Information Storage, Reduction and Dissemination in Sensor Networks: A Survey

Q. Le-Trung, A. Taherkordi, T. Skeie, H. N. Pham
Dept. of Informatics, Univ. of Oslo, Oslo, Norway
E-mail: *{quanle, amirhost, hainp, tskeie}@ifi.uio.no*

P. E. Engelstad
Telenor R&I, Oslo, Norway
E-mail: *Paal.Engelstad@telenor.com*

Abstract— **There is a large body of work addressing information storage, reduction, and dissemination in Wireless Sensor Networks (WSNs), but there seems to be no contribution giving a survey of this literature. In this paper we take on such a challenge and review previous WSN research and further presents a framework that we use for classifying the different proposed approaches and techniques. The framework is structured into 3 tiers: 1) routing and network infrastructure, 2) logical network infrastructure, and 3) information storage and processing. The paper, however, mainly focuses on tier-2 and tier-3. We also discuss future visions and challenges within the WSN community.**

Keywords- Storage; dissemination; sensor; aggregation

I. INTRODUCTION

Wireless sensor networking has been an active research topic for many years in both academia and industry [1]-[3]. *In the first generation* of WSNs, most research concentrates on developing routing protocols and network architectures for WSNs, which are self-organized, self-adapted, and energy-efficient [2]. Sensor readings are often collected by sensors and disseminated to the monitoring sink, normally through the pre-defined data retrieval patterns. Except for the design of MAC energy-efficient protocols [4], the energy efficiency is mostly achieved by the integration of in-network data aggregation [3] techniques into the data dissemination architectures [5].

In the second generation, a WSN is considered as a distributed data-centric database [6]-[7], and users interact with this database via queries. This approach solves the entire problem of service definition and interface to WSNs by mandating SQL queries as the interface. The problems here are in finding energy-efficient ways of executing such queries and of defining proper query languages that can express the full richness of WSN. Fig. 1 shows an integrated view of a WSN database related to two phases: (1) data updates to the database via the data dissemination and aggregation techniques, and (2) data retrievals from the database triggered by the sink(s) through query dissemination. Additionally, Fig. 1 also shows that different WSN applications can pose different data storage models and data/query dissemination architectures, requiring different metrics such as bandwidth, latency, and energy-efficiency on routing protocol and network infrastructure. *In the third generation*, techniques and mechanisms for distributed data-centric storage are extended to accommodate multi-resolution storage [22], multi-query optimization and aggregation [29], as well as multi-demensional range queries [36]. Recently, *in the fourth generation* of WSNs, a new generation NAND flash memory has dramatically improved the capacities and energy efficiency of local flash storage. It is now possible to equip sensor devices with several gigabytes of low-power flash storage, allowing storage to be cheaper than communication and comparable in cost to computation. By exploiting flash-based in-network data storage, a novel database architecture emphasizing local data archival and query processing at sensors is introduced [1].

There is a lot of research on different generations of WSNs and within different areas. However, to the best of our knowledge, there is no extensive survey that reviews all these research areas systematically. In this paper, we classify and link them into an integration framework, consisting of three tiers: 1) routing and network infrastructure, 2) logical network infrastructure, and (3) information[1] storage and processing. Due to the space limitations, we concentrate on tier-2 (not included information reduction techniques), and tier-3 in this paper, together with future visions on WSNs.

This paper is structured as follows. In Section II, different logical network infrastructures for information dissemination are presented. Alternative approaches for information storage and processing are described in Section III. Finally, Section IV shows conclusions and discusses future visions on WSNs.

II. LOGICAL NETWORK INFRASTRUCTURE

Different logical network infrastructures are shown in this Section, see tier-2 in Fig. 2. Typically, logical network infrastructures exploit the hierarchical architectures for the scalability, and the heterogeneity of sensors, e.g., rich-resourced sensors as cluster heads, to perform the architecture construction, maintenance, as well as the data reduction. In the tier-1, the physical network architecture can be flat or hierarchical. The tier-1 and the information reduction techniques in the tier-2 is out of scope of this paper. Further details can be found in [2]-[3].

A. Data Dissemination Architectures

Most architectures for data dissemination in WSNs are hierarchical, which can be classified as *cluster-based* [8]-[9], *chain-based* [10]-[11], *grid-based [12]*, or *tree-based* [13].

[1] In the scope of this paper, information can be sensor readings, i.e., data, or query, or both, depending on specific context.

