06431 Abstracts Collection
Scalable Data Management in Evolving Networks
— Dagstuhl Seminar —

Stefan Böttcher[1], Le Gruenwald[2], Pedro Jose Marrón[3] and Evaggelia Pitoura[4]

[1] Universität Paderborn, DE
[2] University of Oklahoma - Norman, US
ggruenwald@ou.edu
[3] Universität Stuttgart, DE
[4] University of Ioannina, GR

Abstract. From 22.10.06 to 27.10.06, the Dagstuhl Seminar 06431 "Scalable Data Management in Evolving Networks" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Data management, mobile ad-hoc networks, p2p systems, sensor networks

06431 Summary – Scalable Data Management in Evolving Networks

Network technologies like mobile ad-hoc networks, peer-to-peer systems, and sensor networks are getting increasingly popular, but come up with new challenges to scalable data management because data management techniques developed for fixed-wired networks that usually depend on reliable and stable networks can not directly be used.

The goal of this Dagstuhl Seminar was to identify common challenges in these popular network technologies problems, and to get a better understanding of the generality and applicability of solutions to the major problems in scalable data management in these areas.

Keywords: Peer-to-peer, mobile ad-hoc, and sensor networks, atomic transactions, P2P databases

Joint work of: Böttcher, Stefan; Gruenwald, Le; Marron, Pedro; Pitoura, Evaggelia

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/947

06431 Working Group Report on Managing and Integrating Data in P2P Databases

Angela Bonifati (ICAR - CNR - Rende, I)

In this report, to our best recollection, we provide a summary of the working group "Managing and Integrating Data in P2P Databases" of the Dagstuhl Seminar nr. 6431 on "Scalable Data Management in Evolving Neworks", held on October 23–27 in Dagstuhl (Germany).

Keywords: P2P database, data integration

Joint work of: Boncz, Peter A. ; Bonifati, Angela; Illarramendi, Arantza; Janacik, Peter; Konig-Ries, Birgitta; Lehner, Wolfgang; Marrón, Pedro J.; May, Wolfgang; Ouksel, Aris; Römer, Kay; Sapkota, Brahmananda; Sattler, Kai-Uwe; Schweppe, Heinz; Steinmetz, Rita; Türker, Can

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/950

06431 Working Groupe Summary:Atomicity in Mobile Networks

Sebastian Obermeier (Universität Paderborn, D)

We introduce different mobile network applications and show to which degree the concept of database transactions is required within the applications. We show properties of transaction processing and explain which properties are important for each of the mobile applications. Furthermore, we discuss open questions regarding transaction processing in mobile networks and identify open problems for further research.

Keywords: Mobile ad noc networks, mobile databases, mobile transactions, atomicity

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/952

06431 Working Group Summary:P2P, Ad Hoc and Sensor Networks - All the Different or All the Same?

Peter Janacik (Universität Paderborn, D)

Currently, data management technologies are in the process of finding their way into evolving networks, i.e. P2P, ad hoc and wireless sensor networks. We examine the properties, differences and commonalities of the different types of evolving networks, in order to enable the development of adequate technologies suiting their characteristics. We start with presenting definitions for the different network types, before arranging them in a network hierarchy, to gain a clear view of the area. Then, we analyze and compare the example applications for each of the types using different design dimensions. Based on this work, we finally present a comparison of P2P, ad hoc and wireless sensor networks.

Keywords: P2P, ad hoc, wireless sensor networks, database systems

Joint work of: Boncz, Peter A.; Bonifati, Angela; Böse, Joos-Hendrik; Böttcher, Stefan; Chrysanthis, Panos Kypros; Gruenwald, Le; Illarramendi, Arantza; Janacik, Peter; König-Ries, Birgitta; May, Wolfgang; Mondal, Anirban; Obermeier, Sebastian; Ouksel, Aris; Samaras, George; Sapkota, Brahmananda; Steinmetz, Rita; Viglas, Stratis D.

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2007/951

P2P XQuery and the StreetTiVo application

Peter A. Boncz (CWI - Amsterdam, NL)

MonetDB/XQuery* is a fully functional publicly available XML DBMS that has been extended with distributed and P2P data management functionality.

Our (minimal) XQuery language extension XRPC adds the concept of RPC to XQuery, and we outlined our approach to include the services offered by diverse P2P network structures (such as DHTs), in a way that avoids any further intrusion in the XQuery language and semantics.

