Uncertainty-aware Wireless Sensor Networks

Sanchita Mal-Sarkar* and Iftikhar U. Sikder

Department of Computer
and Information Science,
Cleveland State University,
Cleveland, OH 44115, USA
E-mail: s.malsarkar@csuohio.edu
E-mail: i.sikder@csuohio.edu
*Corresponding author

Chansu Yu and Vijay K. Konangi

Department of Electrical
and Computer Engineering,
Cleveland State University,
Cleveland, OH 44115, USA
E-mail: chansuyu@gmail.com
E-mail: v.konangi@csuohio.edu

Abstract: The characterisation of uncertainty and the management of Quality of Service are important issues in mobile communications. In a Wireless Sensor Network, there is a high probability of redundancy, correlation and noise in the sensor features since data is often collected from a large array of densely deployed neighbouring sensors. This article proposes a soft computing approach to manage uncertainty by reasoning over inconsistent, incomplete, and fragmentary information using classical rough set and dominance-based rough set theories. A methodological and computational basis is provided and is illustrated in a real world sensor network application of aquatic biodiversity mapping under uncertainty.

Biographical notes: Sanchita Mal-Sarkar is a Doctoral candidate in the Electrical and Computer Engineering Department and a full-time Term Instructor in Computer and Information Science Department at the Cleveland State University. She received MSc in Physics (with specialisation in Electronics) from the Benaras Hindu University, Benaras, India. She received her second MS in Computer Science from the University of Windsor, Winsor, Canada. Her research interests include data mining, uncertainty management in information systems, wireless communications, sensor networks and fault-tolerant optical networks.

Iftikhar U. Sikder is an Assistant Professor of Computer and Information Science at the Cleveland State University. He holds a PhD in Computer Information Systems from the University of Maryland, Baltimore. His research interests include spatial data warehousing and data mining, uncertainty management in information systems and understanding collaborative aspect of spatial decision support systems. He has authored numerous journal articles, book chapters and presented papers in many national and international conferences.

Chansu Yu received BS and MS in Electrical Engineering from the Seoul National University, Korea, in 1982 and 1984, respectively, and PhD in Computer Engineering from the Pennsylvania State University in 1994. He is currently an Associate Professor in the Department of Electrical and Computer Engineering at the Cleveland State University in Cleveland, Ohio. He has authored/co-authored more than 60 technical papers and numerous book chapters in the areas of mobile networking, performance evaluation and parallel and distributed computing. Yu is a member of the ACM, IEEE and IEEE Computer Society.

Vijay K. Konangi is a Professor of Electrical and Computer Engineering at the Cleveland State University. He was awarded PhD by the Iowa State University. His principal research area is computer networks and his research on aerospace telecommunications and networks has received substantial funding from NASA Glenn Research Center. He is a Co-Author of a book published by the IEEE Computer Society Press.

1 Introduction

In recent years, there is a growing interest in using Wireless Sensor Networks (WSN) in diverse non-deterministic environment (Zou and Chakrabarty, 2004). Sensor networks are being employed in many practical applications that require complex data sampling, analysis and systems integration framework. Many practical applications require deployment of a large number of unattended high-density sensor nodes in a sensor field in frequently changing environment. Moreover, WSNs are required to collect, propagate and integrate data under severe resource (energy, bandwidth and memory) constraints. Hence, uncertainty is an endemic aspect of WSN and its management in terms of data integration, Quality of Service (QoS), performance evaluation is critical. While uncertainty in general affects the system's ability to perform with accuracy and precision, the impact of uncertainty in sensor networks may have additional serious consequences because of its inherent constraints. Managing uncertainty and uncertainty-aware sensor field architecture in a WSN are growing research areas (Wang et al., 2004; Zhao and Guibas, 2004; Zou and Chakrabarty, 2004). While there is an increasing awareness of uncertainty and its aspects and dimensions in the management of sensor networks, little agreement exists among experts on how to characterise them. Existing protocols and algorithms for traditional wireless *ad hoc* networks are inadequate in handling uncertainty. Currently, many sensor networks estimation problems are addressed in a statistical framework where a covariance matrix is used to characterise the uncertainty in a Gaussian-like process or in more general probability distributions for non-Gaussian processes (Zhao and Guibas, 2004). Although probability theory is, by far, a well-known

formalism, apparent limitations, such as probabilistic intolerance to imprecise probability, have paved the way for many rich alternative variations (Parsons and Hunter, 1998) which have yet to be appreciated in sensor network research.

In principle, uncertainty in sensor networks may emerge from ontological constraints (e.g. the lack of specification of what kind of entities exists in the routing path and when). Epistemic uncertainty in WSN springs from the concern over whether such entities are knowable to subjective schemes of sensor nodes in the given spatio-temporal resolution. It involves the characterisation of epistemic parameters and the understanding to what extent these parameters can be represented in the subjective framework. In particular, epistemic uncertainty in sensor networks emerges due to inadequate representation of knowledge that is often incomplete, imprecise, fragmentary and ambiguous. For example, one can imagine a surveillance network for contaminant detection in water distribution systems. A conventional detection problem asks for a trade-off between the probability of detection and the false alarm rate because the consequence of a warning could be the stoppage of the infrastructure service which is quite a costly measure; the false alarm must be avoided as much as possible. However, the major challenge is that there are so many contaminants that even a large array of (contaminant-specific) sensors may not be adequate to detect all contaminants, thus leaving the water distribution system vulnerable to contaminants for which no effective sensor was available or was not employed.

In this article, we propose a computational and methodological basis using rough set theory and dominance-based rough set theory to identify redundant sensor features and select indispensable sensors. A computational formalism is provided that generates minimal covering induction rules among sensor features to improve data quality and to reduce uncertainty in a severely energy-constrained environment. Section 2 describes managerial implications of WSN from the mobile communications point of view and outlines requirements of uncertainty in evaluating QoSs. Section 3 provides an overview of the research challenges of data aggregation in the context of uncertainty management, particularly in resource constrained sensor networks characterised by incomplete and fragmented signal. Section 4 introduces rough set theory as a computational tool for the uncertainty handling mechanism. In Section 5, we illustrate a domain specific application of in-network hierarchical data aggregation techniques. A knowledge-centric domain model of aquatic biodiversity mapping application is integrated with the classical rough set and dominance-based principles to derive minimal covering decision rules. Section 6 briefly outlines the strategic aspects of WSN as well as management implications for mobile communications and discusses the research findings. Finally, Section 7 provides research conclusions with some open issues.

