hash)
peer_id	Required
port	Required
uploaded	Required
downloaded	Required
left	Required
ip	Null
numwant	Null
event	'donate'
size	Required - Bytes available for storing a piece

If the info_hash is null, then the alt-tracker assumes that the requester is an anonymous donor. The response includes a list of peers and the info_hash for the handshake with the seeder.

If the info_hash is not null and is sent in a request to the alt-server, then alt-tracker assumes that the requester is known by the publisher through an OSN, and that the requester is a candidate seeder.

The alt-tracker responds to all requesters with its normal response message to an actual REQUEST, but modified by adding the info_hash to the publishers recently announced content to the response.

DONATING (Response to Intention to Donate)		
failure_reason	(Null | Full error description)	
interval	Required	
complete	Required	
incomplete	Required	
peers	Required	
peer_id	Required	
ip	Required	
port	Required	
info_hash	Required	- Info hash of seeder

On the alt-tracker, when either the number of seeders equals *numwant* + 1 or the number of donors for a torrent becomes zero, then alt-tracker must drop the Request for Donation for the given Torrent.

We stipulate that the candidate seeders will not accept Handshakes for upload requests from anonymous Donors. This solution allows the candidate seeder to participate in this activity, given that it made a social compact with the publisher based on the transference of a .torrent file. Finally, we also leave the piece selection strategy as an open implementation issue for specific P2P clients.

4   ACCOMPLISHMENTS

We have been able to focus our research into the fairly concise problem of how to garner a swarm for content that is unlikely to be very popular.

We have conducted a thorough analysis of the BTP/1.0 protocol [19] to see how we could incorporate OSN incentives into P2P downloads.

We have concluded that an overlay of the OSN onto the P2P is not required for file sharing, although it does have application in information retrieval, specifically in querying and distributing indexes across the P2P network.

Finally, we have been able to study a functioning codebase of the Azureus Java-based BitTorrent client. We have learned that the complexity of modifying the client may take longer than we had originally anticipated.

5 PROBLEMS AND APPROACHES

5.1 Is availability really a problem?

We know for certain that most users will not share large media through OSN's on the web. But, the basic premise of our project is that in this situation P2P is a good alternative but there is a problem with accessing unpopular content due to unavailability. These are generally accepted notions, but their significance and validity remain to be proven.

Our solution is to ask a small number of P2P users who are also users of on-line social networks for how they share videos in their OSN, and whether they have considered using P2P's. In particular, we are interested in how they consider the publishing of their personal content, and most importantly, both how long the content might be available for download and whether they are constantly logged in to the P2P network.

5.2 Role of Social Incentives

We have been able to establish through the existing literature that social incentives are, in general, considered an important consideration in content sharing. But, the degree to which different aspects of social incentives are best-suited for encouraging the sharing of content is not known. Our problem is that we have not established how the social incentives affect user's ratings in the P2P and the OSN.

Our solution is to first establish a simple measure that connects the use of private resources to the public good of the social network as meritorious to the traditional P2P ratings based on downloads and uploads. Further to this, time permitting, we hope to expand our exploration of the various social incentive systems that we have encountered in the literature to show how these can be shared between the OSN and P2P, and used in Peer Selection and Choking Strategies.

5.3 Motivating Example

This research makes a very strong underlying assumption that users in an OSN will accept this new protocol. To these ends, a reasonable motivating example to highlight problems and alternative methods would help to clarify the issue.

5.4 Modeling of Proposed Protocol Extension

A key problem that we will face shortly will be the modeling of the proposed new protocol. We will probably not have time to run any significant simulation.

Our approach to this problem will be to conduct an analysis to ensure that the protocol extension maintains a sound set of states. In addition, we will be able to estimate the consumption of local resources on the different Peers in our system based on the behaviour of Azureus. We will estimate the load that our extension introduces on the P2P network based on content sizes, and swarm size.

5.5 Security and Privacy Issues

We have not explored the implications of this type of system, in terms of either security or privacy. Our system's main weakness is the anonymity of users. We will make a review of the system, and propose solutions to security problems.

5.6 Tracker Software

Our proposal calls for a modified tracker. Up to this point, we have been unable to locate a working codebase for a reasonable BitTorrent tracker. Our desire is to encode the new protocol across a compatible pair of Trackers and Peers to at least show that it is feasible.

We are hopeful that we will find one, but in the event that we do not we will need to find an accommodation for this issue in the paper, or we may rework the protocol to run solely on the Peers.

5.7 Measurement of Availability

We have yet to determine how to measure the availability of content in the P2P.