We also discussed the StreetTiVo application were XRPC is being used for data management in a large P2P environment.

Keywords: Distributed XQuery, P2P, DHT

Joint work of: Boncz Peter A.; Zhang, Ying

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/949

Managing and Integrating Data in P2P Databases

Angela Bonifati (ICAR - CNR - Rende, I)

In this talk I address two main challenges that concern the design and usage of P2P databases. Although such databases recall federated and distributed databases, they open up new fundamental challenges. One of the challenges is the databases design, leading to define a fragmentation strategy for data to be distributed on P2P networks. This is part of our project *XP2P*, XPath for P2P, that also define a query evaluation strategy on the distributed fragments. A second challenge is the integration of heterogeneous data over P2P networks. *HePToX*, standing for Heterogeneous Peers Talk, is a schema matching and querying tool, particularly customized for heterogeneous XML data over P2P networks.

Atomic Transaction Processing in Mobile Ad-Hoc Networks

Joos-Hendrik Böse (FU Berlin, D)

Atomic transaction processing in mobile ad-hoc networks requires for new strategies to overcome nodes failures and minimize blocking situations of mobile nodes. An effective way to increase availability of transaction state information is to distribute such information in controlled way within the MANET allowing recovering nodes to access this information at a high availability. Such redundancy cannot be applied to the coordinator's states as this would introduce additional agreement problems.

In this talk a raw overview on approaches to minimize blocking times due to node or communication failures is given. Failure of participants is addressed by controlled dissemination of the coordinators decision, while failures of coordination nodes is approached by choosing a few reliable backup coordinators to reduce coordination among coordinators as far as possible. The main goal of the presented approaches is to provide for probabilistic transactional atomicity guarantees in MANETs.

Keywords: Atomicity, MANETs, probabilistic guarantees

Secure Multi-Party Data Management

Stefan Böttcher (Universität Paderborn, D)

XML is widely used as data exchange format between different applications, however its use in evolving networks is currently limited and requires solving a variety of new challenging problems ranging from limited energy to limited bandwidth to unpredictable failures to new security and privacy requirements. We present some of the key challenges in these areas and sketch first steps towards a solution. For example, we discuss XML-based access control and privacy violation detection, principles of sovereign information sharing among malicious partners, and cooperative caching for XML clients in mobile ad-hoc networks.

Keywords: Information security, privacy, and access control, sovereign information sharing, XML and XPath

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/948

Power-Aware Data Management and the Third Law of Compu-Dynamics

Panos Kypros Chrysanthis (University of Pittsburgh, USA)

In the last 50 years since the very first electronic computer, we have witnessed many hardware and software evolutions but the shape of computing has always been defined by disruptive technologies.

The best known example of such a disruptive technology from the 80s is the PC.

The advent of mobile and tiny computing devices is the most recent disruptive technology that has impacted every aspect of life from communication, to business, to education, to health, to entertainment. It has also caused a paradigm shift in designing and developing small-sized computing systems, in general and data management protocols, in particular.

In this new paradigm, power or energy consumption has become the new governing law for data management in addition to the two traditional laws of time and space complexities.

In this talk we will focus on this new, third law of Compu-Dynamics, which led to Power-Aware Data Management. We will classify algorithms, protocols and techniques from mobile and sensor databases domains with respect to power-awareness and network awareness, the latter of which is the major source of energy consumption in wireless networks.

Keywords: Mobile Data Management, Sensor Data management, Network-Aware Data Management

Advanced Data Management Technologies Lab @ U. Pittsburgh

Panos Kypros Chrysanthis (University of Pittsburgh, USA)

The Advanced Data Management Technologies (ADMT) Lab in the Computer Science Department of the University of Pittsburgh focuses on user-centric data management for scalable network-centric applications. This talk will provide an overview of the current projects which cover different aspects of the data cycle in evolving network environments: data acquisition with sensors, data stream processing, web data management and data dissemination to stationary and mobile users.

Keywords: Mobile Data Management, Sensor Databases, Stream Data Management, Web & Real-time Data Management, Biological Data Management

DISTRIBUTED (DATA) MANAGEMENT FOR MOVING OBJECTS, SENSORS, AND INDIVIDUALS

Alex Delis (National and Capodistrian University of Athens, GR)

In this talk, I briefly discuss the current situation of aggregated and emerging networks and outline opportunities that may exist for answering spatio-temporal queries for moving objects/users. Deploying such query facilities in large scale geographic areas necessitates the synergy of multiple independent management

systems that have to collaborate in a transparent manner to fast yield results to spatio-temporal queries.