2   Managerial implications for mobile communications

Integrating emerging network structure of modern network enterprises with WSN raises possibility of increasing complexity and uncertainty (Li and Chandra, 2007). The strategic positioning of WSN in the value chain requires an enumeration of a large number of factors, namely risk assessment of potential investment under uncertainty, characterisation of QoS, knowledge transfer of management and control experiences across an extended enterprise. Historically, the recognition of value-centred strategic understanding of mobile IT has been considered the key source of organisational competitive advantages (Sheng, Fui-Hoon and Siau, 2005). This approach identifies the

major strategic implications of mobile IT in improvement of working process, increment of internal knowledge management and enhancement of sales and marketing effectiveness. Wu and Wang (2005) proposed an extended technology acceptance model to explore the determinant factors of m-commerce. Using a fit-viability framework Liang et al. (2007) introduced a model to develop a set of measurement instruments to assess the fit and viability in adopting mobile technology. While there are large numbers of literatures on mobile IT adoption models, the strategic implications of the potential of WSN are not yet fully realised in the business world. Since uncertainty is intrinsically intertwined with business strategic risk management, understanding the implications of WSN at the organisational level should involve assessment of potential threats to manage the spectrum of strategic risks. Such a risk-centred approach could provide long-range strategic advantages to deliver maximum benefits (Slywotzky, 2004).

From a supply chain point of view, the management implications involve extensive surge and innovations in sensor-based product tracking and management in the upstream of supply chain, e.g. in manufacturing and distribution processes. This includes confluence of WSNs, embedded Radio Frequency Identification (RFID), Global Positioning Systems (GPS) and Location-Based Services (LBS) producing enhanced transparency in the business process management (e.g. end-to-end visibility of shipments, production and manufacturing). Unlike RFID, which is related to single hop network, WSNs are rapidly deployable in multi-hop networks, and hence can be used in *ad hoc* situations. Since WSN is capable of storing data locally, it can store data temporarily when RFID tags are out of network until the tags become 'alive' and download the data to a local sink. Thus, the integration of RFID-WSN provides significant transparency in audit trails with timeline under *ad hoc* situations. Such integration provides an opportunity to leverage real-time insight into critical supply chains and inventory management information systems. On the downstream side of supply chain, the ripple effect transcends beyond business-to-business relationships. The ability to track and organise massive information of geospatial and temporal distribution of goods and services to the level of individual consumer i.e. Business-to-Customer (B2C) creates a huge potential for efficient merchandising and marketing strategies in m-commerce (Barnes, 2003; Scornavacca and Barnes, 2006). This aspect creates potential uncertainty in the assessment of perception of risk and privacy in technology adoption. In recent years, there is growing concern about the perception of risk and uncertainty involved in wireless technology adoption (Weis et al., 2003; Günther and Spiekermann, 2005). Thiesse (2007) provides a strategic framework based on risk perception and technology acceptance as well as a set of options for coping with the public perception.

While WSN provides significant competitive advantages in developing vast arrays of strategic applications in the value chain, there is still lack of standards for end-to-end QoS (Das, 2004). In WSN, QoS refers to resource reservation control mechanisms wherein real-time adaptive prioritisation and policies are to be implemented. It requires realisation of application specific demands by ensuring flow of data and performance under constrained and uncertain environments. Currently, organisational requirement specifications and standards on end-to-end QoS are lacking. Particularly, adoption and propagation of WSN technology depend on resources control mechanisms. Therefore, an assessment of strategic implications of WSN should take into account various uncertainty parameters of QoSs, such as server resource constraints, unbalanced traffic, data redundancy, network instability, heterogeneity of traffic type and variable scalability (Chen and Varshney, 2004; Wang, Liu and Yin, 2006). It is anticipated that at an

organisational level, a knowledge management approach to handling uncertainty of quality services could provide desired outputs, as similar experiences have been reported in the industrial manufacturing (Koh and Gunasekaran, 2006).

3   Characterisation of uncertainty in Wireless Sensor Networks

Characterisation of uncertainty in sensor networks involves understanding observational aspects (due to noise, interference and signal conversion), model, model parameters and representational uncertainties. Broadly, two major categories of uncertainty can be identified in dealing with WSNs: ontological uncertainty and epistemological uncertainty. Ontologically, variability, also known as aleatory or objective uncertainty, occurs when the object that needs to be sensed by sensor networks actually exhibits multiplicity across space, time and scale. An empirical quantity measured by sensors may objectively manifest multiple aspects in multiple scales in space-time. Such uncertainties are often handled by a reductionism approach; by a process of aggregation or disaggregation of data; or by estimation of space-time frequency distribution, signature unmixing or decoupling, multiscale and multiresolution analysis. The ontological uncertainty can also be approached by monitoring more general features and parameters to identify a signature of 'normal' conditions and flag anomalies. It requires a significant sensor network component and involves a more substantive 'multi-parameter filtering' problem to estimate the state of the system and derive a conditional probability that the condition is normal or anomalous. While the origin of uncertainty due to objective variability of sensor features are ontological in nature, parameter uncertainty or model uncertainty reflect the epistemic state or lack of knowledge in the system components of sensor network. The classical errors of commission or omission, the choice of sampling scheme, systematic bias introduced in the selection of space-time boundary conditions, level of precision, and other parameters internal into the sensor systems may generate epistemological uncertainty in managing sensor networks. These issues constitute the problem of management of epistemic uncertainty regarding where to sense and how often to sense. This kind of uncertainty can propagate due to the dependency on the resolution of observation and the extent of granularity. For example, the observation of coarser granularity offers less detail while the clumping of information into an aggregate form may prevent finer entities from being distinguished by sensors.

While many mathematical models currently exist to deal with uncertainty in sensor networks, there is scant literature that supports adequate representation of intrinsic uncertainty and ambiguity in data sets being used in the network models. Given the growing demands for complex domain specific applications of sensor networks, extensive research is needed that deals with uncertainty due to model-based data aggregation, fusion and propagation in real-time environments. In particular, the context and the limitations of knowledge discovery techniques of WSN are yet to be understood within the framework of the uncertainty handling mechanism.