This work is supported by the SWISNET project, carried out in Networks & Dist. Systems group, Dept. of Informatics, Univ. of Oslo, Norway.

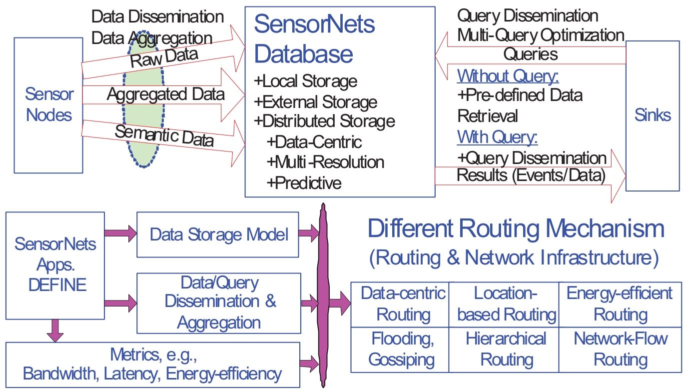

Figure 1. Information storage, routing, and query processing in WSNs.

In energy constrained large-scale WSNs, it is inefficient for sensors to transmit data directly to the sink, especially if there are long distances and multihop routes between source sensors and the sink. In such scenarios, sensors can transmit data to a local data concentrator, i.e., the cluster head, which reduces data from all the sensors in its cluster by one of the information reduction techniques [3], and transmits the results to the sink. This will reduce significantly energy usage of sensors. Multiple cluster-based architectures have been developed for data dissemination such as LEACH [8] and HEED [9].

In cluster-based architecture, if the cluster head is far away from the sensors, it might spend excessive energy on communication. Further improvements in energy efficiency can be obtained if sensors transmit only to its close neighbors. The key idea behind the *chain-based* architecture is that each sensor transmits only to its closest neighbor. In the greedy approach, PEGASIS [10] proposes a chain construction algorithm that starts with the farthest node from the sink. This node is the head of the chain. At each step, the non-chain node which is closest to the chain head is selected and appended to the chain as the new head. The procedure is repeated until all nodes are in the chain. Simulation results in [10] show that PEGASIS outperforms LEACH [8] by 100%-200% in terms of the number of data gathering rounds for different network sizes. An improvement of PEGASIS is presented in [11]. It proposes a minimum total energy algorithm which constructs a chain with minimum $\sum d^2$, where d is the distance between two adjacent nodes in the chain. Simulation results in [11] shows that the minimum total energy algorithm achieves 15%-30% performance improvement compared to the greedy algorithm.

In the *grid-based* architecture, a set of sensors is assigned as the data concentrators in fixed regions of the WSNs. The sensors in a particular grid transmit the data directly to the data concentrator of that grid. A two-tier data dissemination (TTDD) approach to address the multiple, mobile sink problem is presented in [12]. TTDD provides scalable and efficient data delivery to multiple mobile sinks. Each data source in TTDD proactively builds a grid structure which enables mobile sinks to continuously receive data on the move by flooding queries within a local cell only. TTDD exploits the fact that sensor nodes are stationary and location-aware to construct and maintain the grid architecture with low overhead.

In a *tree-based* architecture, sensor nodes are organized into a tree where data reduction is performed at intermediate nodes along the tree and a concise representation of the data is transmitted to the root node, which is usually the sink. Tree-based architecture is suitable for applications which involve in-network data aggregation [3]. The work in [13] proposes heuristics to construct and maintain an aggregation and dissemination tree in sensor networks. The key idea of these heuristics is that sensors with higher residual power have a higher chance to become a non-leaf tree node, i.e., branch point of the tree. To maintain the tree, a residual power threshold is associated with each sensor. When the residual power of a sensor falls below this threshold, it periodically broadcasts help messages within a fixed duration and shut down its radio. Upon receiving a help message, a child node swiches to a new parent.

B. *Query Dissemination Architectures*

This Section is a complementary of Section II.A, in which queries are sent by sink(s) to request data from sensors. Thus, we also call *data dissemination architecture* as *push-arch*, while *query dissemination architecture* as *pull-arch*. In this Section, we discuss *TAG-based tree* [5], its extensions for the fault-tolerance [14]-[16] such as *approximation* and *synopsis diffusion*. New architectures are also introduced, including *tributary-delta* [17], and *sweep* [18], as alternative approaches to construct architectures for query, and/or data dissemination.

