

Research Article

Data-Centric Knowledge Discovery Strategy for a Safety-Critical Sensor Application

Nilamadhab Mishra, Hsien-Tsung Chang, and Chung-Chih Lin

Department of Computer Science and Information Engineering, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-Shan, Tao-Yuan 333, Taiwan

Correspondence should be addressed to Hsien-Tsung Chang; smallpig@widelab.org

Received 15 January 2014; Revised 15 April 2014; Accepted 16 April 2014; Published 24 June 2014

Academic Editor: Hon Tat Hui

Copyright © 2014 Nilamadhab Mishra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In an indoor safety-critical application, sensors and actuators are clustered together to accomplish critical actions within a limited time constraint. The cluster may be controlled by a dedicated programmed autonomous microcontroller device powered with electricity to perform in-network time critical functions, such as data collection, data processing, and knowledge production. In a data-centric sensor network, approximately 3–60% of the sensor data are faulty, and the data collected from the sensor environment are highly unstructured and ambiguous. Therefore, for safety-critical sensor applications, actuators must function intelligently within a hard time frame and have proper knowledge to perform their logical actions. This paper proposes a knowledge discovery strategy and an exploration algorithm for indoor safety-critical industrial applications. The application evidence and discussion validate that the proposed strategy and algorithm can be implemented for knowledge discovery within the operational framework.

1. Introduction

A real-time data-centric sensor network is a network of sensors with explicit timelines to process data. Sensors are smart devices that are capable of sensing data within the permitted coverage, controlling operations, and communicating over a network. This type of network has a greater number of nodes that are distributed and communicated in a given region to measure the physical environmental data (e.g., pressure, temperature, sound, and humidity) with limited resources.

With the advancement of industries and sensors, real-time DSNs have been considered in many contexts to automate an industrial process that involves a high degree of risk [1]. Therefore, based on risk quantification, we classify real-time sensor applications into three different categories: business-critical sensor applications—sensor applications in inventory management and mission-critical sensor application—sensor application in habitat monitoring, and safety-critical sensor application—sensor application in the chemical industry. Among the three applications, the highest degree of risk is measured in safety-critical sensor applications.

The real-time DSN application mainly consists of in-network activities through minimal support of the sensor data accumulation capability, data processing capability, and data storing capability. In a large safety-critical sensor application, where the sensors and actuators are housed together, the actuator cannot perform any instant intelligent actions from the raw sensor data that have been accumulated. Therefore, proper knowledge is required for actuators to perform time-critical risk-driven actions. Improper knowledge may lead to damage to society and social property (e.g., sensors employed to control chemical plants and temperature sensors used in blast furnaces).

As shown in Figure 1, the four basic functional units of a DSN are as follows: the data accumulation unit (DAU), the data processing unit (DPU), the data storage unit (DSU), and the data communication unit (DCU). If $DPU(E)$, $DAU(E)$, and $DCU(E)$ represent the energy consumed in the respective units, then $DPU(E) \ll DAU(E) \ll DCU(E)$, as the radio transceiver is the highest energy consumer. Additionally, in some applications, the DAU consumes more energy compared to the DPU [2]. Therefore, processing with minimum energy consumption encourages researchers to implement

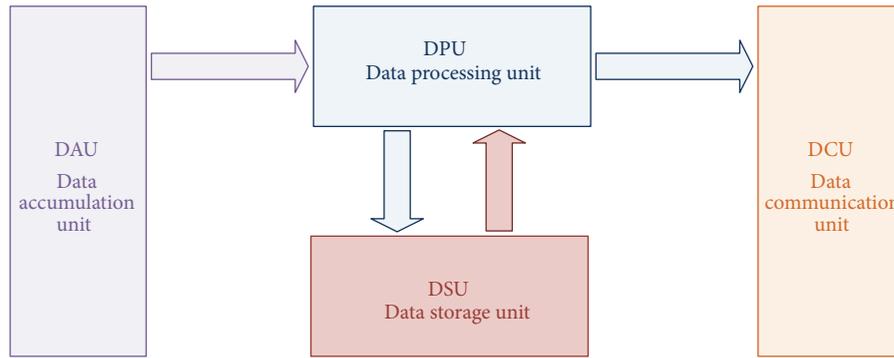


FIGURE 1: Four basic functional units of a DSN.

the knowledge discovery strategy (KDS) within the sensor network environment.

The data-centric KDS creates research challenges when determining the amount of knowledge required for a real-time DSN to preserve knowledge quality, consistency, accuracy, timeliness, integrity, and completeness and produces knowledge of tactical value to actuators. Obtaining semantically enriched knowledge from raw harvested sensor data to meet stringent requirements is a challenge. However, its success depends entirely on the data-centric knowledge discovery and dissemination strategy.

DSNs have gained increasing attention in the industrial automation process because of their high elasticity, dynamicity, and efficiency. Therefore, we may implement our real-time DSN applications in chemical industries to monitor and control various unsuspected scenarios, such as dangerous chemical exposure, a poisoned atmosphere caused by toxic gas leaks, fires and explosions, thermal threats, electrical threats, mechanical failures, and oxygen deficiency to maintain a healthy atmosphere inside the chemical plant [3].

The remainder of this paper is organized as follows. In Section 2, we discuss related work on data collection and a KDS for DSN. In Section 3 we propose a KDS for safety-critical sensor applications. In Section 4, we present the application evidence. In Section 5, we provide a theoretical discussion of our proposed strategy. Finally, we conclude this paper in Section 6.