3.1   *Managing uncertainty in sensor networks: related works*

Research related to uncertainty issues in WSNs addresses broadly two distinct aspects: location or deployment uncertainty and data information uncertainty resulting from data aggregation. Location uncertainty emerges when the placement of sensors is required in a sensor field and the exact locations of the sensors are not known. From the viewpoint of

location uncertainty, routing and location protocols have been proposed for event reporting to mobile sink or target tracking (Howard, Matarić and Sukhatme, 2001; Patwari and Hero, 2003; Zou and Chakrabarty, 2004). Zou and Chakrabarty (2004) developed a model to optimise the number of sensors and their location in a distributed sensor network.

Wang et al. (2004) propose a Bayesian method to describe the lower bound of localisation uncertainty in terms of minimum entropy in sensor networks. The dependency of localisation uncertainty on the sensor network topology is determined by using the Bayesian method and the Cramer-Rao bound. Thus, the algorithm identifies the region where the target is relatively located with some accuracy by assuming Gaussian sensing uncertainty. However, the model did not consider heterogeneous sensors and non-Gaussian sensing. Buttyán, Schaffer and Vajda (2006) propose RANBER, an algorithm for resilient data aggregation in sensor networks by eliminating outliers, based on the well-known RANdom SAmple Consensus (RANSAC) paradigm. The RANBER algorithm is useful even when a large percent of the sample has been compromised by an attacker. The model consists of an aggregator function and a detection algorithm. The detection algorithm analyses the input data before the aggregation function is called and detects unexpected deviations in the received sensor readings. The sample is divided into two halves and the sum for each half is calculated. If the difference of the two sums is greater than a threshold value, it indicates an attack. Reznik and Kreinovich (2004) investigate the issues for improving the reliability, accuracy and uncertainty management of the decisions based on the application of the meta-level models in sensor networks. The meta-level model represents a relationship or association between different sensors. The model depends on expert opinion, data mining techniques (genetic algorithm, neural networks and decision trees) and the type of data collected from sensors. The model attempts to integrate sensor results with the association information available at aggregation nodes and considers both neuro–fuzzy and probabilistic methods to review sensor results and association information. From a database point of view, Cheng and Prabhakar (2003) introduce a data uncertainty framework that represents different levels of uncertainty in information. Depending on the amount of uncertainty in information given to the application, different levels of imprecision are presented in a query answer. They examine the situations when query answer imprecision can be represented qualitatively and quantitatively. An application of range query in a sensor network requires handling interval query and management of uncertainty intervals qualitatively; however, the use of other queries, such as nearest-neighbour queries requires probabilistic threshold information.

3.2 Data redundancy and uncertainty management

The goal of data aggregation in WSN is to fuse data and to conserve energy and bandwidth. The idea behind data aggregation is that data is gathered from multiple sources, combined at a point and periodically transmitted to the sink or base station instead of being sent individually from sources to the sink. Data from several sensor nodes is combined to reduce overall traffic and power consumption in the network to improve the performance and quality of data. Network life time, data accuracy, data freshness and latency are some of the important measures of data aggregation schemes. Timing plays an important role in determining data accuracy and data freshness; the important decision is how long a node should wait to receive data from its downstream

nodes before forwarding to the sink or base station. Longer waiting time increases data accuracy, but decreases data freshness. We can save significant amount of energy by proper selection of data aggregation and forwarding intervals. There are trade-offs among network life time, data accuracy, data freshness and latency (Solis and Obraczka, 2003).

In WSNs, data is gathered from multiple sources and periodically transmitted to the sink or base station for processing. The amount of data gathered at the sink could be overwhelming if all sources send the data directly to the sink. Besides this, there is a high probability of redundancy and correlation in data since data is often collected from densely deployed neighbouring sensors. In order to reduce redundancy and to improve the quality of data, we need to combine data at intermediate nodes or sources and reduce the number of packets transmitted to the sinks. Apart from uncertainty emerging from redundancy in densely deployed neighbouring sensors, a significant amount of data can be lost or corrupted during the transmission from sensor nodes to the sink because of intrusion attacks, node failures or battery depletion in sensor networks.

The redundancy in the sensor network can be attributed to two distinct sources: the uncertainty regarding the frequency of distinct samples to be covered by the network and the number of observable attributes to be monitored by each sensor node. Uncertainty in sampling frequency can be handled by means of statistical estimation, such as simulating space-time sample distribution by using Monte Carlo simulation and domain specific *a priori* distribution. A high dimensional attribute-oriented sensor network can adversely affect the communication as well as data processing (e.g. training in learning systems) performance. Moreover, many real world systems exhibit non-polynomial complexity with respect to attribute dimensionality. For example, a large-scale water treatment plant may require a huge number of attributes to monitor water quality parameters through sensors in order to perform diagnostic detection of faults. Assuming that the diagnostic system's complexity with respect to *n*, the number of variables in the domain is $O(n^{1.75})$; the reduction of five attributes in such systems would result in a 21.9% increase in the speed of the inference process. Without reduction, the inference speed would explode by a factor of 32 (for NP-hard inference engines (given $O(2^n)$). Moreover, the costs associated with connecting sensors and maintaining connections to diagnostic computing equipment can be reduced while points of failure (malfunctioning sensors or overly noisy sensors) are reduced significantly (Shen and Chouchoulas, 2000).

4   Rough Set theory: reasoning under uncertainty

While managing uncertainty in distributed sensor networks requires a rigorous formalism to deal with QoS in terms of inconsistency, ambiguity and fragmentary aspects of data, the traditional perspective of uncertainty management is focused on large, centralised, organisational databases (Pawlak and Slowinski, 1994). Handling uncertainty due to data aggregation and missing information requires space-time syntheses in rigorous formalism. Information granulation or the concept of indiscernibility is at the heart of rough set theory (Düntsch and Gediga, 1998). The rough set theory offers attribute reduction algorithm and the dependency metric for feature selection. In rough set, aggregation or granularity is expressed by partition and their associated equivalence relations on the sets of objects which is also called indiscernibility relations. Indiscernibility leads to the concept of boundary-line cases which means that some elements can be classified according to the concepts or to their complements with the available information, and thus forms the boundary-line cases. The boundary-line cases

imply elements of the universe that cannot be with certainty classified as elements of concepts (Sikder, 2008).