Most common approaches for query dissemination in WSNs are tree-based, and TAG [5] is the most popular. It allows users to express simple, declarative queries, and ensurea that they are distributed and executed efficiently in resource-limited WSNs. TAG processes aggregates in the network by computing and altering the data as it flows through the sensors, discarding irrelevant data, and combining relevant readings into more compact records whenever possible. TAG operates as follows: A user poses aggregation queries from a powered, storage-rich basestation. Operators that implement queries are distributed into the network by piggybacking on the existing ad-hoc routing protocols [2]. Sensors route data back towards the users through a routing tree rooted at the basestation. As data flows up this tree, it is aggregated according to an aggregation function and value-based partitioning specified in the query. However, aggregating along a tree is very susceptible to node and transmission failures, which are common in WSNs. Such failures can result in the loss of entire sub-trees of sensor readings, introducing significant error in the query answer. Efforts to reduce losses by retransmitting packets wastes significant energy and delays query responses. Thus, solutions based on multi-path routing are also proposed in TAG. For aggregates such as MAX/MIN, which are monotonic and exemplary, this provides a fault-tolerant solution. Duplicate-sensitive aggregates like COUNT and AVG, on the contrary, gives incorrect results when the same value is counted multiple times.

Synopsis diffusion [14] is a general framework for achieving significantly more accurate and reliable answers by combining energy-efficient multi-path routing schemes with techniques that avoid double-counting. These techniques use the order- and duplicate-insensitive (ODI) synopses that compactly summarize intermediate results during in-network aggregation. ODI enables the system to adapt message routing to dynamic message loss conditions, even in the presence of asymmetric links. *Approximation* [15] is another approach. In

[15] well-known duplicate-insensitive sketches is extended to handle SUM aggregates. It presents a method to combine duplicate-insensitive sketches with multi-path routing techniques to produce accurate approximations with low communication and computational overhead.

Tributary-Delta [17] is a new approach to in-network aggregation that combines the advantages of both the tree-based architecture and the multi-path routing. Under low loss rates, trees are used for their low or zero approximation error and their short message size. Under higher loss rate or when transmitting partial results accumulated from many sensor readings, multi-path is used for its robustness. Results in [17] show that the tree approach is more accurate than the multi-path approach at very low loss rates, because of its lower approximation error (0% versus 12%). However, at loss rates above 0.05%, trees are much worse than multi-path routing because of their high communication error. On the other hand, Tributary-Delta provides not just the best of both (e.g., from running either tree or multi-path in the entire network), but provides in fact a significant error reduction over the best, across a wide range of loss rate. Recently, a new architecture is introduced and applied for query/data dissemination in WSNs, called *sweep* [18], i.e., a wavefront that traverses the network, passes over each node exactly once, and performs the desired operation(s). The network sweep is implemented by a narrow band of active nodes that moves over the network by issuing invitations to new nodes to join the band, and dropping nodes that have already been processed and serve no other essential purpose. A key feature of sweep that adds to its robustness is that even though tree structures are used for partial aggregation with the sweep active band, these trees are local and are used almost as soon as they are formed. Aggregation paths are deviced on-line as the sweep proceeds and no fixed tree structure is needed over the course of the computation. Thus, this approach is robust to both link volatility and node failures.

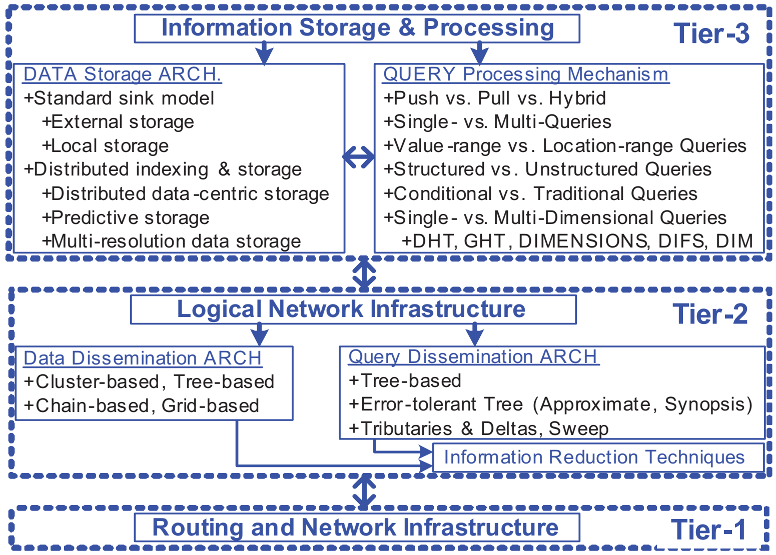

Figure 2. A layer classification of information storage, reduction, dissemination in WSNs.