I will also outline the use of such infrastructure in answering route similarity searches, handling of streaming data to to travelers along a route with similar interests as well as derivation of common characteristics for the movement of points/users in different time periods.

Data Management in Mobile Ad-Hoc and Sensor Network Databases

Le Gruenwald (US National Science Foundation - Arlington, USA)

For data-intensive applications to operate efficiently and effectively in mobile ad-hoc networks and sensor networks, it is necessary to develop data management techniques that take into consideration the specific characteristics of each network environment to improve data availability. This talk presents two such techniques: a data replication scheme for real-time mobile ad-hoc network databases, called DREAM, and a data estimation technique for sensor network databases, called WARM.

DREAM provides a data replication solution that addresses the issues of network partitioning, frequent disconnection, aging of temporal data, timing constraints of transactions as well as battery limitation of mobile nodes. DREAM considers the applications' semantics obtained through various data and transaction types and the remaining energy of nodes to improve data accessibility while reducing overall energy consumption as well as balancing energy consumption among nodes. Experimental results comparing DREAM with other existing replication techniques show that DREAM yields the highest number of transactions that can be executed successfully before their deadlines expired as well as the most balance of nodes' energy consumption.

WARM's purpose is to estimate the values of the sensor data that are missing because they were lost or corrupted or arrived late when they were sent from sensors to servers. WARM employs association rule mining to discover related sensors in a sliding window. It then computes the missing reading of a sensor using the available readings of all sensors that are related to that sensor. To shorten the time required for association rule mining, WARM includes special data structures to store the information that is necessary for the discovery of the relationships among sensors. Experimental results comparing WARM with existing statistical methods using a real-life traffic monitoring application show that WARM yields better data estimation accuracy while having reasonable time and space overheads.

Keywords: Mobile ad-hoc databases, sensor network databases, data replication, data estimation

Joint work of: Gruenwald, Le; Padmanabhan, Prasanna; Halatchev, Mihail

Data Management in Mobile Computing

Arantza Illarramendi (Universidad del País Vasco - Donostia, E)

The attractiveness of computing services that are available anywhere and anytime has given rise to an intense research and development of mobile computing applications. In this abstract I summarize the main features of some data services developed by our research group such as a Locker Rental Service, a Software Retrieval Service, a Location-Dependent Query Processor and finally a Tele- Assistance Service (AINGERU).

Locker Rental Service [VIP 02]: This service incorporates mechanisms that allow mobile users to rent storage space, called lockers, situated in a proxy. The main features of the service are: Data stored in a data locker are available even when the mobile device is disconnected. Specific tasks can be carried out at the locker in the proxy on behalf of the mobile user. The architecture is based on mobile agents. The lockers stay always close to the location of the user (lockers can travel).

Software Retrieval Service [MIRG 06]: This service allows users to select and retrieve software in an easy, efficient and adaptive way. It is based on the use of an ontology and the agent technology. The main features of the service are: Easy: with the help of knowledge-driven agents, users can browse locally the ontology that describes semantically software elements. Efficient: taking advantage of the capability of mobile agents to deal with disconnections and to move to other computers, agents optimize the use of the wireless media. Adaptive: agents take into account the network status and past user actions.

Location-Dependent Query Processor [IMI 06]: This service supports distributed processing of continuous location-dependant queries. Its architecture is based on mobile agents. The main features of the service are: Not only the users issuing queries but also other interesting objects can move; and it provides an efficient performance of the queries in a continuous way.

Tele- Assistance Service (AINGERU) [TIBBG 04]: AINGERU is an agent based data system for a new way of tele-assistance for elderly people. In addition to the functionalities provided by current tele-assistance services it also offers: Vital sign monitoring by using sensors to capture the values of those signs and feed a decision-support system that analyses them and generates an alarm when necessary. Active Assistance by using intelligent agents that can act in the face of anomalous situations without direct intervention of the user. Universal assistance, i.e. irrespective of time or place through the use of wireless communications and PDAs.