Rough set, as a machine learning algorithm, has been extensively used to deal with inexact, uncertain or vague knowledge (Doherty et al., 2006; Sikder and Gangopadhyay, 2007; Sikder, 2008). The main advantage of a rough set is that it is inherently data driven and 'non-invasive' (Polkowski and Skowron, 1998). Unlike fuzzy set theory or statistical analysis, a unique advantage of a rough set is that it does not rely on other model assumptions or external parameters. By utilising the structure of the given data from sensor networks, researchers can develop the numerical value of imprecision or a membership function without requiring any subjective inference on distribution function. Rough set approximations are derived objectively from the given set itself. The data representation in the rough set requires the formation of an information system which is a pair $S = (U, A)$ where U is a non-empty, finite set called the universe and A is a non-empty, finite set of attributes (i.e. $a: U \rightarrow V_a$ for $a \in A$ where V_a is the value set of attribute a). A decision table is an information system of the form $S = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute or class label where we can assume a set V^d of values of the decision d is equal to $\{1, \ldots, r(d)\}$. Decision d determines the partition $\{X_1, \ldots, X_{r(d)}\}$ of the universe U where $X_k = \{x \in U: d(x) = k\}$ for $1 \le k \le r(d)$. This system can be generalised as the decision system $S = (U, A, d_D)$ where $d_D(x) = (d_1(x), \ldots, d_k(x))$ for $x \in U$ (Mitchell, 1997). For an information system $S = (U, A)$, any $B \subseteq A$ is associated with an equivalence relation $\text{IND}_A(B)$ (also called as B-indiscernibility relation, its classes are denoted by $[x]_B$.) defined by: $\text{IND}(B) = \{(x, x') \in U^2$: for every $a \in B$, $a(x) = a(x')\}$. Objects x, x' are indiscernible by attributes B. Given a set of attributes $B \subseteq A$ and $X \subseteq U$, a set of objects, we can approximate X by constructing the B-lower and B-upper approximations of X, $\underline{B}X$ and $\overline{B}X$, respectively, where $\underline{B}X = \{x \in U : [x]_B \subseteq X\}$ and $\overline{B}X = \{x \in U : [x]_B \cap X \ne \varnothing\}$. The set $BN_B(X) = \overline{B}X - \underline{B}X$ represents the B-boundary of X. The accuracy of approximation is measured by $\alpha_B = |\underline{B}(X)| / |\overline{B}(X)|$ where $0 \le \alpha_B \le 1$. A set is rough if $\alpha_B(X) < 1$ (i.e. X is vague with respect to B). Assuming B and Q are equivalence relations in U, the concept of positive region $\text{POS}_B(Q)$ is defined as follows:

$$\text{POS}_B(Q) = \bigcup_{X \in Q} \underline{B}X.$$

Given the attribute set B from sensor nodes, all patterns in U are contained in the positive region that can be mapped to attribute Q. One can derive the notion of degree of dependency of sensor attributes from the concept of the positive region. The degree of dependency (B, Q) of a set B of variables with respect to a set Q is defined as $\gamma_B(Q) = |\text{POS}_B(Q)| / |U|$ which provides a measure of how important B is in mapping the dataset into Q. When $\gamma(B, Q) = 0$, then attributes B are of no use to estimate attribute Q; therefore, these attributes can be ignored by the sensor nodes. If $\gamma(B, Q) = 1$, then B is no longer dispensable and Q completely depends on P. A partial dependency in the range $1 < \gamma(B, Q) < 0$ reflects the tolerance range of sensor attributes. Using this concept, we can estimate the significance of an individual sensor. Assuming that a sensor measures an attribute $x \in B$, the significance of the sensor can be estimated by $\sigma_x(B, Q) = \gamma(B, Q) - \gamma(B - \{x\}, Q)$. A high value of $\sigma_x(B, Q)$ indicates that the removal of the sensor measuring the attribute x significantly affects the overall classification quality.

While if $\sigma_x(B,Q) = 0$, then it indicates that the removal of the sensor does not make any change, and hence the sensor is redundant.

The rough set based data reduction approach allows feature reduction or the identification of redundant attributes and their dependency are derived in rough sets by calculating reducts which allow the decision of whether some of the attributes are redundant or superfluous with respect to decision class. Reducts are all the subset of attributes that are minimal (i.e. that do not include any dispensable attribute). Extraction of reducts requires the construction of an $n \times n$ matrix (c_{ij}), called the discernibility matrix of an information system, such that $c_{ij} = \{a \in A : a(x_i) \neq a(x_j)\}$ for $i, j = 1, \ldots, n$. A discernibility function f_A for an information system is a Boolean function defined by $f_A(\overline{a}_1, \ldots, \overline{a}_m) = \wedge\{\vee \overline{c}_{ij} \quad 1 \leq j < i \leq n, c_{ij} \neq \varnothing\}$ where $\overline{c}_{ij} = \{\overline{a} : a \in c_{ij}\}$ Here, $a_{il}{}^{\wedge} \ldots {}^{\wedge} a_{ik}$ is a prime implicant of f_A. It has been shown that the set of all prime implicants of f_A determines the set of all reducts (Skowron and Grzymalla-Busse, 1994). Computing all possible reducts is a non-trivial task. While computing prime implicants is an NP-Hard (Wroblewski, 1995) problem, it is possible to use heuristic algorithms (e.g. genetic algorithms or dynamic reducts (Polkowski and Skowron, 1998)) to generate a computationally efficient set of minimal attributes. Once the reducts have been computed, then deriving the decision rule is a simple task of laying the reducts over the original decision table and mapping the associated values. Such rules derived from the training set can be used to classify new instances for which the decision classes are unknown. However, it is likely that more than one rule may fire to decide a class for a new object. In that case, strategies are to be adopted (e.g. standard voting) to resolve conflicts among candidate rules that recognise the same object (Greco, Matarazzo and Slowinski, 2002).

