Simulation framework for real-time database on WSNs

Ousmane Diallo\textsuperscript{a,b}, Joel J.P.C. Rodrigues\textsuperscript{a,*}, Mbaye Sene\textsuperscript{b}, Jaime Lloret\textsuperscript{c}

\textsuperscript{a} Instituto de Telecomunicações, University of Beira Interior, Portugal
\textsuperscript{b} Department of Mathematics and Informatics, UCAD, Senegal
\textsuperscript{c} Integrated Management Coastal Research Institute, Universidad Politécnica de Valencia, Spain

\textbf{A B S T R A C T}

Wireless sensor networks (WSNs) have been the focus of many research works. Nowadays, because of the time-critical tasks of several WSN applications, one of the new challenges faced by WSNs is handling real-time storage and querying the data they process. This is the real-time database management on WSN and it deals with time-constrained data, time-constrained transactions, and limited resources of wireless sensors. Developing, testing, and debugging this kind of complex system are time-consuming and hard work. The deployment is also generally very costly in both time and money. Therefore in this context, the use of a simulator for a validation phase before implementation and deploying is proved to be very useful. The aim of this paper is to describe the different specificities of real-time databases on WSN and to present a model for a simulation framework of the real-time databases management on WSN that uses a distributed approach. Then, the model of the simulator is described and developed in Java and a case study with some results demonstrates the validity of the structural model.

\textcopyright{} 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Wireless sensor networks (WSNs) may be defined as a set of smart devices, called sensors, which are able to sense, process and transmit information about the environment on which they are deployed. These devices, distributed in a geographical area, collect information for users interested in monitoring and controlling a given phenomenon and transfer them to a sink node. The latter makes the information available to a gateway where remote users can access. So as to obtain those information, users use applications that communicate with the network through queries (Callaway, 2004; Sacks et al., 2003; Baronti et al., 2007).

Systems based on sensor networks are more and more used in many areas providing various types of WSNs (Mendes and Rodrigues, 2010). These different WSNs involved the development of many applications, which are generally connected to databases treating the amount of data collected from sensors. However, the processing time becomes increasingly critical for certain applications. These applications must query and analyze the data in a bounded time in order to make decisions and to react as soon as possible (Diallo et al., 2011). Some examples of the most popular applications are the following: the control of network traffic (Cranor et al., 2002), transactional analysis (web, banking or telecommunication transactions) (Cortes et al., 2000), human motion tracking application (Chen and Ferreira, 2009), the tracking of actions on dynamic Web pages (Zhu and Shasha, 2002; Chen et al., 2000), monitoring of urban or environmental phenomena (Mainwaring et al., 2002; Ulmer et al., 2003), and the sensors data management (Arasu et al., 2003).

Once the sensors perform their measurement, the problem of data storing and querying arises. Indeed, the sensors have restricted storage capacity (Silva et al., 2004) and the ongoing interaction between network devices and environment results in huge amounts of data.

There are two main approaches to data storage and querying in WSN: distributed and warehousing (Bonnet et al., 2000b). The first approach aims at exploiting the capacities of calculation of sensors. Some queries are distributed and evaluated among the nodes into the network. The objective is to locally calculate in order to limit sending of messages, reducing thus the energy consumption. In the second approach, warehousing, one has a centralized system. Collected data from sensors (in stream) are sent to a central database server, in which user queries are processed. Note that this last technique generates large data flows.

The data collected by the WSN must closely reflect the current state of the targeted environment. However, the environment changes constantly and the data are collected in discreet times. So, the collected data have temporal validity, as time advances they become less and less accurate, until the time where they do not reflect the state of the environment (Idoudi et al., 2009a, 2009b). It is fundamental that responses to application queries...
ensure that returned data comply with logic and temporal constraints. In this context, real-time data management on WSNs is necessary to deal with those constraints. The main goal of real-time data management is to ensure temporal consistence data and process transactions within real-time constraints.

Developing, testing, and debugging this kind of complex system are time-consuming and hard work. The deployment is also generally very costly in both time and money. Therefore in this context, the use of a simulator for a validation phase before implementation and deploying is proved to be very useful (Sundani et al., 2010; Nasreddine, 2012; Egea-López et al., 2005).

The main contributions of this paper are the following:

- Description of the different characteristics of real-time databases management on WSN.
- Proposition of a model for a simulation framework of the real-time databases management on WSN that uses a distributed approach.
- Implementation of the model and performance evaluation.

The remainder of this paper is organized as follows: Section 2 presents the Wireless sensor networks. Section 3 discusses the related work, while the characteristics of real-time database management systems are exposed in Section 4. Section 5 briefly presents the data management on WSNs. In Section 6, the description of the real-time database simulation model is provided. In Section 7, some screenshots of the case study is presented to validate the model and Section 8 concludes the paper.

2. Wireless sensor networks

Generally a WSN has a large number of nodes distributed on an interest area and communicating between them so as to measure a physical quantity (e.g., pollution level in a given area) or to do an event monitoring (e.g., vehicles tracking). WSNs are used with different applications in many areas and are very important for applications that should be deployed in places hostile to human interventions (e.g., volcano monitoring). Each network node is considered smart and embeds these units: a sensor unit which provides a measure of environmental data (such as temperature, humidity, pressure, acceleration, sound, etc.), a processing unit, a storage unit, a communication unit, and an energy unit. The communication unit usually performs data transmission by means of radio (Akyildiz et al., 2002, 2007; Oliveira et al., 2011a; Oliveira and Rodrigues Joel, 2011b).

