
Volume 31, issue 2, 5 February 2008 ISSN 0140-3664
ELSEVIER
computer
communications
Available online at
ScienceDirect
www.sciencedirect.com
Special Issue:
Foundation of Peer-to-Peer Computing
Guest Editors: Javed I. Khan
Adam Wierzbicki

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Computer Communications 31 (2008) 187–189
computer communications
www.elsevier.com/locate/comcom

Introduction

Guest editors' introduction: Foundation of peer-to-peer computing

Peer-to-peer (P2P) computing has quite conspicuously emerged as one of the most innovation rich areas in computer networking. Millions of users now participate in these systems and the user bases are spreading like wildfire. It is perhaps the most significant development in computing since the web. What is interesting about P2P is that although it emerged out of user community but it is increasingly finding its base on rich foundation of computing theory and research. Various aspects of peer-to-peer systems are now being modeled, formalized, and even engineered based on distributed hashing, complex search models, self-organization, complex networking, and graph theories. It has also become a breeding ground of multifaceted technical innovations. Researchers are delving into new territories such as rich overlay networking, replication and caching, publish/subscribe routing, distributed multicast, range query, etc. The first phase of success of peer-to-peer computing is marked by bold deployment, experimentation with radical concepts, and warm embracement by users. Any technology, for its subsequent levels of maturation crucially requires analytic framework surrounding it. Can the design principles, architectural choices, developed techniques proposed and used on contemporary peer-to-peer computing be understood, validated and improved by formal analytical means? With this goal in mind, this special issue has been conceived to collect recent advancements in the area with analytical footing.

The current P2P paradigm has fundamental relationship with computer *communication and networking*. But it also sits in the confluence of three traditionally disjoint areas namely *algorithmics*, *complex systems*, and *social sciences* – each with its own significant body of formal knowledge and analytic tools. First applications of such tools in a confluence area requires substantial amount of innovative thinking. Readers will find the masterful application of these tools by the authors in this compilation.

The response to the call for paper was overwhelming. The special issue has received 102 submissions from researchers from 27 countries. This represents a worldwide spread of serious research interest in the topic. Not to overlook the context – the area lacks patronage from institutional research funding despite its scientific merit and potential for innovation. The papers went through a rather rigorous two stage review process each being sent to 4–7 reviewers. About 200 hundred reviewers participated. The quality of the papers also prompted us to expand the issue into two volumes. Now the special issue highlights 33 selected works. Volume one presents the confluence areas. Volume two presents the core communication and networking area. The common aspect binding these works is that they use innovative analytical means to characterize and validate the principle scientific contributions.

This volume focuses on three confluence areas namely complex systems, social sciences and algorithmics. P2P offers an interesting opportunity to understand and apply the knowledge of complex systems. Peer-to-peer systems are one of the first internet-based applications which have the characteristics of large self-containment systems. Earlier, the observance of power-law properties in various Internet and web graphs has inspired researchers to attempt characterizing various graph expressions of the web and the Internet in terms of complex network properties – such as degree distribution, clustering, diameter, connectedness, etc. [1]. The P2P systems are far more potent. They offer systemic compactness, abundance, and more knowledge and control about the design and architecture. Not only passive observation, the P2P systems let us venture into the deeper understanding of the underlying processes that actually shape what we observe. Besides the classical tools for studying processes, such as Markov theory, stochastic analysis, and fluid flow analysis a body of analytical works available in power-law networking, such as clustering, stability analysis, phase transition analysis, can shed important design guidance for future systems. There are many interesting questions potentially answerable. Some form of hierarchical multi-level organization seem to be a recurrently theme appearing in recent architectures. Does overlapped hierarchical organization offer some fundamental advantage in complex system? Does elimination of free riding really increases the size and inherent vitality of a system or perhaps a parasitic dependence exists between free riders

0140-3664/$ - see front matter
doi:10.1016/j.comcom.2007.10.038

and contributors? These are just few of the fundamental quests about process.