5   Application scenario: aquatic biodiversity mapping under uncertainty

What follows is an illustration of the application of the rough sets rule induction method to deal with inconsistent and missing information in WSN. Using rough set formalism, we show that it is possible to reason over incomplete and inconsistent information received from spatially distributed sensors. The application scenario is an aquatic sensor network-based biodiversity mapping. Underwater sensors can be used to determine the quality of water or biodiversity by measuring its characteristics, such as temperature, density, salinity, acidity, chemicals, conductivity, pH, oxygen, dissolved methane gas and turbidity (Akyildiz, Pompili and Melodia, 2005). In our approach, we assume that different sets of sensors measure different sets of characteristics. The sensors in a cluster are equipped with domain specific function procedures or lookup tables with limited computing capability. The communications among the sensors are only one-hop away, and we assume that all sensor nodes are mobile. A sensor node that can reach the maximum number of sensor nodes in one hop will be selected as a cluster-head, and it broadcasts an advertisement to all other nodes in the network.

The cluster-heads gather data from all non-head nodes, aggregate data, and communicate directly with the sink or base station. Cluster-head nodes consume more energy than non-cluster-head nodes because the cluster-head needs to receive data from all cluster members in its cluster and then send the data to the sink. The cluster-heads are selected in each round to make sure that the energy consumption is evenly distributed among all the sensor nodes and to prolong network lifetime. The scheme uses

Time Division Multiple Access (TDMA) Media Access Control (MAC) for intra-cluster communications and Code Division Multiple Access (CDMA) for inter-cluster transmissions. TDMA has two phases: the setup phase to organise the clusters and the steady-state phase to allow all nodes to transmit periodically during their time slots.

Since the data is processed locally and only the result is sent to the sink, this data aggregation technique decreases energy consumption during data transmission. Figure 1 represents a series of snapshots of a particular cluster taken during different intervals of time. In the first instance, F is selected as the cluster-head because it has the highest number of nearest neighbours (A, B, C, D and E). The scenario is changed during the following instance and E becomes the new cluster-head as nodes are mobile. A new node, G, joins the cluster and it sends the HELLO packet back to its neighbours whenever it hears from them. In the third instance, node E leaves the cluster and joins another cluster, and node C is selected as the cluster-head. When the new cluster-head, node C, does not hear from E in a specified interval of time, it places the node E in a missing node list.

The information system for a cluster $S_c = (U, A \cup \{d\})$ is generated dynamically by the cluster-head which consists of observations of a sensor network (u_{T_n}) representing the global states of sensor nodes in a cluster at different sampling time T_n, and attribute $A = \{S_c(f): s(f) \in Sc\}$ represents a set of parameters derived from each sensor in the cluster. For example, $S_C(P)$ is the aggregated parameter of aquatic species specific suitability rank reflecting physical aquatic conditions such as temperature, pressure and turbidity; and $S_C(C)$ is the chemical suitability rank of the species generated by aggregation from lookup tables involving sensor parameters such as salinity, dissolved oxygen and pH. Similarly, $S_C(B)$ and $S_C(H)$ represent a biological ecosystem index reflecting competition among species and the habitat suitability index, respectively. The suitability rankings are computed from preloaded lookup tables generated by domain experts. As the size of the lookup tables (or domain heuristics) are very small, the sensor nodes require limited processing power to generate the suitability indices from raw information at the sampling location. The advantage of using such domain heuristics is that they obviate the need for storage and for the communication overhead of massive raw information at the sampling level. The decision variable (DI) represents the overall measure of the Diversity Index (DI) in an ordinal rank assigned by the cluster-head.

Figure 1 A series of snapshots for cluster configuration
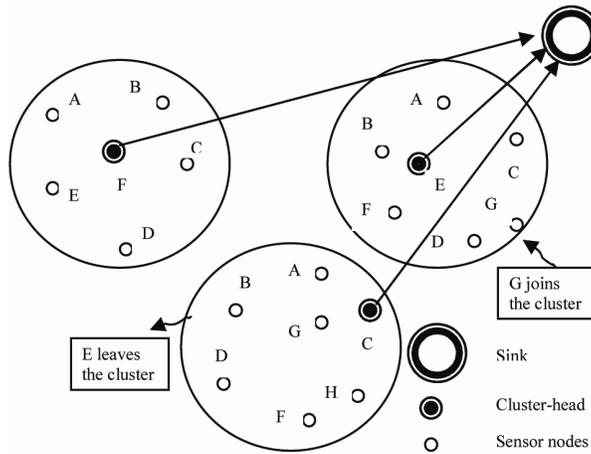
Evidently, the information system S_c (Table 1) shows a realistic and typical scenario of missing information in the data collection process by a sensor network. As the information in S_c is preference ranked in an ordinal scale, it is possible that if a certain observation u_{T_x} outranks the observation u_{T_y} on all criteria, then u_{T_x} can be considered to dominate u_{T_y}. Using dominance-based rough set theory (Skowron and Suraj, 1996; Suraj, 2003), we see that the set of instances or observations which dominate a particular observation u_{T_x} with respect to conditioning attributes $P \subseteq A$ or is dominated by objects formally expressed as follows:

$$D_P^+(u_{T_x}) = \left\{ u_{T_y} \in U : u_{T_y} D_p u_{T_x} \right\}$$
$$D_P^-(u_{T_x}) = \left\{ u_{T_{yx}} \in U : u_{T_x} D_p u_{T_y} \right\}$$

where $u_{T_y} D_p u_{T_x}$ implies that u_{T_y} dominates u_{T_x} with respect to attribute $P \subseteq A$. With the given information, it is possible to derive the lower approximation of $\mathrm{DI}_{high}^{\geq}$ given by $\left(\underline{A}\left(\mathrm{DI}_{high}^{\geq}\right) = \left\{ u_T \in U : D_p^+(u_T) \subseteq \mathrm{DI}_{high}^{\geq} \right\} \right)$ that indicates which observations belong to $\mathrm{DI}_{high}^{\geq}$ with certainty or without any ambiguity. The *P*-upper approximation of $\mathrm{DI}_{high}^{\geq}$ is expressed as $\overline{P}\left(\mathrm{DI}_{high}^{\geq}\right) = \left\{ u_T \in U : D_p^-(u_T) \cap \mathrm{DI}_{high}^{\geq} \neq \varnothing \right\}$ indicating the set of objects that could belong to $\mathrm{DI}_{high}^{\geq}$. Analogously, it is possible to define $\underline{P}\left(\mathrm{DI}_{low}^{\leq}\right)$ and use the similar dominance relation. Figure 2 shows the dominance-based rough set approximation of the diversity index measure with respect to sensor attributes.

