

Semantic Sensor Net: An Extensible Framework

Lionel M. Ni¹, Yanmin Zhu¹, Jian Ma¹, Minglu Li², Qiong Luo¹,
Yunhao Liu¹, S.C. Cheung¹, Qiang Yang¹

¹Department of Computer Science,
Hong Kong University of Science and Technology, Hong Kong

²Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai, P.R. China

¹{ni, zhuym, majian, luo, liu, scc, qyang}@cs.ust.hk,
²li-ml@cs.sjtu.edu.cn

Abstract. Existing approaches for sensor networks suffer from a number of critical drawbacks. First, homogeneous deployments have been commonly assumed, but in practice multiple deployments of sensor nets and heterogeneity of sensor networks are a serious problem. Second, existing approaches are very application-dependent and engineering-oriented. Third, there has been little standard available for WSNs. These drawbacks have significantly limited the further development of sensor networks. To overcome these critical drawbacks, we propose an extensive framework: Semantic Sensor Net (SSN). In brief, a semantic sensor net is a heterogeneous sensor network which enables dynamic tagging of semantic information to sensory data to allow more efficient and systematic monitoring and handling of the environmental dynamics to provide demanded services.

1 Introduction

Recent advances in wireless communications and microelectromechanical systems (MEMS) have led to the wide deployment of large-scale *wireless sensor networks* (WSN), which promises to revolutionize the way we monitor and control environments of interest [1, 2]. WSN was identified as one of the ten emerging technologies that will change the world in MIT Technology Review [3]. A wide variety of attractive applications [4] will come into reality, such as habitat monitoring [5], search and rescue, disaster relief, target tracking, precision agriculture and smart environments.

A sensor node is a low-cost and typically battery-powered device that integrates micro-sensing, onboard processing and wireless communication. A WSN is a self-organizing network composed of a large number of sensor nodes, tightly interacting with the physical world. Such a self-organizing network is able to not only disseminate sensory data across the network, but also provide in-networking real-time processing capability. WSN is a promising technology that effectively bridges the physical world

and the digital world, by which we can extract critical information from physical environments, and therefore better monitor and control dynamics of environments.

Distinct from traditional computer networks, each sensor node in a sensor network plays a minor role. In addition, we are more interested in sensory data. The data-centric nature of sensor networks provides an innovative approach to solve a greater class of applications. However, current work in wireless sensor networks suffers some major drawbacks.

- Most solutions are based on homogeneous sensor array.
- Each solution is usually for a specific application.
- The solution is usually an engineering approach without a common framework.
- There is no standard to allow communication among different sensors at different levels.

This paper proposes a new concept called “semantic sensor net” to alleviate the above drawbacks. A semantic sensor net (SSN) is a heterogeneous sensor network which enables dynamic tagging of semantic information to sensory data to allow more efficient and systematic monitoring and handling of the environmental dynamics to provide demanded services. SSN has the following advantages

- The tagging of semantic information to sensory data allows efficient handling of large-scale distributed heterogeneous sensory data.
- SSN can provide a sound theoretical foundation to research in wireless sensor networks at different levels.
- SSN can help develop a semantic-based framework to systematically solve various applications.

The rest of the paper is organized as follows. Section 2 discusses characteristics and challenges of WSNs. We review existing approaches and their limitations in Section 3. In Section 4, we give an overview of Semantic Sensor Net. Some of our preliminary results are discussed in Section 5. Finally, Section 6 concludes the paper and introduces the directions of future work.

2 Characteristics and Research Challenges

Compared with traditional computer networks and mobile ad-hoc networks (MANET) sharing more similarities, sensor networks present many unique characteristics. Every sensor node is highly resource-constrained, with limited computational capability and small storage. The wireless communication is unreliable. Each sensor node is typically powered by battery and usually not rechargeable. Sensor nodes may be deployed in unattended environments, exposed to unpredictable damages from environments, and hence any sensor node is prone to failure.