However, these resources are generally very limited, particularly those of storage and energy. The sensor network lifetime depends on the available energy in the nodes composing the network (Lin et al., 2011, 2012). This available energy is consumed by three activities (Akyildiz et al., 2002): sensing activity (data acquisition from the environment), communication (sending and receiving packets), which is essential to form a WSN, and data processing, which consists in some operations applied over data by smart sensors (Cho et al., 2001; Madden et al., 2005). However, the sensing and processing activities are much less expensive in energy consumption than the wireless communication activities (Madden et al., 2002). Therefore, saving energy by optimizing the communication activities is the main point attention in many algorithms designed for sensor networks.

Generally, proposed architectures for WSNs are based on the distribution of sensor nodes in a geographical area in such a way that sensors send collected data to a base station by using the routing techniques like in (Li et al., 2011; Al-karaki and Kamal, 2004).

3. Related work

A lot of simulation tools for WSNs have been already proposed. These simulators are generally designed to meet the development constraints (e.g., communication constraints, constraints on the nodes) related to WSNs and can be divided into classes according to the nature of the specified constraints.

The first class of simulators is the oriented network. They emphasize on the behavior of the wireless network and the protocol stack of the operation. These simulators are based on the computer network simulators designed for computers, such as NS-2 (Zhang et al., 2009), OMNeT (Xian et al., 2008; Varga, 2001), J-Sim (Sobeih et al., 2006), etc. Among these simulators it may be noted: SensorSim (Park et al., 2000), Castalia (Boulis, 2007, 2009), J-Sim sensor simulator (Sobeih et al., 2006), and Prowler (Simon et al., 2003). SensorSim is built on top of a NS-2 802.11 network model and it models a software model of the sensor node and a hardware model. The power models of the hardware components have been implemented and the state of the hardware model changes according to the function performed by the software model. Therefore, the power consumption of the network could be simulated. However, the CPU has not been implemented (Park et al., 2000). Castalia is an application-level simulator for wireless sensor network based on OMNeT++. It is parametric and can be used to evaluate various characteristics for specific applications. In Castalia, multiple simple modules can be linked together and form a compound module, e.g. a sensor node. The simple modules implement the atomic behavior of a model, for instance, the network stack layers of a sensor. Node modules are connected to wireless channel and physical process modules (Boulis, 2007, 2009). J-Sim has been developed at the University of Washington by the National Simulation Resource. It is a general-purpose network simulator modeled after NS-2, but unlike NS-2, J-Sim uses the concept of components, replacing the representation of each node as an object. J-Sim provides support for sensors and physical phenomena. Moreover, the energy modeling, except the radio energy consumption, is provided for sensor networks (Sobeih et al., 2006). Prowler is a probabilistic wireless sensor network simulator running on Matlab environment. It is an event-driven simulator that can be set to operate, either in deterministic mode, or in probabilistic mode in order to simulate the non-deterministic nature of certain component of the network such as the communication channel. Prowler is adapted to the development of algorithms and optimized protocols. However, it does not model the energy consumption of the sensor node (Simon et al., 2003). These simulators are oriented network and they do not deal with the databases management techniques, and particularly the real-time databases management techniques.

The second class of simulators is oriented node. They emphasize on the function of a single node with simple communication models. They are specific to targeted nodes and their operating systems. In addition, these simulators are used to verify the compatibility of a node with a given application. Among these simulators include: TOSSIM (Levis et al., 2003), ATEMU (Karir et al., 2004), and SEN (Sundresh et al., 2004). TOSSIM is a discrete event simulator for TinyOS (TinyOS, 2012) wireless sensor networks. TOSSIM can capture the behavior of the network of several TinyOS nodes at bit granularity. It is designed specifically for TinyOS/NeSc applications to be run on MICA Mote platforms. With TOSSIM, designers can easily translate between running an application in the simulation environment and on motes (Levis et al., 2003). ATEMU is an instruction-level cycle-accurate emulator for WSN written in C language. It simulates programs of nodes with an accuracy clock cycle. ATEMU provides support for the AVR processor, and other function units on the MICA2 sensor node platform, such as the transceiver. Moreover, it provides users a...
complete system for debugging and monitoring the execution of their code via a GUI, called Xatdb (Karir et al., 2004). Sensor Environment and Network Simulator (SENS) is a wireless sensor network simulator with a layered architecture, modular. The components of this architecture are customizable and they model the application layer, network communication, and physical environment. SENS allows realistic simulations by using values from real sensors to represent the behavior of component implementations (Sundresh et al., 2004).

All these simulators do not fit within the real-time databases on WSNs. To the best of our knowledge, there is not a tool that takes into account the specific characteristics of the models of real-time databases for WSNs. However, true that our simulator is not a full end-to-end simulator like for instance OMNeT or any other full end-to-end simulator tool with all protocols of WSNs, it is based on a model that focuses on the real-time databases management on WSNs, which are by their characteristics and constraints very difficult to study and model.

