2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing

SECURE SEMANTIC SENSOR WEB AND PERVASIVE COMPUTING

Bhavani Thuraisingham and Kevin W. Hamlen
Department of Computer Science
The University of Texas at Dallas
bhavani.thuraisingham, hamlen@utdallas.edu

***Abstract*—In this paper we discuss issues on developing a secure semantic sensor web. SensorML is the starting point for this work. We explore the layers for a semantic sensor web and discuss security issues. We also discuss secure peer-to-peer computing as it relates to sensor web.**

I. INTRODUCTION

Sensors are everywhere and they have become part of our daily lives. These sensors are continually monitoring the environment, events, activities, people, vehicles, and many other objects, gathering data from these objects, aggregating the data and then making sense out of the data and finally taking actions based on the analysis of the data. For example, we have sensors to monitor the temperature of a manufacturing plant periodically and if the temperature exceeds a certain value, then they should raise an alarm. At the other extreme, we now have sensors monitoring the activities of people and if these activities are considered to be suspicious, the law enforcement officials are notified. We need sensors to be intelligent. That is, sensors must not only be able to monitor the situation, they must also be able to rapidly analyze the data gathered and make decisions. This is because there is a lot of data that is out there and the sensors must process certain critical data and discard certain other data and perhaps store a third set of data for future analysis. Furthermore, the sensors are limited in their storage capabilities. Therefore, the challenge is to develop algorithms for sensors to manage the data and information under massive resource and timing constraints. Essentially, we need semantic web technologies to develop what has come to be known as a web of sensors or a sensor web.

Recently data management system researchers have focused on developing data management techniques for managing sensor databases. While much progress has been made on sensor database management, there is still a lot to be done. For example, very little consideration has been given to security and privacy for sensor databases. Sensors are often operating in an open environment and are vulnerable. The data managed by the sensors may be compromised. We need security techniques to ensure that the sensor data is protected. We must also ensure that the data is not maliciously corrupted. Furthermore, privacy is an added consideration. The people may want the information collected about them to be private. This clearly illustrates the need for applying trustworthy semantic web technologies to develop a secure semantic sensor web.

Closely related to sensor information management is wireless information management pervasive computing. The wireless and mobile devices have embedded sensors and processors. These processors have to communicate with each other to solve problems and carry out activities. That is, we need to apply semantic web technologies for wireless and pervasive computing environments. This paper provides a discussion of the issues involved in developing a secure semantic sensor web. In particular, we discuss various types of security mechanisms, policies and markup languages for sensor data management.

The organization of this paper is as follows. In Section 2 we discuss security for sensor data management. A note on wireless information management and security is discussed in Section 3. Security for moving data systems is given in Section 4. Secure sensor semantic web is discussed in Section 5. Pervasive computing and semantic web are discussed in Section 6. Security for peer-to- peer computing issues as related to sensor networks is discussed in Section 7. The paper is summarized in Section 8. In this network, sensor data managers are connected through some communication subsystem. Background information in data management, sensor networks and sensor data managers are given in [1]. For more details on building trustworthy semantic web we refer to [2].

II. SECURITY FOR SENSOR DATA SYSTEMS

A security policy essentially specifies the application-specific security rules as well as application-independent security rules. Application-independent security rules would be rules such as

* The combination of data from two sensors is always sensitive.
* Sensor operating at level $L1$ cannot read data from sensor operating at level $L2$ if $L2$ is a more sensitive level than $L1$.

The second rule above is usually enforced for multilevel security. Now the main question is how does the policy for secure data management apply for sensor data? We could have sensors operating at different levels. Sensors in the Middle East may be highly classified while sensors in Europe may be less classified. Classified sensors will

978-0-7695-4049-8/10 $26.00 © 2010 IEEE
DOI 10.1109/SUTC.2010.75

gather classified data while unclassified sensors will gather unclassified data. Furthermore, sensor data may be in the form of streams. Therefore, we need access control policies for data streams. Within each level, one could enforce application specific rules. Application specific security rules include the following:

* Only law enforcement officials have authorization to examine data emanating from sensor A.
* Data from sensors A and B taken together are sensitive.
* All the data emanating from sensors in Washington DC federal buildings are sensitive while the data emanating from sensors in North Dakota federal buildings are not sensitive.

Essentially application-specific rules are specified using security constraints. We discuss security constraint processing in a later section.

Security policy integration is a major challenge. That is, each sensor may enforce its own security policy and have its own constraints. The challenge is to integrate the different policies especially in distributed and cluster environments. For example, in the case of a cluster, each cluster of sensors may have its own policy, which is derived from the security policies of the individual sensors. The policies of the clusters will have to be combined to get an integrated policy.