2. Related Works

In a real-time sensor environment, only 49% of the accumulated raw sensor data are useful for knowledge interpretation, and approximately 3–60% of the data are defective [4, 5]. Therefore, the treatment of 51% of the unuseful raw sensor data and 60% of the defective data poses a major hazard to knowledge data discovery (KDD). Hereafter, the DSN treats the data collection problem as one of the most sensitive issues for a large multihop environment, where large sensor nodes are arbitrarily deployed in a random finite region.

During data collection from sensors, the environment may be highly fault tolerant to ensure an adequate data pool [6]. A key problem is associated with maximal data

collection versus minimal node utilization, which increases the network's lifespan [7].

A DSN is broadly considered as the data-producing hub where large and bulk amounts of raw data are produced. Ample literature has focused on numerous ambient issues related to data collection approaches for a DSN, where in-network data processing and management activities are fully emphasized [8, 9]. In the DSN, a flat structure may not be sufficient for active data collection. However, cluster-based data collection approaches have gained improved performance [10, 11].

Quality, reliability, and accuracy are the most important parameters considered during sensor data collection to achieve quality management [12, 13]. These three parameters are highly desired to identify the adequacy of the sensor data.

Energy-efficient data collection issues are emphasized to increase the overall network lifespan. Several studies suggest the use of a power-saving strategy to accumulate data from the sensor environment [14–16].

DSNs are typically studied for major mission-critical and business-critical sensor applications. Additionally, DSNs have key contributions for safety-critical applications to accumulate the real-time data with strict timelines and thresholds [17–19]. An association has been established between sensor applications and safety-critical applications to ensure a safety-critical sensor application in which the sensor nodes are configured as data collection nodes that accumulate real-time data from the physical environment [20, 21].

Real-time sensor data can be made applicable through an effective KDS in which data mining, data warehousing, and computational intelligence are combined to achieve the common goal of DSN applications [22–24]. Data warehousing anticipates the storage of a large amount of sensor data. However, data mining finds useful data patterns from the storage of large data. Between data mining and data warehousing, computational intelligence acts as a catalyst for data mining to find these useful patterns in a timely manner.

From the associated studies, we collect sensor data and issues related to sensor data collection. Additionally, we investigate the use of various parameters that should be considered during sensor data collection, such as quality, energy efficiency, and fault tolerance. Because sensor data collection is the backbone of the KDS and because effective

knowledge discovery depends on an appropriate data collection strategy, after scrutinizing data collection issues and hazards, we proceed to establish a mechanism to utilize these data by integrating three distinct areas (i.e., data mining, data warehousing, and computational intelligence) to ensure a complete KDS [25, 26]. In addition, we contemplate several safety-critical sensor applications (e.g., fire monitoring, gas leak monitoring, and other critical applications) that are assumed to apply to the industrial automation process to supervise the safety and security of different industries, such as the chemical industry, nuclear power reactors, and other highly sensitive areas that have a lower degree of human supervision.

Hence, to achieve a unified KDS in a sensor platform, this study has performed tasks ranging from data collection to knowledge discovery by considering real-time applications with the broad aim of obtaining knowledge from data.

3. Strategy Declaration and Anticipation

The discovery of a knowledge pattern from data is an emerging problem in DSNs; three diversified areas (i.e., data mining, data warehousing, and soft computing) are integrated together to target the common goal of a safety-critical sensor actuator application. Three activities are mainly synchronized within strict timelines for any real-time DSN application—data collection from sensors and knowledge discovery and knowledge dissemination to actuators—and it is essential to combine all three activities into a common frame. The following norms and strategies are used to illuminate the detailed scenarios.

Norms and Guidelines

- (i) The real-time DSN application is configured to monitor safety-critical applications whose failure leads to serious environmental damage.
- (ii) The application usage is purely indoor industrial applications for safety monitoring and control.
- (iii) The local database node is an autonomous microcontroller that acts as a dedicated cluster head (CH) to which the required number of sensors and actuators is connected within a one-hop network.
- (iv) The sensor from each regular moment senses the data from the environment and sends them to the CH for further processing.
- (v) Energy should not be a constraint for indoor safety-critical applications: sensors and actuators should be able to store energy from electricity; however, we still aim to minimize the overall energy consumption.
- (vi) The actuator cannot function intelligently with the raw sensed data; thus, an appropriate strategy is required to yield filtered knowledge.
- (vii) Because it is a safety-critical application, the real-time deadline must be met irrespective of the local server (CH) loads.

- (viii) Because the DSN is implemented for a safety-critical application that anticipates high risks, all constraints of the DSN and safety-critical system must be carefully considered to design this real-time system.
- (ix) Knowledge users can send a query to the DSN through the appropriate channel to determine the current situation.
- (x) The CH must be capable of determining the emergency to generate a quick awareness regarding the emergent situation.

Strategy

- (i) Sensors send raw data to the CH without any manipulations and thus harvest the sensor's energy.
- (ii) The CH receives large continuous raw data from sensors and processes those data using a KDS to produce the desired knowledge.
- (iii) Next, the CH transfers the knowledge to the actuators to take immediate action and simultaneously communicate the knowledge to a knowledge warehouse server that knowledge users can access for further actions.
- (iv) Actuators should report to local servers (CH) to initiate their performed actions so that the knowledge users can query those data for further situation analysis and counteractive actions.

