Indian J. Innovations Dev., Vol. 1, No. 3 (Mar 2012) ISSN 2277 - 5390

Design, implementation and evaluation of SemHD: a new semantic hierarchical sensor data storage

Gheisari M

Department of Computer Science and Research Branch, Islamic Azad University, Tehran, Iran
mehdi.gheisari61@gmail.com

Abstract

Sensor networks are dense wired or wireless networks for collecting and disseminating environmental data. Sensors enable machines to capture and observe characteristics of physical objects and features of natural incidents. Most of the current efforts on sensor networks are focused on networking and service development for various applications, but less on processing the emerging data. Sensor networks generate immense amount of data, which requires advanced analytical processing and interpretation by machines. Processing and interpretation of huge amounts of heterogeneous sensor data and utilizing a coherent structure for this data is an important aspect of a scalable and interoperable sensor network architecture. In this paper, we describe a new semantic hierarchical sensor data storage named SemHD, which arranged sensors in hierarchical form and each sensor send their data to cluster head in semantic model.

Key Words: Sensor Networks; Ontologies; Knowledge Modeling; SensorML; SemHD; Sensor data storage.

Introduction

Many sensor network applications that are related to pervasive computing, e.g., monitoring learning behavior of the children, senior care system, environment sensing, etc, generate a large amount of data continuously over a long period. Often, the large volumes of data have to be stored somewhere for future retrieval and data analysis. One of the biggest challenges in these applications is how to store and retrieve the collected data (Singh *et al*. 2008). We store sensor data's in a form that machines can collect and understand the data provided by the various types of sensors and networks. Section-2 describes background studies, semantic web technologies that include XML (Extensible markup language), RDF (Resource Description Framework), SWE (Sensor Web Enablement) and ontology. In section-3, we describe a novel sensor data storage based on SWE standards, we use a universal language to provide semantic data modeling for sensor networks. In section-4, we say the state-of-the-art of sensor data storage discussion. Section-5 concludes the paper and discusses the future work.

Background

In this section, we review some import ant background knowledge.

Semantic web

Semantic web is an extension to the current web in which the meaningful relationships between resources is represented in machine process able formats (Swartz, 2007). The main idea in the semantic web is to provide well-defined and machine accessible representation of the resources and their relationships rather than simple links as they are offered by the link structure on the current web (i.e. ref links in HTML).

The World Wide Web Consortium (W3C) has defined different standards for representing the semantic web data in machine accessible and process able formats. The primary technologies for the semantic web include the Extensible Markup Language (XML), Resource Description Framework (RDF), RDF Schema (RDF-S), and the Web Ontology Language (OWL).

Extensible Markup LANGUAGE (XML)

XML is actually a set of syntax rules for creating semantically rich markup languages in a particular domain. The fundamental construct in an XML document is the element. An element is simply a pair of matching start- and end-tags, and all the text that appears between them (Akmal *et al*. 2003).

XML documents must have a single root element that encompasses all other elements in the document. Elements may have sub elements nested within them, to any level of nesting. Elements may also have attributes. The following example shows an XML document.

```
<Account >
<account-number> A-101 </account-number>
<branch-name> Downtown </branch-name>
<balance> 500 </balance>
</Account>
```

This example generates an account with account number A-101, its branch name is Downtown, and amount of its balance is 500.

Resource Description Framework (RDF)

At the simplest level, the resource description framework is an XML-based Language to describe resources. While XML documents attach *meta* data to parts of a document, one use of RDF is to create *meta* data about the document as a standalone entity. The RDF is a framework that allows data within a domain to be linked through named relationships. An RDF graph is encoded as a set of subject-predicate-object triples, which resemble the subject, verb, and object of a sentence. The subject and object are nodes in the graph and the predicate is a directional named link between the subject and object. This simple triple structure turns out to be a natural way to describe a large majority of the data processed by machines. A Universal Resource Identifier (URI), an address just like that used for web pages, identifies each the subjects, verbs and objects.

Thus, anyone can define a new concept, or a new verb, by defining a URI for it on the Web.

Sensor Web Enablement (SWE)

The Open Geospatial Consortium recently established the Sensor Web Enablement as a suite of specifications related to sensors, sensor data models, and sensor web services that would enable sensors to be accessible and controllable via the web.

The core suite of language and service interface specifications includes the following:

(1) Observations & Measurements (O&M) - Standard models and XML Schema for encoding observations and measurements from a sensor, both archived and real-time.