4. Characteristics of data and transactions in real-time database systems

Like a traditional database management system (DBMS), a real-time DBMS (RT-DBMS) must process transactions and ensures that the logical consistency of the data is not violated. However, unlike a traditional DBMS, a RT-DBMS emphasizes on the temporal validity of the data and the time constraints or deadlines for transactions (Idoudi et al., 2009a, 2009b; DiPippo and Wolfe, 1997).

The main purpose of a RT-DBMS is to process transactions on time, while maintaining logical and temporal consistency of data. The temporal consistency expresses the need to maintain consistency between the current state of the targeted environment and the state as reflected by the database contents. The temporal consistency can be measured in two ways (Ramamritham, 1993):

- Absolute consistency, which deals with the need to maintain the view representing the state of the targeted environment consistent with the real state of the environment.
- Relative consistency, which concerns data derived from other ones.

To satisfy these temporal constraints, the structure of the data must include these attributes: (i) timestamp, which indicates the instant when the observation relating the data was made; and (ii) absolute validity interval (avi) that denotes the time interval following the timestamp during which the data are considered valid. Another attribute can be considered: the imprecision or data error (DE), which refers to how the current state of the targeted environment may differ from the measured data (Ramamritham, 1993). The data error on a data version d is defined by:

$$DE(d) = \frac{100 \times (current\ value\ (d) - update\ value\ (d))}{current\ value\ (d)} \%$$

There are two kinds of transactions in real-time databases: update transactions and user transactions. Update transactions are executed periodically to update real-time sensor data, or sporadically to update the derived data in order to reflect the state of the real world. Derived data are the data computed using sensor data. User transactions, representing user queries, arrive aperiodically. They may read or write non real-time data, but only read real-time data (Idoudi et al., 2009a, 2009b). These update transactions also to comply with temporal constraints must have these following attributes: (i) liberation time that represents the moment on which all the resources for the transaction processing is available; (ii) computing time that indicates the execution time needed for the transaction; and (iii) maximum time, which indicates the maximum time limit for the transaction execution and the periodicity that refers to the frequency with which the transaction happens (Chagas et al., 2010).

To satisfy the logical consistency of the data, transactions must be processed with Atomicity, Consistency, Isolation and Durability (ACID) properties. But unlike the conventional databases, in real-time databases these properties are relaxed. Firstly, the atomicity may be relaxed. It is only applied to the sub-transaction that wants to deal with completely data consistency. Secondly, since timeliness is more important than correctness, in many situations, correctness can be traded for timeliness. Thirdly, the isolation allows transactions to communicate with others to better perform control functions. Similarly, in real-time databases, not all data must be permanent and some of them are temporal (DiPippo and Wolfe, 1997; Chagas et al., 2010; Ramamritham, 1993).

5. Data management on wireless sensor networks

Among the activities of wireless sensor nodes, the communication activities are more expensive in terms of energy consumption (Madden et al., 2002, 2005). Therefore, saving energy by optimizing the communication activities is the main point attention in many algorithms and architectures designed for sensor networks (Madden et al., 2005). There are two main approaches to data storage and query in WSNs (Bonnet and Seshadri, 2000; Bonnet et al., 2000): the warehousing approach and the distributed approach.

The warehousing approach is a centralized system (Bonnet et al., 2001). The data gathered by the sensors are sent to a central database server where user queries are processed. In this case, the sensors act as simply collectors. The warehousing approach is the most used one in data storage and query processing. However, it has some drawbacks, such as, the huge number of generated data can easily create a bottleneck on the central server; and the huge amount of information transferred wastes the resources. The wireless data transmission decreases consequently the WSN lifetime. Indeed, it is shown in (Madden et al., 2002) that the wireless data transmission is more expensive in terms of energy consumption than the data processing and the sensing activities. Moreover, it is clear that this approach is not adequate for non-historical queries because it involves time delay for the results.

In the other hand, the distributed approach exploits the capabilities of sensors calculation and the sensors act as local databases (Bonnet and Seshadri, 2000). It aims to locally calculate in order to limit sending of messages, reducing thus the energy consumption in the network. In this approach, the data are stored in a central database server and in the devices themselves, enabling one to query both. Here, the devices act as part of the database. This approach can offer several advantages: on sensors, the processing of queries is done on quasi real-time. In fact, query processing inside the devices themselves means that the most current data will be acquired. The data will certainly arrive to the users with their temporal validity. It supports long-running queries and instant queries. Furthermore, the distributed approach increases the lifetime of the network by reducing the communication activities among the devices.

There are three sorts of queries in sensor networks: (i) historical data queries, which are run against the server; (ii) instant queries, which are run against a device in an instant of time; and (iii) long-running queries, which refer to queries run against a device during a time interval (Bonnet and Seshadri, 2000; Neto et al., 2004).
6. Real-time database simulation model for WSNs

The design of the simulator involves the development of a library of models from which it will be possible to build our architecture of real-time databases management for WSN. These models are descriptions of behaviors and functions supported by each component of the system and can replace, in the simulation, the actual components of the architecture to be modeled.