Table 1   Rough set Dynamic Information system generated by the sensor cluster-head

	$S_C(P)$	$S_C(C)$	$S_C(B)$	$S_C(H)$	*Diversity Index*
u_{T_1}	2	1	2	1	high
u_{T_2}	1	?	2	1	low
u_{T_3}	?	?	1	1	low
u_{T_4}	2	?	2	2	high
u_{T_5}	1	?	2	1	high
u_{T_6}	1	1	1	1	low

Figure 2 Rough set approximation of information under uncertainty by using lower and upper approximation of classes

(a)   (b)
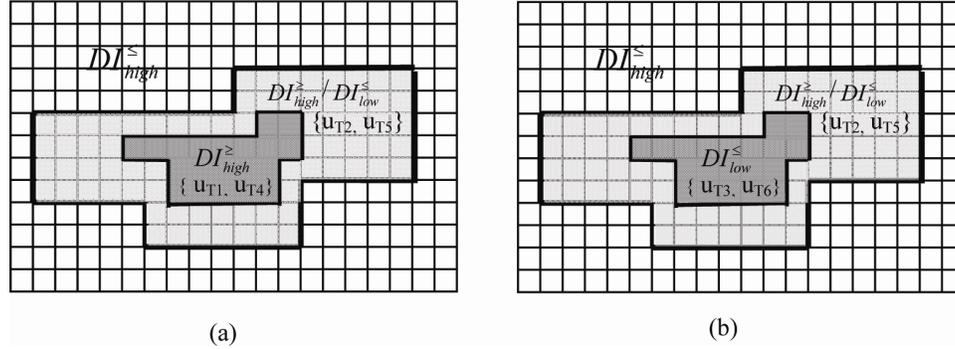
(a) representing at least high diversity and (b) at most low diversity; The inner core area represents absolute certainty in classification; the grey outer area represents the boundary region where classification is uncertain.

Table 2 The minimal covering rules generated by the rough set rule induction algorithm

$S_C(P)$	$S_C(B)$	Diversity Index	Support	Relative strength (%)
	$<= 1$	$DI^{\leq}_{low}$	2	100
$>= 2$	$>= 2$	$DI^{\geq}_{high}$	2	100
$<= 1$	$>= 2$	$DI^{\leq}_{low} \vee DI^{\leq}_{high}$	2	100

Using the above information, one can calculate the reducts to remove redundant sensors not contributing to the content of the lower approximation. The reduct was computed as $\Re = \{S_c(P), \ S_c(B)\}$, showing that information retrieved from $S_C(C)$ and $S_C(H)$ is dispensable and does not add any value to increase certainty in the calculation of the current cluster. We used a rule-induction algorithm Dominance-Based Learning from Example Module (DOMLEM) (Greco et al., 2000), built on the idea of Learning from Example Module 2 (LEM2) (Grzymala-Busse and Hippe, 2001), for generating a minimal description of the preference systems. Using the algorithm, we obtained three minimal covering decision rules. Since there are six records in the table, three minimal covering represents approximately a 50% reduction of the knowledge systems. We obtained two deterministic rules describing (at most low) $DI^{\leq}_{low}$ and (at least high) $DI^{\geq}_{high}$. The minimal covering rules derived are shown in Table 2.

The first rule indicates that if $S_C(B)$ is at most 1, then the diversity index is at its lowest point. This rule applies to u_{T_3} and u_{T_6} with 2 supports and 100% relative strength. The second rule dictates that if $S_C(P)$ is at least 2 and $S_C(B)$ is at least 2, then the diversity index is at least high. The support for this rule is derived from u_{T_1} and u_{T_4}. The third rule is an example of a non-deterministic rule due to an ambiguous observation of u_{T_2} and u_{T_5}. The extracted rules do not include redundant sensors $S_C(C)$ and $S_C(H)$.

The cluster-heads are in charge of indexing and caching rules generated from the information system using the dominance-based rough set and dispatching the rule streams to the sink. This process can be addressed by transforming rough set rules to Petri Nets. A parallel algorithm for constructing Petri Nets or coloured Petri Nets from real-time rough set rules and decision tables has been proposed (Skowron and Suraj, 1996; Suraj, 2003). In real-time concurrent environments, this approach holds promise in managing sensor information dynamically. Clearly, this scheme allows an energy-efficient mechanism to compress the data stream by removing redundant sensors and encoding sensor signals in terms of a set of rules. Thus, this approach improves energy efficiency by coordinating the sleep schedules of sensor nodes. We can achieve this goal by finding out the maximum number of subgroups of disjoint nodes such that each subgroup can predict the entire set of sensors. In the WSN, since there could possibly be numerous concurrent cluster-heads sending information to sink, the concurrent decision rules need to be properly space-time indexed.

6   Discussion

The strategic positioning of WSN in the value chain requires explicit characterisation of uncertainty parameters. It provides a significant insight in business decision making and technology adoption. While strategic benefits of WSN entails enhanced transparency and end-to-end visibility in business process management, there are a number of specificities of WSN that requires an uncertainty management framework. In particular, strategic decisions regarding WSN implementation call for assessments of risk factors that define uncertainty parameters of QoSs. From a supply chain point of view, the implications of WSN include enhanced transparency in product tracking and management in the upstream of supply chain as well as downstream tracking at the consumer level creating a huge potential for efficient merchandising and marketing strategies in m-commerce. An assessment of managerial implications should include the perception of risk, uncertainty and privacy in technology adoption.

From a management point of view, an explicit characterisation of uncertainty is essential in mission planning, resource deployment, resource planning, sensor scheduling and sensor control. For example, questions like how frequently to measure sensor features which area of the environment to focus on and what criteria are relevant for sensor selection cannot be modelled *a priori*. Deterministic resource planning for WSN requires explicit assignment of the responsibility to each sensor, such as sensor selection, multi-target tracking and sensor allocation for classification, negotiation and cooperation between sensors in WSN. Similar requirements can be demanded for sensor scheduling and sensor control (Xiong and Svensson, 2002). Although the redundant deployment of sensors allows for sampling of a signal at finer resolution and provides high fault tolerance and noise immunity, there is an inherent trade-off between QoS and resource constraints which must be addressed from an uncertainty-aware WSN framework. Therefore, the need for uncertainty-driven WSN management and strategic control is apparent.

