Report on the Fourth International Workshop on Data Management for Sensor Networks (DMSN 2007)

Magdalena Balazinska
Univ. of Washington
magda@cs.washington.edu

Amol Deshpande
Univ. of Maryland
amol@cs.umd.edu

Alexandros Labrinidis
Univ. of Pittsburgh
labrinid@cs.pitt.edu

Qiong Luo
Hong Kong Univ. of Sci. & Tech.
luo@cse.ust.hk

Samuel Madden
MIT
madden@csail.mit.edu

Jun Yang
Duke Univ.
junyang@cs.duke.edu

Introduction

Sensor networks enable an unprecedented level of access to the physical world, and hold tremendous potential to revolutionize many application domains. Research on sensor networks spans many areas of computer science, and there are now major conferences, e.g., *IPSN* and *SenSys*, devoted to sensor networks. However, there is no focused forum for discussion of early and innovative work on data management in sensor networks. The *International Workshop on Data Management for Sensor Networks (DMSN)*, inaugurated in 2004, aims to fill this significant gap in the database and sensor network communities.

DMSN 2007, the fourth DMSN workshop, was held on September 24, 2007 in conjunction with the *33rd International Conference on Very Large Database (VLDB 2004)* in Vienna, Austria. Building on the successes of the three previous DMSN workshops, DMSN 2007 aimed at bringing together researchers working on all aspects of sensor data management: from data processing in networks of remote, wireless, resource-constrained sensors to managing heterogeneous, noisy, and sometimes sensitive sensor data in databases. The resource-constrained, lossy, noisy, distributed, and remote nature of sensor networks means that traditional database techniques often cannot be applied without significant re-tooling. Challenges associated with acquiring and processing large-scale, heterogeneous sets of live sensor data also call for novel data management techniques. Finally, in many applications, collecting sensor data raises important privacy and security concerns that require new protection and anonymization techniques.

As the field of sensor networks continues to develop, we have expanded the scope of DMSN 2007 from previous workshops in the series, by encouraging contributions on a broader set of topics, including: database languages for sensor tasking; distributed sensor data storage and indexing; data replication and consistency in noisy and lossy environments; energy-efficient data acquisition and dissemination; in-network query processing; networking support for data processing; query optimization and deployment planning in sensor networks; database techniques for managing loss, uncertainty, noise, and ambiguity; model-based sensor data processing; challenges and techniques for new types of sensor data, e.g., RFID, images and videos, data from scientific and medical instruments; personal, ubiquitous applications of sensor-based infrastructures; integration of sensor data of different modalities and from different sources; integration of sensor data in traditional databases and streaming systems; techniques for secure sensor data collection and processing; and privacy protection techniques for sensor data.

Program

The workshop program includes a keynote address, three research sessions with a total of seven papers, and a panel discussion on the present and future of sensor data management research. As a response to the Call for Papers this year, we received 15 full paper submissions. During the review process, each paper was reviewed by three or four members of the program committee or external reviewers and was also carefully discussed, resulting in the acceptance of seven papers.

Keynote Address The keynote address was delivered by Prof. Gustavo Alonso from ETH Zurich, with the title: "Myths and Realities of Sensor Network Data Management." After about a decade of research on sensor networks, there is a growing body of expertise and experience in real deployments, and a number of companies have begun to commercialize sensor networks. It is now the high time that we reexamine our research efforts in perspective. Gustavo started his talk by questioning the assumptions about wireless sensor

networks that are commonly made by the research community, and gave the audience a healthy dose of reality check: Over and over again, real deployments show that some of the claimed properties of sensor networks (easy deployment, low cost, lack of alternative solutions, self organization, large-scale monitoring) are still not attainable in fields today. Gustavo then proposed a number of more realistic, but nevertheless ambitious, targets for research in sensor networks. First, existing programming tools for sensor networks still require considerable expertise and provide little support for reliability; for sensor networks to make a real impact, we must provide tools for real users instead of novelties for nerds, and develop turnkey solutions that address all aspects of an application, including production, deployment, maintenance, and reuse, as well as data collection, storage, cleansing, and analysis. Gustavo also argued for an approach to system design that is driven by real applications and deployments. There exists no one-size-fits-all solution; instead, we need to evaluate alternative designs and techniques in the context of each application scenario's requirements and constraints. Gustavo's call for tackling real, non-tutored, long-term deployments of sensor networks and linking them to the rest of the IT infrastructure was met with much support from audience.

Research Sessions The seven accepted papers are organized into three sessions: 1) in-network processing, 2) novel sensing modalities, and 3) modeling and programming sensor networks.

