Peer-to-Peer Systems: "A Shared Social Network"

Nguyen Hoang Anh
Helsinki University of Technology
hanguyen@cc.hut.fi

Abstract

In the last few years, the success of the Napster online music sharing program has drawn much attention to an emerging paradigm for communication on the Internet namely Peer-to-Peer (P2P) technology. In P2P, file sharing is probably the most popular application. But in fact, computers in P2P networks - or peer nodes - can share various types of data and resources, i.e. files, bandwidth, storage space and processor cycles. In this paper, I will present one promising application of P2P technologies in building social networks.

KEYWORDS: P2P, social networks, P2P security

1 Introduction

Peer-to-Peer is emerging as a prominent design pattern for future systems with regard to scalability and resilience due to its decentralized and self-organizing approach. P2P applications currently contribute to the main portion of traffic volume on Internet: more than 50 percent of all download traffic and more than 75 percent of all upload traffic is P2P related [2]. It seems that more people are using P2P today, in 2004, 1 CacheLogic-server registered 3 million IP addresses in 30 days and in 2006, it just needed 8 days to reach that number.

In its early days, P2P was mostly used for file sharing systems, but gradually, grid computing and VoIP have also been successfully implemented. P2P architecture is different from Internet's traditional Client-Server paradigm. The continuous growth of the Internet in the terms of users and bandwidth is accompanied by increasing requirements of rich feature Internet application, which are Scalability, Security and Quality of Service [8]. In this situation, P2P promises to provide prospective solutions through a fundamental change of paradigm. P2P networks forms a de-centralized and self-organizing overlay network on top of the Internet, on which, peers are interconnected and directly interact with each other to share resources, without central services.

Also in the last few years, social network services have appeared as an outgrowth focusing on the building of online social networks for communities of people who share interests and activities, or who are interested in exploring the interests and activities of others [1]. The trend of combining P2P and social network is now considered the next big thing in the Internet.

In this paper, I want to contribute to the discussion on the possible directions and the problems of using P2P to build social networks; an attention is dedicated to security concerns. In section 2, a short overview of the topic's background discussed here is given. The advantages/drawbacks of the P2P approach to build social networks are presented in section 3, how to use P2P technologies to develop social networks is discussed in section 4. In section 5, security concerns are discussed, followed by conclusions in section 6.

2 Background

2.1 Social Networks

Recently, Web 2.0 has started emerging as the next generation of the Internet. The term Web 2.0 started to be notable after the first O'Reilly Media Web 2.0 conference in 2004. There are different definitions for Web 2.0. Tim O'Reilly [7] outlined 8 themes that he thinks it is crucial when identifying Web 2.0. According to Tim O'Reilly, Web 2.0 is about Web as Platform, on which, developers create software applications by using Application Programming Interfaces, or APIs. In Web 2.0, Web is seen as a two-way communication where people are both readers and writers. Web 2.0 means many things, two of the essential pillars are user-generated content and social network sites like MySpace, Facebook, and Hi5, which have an increasing number of people joining. These systems allow participants to register for an account and when logged in, they are asked to create a profile to represent them online. These profiles are collected together into "friend list" of a participant. Participants can invite other as "friend", and if that person accepts the invitation, a photo of each is displayed on the profile of the other. This way, participants develop their friend lists. The process is similar to the way we make friend in the real world, but the difference is, if in the real world, we may not see the friends of our friends, in these systems, the relationship are visible: participants are able to view their friends' friend list. A participant can choose one profile to make a new friend, therefore, the collection of "friends" is not simply a list of close ties - what we normally call "friend" - but instead, this feature allows participants to be seen by the rest of the world. A participant can have hundreds of "virtual" friends even though he or she even does not know any of them as a real friend in the real world.

Profiles and friends lists are two key features on social network sites. The third is the commenting feature which allows users to comment on their friends' profiles. These comments are seen by anyone who is able to access to that profile.

These three features - profiles, friends lists, and comments - make up the primary structure of social network sites. Fur-

thermore, each social network can provide additional features to attract participants. Social network sites allow users to traverse from Friend to Friend and communicate with anyone who has a visible profile. The purpose is to allow users to create their small world online by building friends list of others who share some kinds of interests and ideas.

