COMAD 2006 KEYNOTE TALK

Information Search in Peer-to-Peer Systems

Gerhard Weikum
Max-Planck Institute for Informatics
Saarbruecken, Germany
weikum@mpi-inf.mpg.de
http://www.mpi-inf.mpg.de/~weikum

ABSTRACT

The peer-to-peer (P2P) computing paradigm has been very successful in the proliferation of global applications like file sharing in Internet-wide communities (e.g., Gnutella, BitTorrent) or IP telephony (e.g., Skype). P2P systems should be completely decentralized and should work without any centralized components that could become bottlenecks in terms of performance, availability, or vulnerability to attacks. They should be scalable without any limitations, by being able to grow from a few nodes to many millions of computers. They emphasize the autonomy of the underlying computers and should tolerate frequent node failures, high dynamics in terms of rapidly changing data and load characteristics, and high churn by allowing nodes to join and leave the network without prior notice. They should even be robust to misbehaving peers which may span egoistic, cheating, and malicious peers. None of these salient properties should require any global planning, administration, or control. So P2P systems should be completely self-organizing. In addition, P2P systems should have a software architecture that is much simpler than that of big monolithic systems, to enable scalability and self-organization.

Despite the impressive success of P2P file-sharing applications, the question is still valid and, in my opinion, widely open whether the outlined P2P utopia will become practically viable also for more advanced forms of global data management and information search with more sophisticated functionality.
This talk discusses this question, aiming to identify design principles and building blocks towards the P2P dream of simple, scalable, and self-organizing data management on an Internet scale. Advanced applications that drive the discussion include Google-style Web search implemented in a P2P manner, decentralized Web archiving with time-travel query support, and P2P publish-subscribe functionality for scholarly information (e.g., publications, projects, conference sites and reports, etc.) and similar social communities.

Speaker's Profile: Gerhard Weikum is a Scientific Director at the Max-Planck Institute for Informatics in Saarbruecken, Germany, where he is leading the research group on databases and information systems. Earlier he held positions at Saarland University in Germany, ETH Zurich in Switzerland, MCC in Austin, Texas, and he was a visiting senior researcher at Microsoft Research in Redmond, Washington. His recent working areas include implementation, optimization, and self-organization aspects of distributed information systems such as peer-to-peer systems, and intelligent search and organization of semi-structured data on the Web and in digital libraries. Dr. Weikum has received several best paper awards including the VLDB 2002 ten-year award, and he is an ACM Fellow. He has served on the editorial boards of various journals and book series, including ACM TODS, IEEE CS TKDE, and the Springer LNCS series, and as program committee chair for international conferences like ICDE 2000 and ACM SIGMOD 2004. He is currently the president of the VLDB Endowment.

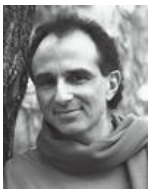
COMAD 2006 KEYNOTE TALK

StrangerDB: Safe Data Management with Untrusted Servers

Dennis Shasha
Department of Computer Science
Courant Institute of Mathematical Sciences
New York University, New York, NY 10012, USA
shasha@cs.nyu.edu
http://cs.nyu.edu/shasha

ABSTRACT

Imagine that you and your friends want to share information in a database because you want concurrency control, recovery, and query processing, but you don't trust the database administrator. You want to protect data from being observed (privacy). You want to make unauthorized modifications evident (a form safety). You want to force the server to deliver a consistent picture to all honest users or be discovered (a form of liveness). Encryption and signatures make the first two possible. Liveness is another matter since the database administrator could "fork" the database into several copies, keeping some of your friends ignorant of your latest updates and you ignorant of theirs. In joint work with David Mazieres and some great students, we have worked out how to achieve these properties for file systems. This talk presents a design for database systems that integrates these goals with query processing, concurrency control, and recovery.



Speaker's Profile: Dennis Shasha is a professor of computer science at the Courant Institute of New York University where he works with biologists on pattern discovery for microarrays, combinatorial design, and network inference; with physicists, musicians, and financial people on algorithms for time series; and on database applications in untrusted environments. Other areas of interest include database tuning as well as tree and graph matching. Because he likes to type, he has written five books of puzzles, a biography about great computer scientists, and technical books about database tuning, biological pattern recognition and time series. He has co-authored fifty journal papers, sixty conference papers, and seven patents. For fun, he writes the puzzle column for Scientific American. Until July of 2007, he is at INRIA, Rocquencourt (near Paris, France) with the group of Philippe Pucheral.

COMAD 2006 KEYNOTE TALK

Taming the Dynamics of Distributed Data

Krithi Ramamritham
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay, India
krithi@cse.iitb.ac.in
http://www.cse.iitb.ac.in/~krithi

ABSTRACT

Data gathered from (wireless) sensor networks and those delivered or streamed today via the internet reflect rapid and unpredictable changes in the world around us. Clearly, the Quality of Service needs for such delivery are much more stringent than for static data. This talk will examine the nature of dynamics of distributed data, study the suitability of the current infrastructure for disseminating time varying information, and discuss fresh approaches to maintain the temporal coherency of dynamic data and of queries over such data. We argue that executing user queries over dynamic data calls for the careful design of techniques for change dissemination, dynamic and cooperative caching, in-network filtering and aggregated query processing. We show how exploiting the characteristics of data and the correctness requirements associated with query results leads to effective and efficient solutions that improve scalability and reduce overheads.

Speaker's Profile: Krithi Ramamritham received the Ph.D. in Computer Science from the University of Utah and then joined the University of Massachusetts. He did his B.Tech. in Electrical Engineering and M.Tech. in Computer Science, both from the Indian Institute of Technology Madras. He is currently at the Indian Institute of Technology Bombay as the Vijay and Sita Vashee Chair Professor in the Department of Computer Science and Engineering.

His areas of interest include database systems, real-time systems and internet computing. He has co-authored two IEEE tutorial texts on real-time systems, a text on advances in database transaction processing, and a text on scheduling in real-time systems. He is an Editor-in-Chief of Springer's Real-Time Systems Journal. His other editorial board contributions include IEEE Transactions on Mobile Computing, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Parallel and Distributed Systems, IEEE Internet Computing, the WWW Journal, the Distributed and Parallel Databases journal, and the VLDB Journal. Prof. Ramamritham is a Fellow of the IEEE, a Fellow of the ACM, and a Fellow of the Indian National Academy of Engineering. He is a recipient of the Distinguished Alumnus Award from IIT Madras.