III. Information Storage and Processing

This Section reviews and classifies research directions in the second and third generation of WSNs, as specified in Section I. These sub-classes include *data storage architecture*, and *query processing mechasnim*. The tier-3 in Fig. 2 illustrates this classification in more details. Note that the query processing mechanism is actuallly a component of data storage architecture, but it is classified as a separate branch due to its richness in styles and attributes.

A. Data Storage Architectures

In the data storage approach, a WSN is considered as a database, and users access to this database through queries. It can be classified as *standard sink model*, and *distributed indexing and storage* [19].

The standard sink model normalizes the research directions in the first generation of WSNs, see Section I, to the database approach, in which data collected by sensors is stored either locally at source sensors, called *local storage*, or externally at the sink(s)/basestation(s), called *external storage*. The former is not the research focus in the first generation of WSNs due to the limited resources of sensors, including limited energy and storage capacity. However, the local storage is triggered recently in the fourth generation of WSNs by the appearance of a new generation NAND flash memory [1]. In opposition to the local storage, *external storage* is typically assumed for the reseach on the first generation of WSNs. Upon detection of events, the relevant data is sent to external storage where it can be further processed as needed. Thus, there is no cost of external queries since the event information is already external.

Distributed indexing and storage uses in-network storage, builds distributed indexes and stores partial aggregates to facilitate user queries. For related work in this field of the second and third generation of WSNs, we classify it as *distributed data-centric storage* [6]-[7], [20], *multi-resolution data storage* [22], and *predictive storage* [23]. In the *distributed data-centric storage* relevant data is stored, by name, at nodes within WSNs, i.e., all data with the same general name will be stored at the same sensor nodes. Queries for data with a particular name can then be sent directly to the node storing that named data, thereby avoiding the flooding. This storage archirecture is based on the use of GPSR [21] for low-level routing, and a distributed hash-table (DHT) built on top of GPSR. Distributed data-centric storage is suggested to be preferable for large-scale WSNs [21], where there are many detected events and not all event types are queried.

However, *distributed data-centric storage* requires that queries and associated processing be clearly defined a priori. When details about queries are not known a priori, such architecture is inappropriate, leading to the *multi-resolution data storage* [22]. This architecture provides storage and search for raw sensor data in data-intensive scientific applications. The key idea behind this storage architecture is spatio-temporal summarization: muti-resolution summaries of sensor data are constructed and stored in the network in a spatially and hierarchically decomposed distributed storage structure optimized for efficient quetying. This storage model is lossy, and progressively degrading, where multi-resolution summarization uses wavelets, and where queries on data are posed in a drill-down manner. This means that they are first processed on coarse, highly compressed summaries corresponding to larger spatio-temporal volumes, and the approximate result obtained is used to focus on regions in the networks that are most likely to contain result set data. Recently, a new trend in low-cost low-energy NAND flash memory makes it feasible to build cost-effective sensors with more than a gigabyte of flash memory [1], which leads to the *predictive storage*. The work in [23] proposes PRESTO, a predictive storage architecture for emerging large-scale, hierarchical WSNs. It is a proxy-centric architecture, where tethered proxies balance the need for interactive querying from users with the energy optimization needs of the remote sensors. PRESTO exploits the presence of a large local store to reduce communication by storing data locally at remote sensors whenever possible, and intelligent caching at proxies. The system also addresses user needs for querying such WSNs by exposing a unified, easy to use data abstraction across numerous proxies and remote sensors.

B. Query Processing Mechanisms

1) Query Styles and Attributes

This Section distinguishes among alternative approaches to categorize query styles and attributes.