Query processing and optimization represent another major research issue in a large distributed sensory database for achieving excellence in terms of modeling optimal query execution plans. However, choosing an efficient query execution strategy is important in order to minimize the query response time. Additionally, implementing the KDS at the CH provides the operations of the sensor association rule mining to extract the tactical knowledge patterns that can be used by the actuators and knowledge users to perform actionable insights [27, 28].

3.1. Proposed Cluster-Level KDS. Our proposed cluster-level data-to-knowledge (D2K) migration strategy sends the optimal knowledge from the CH to the actuators within strict timelines. Only a fraction of a second may be required to meet the real-time requirements to provide the necessary processing at the cluster level.

Additionally, the KDS intends to design an adequate knowledge database that accommodates the events and corresponding actions along with time stamps so that several probabilistic event detection models can be inherited for future event detections.

Figure 2(a) describes the black box process of the D2K migration strategy. This strategy acts as an entire process that accepts raw sensor data as the input and crops knowledge as the desired output. This label-0 architecture is also known as a context model for the D2K migration strategy. As described in Figure 2(b), five logical processes or functions are performed by the CH's processing unit to obtain the

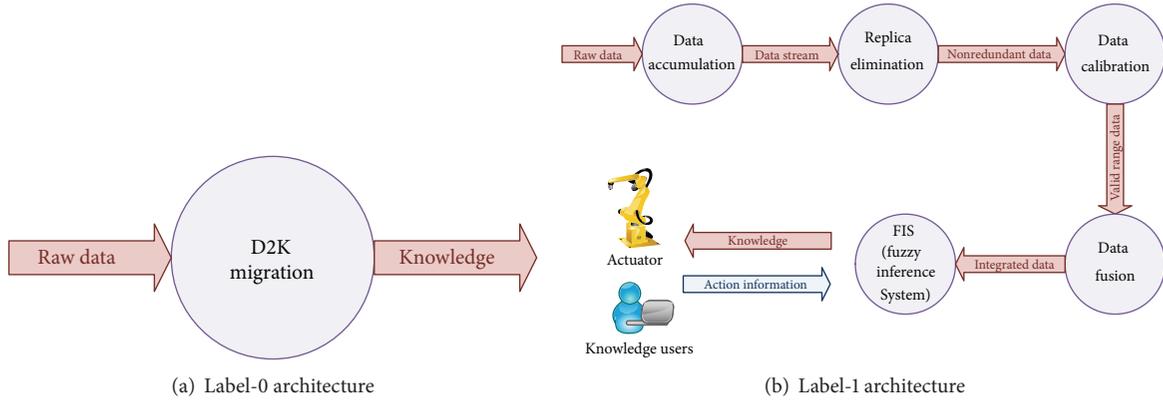


FIGURE 2: (a) D2K migration label-0 architecture and (b) D2K migration label-1 architecture.

required knowledge from the data and ensure a complete KDS before transmitting the data to either the actuators or knowledge users.

Step 1. The data accumulator collects raw sensed data from the sensors and passes the raw data to the replica eliminator.

Step 2. The replica eliminator removes the redundant data and immediately passes them to the data calibrator.

Step 3. The data calibrator only filters the useful data based on a valid range and passes them to the data fuser to produce integrated data.

Step 4. The data fuser produces the integrated database using the fusion operation and passes the integrated database to the fuzzy controller.

Step 5. The fuzzy controller infers the knowledge from the integrated database using a FIS (fuzzy inference system) and ensures that the inferred knowledge can be well perceived by the actuators to perform the intelligent actions. The configuration of a typical fuzzy controller depends on the application requirement of the deploying environment.

The D2K strategy is a power-efficient KDS designed to maximize the DPU(E) and minimize the DCU(E). To limit the communication energy of the DCU, the sensor nodes must send data to the CH in regular time intervals $(T_1, T_2, T_3, \dots, T_n)$ rather than sending data spontaneously. Based on industrial application scenarios, a certain threshold may be applied at the node end to limit the data sent, and a priority-based data-sending policy should be adapted to send crucial data to the local server (CH) without any failures. Additionally, the CH may require an estimation of the average number of data units collected and stored in each sensor buffer to meet the emergency needs of the knowledge users at time t .

Figure 3 presents the load equilibrium between data sensation and data consumption at any time t with a constraint that the data sensation rate (α) must be greater than or equal to the data consumption rate (μ). If the data sensation rate and consumption rate are known, then we can estimate the



FIGURE 3: Sensor with the data sensation/consumption process.

approximate number of data units stored in the sensor buffer at any time t for any sensor. Let α be the data sensation rate and μ be the data consumption rate. Then, the average number of data units at each sensor buffer is L , where $L = \{\alpha/(\alpha - \mu)\}$.

3.2. D2K Strategy Analysis. Five key processes are involved in the transformation of data to knowledge: the data accumulation process, replica elimination process, data calibration process, data fusion process, and knowledge filtration process.

(i) Data Accumulation Process. The data accumulation process is the process of collecting raw data from the sensor nodes on a pull or push basis to the local database node that acts as a CH without any data manipulation at the node end, which ultimately yields the sensor's residual energy. The CH may employ numerous data collection models, such as the probabilistic data collection model, deterministic data collection model, and threshold sensitive data collection model, to assemble the correctly sensed data.

In practical applications, some sensors become faulty during operation, and the faulty value creates a heavy burden on the DSN application to produce tactical knowledge. Therefore, the CH must confirm that the sensed raw data are correct in all aspects before transforming the data into knowledge [29]. A data agreement protocol should be used by the CH to resolve this data inaccuracy problem. A set of rules are associated with this protocol.

