Real Time Tactic for Knowledge Placement & Management in Wireless Sensor Networks

Arabinda Nanda*, Saroj Kumar Rout*, Amiya Kumar Rath †

* Department of Computer Science, Krupajal Engineering College, Bhubaneswar, India
† Department of Computer Science & IT, College of Engineering , Bhubaneswar, India
aru.nanda@rediffmail.com , rout_sarojkumar@yahoo.co.in , amiyaamiya@rediffmail.com

Abstract: Knowledge Placement & Management in wireless sensor networks has been a research area in recent years. Many existing Sensor Knowledge Management Systems view sensor data as a continuous stream that is sensed, filtered, processed, and aggregated as it "flows" from sensors to users. We argue that the Knowledge Management infrastructure will need substantial redesign to fully exploit the presence of local storage and processing capability in order to reduce expensive communication. In this paper we propose an architectural framework for Knowledge Placement and Management. We propose a reliable and real time tactic for Knowledge placement and achieving data integrity using self organized sensor clusters. Instead of storing information in individual cluster heads as suggested in some protocols, in our architecture we suggest storing of information of all clusters within a cell in the corresponding base station and finally we are trying to calculate Sensor Network Knowledgebase Query Processing time and space complexity.

Keywords: Cluster Head, Data Reliability, Knowledge Management, Real Time Communication, Wireless Sensor Networks.

1. Introduction

The fact of knowing about the world is known as knowledge. Data store in a database where as knowledge store in knowledge base. Knowledge is closely related with intelligence. A person having more knowledge is called highly intelligence person. The artificial intelligence area has developed techniques for representing knowledge in forms that can be exploited by computational procedures and heuristics. Database management systems research produced techniques that support the representation and management of large amounts of relatively simple
Knowledge. Knowledge management is concerned with the representation, organization, creation, acquisition, usage, and evolution of knowledge in its many forms. To build successful technologies for knowledge management, we need to further our understanding of how individuals, groups and organizations use knowledge. Given that more and more knowledge is becoming encoded in computer readable forms, we also need to build tools which can effectively search through databases, files, web sites and the like, to extract information, capture its meaning, organize it and make it available. The main goal of our research paper is to development of a Knowledge Placement & Management System (KPMS) to store knowledge. Wireless Sensor Networks (WSNs) provide a new model for sensing and collecting information from various environments, with the possible to serve many and different applications. Current WSNs typically communicate directly with a centralized controller or satellite. On the other hand, a elegant WSN consists of a number of sensors spread across a geographical area; each sensor has wireless communication capability and sufficient intelligence for signal processing and networking of the data. Accessing and processing knowledge and data produced in a WSN using a database-like approach [2]–[3] has several advantages. Sensors can be deployed in the physical environment and applications that manipulate their knowledge and data can be created, refined, and modified afterwards without any physical involvement on the sensors themselves. The data management activity performed in the network can be remotely controlled by interactively issuing queries, expressed in a high level language, which specify what knowledge and data are of interest for a certain task, and how they should be manipulated. In this paper we propose an architectural framework for reliable and real time placement and dissemination of knowledge in WSN. We focus on middleware technology, and describe details of some existing research prototypes, then address challenges and future perspectives on the middleware. This study highlights that middleware needs to provide a common interface for various functional components of WSN: Detection and knowledge collection, signal processing, knowledge and data aggregation, and notification. By integrating sensing, signal processing, and communication functions, a WSN provides a natural platform for hierarchical information processing.

2. Related Work

[4] and [1] are two proposals providing real time data delivery services in sensor networks. They are based on an idealized model that takes into account both the remaining