Query execution is the first category. It can be classified as *push*, *pull*, or *hybrid*. The *push* approach is the default query mechanism for the first generation of WSNs, where all possibly relevant sensor readings are proactively pushed out of the network, e.g., to an external storage, upon the pre-defined query or data retrieval patterns. In the *pull* approach, queries are reactively sent into the network for pulling relevant results out of the network. Finally, in the hybrid *push-pull* approach [24]-[26], sensor readings are proactively pushed to selected nodes in the network, where they are later pulled when queries are issued. Typically, *pull* and *hybrid* approaches are used in the second and third generation of WSNs. *Query structure* is the second category [24]. With *unstructured* queries, the node issuing the query does not know in advance where any copy of the requested information can be found. The query dissemination is therefore a form of blind search, which can take the form of an expanding ring search or a sequential trajectory search. With *structured* queries, a hash or an index, is used so that the querying node knows exactly where the nearest copy of the requested event information can be found (See the *indexing techniques* in Section III.B.3). Thus, there is a trade-off between the energy cost of replicated storage and of querying. The trde-off is determined by the number of replicas created by each event. While the first generation of WSNs normally uses unstructured query mechanisms, both mechanisms are mentioned in the second and third generation of WSNs. *Query plan* is the third category, where *traditional* query plans that use different ordering of query predicates depending on different attributes, are called *conditional* query plan. A conditional query plan is particularly relevant in acquisitional systems where the cost of acquiring some of the attributes is non-negligible and where correlations exist between one or more attributes. As seen in [27], the query evaluation can become cheaper by observing additional attributes, if such additional observations are low-cost and allow the query processor to determine with a high confidence that a query predicate can offer substantial performance gains, while avoiding the cost of acquiring unnecessary and expensive attributes. It also shows how to build *conditional* plans that branch into one or more sub-plans, each with a different ordering for the expensive query predicates, based on the runtime observation of low-cost attributes.

The final category is based on summary data-structure [28]. It considers two kinds of aggregate range queries: *value-range* and *location-range*. The former is the query that asks for the number of sensors that record values in the specified value range, while the latter ask for the sum of values recorded by sensors located in a particular rectangle in the field. In [28] summary data-structures, called linear sketches, are constructed over the sensor data using in-network aggregation. They are used to answer aggregate queries in an approximate manner at the basestation.

2) Multi-Query Processing and Optimization

When multiple queries are posed to the resource-limited WSNs simultaneously, it is critical to process them efficiently by sharing their data acquisition, computation, and communication cost, as well as reducing the bandwidth contention and data loss as a result of transmission collision. However, there are only a few studies on multi-query optimization and aggregation in WSNs [29].

Multi-query processing and optimization is motivated by the share of sensor readings and data transmissions among different queries if the query regions of different queries overlap. Energy then can be efficiently conserved by sharing the partial aggregate results of common regions among different queries. The most general method for the work sharing optimization for multiple queries is to devide the intersected aggregation regions (bounding box) into many separated groups, and calculate the aggregate values for each separated groups using the “GROUP BY” processing technique introduced in TAG [5]. This approach can effectively share the common partial aggregate results among different queries to avoid the redundant data acquisition and transmission load. However, when the number of aggregate queries increases, the number of divided separated groups can grow into a quite huge number, so the total communication cost might become unacceptably high [29].

Algorithms to solve this problem are presented in [30]. A linear reduction apgorithm is first proposed to reduce the number of the separated groups divided from the query bounding boxes, and then two hybrid algorithms are presented to approximate the optimal cost, while avoiding the high computational requirements for linear reduction. Another approach is introduced in [31]. In this approach, the bounding box is considered as a black box, and the aggregate value of the bounding box can be found simply by subtracting the sum of the incoming partial aggregate values from the sum of the outgoing partial aggregate values. This approach can reduce the total communication cost more efficiently than the former solution using separated groups, utilizing the sharing work of partial aggregate values. However, this approach does not consider further partial aggregate merging in network and utilizes the simple routing construction scheme that is not adaptive to the multi-query processing. Two improvements of this approach are presented in [29]. They include an equivalence class based merging algorithm for in-network merging of partial aggregate values of multi-queries, and an adaptive fusion degree based routing scheme as a cross-layer designing technique. A two-tier multiple query optimization (TTMQO) scheme was recently proposed in [32]. It represents further improvements to the work presented above, as it supports more types of queries and data acquisition queries. The first tier of TTMQO, called the basestation optimization, adopts a cost-based approach to rewrite a set of queries into an optimized set that shares the commonality and eliminates the redundancy among the queries in the original set. The optimized queries are then injected into the network. In the second tier, called in-network optimization, it efficiently delivers query results by taking advantage of the broadcast nature of radio channel, sharing the spatio-temporal correlation of sensor readings among similar queries at a finer granularity.