The first session features three papers on in-network processing, an idea that has been proven effective in reducing the cost of communicating data to base stations. The paper presented by Shili Xiang (National Univ. of Singapore) shows how to optimize a large number of queries across multiple base stations attached to the same sensor network, exploiting similarities among queries allocated to the same base station during in-network processing. The paper by Yongxuan Lai (Renmin Univ. of China) et al. proposes an in-network data-centric storage scheme that is dynamically balanced to avoid hotpots caused by data skewness, as well as to reduce the cost of data reporting and storage. Finally, the paper presented by Demetrios Zeinalipour-Yazti (Open Univ. of Cyprus) investigates how a swarm of moving sensor nodes can collaboratively compute a perimeter; data is acquired from the perimeter and aggregated and replicated by nodes inside the perimeter for reliability.

The second session features two papers that investigate novel sensing modalities that are much more complex than numeric readings such as temperature and light. Josh Hyman (UCLA) presented work on modeling the rate of CO_2 uptake of drought-tolerant moss using a series of images taken of the plant; the work shows that regression tree models based on color features of the images taken by inexpensive image sensors can effectively predict CO_2 uptake, which is very difficult to measure directly. The second paper, presented by Saket Navlakha (Univ. of Maryland), studies vehicle tracking using video streams produced by traffic cameras; their novel approach, which represents image feature as a graph capturing neighboring relationships among vehicles, makes tracking feasible even with low-quality video streams.

In the third session on modeling and programming sensor networks, Pablo Guerrero (TU Darmstadt) made a case for using workflows to express the application logic of a wireless sensor and actuator network. This approach is better at capturing the actuation aspect than previously proposed programming paradigms, which focused primarily on sensor data collection. The second paper, presented by Yanif Ahmad (Brown Univ.), proposes a framework called *Pulse* for processing continuous queries over continuous-time data models, appropriate for many physical phenomena monitored by sensor networks; optimizations possible under Pulse can lead to significant performance savings compared with the traditional strategy of processing sensor data as streams of discrete tuples.

Panel Discussion The workshop concluded with a roundtable discussion featuring four panelists: Minos Garofalakis (Yahoo! Research and UC Berkeley), Zachary Ives (Univ. of Pennsylvania), Samuel Madden (MIT), and Sunil Prabhakar (Purdue Univ.). All panelists answered a resounding "no" to the question posed in title of the panel: "Sensor Data Management: Are We Done?"

Sam expressed optimism on the impact of database research on sensor network applications, but suggested that the DMSN community should reach out further—not only to the larger sensor networking and the more general distributed systems communities, but also to domain-specific communities with sensor applications. Echoing Gustavo's point in the keynote address, Sam cautioned against unrealistically simple simulations as well as overly ambitious deployment projects based on unproven technology.

The database community today is still mostly wedded to exact answers. Minos, Sunil, and Zack all identified support for approximation, uncertainty, and statistical methods as an important challenge. The inherently noisy and incomplete nature of sensor data makes sensor data management an ideal application domain for emerging probabilistic and uncertain databases.

Minos also highlighted distributed stream processing as an area with many open challenges motivated by sensor data management, including distribution of data and processing, and support for complex "queries" such as probabilistic inference.

Zack likened in-network sensor data processing to the old problem of distributed query processing, but with "extreme" degrees of distribution and constraints that make the problem more interesting. He also pointed out an analogous connection between sensor data fusion and the data integration problem.

Past and Future

The series of DMSN workshops has proved very successful at its goal of providing a forum for database researchers to present their innovative ideas on sensor data management. Over the last four years, we have had a collection of great papers on topics ranging from in-network storage and querying to building middleware systems for processing sensor data. We have also been fortunate to have had excellent keynotes and invited talks - Dr. Wei Hong from Intel Research (2004), Dr. Henry Tirri from Nokia Research (2005), Prof. Mike Franklin from UC Berkeley (2006), and Prof. Gustavo Alonso from ETH Zurich (2007). The themes of the workshop over the years mirror the development of sensor data management field itself. In 2004 and 2005, the workshops focused more on understanding the needs of sensor network applications and the role of sensor data management, as evidenced by the talks by Wei Hong and Henry Tirri. In the last two years, however, the focus shifted to higher-level issues such as data streams, uncertain data management, and infrastructures for large-scale data management and sharing.

Research in sensor data management is at a crossroads. There has been much progress in solving the problems, but many hard research challenges remain open, as highlighted by the participants of this year's panel. We see an ongoing need for a forum like DMSN. The interest in the workshop remains high, as witnessed by the consistently high registration numbers and paper submissions from all over the world. Hence, we plan to continue to organize DMSN workshops in the near future.

Acknowledgements

Many people have contributed to the success of DMSN 2007. We would like to thank all authors who have submitted papers, as well as members of the program committee and the external reviewers; without them, it would have been impossible to put together a high-quality program for DMSN 2007. We would like to thank the Conference Management Toolkit (CMT) Support Team at Microsoft Research for making CMT available to us and Intel Corporation for their generous support of the workshop. Finally, we would like to thank the local conference organization chair, Bernhard Schandl, and the VLDB 2007 workshop organizing committee—in particular Georg Gottlob, H. V. Jagadish, and Krithi Ramamritham—for their help.