Security has an impact on all of the functions of a sensor data manager. Consider the query operation. The query processor has to examine the access control rules and security constraints and modify the query accordingly. For example, if the fact that the existence of Operation X is classified, then this query cannot be sent to an unclassified sensor node to monitor the situation. Similarly the update process also examines the access control rules to see if the data coming from a particular sensor can be inserted into the sensor database. That is, say data coming from a sensor that manages an operation in the Middle East may not be entered into a sensor data manager in South East Asia. Secure sensor transaction processing is another issue. First, what does transaction processing mean? One could imagine a sensor at site A and a sensor at site B monitoring the environments and making simultaneous updates to the sensor database. Both updates have to be carried out as a transaction. This is conceivable if both, say, temperature values depend on the final computation of some parameter. So assuming that the notion of a transaction is valid, what does it mean to process transactions securely? There has been a lot of work on secure transaction processing both for single level and multilevel transactions. In the case of a single level transaction, it is assumed that the transaction is processed at a single security level. In the case of multilevel transactions, the transaction may operate at multiple security levels. The main challenge is to ensure that information does not flow covertly from a higher level to a lower level. Sensor transaction processing will also have similar challenges. We need to examine the techniques from secure transaction processing and real-time transaction processing to see if we can develop techniques specific for dependable sensor transaction processing (see also [3] and [4]).

Next consider the storage manager function. The storage manager has to ensure that access is controlled to the sensor database. Storage manager may also be responsible for partitioning the data according to the security levels. The security impact of access methods and indexing strategy for sensor data are yet to be determined. Metadata management is also another issue. For example, we need to first determine the types of metadata for sensor data. Metadata may include descriptions about the data, the source of the data as well as the quality of the data. Metadata may also be classified. In some cases, the metadata may be classified at a higher level than the data itself. For example, the location of the data may be highly sensitive while the data could be unclassified. We should, also ensure that one cannot obtain unauthorized information from the metadata.

In a shared sensor data processing environment, because a lot of data has to be aggregated and fused, there could be a potential for inference problems. That is, the aggregated data from sensor nodes A, B and C could be highly sensitive. For example, one sensor could monitor the situation is the Middle East and another sensor could monitor the situation is Asia and the combined sensed information could be highly sensitive. The inference controller has to examine the constraints and prevent such sensitive information from being released to individuals who are not authorized to acquire this information.

Sensors are especially vulnerable to attacks as they function in an open environment. That is, sensors could be anywhere in the world and the sensor data managers could be compromised and the results aggregated. As a result the most private information could be divulged. The sensors could monitor, say, people in a shopping mall and the activities of the people may be monitored by law enforcement agencies as well as hackers who have hacked into the sensor system. Furthermore, based on the events monitored, the law enforcement agencies could arrest innocent individuals especially if the analysis tools do not give accurate information.

There has been some work recently on privacy-preserving sensor surveillance by Mehrotra et al [5]. The idea here is for users to wear RFID tags (Radio Frequency IDs) and these tags are detected by sensors. If it is a tag that the sensor can recognize then the persons identity is hidden. If the sensor cannot recognize the tag, then the person is displayed and privacy is not maintained about him or her. Note that as stated in a recent special issue of IEEE Spectrum Magazine (see [6]), RFID technology raises many interesting security and privacy questions. Furthermore, it is stated that sensor and wireless technology will revolutionize Information Technology. Some research on security and

surveillance is also reported in [7] and [8].

III. Secure Wireless Data Management

Wireless devices include telephones, PDAs (personal digital assistants) and more recently laptop computers. These devices have sensors embedded in them. Therefore, all of the security issues discussed for sensor data and information management apply for wireless information management. There are also additional considerations for wireless information management.

An excellent introduction to security in wireless networks was given recently in [9]. The authors state that these networks are susceptible to various attacks including denial of service, node capture and physical tampering. These networks are used for various applications including monitoring building safety and earthquakes as well as for military applications. Various security solutions including encryption, authentication and privacy are discussed in [9]. In addition, challenges for secure group communication as well as intrusion detection are discussed.

In the case of secure data management for wireless networks, the challenges include secure query processing, transaction management, storage management and metadata management. For example, how can queries be optimized efficiently? The users of the mobile devices may not stay in one location and therefore the query response has to be routed to the users location. Is the user authorized to see the response? How can access control be enforced? How can the user be authenticated? How can the aggregation problem be handled? As stated in [9], the sensor data has to be aggregated before sent to the base station so that the base station is not flooded with data. Does the user have access to the aggregated data?