Hypothesis. For specific applications, if more than one sensor is deployed to perform a specific task in a cluster region, then at any time instance, those sensors accumulate the same data

value, and, from those sensors, a certain data value may be faulty due to the inaccurate functioning of a few sensors.

Guidelines for Data Correctness

- (i) Every sensor in a cluster senses the individual physical data value, and all nonfaulty sensors must agree on that data value.
- (ii) In a cluster, all sensors should send the data values to the CH, and the CH identifies the most trusted correct data through the major sensed values. For a safety-critical system, highly trustworthy correct data are highly desired to resolve all possible risks.

(ii) *Replica Elimination Process.* In a typical real-time sensor environment, data redundancy is a common problem because within a cluster, more than one sensor may send the same data value to the CH. This process ensures that we obtain a distinct data value at the CH to further minimize the transmission energy by reducing the large data volume.

(iii) *Data Calibration Process.* This process helps obtain the inbound data values by eliminating all outbound data values. Due to an internal fault, some sensors may sense a value exceeding the anticipated value, which creates a major problem at the CH to transform those data into tactical knowledge.

(iv) *Data Fusion Process.* The data fusion process is the process of integrating the distinct, valid, and trustworthy data values into a relational database in which the data are organized in a row-column structure. Therefore, this process, which is simply known as the data integration process at the CH, uses numerous data integration approaches.

One-Step Integration Approach. In this approach, all data values are directly mapped to the relational database at the CH, which is not feasible for continuous, incoming agile data.

Phase-In Integration Approach. At a particular time, some data values can be mapped to a relational database, which is also known as an incremental integration approach or an evolutionary integration approach. This approach builds the most suitable sensor relational database at the CH.

(v) *Knowledge Filtration Process.* This process, the backbone of the KDS, discovers potential knowledge from unstructured raw data; at the CH, a fuzzy controller is employed along with a FIS to yield relevant knowledge and standard knowledge representation formats, such as decision trees, decision tables, association rules, and classification rules that can be used to symbolize the knowledge.

3.3. D2K Functional Exploration

3.3.1. D2K Algorithm. See Algorithm 1.

3.3.2. *D2K Algorithm Analysis.* The entire D2K algorithm should run at an individual CH to transform actionable

stimulates (knowledge) from large sensory datasets in a timely manner.

For the data accumulation function, the data from the corresponding sensors are placed into the CH storage. However, before placement, a suitable data validation scheme may be employed to regulate the data's lawfulness and precision.

The replica elimination function suggests a suitable method of reducing data redundancy at a CH with a concern that an individual subnetwork tends to centralize the approach. However, in the entire distributed sensor database approach, data replication is highly desired to ensure global data availability and parallel access. Therefore, maintaining data replicas at individual CH points has gained research attention.

A simple data calibration logic is presented in the algorithm with the aim of producing data within the valid threshold values. For instance, among large data values, if only a small portion of data values has unexpectedly high deviations compared with the major data values, then a sensitive mechanism may be adapted to handle those highly deviated data.

The data fusion function suggests the mapping of incremental data to a sensor relational database so that the access key to an individual record can be well defined and can act as a unique identifier. However, the object relational model is suitable for handling large dynamic graphics data.

Finally, the knowledge filtration function advocates the use of a fuzzy circuitry embedded with a specific FIS for the purpose of knowledge inference, optimization, and actionable controls. We consider that our D2K algorithm is more applicable if it is capable of producing actionable stimulates at the appropriate time irrespective of the data loads and congestion level.

4. Application Evidence

Consider a DSN environment in which all static sensors and actuators are wirelessly connected through several dedicated local servers (CHs) within a one-hop distance such that each CH acts as a permanent dedicated leader to the partially distributed network. All sensor nodes are battery powered but can harvest energy from electricity. Each local database node (CH) is an autonomous microcontroller that is equipped with powerful processing, communications, and storage units to perform consistent database operations. Here, the sensors communicate with CH, and the CH communicates with the base station and actuators to disseminate the knowledge.

Figure 4 outlines a four-layer architecture in the distributed sensor platform in which layer-1 is responsible for data production, knowledge discovery, and action; layer-2 is responsible for knowledge transportation; layer-3 is responsible for knowledge storage; and layer-4 is responsible for knowledge application and utilization. Layer-1 comprises an industrial operational platform to implement a D2K migration strategy at the CH. In layer-3, effective extraction transformation loading (ETL) operations may be adopted to symbolize the knowledge in several layouts so that the knowledge users can properly exploit the knowledge for

```

D2K_magration_strategy ()
{
  Functional analysis of our D2K strategy can be performed through a triplet format,
  that is, (I, P, O), where I is the input to the functions, P is the process
  to describe the internal operations, and O is the output of the functions.
  Data_accumulation ()
  {
    Input: raw data at motion (flow)
    Output: raw data at rest (storage)
    Process: use pull/push method to accumulate data
    If (method = push)
      {Sensors periodically sense environment and send data to CH}
    If (method = pull)
      {CH injects SQL into sensor environment to extract data}
    CH performs data validations to check the correctness of the data
  }
  Replica_elimination ()
  {
    Input: raw sensor dataset
    Output: distinct dataset
    Process: data matching
    If (old data value = new data value)
      {CH discards to store the new value and the old value remains in the CH storage}
    If (old data value != new data value)
      {CH stores the new data value into its storage}
  }
  Data_calibration ()
  {
    Input: distinct dataset
    Output: valid range dataset
    Valid data range = [lb, ub]
    // ub defines the upper threshold value
    // lb defines the lower threshold value
    Process: equivalence class partitioning
    If (ub < distinct data < lb)
      {CH treats these data as an invalid data class}
    If (lb <= distinct data <= ub)
      {CH treats these data as a valid data class}
    CH rejects the invalid data class from its storage
  }
  Data_fusion ()
  {
    Input: valid dataset
    Output: integrated database
    Process: database operation
    While (arrival of valid data = true)
      {
        CH maps the data into the database with its access key
      }
  }
  Knowledge_filtration ()
  {
    Input: integrated database
    Output: control instructions to be stored in the knowledge base
    Process: fuzzy controller operation with a FIS
    While (availability of integrated database = true)
      {CH permits the database into a fuzzy controller circuitry embedded
      with a FIS to extract the useful knowledge}
  }
}