3) *Indexing Techniques for Multi-Dimensional Queries*

The multi-dimentional queries can be solved by using indexing techniques, which essentially trade-off some data insertion cost to enable efficient querying. Most research in this field can be applied for *distributed indexing and storage architecture* of WSNs shown in Section III.A.

Geographic Hash Table (GHT) [33] is the indexing technique used in distributed data-centric storage (DCS) [7], [20]. GHT hashes keys into geographic coordinates, and stores a key-value pair at the sensor node geographically nearest the hash of its key. The system replicates stored data locally to ensure persistence when a node fails. It uses an efficient consistency protocol to ensure that key-value pairs are stored at the appropriate nodes after topological changes, and it distributes load throughout the network using a geographic hierarchy. DIMENSIONS [34], DIFS [35], and DIM [36] can be thought of as using the same set of primitives as GHT (i.e., storage using consistent hashing), but for different ends: DIMENSIONS allows drill-down search for features within a sensor network, DIFS allows range queries on a single key in addition to other operations, and DIM uses a locality-preserving hash to store data, enabling efficient multi-dimensional range queries. In particular, DIMENSIONS takes time-series data as input and compresses it using wavelets while retaining significant features. This compressed data is then stored with the WSNs to produce a multi-resolution map, i.e., a hierarchical architecture. Such maps allows users to drill down into areas that appear to contain significant phenomena. For DIFS, it extends GHT to support efficient range queries while maintaining balanced load across nodes. DIFS achieves this by constructing a multiply rooted hierarchical index in which non-root nodes can have multiple parents. Nodes store event information for a particular range of values detected within a particular geographic region. Higher-level nodes cover smaller value ranges detected within large geographic regions while lower level nodes cover a wider range of values from within a smaller geographic region. The key idea behind the construction of this hierarchy is to incorporate the value of an event, as well the location of the detecting node in determining the storage node for that event occurrence. Using this index, DIFS can efficiently support range queries, queries related to the distributions of values in space and so forth. For DIM, it uses a technique whereby events whose attribute values are close are likely to be stored at the same or nearby nodes. DIM then uses GPSR [21] to routes events and queries to their corresponding nodes in an entirely distributed fashion.

However, any hierarchical architecture has the problem that sensors that hold information corresponding to high levels in the hierarchy become traffic bottlenecks. This overload can be avoided in GHT [33] by using replication of structured information. Another approach taken by DIFS [35] is to partition the information about large geographic regions into subsets according to smaller allowed ranges of the field value, and store these subsets in different nodes. *Fractional cascading* [37] avoids the bottleneck created by higher level nodes in the hierarchy by recursively partitioning the sensor field by a standard quad-tree. Aggregates from each quad in the tree are computed and stored at all sensor nodes in the quad. Each node has the values of itself and aggregates of all the quads in which it resides. This improves data survivability and query efficiency as important information are naturally replicated mode widely. *Hierarchical spatial gossip* [19] can be considered as an alternative way to achieve fractional cascading. Instead of a fixed quad-tree partitioning, it keeps the data summarization hierarchy of each node adaptive and centered on the node itself. Thus, any two nodes will have slightly different world views at each scale, as their multi-resolution ranges differ, while two leaf nodes in a fixed quad-tree may share the same data of many high-level quads.

IV. CONCLUSIONS AND FUTURE VISIONS

To the best of our knowledge, there are no extensive survey linking related research work in the field of information storage, reduction, and dissemination together. To address this shortcoming, this paper reviews and classifies WSN research work into a three-tier framework. Due to space limitations, only tier-2 and tier-3 were discussed. Different alternative approaches and techniques have been described and discussed in each sub-class within each tier. The main contribution is the linking of these sub-classes into a general framework. This is general enough to be applied for any real applications in WSNs by selecting the corresponding architectures and techniques. In this way, different researchers, developers, and implementers can benefit from this paper. As a demonstration, in the SWISNET project [38], we are currently working on a middleware solution for providing adaptivity and (re)-configurability to be used in the numerous dynamic WSN applications. This middleware is placed on top of this integration framework. In fact, the middleware services interact directly with some components within this framework, such as query and data dissemination architectures for the code distribution and data propagation, respectively.