Unlike computers in traditional computer networks, most sensor nodes have no global ID due to low-cost mass production. In addition, global IDs introduce too much overhead which are not affordable for resource-constrained sensor nodes. Because of the lack of global IDs, traditional networking methodologies are not appropriate for WSNs. Since any sensor node may become unavailable at any time and wireless communications are unreliable, the network topology may change constantly over

time. Thus, the capability of self-organizing and self-configuring is fundamentally important. In many applications, sensor nodes are usually deployed without per-node placements (e.g., dropping sensor nodes from a flying airplane). Given the dynamic nature of sensor nodes, it is essential that these sensor nodes can cooperate with each other, form a network automatically, and work as a whole without human intervention.

With the rapid development of WSNs and its growing commercialization, it is probable that in an environment of interest different types of sensor nodes are deployed. Therefore, a sensor network could be very heterogeneous. Heterogeneity exists in both individual sensor nodes and sensor networks as a whole. So far there have been few standards available for WSNs. Different manufacturers produce distinct sensor network systems, adopting different hardware and software components. Even if we do not consider the heterogeneity caused by different manufacturers, sensor nodes from the same manufacturer can still be very heterogeneous in terms of function, capability, and so on. Sensors with different functions have been available, such as temperature, pressure, light, motion, and their combinations. Sensor networks as a whole can also be heterogeneous, since different sensor networks may employ different network-organizing strategies, routing algorithms, and aggregation methods. How to integrate heterogeneous sensor nodes to develop various, flexible and extensible applications has been a great challenge!

3 Existing Approaches and Their Limitations

In the past several years, sensor networks have received considerable research efforts, covering different aspects of design. In this section, we give an overview of existing approaches and study their limitations.

Because of resource constraint, the popular layered architectures used in traditional system design are not appropriate for sensor networks. Although a layered architecture can provide better organization and is more extensible, it introduces too much overhead since a packet may be added multiple headers of protocols, dominating the size of a packet. Existing research of sensor networks can still be roughly classified into several categories according to their functions.

- **Hardware and Wireless Communication** include sensor architecture, radio module design, MAC [6] and power control .
- **Infrastructure Establishment** includes deployment, localization [7], time synchronization [8], ID assignment and calibration, and middleware.
- **Network Organization** includes topology control [9], density control and cluster management.
- **Data Dissemination** includes routing [10], aggregation, compression, diffusion and query processing [11].
- **Applications** such as habitat monitoring, target tracking, battlefield surveillance, pollution monitoring [12], industry control, and so on.

Although extensive research has been conducted and some real applications have been in place, existing approaches suffer from several critical drawbacks which have

significantly restricted the wide deployment of sensor networks in practical applications.

- **Homogeneous sensors are commonly assumed.** In a small-scale sensor network, it may be reasonable to have homogeneous sensors which are usually identical or similar in terms of function, node architecture and software. With the homogeneous assumption, solutions can be greatly simplified. However, in practice sensor networks are inherently very heterogeneous. Existing solutions based on the homogeneous assumption can hardly work in such heterogeneous systems.
- **Existing solutions are very application-dependent.** Applications of sensor networks are very diverse and could have very different requirements and objectives. For example, the battle field surveillance application requires sensor nodes report to the gateway only when some events are detected. In contrast, for a building temperature monitoring application, sensor nodes report to the gateway regularly. Given such two distinct application scenarios, the respective designs have been very different. Such application-specific solutions are not extensible and cannot be reused. Most existing applications of sensor networks adopted tailor-made designs.
- **Existing solutions are engineering-oriented.** In contrast to traditional computer networks, so far there has been few standard, like TCP/IP, available for WSNs. Due to the lack of widely-accepted standards, we have to re-design most building blocks when developing a new sensor network, such as topology control, routing algorithm, and query processing. Such engineering-oriented approaches are particularly inflexible, and pose developers a big burden of re-engineering. In order to avoid unnecessary re-engineering, it is essentially important to develop core standards for sensor networks.