```

ALGORITHM 1: D2K algorithm.

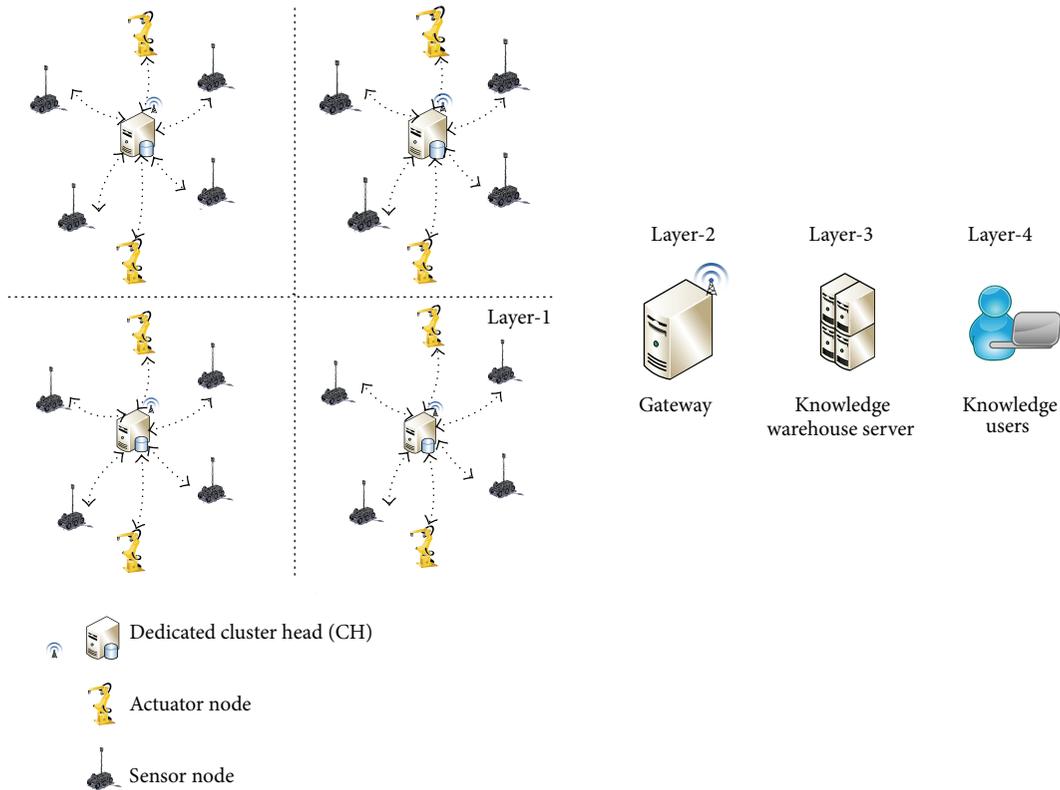


FIGURE 4: One-hop operational framework for an indoor safety-critical industrial application.

monitoring industrial safety and security. Here, a dedicated CH acts as a local database node. The data flow from sensor nodes to the CH and the CH data are transformed into knowledge (actionable data). The CH passes the knowledge to the actuator to perform the intelligent actions with immediate effects and then sends the reported knowledge and events to the knowledge warehouse server via a gateway, where the knowledge users perform further analysis and take the necessary actions and precautions.

This operational platform has real-time implementations in safety-critical industrial applications, such as fire monitoring and the monitoring of highly unanticipated events inside the chemical industry, as described in Section 1. In a fully automated chemical industry, several machines are independently operated by a computer-aided manufacturing (CAM) microcontroller. In each machine, several sensors, such as electrical sensors, motion sensors, and gas sensors, should be fabricated so that whenever the machine breaks during operations, the motion sensor immediately reports to the local server (CH) regarding the machine's failure status. The CH instantly generates the instructions that will be executed by the CAM microcontroller to stop the machine's operation, directs the electrical sensor to disconnect the power supply from that machine, and guides the gas sensors to detect any toxic gas leaks to avoid any further incidents. Both incidents and actions are communicated to the knowledge warehouse server via a gateway, where the knowledge users can have easy access to perform further legal actions. In risk-driven industrial applications, the one-hop topology is preferred due

to its nonstop delivery of confidential data from the sensor to the CH without any interference. Based on this requirement, the knowledge users send their request to the CH through the knowledge warehouse server to extract specific information related to safety, security, and the internal status of the industrial environment.