This is an open issue at this point. We hope to resolve the issue by reviewing the literature. A naive approach would be to use a minimal probability that a given user makes their content available at any given time. By increasing the number of users, and varying the minimal probability, some characterization of availability could be made.

6 PROPOSED SCHEDULE

Week of Nov 8

Complete a draft of protocol

Develop a motivating example

Review the security and privacy issues of approach

Conduct user study for how user's share video and large files in OSN

Week of Nov 15

Find a codebase for a compatible Tracker and Client, preferably not in Python.

Code new protocol

Analyse the new protocol for estimated resource use and bandwidth

Week of Nov 22

Run preliminary tests to observe Tracker and Client(s)

Explore possible introduction of shared OSN/P2P social incentives

Write up for final paper

Nov 29

Project Completed

BIBLIOGRAPHY

[1]   AllPeers. http://www.allpeers.com/, 2007.

[2]   Stephanos Androutsellis-Theotokis and Diomidis Spinellis. A survey of
        peer-to-peer content distribution technologies. ACM Comput. Surv.,
        36(4):335–371, 2004.

[3]   Azureus Inc. http://www.azureus.com/, 2007.

[4]   N. Borch. Social peer-to-peer for social people. In The Int'l Conf. on
        Internet Technologies and Applications, Sep 2005.

[5]   del.ico.us. http://del.ico.us/, 2007.

[6]   DIGG. http://www.digg.com/, 2007.

[7]   FaceBook. http://www.facebook.com/, 2007.

[8]   Andrew Fast, David Jensen, and Brian Neil Levine. Creating social
        networks to improve peer-to-peer networking. In KDD '05: Proceeding of
        the eleventh ACM SIGKDD international conference on Knowledge
        discovery in data mining, pages 568–573, New York, NY, USA, 2005.

[9]   FoxTorrent. http://www.foxtorrent.com/, 2007.

[10]   A. Habib and J. Chuang. Service differentiated peer selection: an incentive
        mechanism for peer-to-peer media streaming. Multimedia, IEEE
        Transactions on, 8(3):610–621, 2006.

[11]   G. Hardin. The tragedy of the commons. 162(3859):1243–1248, 1968.

[12]   Chen Hua, Yang Mao, Han Jinqiang, Deng Haiqing, and Li Xiaoming.
        Maze: A social peer-to-peer network. In CEC-EAST '04: Proceedings of
        the E-Commerce Technology for Dynamic E-Business, IEEE International
        Conference on (CEC-East'04), pages 290–293, Washington, DC, USA,
        2004. IEEE Computer Society.

[13]   IMEEM. http://www.imeem.com/, 2007.

[14]   Robert J¨aschke, Leandro Balby Marinho, Andreas Hotho, Lars Schmidt-
        Thieme, and Gerd Stumme. Tag recommendations in folksonomies. In
        Joost N. Kok, Jacek Koronacki, Ramon L´opez de M´antaras, Stan
        Matwin, Dunja Mladenic, and Andrzej Skowron, editors, Knowledge
        Discovery in Databases: PKDD 2007, 11th European

Conference on Principles and Practice of Knowledge Discovery in Databases, Warsaw, Poland, September 17-21, 2007, Proceedings, volume 4702 of Lecture Notes in Computer Science, pages 506–514. Springer, 2007.

[15]   B. Yun L. HongTao, H. ZhiXing and Q. Yu Hui. Auction incentive mechanism in p2p. In 2007 International Conference on Multimedia and Ubiquitous Engineering , year = 2007, journal = mue, volume = 00, isbn =0-7695-2777-9, doi =http://doi.ieeecomputersociety.org/10.1109/MUE.2007.3, publisher = IEEE Computer Society, address = Los Alamitos, CA, USA.

[16]   Richard T. B. Ma, Sam C. M. Lee, John C. S. Lui, and David K. Y. Yau. An incentive mechanism for p2p networks. In ICDCS '04: Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS'04), pages 516–523, Washington, DC, USA, 2004. IEEE Computer Society.

[17]   J.A. Pouwelse, P. Garbacki, J. Wangand A. Bakker, J. Yang, A. Iosup, D. Epema, M.Reinders, M.R. van Steen, and H.J. Sips. Tribler: A social-based based peer to peer system. In 5th Int'l Workshop on Peer-to-Peer Systems (IPTPS), Feb 2006.

[18]   Sujay Sanghavi and Bruce Hajek. A new mechanism for the free-rider problem. In P2PECON '05: Proceeding of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems, pages 122–127, New York, NY, USA, 2005. ACM.

[19]   J. Fonseca, B. Reza, L. Fjeldsted. BitTorrent Protocol -- BTP/1.0. http://jonas.nitro.dk/bittorrent/bittorrent-rfc.html, April 2005.