P2P has also led the advent of social computing [2]. It seems to be one of the first computing systems where the architecture is being evolved with directs response to social dynamics. In classical computer engineering architecture and design meant application of technical knowledge from area such as digital electronics, circuit design, signal communication, automata and algorithmics. Now it means almost in the same sense of technical rigor – application of social engineering. The emergence of Napster for 'illegal' file sharing, 'legal' attack on its central component, and its 'architectural' evolution to a query-protocol-based as opposed to a centralized directory based Gnutella like systems vividly manifests the arrival of social engineer in the design of computing systems. Social engineering is now playing increasingly bold role in areas such as fair resource provisioning, security, load-balancing to network topology optimization. *Is sociology turning into a hard technology?* It suddenly seems many of the existing socio theoretical works will come handy to the designers of this new breed of computing systems. It is highly likely that game theory, decision theory, theories of social organization, theories of competition and cooperation and a whole body of knowledge on social mechanics will contribute to the development of next generation computing systems. Not to mention the benefits are two way. The emergence of peer-to-peer type systems for the first time in history provides a structured observation platform to social scientists.

P2P algorithms are primarily involved with large scale search in autonomous systems. Search is a fundamental process in peer-to-peer systems. P2Ps tenet is to maintain minimum centralizations in any form including not having any central directory. In structured architectures the challenge is that nodes can only be organized using one order. Thus, single key search is rather trivial on structured P2P. However, special searches – such as partial search, range query, or multi-dimensional search are the real challenge. Any search – single or multi-key – is a complex problem over unstructured P2P since there is no available explicit order. The special issue presents some ingenuous techniques in the later three non-trivial categories. Below we present a short introduction to the works included in this volume in the above three confluence areas.

Analysis of large and autonomous systems: The power-law has been speculated on various web and internet systems. In this work, Xie, Chen, Vandenberg, and Pan measure Gnutella network and report that even if it were a power-law network earlier, its current topology is possibly evolving out. The authors attempt a more complex two-layer characterization of Gnutella topology. Goal of many P2P social engineering is to discourage free riding. Though systems like Bit-torrent use tit-for-tat principles but it is still observed that some users receive uneven bandwidth even after contributing. In their particularly interesting work Eger and Killat pose a deeper question. Up to what extent it matters? Armed with the Game Theory the work anticipates strategies that can improve system-wide stability and achieve Nash equilibrium even when some individuals act selfishly. Qiu and Sang in their paper delve further into such systemic questions regarding global stability of P2P society using fluid dynamics like analysis. It seems P2P networks goes through three phases during their lifecycle – a growing phase, a stabilizing phase, and a decaying phase. Most performance analysis presented has tacit assumption of stabilizing. However, are networks really enter into stable state or oscillate? P2P systems add one type of robustness via its distributedness. It also adds another type of unreliability due to churn in such large autonomous collection. Wu, Tian, Ng, and Datta in their work attempts formal modeling of a P2P storage using stochastic methods to analyze the time-evolution of peer-to-peer storage systems under churn. It can be used to many purposes such as optimizing bandwidth usage, provisioning for bandwidth spike, improving system capacity.

Social engineering: To date the principal concern of P2P social engineering has been to encourage cooperation over greed and discourage free riding. Karakaya, Korpeoglu, and Ulusoy propose an overlay adaptation scheme which would bring contributing peers closer to each other on the adapted topology and to push the free riders away from the contributors. The scheme seems to be decentralized and does not require any central entity to control and coordinate or maintenance of any global reputation system. Bit-torrent demonstrated the magic of parallel upload download. Creus-Mir, Masanel, and Hervas in their paper provide an analytical foundation to the general situation of such stylized distributed file sharing and show a bandwidth mechanics to encourage cooperation over greed. Palomar, Tapiador, and Castro expose another important avenue of social engineering – security. Enforcing access control and integrity for contents distributed over a P2P system is not trivial. One can easily fake the name of a file and serve a completely different content over today's P2P systems. A P2P system can neither rely on a centralized authority (such as PKI) for basic security services such as authentication and integrity. In this paper, Palomar, Tapiador, Castro, tackle this hard problem of distributed security and attempts an interesting scheme which can help an author to specify access control policy and ensure authenticity even when the document is served by others in the community. The scheme is founded on cooperation found in a community – where at least a fraction is expected to work honestly.

