



Secure Sensor Information Management and Mining

An overview of infrastructures and data managers for dependable sensor networks

This article describes issues and challenges for secure sensor information management. In particular, we will discuss data management for sensor information systems including stream data management, distributed data management for sensor data, sensor information management including mining sensor data, security for sensor databases, and dependable sensor information management such as tradeoffs between security, real-time processing, and fault tolerance. Finally we will discuss object-based infrastructures for sensor systems as well as directions for sensor information management.

Introduction

Sensor networks and sensor information management are critical technologies for many applications, including process control and manufacturing, and more recently for detecting and preventing terrorism. Sensors are embedded in a myriad of devices and locations including public buildings such as shops, theaters, airports, railway stations, and hospitals. Data in the form of streams emanate from sensors, and these streams have to be aggregated, fused, stored, managed, and analyzed for various applications. For example, streams that emanate from the sensors will have to be mined to extract useful information. Sensor data could be information about tracking various individuals or information from sensors tracking the temperature in a manufacturing plant. We need to mine the data to determine potential problems. If we see that an individual is making frequent trips to a shop that sells chemicals and then also makes visits to a shop that sells firearms, then we may want to place that

individual on a suspicious individuals list. That is, with appropriate mining of the data emanating from the sensors placed in the various shops we could make connections, links, and associations. Another example is mining the information emanating from video cameras installed in various shopping malls and airports and other public places. This video data may also be in the form of streams and will have to be mined to detect suspicious behavior [1].

One of the emerging technologies is biological sensors. The idea is to develop sensors that could detect terrorists carrying biological agents and chemical agents. The spread of biological agents will also have to be detected. The information gathered has to be mined possibly in real time to determine if similar attacks have occurred and what emergency response measures should be taken.

Bhavani Thuraisingham

To analyze and mine sensor data effectively, we need to be able to build supporting infrastructures and data/information managers for sensor networks. Sensor networks are essentially a collection of networked sensors; each sensor may carry out some local processing, and the sensors have to work together to solve a particular problem. There are many special data management and information management considerations for sensor data management and mining. These include special data modeling techniques, query processing, and index strategies. We also need to build infrastructures including special purpose operating systems and middleware for sensor networks. Essentially we need to build a dependable environment to host sensor applications to carry out tasks like multisensor data fusion and sensor data mining.

This article discusses infrastructures and information/data managers for dependable sensor networks. By dependable we mean sensor networks that are secure, survivable, and fault tolerant and can process data in real time. There are quality of service tradeoffs that one needs to make to build dependable sensor networks. These tradeoffs will also be discussed. Our previous experience in building next-generation command and control systems using real-time objects will form the basis for the discussion on sensor networks. Our infrastructure for the sensor network will consist of sensor objects that have to perform many functions such as interprocess communication as well as memory management. The sensor objects have to incorporate security and fault tolerant processing capability. They also need to have the capability to manage data in real time.

The sensor data/information manager will manage sensor/stream data. We will examine various issues like data models and query strategies for sensor data processing. The sensor data/information manager will be hosted as an application on an object-based infrastructure. The processing may be distributed among multiple sensor nodes, and the activities of these nodes have to be coordinated. Essentially we need a distributed sensor information management system. One can also think of such a system as a peer-to-peer sensor system where the nodes are peers that have to work together to solve a problem. Our goal is to use existing technologies and develop new ones as needed to build an environment for sensor information management. Sensor information management and sensor data mining are becoming critical for many applications including detecting and preventing terrorism. Therefore, it is important to build dependable and survivable infrastructures and data managers for sensor networks.

Terminology

There is some confusion in the literature among the terms sensor networks, sensor information management, sensor data management, and sensor information systems. We will discuss our assumptions in this section.

Essentially sensor networks encompass everything. A sensor network may be considered to be a collection of networked sensors. These sensors collaboratively work together and solve problems such as collaborative situation monitoring. Integrated systems are needed to manage the sensor networks. These integrated systems consist of various components including sensor information management systems, sensor data management systems, communication and networking subsystems, sensor operating systems, and sensor middleware. Sensor networks are illustrated in Figure 1.

