Looking Beyond the Legacy of Napster and Gnutella.

Kiran Nagaraja, Sami Rollins, Mujtaba Khambatti
kiran@nec-labs.com, srollins@mtholyoke.edu, mujtabak@microsoft.com

Trace back to the year 1999. This was when Napster took the music world by storm. The application brought together millions of music enthusiasts eager to share and download each other's song collections. The ease with which two 'peers' could share a file was refreshing. Eventually, this resulted in a network of 60 million Napster users! It was also around this time that the phrase 'peer-to-peer' (P2P) came to be associated with systems like Napster. The phenomenon not only stirred the social sphere, but also created a buzz within the academic and industrial worlds. Anyone who knew anything about distributed systems asked the question 'Is peer-to-peer computing really a new paradigm?' The curt answer was NO. Usenet news and the email system based on the Simple Message Transfer Protocol (SMTP) were examples of systems, invented as far back as the 1980s, which employed identical concepts of decentralization. The detailed answer is that the significant number of novel contributions in data structures, algorithms, and mechanisms that have emerged from this field over the past five years justify its distinction from distributed systems in general. However, despite this innovation, a summary look at the applications that have emerged from this field tells a mostly monotonic story.

The Legacy Owing to the popularity of Napster and its successors, including Gnutella, Kazaa, Morpheus, and E-Donkey, file sharing became by far *the* killer P2P application. In fact, its popularity almost completely eclipses other P2P applications. Moreover, many of the research contributions in the P2P space have been targeted toward addressing challenges facing these popular file-sharing networks. Gnutella did address Napster's shortfall of complete decentralization. But, it's unstructured nature brought forth concerns over the efficiency and scalability of its search mechanism. In 2001, the use of 'superpeers', which are a set of more powerful nodes in a heterogeneous network, was utilized to transform the flat topology of existing networks into a hierarchy. Superpeers are considered faster and/or more reliable than normal peers and take on server-like responsibilities. For example, in the case of file sharing, a superpeer builds an index of the files shared by its 'client' peers and participates in the search protocol on their behalf. This improves scalability by limiting the flood of search traffic. Simultaneously, various projects such as Chord, Pastry, Tapestry, and CAN proposed the use of Distributed Hash Tables (DHTs) to provide an efficient way to locate objects within the network. DHTs proposed a structured approach to building an overlay that would efficiently support mapping of an object identifier to its location. Even though such core contributions had their root in solving the decentralized search problem of file sharing systems, it became widely accepted that the core operation in most P2P systems was efficient location of data objects. Apart from services for content location and query routing, approaches to provide security and availability for content, and studies to establish authentication, trust and reputation of individual peers were also actively pursued. In summary, advances along these directions have helped establish a firm base to explore a complex set of applications beyond file sharing.

Other Domains The hype did raise enough curiosity among researchers and developers to apply P2P mechanisms to domains other than file sharing. Some of the early explorations were in the areas of distributed storage, content distribution, communication and collaboration, and to an extent even decentralized gaming [1]. Reduced management costs per entity, absence of central points of failure, incremental scalability, a potentially larger resource pool, reduced individual scrutiny, and potentially higher fault resilience were the primary advantages over centralized solutions. However, except for a few real products such as Groove Network's Virtual Office for enterprise-wide collaboration, most of these efforts remain academic. In fact, only the use of P2P networks for

[1] We purposely exclude distributed computing applications such as grid computing. These systems do employ P2P characteristics such as decentralization and incremental scalability, however, their usage, at least in its current form, is often managed either at a single location or at multiple ones in a federated manner.

distributed storage services can match the level of interest generated by file sharing. Oceanstore, PAST, Pastiche, PeerStore, and CFS are among the notable academic projects that propose the use of P2P networks for universal access to or archival of content. Content distribution also received interest. Networks such as Akamai and BitTorrent optimize network usage for fast and efficient download of content by end hosts. Essentially, despite all of the advances in developing core services, the application of P2P has remained well within the domain of content sharing. This field has yet to see revolutionary applications beyond what Napster demonstrated.

The Revolution is Coming? P2P has democratized the way people use computing. It links people irrespective of their location and affiliation and it raises numerous possibilities of interaction and collaboration not only in social activities, but also on a commercial level. First, it presents an opportunity for sharing digital content. This is actively being pursued by various efforts mentioned above. It also presents an opportunity for sharing computing and other resources. Though we have seen applications, such as SETI@Home, that harness idle computing cycles from a pool of Internet hosts, there are none that allow individuals to gather and use such resources by mutual consent. Second, it makes unbounded social interaction and collaboration possible without having to depend on centralized rendezvous points. P2P communication, conferencing, network gaming (such as first person shooter - FPS), voting, opinion polls, and synchronized viewing of video streams are some examples of applications that can leverage this feature of P2P. Existing applications such as NetMeeting and some features within today's instant messaging systems (MSN, Yahoo, etc.) are glimpses of what is possible. There are also a few P2P applications that support collaboration between people within an enterprise, but they are specialized to deal with data and processes within a work setting. Finally, the ultimate test for P2P's success will be its ability to support applications of commercial value over the network. Establishing market places and auctions in a P2P setting would create the equivalent of today's Amazon and eBay without the negatives associated with centralization.