2.2 P2P technologies

A peer-to-peer (or P2P) computer network refers to any network that does not have fixed clients and servers. In a P2P network, every peer node will function as both clients and servers to the other nodes on the network [8]. This is different from client-server model, with millions of consumer clients communicating with a relatively privileged set of servers. Nowadays, a large number of Internet users start using their powerful home computer and broadband Internet connection for more than just browsing the web and sending/receiving emails. Instead, they joint P2P networks, connecting directly to each other to form groups and working together to become user-created search engines, virtual computers, and filesystems. There are several P2P technologies. These technologies differ in how they search for data contained in their peers.

2.2.1 P2P approaches

Some systems, such as Napster [6], use client-server approach for some tasks, and P2P approach for others - this approach is hybrid P2P approach. Napster is an application for sharing mp3 files. It includes a software that allow people to download file. For indexing, Napster provides a directory of music files located on many personal computers throughout the world. This system works as follow. Anyone can publish a list of music files on their computers that they are willing to share on the Napster server. The server matches up requests for files with the list of providers, but the files themselves are transferred directly from personal computer to personal computer. These files do not pass via the Napster server as they would in the client-server model. This is possible because Napster used P2P algorithm to work around the structure of the Internet and allow computers that do not have domain names to locate each other and exchange specific files.

Gnutella [4] is another P2P application but it employs the pure P2P approach for every task. Gnutella, just like Napster, is for music sharing, but without the central server directory. Gnutella users need to download the software start using the system. When someone makes a request for a file, the request is checked by other personal computers and relayed onwards to yet other personal computers down the chain, until the requested file has been located.

FreeNet [3] also uses pure P2P approach, but it differs from Gnutella in two important aspects. First, the information in FreeNet is encrypted, so the person who originally put it on to FreeNet can be anonymous. Once the information has been posted, it moves randomly to another computer. The user of that computer will not know what information is on his machine. Because FreeNet has no central directory, a search engine looks through the entire network each time anyone seeks a file. Second, FreeNet tries to be efficient.

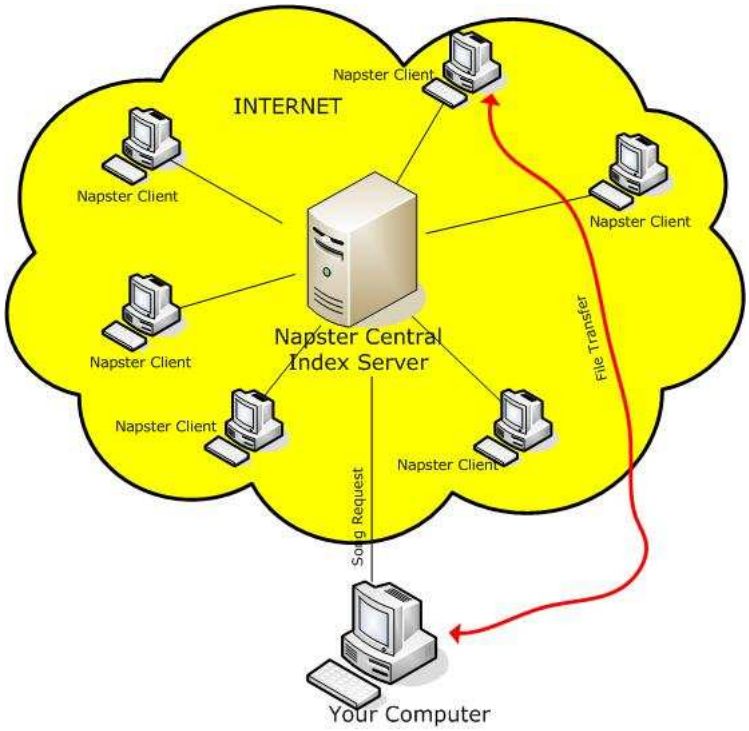

Figure 1: Napster's architecture

Unlike Gnutella, it looks out for popular files and ensures that a number of copies exist in various places. FreeNet will also move the information close to the place where it is in demand. This helps ensure that computers containing the information are not overloaded with requests. This also enables access to files even if the originating computer is off-line.

2.2.2 Distributed hash table

Napster had a central index server: each node, upon joining, would send a list of locally held files to the server, which would perform searches and refer the querier to the nodes that held the results. This central component left the system vulnerable to attacks and legal issues. Gnutella moved to a flooding query model - each search would result in a message being broadcast to every other machine in the network. While avoiding a single point of failure, this method was less efficient than Napster. Finally, Freenet was also fully distributed, but employed a heuristic key based routing [1] in which each file was associated with a key, and files with similar keys tended to cluster on a similar set of nodes. Queries were likely to be routed through the network to such a cluster without needing to visit many peers. However, Freenet did not guarantee that data would be found.