In the case of transaction management, we need to first define the notion of a transaction. Multiple users may be updating the data managers both attached to the wireless nodes as well as to the base stations. Do the transactions have to be serializable? Can we live with weak serializability? Now can access control be enforced during transaction processing?

Storage management issues include managing the storage attached to the wireless nodes as well as the base stations. The wireless nodes may have limited storage capability. In addition, they also have limited power capability. Therefore, the challenge is what data should be maintained by the wireless node and what data should be stored at the base station? How can replicated copies be kept consistent? How can the integrity and security constraints be enforced? What is the security impact on access methods and index strategies?

Finally, in the case of metadata management, we need to first define what metadata is for wireless networks. Typically, metadata for wireless networks would include information about the nodes, their capacity, power consumption-related information as well as security policies enforced. Storing and managing the metadata is a challenge. If the data storage is limited at the wireless nodes, then where is the metadata to be stored? Is it feasible to store the metadata at the base stations and retrieve it each time it is needed? Can coaching strategies used for data be used for metadata?

Other attacks include intrusions, denial of service, eavesdropping, and spoofing. While we need security solutions for wireless networks, are there additional challenges for data management? Many organizations including university campuses have gone wireless. Therefore, building security solutions both for wireless networks and data managers as well as other information management technologies is critical. An overview of security for wireless sensor networks we refer to [9].

IV. Secure Mobile and RFID Data Management

Recently there have been many efforts on developing moving data management. For example, RFID tags on people and objects may be moving from place to place. Any data management system that manages such tags should have the capability to manage moving data. First of all, we need to develop secure query processing strategies. In mobile databases, users are moving continuously and data may also be migrating. Therefore, we need dynamic query processing strategies as the size of the databases as well as communication distances may vary from time to time. We also need special techniques to handle the movement of data and users. Special transaction processing techniques are also needed. We may have to sacrifice strict serializability of transactions, as the data is dynamic in nature. Furthermore, we need to examine security for transaction management in a mobile environment.

Much of the algorithms will depend on the security policies enforced. For example, what are the application specific security constraints? How can we handle missing information? How can we enforce flexible security policies? How can we securely route the information to the mobile users? How do we handle data that migrates from place to place? Some directions are provided in [10].

V. Secure Semantic Sensor Web

The sensor nodes in a network have to interoperate with each other, enforce various security policies and carry out activities in real-time. Therefore, there is now much interest to integrate sensor networks with semantic web technologies. In this section, we will discuss some of the directions and in the next section we will go beyond sensor webs and discuss how semantic web is integrated with pervasive computing infrastructures.

In the recent workshop on Semantic Sensor Networks in Athens, Georgia [11], there were discussions on the

use of semantic webs to address sensor networking applications using RFID technologies as well as complex, cross-jurisdictional, heterogeneous, dynamic information systems. As stated in the worship objective, the goal was to "to develop an understanding of the ways semantic web technologies, including ontologies, agent architectures and semantic web services can contribute to the growth, application and deployment of large-scale sensor networks."

Hendler and his team at the University of Maryland have developed techniques for the interactive composition of web services that can be used in a sensor network environment. They state that "as web services become more prevalent, tools will be needed to help users find, filter and integrate these services." Therefore, they have developed ways to compose existing web services that include "presenting matching services to the user at each step of composition, filtering the possibilities by using semantic descriptions and directly executing the services through WSDL".

A significant development connecting semantic web with senior technologies is OGCs SensorML standard. As stated in [12], "SensorML is an Open Geospatial Consortium standard markup language (using XML schema) for providing descriptions of sensor systems. By design, it supports a wide range of sensors, including both dynamic and stationary platforms and both in-situ and remote sensors." It supports many features including sensor discovery, sensor geolocation, processing of sensor observations, a sensor programming mechanism and subscription to sensor alerts.

While there is much progress, we need to develop ontologies and web services, as well as sensor RDF so that we have the capability for automated machine understandable sensor data and web pages. Furthermore, very little research on incorporating security into sensor webs has been reported. The policies that we have discussed in earlier sections can be expressed in languages like SensorML. Furthermore, we need to build reasoning systems to reason about the policies as well as make deductions and learning. The University of Maryland Baltimore County is leading the way for such research. We discuss some of their work in connecting semantic webs with pervasive computing in the next section.