Current research is certainly targeting a broader P2P application space. There are several proposals to build more complex application using the existing base of efficient and scalable core services. Some examples include the creation of a digital library [3], a complex massively multi-player online game that uses Chord for object location [1], and a commercial content distribution network that assumes the presence of neutral superpeers to perform bookkeeping for downloads between the provider and consumer [4]. There are also ongoing research efforts to construct network-wide infrastructure, such as a P2P-based DNS [2] and a spam-filtering service [5], that can help improve overall performance of legacy and P2P applications alike. However, for all of this to translate in to practice, developers need to be able to benefit from the solid platform established by the research community. Consider the breakdown shown in Figure 1 of the research topics addressed by papers presented over the last four years at the International Workshop on Peer-to-Peer Systems (IPTPS) – a prominent P2P

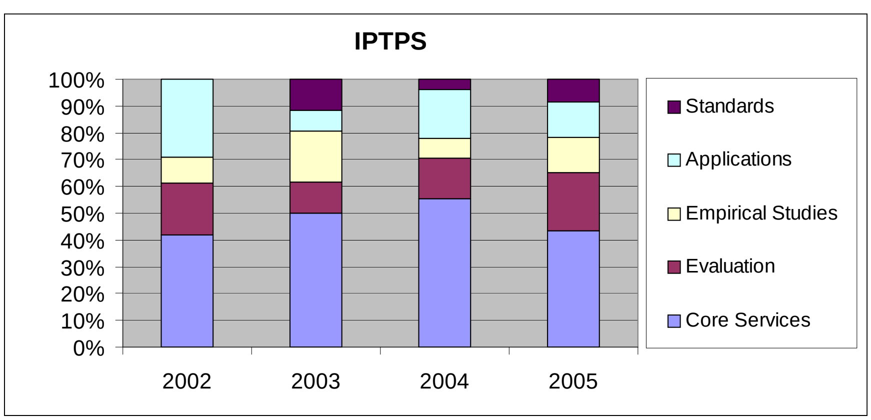

Figure 1: *Focus of P2P researchers as evident from accepted papers at the IPTPS workshop.*

workshop. This forum encourages submissions to discuss state-of-the-art in P2P and to identify key research challenges. Expectedly, core services have received the most attention over the years and we expect this trend to continue. Applications, however, are neglected. More disappointing is the low fraction of submissions that address concerns of application developers. Represented by the key 'Standards' in the graph, this intends to cover topics such as "What is the 'right' set of interfaces to be exported by core services?", and "How does one choose the appropriate data structure, algorithm, or mechanism under the specific condition presented by the application?". To repeat a point made by one of the papers at this very venue, all of the research will receive neither feedback nor validation unless there is an active set of clients for the technology. This appears to be a major problem. Efforts such as JXTA, which invite the community to be involved in the process of open source development of P2P specifications, standards, and technology, are steps in the right direction. The systems community involved in P2P research has to step up to this challenge as well.

In conclusion, over the last five years the accumulated advances in the area of P2P have helped establish a firm set of core structures, algorithms and mechanisms to address various concerns of a P2P network including performance, availability, security, and trust. However, the field is yet to realize its full potential in the application domain. Its time to move beyond file-sharing and fill in the void.

References

[1] S. Douglas, E. Tanin, A. Harwood, and S. Karunasekera. Enabling Massively Multi-Player Online Gaming Applications on a P2P Architecture. In *Proceedings of International Conference on Information and Automation*, Dec. 2005.

[2] K. Park, V. S. Pai, L. Peterson, and Z. Wang. CoDNS: Improving DNS Performance and Reliability via Cooperative Lookups. In *Proceedings of the Sixth Symposium on Operating Systems Design and Implementation(OSDI '04)*, Dec. 2004.

[3] J. Stribling, I. G. Councill, J. Li, M. F. Kaashoek, D. R. Karger, R. Morris, and S. Shenker. OverCite: A Cooperative Digital Research Library. In *Proceedings of Fourth International Workshop on Peer-to-Peer Systems (IPTPS '05)*, Feb. 2005.

[4] A. Wierzbicki and K. Goworek. Peer-to-peer Direct Sales. In *Proceedings of Fifth International Conference on Peer-to-Peer Computing (P2P '05)*, August-September 2005.

[5] F. Zhou, L. Zhuang, B. Y. Zhao, L. Huang, A. D. Joseph, and J. Kubiatowicz. Approximate Object Location and Spam Filtering on Peer-to-peer Systems. In *Proceedings of International Middlware Conference (Middlware '03)*, June 2003.