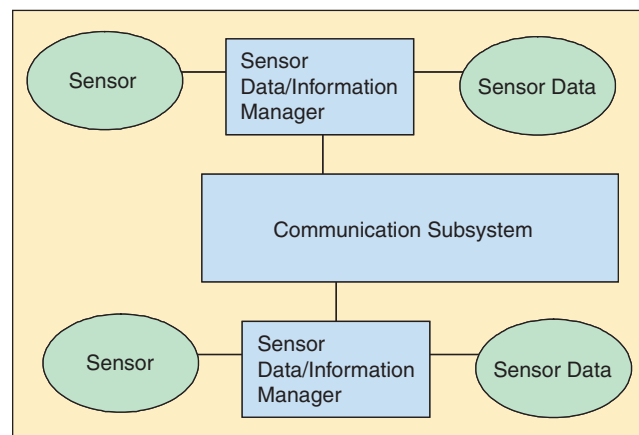
A sensor information management system is a component of a sensor network; it is the information management component. Sensor information management systems essentially extract information out of the sensor data and manage the information. Sensor data management systems manage the sensor databases. These data-

bases may not be full-fledged secondary storage databases as in the case of relational databases being used, for example, for banking and airline reservation applications. These databases could be main memory databases. Essentially sensor information systems encompass both sensor information management systems and sensor data management systems.

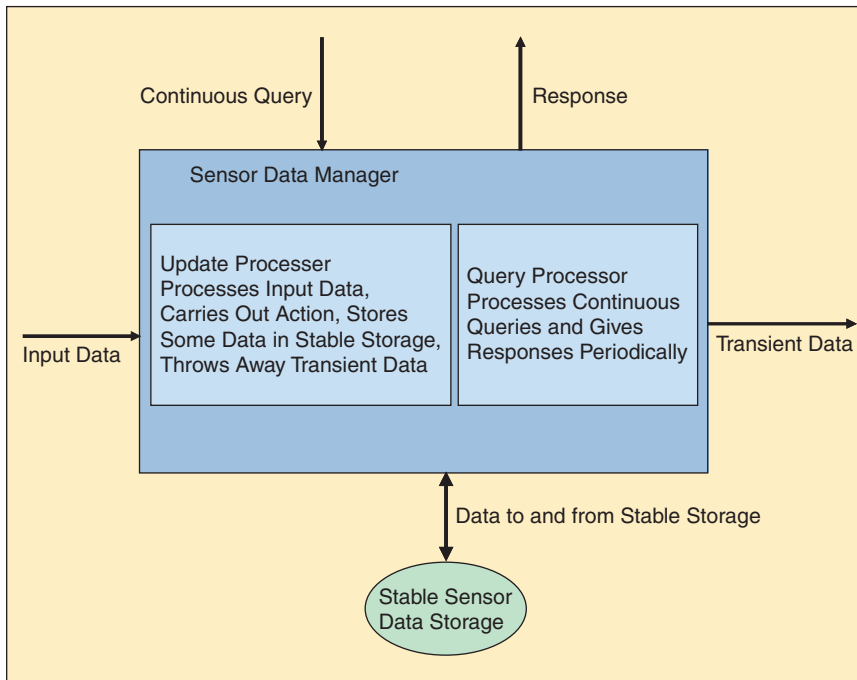
Sensor Data Management

Various research efforts are under way to develop sensor data management systems. These include the efforts described in [2]–[4] (see, for example, the work at Stanford University, Cornell University, MIT, University of California Berkeley, Brown University, and Brandeis University). Sensor data may be in the form of streams. Special data management systems are needed to process stream data. For example, much of the data may be transient. Therefore, the system has to rapidly analyze the data, discard data that is not needed, and store the necessary data. Special query processing strategies including query optimization techniques are needed for stream data management. Many of the queries on stream data are continuous, and we need special query strategies for processing them. Researchers are also examining special data models as well as extensions to the relational data model for stream data management. Query languages such as SQL are being extended with constructs for querying stream data. Research efforts are also under way for extending XML with constructs for sensor data management. We also need to determine the types of metadata to collect as well as develop techniques to store and manage the metadata. Figure 2 illustrates sensor data management.