lifetime before a data report expires and the remaining distance the report has to travel. Given the physical distance to the destination and the lifetime of a packet, a desired velocity can be calculated. If the packet travels along a straight line towards the destination at desired speed it will not miss the deadline. They both use greedy geographical forwarding, but differ in the exact mechanism used. In RAP, the required velocity is updated at each hop to reflect the urgency of the report. A node uses multiple FIFO queues with different priorities to schedule reports of varying degrees of urgency. Each queue accepts reports of velocities within a certain range. RAP further adjusts the waiting and back-off times at MAC layer based on the priority of the report being transmitted, so that one with higher priority has greater probability of accessing the channel. SPEED is an adaptive, location-based real-time routing protocol, which can be effectively used if the location information is available in all sensor nodes and the location updates can be delivered to the source sensors regularly. In SPEED, a node actively controls the data rate to avoid congestion by maintaining a relatively stable relay-speed to each neighbor. The node measures the delay of sending a report to each neighbor using exponential weighted moving average. Given a report with velocity v to be maintained across the sensor network, it computes the speed vi of the report, if neighbor Ni is chosen as the next hop. It chooses one neighbor from the group of neighbors with vi>v to forward the report to the next node. If no such neighbor exists, the report is forwarded to the next node with some probability. Nodes in a congested region also feedback back pressure messages upstream, so that data are detoured to bypass the region.

3. Design Approach

The basic requirements for a knowledge management system mean to build a so called information workspace [4], which can be seen as a special extension of the classical information retrieval systems [5]. The differences can be summarized as follows:

- Easy knowledge management (information work) for both content creators and information consumers (no more effort is required for storing and modifying content than for retrieving information)
- The information retrieval is not restricted to term-based queries
- The collection structure or document hierarchy is not necessarily implied by terms
- There is a possibility for automatic information exchange between the actors (e.g., by agents)

WSN Constraints

- Available bandwidth;
- Processing and memory capabilities of WSN nodes;
- Availability of energy resources;
- High error rates (from 2% to 30% per link).
- Number of nodes (network size and depth).

4. Network Topology

One important property of a WSN is its diameter, that is, the maximum number of hops between any two nodes in the network. In its simplest form, a WSN forms a single hop network, with every sensor node being able to communicate directly with every other node. An infrastructure based network with a single base station forms a star network with a diameter of two. A multi-hop network may form an arbitrary graph, but often an overlay network with a simpler structure is constructed such as a tree or a set of connected stars. The topology affects many network characteristics such as latency, robustness, and capacity. The complexity of data routing and processing also depends on the topology. Given the large number of nodes and their potential placement in difficult locations, it is essential that the network is able to self organize; manual configuration is not feasible.

4.1 Density and Network Size

The effective range of the sensors defines the coverage area of a sensor node. The density of the nodes indicates the degree of coverage of an area of interest by sensor nodes. The network size affects reliability, accuracy, and data processing algorithms. The density can range from a few sensor nodes to a hundred in a region, which can be less than 10m in diameter.

4.2 Connectivity

The communication ranges and physical locations of individual sensor nodes define the connectivity of a network. If there is always a network connection (possibly over multiple hops) between any two nodes, the network is said to be connected. Connectivity is broken if the network may be partitioned occasionally. If nodes are isolated most of the time and enter the communication range of other nodes only occasionally, we say that communication is irregular. Connectivity influences the design of communication protocols and data dissemination mechanisms

4.3 Node Addressability

This indicates whether or not the nodes are individually addressable. Ex: - the sensor nodes in a parking lot network should be individually addressable, so that the locations of all the free spaces can be determined. Thus, it may be necessary to broadcast a message to all the nodes in the network. If one wants to determine the temperature in a corner of a room, then addressability may not be so important. Any node in the given region can respond.

4.4 Data Aggregation

Data aggregation is the task of data summarization while data are traveling through the sensor network. An excessive number of sensor nodes can easily congest the network, flooding it with information. The popular solution to this problem is to aggregate or fuse data within the WSN then transmits an aggregate of the data to the controller.

There are three major ways of performing data aggregation:

1. Diffusion algorithms assume that data are transmitted from one node to the next, thus propagating through the network to the destination. Along the way data may be aggregated, mostly with simple aggregation functions and assuming homogeneous data.
2. Streaming queries are based on SQL extensions for continuous querying. Here data are considered to be transient while the query is persistent.
3. Event graphs work on streams of events and compose simple events into composite events based on event algebra

4.5 Query Capability and Propagation

There are two types of addressing in sensor network; data centric, and address centric. In a data centric paradigm, a query will be sent to specific region in the network, whereas in addressing centric, the query will be sent to an individual node. A user may want to query an individual node or a group of nodes for information collected in the region. Depending on the amount of data fusion performed, it may not be feasible to transmit a large amount of the data across the network. Instead, various local sink nodes will collect data from a given area and create summary messages. A query may be directed to the sink node nearest to the desired location.