VI. Pervasive Computing and the Semantic Web

Pervasive computing is essentially about a collection of sensors, wireless devices, embedded processes all interacting with each other to carry out various activities such as video surveillance, detecting supplies in the registers, automatic washing machine operation, and monitoring the heart beats of patients. Semantic web technologies are being integrated with pervasive computing to understand the data and carry out various pervasive computing operations.

Lalana Kagal and others at the University of Maryland Baltimore County have examined semantic framework for senior web [13]. Their solution is based on webs services and distributed trust management. Key to their architecture is the "The Service Manager" that acts as a mediator between the services and the users. All clients of the system, whether they are services or users, have to register with a Service Manager in what they call the "Smart Space", The Service Manager (SM) is responsible for processing Client Registration/De-Registration requests, responding to registered Client requests for a listing of available services, for brokering Subscribe/Un-Subscribe and Command requests from users to services, and for sending service updates to all subscribed users whenever the state of a particular service is modified. They state that the service manager is arranged in a tree-like hierarchy and messages are routed through to other service managers through this tree. The essential points of their trust management approach are as follows: Each client establishes trust with its service manager, and SMs across the hierarchy establish trust among them, hence trust now is a concept that is transparent between all clients in the system. They have also defined a security agent that carries out the security activities. Semantic web technologies are utilized for policy specification as well as reasoning by the security agent.

Some of the early work on integrating pervasive computing with semantic web was carried out at Nokia research labs by Lassila et al [14]. In his presentation on "pervasive computing meets the semantic web", Labial introduced the notion of "semantic gadgets" which combine semantic web with ubiquitous computing. He states that: 1) device capabilities and service functionality are explicitly represented, 2) everything is addressable (using URIs), and 3) semantic web is the basis for "semantic interoperability". In his approach, agents will discover services, carry out reasoning, as well as learn and plan. Lassila also introduced the notion of device coalitions where all devices advertise their services and a device extends its functionality by discovering missing functionality offered by another device, and contracting the use of the service. He further states that everything can be discovered including reasoning services and planning services. One needs to integrate the work, say, by Lassila with the work by Kigali to integrate trustworthy semantic web and ubiquitous computing.

In some other work at Fujitsus Labs, the researchers introduce the notion of task computing as the technology to integrate semantic web with pervasive computing. They state that task computing which shifts the focus to what users want to do as opposed to the specific means of users doing the task. They then state that task computing offers device manufacturers an incentive to incorporate semantic web technologies into their devices. They have implemented a task computing environment using the semantic web technologies such as RDF and OWL [15].

There is also now research on integrating sensor webs with RFID technologies and semantic webs. For example, the research of Nabil Adam and his team at Rutgers has developed techniques for the composition of web services

for RFID data management and interoperability for border patrol [16]. This is an area that is critical for homeland security. In summary, while there is now research on integrating semantic web with pervasive computing and RFID data management, we need to also focus on security issues. As we have stated, policies could be expressed in semantic web languages. Furthermore, we need to develop reasoning engines that can reason about the policies and handle problems like the inference and privacy problems.

VII. Secure Peer-To-Peer Computing

Deployed sensor networks typically consist of a large number of low-power, lightweight devices with limited computing and communication capabilities. The sensor nodes therefore typically lack the computing power and physical security necessary to defend against a capable and determined adversary. They must collaborate to effectively aggregate data, make decisions, and communicate relevant information to trusted data collection points.

Fully decentralized peer-to-peer (p2p) networking paradigms are an obvious choice to facilitate these objectives. While peer-to-peer networking is perhaps most popularly known for internet file-sharing, an explosion of peer-to-peer research over the past decade (c.f., [17]) has established the technology as highly applicable to a broad range of networking applications and environments; and its low computing costs and adaptability to network churn makes it well-suited to sensor networking applications.

At a high level, the appeal of p2p stems from its lack of any central authority, and therefore any central point of possible failure. Every node acts as both server and client, using highly fault-tolerant, lightweight communication protocols to cooperatively route and store information across the network. This extreme decentralization tends to result in high resilience against data integrity violations (since data tends to be massively replicated) and standard denial of service attacks (since the network can only be disconnected by compromising a large percentage of the network within a short span of time).

However, high decentralization leaves current-day p2p networks vulnerable to new classes of attacks that are inapplicable to traditional centralized networks, and therefore less well studied. For example, Sybil attacks [18] involve attackers who contrive to join victim p2p networks under many false identities rather than attacking the network directly. Over time and with judicious use of computing resources, they can manage to occupy a significant portion of the network topology and thereby intercept a disproportionate share of network traffic. This can lead to successful confidentiality and denial of service attacks even in these highly decentralized settings.