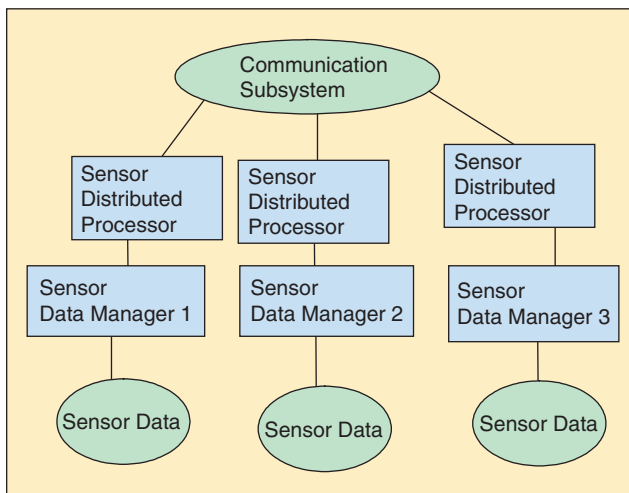
While many of the efforts are extending or enhancing current data management systems to process sensor data, the main question is: do we need a radically different kind of data model and data management system? Research is also needed on developing access methods and index strategies for stream/sensor data management systems. One also needs to examine the notion of a transaction and determine the type of transaction model suitable for stream databases. Finally we need to examine



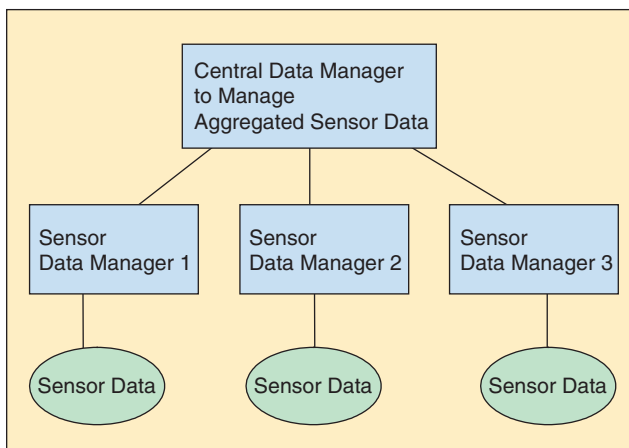
▲ 1. Sensor networks.



▲ 2. Sensor data management.



▲ 3. Distributed sensor data management.



▲ 4. Distributed data and central control.

techniques for managing main memory databases and real-time databases and determine if they are applicable to stream data management. While there is much progress during the last few years, much research still remains to be done.

Data Distribution

Sensors are often distributed and in many cases embedded in several devices. We need distributed data processing capabilities for managing the distributed sensors. Data possibly in the form of streams may be emanating from multiple sensors. Each sensor may have its own data management system, and the various data management systems may be connected. The distributed data management system may process the sensor data emanating from the sensors. In some cases the data may be aggregated and sent to a central data

management system. We need tradeoff studies between developing distributed sensor data management systems and aggregating the data and managing at a central location. Figures 3 and 4 illustrate the various architectures.

Aggregating the sensor data and making sense out of the data is a major research challenge. The data may be incomplete or sometimes inaccurate. We need the capability to deal with uncertainty and reason with incomplete data. We also need to examine various distributed data management strategies including distributed query processing and managing metadata in a distributed environment. For example, we need to develop distributed query optimization strategies as well as techniques for data aggregation. Each sensor may have limited memory. Therefore, we need to examine techniques for managing distributed main memory sensor databases as well as examine distributed real-time data management and scheduling techniques for sensor data management.

Sensor Information Management

Information management includes extracting information and knowledge from data as well as managing data warehouses and mining and visualizing the data. Much work has been carried out on information management the last several years. We now need to examine the applicability of various information management technologies for managing sensor/stream data.

For example, sensor data has to be visualized so that one can better understand the data. We need to develop visualization tools for the sensor data. One may also need to aggregate the sensor data and possibly build repositories and warehouses. However, much of the sensor data may be transient data, and therefore we need to determine what data to store and what data to discard. Data

may also have to be processed in real time. Some of the data may be stored and possibly warehoused and analyzed for conducting analysis and predicting trends. That is, the sensor data emanating from surveillance cameras has to be processed within a certain time. The data may also be warehoused so that one can later analyze the data.

Sensor data mining is becoming an important area (see [5]). We need to examine the data mining techniques such as association rule mining, clustering, and link analysis for sensor data. As we have stressed, we need to manage sensor data in real time. Therefore, we may need to mine the data in real time also. This means not only building models ahead of time so that we can analyze the data in real time, but we may also need to build models in real time. That is, the models have to be flexible and dynamic. This is a major challenge. We also need many training examples to build models. For example, we need to mine the data emanating from sensors and detect and possibly prevent terrorist attacks. This means that we need training examples to train the neural networks, classifiers, and other tools so that they can recognize in real time when a potential anomaly occurs. Sensor data mining is a fairly new research area, and we need a research program for sensor data management and data mining. Figure 5 illustrates sensor information management.