References

[VIP 02] Y. Villate, A. Illarramendi, E. Pitoura Keep your Data Safe and Available While Roaming Mobile Networks & Applications (MONET) V.7, N.4, 315-328, 2002

[MIRG 06] E.Mena, A. Illarramendi, J.A. Royo, A. Goñi A Software Retrieval Service based on Adaptive Knowledge-Driven Agents for Wireless Environments

ACM Transactions on Autonomous and Adaptive Systems (TAAS) V.1, 67-90, 2006

[IMI 06] S. Illarri, E.Mena, A. Illarramendi Location-Dependent Queries in Mobile Contexts: Distributed Processing Using Mobile Agents IEEE Transactions on Mobile Computing V.5, N. 8, 1029-1043, 2006

[TIBBG 04] A. Tablado, A. Illarramendi, M. Bagues, J. Bermúdez, A. Goñi Aingeru: an Innovating System for Tele Assistance of Elderly People The Journal on Information Technology in Healthcare V.2, N.3, 205-214, 2004

Data Management in Mobile Wireless Sensor Networks

Peter Janacik (Universität Paderborn, D)

Mobile wireless sensor networks enable a wide range of novel application areas, such as habitat monitoring, scientific exploration, medicine and surveillance. Most applications from these areas strongly rely on an efficient data management, which comprises functions such as efficient storage, reduction, fusion, retrieval, replacement and maintenance of sensor data, as well as, strategies to provide adequate reactions to changes in the node topology. For illustration, consider a typical example from the scientific exploration application area: a Mars sensor network consisting of a large number of sensor nodes, in which some nodes, the Mars rovers, are mobile while others are just stationary. During their operation, the nodes collect measurements and pass them to the network's data management, which is responsible for their storage and organization in the network (among the other functions identified above). In order to analyze the surface of the planet, the nodes further have to combine their locally measured data with data already present in the network, which serves as a large database. The goals of data management in the context of this scenario are, for example, to reduce the energy consumed during the operations, to extend the lifetime of the data, to maximize the amount of information stored, to increase the rate of query matches or to reduce the amount of overhead incurred by topology changes.

Keywords: Data Management, Mobile Wireless Sensor Networks

Service Orientation in Mobile Computing

Birgitta König-Ries (Universität Jena, D)

Service Orientation in Mobile Computing - An Overview

Service-orientation is widely regarded as a promising paradigm to enable resource sharing among loosely coupled systems. In this talk, we argue that this makes service-orientation also a very suitable paradigm to support mobile applications. After giving some motivation on why this might be the case, we

identify extensions to "standard" service orientation that are needed to fully support all the needs of mobile applications.

Mobile devices are inherently less powerful than their stationary counterparts; they lack processing power, periphery and/or network connections. On the other hand, users of mobile devices accept more and more "desktop like" behavior from their device. To overcome the gap between its own capabilities and the user's expectations, a mobile device might use functionality provided by others. A particularly elegant way to ensure this functionality sharing is through a service-oriented architecture.

Mobile environments have, however, a number of characteristics, that make it hard to use standard service-orientation, e.g. in the form of web services, alone. Rather, more advanced approaches are needed: Mobile devices are used in continously changing environments. Therefore, it is impossible to use static binding to external services. A service that is there and suitable today may no longer be available or suitable tomorrow. We therefore need the possibility to dynamically (and automatically, no user will be willing to manually ensure correct service binding) bind services at run-time. To achieve this goal, we need semantic, machine-understandable descriptions of offered services, semantic, preference-encoding descriptions of requested services, means to compare offers and requests, means to compose several offers to fulfill complex requests and means for efficient execution of complex services.

The talk gives an overview of existing approaches in this area as well as presenting our own work.

The material presented is mostly based on:

Mohamed Hamdy, Birgitta König-Ries: Service-Orientation in Mobile Computing - An Overview. In: Proceedings of the 2006 TAMC workshop (held at MDM 2006).

Robstness and Error Propagation Model in Data Stream Environments

Wolfgang Lehner (TU Dresden, D)

Research Issues in Data Streaming

The basic idea of data streaming consists in providing an infrastructure where permanently incoming data (e.g. from sensors) is accepted, transformed, joined, and given to potentially multiple consumers. These consumers register standing queries and accept periodically events from the data stream system which are compliant to their query specification. From an overall perspective, data streaming provides a huge potential of research facilities. In this context, we focus on two issues, namely robustness and error propagation in (distributed) streaming environments.