Regarding the hardware configuration of the CH, we nominate the CH as an authoritative microcontroller device equipped with various hardware components. Therefore, the D2K algorithm can be efficiently embedded into the circuitry of the CH for real-time operations. In the CH, the fuzzy controller circuits are fabricated along with a FIS to generate control instructions to monitor and control the actuators to perform safety-critical actions.

5. Discussion

We defined an operational platform for an indoor safety-critical industrial application to detect and resolve potential threats in the chemical industry, as described in Section 1. We may implement the IEEE 1451.5-802.11 standard-based DSN to construct a one-hop clustered star network in which a dedicated CH is an autonomous microcontroller device and each node contains an 802.11 wireless radio to wirelessly communicate with the central CH [31]. We also proposed a D2K algorithm that runs at an individual CH to filter useful knowledge that is disseminated to actuators and knowledge users to perform timely intelligent actions to protect

TABLE I: Comparison of various works based on the implemented strategies and algorithms for sensor data.

Names of authors	Summary	Strategy/algorithm	Unified approach
Dinesh Kumar et al., 2012 [30]	(i) Sensor data filtration using neural network tools (ii) Gateway level data filtration and compression	(i) KDS (ii) Back propagation algorithm	KDS-NN (neural network)
Khanna and Liu, 2012 [19]	(i) Genetic approach-based pattern identification and activity monitoring (ii) Dynamicity of data collection tree	(i) KDS (ii) Genetic algorithm	KDS-GA
Wu and Clements-Croome, 2007 [22]	(i) Sensor data mining (ii) Knowledge discovery (iii) Data calibration (iv) Data cluster analysis (v) Data replication control	(i) KDS (ii) Data mining algorithm	KDS-DM
Boukerche et al., 2008 [23]	(i) Knowledge discovery for chronological patterns (ii) Pattern identification (iii) Pattern modeling (iv) Pattern mining (v) Behavioral pattern extraction	(i) KDS (ii) Data mining algorithm	KDS-DM
He et al., 2010 [24]	(i) Knowledge discovery for real-time monitoring (ii) Real-time data mining (iii) Data calibration (iv) Data reduction	(i) KDS (ii) Data mining algorithm	KDS-DM
Defeng et al., 2010 [21]	(i) Sensor real-time application monitoring (ii) Intellect at base station to monitor critical activities (iii) Data processing operations	(i) KDS (ii) Data mining algorithm	KDS-DM
Rodrigues et al., 2009 [25]	(i) Data mining and knowledge discovery from sensor data (ii) Cluster-based data analysis	(i) KDS (ii) Data mining algorithm	KDS-DM
Aggarwal, 2013 [26]	(i) Data elicitation (ii) Data cleaning (iii) Data management (iv) Data mining	(i) KDS (ii) Data mining algorithm	KDS-DM
Das, 2012 [27]	(i) Sensor association rule mining (ii) Event-based knowledge pattern discovery	(i) KDS (ii) Data mining algorithm	KDS-DM
Boukerche and Samarah, 2008 [28]	(i) Sensor association rule mining (ii) Behavioral pattern discovery (iii) Distributed data extraction	(i) KDS (ii) Data mining algorithm	KDS-DM
Our work	(i) Promotes in-network KDS through D2K migration process (ii) Implements fuzzy inference system for knowledge optimization and control at the CH level (iii) Aims to be applied in safety-critical industrial applications	(i) KDS (ii) D2K algorithm	KDS-FIS

the chemical plant from highly unexpected mishaps. Therefore, our approach promotes industrial safety by resolving threats.

This study reveals that the volume of raw data at the sensor is substantially higher than the volume of filtered knowledge [30]. Out of 15,000 bytes of raw data, only 400 bytes of packed filtered knowledge can be produced, that is, only 3%;

however, transmitting this voluminous raw data would lead to massive traffic congestion and energy consumption.

Therefore, our proposed D2K algorithm claims to preserve 97% of the transmission energy, and 3% of the packed knowledge can be well perceived by the actuators to perform the safety-critical actions required to resolve threats in the chemical industry.

Over the sensor data, the following unified approaches can be suggested for the KDS, as described in Table 1:

- (i) KDS-NN: KDS applied with a neural network algorithm;
- (ii) KDS-GA: KDS applied with a genetic algorithm;
- (iii) KDS-DM: KDS applied with a data mining algorithm;
- (iv) KDS-FIS: KDS applied with a fuzzy inference system algorithm.

The KDS for sensor data can be widely implemented through data mining tools or computational intelligence tools. Data mining tools stimulate the use of data clustering, classification, calibration, and association, whereas computational intelligence tools promote the use of the fuzzy inference system, artificial neural system, and genetic algorithm to build an effective KDS for different safety-critical sensor applications.

Study Analysis. An analysis of the current research trend of the KDS for sensor data demonstrates that the KDS-DM approach is more widely applied than other approaches. However, the KDS-GA, KDS-NN, and KDS-FIS play vital roles in enriching the discovery efficiency of safety-critical sensor data.

P_1 , P_2 , P_3 , P_4 , and P_5 are 5 performance dimension parameters that measure the relative performance of KDS-GA, KDS-NN, and KDS-FIS, such that

- (i) $P_1 \leftarrow$ reasoning mechanism,
- (ii) $P_2 \leftarrow$ knowledge discovery efficiency,
- (iii) $P_3 \leftarrow$ decision-making capability,
- (iv) $P_4 \leftarrow$ computational efficiency,
- (v) $P_5 \leftarrow$ industrial applicability.