From a resource management point of view, a larger sample size requires more energy, bandwidth and latency. However, stochastic randomness and fuzziness can be reduced by collecting more samples. Hence, it is important to develop a strategy that requires the fewest possible selections of sensors sufficient to reconstruct data for the

entire WSN for a particular decision making while keeping the energy requirement, fault tolerance and noise immunity within a predefined boundary. Since traditional data compression techniques have limited use for dense sensor networks, avoiding oversampling beyond network capacity for a sensor network is necessary to prevent data transmission from some sensors (Marco et al., 2003; Gupta et al., 2008). In our article, we have introduced rough set theory to identify redundant sensors and to select indispensable sensor features from inconsistent and incomplete information. A methodological and computation foundation is provided with an emphasis on uncertainty management and risk assessment. Using a Dominance-based Rough Set Approach (DRSA) – an extension of rough set – to reason with incomplete and imprecise data sets, we have illustrated an application for intergradations of experts' preference-based knowledge systems in WSN. Given the rough set's generic and non-invasive modelling advantage, a knowledge integration framework can be designed to handle some of the challenges in the complex network management of various applications.

7 Conclusions

Strategic positioning of WSN in the value chain should involve a risk assessment of the QoSs as well as the characterisation of uncertainty in the management of WSN. Given the growing demands for complex domain specific applications of sensor networks, research on uncertainty-aware sensor networks is needed to deal with data aggregation, fusion and propagation. Using rough set theory, we propose a soft-computing methodological basis to deal with data redundancy and inconsistency in a resource constrained environment. Unlike traditional probabilistic approaches, this methodology is computationally less resource intensive and allows for real-time management of the sensor data stream from the local clusters without requiring substantial communication overhead. Using the formalism of the dominance-based rough set approach, we illustrate an application domain where a dynamic information system is generated for aquatic biodiversity mapping. It provides a methodology to identify redundant and indispensable sensor features, and to determine feature dependencies from inconsistent and incomplete information. Unlike traditional compression algorithms, the rough set based dynamic feature selection algorithm allows the compression of the data stream without altering the underlying data semantics. The discovery of minimal covering 'if.., then…,' decision rules from sensor features offers additional advantage of using domain heuristics. It obviates the need for storage for the communication overhead of massive raw information at the sampling level. In addition, the decision rules induced from the rough set approximations can be transformed to Petri Nets for managing concurrent processes (e.g. space-time indexing of features) from multiple cluster-heads in real-time decision making in WSN.

References

Akyildiz, I.F., Pompili, D. and Melodia, T. (2005) 'Underwater acoustic sensor networks: research challenges', *Ad Hoc Networks*, Vol. 3, pp.257–279.

Barnes, S. (2003) 'The wireless application protocol as a platform for mobile services', *Mbusiness: The Strategic Implications of Wireless Communications*, pp.1–264, Amsterdam: Elsevier.

Buttyán, L., Schaffer, P. and Vajda, I. (2006) 'RANBAR: RANSAC-based resilient aggregation in sensor networks', Paper presented in the Proceedings of the *Fourth ACM Workshop on Security of AD HOC and Sensor Networks*, pp.83–90, Alexandria, VA, October 30.

Chen, D. and Varshney, P.K. (2004) 'QoS support in wireless sensor networks: a survey', *International Conference on Wireless Networks*, pp.227–233, Las Vegas, Nevada, June 21–24.

Cheng, R. and Prabhakar, S. (2003) 'Managing uncertainty in sensor databases', *ACM SIGMOD Record*, Vol. 32, pp.41–46.

Das, S.K. (2004) 'Coping with uncertainty in wireless mobile networks: an information theoretic framework', Paper presented in the Proceedings of the *Second International Workshop on Mobility Management and Wireless Access Protocols*, p.1, Philadelphia, PA, 1 October.

Doherty, P., Lukaszewicz, W., Skowron, A. and Szalas, A. (2006) 'Knowledge representation techniques: a rough set approach', *Series: Studies in Fuzziness and Soft Computing* (Vol. 202, pp.1–342). New York, NY: Springer

Düntsch, I. and Gediga, G. (1998) 'Uncertainty measures of rough set prediction', *Artificial Intelligence*, Vol. 106, pp.77–107.

Greco, S., Matarazzo, B. and Slowinski, R. (2002) 'Rough approximation by dominance relations', *Int. J. Intelligent Systems*, Vol. 17, pp.153–171.

Greco, S., Matarazzo, B., Slowinski, R. and Stefanowski, J. (2000) 'An algorithm for induction of decision rules consistent with the dominance principle', in W. Ziarko and Y. Yao (Eds), *Rough Sets and Current Trends in Computing* (pp.304–313). Berlin, Germany: Springer-Verleg.

Grzymala-Busse, J.W. and Hippe, Z.S. (2001) 'Melanoma prediction using K-nearest neighbor and LEM2 algorithms', *Intelligent Information Systems* (pp.43–55). Poland, Zakopane: Springer.

Günther, O. and Spiekermann, S. (2005) 'RFID and the perception of control: the consumer's view', *Communications of the ACM*, Vol. 48, pp.73–76.

Gupta, H., Navda, V., Das, S. and Chowdhary, V. (2008) 'Efficient gathering of correlated data in sensor networks', *ACM Transactions on Sensor Networks*, Vol. 4, Article 4, pp.1–31.

Howard, A., Matarić, M.J. and Sukhatme, G. (2001) 'Relaxation on a mesh: a formalism for generalized localization', Paper presented at the Proceedings of the *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1055–1060, Wailea, Hawaii, October.

Koh, S.C.L. and Gunasekaran, A. (2006) 'A knowledge management approach for managing uncertainty in manufacturing', *Industrial Management and Data systems*, Vol. 6, pp.439–459.

Li, X. and Chandra, C. (2007) 'A knowledge integration framework for complex network management', *Industrial Management and Data Systems*, Vol. 107, pp.1089–1109.

Liang, T-P., Huang, C-W., Yeh, Y-H. and Lin, B. (2007) 'Adoption of mobile technology in business: a fit-viability model', *Industrial Management and Data Systems*, Vol. 107, pp.1154–1169.

Marco, D., Duarte-Melo, E.J., Liu, M. and Neuhoff, D.L. (2003) 'On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data', Paper presented in the Proceedings of the *International Workshop on Information Processing in Sensor Networks (IPSN)*, pp.1–16.