Robustness in Data Stream Processing

In classical database systems, incoming queries are delayed, if the system is busy - the users do notice a potential overload by increasing query response

time. In data stream environments, the system only has 'one look' at the data item; otherwise the streaming data item is gone. Therefore, we have to provide a mechanism which (a) is able to accept incoming data items and (b) provide a quality of service guarantee to the consumer. Within our current research prototype QStream, we developed a data stream management system on top of a real-time operating system so that we are able to compute an optimal schedule for a given query load once a new query is registered within the system. A new query is therefore only accepted if the system is able to produce a feasible schedule. Obviously, the computation is based on multiple assumptions subsumed by a parameterized cost model. In order to avoid re-computations of a schedule, the system is based on a jitter-constraint periodic stream model which is able to trade memory (for internal buffers) for quality of service. However, if the changes in the environment are so huge (e.g. changes in the selectivity or number of rows produced by a time-based window on the data stream), we trigger a re-computation of the schedule. A good overall compromise is therefore needed incorporating the internal resource requirements and the number of re-computations of the schedule. Within our current research activities, we are proposing a mathematical model and a prototypical implementation for QStream to permanently monitor and automatically adjust the necessary parameters.

Error Propagation Model

Since data stream management systems are usually build to analyse sensor data and combine intermediate results with static data using statistically complex transformations and forecast methods, a system has to provide a notion of 'reliability' of the final result. Within our second research project in the area of data stream management systems, we design an error propagation model, which is annotates the raw data with additional information about the statistical validity of the specific data item. For example, we defined window-based measures to compute a notion of completeness, correctness, and confidence intervals and gave transformation rules for the individual operators, e.g. the impact of a selection operation on the confidence interval.

Summary and Relationship to Dagstuhl Seminar

We defined a model for robustness und error propagation in the context of general and higher-level data stream management and received valuable feedback from the community. With regard to the Dagstuhl seminar, we are interested in exchanging our ideas with the sensor network community: are similar issues relevant for sensor networks; how can we combine the notion of robustness with the optimality criteria in that context (e.g. to save energy)? How can we implement a resource-sensitive mechanism of error-propagation in sensor networks? These questions will hopefully spark interesting discussions.

Sensor Network Group Activities

Pedro Jose Marrón (Universität Stuttgart, D)

In this talk I gave a very brief overview of the set of current activities of the

sensor network group that I lead as part of the distributed systems department of the University of Stuttgart. The talk describes the work of the TinyCubus system as well as a series of system algorithms, data management algorithms and system properties addressed as part of the research in the group.

Keywords: Sensor network, research activities

Middleware Approaches for Wireless Sensor Networks

Pedro Jose Marrón (Universität Stuttgart, D)

Wireless Sensor Networks are a canonical example of a wider field dealing with Cooperating Objects that attempts to create the necessary technologies to make Weiser's vision of the disappearing computer a reality. Cooperating Objects are, in the most general case, small computing devices equipped with wireless communication capabilities that are able to cooperate and organise themselves autonomously into networks to achieve a common task.

In this presentation, Wireless Sensor Networks are presented as an enabling technology for Pervasive Computing. We then classify the different approaches used for the design and implementation of middleware solutions existing in the literature: "classic", data-centric, virtual machines and adaptive middlewares. The presentation concludes with a comparison of approaches and the formulation of a series of open problems that should be tackled in the near future.

Keywords: Middleware, Wireless Sensor Networks

Distributed Processing of Active Rules over Heterogeneous Component Languages in the Semantic Web

Wolfgang May (Universität Göttingen, D)

The talk gives a short overview of an ECA Framework for active rules over heterogeneous component languages for the Semantic Web. Rules can use different languages in their components. These languages are identified by URIs that are in turn associated with appropriate processors. When processing ECA rules, the ECA language processor determines the appropriate service and initiates the data exchange (using an agreed XML format - except this there is no restriction on the participating services). Language services, as well as domain application services and domain brokers can join and leave the network.

