

# Preserving Data Integrity in Peer-to-Peer Networks

Mema Roussopoulos  
Gordon McKay Endowment, Harvard University

Peer-to-peer systems have become synonymous with file-sharing systems. Much of the focus of research in this area has been on providing algorithms to improve the efficiency, robustness, and security of routing in peer-to-peer systems, or designing services such as indexing and search for use by file-sharing applications running on these systems. There has been less focus on discovering new applications or enumerating the characteristics of applications for which peer-to-peer systems provide a viable, if not the only, solution.

The purpose of this talk is to show that peer-to-peer is more than just file-sharing. I will review the definition of peer-to-peer and describe a set of characteristics of applications for which peer-to-peer networks are a necessary solution. I will then describe an example application, a peer-to-peer digital preservation system called LOCKSS (Lots of Copies Keep Stuff Safe).

The LOCKSS system is a tool librarians can use to preserve long-term access to content published on the web. The system is currently being deployed at about 100 libraries around the world with the support of publishers representing 2000 electronic titles. LOCKSS has three main functions. It collects the content by crawling the publisher's web sites, it distributes the content by acting as a proxy for reader's browsers, and it preserves the content through a cooperative, peer-to-peer process of damage detection and repair. I will describe our experience designing and evaluating this system.

## Short Biography

Mema Roussopoulos is an Assistant Professor of Computer Science on the Gordon McKay Endowment at Harvard University. Before joining Harvard, she was a postdoctoral Fellow in the MosquitoNet Group at Stanford University. She received her PhD and Master's degrees in Computer Science from Stanford, and her Bachelor's degree in Computer Science from the University of Maryland at College Park. Her interests are in the areas of distributed systems, networking, mobile computing, and digital preservation.