Besides the above, existing solutions are not extensible in the sense that once a sensor network for a specific application has been deployed, it is extremely hard to accommodate application dynamics and new application additions over the same sensor network. It results from two reasons. First, the current hardware limitation does not allow frequent updates of software burned in sensor nodes, as it leads to unreliability and high power overhead. Second, there is no effective mechanism to support such application dynamics and new application additions, which in practice are very desirable to make applications better meet real needs.

Having suffered much from these major drawbacks, we come to realize that a solid foundation and framework for WSNs is highly necessary, through which we can overcome these drawbacks, and hence promote the further development of sensor networks.

4 Semantic Sensor Net: An Overview

To alleviate the drawbacks experienced by existing approaches, we propose an extensible systematic framework: *Semantic Sensor Net* (SSN). In brief, a SSN is a heterogeneous sensor network which enables dynamic tagging of semantic information to sensory data to allow more efficient and systematic monitoring and handling of the environmental dynamics to provide demanded services. The important concept *semantics* is introduced to address various challenges, which exists in different levels of

designs of sensor networks, effectively enabling the integration, exchange, and reuse of sensory data across various applications and multiple sensor nets.

As mentioned in the previous section, any single sensor node is of negligible importance, and instead sensory data is what we are concerned the most. It suggests the data-centric principle, which is fundamentally different from the node-centric principle for traditional computer networks. The data-centric principle should be incorporated throughout designs of sensor networks. However, it is very challenging. Sensory data have unique characteristics and can be utilized in very different ways. High level applications usually require the integration of various sensory data. We believe the semantics-based framework can well address these challenges, and provide a solid foundation for WSN. Although sensory data can be very diverse, semantics inherently associated with sensory data can enable the integration and exchange of various sensory data, and accommodate different requirements of high-level applications.

The concept of semantics has been successfully introduced in semantic web, which is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation [13]. Traditionally, web pages are composed mainly for human's comprehension. Being intelligent, human beings have the ability to understand web information. However, it is difficult for digital computers to understand the meaning behind web pages. The semantic web brings structure to the meaningful content of web pages, so called semantics, enabling computers to carry out sophisticated tasks for users.

In SSN, semantics presents more flexible usage, which not only allows sensory data to be shared and integrated across various applications, but also provides a powerful framework for designs of sensor networks. Basically, *semantics refers to the critical meaning of sensory data, sensor nodes and application requirements, which we believe can help better decision making in various designs of sensor networks*. Within SSN, semantics can exist in different levels from bottom to top, as shown in Figure 1. Semantics in sensory data is the most basic, effectively supporting the realization of semantics in upper levels. Semantics in various applications is the most complex and can be factorized into much simpler forms. Semantics can be converted and form new semantics, and support efficient operations in different levels.

To demonstrate how semantics helps, let's take an example. Suppose in a building, a large number of heterogeneous sensor nodes are deployed for monitoring the environment inside the building. Example sensors include temperature, light and humidity. Now we may be concerned with whether there is a fire emergency inside the building. The 'fire emergency' certainly encompasses much semantics that can only be understood by human, and it resides on the service level. A fire can be roughly interpreted as a combination of strong light detections and high temperature detections in the same area. So the fire emergency is converted to a query with more specific semantics, "a strong light detection (≥ 10 candlepower) plus a high temperature detection (≥ 80 °C) in the same region (distance ≤ 1 m) within 10 seconds". The query is then sent across the sensor network. On receiving the query, a sensor node will be able to interpret and then set up new routing rules. The basic semantics of the new routing rule, which resides on the data dissemination level, may act like the following. This query is prioritized to be forwarded to temperature and light neighbors such that other sensors can avoid being involved in, which is a power-efficient design. And, if a

neighbor within ten meters reports a strong light detection and the node itself detects a high temperature event, it will form a fire alarm event and reports it to the gateway.