Distributed hash tables (DHT)[1] use a more structured key based routing in order to attain both the decentralization of Gnutella and Freenet, and the efficiency and guaranteed results of Napster. One drawback is that, like Freenet, DHTs only directly support exact-match search, rather than keyword search, although that functionality can be layered on top of a DHT. DHT technology has been adopted as a component of BitTorrent [1].

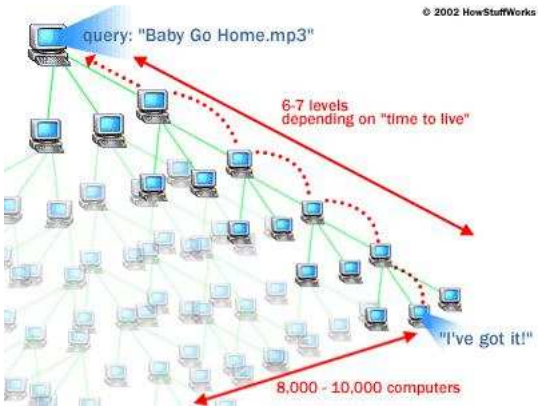

Figure 2: Gnutella's architecture

3 P2P in building social networks

The year 2007 witnessed the explosion of social networking. In October 2007, Alexa found that 7 over the top 10 most visited Websites were Web 2.0 sites. A month earlier, ComScore reported that MySpace, which has more than 55 million unique visitors, is still growing at a constant rate of 23% over the year. At the same time, Facebook increased its number of unique visitors by 129% and Bebo grew 83%. The healthy growth rate creates new challenges. Providers of social networking sites need to be aware of challenges posed by Web 2.0.

- Large Volumes of Content: Social networking sites are attractive to users mostly by allowing them creating personal profiles and sharing personal content. Thus, social networking sites must be capable of supporting photo and video sharing. This leads to massive volumes of rich media content needing to be stored, but still quickly accessible.

- Heavy interactivity: Social networking sites experience higher levels of interaction. It is due to the fact that visitors to social networking sites typically spend more time on them than on most other kinds of sites. Visitors to social networking sites also tend to click-through to more pages, thereby generating a higher number of page views. This is because in social networking sites, different messages, blogs, posts, photographs, and videos are all combined together on one page in thumbnail or summary form, requiring users to click through to view them in full.

  As a result of these factors, social networking sites can experience significant growing pains. The primary challenge of all problems is managing scalability and performance in unpredictable environment.

The limitations of client/server approach are seen in social networks. P2P systems offer an alternative by distributing parts of the system or even the whole system physically or logically. Decentralization is an interesting feature of P2P technologies in order to avoid single point of failures or performance bottlenecks in the system.

Using P2P technologies to develop social networks is good, absolutely for service providers. But does this approach give the same level of benefit for end users? Joining social networking sites based on P2P, users can have parts of their data at home - where they can control data. They don't have to depend on service providers. Also, users can define which information they want to share with which individual person or user-defined groups, instead of policies pre-defined by specific service.

P2P technology expects the end users to contribute bandwidth, hard drive storage and processing power. It will work well if participants are offering their bandwidth and open to seed all content. But in real world, there are multiple problems with P2P systems that could affect their adoption. Most users don't understand how P2P works, and once they do, they get concerned about giving up bandwidth. If just a few participants accept to seed data, it can slow down the social network's performance.

So if users are now happy with just the content available on the other social networks, they aren't going to be happy about their bandwidth being in constant use to save service provider money. Also, how to attract users to new social network sites when they stick to their familiar sites is common problem among new social network sites.

4 Challenges and solutions

Using P2P approach for building social networking, DHT can be used to transfer the core information like user profile and display photos. For transferring large files, we can use a P2P algorithm like BitTorrent.

The challenges here are how to distribute the social network graph [9], and then how to allow nodes to communicate with each other.

4.1 How to distribute the social network graph

The partitioning of the social network graph can cause problems, because the graph can be used for all kinds of purposes, for example spamming and tracking users. Current centralized services, such as Facebook and Google (Google has their social graph API: http://code.google.com/apis/socialgraph), don't have to worry about this. But in a totally decentralized and distributed environment, it is much more difficult to be able to ensure proper operation.

