Report on the First IEEE International Workshop on Networking Meets Databases (NetDB'05)

Cyrus Shahabi, Ramesh Govindan   and   Karl Aberer
University of Southern California   EPFL
Department of Computer Science   Computer & Communication Science
Los Angeles, CA 90089-0781   Bâtiment BC, Station 14
U.S.A.   CH-1015 Lausanne, Switzerland
[shahabi , ramesh]@usc.edu   karl.aberer@epfl.ch

In this report, to the best of our ability, we try to summarize the presentations and discussions occurred within the First IEEE International Workshop on Networking Meets Databases (NetDB) which was held in Tokyo Japan on April 8th and 9th, 2005. NetDB was one of the many (11 to be exact) satellite workshops of the IEEE ICDE (International Conference on Data Engineering) 2005 conference. This workshop is part of the very few initiatives in bringing the networking and database communities together. The focus research areas of NetDB 2005 were sensor and peer-to-peer networks.

1. Objectives and Justifications

The objective of NetDB is to bring together researchers in the networking and database communities to debate emerging research directions at the intersection of these two fields. We are witnessing the blurring of the traditional boundaries between these two disciplines, especially in the emerging areas of sensor and peer-to-peer networks. We believe the time is ripe for these two communities to get together and discuss the common interests, share and exchange expertise and results, and appreciate each other's terminologies and contributions. The goal of this year's workshop was to promote discussion of ideas that will influence and foster continued research in the areas of sensor and peer-to-peer networks. The workshop provides a venue for researchers to present new ideas that can significantly impact both communities and perhaps give birth to a new community in the long term.

The main reason we selected peer-to-peer and sensor networks as the focused areas for this year's workshop was due to growing number of research papers in these two areas authored by both the database and networking communities. More importantly, unlike other target applications, the research papers appearing in each community's forums are not always addressing complementary research issues but mostly the same exact problems, sometimes following the same

approaches. Hence, the question was whether it makes sense to have joint forums, such as NetDB, to just consolidate terminologies and encourage closer interactions. And the broader question is if a new area is emerging such as what happened when some of the database, signal processing and computer vision researches interacted and gave rise to the "multimedia" community or whether this is just a temporary trend that will pass.

2. Logistics and Statistics

The workshop was spread into two half-days in the afternoon of April 8th and the morning of the April 9th. We had an impressive program committee with 17 expert members from both areas of networking and databases. They professionally reviewed seventeen papers (at least 3 reviews per submission), out of which nine papers were accepted for presentation in the workshop.

About thirty people registered for the workshop that given the date of the workshop (the last day of the conference), the competition (all the other workshops running in parallel) and the number of accepted papers (nine) was very encouraging and to be honest pleasantly surprising. We think this should partly be attributed to our excellent keynote speaker (Mike Franklin) and to our celebrity panel members (Karl Aberer, Amr El Abbadi, Ramesh Jain, Dimitrios Gunopulos and Wei Hong). Also in the audience, in addition to the chairs and panel members, we had the following PC members: Gustavo Alonso and Ugur Cetintemel. Our special thanks to all of them as well as to the authors, the rest of the PC and the audience for making this first NetDB workshop a success. Finally, our gratitude goes to Prof. Masaru Kitsuregawa for his unlimited support for this workshop.

3. Technical Presentations

The nine paper presentations were distributed into four sessions, two sessions on April 8th and two sessions on April 9th. The papers' topics were very much consistent with the ideas behind the workshop's objectives and interestingly in a very balanced way. Consequently, we evenly distributed the papers between the following four sessions: networking, databases, peer-to-peer networks and sensor networks. The entire program including links to the papers is available at: http://infolab.usc.edu/netdb05/ .

The focus of the papers under the **database session** was mainly on query processing issues. The first paper of this session, by a group of authors from Harvard University, discusses ways to optimize queries for distributed stream-based applications where the network is large in scale and dynamic. The second paper, from Brown University, focuses on evaluating the database join operator in heterogeneous networks with the objective of expediting the output of the result.

The papers under the **networking session** considered the application of database techniques to network measurement and monitoring. In the first paper in this session, Li et al. argue that wide area network monitoring can be naturally supported by a distributed indexing system that enables multidimensional range queries. The second paper in the session pertains to stream processing, and argues that relatively simple information theoretic approaches work well in helping to detect qualitative changes in the Internet traffic streams.