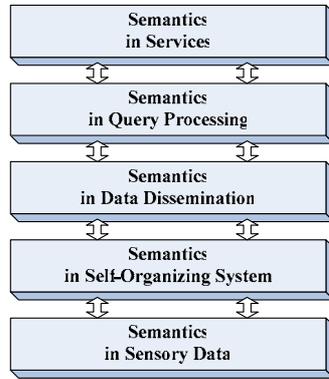


Figure 1: Semantics exists in different levels of sensor networks

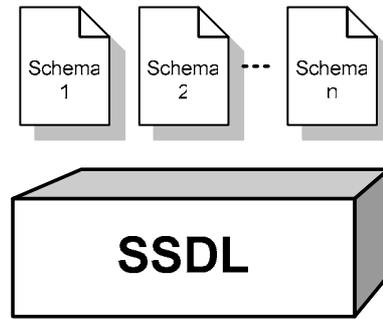


Figure 2: Sensory-data Semantic Modeling Framework (SSDL)

Besides semantics in the upper levels, semantics in the lower levels are also essential. A sensor node must maintain its own semantics, such as ID, location, sensing type, and sensing accuracy. And when a sensory data is available for transmission, its semantics should be enclosed, enabling other sensors to interpret it; otherwise, other sensors may have no idea what the received data is. The semantic is on the bottom sensory data level. Given the highly heterogeneous sensors in the environment, an integrated network is expected to set up, including all possible sensors. Since different sensors may adopt different wireless technologies, direct communication between some sensors may not be possible. In this case, some nodes should act as bridges for these sensors. Bridges are on the networking level, and may require new semantics to annotate for themselves.

The success of SSN requires the following be well addressed.

- **Semantic Sensory Data Modeling.** Sensory data can be very diverse. To enable the integration and exchange of various sensory data, an expressive data model is highly needed.
- **Semantic Sensor Network System Architecture.** To facilitate the development of sensor network systems, extensible system architecture plays a fundamental role.
- **Semantic-based Data Dissemination.** Data dissemination is at the core of sensor networks. Semantic-based data dissemination protocols are able to enable dissemination of various sensory data over heterogeneous sensor networks.
- **Semantic Query Processing.** With the unique abilities of built-in computation and data storage, sensor networks are very promising in data management. Queries are the effective way to acquire useful information from sensor networks. Due to sensory data distributed across the whole sensor network as well as the heterogeneity of sensory data, query processing is challenging. The semantics-based framework could help more efficient query processing.
- **Services.** We need abstract and form services. Based on which various applications can be built conveniently.

5 Some Preliminary Work of SSN

In this section, we present some of our preliminary work in SSN. More advanced and detail work will be carried out in our future work.

5.1 Semantic Sensory Data Modeling

Each sensor generates some kind of raw data. To make the raw data meaningful, we have to tag it, i.e., we should attach the semantics to it. The semantics of sensory data is the necessary description about data generation environment where the raw data was generated. The description should include:

- **Meta data.** These are the necessary description about the raw data itself. Take raw data produced by a temperature sensor for example. We have to make it clear that the raw data is a temperature measurement, how accuracy it is, and in what conditions it is valid. Meta data usually depends on the capability of sensing devices. Different sensing devices may have different kinds of meta data.
- **Context information.** These are about the context information in which the raw data was generated, which are usually related to the sensor node which the sensing device is attached to [14]. Take the above for example. In general, it should be made clear that where the temperature measurement was made (i.e., location of sensor node), which node took the measurement if applicable (i.e., ID of sensor node), and when it was captured (i.e., timestamp).

Without these necessary descriptions, raw data itself is meaningless. Therefore, raw data should be attached with respective semantics. We have the intuition that the sensory data produced by two identical sensors should have the same presentation pattern of semantics. We term presentation pattern as schema. The same type of sensory data follows the same schema and different types of sensory data follow different schemas. As long as one has the right schemas, he is able to dynamically interpret the sensory data for various usages, despite how heterogeneous these sensory data are. We propose an expressive Sensory-data Semantic Modeling Framework (SSMF), as depicted in Figure 2. Sensor-data Semantic Description Language (SSDL) is the language for defining schemas. Such a framework is very expressive and extensible. Users can conveniently define customized schemas using SSDL.