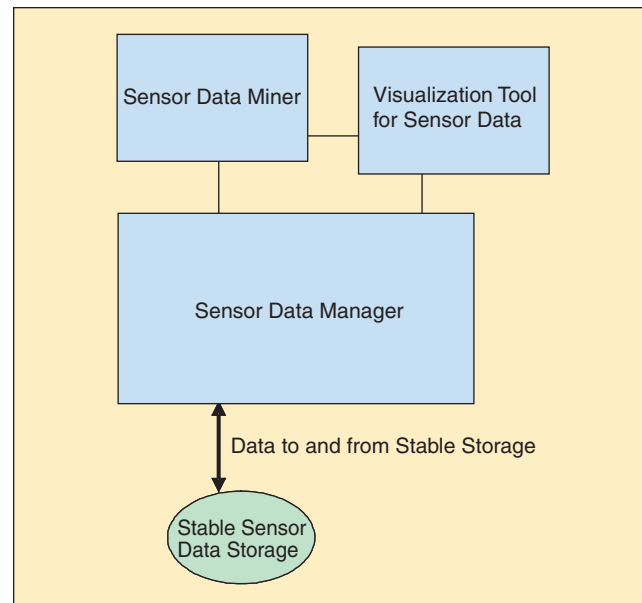
Security Issues

Much work has been carried out in securing data management systems. The early work was on access control, and later researchers focused on multilevel secure database management systems. More recent research focuses on role-based access control models as well as examining security for new kinds of databases as well as applications such as e-commerce and medical information systems (see, for example, [6] on data and applications security).

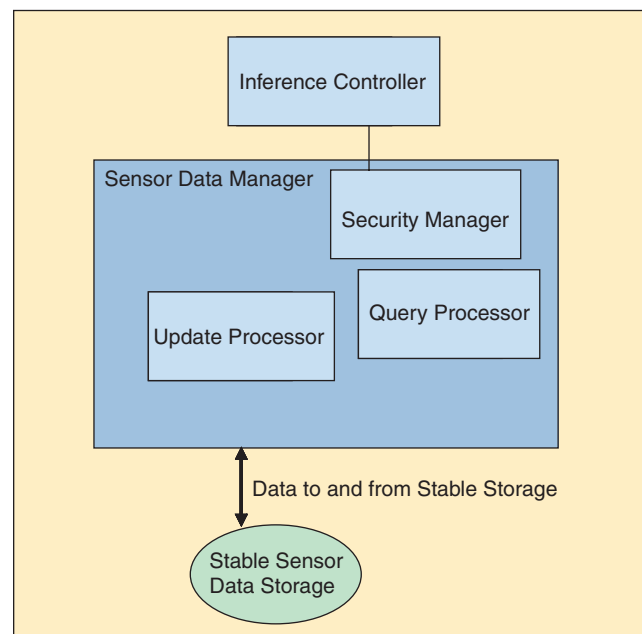
We need to conduct research on security for sensor databases and sensor information systems. For example, can we apply various access control techniques for sensor and stream databases? That is, can we give access to the data depending on the roles of the users, such as the airport security officer who has access to all of the sensor data emanating from the sensors while the airport ticketing agent may have limited access to certain sensor data. Another challenge is granting access to aggregated data. Individual data may be unclassified while the aggregated data may be highly sensitive. This is, in a way, a form of the inference problem in database systems. Note that inference is the process of posing queries and obtaining unauthorized information from the legitimate response received. Due to the aggregation and fusion of sensor data, the security levels of the aggregated data may be higher than those of the individual data sets. We also need to be aware of the privacy of the individuals. Much of the sensor data may be about individuals such as video streams about activities and other personal information. This data has to be protected from the general public and from those who are unauthorized to access the data. We

have looked at privacy as a subset of security (see, for example, [7]). There is also research on privacy preserving data mining and the techniques have to be examined for sensor data mining [8].