The **peer-to-peer networks session** was dedicated to two papers focused on the data access issues with DHT-based peer-to-peer databases. The first paper, which is a result of collaboration between researchers from two German universities (TU Ilmenau and UniMagdeburg), promotes query approximation as a solution for he problem of uncertain data availability in dynamic peer-to-peer databases. Particularly, this paper defines approximation semantics for aggregate queries and similarity queries, and introduces two query evaluations methods to answer such approximate queries. The second paper, from EPFL, considers the problem of non-uniform data distribution in peer-to-peer data networks, and inspired by small-world models proposes a generalization for the DHT family that preserves the efficiency of DHT-based data access even under skewed data distribution scenarios.

Finally, the **sensor networks session** examined the application of database techniques to networked sensing. The energy and storage constraints in this domain give rise to novel approaches. Two of the papers consider the energy efficient processing of historical queries; Coman et al. analytically examine the efficacy of a variety of query routing strategies, while Zeinalipour-Yazti et al. discuss the query processing opportunities and storage management challenges presented by low-power sensing platforms having several megabytes of storage. The other paper in this session argues that sensor network data must be named by its *provenance*--- the origin of the data and the history of operations on it. Such an approach leads to novel data management and query processing issues, and the paper delineates them well.

4. Panel Discussion and Keynote Talk

The panel was the last event of the first day of the workshop in order to keep the audience around until the last minute! The topic of the panel, *"Networking Meets Databases: Do we meet or merge?"* was one of the main questions we wanted to discuss in this workshop. To be specific, whether it is sufficient for these two communities to have occasional joint conferences such as NetDB or a new community should emerge addressing the common challenges?

This is both a technical question, whether technical problems exist that require joint research efforts by both communities, and a cultural question, in how far the existing communities with their different habits, terminologies and backgrounds are fit to interact.

Panel statements were given by Amr el Abbadi, Wei Hong, Dimitrios Gunopolous and Ramesh Jain. All four panelists were of the opinion that sensor networks are one of the driving forces of bringing the communities closer together. Wei Hong claimed that convergence is already here and that it is an important development to study important questions such as layering vs. cross-layer optimization, declarativeness vs. expressiveness and performance vs. predictability. Dimitrios Gunopolous emphasized the role of P2P as a powerful model for developing infrastructure-less Internet-Scale systems. For example P2P could overcome many problems of current search engines. Interdisciplinary research is the way to go there, but the fields have to keep their core strength. Amr el Abbadi pointed out that content-based addressing is already a common concern. However, he observed that conferences of each community have their own idiosyncrasies, which prevent cross-fertilization. Ramesh Jain elaborated on his vision of a document and an event Web. This development will generate many new research problems which in his opinion are neither solely addressable by the networking nor database communities.

In the subsequent discussion, the issue of publishing across communities was hotly debated. It became clear that there are cultural problems to be overcome, but also that the trend is towards opening, as shown by many concrete examples of cross-fertilization and cross-publishing.

The second day of the workshop started with the keynote talk (again to bring the audience to the workshop from the first minute!) delivered by Mike Franklin. In his talk, Mike defined the concept of *high fan-in architectures* for large scale applications where vast numbers of events measured at the edges of the network are continually refined, summarized, augmented, and aggregated as they flow towards the interior. Subsequently, he discussed the key characteristics and challenges presented by high fan-in systems, and argued for a uniform, query-based approach towards addressing them.

Some of discussions following Mike's talk were more on the differences between peer-to-peer and sensor networks. Some in the audience believed that there are more funding opportunities for sensor network projects while projects in peer-to-peer networks usually get labeled with illegal music/video sharing and hence has a negative spin associated with them. Otherwise, it seemed that the audience agrees that both applications share some common underlying challenges.

5. Where we go from here?

Obviously, many of the questions raised by this first workshop have not been addressed. Most importantly, it is not clear whether a new community is emerging or not. However, one clear conclusion is that the workshop should continue because of its huge success in its first year. The importance of the overlapping areas of networking and databases is very clear and the research in this overlapping area is healthy, active and well-funded.

Encouraged by the success of the first workshop, we have created a steering committee for NetDB constituting of: Hari Balakrishnan (MIT), Michael Franklin (UC Berkeley), Ramesh Govindan (USC) and Cyrus Shahabi (USC). Obviously, the steering committee consists of two networking and two database researchers. In addition, following the first workshop, the 2nd NetDB workshop will be held with ICDE 2006 at Atlanta, GA. It also seems to become a tradition that the workshop be co-chaired by a networking and a database researcher. Hence, NetDB'06 is co-chaired by Ugur Cetintemel and John Jannotti from Brown University. The NetDB'06 website is at:

http://www.cs.brown.edu/research/db/netdb06/