5.2 Semantic-based Data Dissemination

Data dissemination plays a major role in enabling commands to be propagated from gateway nodes to sensor nodes and collecting sensed data of interest from sensor nodes back to the gateway nodes for further study. It is very challenging, however, to design effective data dissemination protocols for sensor networks. It would become even more difficult if we consider the heterogeneity in both sensors and sensor networks as a whole. Heterogeneous sensors in a sensor network produce heterogeneous sensory data. Based on the sensory data, different actions may be required. Such semantic-based routing requirements make the design of data dissemination protocols

greatly complicated. In addition, heterogeneous sensor networks may deploy different network protocols and network organization. Some applications may require cooperation among these networks. Such inter-network heterogeneity makes the design of data dissemination protocols even more challenging.

We need to propose **semantics-aware routing**. It is not a specific routing algorithm. Instead, it is rather a framework, which aims to enable efficient data dissemination over large-scale heterogeneous sensor networks. Despite of the heterogeneity, it is certainly desirable that all of the sensor nodes work as a whole network, other than several separate networks working independently. Working as a whole can provide longer network lifetime, more efficient data dissemination, more complete sensory data and more flexible application integration.

Built over various existing routing algorithms, semantic-based routing should be extensible in the sense that it can address new emerging semantics possibly added to the sensor networks. It should take the advantage of available semantics in both sensory data and sensor nodes. Some possible scenarios are summarized as follows.

- **Semantics in sensor node.** After receiving a packet from its neighbors, the sensor node should take proper actions based on its capability (one kind of semantics). Suppose the destination of the packet is defined by a geographic region, and the node has no knowledge of this physical location. In this case, the sensor node should forward the packet to one of this neighbor which knows its physical location, or simply drop the packet.
- **Semantics in sensory data.** To save energy, data aggregation is a popular technique associated with routing in sensor networks. Rather than routing back every piece of sensory data, some sensory data are aggregated at some nodes, and the resulting data are then routed back to the gateway. Such in-network decisions should be based on semantics in the sensory data.
- **Semantics in query.** In general, queries are about some specific kind of sensory data. If a sensor node is able to interpret semantics in queries, irrelevant sensor nodes can avoid being involved so that more power can be saved. For example, if a query is issued to ask about temperature information, it is desirable that light sensors are not involved. After receiving a query, a sensor node can selectively forward the query to those neighbors which are also temperature sensors. By this means, light sensors will less be involved, and therefore such a query is more power-efficient.

5.3 Semantic Query Processing

The advent of wireless sensor networks provides a unique distributed platform to acquire and query streams of data. Traditional computational models transmit and process the data at each time instant and each individual record at a time. These models can no longer hold true for sensor networks. Furthermore, each sensor node only provides a small piece of the picture. To obtain an overall picture of the area, we must be prepared to answer queries about high-level patterns in place, instead of answering queries that concern low-level information that concern only limited space and time

[15]. In other words, we must be able to provide *semantic-level answers* to *pattern-related queries*.

Semantic queries distinguish themselves from the traditional queries with two major characteristics:

- Queries can involve aggregated environmental conditions, are location-context dependent and related to the tracking and monitoring of moving objects
- Queries are answered by taking into consideration of the device semantics on sensor-distribution topology, state and capabilities and answers can depend on power, sensors, and levels of confidence

Equipped with a variety of sensors, a wireless sensor network could make semantic inference about its environment, with queries on such conditions as the overall temperature, humidity, vibration, sound and lighting. All these queries can be answered dependent on location context [15]. A query planning system must be intelligent in order to provide timely and confident answers using a minimum amount of energy.