Mitchell, T.M. (1997) *Machine Learning* (pp.1–414). Portland, OR: McGraw-Hill.

Parsons, S. and Hunter, A. (1998) 'A review of uncertainty handling formalisms', *Applications of Uncertainty Formalisms* (Vol. 1455, pp.8–37). London, UK: Springer-Verlag.

Patwari, N. and Hero, A.O. (2003) 'Relative location estimation in wireless sensor networks', *IEEE Transactions On Signal Processing*, Vol. 51, pp.2137–2148.

Pawlak, Z. and Slowinski, R. (1994) 'Rough set approach to multi-attribute decision analysis', *European Journal of Operational Research*, Vol. 72, pp.443–459.

Polkowski, L. and Skowron, A. (1998) 'Rough sets in knowledge discovery I: methodology and applications', in L. Polkowski and A. Skowron (Eds), *Studies in Fuzziness and Soft Computing* (pp.1–576). Heidelberg, Germany: Physica-Verlag.

Reznik, L. and Kreinovich, V. (2004) 'Fuzzy and probabilistic models of association information in sensor networks', *13th IEEE International Conference on Fuzzy Systems*, Vol. 1, pp.185–189, Budapest, Hungary, 25–29 July, 2004.

Scornavacca, E., Jr. and Barnes, S.J. (2006) 'Barcode enabled m-commerce: strategic implications and business models', *Int. J. Mobile Communications*, Vol. 4, pp.163–177.

Shen, Q. and Chouchoulas, A. (2000) 'A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems', *Engineering Applications of Artificial Intelligence*, Vol. 3, pp.263–278.

Sheng, H., Fui-Hoon, F. and Siau, K. (2005) 'Strategic implications of mobile technology: a case study using value-focused thinking', *Journal of Strategic Information Systems*, Vol. 4, pp.269–290.

Sikder, I.U. (2008) 'Discovering decision heuristics in collaborative planning', *Int. J. Management and Decision Making*, Vol. 9, pp.1–15.

Sikder, I.U. and Gangopadhyay, A. (2007) 'Managing uncertainty in location services using rough set and evidence theory', *Expert Systems with Applications*, Vol. 32, pp.386–396.

Skowron, A. and Grzymalla-Busse, J. (1994) 'From rough set theory to evidence theory', in R. Yager, M. Fedrizzi and J. Kacprzyk (Eds), *Advances in the Dempster-Shafer Theory of Evidence* (pp.192–271). New York, NY: John Wiley and Sons, Inc.

Skowron, A. and Suraj, Z. (1996) 'A parallel algorithm for real-time decision making: a rough set approach', *Journal of Intelligent Information Systems*, Vol. 7, pp.5–28.

Slywotzky, A. (2004) 'Exploring the strategic risk frontier', *Strategy and Leadership*, Vol. 32, pp.11–19.

Solis, I. and Obraczka, K. (2003) 'In-network aggregation trade-offs for data collection in wireless sensor networks', *Technical Report No. INRG 102*, pp.1–18. Santa Cruz, CA: University of California Santa Cruz.

Suraj, Z. (2003) 'Information systems as a tool for specification of concurrent systems', Paper presented in the Proceedings of the *2003 IEEE International Conference on Robotics and Automation*, Vol. 3, pp.4324–4329, Taipei, Taiwan, 14–19 September.

Thiesse, F. (2007) 'RFID, privacy and the perception of risk: a strategic framework', *The Journal of Strategic Information Systems*, Vol. 16, pp.214–232.

Wang, Y., Liu, X. and Yin, J. (2006) 'Requirements of quality of services in wireless sensor network', Paper presented in the Proceedings of the *International Conference on Networking*, p.116, April.

Wang, H., Yip, L., Yao, K. and Estrain, D. (2004) 'Lower bounds of localization uncertainty in sensor networks', Paper presented in the Proceedings of the *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp.917–920, Quebec, Canada, 17–21 May.

Weis, S.A., Sarma, S.E., Rivest, R.L. and Engels, D.W. (2003) 'Security and privacy aspects of low-cost radio frequency identification systems', *Lecture Notes in Computer Science*, Vol. 2802, pp.454–469, London, UK: Springer.

Wroblewski, J. (1995) 'Finding minimal reducts using genetic algorithms', in P.P. Wang (Ed), Paper presented in the Proceedings of the *International Workshop on Rough Sets Soft Computing at Second Annual Joint Conference on Information Sciences (JCIS'95)*, pp.186–189, Wrightsville Beach, NC.

Wu, J.H. and Wang, S.C. (2005) 'What drives mobile commerce? An empirical evaluation of the revised technology acceptance model', *Information and Management*, Vol. 42, pp.719–729.

Xiong, N. and Svensson, P. (2002) 'Multi-sensor management for information fusion: issues and approaches', *Information Fusion*, Vol. 3, pp.163–186.

Zhao, F. and Guibas, L.J. (2004) *Wireless Sensor Networks: An Information Processing Approach* (pp.1–376). Boston, MA: Morgan Kaufmann.

Zou, Y. and Chakrabarty, K. (2004) 'Uncertainty-aware and coverage-oriented deployment for sensor networks', *Journal of Parallel and Distributed Computing*, Vol. 64, pp.788–798.

Wireless sensor network (WSN) is a high and new technology consists of spatially distributed autonomous sensors to monitor physical or environmental conditions. With the development of computer and communication, wireless sensor network has been a research hot point. WSN is an international refereed journal dedicated to the latest advancement of wireless sensor network and applications. Wireless sensor network (WSN) refers to a group of spatially dispersed and dedicated sensors for monitoring and recording the physical conditions of the environment and organizing the collected data at a central location. WSNs measure environmental conditions like temperature, sound, pollution levels, humidity, wind, and so on. These are similar to wireless ad hoc networks in the sense that they rely on wireless connectivity and spontaneous formation of networks so that sensor data can be transported We consider the sensor deployment problem in the context of uncertainty in sensor locations subsequent to airdropping. Sensor deployment in such scenarios is inherently non-deterministic and there is a certain degree of randomness associated with the location of a sensor in the sensor field. We present two algorithms for the efficient placement of sensors in a sensor field when the exact locations of the sensors are not known.