We finally discussed the future visions of WSN research within the field of information storage, reduction, and dissemination. We believe that the below metrics will affect the research trends in WSNs. Firstly, advances in the WSN technologies, such as NAND flash memory [1], will favour the local data storage approach. Thus, WSN research directions focusing on new architectures and protocols for data/query dissemination, as well as new database concepts, will be dominant. Secondly, advances in the wireless communication technologies, e.g., the ultra-wideband radio technology (UWB-RT) [39], will lead to the new design of new MAC protocols for WSNs. This will again affect the new design of architectures and protocols for data/query dissemination in WSNs, where the cost of bandwidth is a constraint. Finally, we believe that techniques and mechanisms for distributed indexing and storage to support for multi-demensional range queries, multi-query optimization and aggregation, as well as multi-resolution storage, for the third generation of WSNs, will also be a fruitful approach. But this direction will consider cross-layer optimization, especially those new characteristics from the MAC and network layers [4].

References

[1] Y. Diao, D. Ganesan, G. Mathur, and P. Shenoy, "Rethinking data management for storage-centric sensor networks," *CIDR2007*, pp. 22-31, Asilomar, CA, USA, Jan.7-10, 2007.

[2] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks," *Elsevier Ad Hoc Network Journal*, Vol. 3/3, pp.325-349, 2005.

[3] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks: a survey," *IEEE Comm. Surveys & Tutorials*, pp.48-63, Fourth Quarter 2006.

[4] V. I. Zadorozhny, P. K. Chrysanthis, and P. Krishnamurthy, "A framework for extending the synergy between MAC layer and query oprimization in sensor networks," *DMSN'04*, pp.68-77, Toronto, Canada, Aug.30 2004.

[5] S. Madden, M. J. Franklin, and J. M. Hellerstein, "TAG: a tiny aggregation service for ad-hoc sensor networks," *ACM SIGOPS Operating Systems Review*, pp.131-146, Vol. 36, Issue SI, 2002.

[6] J. Gehrke and S. Madden, "Query processing in sensor networks," *IEEE Pervasive Computing*, pp.46-55, Jan-Mar.2004.

[7] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: An acquisitional query processing system for sensor networks," *ACM Trans. On Database Systems*, pp.122-173, Vol. 30, Issue 1, Mar.2005.

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Sensor Networks," *Hawaii Intl. Conf. System Sciences*, Hawaii, Jan.2000.

[9] O. Younis, and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks," *IEEE Trans. on Mobile Computing*, Vol. 3, Issue 4, Oct.2004, pp.366–379.

[10] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Trans. on Parallel and Dist. Systems*, pp.924-935, Vol. 13, No. 9, Sept.2002.

[11] K. Du, J. Wu, and D. Zhou, "Chain-based protocols for data broadcasting and gathering in the sensor networks," *IPDPS'03*, pp.260-267, Nice, France, Apr.22-26 2003.

[12] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang, "A two-tier data dissemination model for large-scale wireless sensor networks," *MobiCom'02*, pp.148-159, Geogia, USA, Sept.23-28 2002.

[13] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," *VTC2003-Fall*, Vol. 4, pp.2168-2172, Oct.2003.

[14] S. Nath, P. B. Gibbons, S. Seshan, and Z. R. Anderson, "Synopsis diffusion for robust aggregation in sensor networks," *SenSys'04*, pp.250-262, Maryland, USA, Nov.3-5 2004.

[15] J. Considine, F. Li, G. Kollios, and J. Byers, "Approximate aggregation techniques for sensor networks," *ICDE'04*, pp.449-460, Boston, USA, Mar.30-Apr.02 2004.

[16] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregation information," *FOCS'03*, pp.482-491, Ithaca, NY, USA, Oct.11-14 2003.

[17] A. Manjhi, S. Nath, and P. B. Gibbons, "Tributaries and deltas: efficient and robust aggregation in sensor network streams," *SIGMOD'05*, pp.287-298, Maryland, USA, Jun.14-16 2005.

[18] P. Skraba, Q. Fang, A. Nguyen, and L. Guibas, "Sweeps over wireless sensor networks," *IPSN'06*, pp.143-151, Tennessee, USA, Apr.19-21'06

[19] R. Sarkar, X. Zhu, and J. Gao, "Hierarchical spatial gossip for multi-resolution representations in sensor networks," *IPSN'07*, pp.420-429, Cambrige, Massachusetts, USA, Apr.25-27 2007.