4.6 Data Dissemination

The ultimate goal of a WSN is to detect specified events of interest in a sensor field and to deliver them to subscribers. Because of the overlap in the proximity ranges of sensors, the same phenomenon might be recorded by multiple sensor nodes. Alternatively, systematic aggregation might lose all the data on the same phenomenon. End-to-end event transfer schemes that fit the characteristics of WSNs are needed, in the same way that delivery semantics of asynchronous communication, such as publish/subscribe, is needed for wired distributed systems. The electric power consumed depends substantially on how the sensor data is handled and communicated. Because the capacity of the battery of the sensor node is very limited, it is necessary to consider the extent to which the demands of applications can be met. Adaptive communication protocols (Power aware Protocols that consider power consumption, are actively being researched.

4.7 Real-Time Support

Object tracking applications may need to correlate events from different source nodes in real-time. Real-time support (e.g. a physical event must be reported within a certain period of time) may be critical in WSNs. This aspect affects time synchronization algorithms, which may be affected by the network topology and the communication mechanism deployed.

4.8 Reliability

The reliability or fault tolerance of a sensor node is modeled [2] in using the Poisson distribution to capture the probability of not having a failure within the time interval (0, t). The fault tolerance level depends on the application of the WSN.

4.9 Security

Threats to a WSN are described in [3] and classified into the following categories:

- Passive Information Gathering: If communications between sensors or between sensors and intermediate nodes or collection points are in clear, then an intruder with an appropriately powerful receiver and well designed antenna can passively pick off the data stream.
- Subversion of a Node: If a sensor node is captured it may be tampered with, interrogated electronically and perhaps compromised. Once compromised, the sensor node may disclose its cryptographic keying material, and access to higher levels of communication and sensor functionality may be available to the attacker. Secure sensor nodes, therefore, must be designed to be tamper proof and should react to tampering in a fail-complete manner where

cryptographic keys and program memory are erased. Moreover, a secure sensor needs to be designed for its emanations not causing sensitive information to leak from it.

• False Node: An intruder might add a node to a system and feed false data or block the passage of true data. Typically, a false node is a computationally robust device that impersonates a sensor node. While such problems with malicious hosts have been studied in distributed systems, as well as ad-hoc networking, the solutions proposed there (group key agreements, quorums and per-hop authentication) are in general too computationally demanding for sensors.

• Node Malfunction: A node in a WSN may malfunction and generate inaccurate or false data. Moreover, if the node serves as an intermediary, forwarding data on behalf of other nodes, it may drop or garble packets in transit. Detecting and culling these nodes from the WSN becomes an issue.

• Node Outage: If a node serves as an intermediary or collection and aggregation point, what happens if the node stops functioning? The protocols employed by the WSN need to be robust enough to mitigate the effects of outages by providing alternate routes.

• Message Corruption: Attacks against the integrity of a message occur when an intruder inserts itself between the source and destination and modifies the contents of a message.

4.10 Knowledge accumulation and dissemination

In real meaning, management can be said to be based on the gathering and transforming knowledge into marketable products and services. Thus the management of this knowledge collection and transformation is critical to success. A good starting point for the exploration of this area is Wiig's (1995) three volume discourse "Knowledge Management Methods". Of especial interest to us was the first volume describing the way in which people and organizations create, represent, disseminate and use knowledge? Wiig argues that it is essential that the key knowledge domains are identified and incorporated within the organization's activities and structures. Wiig defines this as the concept of Critical Knowledge Function (CKF) and goes on to describe various CKFs such as Business, Constraints, Vulnerabilities, Opportunities etc. Wiig describes this panoply of CKFs as "multiple CKFs" leading to complexity escalation, causing problems in the management and organization of people and activities in order to extract maximum value from the knowledge base. The problem of identifying the key areas of knowledge conversion and dissemination has also been addressed by Hall and Andriani (1998) in their knowledge capability framework.