6 Conclusions and Future Work

For wireless sensor networks, any individual sensor node is by itself unimportant. Instead, sensory data collected from a group of sensors are what we are most concerned about. This suggests a data-centric principle in data processing, which is distinct from the node-centric principle in traditional computer networks. In response to the new challenges posed by the sensor networks, in this paper we have proposed an extensible framework known as the Semantic Sensor Networks. By explicitly exploiting the semantic information, which uncovers the machine-understandable meaning embedded in low-level sensory data, sensor nodes and application requirements, SSN enables the integration, reuse, and exchange of sensory data across various applications.

In this paper, we have just begun to touch the tip of the iceberg in SSN research. Our future work aims to make SSN practical for real developments of sensor networks, especially for large-scale heterogeneous sensor networks. To this end, a wide range of topics should be extensively studied, including an extensible architecture to address the highly heterogeneous nature of WSNs used in practice, the practical semantics-based data dissemination protocols, the semantics-based query processing methods and service methodologies for developing sophisticated applications.

Acknowledgement

We thank HKUST, Hong Kong Research Grant Council, and National Natural Science Foundation of China for providing support to this research. L. Ni, Q. Yang, Q. Luo and S.C. Cheung are supported in part by RGC grants HKUST 6264/04E, HKUST 6187/04E, HKUST 6263/04E, and 6167/04E, respectively. M. Li is sup-

ported by National Natural Science Foundation of China (No. 60473092) and Program for New Century Excellent Talents in University (No. NCET-04-0392).

References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, vol. 38, pp. 393--422, 2002.
- [2] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *ACM Communications*, vol. 43, pp. 51-58, 2000.
- [3] MIT Technology Review, Feb. 2003, <http://www.techreview.com>.
- [4] D. Estrin, G. P. L. Girod, and M. Srivastava, "Instrumenting the world with wireless sensor networks," presented at ICASSP, Salt lake City, UT, 2001.
- [5] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," presented at the 1st ACM international workshop on Wireless sensor networks and applications, 2002.
- [6] E. Shih, S. H. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan, "Physical Layer Driven Protocol and Algorithm Design for Energy-Efficient Wireless Sensor Networks," presented at 7th Annual International Conference on Mobile Computing and Networking (MOBICOM), 2001.
- [7] L. Doherty, K. S. J. Pister, and L. E. Ghaoui, "Convex Position Estimation in Wireless Sensor Networks," presented at INFOCOM'01, Anchorage, AK, 2001.
- [8] J. Elson and D. Estrin, "Time Synchronization for Wireless Sensor Networks," presented at the 15th International Parallel and Distributed Processing Symposium, 2001.
- [9] A. Cerpa and D. Estrin, "ASCENT: Adaptive Self-Configuring sEnsenor Networks Topologies," presented at 21 Annual Joint Conference of the IEEE Computer and Communications Societies, 2002.
- [10] A. Salhieh, J. Weinmann, a. M. Kochhal, and L. Schwiebert., "Power Efficient Topologies for Wireless Sensor Networks," presented at International Conference on Parallel Processing (ICPP). Valencia, Spain, 2001.
- [11] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "The Design of an Acquisitional Query Processor for Sensor Networks," presented at SIGMOD'03, San Diego, CA, 2003.
- [12] H. Ngan, Y. Zhu, and L. M. Ni, "SAS: Stimulus-based Adaptive Sleeping for Wireless Sensor Networks," presented at the 34th International Conference on Parallel Processing (ICPP'05), Norway, 2005.
- [13] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, 2001.
- [14] C. Xu, S. C. Cheung, C. Lo, K. C. Leung, and J. Wei, "Cabot: On the Ontology for the Middleware Support of Context-Aware Pervasive Applications," presented at Building Intelligent Sensor Networks (BISON'04) in conjunction with IFIP International Conference on Network and Parallel Computing, Wuhan, China, 2004.
- [15] A. Deshpande, C. Guestrin, S. Madden, J. M. Hellerstein, and W. Hong, "Model-Driven Data Acquisition in Sensor Networks," presented at VLDB'04, Toronto, Canada, 2004.