Many researchers argue that the object-oriented model is more suitable than the relational model (Kim, 1995) to model real-time data, because of the nature of several real-time applications, which handle complex real-world objects with time constraints. Thus, many projects on real-time databases have chosen the object-oriented model for their systems (Terrier et al., 1997; Wolfe et al., 1997). However, the relational model (Madden et al., 2005) is the most used on real-time or distributed databases modeling.

The global architecture of our simulator model of real-time database management techniques on WSNs (see Fig. 1) uses the distributed data storage and the query approach and is inspired by the works made in (Neto et al., 2004). Thus, our model considers the gateway more general by integrating into it a database warehousing (DBW) that stores real-time data and a real-time database management system (RT-DBMS) that handles real-time transactions. The model of the wireless sensor network (WSN) nodes is built by using an event-based model, particularly an event-advance design (Granat, 2006). Thus, the nodes are programmed as follow: each time an event (user transaction) occurs in the network; all the nodes are triggered to compute the previous acquisitions, store the sensor data, and send all the underlying periodic update transactions to the database server. Moreover, each node targeted by the user transaction at the same time, after completion of its previous works, sends the results on required time. All these actions are scheduled on time based on the instant time that the network was initialized, the periodicity of acquisition and database update transactions, and the instant time that the user transaction arrives in the network. Therefore, the devices act as local databases, including a software component that manages the data storage and user queries with time constraints. The database server incorporates also an application that deals with time-constrained data and time and logic constrained transactions. We distinguish two kinds of transactions: update transactions and user transactions. The update transactions can be local updates for periodically updating the local sensor databases or server updates that periodically update the real-time data in the database server. The user transactions are real-time user queries that only read real-time data or historical queries that may read or write historical data.

Below, the description of the behavioral models of various elements that compose the simulator model of real-time database management techniques for WSNs is made, by explaining clearly the principle of object-oriented modeling used in the simulator developed in Java.

6.1. Wireless sensor network

The WSN is composed by smart sensors (Ruiz et al., 2004), which are able to sense, process and transmit information about the environment on which they are deployed. These sensors act as local databases, named sensor database (SDB). Thus, the WSN forms a distributed database. An illustration of a generic sensor model with its basic components is given in Fig. 2.

The sensor unit is formed by the sensor and the Analog-to-Digital Converter (ADC). The sensor senses a change in the physical condition and transforms it to electrical signal. This signal is converted by the ADC to digital data, which can be stored in the memory and processed by the processor. The processor and the memory compose the treatment unit. The data communication into the network is performed by the transceiver located in the
communication unit. This communication can be performed with or without wire. In the case of wireless communication, the most used technique is the radio frequency transmission (Adelstein et al., 2005). The batteries are used as energy sources.

For modeling our generic sensor as a local database, we use an object-oriented modeling as follow: a SensorNetworkNode class encapsulates all the properties and functionalities of the device. Thus, the measured physical phenomenon, of course, in our case study uses temperature measures, is modeled by the nextGaussian() function of the class Random that will generate pseudo-random available temperature measures (Deshpande et al., 2004) with best precision. This function is used in the acquisition phase (acquisition() function in our model). However, depending in the case study, others methods can be used, for example, a memory that contains the measures of the physical phenomenon (Nasreddine, 2012). Depending of the sampling frequency of the local updates, the acquisition() function is called in each period and the temperature value returned by the nextGaussian() function is stored in coherent way in the memory (list of sensor data) of the sensor that stores sensor data. Moreover, functions that periodically send the acquired data, according to the periodicity of the database updates, to the database server are implemented. Concerning the other components of the sensor model, functions that receive queries, process them in real-time according to their time constraints, and send the data to the users are added. Thus, these functions model the treatment unit and the communication unit. For the energy consumption, functions are also used to retrieve an amount of energy from the global energy remained on the sensor according to the data sensing, processing, transmission, and the sensor’s duty cycle operation. Formula (2) below gives the energy consumption of the sensor node. For more details about the calculation of each term of the formula, readers can refer to (Briand et al., 2010).

\[ E_{\text{node}} = E_{\text{sensor}} + E_{\text{ADC}} + E_{\text{c}} + E_{\text{transceiver}} \]  

(2)

To capture the actual energy consumption of each sensor’s activity in the simulation, each activity needs to be calibrated by real measures. For that, the current consumptions of the main sensor’s activities are configured according to the TelosB datasheet (CrossBow Inc, 2013; Nguyen et al., 2011), as presented in Table 1. In addition, the simulation used AA-size batteries with each nominal ampere-hour capacity of 2.4 Åh. However, the values of the parameters (the initial battery capacity according to its type, the cost of the operations with respect to the sensor type, etc.) can be configured in the main configuration file.