This growth of internal complexity is extended with external influences. Markets shift, uncertainty dominates, technologies proliferate, competition multiplies and products and services have shorter life cycles. Thus knowledge management becomes key to organizational development, learning and innovation (McElroy (2003)). At the heart of this organizational development is the relationship between tacit knowledge (informal), which is held within individuals and explicit knowledge (formal), which is articulated, codified and distributed (Davenport & Prusak (2000), Nonaka & Ta-kuechi (1995)). Knowledge creation, dissemination and use within an organization are dependent on the conversion of tacit to explicit knowledge. Nonaka and Takuechi (1995) propose four ways in which this happens:

- Socialisation (tacit to tacit)
- Externalisation (tacit to explicit)
- Internalisation (explicit to tacit)
- Combination (explicit to explicit).

Whilst technology plays a role in this conversion (e.g. use of an Intranet system), it is essentially a human based process. Von Krogh et al (2000) argue that knowledge exchange and development is best achieved formally through clear organizational goals. In contrast SainteOnge and Wallace (2003) assert that this conversion issue is best addressed informally through human processes with groups of like-minded individuals coming together in Communities of Practice (CoP) to share knowledge. It is this informal aspect that we have tried to develop within our complexity model (see later). As if ensuring that the organization "knows what it knows" is not complex enough, defining what it "needs to know but does not know" adds yet another layer of complexity. Koulopoulos (1997) proposes the concept of knowledge gaps, which can be evaluated by using the "knowledge chain" comprising:

- Internal awareness leading to
- Internal responsiveness
- External awareness leading to
- External responsiveness.

Again there is both a formal (e.g. patent searching) and informal (e.g. networking) elements to this problem. What is essential in either case is the ability for the organization/individuals to learn, develop and adapt. Within organizations there are critical knowledge functions, with people accumulating (from internal and external sources), converting and disseminating knowledge for practical application. Some of this is achieved through formal systems, but the scale and complexity of the processes means that much of this is achieved through informal processes such as CoPs. It is the combination of for-mal and informal (based on complexity theory) organizational structures and systems that is critical. This provides the balance between freedom and control required within a commercial operation. The informal aspects are too important to be left to chance and need to be encouraged and “incorporated” into formal organizational processes.