We now compare those three united approaches to determine the effectiveness of KDS-FIS compared to KDS-GA and KDS-NN. Here, we use two fuzzy quantifiers, that is, “more” and “less,” to assign an approximate weightiness to the above parameters. We studied the advantages and disadvantages of the three united approaches that can be individually applied to safety-critical sensors.

We compare those three unified approaches by considering the performance dimension matrix presented in Table 2.

The reasoning mechanism and knowledge discovery efficiency are greater in KDS-FIS compared with KDS-GA and KDS-NN when using a logic base along with an effective inference process. Due to its reduced processing complexity, the KDS-FIS has a higher computational efficiency that yields faster decision-making capabilities compared with the other approaches. The fuzzy approach has considerable industrial applicability due to its high embedding compatibility with heterogeneous microcontroller devices.

For definite instances over sensor data, the approximate crisp range of two fuzzy quantifiers can be expressed in terms of the membership functions as follows:

$$(i) \quad \mu_{(R)}^{(\text{less})} : \text{less} \longrightarrow [0.1, 0.4], \quad (1)$$

TABLE 2: Performance dimension matrix of KDS-GA, KDS-NN, and KDS-FIS for sensor data.

	P_1	P_2	P_3	P_4	P_5
KDS-NN	Less	Less	More	Less	Less
KDS-GA	Less	Less	Less	More	Less
KDS-FIS	More	More	More	More	More

TABLE 3: Performance matrix (R) based on definite instances over sensor data.

	P_1	P_2	P_3	P_4	P_5
KDS-NN	0.4	0.4	0.8	0.4	0.4
KDS-GA	0.2	0.3	0.4	0.6	0.2
KDS-FIS	0.9	0.9	0.8	0.8	0.9

(ii)

$$\mu_{(R)}^{(\text{more})} : \text{more} \longrightarrow [0.6, 1], \quad (2)$$

where $\mu_{(R)}^{(\text{less})}$ and $\mu_{(R)}^{(\text{more})}$ are the membership functions or the degree of belongingness.

The defuzzification process provides a method of extracting a crisp value from the fuzzy quantifiers as approximate representative values. For instance, let R be the output in Table 3.

The relative performance evaluation graph (see Figure 5) clearly demonstrates the general productivity of KDS-FIS over KDS-NN and KDS-GA for any industrial safety-critical applications.

Our D2K algorithm can directly use the KDS-FIS approach to accelerate the decision-making process at the CH to regulate industrial safety-critical events.

For industrial applications, the monitoring quality of KDS-FIS is better than those of KDS-GA and KDS-NN because KDS-FIS uses an expert database or knowledge base directly to effectively generate the action commands for the actuators.

Several characteristics relating to the KDS-FIS implementation, such as dependability, portability, timeliness, and interoperability, may be considered throughout knowledge discovery. The online KDS-FIS presents additional research challenges compared with the offline KDS-FIS in the generation of tactical knowledge value. Computational efficiency is a major hazard for real-time data mining and analytics. The faster and more reliable tactical decisions made through KDS-FIS for safety-critical environments may be considered a problem in DSNs. Moreover, the achievement of KDS-FIS for safety-critical applications requires efficient data elicitation, data processing, and knowledge production.

6. Conclusion

In this paper, we proposed a D2K algorithm and presented application evidence for the effective implementation of KDS in an industrial environment to resolve industrial safety issues. Our proposed D2K discovery algorithm operates in

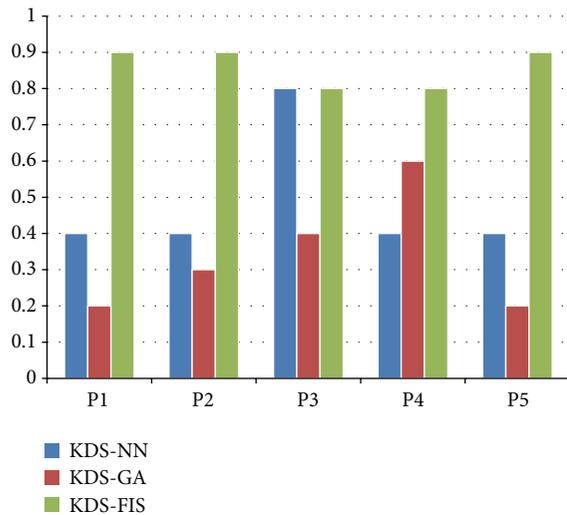


FIGURE 5: Relative performance evaluation based on approximate defuzzification reasoning.

a real-time environment to produce tactical knowledge value for actuators to perform safety-critical actions. Furthermore, the involvement of the fuzzy controller circuitry along with an inference system (KDS-FIS) makes the knowledge representation and control operations effective and efficient.

In subsequent research, we will further study the firmware implementations and updating mechanism of the CH for industrial environments to enhance overall safety and security.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

Financial support provided by the Ministry of Science and Technology, Republic of China, through Grant NSC 102-2221-E-182-060 of Chang Gung University is gratefully acknowledged.