[20] S. Ratnasamy, D. Estrin, R. Govindan, B. Karp, and S. Shenker, "Data-centric storage in sensornets," *ACM SIGCOMM CCR*, pp.137-142, Vol. 33, Issue 1, Jan.2003.

[21] B. Karp, and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," *MobiCom'00,* Boston, MA, Aug.2000.

[22] D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R. Govindan, "Mutiresolution storage and search in sensor networks," *ACM Trans. On Storage*, pp.277-315, Vol. 1, No. 3, Aug.2005.

[23] P. Desnoyer, D. Ganesan, H. Li, M. Li, and P. Shenoy, "PRESTO: a predictive storage architecture for sensor networks," *HotOS'05*, Santa Fe, NM, US, Jun.12-15 2005.

[24] S. Kapadia and B. Krishnamachari, "Comparative analysis of push-pull query strategies for wireless sensor networks," *DCOSS'06, LNCS 4026*, pp.185-201, 2006.

[25] N. Trigoni, Y. Yao, J. Gehrke, and R. Rajaraman, "Hybrid push-pull query processing for sensor networks," *Workshop on Sensor Networks*, Berlin. Germany, Sept.2004.

[26] X. Liu, Q. Huang, and Y. Zhang, "Combs, needles, haystacks: balancing push and pull for discovery in large-scale sensor networks," *SenSys'04*, pp.122-133, Baltimore, Maryland, USA, Nov.3-5 2004.

[27] A. Deshpande, C. Guestrin, W. Hong, and S. Madden, "Exploiting correlated attributes in acquisitional query processing," *ICDE'05*, pp.143-154, Washington, DC, USA, Apr.5-8 2005.

[28] K. Kalpakis, V. Puttagunta, and P. Namjoshi, "Accuracy vs. lifetime: Linear sketches for approximate aggregate range queries in sensor networks," *DIALM-POMC'04*, pp.67-74, Pennsylvania, USA, Oct.1'04.

[29] Lei Xie, L. Chen, S. Lu, Li Xie, and D. Chen, "Energy-efficient multi-query optimization over large-scale sensor networks," *WASA'06, LNCS 4138*, pp.127-139, 2006.

[30] N. Trigoni, Y. Yao, A. Demers, J. Gehrke, and R. Rajaraman, "Multi-query optimization for sensor networks," *DCOSS'05*, pp.307-321, Marina del Rey, CA, USA, Jun.30-Jul.01 2005.

[31] F. Emekci, H. Yu, D. Agrawal, and A. E. Abbadi, "Energy-conscious data aggregation over large-scale sensor networks," *Tech. Report ID: 2003-32*, Dept. of CS, Univ. of California, Santa Barbara, Nov.2003.

[32] S. Xiang, H. B. Lim, K-L. Tan, and Y. Zhou, "Two-tier multiple query optimization for sensor networks," *ICDCS'07*, Toronto, Ontario, Canada, Jun.25-29 2007.

[33] S. Ratnasamy, B. Karp, L. Yin, and F. Yu, "GHT: A geographic hash table for data-centric storage," *WSNA'02*, Atlanta, Georgia, Sept.2002.

[34] D. Ganesan, D. Estrin, and J. Heidemann, "Dimensions: why do we need a new data handling architecture for sensor networks?," *ACM SIGCOMM CCR*, pp.143-148, Vol. 33, Issue 1, Jan.2003.

[35] B. Greenstein, D. Estrin, R. Govindan, S. Ratnasamy, and S. Shenker, "DIFS: A distributed index for features in sensor networks," *In'tl Conf. on Sensor Network Protocols and Applications*, pp.163-173, Los Angeles, CA, USA, May.11 2003.

[36] X. Li, Y. J. Kim, R. Govindan, and W. Hong, "Multi-dimensional range queries in sensor networks," *SenSys'03*, pp.63-75, Los Angeles, CA, USA, Nov.5-7 2003.

[37] J. Gao, L. J. Guibas, J. Hershberger, and L. Zhang, "Fractionally cascaded information in a sensor network," *IPSN'04*, pp.311-319, Berkeley, CA, USA, Apr.26-27 2004.

[38] Networks and Distributed Systems group, IFI/UiO, SWISNET project, http://www.ifi.uio.no/forskning/grupper/nd/projects/swisnet/.

[39] D. Porcino, and W. Hirt, "Ultra-Wideband Radio Technology: Potential and Challenges Ahead," IEEE Communications Magazine, July 2003.