4.2 How to map discovered peer to content

For content discovery, we need a mechanism to map the discovered nodes to content. We have two directions to approach this. One solution is: first, find an element in the social network, then find what content is available. The other solution is: first, search for available content, then select a peer that is part of the social network.

4.3 Successful social networks based on P2P

4.3.1 Skype

Skype is considered the most successful VoIP service using P2P technology, based on Kazaa file sharing [1]. Skype allows its users to make better quality calls, regardless of their location, send instant messages, make video calls, conference calls, transfer files, call traditional telephone for a much less cost than traditional calls. Since its 1.4 release, Skype added social networking like features and is considered a social networking system. The user's profile in Skype shows how many people are in user's contact list. These contact lists form social network. Users can decide with whom they share their contact list or not.

There are three main entities in a Skype network: super nodes, ordinary nodes, and the logon (login) server. A Skype node is simply a computer running the Skype application (also sometimes called a Skype client). A super node is also simply a computer running the Skype application, but Skype application has selected it to take on some of the administrative and coordinating activities of its P2P network. As Skype has not made public the rules under which a node can become a super node, you cannot control whether you become a super node or not; but it is clear that your chances of becoming a super node are greater if you have a better broadband Internet connection. There are millions of nodes and thousands, or perhaps tens or hundreds of thousands, of super nodes.

The main functions of a Skype client (SC) are: login, user search, start and end calls, media transfer, presence messages. A node must register and authenticate itself with a Skype server during login. Skype login servers store information about usernames and passwords to authenticate users' login. There are a lot of different Skype login servers using different ports, and an obfuscated list of servers is hardcoded in the Skype executable. Besides Skype login servers, Skype has no central server.

In order to join Skype community, an SC must connect to a super node to authenticate the username and password with a central server. An SC must know to which super node it has to connect, therefore, SC need to maintain a local table that contains the IPs and corresponding ports of super nodes. This is called host cache (HC) and is stored in Windows Registry of the given SC. SC builds and refreshes HC periodically. Skype has several built-in addresses of different nodes, which are called bootstrap super nodes.

Every time you start Skype, it reads the data from the host cache, takes the first IP and port from there and tries to connect to this SN. If the connection fails for some reason (the SN is offline; it is no longer part of the network...) then it reads the next line from the table. In case it fails to connect to any of the IPs listed, the Skype returns a login error upon start-up. Hence, the host cache must contain at least one valid entry in order for the application to connect to the network and work properly. Valid entry means an IP address and port number of an online Super Node. Not only offer good voice quality calls with low data-rate requirements, Skype also secure the calls. Everything that is being transferred across the network is being encrypted by AES (Advanced Encryption Standard) to ensure privacy.

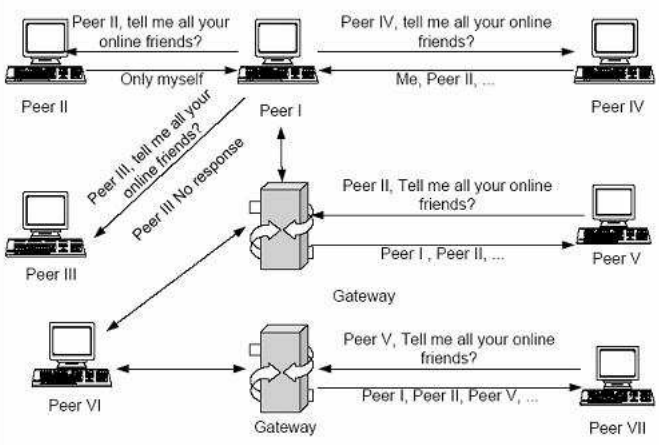

Figure 3: Peers discover in Maze [5]

4.4 Maze

Maze [5, 11] is a P2P file-sharing system and social network which attracts a large number of users. Maze is developed and operated by an academic research team.

For authentication, Maze uses Keberos [1] - like mechanism. Maze uses a central server namely Heartbeat as a directory of online peers. When loading up, a peer sends and "online" message to the Heartbeat server to inform its online status. Then the peer tries to connect to its friends - whose addresses are saved in central server. The peer sends an UDP message "who are you", which includes its unique ID and its ticket, directly to the friend's address. The receiver will reply and "I am somebody" message, which includes its unique ID and its ticket. If the receiver's ticket is valid and its unique ID is one of the requester's friends, the sender peer knows that this friend's status is "online". Symmetrically, the receiver will know that the sender is "online" as well.