An Economic Incentive Model for Efficient Dynamic Replication in Mobile-P2P networks

Anirban Mondal (University of Tokyo, J)

In mobile ad-hoc peer-to-peer (M-P2P) networks, frequent network partitioning leads to typically low data availability, thereby making data replication a necessity. This work proposes EcoRep, a novel economic incentive model for dynamic replica allocation in M-P2P networks. EcoRep performs replica allocation based on a data item's relative importance, which is quantified by the data item's price in terms of a virtual currency. The price of a data item depends on its access frequency, the number of users who accessed it, the number of its existing replicas, its (replica) consistency and the average response time required for accessing it. EcoRep ensures fair replica allocation by considering the origin of queries for data items. EcoRep requires a query issuing user to pay the price of his requested data item to the user serving his request. This discourages free-riding and encourages user participation by providing an incentive for users to become service-providers. Moreover, the architecture of EcoRep includes broker mobile peers, which facilitate indexing of the data items in the network. EcoRep also considers other issues such as load, energy and network topology as replication criteria. Our performance study indicates that EcoRep is indeed effective in improving query response times and data availability in M-P2P networks.

Keywords: Mobile, Peer-to-peer, Incentive model, Economic model, Dynamic replica allocation, free-riding

Transaction Processing in Mobile Ad-Hoc Networks

Sebastian Obermeier (Universität Paderborn, D)

Whenever database applications that make use of transactions should be transferred to mobile networks, message costs are high and network failures, like network partitioning or node failures, make global knowledge concerning the operational status of devices difficult or even impossible.

Therefore, within MANETs, there are some interesting new challenges, among which my research focuses on the following:

* Which kind of atomic guarantees can be given for distributed transactions?

* Which requirements must be fulfilled by mobile atomic commit protocols regarding compensation, transaction models, and blocking time?

* What are scenarios worth to simulate?

In my talk, I will show a possible solution outline and give experimental results for one scenario.

Keywords: Transactions, mobile networks

Scalable Data Management in Mobile Environments: Protocol design in MANETs and data management in mobile sensors

Aris Ouksel (University of Illinois at Chicago, USA)

My presentation addressed with two areas of my current research interests: (1) Protocol design in Mobile Ad-hoc Networks, and (2) Sensor data management. It describes briefly the motivating challenges and our approach to dealing with them.

Keywords: MANETs, mobile sensor networks, economic models, information discovery and dissemination, localization, indexing, query processing

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/953

Full Paper:
http://www.uic.edu/∼aris

Usable Sensor Networks

Kay Römer (ETH Zürich, CH)

Being deeply embedded into the physical world, sensor networks are very dynamic and uncertain distributed systems. Unfortunately, sensor networks offer only very limited resources to deal with this complexity. Hence, today only skilled experts can program, deploy and use these systems. In this brief talk, we present concepts and tools to make sensor networks more accessible to non-expert users. We focus on three phases on the sensor network lifecycle: implementation, deployment, and actual use of the sensor network.

Regarding implementation, we present a programming abstraction called "generic role assignment", which supports the automatic assignment of functions to sensor nodes, such that properties of a sensor node (e.g., remaining energy, network neighbors) match the requirements of the assigned function. Essentially, sensor nodes take on certain "roles" in the network.

With generic role assignment, roles and rules for their assignment can be easily specified using a declarative configuration.

We briefly discuss such a role specification language and distributed algorithms for role assignment according to such specifications.

Recent experience with the deployment of sensor networks demonstrates that it is far from trivial to setup a working larger-scale sensor network in the field. Even though simulations and experiments with lab testbeds suggest a working system, subtle real-world influences lead to frequent failures in the field. Identifying and fixing these problems is currently a difficult and cumbersome task due to the lack of appropriate concepts and tools. We classified common problems that had been encountered during deployment and could show that many

of these problems can be detected by overhearing and analyzing sensor network traffic without need for an instrumentation of sensor nodes.

Based on this observation, we develop a tool called "SNIF" to inspect a deployed sensor network, consisting of a distributed network sniffer and a data-stream-based framework for online traffic analysis.

Finally, we are investiating ways to use sensor networks to enable users to search the real world. Such a "Google for the Real World" should allow users to find answers to questions such as "which meeting room is currently not occupied?" or "which cafeteria currently has the shortes queue?"

