



**Peer-to-Peer: A tool for stealing music?! Many people would be tempted to argue that yes! No doubt, sharing music and video files has been the catalyst in helping the P2P term spread its reputation to the wide public, with frequent references on the TV and the general press.**

Contrary to common belief, however, there's a whole lot more in P2P than file-sharing. What is it then? It's not limited to any specific application. It's a whole new philosophy in doing things decentralized. It's about enabling a large set of nodes cooperate without central coordination to collectively carry out some task. It's about exploiting the combined power of a large collection of nodes. It's about distributing the execution of a service. It's about utilizing the resources at end nodes. It's about pushing the responsibilities of a service to the end nodes. And these nodes can be highly heterogeneous, distributed across all continents, under various administrative domains, non-dedicated, and unstable.

Consequently, participants in a P2P system are not mere consumers, passively using the service provided by some server. There is no server after all.

Instead, they interact, participate, and contribute their resources to the service they make use of. In the P2P model, ordinary computers are promoted from passive entities to active components. The main philosophy behind these systems is communal collaboration: Computers share both duties and benefits!

The interesting property of this model is that system load and system resources grow aside, proportionally to the number of participating computers. As a result, P2P systems tend to be extremely scalable, and can spontaneously adapt to unexpected increase or decrease in demand.

Getting back to the file-sharing example, these P2P notions come in different flavors. In the first massively used system, Napster (ca. 1999), searching for files was centralized: a server kept track of the music libraries (just the titles and metadata, not the actual files) of all online peers, processed incoming queries, and matched downloaders to the appropriate uploaders. Clearly, searching was not done in a P2P way. The functionality spread across end-nodes was downloading. The actual downloading of songs was done directly between peers in a pure P2P fashion, effectively harnessing the enormous

aggregate bandwidth capacity of all participating nodes. Indeed, the alternative of matching that bandwidth capacity by a dedicated cluster of servers would have been enormously expensive and difficult to implement, if not infeasible. Later systems, such as Gnutella, eDonkey2000, FastTrack (KaZaA), additionally distributed the search functionality on multiple nodes. The motivation this time was to spread the administration on multiple domains and countries, to evade law enforcement for copyright violation. Still, exploiting aggregate bandwidth has been the cornerstone of file-sharing systems, most notably BitTorrent.

Bandwidth is the most typical resource P2P applications tend to harness, particularly the ones focused on disseminating information to large node populations. Various forms of such applications exist. Content Delivery Networks (CDNs), and broadcast / multicast systems disseminate information to large sets of nodes, propagating the data through a P2P network. Other applications specialize in delivering streams of data to the end nodes, utilizing the nodes themselves for the propagation of the stream. Such systems include video or audio streaming, and video on demand. Special note should be made for VoIP (Voice over IP) systems, such as Skype, which provide internet telephony. Such systems form a P2P network, which they use to route phone call streams.

Other P2P systems aim at harnessing different types of resources. Processing power is a resource that comes in abundance, considering the vast number of unused CPU cycles on multiple ordinary computers around the globe. Some projects (e.g., Seti@home, the World Community Grid, etc.) aim at harnessing this power, by splitting heavy computations in smaller parts and distributing them to the participating peers. Although

coordination in such systems is often centralized, it is the contribution of resources of the end nodes that qualify them as P2P.

Storage is another resource of interest. Some P2P systems were designed for exploiting unused storage space on ordinary computers to provide a distributed network storage service. Typically, they split files on small chunks, and distribute them on various nodes across the network. Moreover, they often replicate each chunk, to provide data persistence against unpredictable hardware failures, natural disasters, human mistakes, etc. Other resources of interest include people's presence (think of Instant Messaging applications), and collaboration between users (shared whiteboards, co-reviewing, co-editing, etc.). Forming semantic networks, clustering users based on their interests and preferences can prove very useful for recommendation services. Last, but not least, gaming has taken a massively distributed direction in the last few years, involving large communities of players interacting in a virtual environment.

The P2P model of communication poses a number of challenging issues. Most important is the unprecedented scale of such systems. The size of decentralized systems has grown from large to massive, in certain cases involving millions of nodes spread all around the globe. And with massive scale comes massive instability, due to the highly dynamic nature of systems of that size. Typically, nodes in P2P systems provide no guarantees regarding their participation patterns. Instead, they join and leave at any time at will, let alone unpredictable ungraceful failures of nodes and links. Finally, nodes in P2P systems are generally heterogeneous with respect to their hardware architectures and software platforms, as well as regarding their computational, memory, storage, and network resources.

The aforementioned issues place a critical burden on the administration and management of such systems. Massive scale in combination with highly dynamic environments renders explicit control of such systems infeasible, or at least very hard to implement. Centralized systems impose severe limitations in keeping track of millions of nodes coming and going. Scale issues can be alleviated by means of a more sophisticated architecture, such as a hierarchical structure. However, this increases complexity and administration costs, and the system depends on the availability of certain key nodes. Solutions introducing redundancy by means of replication exist, however they inflict additional complexity which is nontrivial at such scale. On top of that, all these solutions (i.e., both centralized and hierarchical ones) require an educated guess of the expected system size in order to appropriately allocate adequate resources. Generally, trying to impose explicit control on this class of systems and tracking them by traditional deterministic methods becomes increasingly complex.

The Peer-to-Peer and Self-Organizing Networks course, offered in the current semester to Master's students at ETH, explores such issues of P2P systems. The focus is to address the challenges in scale and autonomous coordination, study the most notable and recent protocols in the area, and comprehend the mechanisms behind building robust P2P systems. Particular attention is given in developing a critical thinking stimulating research initiatives.

If you are motivated by P2P systems and feel you should have joined this course, we have some good news for you! Our current research focuses on such issues too! We step away from the deterministic and explicit control of massive-scale systems, exploring methods that enable their autonomous

management by means of self-organization. Our research aims at two directions. First, introduce new protocols, and explore their usage in a number of current or future applications. Second, come up with new application models, new ideas that could offer novel useful functionalities, either based on existing P2P solutions, or triggering new research. An example could be P2P on wireless network, but not only. So, if you feel enthusiastic about this area, if you have some wild ideas on revolutionary directions that may sound exaggerated today, or if you would just like to have a chat on related topics, you are very encouraged to contact me at [spyros@inf.ethz.ch](mailto:spyros@inf.ethz.ch), and come discuss such issues in person.

△