(2) Sensor Model Language (SensorML) - Standard models and XML Schema for describing sensors systems and processes; provides information needed for discovery of sensors, location of sensor observations, processing of low-level sensor observations, and listing of task able properties.

(3) Transducer Model Language (TransducerML) – Standard models and XML Schema for describing transducers and supporting real-time streaming of data to and from sensor systems.

(4) Sensor Observations Service (SOS) - Standard web service interface for requesting, filtering, and retrieving observations and sensor system information. This is the intermediary between a client and an observation repository or near real-time sensor channel.

The following example shows a timestamp encoded in O&M and semantically annotated with RDFa.

The timestamp's semantic annotation describes an instance of *time: Instant* (here, *time* is the namespace for OWL-Time ontology):

```
<swe:component rdfa:about=“time_1”
rdfa:instanceof=”time:Instant”>
<swe:Time rdfa:property=“xs:date-time”>
2010-0308T05:00:00
</swe:Time>
```

</swe:component>

This example generates two RDF triples. The first, *time_1 rdf:type time:Instant*, describes *time_1* as an instance of *time:Instant* (subject is *time_1*, predicate is *rdf:type*, object is *time:Instant*). The second, *time_1 xs: date-time "2010-03-08T05:00:00,"*describes a data-type property of *time_1* specifying the time as a literal value (subject is *time_1*, predicate is *xs:date-time*, object is *"2008-03-08T05:00:00"*).

Ontology

Ontologies are typically defined as an abstract model of a domain of interest with a formal semantics in the sense that they constitute a logical theory. These models are supposed to represent a shared conceptualization of a domain as they are assumed to reflect the agreement of a certain community or group of people. In the simplest case, ontologies consist of a set of concepts or classes, which are relevant for the domain of interest, as well as a set of relations defined on these concepts. The general idea is that data and services are semantically described with respect to ontologies, which are formal specifications of a domain of interest, and can thus be shared and reused in a way such that the shared meaning specified by the ontology remains formally the same across different parties and applications. Ontologies are utilized by the semantic Web Applications to offer conceptualized representation of domains and to specify meaningful relationships between the resources (Barry Burd, 2002). Ontologies provide a common and shared understanding of different domains. OWL is a language that is based on description logic and facilitates construction of ontologies for different domains. The OWL representation of data enables expression of semantics and meaningful relationships between resources and amongst different attributes of complex data.

Software agents for reasoning and inference purposes are to enable systems to derive additional knowledge from the represented data can access the OWL data. There are common query languages such as SPARQL available for the OWL data, in other words the stored ontology can be accessed via SPARQL queries. There are also widely used software systems such as Jena and Sesame to deploy and manage the constructed ontologies (Estrin and Mani Srivastava, 2004).

XLink

The XML linking language, or XLink, is an XML markup language used for creating hyperlinks in XML documents. XLink is a W3C specification that outlines methods of describing links between resources in XML documents, whether internal or external to the original document. XLink defines a set of attributes that may be added to elements of other XML namespaces. XLink provides two kinds of hyper-linking for use in XML documents. Extended links are out of band hyper-links that, in a link base document, can link resources over which the link editor has no control. Simple links offer similar functionality to HTML links, which are in band links.

Related Work

Russomanno discuss a broad sensor ontology, which is called Onto-Sensor. Onto-Sensor primarily adapts parts of SensorML descriptions and uses extensions to the IEEE Suggested Upper Merged Ontology (SUMO) to describe sensor information and capabilities. The ontology is developed to support sensor information system applications in dynamic sensor selection, reasoning and querying various types of sensor (Lewis *et al.*, 2006). Onto-Sensor relies on deep knowledge models and provides extensive information about different aspects of the sensor nodes and devices. The ontology is represented in OWL format, the authors have discussed the advantages of the proposed approach compared to SensorML and XML based solutions. The main enhancement is providing self-descriptive meta-data for the transducer elements and embedded semantics in the descriptions, which could be utilized in various sensor discoveries, and reasoning applications (Adida, 2009). Although Onto-Sensor illustrates a semantic approach to sensor description and provides an extensive knowledge model, there is no distinctive data description model to facilitate interoperable data representation for sensors observation and measurement data.

A universal sensor observation and measurement data model in collaboration with a sensor specification model create semantic sensor network architecture. Semantic sensor network will utilize semantic web technologies and reasoning mechanisms to interpret sensor data from physical devices that perform observations and measurements. This will support building automated sensor information processing mechanisms to extract additional knowledge from real-time or archived sensor data (Antoniou and van Harmelen, 2004).