The key to protecting against such attacks seems to be to harness the considerable distributed computing power of p2p networks to cooperatively learn and enforce strong security properties. For example, our past work [19], [20] has led to p2p variants that use distributed hash tables and economic incentive schemes to discover and avoid malicious nodes who have managed to infiltrate the network. Future research should apply similar techniques to enforce strong, end-to-end data confidentiality properties, which continue to be a prominent open problem in the field. Such work is essential for the development of secure, fault-tolerant, distributed networking paradigms that are resilient against data integrity, data confidentiality, and data availability attacks in wireless sensor networks.

VIII. Summary and Directions

We have discussed a number of security issues for sensor data management and also gave privacy some consideration. We also discussed secure wireless information management. Much of our focus has been on access control. We then discussed the integration of semantic web technologies with sensor and pervasive computing technologies. We also briefly discussed secure peer-to-peer computing with respect to sensor networks.

There are several areas for future research. First, we need to develop appropriate policies for sensor data and represent these policies in a suitable language. SensorML is a start. We need to develop languages comparable to RDF such as Sensor. We also need to build reasoning engines to reason with sensor data and the policies. Finally, we need to integrate pervasive computing with semantic web, sensor web and security.

References

[1] Carney, D., et al, "Operator Scheduling in a Data Stream Manager", *Proceedings of the 29th International Conference on Very Large Data Bases*, Berlin, Germany, 2003.

[2] Thuraisingham, B. *Building Trustworthy Semantic Web*. CRC Press/Taylor and Francis, 2007.

[3] Thuraisingham, B., "Secure Sensor Information Management", *IEEE Signal Processing*, May 2004.

[4] Thuraisingham, B., "Security and Privacy for Sensor Databases", *Sensor Letters, Inaugural Issue (American Scientific)*, March 2004.

[5] Mehrotra, S., et al, "Privacy Preserving Surveillance", Demonstration, University of California, Irvine, 2004.

[6] *IEEE Spectrum*, July 2004.

[7] Thuraisingham, B. et al, "Access Control for Video Surveillance", *Proceedings ACM SACMAT*, 2006.

[8] S. Chaitanya and B. Thuraisingham, "Automatic Face Detection for Privacy Preserving Surveillance", *Technical Report, The University of Texas at Dallas*, 2006.

[9] Perrig, A. et al, "Security in wireless sensor networks". *Communications of the ACM*, 2004.

[10] Lubinski, A., "Security Issues in Mobile Database Access", *Proceedings of the IFIP Database Security Conference*, Chalkidiki, Greece, July 1998 (formal proceedings published by Kluwer 1999).

[11] *Semantic Sensor Networks Workshop*, November 5-9, Athens, Georgia, USA 2006.

[12] SensorML http://en.wikipedia.org/wiki/SensorML

[13] L. Kagal et al, "A Security Architecture Based on Trust Management for Pervasive Computing Systems", http://ebiquity.umbc.edu/_file_directory_/papers/15.pdf

[14] O. Lassila, "Semantic Gadgets: Pervasive Computing Meets the Semantic Web", http://www.lassila.org/publications/2002/lassila-nist-pervasive-2002.pdf

[15] Ryusuke Masuoka, Bijan Parsi and Yannis Labrou, "Task Computing- the Semantic Web meets Pervasive Computing" - http://www.flacp.fujitsulabs.com/~rmasuoka/papers/Task-Computing-ISWC2003-202-color-final.pdf

[16] Aabhas V Paliwal, Nabil Adam, Christof Bornhövd, Joachim Schaper, "Semantic Discovery and Composition of Web Services for RFID Applications in Border Control", http://www.dvs1.informatik.tu-darmstadt.de/staff/bornhoevd/ISWC'04.pdf

[17] J. Risson and T. Moors. "Survey of research towards robust peer-to-peer networks: Search methods". *Computer Networks*, vol. 50, pp. 3485-3521, 2006.

[18] J. R. Douceur. "The Sybil attack". *In Proc. 1st International Workshop on Peer-to-peer Systems*, March 2002, pp. 251-260.

[19] N. Tsybulnik, K. W. Hamlen, and B. Thuraisingham. "Centralized security labels in decentralized p2p networks". *In Proc. Annual Computer Security Applications Conference*, December 2007, pp. 315-324.

[20] K. W. Hamlen and B. Thuraisingham. "Secure peer-to-peer networks for trusted collaboration, invited paper." *In Proc. 2nd IEEE International Workshop on Trusted Collaboration*, November 2007.