5. System Structure

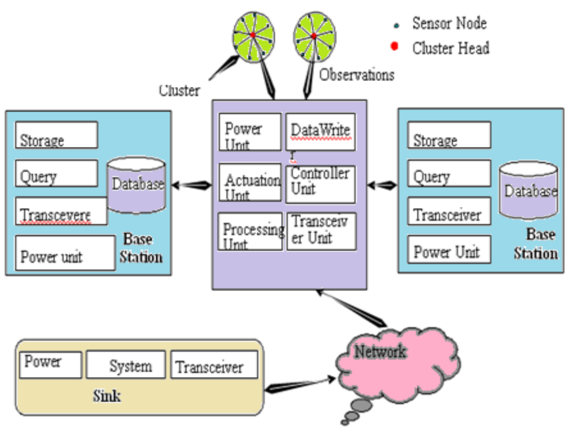

Fig-1 System Structure

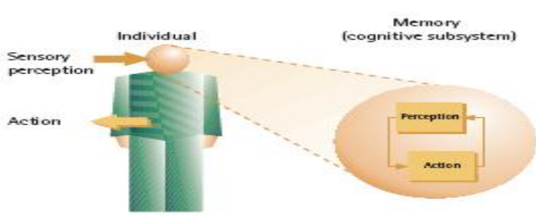

Fig-2 Individual Knowledge

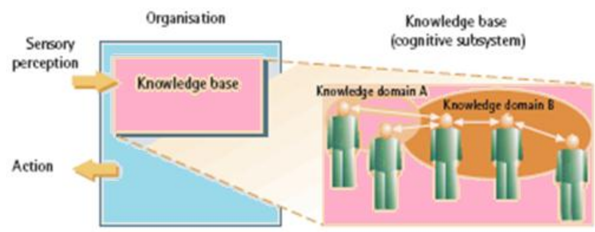

Fig-3 Organizational Knowledge

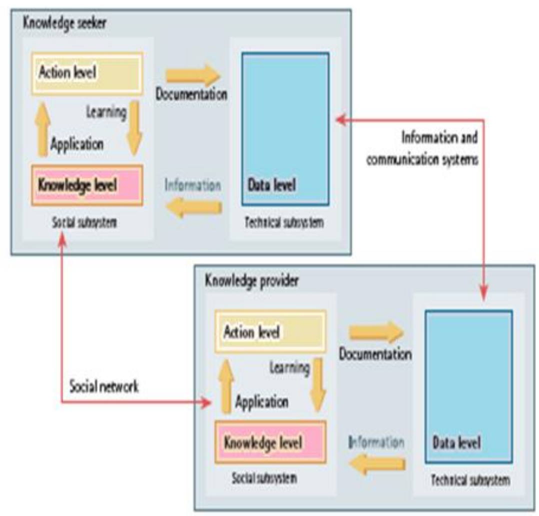

Fig-4 Knowledge transfer between knowledge seeker and knowledge provider

The main purpose of storing the entire information of the cell in the base station is that better data integrity can be achieved as ARS need not spend time on data aggregation before taking an effective decision.

The main advantage of replication is that in case due to some severe disaster some base station collapses. The ARS can obtain the information from the adjacent cell’s base station. Our main aim is to prevent the collapse of the communication system.

The functional components of the base station can be broadly divided into storage manager component, query processor component, transceiver unit and power unit. The storage manager component provides the interface between the low level data stored in the database and queries submitted to the system. Its main goal is to simplify and facilitate access to data. The storage manager component may include authorization and integrity manager, file manager, buffer manager etc. The query manager component handles queries received from ARSs; it may further include DML compiler, metadata manager, query evaluation engine etc.

5.1 Action and Relay Station

The ARS are resource rich nodes equipped with better processing capabilities, higher transmission powers and longer battery life. The ARS nodes are placed on the bordering areas of cells and are responsible for data dissemination in a time efficient manner. An ARS unit consists of six basic components: an actuation unit, a processing unit, controller (decision unit), data writer, a power unit, and a transceiver unit. The decision unit functions as an entity that takes sensor readings as input and generates action commands as output. These action commands are then converted to analog signals and transformed into actions via the actuation unit. The ARS nodes are placed on each pair of shared edges along the border between two cells. The total number of seed ARS nodes needed for a N-cell system are computed using (5). Every ARS supports two types of interfaces: ad-hoc relay interface and cellular interface. By ad-hoc interface, ARS may communicate with other ARSs and sink nodes. It uses cellular interface to communicate with base stations of cellular network. During disaster any ARS may be collapsed. But there is little chance of collapsing all ARSs of a cell. Only one ARS is enough to convey data from sensor network of a cell to a base station. The data writer component of ARS can pass the information received from various CHs of WSN immediately to the corresponding base stations or it may store the information received in its local buffer and after some time send the combined information to further conserve power and communication bandwidth.

6. Sensor Network Knowledge Based Query processing time and space complexity

The issue of estimating database query space and time complexities. Initially, queries without joins are considered and classified into five categories in accordance with complexity (type and number of clauses) in a progressive manner. The storage space and execution time complexity measures for each category are then derived after translating the queries into their algebraic representations and then deriving possible relations that accounts for the different factors (i.e., clauses found in the statement). Joins were then considered and similar complexity expressions were derived.

Modes of evaluation.- Important modes of query evaluation: *filtering*, *full-fledged*, and *pattern matching*. The primary notion in all modes is a *match*. For a query Q and a node r∈Q, we denote by Qr the sub-query rooted at r. Similarly, for a document D and an element e∈ D, we denote by D the sub-document rooted at o.

6.1 Query Complexity

On the World Wide Web, effective query processing used to be impossible due to the lack of data structures and scheme information. The time is proportional to the size of QUERY, the sum of the times required for each iteration:

$$\sum_{i=1}^{x} \mathbf{QUERY}$$

The change in QUERY with each iteration is dependent on the data structure, QUERY grows exponentially, and the total number of iterations equals log2C, where C is the number of concepts in E-R (Entity Relation) Concepts that are related to one another in E-R Relationships. For a list, QUERY remains constant, while the total number of iterations equals C. But any particular concept will be queried for atmost once, and every concept can potentially be related to the concept(s) of interest.
The time when the size of QUERY is one tuple is equal to Q, then the upper bounds on the time required for the algorithm to terminate becomes:

$$C*Q.$$

For large databases that are searched on indexed fields, Q may be approximately on the order of Log R, where R is the number of tuples in the table being searched. Q, being constant, is dropped, resulting in:

$$0(C).$$

The amount of space required to run the Query O(R), where R equals the number of tuples in table E-R Relationships, since TUPLES may equal the size of R, but cannot exceed R*(transitive closure).