Algorithmics: Search and query: For structured P2P-hashing has been embraced as a corner stone which distributes the search key-space uniformly and randomly over the peers. This essentially leaves out partial keyword search, wildcard search and range query. Schutt, Schintke, and Reinefeld abandon hashing for a key-order preserving function and demonstrates CHORD# which not only supports the above search types but also preserves the logarithmic performance. They also show a multi-dimensional generalization of the scheme. Complex search turns into a set of blind search even on structured P2P. Shim, Gelman, Vladmir Vishnevsky, Safonov, and Yakimov presents a recursive partitioning scheme for blind search on structured P2P which can avoid message duplication thus making blind searches scalable over structured P2P. How to improve search on an unstructured system? Chen, Zhang, Chen, and Shi paper presents a technique where instead of one large flooding multiple restrained phases are used. Each flooding is controlled by a 'ticket' which provides guidance for subsequent phases to expand. Shi, Han, Liu, Meng, and Yu present another way of improving search – the scheme uses reinforcement learning to exploit mass peer behavior. From expected number of returned results it gradually learns the content distribution. Banaei-Kashani and Shahabi's approach poses any query into unstructured P2P search as a partial search query with a specification of desired completeness. Using percolation analysis – similar to the one found in epidemic dissemination they show how to tune the spreading and get the answer rather correctly and efficiently. To support advanced query over unstructured P2P, various interesting tree data structures has been tried – such as B-Trees, AVL, Trie, etc. Tran and Nguyen present another interesting tree called zig-zag hierarchy for various advanced searches with low overhead including similarity search (k-nearest neighbor) and range query. Gonzalez Beltran, Milligan, and Sage present another innovative data structure called Skip Tree Graph suitable for exact match and range query. Content-based search is one of the most complex forms of search and any P2P system is yet to tackle it. Mueller, Boykin, Roychowdhury, and Sarshar take on this challenge and investigate the problem of content-based-image retrieval (CBIR) on P2P.

Systems: Finally in this volume we also include two papers on P2P middleware. An important signature of a new paradigm is the existence of a distinguishable abstraction of its core. Nothing captures the existence of such a fundamental core than an application development framework. Its design encapsulates the core architecture of a paradigm. If many applications can be built by repeat use of the framework that demonstrates the versatility and value of the paradigm. We present two such development frameworks – one for unstructured P2P is by Walkerdine, Hughes, Rayson, Simms, Gilleade, Mariani, and Sommerville, and another for structured P2P by Shudo, Tanaka, and Sekiguchi. Both demonstrate that how a smorgasbord of P2P applications – to name a few – instant messenger, video distributed video encoder, digital library, distributed virtual world, etc. can be built on top.

References

[1] Newman, Mark, The Structure and Function of Complex Networks, SIAM Review 45 (2003) 167–256.

[2] Douglas Schuler (Ed.), Social computing special edition of the Communications of the ACM, vol. 37, issue no. 1, January 1994, pp. 28–108.

Javed I. Khan
Networking and Media Communications Laboratories,
Department of Computer Science,
Kent State University, USA
E-mail address: javed@kent.edu

Adam Wierzbicki
Polish-Japanese Institute of Information Technology,
ul. Koszykowa 86, 02-008 Warsaw, Poland
E-mail address: adamw@pjwstk.edu.pl

Available online 20 November 2007