Finally we need to examine security policies for sensor data. These security policies may be dynamic, and we need to develop ways to enforce security constraints that vary with time. We also need techniques for integrating security policies especially in a networked and distributed environment. For example, different policies may apply for different sensor databases. These policies have to be integrated when managing distributed databases. Figure 6 illustrates a system architecture for managing security in a sensor data management environment. Integrating security policies is illustrated in Figure 7. One of the



▲ 5. Sensor information management.



▲ 6. Secure sensor data management.

major questions here is what are the special considerations for security for sensor and stream data? Do the access control models that have been developed for business data processing applications work for stream data? We need to start a research program on secure sensor networks and secure sensor information management.

Dependable Sensor Information Management

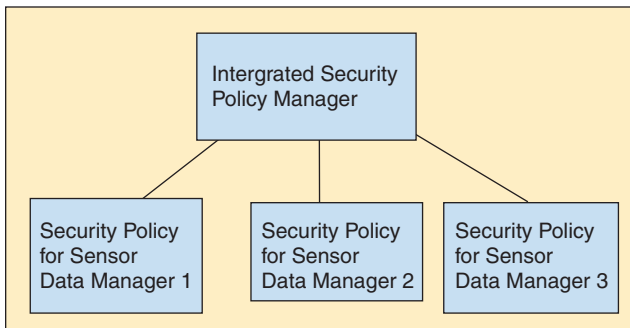
While the previous section focused on secure sensor data management, in this section we will discuss some other issues such as survivability, fault tolerance, and real-time sensor data management. For many applications, including command and control, intelligence, and process control, data are emanating from sensors. Timing constraints may be associated with data processing. That is, the data may have to be updated within a certain time, or it may be invalid. There are tradeoffs between security and real-time processing. It takes time to process the access control rules, and, as a result, the system may miss the deadlines. Therefore, we need flexible security policies. In certain cases real-time processing

may be critical. If we are to detect anomalies from the time a person checks in at the ticket counter until he boards the airplane, then this anomaly has to be detected within a certain time. In this case we may have to sacrifice some degree of security. In other cases, we may have a lot of time to say analyze the data, and in this case only authorized individuals may access the data.

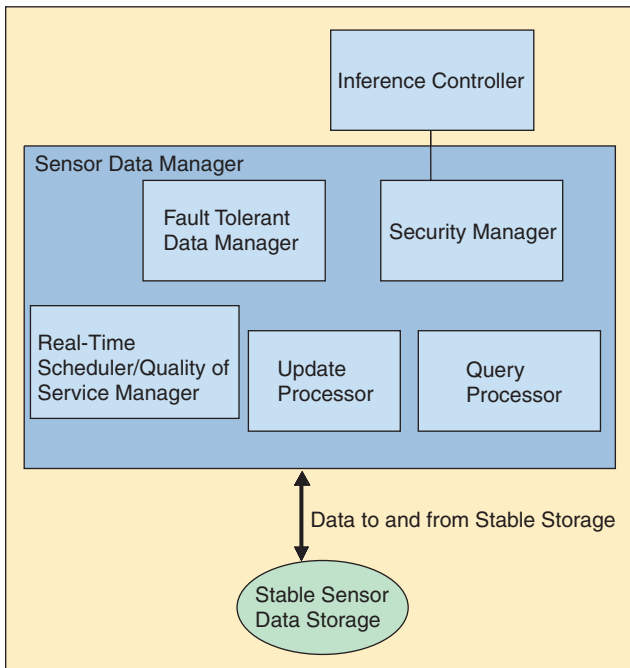
Other issues include fault-tolerant sensors and survivable sensors. Much work has been carried out on fault-tolerant data management. We need to examine the fault-tolerant data processing techniques for sensor data. Furthermore, these sensor databases have to survive from failures as well as from malicious attacks. Many of our critical infrastructures, such as our telephones, power lines, and other systems, have embedded sensors. The data emanating from these sensors may be corrupted maliciously or otherwise. We need to develop techniques for survivable sensor data management. Figure 8 illustrates additional components for sensor data management to manage faults as well as to enforce quality of service parameters.