In real-time databases, the data should comply with logic and temporal constraints. Thus to comply with temporal constraints, the structure of the data should have these properties: (i) timestamp, (ii) absolute validity interval (avi), and the imprecision (Ramamritham, 1993). All these attributes include the value of the data and the location-identification of the sensor that acquires the data are encapsulated in our model of sensor data with a class named SensorData, which implements the data acquired by a sensor. Because of the temporal constraints of sensor data, they become quickly invalid and should be periodically updated. Thus to deal with these constraints, transactions have logical and temporal constraints. In the sensors we distinguish two kinds of transactions: the local update transactions and the read-only transactions. The local update transactions, represented in our implementation by the LocalUpdate class, represent a new acquisition of a data that will be stored in the sensor database. These transactions are periodic and firm (Pailler, 2006; Neto et al., 2004) and they are executed by call of the acquisition() function. The read-only transactions are the transactions sent by the user applications to read the local databases. In our case study, the transactions are the instant query transactions and the long-running transactions. They can be executed by calling the instantQueryExecution() and longRunningQueryExecution() functions respectively. These kinds of transactions are aperiodic.

As already mentioned above, the processing model in the sensor nodes is built by using an event-based model, particularly an event-advance design (Granat, 2006). Thus, the nodes are programmed as follow: each time an event (a read-only transaction) arrives in the network, all the nodes are triggered to compute the previous acquisitions, store the sensor data, and send all the underlying periodic update transactions to the database server. Moreover, each node targeted by the user transaction at the same time, after completion of its previous works, sends the results on required time. All these actions are scheduled on time based on the instant time that the network was initialized, the periodicity of acquisition and database update transactions, and the instant time that the user transaction arrives in the network.

The implementation of this event-advance design technique is done as follow: we know the instant time that the network was initialized, let call it initialize-time. We know the arrival time of the user query, let call it now, and finally we know the periodicity (let call it T) of the local update transactions. Thus, the difference between now and initialize-time, divided by the periodicity (T) gives us the number of local updates (let call it Q) that the sensor should perform before processing the query. Finally, having the release time (the date on which the first instance of the query should be activated) included in the local update transaction, let call it r, we can compute the activation dates for each instance of the local update transaction with this formula (Pailler, 2006):

\[ r_k = r + (k-1)T \]  

(3)

The same technique is used to calculate the activation date for all the instances of the server update transactions (explained hereafter) that should be done before the arrival of user transactions. Thus, by scheduling all the instances of the local and server update transactions, and the user transaction with respect to their activation dates, we can manage the data storage and transaction processing with respect to the logic and time constraints of data and transactions.

6.2. User transactions generator

The transactions generator is a component, which is responsible for generating the user transactions aperiodically. Thus, it is composed of an interface where the user transactions will be parameterized to be sent into the system and a pseudo-random function that provides the time intervals representing the sending moments of transactions. As mentioned above, the properties of a specific transaction are also encapsulated into a specific transaction class. For example using the transaction class, a long-running query is parameterized as: Transaction < idTrans=12 startTrans="Dec 03, 2012 11:09:34" perTrans=1 time unit (t.u.), idLocSensor=0...5, endTrans="Dec 03, 2012 11:09:40">. In this example, the query

**Table 1**

Nominal power consumption of TelosB nodes (Nguyen et al., 2011).

<table>
<thead>
<tr>
<th>Components</th>
<th>Mode</th>
<th>Current draw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module</td>
<td>Active</td>
<td>1.8 mA</td>
</tr>
<tr>
<td></td>
<td>Sleep</td>
<td>5.1 µA</td>
</tr>
<tr>
<td>RF-Transceiver</td>
<td>Receive</td>
<td>19.7 mA</td>
</tr>
<tr>
<td>Transmit(0 dBm)</td>
<td></td>
<td>17.4 mA</td>
</tr>
<tr>
<td></td>
<td>Sleep</td>
<td>0.01 mA</td>
</tr>
</tbody>
</table>
6.3. Gateway

The gateway usually is used to make the connection between the wireless sensor network and the external networks (e.g., IP). It has the same power as a PC. In our model, the gateway is more general by including a database warehousing (DBW) that stores real-time data and a real-time database management system (RTDBMS) that handles in real-time the transactions.

6.3.1. RT-DBMS

The admission controller (AC) has the task of controlling user transactions that are accepted in the system. Thus, the user transactions that represent instant or long-running queries will be sent to the WSN, whereas the real-time user transactions and the historical queries will be directed towards the server database.

In that phase, we consider now the real-time user transactions that arrive aperiodically to read the real-time data stored in the database server, modeled by the $R T r e a d T r a n s a c t i o n$ class and the real-time update transactions that arrive periodically from the WSN for updating the real-time data to the database server, modeled by the $S e r v e r U p d a t e$ class (Idouidi et al., 2009b). Thus, the real-time user transactions include among other properties their computing time (Pailier, 2006) and the real-time update transactions have among other attributes (i) the liberation time, (ii) the computing time, (iii) the deadline, and (iv) the periodicity (Chagas et al., 2010).

The real-time user transactions admitted in the system and the server update transactions from the sensors are placed in a waiting queue before being sent to the transaction manager. All these transactions in wait are scheduled by the scheduler according to their priorities, particularly their deadlines. In fact, we use the Earliest Deadline First (EDF) protocol (Liu and Leyland, 1973), which orders the transactions according to the principle that the transaction that has the nearest deadline must be executed in priority.

