

Computer scientists develop P2P system that promises faster music, movie downloads

April 10 2007

A Carnegie Mellon University computer scientist says transferring large data files, such as movies and music, over the Internet could be sped up significantly if peer-to-peer (P2P) file-sharing services were configured to share not only identical files, but also similar files.

David G. Andersen, assistant professor of computer science at Carnegie Mellon, and Michael Kaminsky of Intel Research Pittsburgh have designed such a system, called Similarity-Enhanced Transfer (SET). By identifying relevant chunks of files similar to a desired file, SET greatly increases the number of potential sources for downloads. And boosting the number of sources usually translates into faster P2P downloads, Andersen explains. How much SET could speed up downloads varies based on a number of factors, including the size and popularity of a given file. In some cases, SET might speed transfers by just 5 percent; in others, it might make downloads five times faster.

The researchers, along with graduate student Himabindu Pucha of Purdue University, will present a paper describing SET and release the system code at the 4th Symposium on Networked Systems Design and Implementation, April 11 in Cambridge, Mass.

"This is a technique that I would like people to steal," Andersen said. Though he and his colleagues hope to implement SET in a service for sharing software or academic papers, they have no intention of applying it themselves to movie- or music-sharing services. "But it would make P2P transfers faster and more efficient," he added, "and developers

should just take the idea and use it in their own systems."

"In some sense, the promise of P2P has been greater than the reality," Andersen said. By creating many more sources for data files, P2P reduces bottlenecks for data transfers. But residential Internet service providers allot far more bandwidth for downloading than they do for uploading files, an imbalance that continues to slow P2P data transfers. And members of P2P services often limit their computer's upload capacity so it is not tied up with other peoples' uploads.

Like P2P services such as BitTorrent, Gnutella and ChunkCast, SET speeds up data transfers by simultaneously downloading different chunks of a desired data file from multiple sources, rather than downloading an entire file from one slow source. Even then, downloads can be slow because these networks can't find enough sources to use all of a receiver's download bandwidth. That's why SET takes the additional step of identifying files that are similar to the desired file.

No one knows the degree of similarity between data files stored in computers around the world, but analyses suggest the types of files most commonly shared are likely to contain a number of similar elements. Many music files, for instance, may differ only in the artist-and-title headers, but are otherwise 99 percent similar.

Different versions of software packages likewise remain highly similar.

Taking advantage of those similarities could speed downloads considerably. If a U.S. computer user wanted to download a German-language version of a popular movie, for instance, existing systems would probably download most of the movie from sources in Germany. But if the user could download from similar files, the user could retrieve most of the video from English versions readily available from U.S. sources, and download only the audio portion of the movie from the

German sources.

SET's basic operation is similar to that of BitTorrent. Once the download of a data file is initiated, the source file is divided into smaller, unique chunks — SET divides a one-gigabyte file into 64,000 16-kilobyte chunks, for instance. Different chunks are downloaded simultaneously from numerous sources that have the identical file, and then the chunks are reassembled into a single file.

But while that process of downloading is under way, SET continues to search for similar files using a process called handprinting, inspired by techniques that have been used for clustering search results or detecting spam. A sampling technique is used to see if non-identical files contain chunks matching those of the desired file. Relevant chunks can then be downloaded from the similar files identified by this method.

In tests based upon real files downloaded from today's peer-to-peer networks, SET improved the transfer time of an MP3 music file by 71 percent. A larger 55-megabyte movie trailer went 30 percent faster using the researchers' techniques to draw from movie trailers that were 47 percent similar. The researchers hope that the efficiency gains from SET will enable the next generation of high-speed online multimedia delivery.

Source: Carnegie Mellon University

Citation: Computer scientists develop P2P system that promises faster music, movie downloads (2007, April 10) retrieved 10 April 2024 from <https://phys.org/news/2007-04-scientists-p2p-faster-music-movie.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
