Investigations in Peer-to-Peer Systems

Mujtaba Khambatti {mujtaba} @asu.edu

1. ABSTRACT

Peer-to-peer systems have been enjoying increasing popularity due to claims of being able to harness huge amounts of storage without the drawback of having to trust and rely on a central point of control. These systems create various issues that are not easy to solve. At this time, existing peer-to-peer systems offer solutions for some of the issues. I am engaged in exploring different, efficient and correct techniques that can provide a solution to many of these issues. On a larger scale, I am attempting to define solutions and best practices that can be employed for creating, maintaining and using peer-to-peer systems. My research is currently focused on ways to retrieve information from a peer-to-peer system when there is only some knowledge about the content of that information.

2. INTRODUCTION

There is a growing amount of attention within both research groups and industry that is focused on peer-to-peer (or P2P) systems. Since its recent tryst with the public eye as a result of Napster's [1] P2P style of file sharing, researchers and corporations alike are finding themselves using the term "P2P" more and more frequently. With the subsequent explosion in the number of P2P implementations that have already become available [1, 2, 3, 4, 5, 6, 7], there is a need to formalize the definition of P2P systems along with a list of significant attributes and best practices. Rowstron and Druschel [2] attempt to define P2P systems as distributed systems in which all nodes have identical capabilities and responsibilities and all communication is symmetric.

P2P systems have harnessed huge amounts of storage with little investment and almost no central authority [1, 4, 7]. However these systems still suffer from various problems of scalability, security, fault tolerance and information retrieval. In the traditional client-server model, the server is the central authority for control and information and thus the single point of failure that also requires the trust of its clients. The serverless P2P system establishes a decentralized model where control is a collaborated effort and information is distributed amongst the participating peers so that there is no single point of failure or central authority that has to be trusted. This new model makes it challenging to search for information, share it amongst many peers, make collective decisions, adapt to peer mobility, discover resources, authenticate new peers, set and control access rights for a collection of resources, etc.

One way to characterize P2P systems is by their model. Not all implementations are completely serverless. Some systems, known as hybrid P2P systems, employ a technique using leaders or superclients [8] that play the role of a central authority, but these do not function exclusively as servers. These systems attempt to include the best of both the purely centralized and the purely decentralized approaches.

3. MOTIVATION

The deployment of P2P systems can be envisioned using the following scenarios:

- In an era when personal digital assistants (PDA) are as common an accessory as a wallet or a purse, one can imagine a person walking into a shopping mall with a PDA and discovering the presence of her friends or other users that are willing to share information on the latest bargains and sales. Similarly, a family that decides to shop at the same mall but at different locations can not only maintain a communication link but also discover their relatives or other family groups that might be willing to share information. Stores that sell products of use to families can discover such groups and advertise information about their product.
- Students of a cyber university who are geographically distributed can be pictured working in a group on a semester project that requires programming. By employing a P2P system they exchange source code files, maintain replicas of files and search for new files. Occasionally, the course instructor joins the group to communicate with the group members and also to evaluate the group's status. Due to the lenient

rules of this cyber university, students may join or leave this course at any time requiring that these peer groups be adequately flexible.

4. WORK-IN-PROGRESS

My research is currently focused on ways to retrieve information from a P2P system when there is very little knowledge about the content of that information. Successful attempts to find files using the name of the file or part of the name already exist. However, none of those methods can search for files or other information based on the content or attributes of that information. The dissemination of the results of such queries will be an important issue to solve so that similar queries will be answered quicker. I will simulate my technique and test it for correctness and performance.

In addition to this, I am involved in simulating a set of procedures to efficiently share files through replication amongst a group of peers that will want to read and write to those files. Again these simulations are being built to expose the validity of the approach, the overhead involved and any limits such a system might have.

In the future, my investigations will extend into some security aspects of P2P systems.

5. REFERENCES

[1] Napster.http://www.napster.com/.

[2] A.Rowstron and P.Druschel. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In *Proc. ACM SOSP'01*, Banff, Canada, Oct. 2001.

[3] W.J.Bolosky,J.R.Douceur,D.Ely,and M.Theimer. Feasibility of a serverless distributed file system deployed on an existing set of desktop PCs. In *Proc.SIGMETRICS '2000*, Santa Clara,CA,2000.

[4] I.Clarke,O.Sandberg,B.Wiley,and T.W.Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Workshop on Design Issues in Anonymity and Unobservability*, pages 311 –320, July 2000. ICSI, Berkeley, CA, USA.

[5] R.Dingledine,M.J.Freedman,and D.Molnar.The Free Haven project:Distributed anonymous storage service.In *Proc.Workshop on Design Issues in Anonymity and Unobservability*, Berkeley, CA, July 2000.

[6] J.Kubiatowicz, D.Bindel,Y.Chen, S.Czerwinski, P.Eaton, D.Geels, R.Gummadi, S.Rhea, H.Weatherspoon, W.Weimer, C.Wells, and B.Zhao. Oceanstore: An architecture for global-scale persistent store. In *Proc. ASPLOS '2000*, Cambridge, MA, November 2000.

[7] The Gnutella protocol specification, 2000. http://dss.clip2.com/GnutellaProtocol04.pdf.

[8] B.Yang, H.Garcia-Molina. Comparing Hybrid Peer-to-Peer Systems (extended), *Technical report*, 2000.