Ontology-based description of a service oriented sensor network is discussed in Barnaghi *et al*. (2009). The SWE and Geography Markup Language (GML) classes and properties in collaboration with SensorML, Suggested Upper Ontology (SUMO) and Onto-Sensor are used to develop ontology for sensor service description (Sheth *et al*. 2008). The ontology consists of three main components service property, location property and physical property.

Service-property explains what a service does and properties in the other two components describe the contextual and physical characteristics of the sensor nodes in wireless sensor network architecture. The ontology is represented in OWL form and some initial consistency checking and query results are provided to evaluate the validity of the proposed solution (Lewi, 2004). The system, however, does not specify how complex sensor data will be described and interpreted in a sensor network application (Heinzelman *et al*., 2000).

The proposed framework concentrates on building sensor description ontology for sensor discovery and description of sensor *meta* data in a heterogeneous environment (Lin *et al*., 2007). Although sensor device and service description will contribute to build, more autonomous sensor networks providing an interoperable data description model would be also an essential requirement in architecture for semantically enabled sensor networks.

Henson *et al*. (2008) describe a prototype application for the sensor web by using annotated video data. The dataset contains YouTube videos annotated with SensorML and XLINK models with reference to time ontology. The authors discuss how utilizing the semantic leads to retrieve videos by specifying temporal concepts such as “within”, “contains”, or “overlaps” during a time interval query submission. The proposed application demonstrates the main benefits of adding semantics to the sensor network and sensor data. The authors use keyword tagging and meta-data description to provide references to temporal concepts and domain ontologies. An extension to this idea could be seen as providing a universal *meta*-data structure with a broader scope to accommodate various sensor data types.

Semantic hierarchical sensor data storage (SEMHD)

In this section, we have introduced new sensor data storage.

At first, we arrange sensor nodes into some clusters. A sensor node in a cluster; plays role of a cluster head; collects sensor data from sensors that relies on relevant cluster, then aggregate data and send them to sink for future querying. Sensors send their data in XML form. In Fig.1, we see a snapshot of a sensor network that sensors are arranged in hierarchical mode.

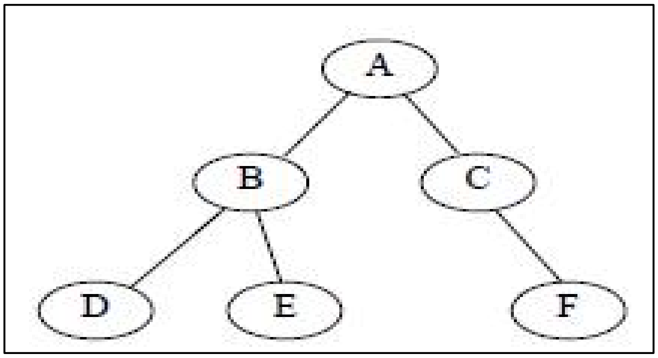

Fig.1. Network view

The network in Fig.1 is divided into two parts. Node B and C plays the role of cluster head in the network. Node B and node C aggregates received data. Then send them to sink node, which in this example is Node A. In other words, parent nodes should done data aggregation for sending to sink node. We have done our simulation using j-sim sensor networks simulator and protégé 2000 software. We also use LEACH algorithm that is a hierarchical protocol for clustering of sensors.

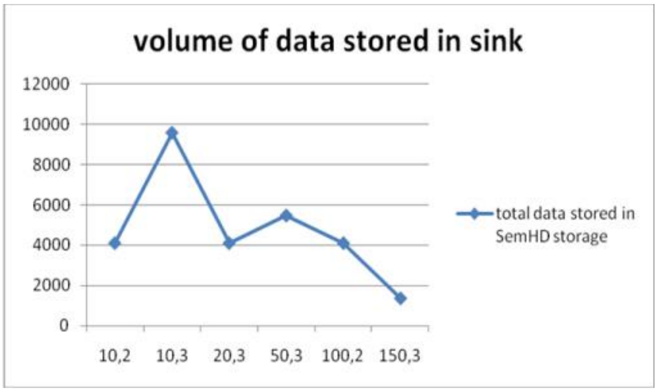

Fig.2. Amount of received data in sink node

In Fig.2, we can see amount of received data in different situation. Horizontal axis shows an ordered pair; (X, Y); X is the number of sensors and Y shows the number of clusters that the sensor network divides into. For example, in (10, 3) point we have 10-sensors that divides into 3-clusters. As we can see, usually increasing sensors in sensor networks causes fewer amount of data received in Sink node. One of possible reason is data aggregation overhead because fewer data are transmitted in the network. We should have a trade-off in total amount of received data and number of clusters.