Maze peers can also share their friend list and friend status. A peer can get other peers' online friends by sending message "Tell me all your online friends", as shown in the above figure.

In Maze, it is most likely that two peers are friends if they share the same interest. When a peer accepts a file download request, the request will be transmit to friends of this peer to find different locations of the requested file. Thereby, download progress can be speed up by downloading file from multiple locations at the same time.

5 Security concerns

As mentioned in the section 3, P2P technologies would bring bright perspective to develop social networks. This approach has many advantages against traditional client-server approach. But what can happen to the security of networks? In addition to the risks posing in social networking sites, like identity thieves , there are the risks due to the essence of P2P. Following are the risks. Individual malicious peer

In a social network, it is so easy to steal an individual identity and open accounts in their names. Those malicious peers may give incorrect responses to requests: the content

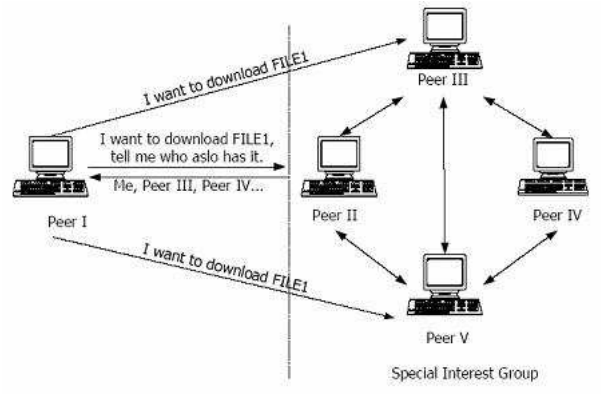

Figure 4: File download in Maze [5]

of data is not what their names or metadata suggest. Thus users get annoyed by contents that they were looking for, and wasted their resources to download them or cache them. This problem can subvert the reputation of P2P social networks. Furthermore, peers may expose themselves to infection with viruses when downloading content from malicious peers.

Groups of malicious peers

When a peer joins a P2P network, it might falsely join a malicious network and become a zombie of a botnet. Another threat of malicious peer is called Sybil attack [**?**]. In which, an attacker creates a large number of identities, thereby can control a substantial fraction of the system.

P2P Worms

Unlike other high-profiles worms, P2P worms [10] in social networks may use a much simpler propagation method: locating potential targets via the "friend list". Later on, these worms can cause DoS attacks and install/run malicious code remotely.

Folder sharing

Misconfigurations of client peer can cause unintentional sharing of local resources. This will lead to possible loss of data confidentiality and data integrity if the wrong people gain access. Even malicious users can exploit P2P vulnerabilities to gain access to local resources.

6 Conclusion

Social networking is growing explosively and attracts a large number of users. As a result, infrastructural challenges become more significant for social networks. As discussed in this paper, the approach of using P2P to build social networking can be the promising answer to the scalability problem. At the same time, like the other P2P applications, this approach faces the same security issues.

References

[1] Wikipedia. At `http://en.wikipedia.org/wiki/`.

[2] Enesto. Peer-to-peer traffic statistics. At `http://torrentfreak.com/peer-to-peer-traffic-statistics/`.

[3] Freenet project. At `http://freenetproject.org/`.

[4] Gnutella. At `http://www.gnutella.wego.com/`.

[5] C. Hua, Y. Mao, H. Jinqiang, D. Haiqing, and L. 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*, pages 290–293, Washington, DC, USA, 2004. IEEE Computer Society.

[6] Napster. At `http://www.napster.com/`.

[7] T. O'Reilly. What is web 2.0. At`http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html`.

[8] K. W. Ralf Steinmetz. *Peer to Peer Systems and Applications*. Springer, 2005.

[9] Thoughts on the social graph. At `http://bradfitz.com/social-graph-problem/`.

[10] L. Zhou, L. Zhang, F. McSherry, N. Immorlica, M. Costa, and S. Chien. A first look at peer-to-peer worms: Threats and defenses. In *Proceedings of Peer-to-Peer Systems IV, 4th International Workshop (IPTPS)*, pages 24–35, February 2005.

[11] M. Zhou, Y. Dai, and X. Li. A measurement study of the structured overlay network in p2p file-sharing systems. *Adv. MultiMedia*, 2007(1):10–10, 2007.