Personalization for the Wireless user

George Samaras (University of Cyprus, CY)

Mobile clients present a new and more demanding breed of users. Solutions provided for the desktop users are often found inadequate to support this new breed of users. Personalization is such a solution. The moving user differs from the desktop user in that his handheld device is truly personal. It roams with the user and allows him access to info and services at any given time from anywhere. As the moving user is not bound to a fixed place and to a given time period, factors such as time and current experience becomes increasingly important for him. His context and preferences are now a function of time and experience and the goal of personalization is to match the local services to this time-depended preferences. In this paper we exploit the importance of time and experience in personalization for the moving user and present a system that anticipates and compensates the time-dependant shifting of user interests. A prototype system is implemented and our initial evaluation results indicate performance improvements over traditional personalization schemes that range up to 173

Keywords: Personalization, time, experiance, mobile user, wireless user

Semantic P2P Overlay Networks

Brahmananda Sapkota (Nat. University of Ireland - Galway, IRL)

In this paper, a Semantic approach for creating a semantic P2P Peer-to-Peer overlay network is proposed. Overlay networks are created at application layer, i.e., on top of transport layer, to form a structured virtual network topology such that a deterministic search can be guaranteed. The overlay network facilitate message routing to the locations not specified by IP (Internet Protocol) addresses. The peers in the overlay network are organized in such a way that peers having semantically similar data form a semantic cluster. Peers can participate in more than one cluster forming a coordination cluster and facilitating communication between clusters.

Keywords: Peer-to-Peer, Semantic Routing, Semantic Similarity, Semantic Routing

Joint work of: Sapkota, Brahmananda; Nazir, Sanaullah; Toma, Ioma; Hauswirth, Manfred

A Universal Storage Approach to Distributed Data Management

Kai-Uwe Sattler (TU Ilmenau, D)

Many new applications, for example Wikis, social networks, and distributed recommender systems, require the efficient integration of decentralized and heterogenous data sources at a large scale. In the talk, we present our vision of a universal storage for RDF-like triple data based on a structured P2P system and the universal relation model as its key enabling technologies to achieve flexibility, robustness, and efficiency for large-scale distributed data storage and query processing.

We outline results of our work on similarity-based query operators exploiting the features of the underlying DHT infrastructure as well as on large-scale distributed processing of complex query plans.

Transactions on replicated data in mobile Ad Hoc nets - and some further issues of large scale networks

Heinz F. Schweppe (FU Berlin, D)

In the Coco/Da project1 data centric middleware functions for Mobile Ad Hoc Nets have been investigated and developed. Transactional support and data replication are important functions of traditional data management systems. Using a volatile communication infrastructure, replication plays an even more important role than in fixed wired systems in order to increase availability of data. Keeping replica consistent, however, is more involved with unreliable communication.

Keywords: Transaction, mobile ad hoc net, replication, uncertainty in evolving networks

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2007/954

Queryable DTD-aware compression of XML data streams

Rita Steinmetz (Universität Paderborn, D)

Whenever data has to be distributed in a network, all nodes of the network might benefit from disseminating the data in a compressed form in order to save communication costs.

But often, if the compression algorithm does not allow to evaluate queries on the compressed data, those saves in communication costs are bought for the price of higher computation costs.

Therefore, there is a need for compression algorithms that allow to evaluate queries on the compressed data and that are able to compress and decompress partial data, i.e., that are able to compress and decompress infinte data streams.

In my talk, I presented an overview on such a compression system for XML data and I showed some evaluation results concerning the compression ratio of our system. Finally, I presented an overview of the architecture of the future system: a modular system, that allows to design a custom-made compression system according to the user's needs.

Keywords: XML, data streams, compression

Joint work of: Steinmetz, Rita; Böttcher, Stefan

Distributed Data Management

Stratis D. Viglas (University of Edinburgh, GB)

In this talk I will present my understanding of the problems involved in building a scalable data manager in an evolving network. I will first briefly present my own contributions in the area, and how they can be tied to the issues addressed by this seminar. I will then present a list of open problems, approaches to solutions of which will hopefully be among the outcomes of the seminar.

The key issues of scalable, decentralised data management, can be better described as the problems arising when one of the basic assumptions of traditional data management is dropped: a local hard disk drive is not there to reliably store and maintain data. The data therefore resides on the network and one has to revisit all the problems associated with storage, retrieval, and processing of network-bound data. The differences to traditional data management are numerous: (i) there may not be any notion of fixed location; rather, identifying the location of a particular data item becomes an issue of querying the entire network. (ii) There is no notion of a multi-pass algorithm; rather, data is seen and acted upon only once. (iii) Transaction processing and fault tolerance has to deal not only with hardware failures, but also with connectivity and/or network failures. These dimensions move data processing to completely new realms and ask of the researchers of the community to merge ideas from other disciplines with traditional data processing ones. Identifying a common framework to address all these issues in transparent and scalable ways will be, at least according to my opinion, the extensive research agenda of the next few years.

Keywords: Data management, decentralization, distribution