Executive Summary of the Dagstuhl Seminar 06131
"Peer-to-Peer-Systems and -Applications"
26.03.06 - 29.03.06

Organizers:
Anthony D. Joseph (Univ. California - Berkeley)
Ralf Steinmetz (TU Darmstadt)
Ion Stoica (University of California at Berkeley)
Klaus Wehrle (Universität Tübingen)

Classification:
- Networks

Keywords:
- Peer-to-Peer
- Self-organization
- Massively distributed systems

Abstract

This is the executive summary of Dagstuhl Seminar 06131, "Peer-to-Peer-Systems and –Applications". The seminar was held from March 26th to March 29th, 2006, at the International Conference and Research Center for Computer Science at Castle Dagstuhl, Germany.

Motivation

Under the term "Peer-to-Peer", a very promising paradigm for communication in the Internet arises. Though it was originally used for pragmatic and not always legal file-sharing activities, the Peer-to-Peer technology offers interesting opportunities for highly distributed and scalable systems and applications.

According to recent reports from ISPs, a major amount of Internet traffic is governed by Peer-to-Peer applications. Due to the continuing growth and diversification of the Internet and its applications, it becomes exceedingly difficult to meet the resource demands by traditional Client-Server solutions. These centralized approaches can be hardly realized for, e.g., file sharing applications, distributed file systems, or grid computing environments.

Given this persistent and long-term development, there are three fundamental challenges for current and future Internet applications:

- Scalability is of utmost importance in order to cope with user bases and resource consumption of applications (in terms of bandwidth, storage, processing, etc.) growing by several orders of magnitude.
- Only through security and reliability it is possible to maintain the availability of centralized services in the face of distributed denial-of-service attacks. Data privacy and censorresistance are also of growing concern.
- Flexibility and quality of service allow the rapid deployment of new technologies throughout the Internet, e.g. to realize long-promised multicast and host mobility features.

The Peer-to-Peer paradigm shows the potential to meet these challenges. Peer-to-Peer systems share distributed resources and address services based on content rather than location. Without the necessity for central entities, they organize themselves into cooperating infrastructures of symmetric peers. The two approaches of structured vs. unstructured P2P systems allow for different and sometimes complementary trade-offs and still bear a wide range of ongoing and future research in the P2P area.

The goal of the second Dagstuhl seminar on Peer-to-Peer-Systems and –Applications was to assemble researchers being highly active in the area of Peer-to-Peer mechanisms and networking

1. to reflect on recent research activities,
2. to identify future research issues, i.e. major challenges and
3. to strengthen the Peer-to-Peer community in research

Organization

About 63 researchers accepted the invitation to come together in Dagstuhl for this meeting. 19 of them had the chance to present their results in a 30 minutes plenary talk, separated into six sessions.
The 19 talks covered a wide range of topics; for example the discussion what Self-Organization is to applications for P2P-Systems. Additionally, 7 posters were presented during a coffee break. There was also a teamwork session, where small groups discussed a topic in detail and presented the results in the plenum. The social event was a nice city tour to Trier.

Teamwork

We had six working groups discussing different topics in the area of P2P. The results of the working groups can be found in the working groups' slides.

Teaching "Peer-to-Peer Systems and Applications"
Shepherd: Klaus Wehrle
The aim of this team was to discuss what topics are interesting for teaching courses on Peer-to-Peer systems and applications. Various participants of the seminar already gave courses on topics related to this seminar. Based on these experiences, the group wanted to find a common understanding of which topics should be covered in a course.

Information Retrieval in Peer-to-Peer Networks
Shepherd: Wolf-Tilo Balke
Today's information provisioning in peer-to-peer systems has to deal with several types of information. First there are data files or textual content for sharing that can be more or less described by adequate metadata. In contrast there is process data, which is continuously created, like for instance log information, intermediate results or state information. This data is essential in supporting long-running applications, where users interact with the infrastructure over a significant time-span, e.g., media streaming for IPTV or online experiments. Such data can easily fill 5-10 MB of sequential, non-indexed data files and has to be monitored or searched at run time to detect events that enable the intelligent adaptation of parameters for quality of service (QoS) control or graceful recovery from failures.