Fig.3 shows the lifetime and remaining energy of a sensor network in variety of situations. As we can see, energy consumption of cluster head nodes is more than other nodes in the network but we are getting powerful features like more scalability, better management and less consume of bandwidth. We should have a decision in which approach is better for our purpose and applications between if arrange sensors or not.

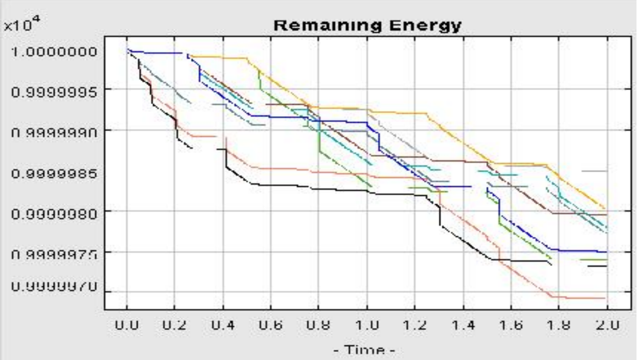

Fig.3. Energy consumption of the network

Conclusion and future work

In recent years, advances in energy efficient design and wireless technologies have enabled exciting new applications for wireless devices. These applications span a wide range including real time and streaming video and audio delivery and remote monitoring using networked micro sensors personal medical monitoring and home networking of everyday appliances. While these applications require high performance from the network they suffer from resource constraints that do not appear in more traditional wired computing environments. In particular, wireless spectrum is scarce often limiting the bandwidth available to applications and making the channel error prone and the nodes are battery operated often limiting available energy. If we can store sensors data more effectively, we have sensor networks more effective and lifetime. In this paper, we introduced and formalized a new Semantic Hierarchical Sensor Data Storage that divides sensors into some clusters. A node in a cluster that collects sensor data called cluster head. Sensors send their data in XML form; then cluster head aggregate received sensor data, then send them into sink; sink node collect data for further process like response more variety of queries, etc. For future work, we plan to explore a new mechanism to deal with link failures between sensors in the network. Sending data more semantically will be also another step. Another step is evaluating this method when sensor sends their data in stream.

References

1. Adida B (2009) RDFa in XHTML: Syntax and Processing-W3C Recommendation.
2. Akmal B Chaudhri, Awais Rashid and Roberto Zicari (2003) XML Data Management: Native XML and XML-enabled Database Systems. Addison Wesley.
3. Antoniou G and van Harmelen F (2004) A Semantic Web Primer (Cooperative Information Systems). The MIT Press.
4. Barry Burd (2002) Java & XML for Dummies. John Wiley & Sons.
5. Barnaghi P, Meissner S, Presser M and Moessner K (2009) Sense and Sens'ability: Semantic Data Modelling for Sensor Networks. Proceedings of ICT-Mobile Summit conference.
6. Estrin DCD and Mani Srivastava (2004) Overview of Sensor Networks. IEEE Computer Society.
7. Heinzelman WR, Chandrakasan AP and Balakrishnan H (2000) Energy-Efficient Communication Protocol for Wireless Microsensor Networks. Proc. of the 33rd IEEE International Conference on System Sciences, Honolulu, USA, 1–10.
8. Henson CA, Pschorr JK, Sheth AP and Krishnaprasad Thirunarayan (2008) SemSOS: Semantic Sensor Observation Service. IEEE computer society.
9. Lewi FL (2004) Wireless Sensor Networks. John Wiley; New York.
10. Lewis M, Cameron D, Xie S and Arpinar B (2006) Es3n: A Semantic Approach to Data Management in Sensor Networks. Semantic Sensor network workshop. The 5th International Semantic Web Conference ISWC, Athens, Georgia, USA.
11. Lin S, Arai B and Gunopulos D (2007) Reliable Hierarchical Data Storage in Sensor Networks. 19th International Conference on Scientific and Statistical Database Management IEEE, SSDBM, 26.
12. Sheth A, Henson C and Sahoo S (2008) Semantic sensor web, Internet Computing. *IEEE*, 12, 78–83.
13. Singh CP, Vyas OP and Tiwari KU (2008) A Survey of Simulation in Sensor Networks. In Proceeding of CIMCA.
14. Swartz A (2007) The Semantic Web in Breadth. Available at: http://logicerror.com /semanticWeb-long.