References

- [1] M. Neugebauer, J. Ploennigs, and K. Kabitzsch, "Evaluation of energy costs for single hop vs. multi hop with respect to topology parameters," in *Proceedings of the IEEE International Workshop on Factory Communication Systems (WFCS '06)*, pp. 175–182, June 2006.
- [2] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.
- [3] <http://EzineArticles.com/4533265>.
- [4] K. Ni, N. Ramanathan, M. N. H. Chehade et al., "Sensor network data fault types," *ACM Transactions on Sensor Networks*, vol. 5, no. 3, article 25, 2009.
- [5] R. Szweczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pp. 214–226, November 2004.
- [6] L. Zhang, Q. Ye, J. Cheng et al., "Fault-tolerant scheduling for data collection in wireless sensor networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 5345–5349, usa, December 2012.
- [7] C. Wang and H. Ma, "Data collection in wireless sensor networks by utilizing multiple mobile nodes," in *Proceedings of the 7th International Conference on Mobile Ad-hoc and Sensor Networks (MSN '11)*, pp. 83–90, Beijing, China, December 2011.
- [8] C. Cheng, C. K. Tse, and F. C. M. Lau, "A delay-aware data collection network structure for wireless sensor networks," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 699–710, 2011.
- [9] Z. Li, Y. Liu, M. Li, J. Wang, and Z. Cao, "Exploiting ubiquitous data collection for mobile users in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 312–326, 2013.
- [10] B. Gedik, L. Liu, and P. S. Yu, "ASAP: an adaptive sampling approach to data collection in sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 12, pp. 1766–1783, 2007.
- [11] H. Jiang, S. Jin, and C. Wang, "Prediction or Not? An energy-efficient framework for clustering-based data collection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 6, pp. 1064–1071, 2011.
- [12] W. Lou and Y. Kwon, "H-SPREAD: a hybrid multipath scheme for secure and reliable data collection in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1320–1330, 2006.
- [13] J. Zhu, S. Papavassiliou, and J. Yang, "Adaptive localized QoS-constrained data aggregation and processing in distributed sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 9, pp. 923–933, 2006.
- [14] C. Liu, K. Wu, and J. Pei, "An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 7, pp. 1010–1023, 2007.
- [15] X. Tang and J. Xu, "Adaptive data collection strategies for lifetime-constrained wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 6, pp. 721–734, 2008.
- [16] C. Hung, W. Peng, and W. Lee, "Energy-aware set-covering approaches for approximate data collection in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 1993–2007, 2012.
- [17] R. A. Roseline and P. Sumathi, "Local clustering and threshold sensitive routing algorithm for wireless sensor networks," in *Proceedings of the IEEE sponsored International Conference on Devices Circuits and Systems (ICDCS '12)*, pp. 365–369, March 2012.
- [18] R. A. Rosaline and P. Sumathi, "A decentralised routing algorithm for time critical applications using wireless sensor networks," in *Advances in Computing and Information Technology*, vol. 177 of *Advances in Intelligent Systems and Computing*, pp. 125–133, Springer, Berlin, Germany, 2013.
- [19] R. Khanna and H. Liu, "Machine learning approach to data center monitoring using wireless sensor networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 689–694, December 2012.

- [20] J. Åkerberg, M. Gidlund, T. Lennvall, J. Neander, and M. Björkman, "Efficient integration of secure and safety critical industrial wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, article 100, 2011.
- [21] T. Defeng, L. Shixing, X. Wujun, and Z. Yongming, "A fire monitoring system in ZigBee wireless network," in *Proceedings of the 2nd International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC '10)*, pp. 48–51, October 2010.
- [22] S. Wu and D. Clements-Croome, "Understanding the indoor environment through mining sensory data-A case study," *Energy and Buildings*, vol. 39, no. 11, pp. 1183–1191, 2007.
- [23] A. Boukerche, S. Samarah, and H. Harbi, "Knowledge discovery in wireless sensor networks for chronological patterns," in *Proceedings of the 33rd IEEE Conference on Local Computer Networks (LCN '08)*, pp. 667–673, Quebec, Canada, October 2008.
- [24] W. He, S. Yang, and L. Yang, "Real-time data mining methodology and emergency knowledge discovery in wireless sensor networks," *PGNet-2010*.
- [25] P. P. Rodrigues, J. Gama, and L. Lopes, "Knowledge discovery for sensor network comprehension," in *Intelligent Techniques for Warehousing and Mining Sensor Network Data*, pp. 179–212, 2009.
- [26] C. C. Aggarwal, "An introduction to sensor data analytics, and mining sensor data," in *Managing and Mining Sensor Data*, pp. 1–8, Springer, New York, NY, USA, 2013.
- [27] A. Das, "A novel association rule mining mechanism in wireless sensor networks," in *Proceedings of the 3rd National Conference on Emerging Trends and Applications in Computer Science (NCETACS '12)*, pp. 274–276, IEEE, Shillong, India, March 2012.
- [28] A. Boukerche and S. Samarah, "A novel algorithm for mining association rules in wireless ad hoc sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 7, pp. 865–877, 2008.
- [29] J. Gao, J. Wang, and X. Zhang, "HMRP-based distributed fault detection for wireless sensor networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 640–644, December 2012.
- [30] L. P. Dinesh Kumar, S. Shakena Grace, A. Krishnan, V. M. Manikandan, R. Chinraj, and M. R. Sumalatha, "Data filtering in wireless sensor networks using neural networks for storage in cloud," in *Proceedings of the International Conference on Recent Trends in Information Technology (ICRTIT '12)*, pp. 202–205, April 2012.
- [31] E. Y. Song, K. B. Lee, S. E. Fick, and A. M. Donmez, "An IEEE 1451.5-802.11 standard-based wireless sensor network with embedded WTIM," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC '11)*, pp. 1–6, Binjiang, China, May 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