The type information in such process data, however, is quite diverse ranging from simple structured information (like CPU usage of a certain node, delays within the network or concise information about a video frame, etc.) to largely unstructured data (like verbose failure descriptions or newly discovered service functionalities). Keeping this data readily accessible is a crucial challenge. Major problems are on one hand the creation of light-weight structures for logging or keeping intermediate states, which allow supporting IR-style retrieval especially with respect to online-algorithms and real-time IR techniques. Moreover, the support of functions for aggregation of data, finding correlations in data and conditional expressions, as well as effectively indexing (highly discriminating) terms or phrases and ranges of values. On the other hand, peers need to offer this information (possibly in different resolutions) and all information needs to be persistent even in the presence of churn (i.e. a certain degree of replication is necessary) to allow for graceful recovery or checkpointing.

For these ends lightweight ontologies and dictionaries, as well as indexed aggregated information in terms of ranges or sets (for instance kept in a DHT structure) can be expected to be useful. This is primarily because for different applications they enable the explicit formulation of a trade-off between the performance and necessary precision of queries.

Peer-to-Peer and Industry

Shepherd: Oliver Heckmann

Aim of this teamwork was to identify commercial applications, business models and those markets, where the application of P2P architectures makes most sense and of those, where we can assume that client-server architectures will remain dominant. Further, the group discussed other aspects where P2P affects industry, for example Internet service providers that have to deal with P2P traffic and whether they benefit from it or not.

Self-Organization: Hype or Key Technology?

Shepherds: Hermann de Meer, Thomas Fuhrmann

Aim of this teamwork was to discuss the basics and definition of Self-Organization. Self-organization has been explored in the natural and social sciences for some time now. Yet, it is still open how self-organization can be defined in computer science. During the Dagstuhl discussion, the group named complexity and emergent properties as important criteria. Self-organizing systems are distributed systems that are governed by rules operating on locally available information. Thereby, the human is kept out of the loop (cf. autonomic and organic computing). Self-organizing systems in computer science should serve a purpose, but the system's behaviour is not entirely predictable. As a result, self-organizing systems are robust and can adapt to changing conditions and environments. However, many open questions remain:

- Can we identify general principles of self-organization?
- Can we establish engineering rules to build (composable) self-organizing systems?
- Can we name objective criteria when to employ self- organization?
- Which would be good examples to further explore self-organization in computer science?

Mobile Peer-to-Peer

Shepherds: Wolfgang Kellerer, Anthony D. Joseph

Mobile and wireless systems can especially benefit from the self-organizing characteristics of P2P systems, e.g., with respect to robustness, however the heterogeneity and dynamicity of those systems also poses severe challenges to P2P concepts. Moreover exploiting mobile systems features, new applications for P2P systems are opened such as location based services. The degree of controllability as an operator requirement versa self organization has to be considered as another important challenge for mobile P2P approaches. The focus was to elaborate on whether the current state of the art in P2P solutions is sufficient for mobile systems and discusses open research issues as well as emerging application classes.

What P2P can do and what P2P can't do

Shepherds: Ernst Biersack, Ralf Steinmetz

The aim of this team was to discuss what the limits of P2P are.

P2P is one form of distributed computing with the following characteristics: (i) commoditized resources, (ii) massive/world-wide scale, and (iii) anybody can contribute some resources, which makes for the non-commercial aspect of many p2p applications with p2p networks having the flavor of "social organizations".

P2P has the advantage that "everybody" can come develop and deploy a new p2p application (e.g BitTorrent, Gnutella, etc), which is due to the externalization of the cost of deployment of a service (see e.g. Skype) that allows for a fast deployment with little initial investment.

P2P systems can have distinct advantages in certain cases, for instance when used for

- Archival storage where the fragments of a document are spread over many machines, which makes it hard for an adversary to alter/destroy the document
- Assuring anonymity since identities are cheap to create and difficult to trace back to a real person

The fact that identities are cheap to create poses however problems in the area of security and trust that need to be addressed.

P2P also has some limitations that may make it very difficult or impossible to

- assure transactional ACID-like guarantees or make it hard to
- build applications that must be able to handle frequent and fine-grained mutation of state.

P2P is still very much in its infancy and much more research is needed to develop

- Tools to model and analyze large scale p2p systems and understand tradeoffs between conflicting goals
- A “theory” of p2p systems as there exists one for distributed systems, which is needed to prove certain properties .