“A P2P Storage System Based on Erasure Codes”
Dina Katabi

Peer-to-Peer systems, almost unheard of a few years ago, are now becoming one of the most popular Internet applications. The most important use of P2P systems is to store and share files. In a P2P network, nodes may join and leave the system at any time. As a result of the intermittent availability of the nodes, ensuring high availability of the stored data is an interesting and challenging problem. In particular, the tradeoffs between the availability of the files and the redundancy involved in storing the data is not yet well-understood.

To ensure the availability of the stored data, one needs to add a high degree of redundancy. Most of the current proposals provide redundancy through replication. For example, one might replicate the whole file as in Pastry [1] or replicate blocks of the file separately as in CFS [2]. In contrast to previous work on P2P storage, in this project, we will explore the benefits of providing high availability for stored files through redundant coding of the data using some form of erasure codes (Reed-Solomon codes [3]], Tornado codes [4]], etc.).

Erasure codes encode an object (e.g., a file) of size K into N fragments, where N > K. Although each one of the N fragments differs from the others, the retrieval of any K out of these N fragments allows a correct reconstruction of the data object. Erasure codes are traditionally used to send a data object over a lossy channel. A P2P system resembles a lossy channel because any node can leave the system, which results in the loss of the data its stores.

The advantage of redundant encoding over replication is that it provides a higher availability for the same amount of redundancy (this is why it is used over lossy channel rather than just sending multiple copies of the data). Further, by using a small fragment size and distributing a fragment to each node in the P2P network, one can eliminate the need for searching or maintaining a directory of who stores which files, which could be a complex task given the size of the system and its quick dynamics. A third advantage of using erasure codes results from the fact that the fragments will be distributed to many machines. Thus, a user retrieving the file can simultaneously download the fragments. This allows the user to benefit from the bandwidth available on all paths to the machines storing the fragments, which decreases the time taken to retrieve the file.

We will explore all of the above issues and evaluate the benefits and the drawbacks of using erasure codes in P2P storage systems.

References

1. Antony Rowstron, Peter Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. Lecture Notes in Computer Science, 2001.

2. F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica. Wide-area cooperative storage with CFS. In Proceedings of Chateau Lake Louise, Canada, October 2001.

3. J. Plank "A tutorial on Reed-Solomon coding for fault-tolerance in raid-like systems", Software Practice and Experience, 27(9):995-1012, September 1997.

4. John Byers, Michael Luby, Michael Mitzenmacher. Accessing Multiple Mirror Sites in Parallel: Using Tornado Codes to Speed Up Downloads. INFOCOM 1999