7. Performance Evaluation

The performance is measured by average energy indulgence, system lifetime, successful data delivery and number of live nodes. To improve the longevity of the system it may be desired that only minimal set of sensors sense and report the environment. It is expected that our architecture will achieve greater energy savings, enhanced availability and fault tolerance. The scalability is very easy to achieve is our architecture and it will prolong network operation lifetime.

8. Conclusion

We assert that the best implementation for reliability in Real time Knowledge Placement & Management in Wireless Sensor Networks architectures involves both the knowledge based Query processing and network management. We propose a reliable and real time tactic for knowledge placement and achieving data integrity using self organized sensor clusters. In our proposed architecture we suggest storing of information of all clusters within a cell in the corresponding base station.

References

[1] C. Lu, B. M. Blum, T. F. Abdelzaher, J. A. Stankovic, and T. He. “RAP:A real-time communication architecture for large-scale wireless sensor networks,” in 8th IEEE Real-Time and Embedded Technology and Applications Symposium, 2002, pp. 55- 66.

[2] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "The design of an acquisitional query processor for sensor networks", Proc. of the 2003 ACM SIGMOD Int. Conf. on Management of Data, 2003, pp. 491- 502.

[3] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad-hoc sensor networks", Proc. of the 5th symposium on Operating Systems Design and Implementation, vol. 36, 2002, pp. 131 – 146.

[4] T. He, J.A. Stankovic, C. Lu, T. Abdelzaher. “SPEED: A stateless protocol for real-time communication in sensor networks.” Proc. of 23rd Int. Conf. on Distributed Computing Systems (ICDCS'03), Providence, Rhode Island, USA, May, 2003

[5] Y. Yao and J. E. Gehrke, "The Cougar approach to in-network query processing in sensor networks", ACM Sigmod Record, vol. 31, no. 3.

Authors Profile

Prof. Arabinda Nanda: Received M. Tech (CS) from Utkal University in the year 2007.Currently working as Assistant Professor in the Department of Computer Science & Engineering at Krupajal Engineering College, Bhubaneswar, Orissa, India. Contributed more than 12 research level papers to many National and International journals and conferences. Having research interests include Sensor Network, Adhoc Network, Soft Computing, Artificial Intelligence and Data Mining.

Prof. Saroj Kumar Rout: Received M. Tech (CS) from Utkal University in the year 2007.Currently working as Assistant Professor in the Department of Computer Science and Engineering at Krupajal Engineering College, Bhubaneswar, Orissa, India. Contributed more than 08 research level papers. Having research interests include Sensor Network, Adhoc Network, Embedded System, and Network Distributed System.

Prof (Dr) Amiya Kumar Rath: Obtained B.E. degree in Computer Science & Engg. from Marathwada University, Maharastra in the year 1990, MBA degree in Systems Management from Shivaji University in the year 1993, M.Tech in Computer Science from Utkal University in year 2001 and Ph.D in Computer Science in the year 2005 from Utkal University for the work in the field of Embedded system. Served in various positions in different premier institutes namely College of Engineering, Osmanabad, Maharastra, Orissa Engineering College, Bhubaneswar, Kalinga Institute of Industrial technology (KIIT), Bhubaneswar, Krupajal Engineering College, and Bhubaneswar in the Department CS & IT Engg. Presently working with College of Engineering Bhubaneswar (CEB) as Professor of Computer Science & Engg. Cum Director (A&R) and is actively engaged in conducting Academic, Research and development programs in the field of Computer Science and IT Engg. Contributed more than 35 research level papers to many national and International journals. and conferences Besides this, published 4 books by reputed publishers. Having research interests include Embedded System, Adhoc Network, Sensor Network, Power Minimization, Biclustering, Evolutionary Computation and Data Mining.