The transaction manager consults regularly the waiting queue, updates it, and checks the validity of transactions (i.e., transactions that have not yet missed their deadline in relation to the current time). The transactions that have lost their deadline are aborted, the others are still handled by the transaction manager for a possible execution: the freshness manager checks the freshness of data that will be accessed by a transaction. If the data is outdated then the transaction is waiting in a queue. Since several transactions can access the data stored on the database server, a concurrency control protocol is needed to deal with the cases in which concurrent transactions may lead to inconsistent data in the database. In our model we use the Epsilon Serializability (ESR) criterion (Pu, 1991) to deal with concurrent transactions. Indeed, the Epsilon Serializability (ESR) criterion is well adapted in the processing of firm (the deadline of the transaction must be met, otherwise the transaction is rejected) real-time transactions (Bouazizi et al., 2004).

The ESR relaxes the severity of the classic serializability (SR) in the transaction processing by allowing a limited inconsistency in the database. This limited inconsistency is automatically maintained by the divergence control algorithms (DC), in the same way that the SR is managed by the concurrency control techniques (CC). In our model, we use the DC algorithm with two phases (Bouazizi et al., 2004) named 2PLDC, which is an extension of the algorithm of 2PL concurrency control of the classic serializability.

In this algorithm, the update transactions called write Epsilon-transaction ($E T w$) export an inconsistency in the database, while user transactions called read Epsilon-transaction ($E T r$) import an incoherence from the base. These inconsistencies are gathered in two accumulators and the transactions can continue their execution as long as these accumulators do not exceed a certain given limit value.

Thus, in addition to the attributes mentioned early in this paragraph, the real-time update transaction becomes an $E T r$ and incorporates two other attributes in our model: an attribute that records the total amount of inconsistency exported in the database and an attribute that records the maximum imprecision that the real-time update transactions ($E T w$) can export in the database. These attributes are modeled respectively by $a c c u m u l a t e I m p o r t$ and $l i m i t I m p r o n c e$ in the model. Similarly, the real-time user transaction becomes an $E T r$ and has two more attributes: an attribute that records the total amount of inconsistency imported by the $E T r$ modeled by $a c c u m u l a t e I m p o r t$ and an attribute that records the maximum imprecision that the $E T r$ can import modeled by $l i m i t I m p r o n c e$ in the model. All such characteristics will permit to deal with both the time and logic restrictions.

This divergence control algorithm is implemented to allow conflicting transactions (Read/Write in the serializability classic) to execute concurrently in every way that their scheduling does not lead an imprecision that is higher than the one accepted in the data. Thus, when an $E T r$ wants to access a data being running by an $E T w$, we increment the value of the variables $a c c u m u l a t e I m p o r t$ and $a c c u m u l a t e E x p o r t$, respectively of the $E T r$ and the $E T w$ and verify if

$$ac c u m u l a t e I m p o r t \leq l i m I m p r o n c e$$

(4)

$$ac c u m u l a t e E x p o r t \leq l i m I m p r o n c e$$

(5)

In the same fashion, when an $E T w$ demands to access a data being executed by several $E T r$, all the $E T r$ increments by one unit their $a c c u m u l a t e I m p o r t$ and verify if formula (4) is true. However, the $E T w$ should verify its $a c c u m u l a t e E x p o r t$ regarding the incrementation by the number of $E T r$ in conflict with it. If condition (5) is false or if any $a c c u m u l a t e I m p o r t$ of one $E T r$ exceeds its $l i m I m p r o n c e$, then we make wait or abort the transaction which demands to access the data in processing. Thus, this mechanism allows several real-time transactions to meet their deadlines.

6.3.2. Database warehousing

The sensor data are modeled by the $S e n s o r D a t a$ class, which encapsulates these attributes: the location identification of the sensor that acquires the data, the timestamp, absolute validity interval (avi), the imprecision, and the value of the data. Thus, the database warehousing (DBW) is a list of a sensor data and is modeled by the $M e m o r y C l a s s$. The size of this memory can be fixed on the beginning of the simulation experience. The coherency of the DBW is maintained by the real-time update transactions and the Epsilon Serializability (ESR) criterion. This allows to simulate a real-time database as in the real databases.

Fig. 3 illustrates the class diagram of the implementation of the whole simulation framework.

7. Performance evaluation of a simulator model for real-time databases management techniques in WSN based on a distributed approach

The model considers a WSN where the sensor nodes, by using an event-based design, receives user queries, computes the previous acquisitions of temperature, sends all the underlying periodic update transactions to the database server, and process user queries with respect to time and logic constraints as explained...
early. The RT-DBMS receives the real-time update transactions and processes them according to logic and time constraints of data and transactions. Thus, the RT-DBMS schedules the transactions with respect to the Earliest Dead-line First (EDF) protocol (Liu and Leyland, 1973) and uses the Epsilon Serializability criterion to deal with concurrent transactions.

The execution of the simulator is made by setting up some parameters according to a particular experiment. Indeed, the simulator is implemented in Java programming language and in the file containing the MainSimulation class one can fix almost all the parameters of the simulation, e.g., the number of sensor nodes composing the network. Each sensor node has to be configured also in this file by setting up the attributes (identification of the sensor location, the maximum power of the batteries, the parameters of the local and sever updates) of the SensorNetworkNode class. In addition, the practitioner has to configure here in this file by setting up the attributes (identification of the transaction location, the maximum number of user transactions (instant query transactions and long-running query transactions) and their characteristics to be sent to the network nodes. The real-time user transactions (RTreadTransaction class) that want to access the real-time data storing in the database server has to be configured also in that file. Thus for analysis the validity of the model, five (5) sensors forming the sensor network were considered in the configuration file. Table 2 summarizes the configurations of the local updates (periodic acquisitions) and the server updates (periodic updates of the database server) of each sensor node.

