Peer sharing behaviour in the eDonkey network and Design of Server-less File Sharing Systems

Anne-Marie Kermarrec

IRISA, Rennes

Peer-to-peer file sharing systems have grown to the extent that they now generate most of the Internet traffic, way ahead of Web traffic. Understanding workload properties of peer to peer systems becomes necessary to optimize their performance. In this talk, we present the analysis of an eDonkey peer to peer file sharing system workload of over 350 terabytes of data, gathered mainly in Europe between December 2003 and February 2004. The trace is based on an active crawl of over 50,000 client caches per day, where a cache consists of files a peer agrees to share. We first present summary statistics on file replication and popularity, and peer geographic distribution. We also analyze temporal properties of the workload, namely the evolution of file popularity, and of peer cache contents. We are able to analyze fine properties such as the evolution of the overlap between distinct peer caches thanks to the massive information obtained from the crawler. We focus on the emergent clustering properties of such workloads, in terms of similarity of interest between peers. We assess such clustering properties via both direct metrics and an indirect operational index. The latter measures how frequently peers could retrieve files they sought from caches of appropriately selected "semantic neighbours". We observe the presence of a significant amount of interest-based clustering. This suggests several potential optimizations of current peer to peer networks by exploiting this property.