The parameters in the above table are specific to those transactions. The transactions are like the Query Language for Real-Time Database (QL-RTDB) (Chagas et al., 2010) with properties encapsulated in a class as already explained early in the paper. This class, depending of the transaction type, is composed of attributes that indicate the identification of the transaction (idTrans), the generation time (Li), the computational time (Ct), the periodicity (Pr), the deadline (Di), the total amount of inconsistency exported in the database or imported from the database (accumu-Export/Import), the imprecision limit (Lim-Imp/Exp), and the location identifications (idLocSensor) of the temperature sensors targeted by the transaction. The syntax is like: Transaction < idTrans, Li, Ct, Pr, Di, accumu-Export/Import, Lim-Imp/Exp, idLocSensor0..idLocSensorN >.

This experiment considers also two real-time user transactions (rt-uT): a real-time user transaction with an identification 500, a computational time 6 t.u., a deadline 12 t.u., and a real-time user

![Fig. 3. Class diagram of the model.](image-url)

### Table 2

Summary of the configuration parameters of the sensors.

<table>
<thead>
<tr>
<th>Server updates</th>
<th>Li In time unit</th>
<th>Ct</th>
<th>Di</th>
<th>Pr</th>
<th>Li</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>idLocSensor</td>
<td>(t.u.)</td>
<td>(t.u.)</td>
<td>(t.u.)</td>
<td>(t.u.)</td>
<td>(t.u.)</td>
<td>(t.u.)</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
<td>18</td>
<td>18</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>14</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>15</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>12</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>
transaction with an identification 600, a computational time 5 t.u., a deadline 12 t.u. The user transaction with an identification 500 demands to read a data sent by the sensor in location 0 from the real-time database server and the user transaction with an identification 600 wants to read the data item sent by the sensor 1. These transactions are aperiodically sent towards the database server.

According to the Epsilon Serializability criterion, the server update transactions and the real-time user transactions include respectively an attribute that stores a value representing the amount of inconsistency exported in the database and an attribute that stores a value representing the amount of inconsistency imported from the database. The structure of these transactions has also a value that represents the limit of imprecision exported in the database and a value that represents the limit of imprecision imported from the database. For more efficient analysis, these values are generated by a random function in the model.

Depending on the parameters and transaction used, the results are given by numerical values.

### 7.1. Case where an instant query is sent to the network to query sensor data

We first analyze the case where an instant query is sent in the network. The properties of these queries are encapsulated in class that is composed of attributes that indicate the identification of the query and the locations of the temperature sensors that the user wishes to query. The syntax is like: Transaction <idTrans, idLocSensor0..idLocSensorN>. As the case study, the instant query <idTrans=0, idLocSensor0..idLocSensor1> gives the results in Table 3.

An instant query is a query that is run against a device in an instant of time. Thus, according to Table 3, the query arrives in the network at Oct 31, 2012 23:17:22 and is immediately answered by the targeted sensors. By using an event-advance desing, the sensor nodes are based on formula (3) and compute all the previous local updates and send all the underlying server updates towards the database server before returning the required data with their temporal parameters. Thus, the simulation model shows also for each sensor the number of local updates and server updates to compute before processing the query. Here, only the sensors in locations 0 and 1 reply because they were targeted by the query. However, the remaining sensors perform their local and server updates.

Once the server updates (sup) arrive in the Gateway, they are handled with the real-time user transactions (rt-uT) by the RT-DBMS with respect to time and logic constraints (see Table 4).

Table 4 presents the sequence of both real-time user transactions and real-time update transactions in the server database. It is also possible to observe the time properties for each transaction, the sensor performing the updates or the sensor data targeted by the user transaction, the parameters to deal with the Epsilon Serialisability, and the data value added or read including its eventual error. The liberation times (Li) indicate the executions of the transactions according to their periods. Considering the quality of service management, the model uses the Earliest Deadline First (EDF) protocol (Liu and Leland, 1973) to schedule the transactions and performs the Epsilon Serializability techniques to deal with concurrent transactions. Formulas (4) and (5) are used here to allow or forbid the simultaneous execution of transactions. Thus as illustrated in the table, the user transaction 500 allows to read the data with an imprecision 0.241%, while the user transaction 600 does not allow any imprecision in the read value.

The model allows also to compute the energy consumed in the network, see Table 5. Table 5 above presents the percentage of energy consumed and remaining energy for each sensor node according to its activities performed. These activities are mainly data sensing, processing, communication, and the sensor's duty cycle operation. In the table, one can observe that the sensors in locations 0 and 4 have spent most of the energy. This is due to the fact that they perform more server updates (see Table 3), leading then to more data transmission. In fact generally in WSNs and with respect to the TelosB datasheet (see Table 1), the wireless communication consumes more energy than the other activities performed by the sensors. The energy consumed by the sensor 0 is the biggest because, regarding the sensor 4 of which it has the same number of local updates and server updates, it executes the instant query and sends the result to the base station. This leads to more energy consumption. The energy consumption is computed based on formula (2) in (Briand et al., 2010).

### 7.2. Case where a long-running query is sent to the network to query sensor data

The configurations of the network and the two real-time user transactions addressed to the database server are not changed. However, we sent a long-running query to trigger the process

---

**Table 3**

Summary of the results of the instant query that queries the temperature sensors in locations 0 and 1.

<table>
<thead>
<tr>
<th>IdLocSensor</th>
<th>Access date</th>
<th>Returned data (°C)</th>
<th>AVI (t. u.)</th>
<th>Timestamp</th>
<th>Number of local updates (sup)</th>
<th>Number of server updates (sup)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>X X X X X X X X X</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X X X X X X X X X</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X X X X X X X X X</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4**

Summary of the sequence of executions of real-time transactions in the database server.

<table>
<thead>
<tr>
<th>idTrans</th>
<th>idLocSensor</th>
<th>Li</th>
<th>Ct (t.u.)</th>
<th>Pr (t.u.)</th>
<th>Di (t.u.)</th>
<th>LimExp</th>
<th>LimImp</th>
<th>Value (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 (sup)</td>
<td>4</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>X</td>
<td>21.152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000 (sup)</td>
<td>0</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>X</td>
<td>27.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4000 (sup)</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>4</td>
<td>X</td>
<td>21.914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000 (sup)</td>
<td>3</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>X</td>
<td>26.452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9000 (sup)</td>
<td>1</td>
<td>4</td>
<td>18</td>
<td>8</td>
<td>X</td>
<td>28.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000 (sup)</td>
<td>4</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>X</td>
<td>21.151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000 (sup)</td>
<td>0</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>X</td>
<td>28.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 (rt-uT)</td>
<td>0</td>
<td>6</td>
<td>X</td>
<td>6</td>
<td>X</td>
<td>28.032 ± 0.241%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>600 (rt-uT)</td>
<td>1</td>
<td>5</td>
<td>X</td>
<td>12</td>
<td>X</td>
<td>28.033 ± 0.0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of the network. A long-running query refers to a query runs against a device during a time interval (Bonnet and Seshadri, 2000; Neto et al., 2004). The characteristics of the query are encapsulated in class that is composed of attributes that indicate the identification of the query (idTrans), the initiate date of the query execution (startTrans), the period of the data sampling (perTrans), the terminate date (endTrans) of the query execution, and the locations of the temperature sensors (idLocSensor) that the query targets. The syntax is: Transaction < idTrans, startTrans, perTrans, idLocSensor0..idLocSensorN, endTrans >. For the case study, the long-running query parameters are: < idTrans=2, startTrans=“Sep 03, 2012 11:09:34”, perTrans=1 t.u., idLocSensor=0.01, endTrans=“Sep 03, 2012 11:09:37” >. The results are summarized in Table 6.

Table 6 illustrates the execution of the long-running query, which returns the temperature values stored in the sensors 0 and 1. This query samples the temperature values each second from “Sep 03, 2012 11:09:34” to “Sep 03, 2012 11:09:37”. Moreover, all the nodes perform their local updates and sends the update transactions towards the database server.

8. Conclusion and future work

This paper presented a simulation framework for real-time database management on wireless sensor networks. Thus, the description of the different characteristics to design real-time databases on WSN has been done and we proposed a model for a simulation framework of real-time databases on WSN that uses a distributed approach.

The distributed approach was chosen considering its advantages, unlike the warehousing approach. Thus, with the distributed approach the queries could be processed inside the network that diminishes the data flow in the network and increases the lifetime of the network. This is due to the fact that the communication activities use more power than the local process activities. Moreover, by using the distributed approach it is possible to process instant-queries and long-running queries, which are quasi-real-time queries.

In real-time database management on wireless sensor network, the data and transactions have temporal restrictions. For that, the proposed model uses the Earliest Dead-line First (EDF) protocol to schedule transactions and the Epsilon Saliency techniques to allow conflicting transactions to execute simultaneously in way that their scheduling does not lead an imprecision that is higher than the one accepted in the data. Unlike to usual sensor network simulators oriented network that study the network in the point of view communication or simulators oriented node, which study the functions of a single node, this simulator model emphasizes on the real-time database management techniques for WSNs that use a distributed approach. Thus, this model integrates all the components of a sensor network and the characteristics of real-time database techniques. The full model has been implemented in Java. The object-oriented model was chosen because it is more suitable to model the complex real-world objects with time constraints than the relational model (Kim, 1995). Moreover, the simulator is configurable and the protocols may be changed as required.

A case study shows the execution of real-time transactions and the energy consumed in the network that demonstrates, thus, the validity of the model.

We are currently working on a probabilistic study of real-time transactions with an architecture of real-time database for WSN that uses a distributed approach. This study should allow us to optimize the performance of real-time transaction executions and network resources.

Acknowledgments

This work has been partially supported by the Instituto de Telecomunicações, Next Generation Networks and Applications Group (NetGNA), Portugal, by National Funding from the FCT.