Infrastructures for Sensor Information Systems

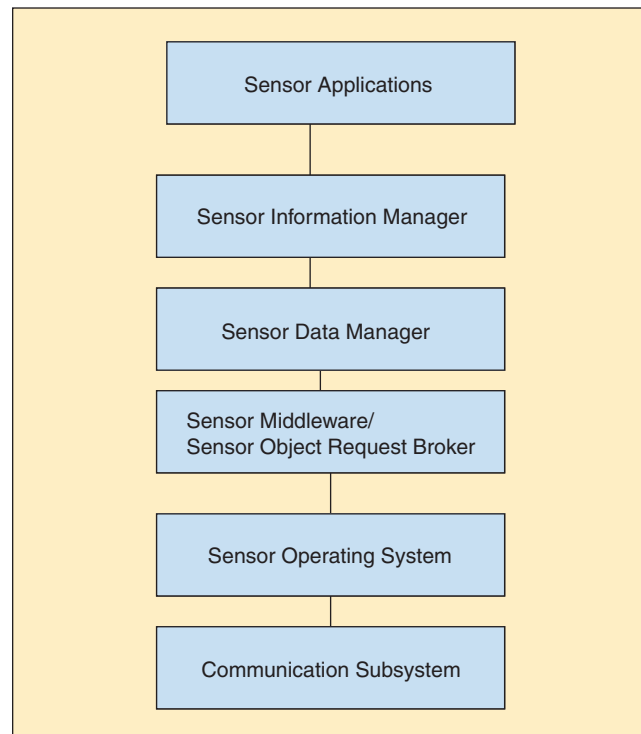
The previous sections discussed various aspects such as data management, information management, security, and survivability for sensor databases. We need to address all these issues and eventually build integrated systems. One possibility is to follow the approach we have developed for real-time command and control systems such as AWACS (advanced warning and control system). Here we developed an infrastructure consisting of a real-time object request broker and services using commercial real-time



▲ 7. Integrating security policies.



▲ 8. Survivable sensor data manager.



▲ 9. Integrated system.

operating systems. We then developed a real-time data manager and applications hosted on the infrastructure. We used object technology for integrating the various components. We also showed how such an infrastructure could be used to migrate legacy applications [10].

We can take a similar approach to build an integrated system for sensor information systems. We need appropriate operating systems and infrastructures possibly based on object request brokers. We need to host sensor data managers and applications such as multisensor data integration and fusion on the infrastructures. We need to ensure that the system is secure and survivable. We also need to ensure that the infrastructure is secure; that is, security has to be built into the system and not considered as an afterthought. Figure 9 illustrates a layered architecture for the integrated system of a sensor network. Figure 10 illustrates the use of object technology for integration.

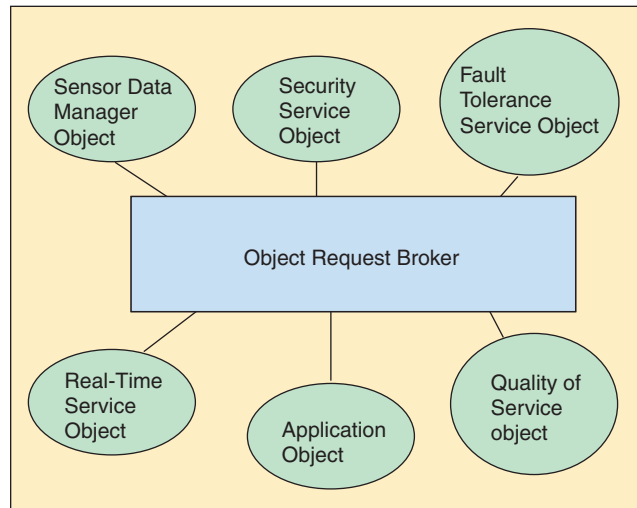
We also need to examine both centralized and distributed architectures for sensor data management. On the one hand we can aggregate the data and send it to a centralized data management system, or we can develop a full-fledged distributed data management system. We need to conduct simulation studies and determine tradeoffs between various architectures and infrastructures.

Summary and Directions

This article has provided an overview of secure sensor information management. First we discussed the need for sensor information management and then provided an overview of the issues involved in managing sensor databases. Then we discussed distributed data management issues for sensor data and provided an overview of sensor information management and sensor data mining. Next we discussed security issues such as access control for sensor data. Then we discussed quality of service issues such as tradeoffs among security, real-time processing, and fault-tolerant computing. Finally we discussed infrastructures and architectures for sensor information systems.

While sensor technology has been around for many decades, it is only recently that information technology is being integrated with sensor technology. Sensor information management has received much attention especially due to the fact that large amounts of sensor data are being collected, managed, and analyzed. Sensor data may be emanating from surveillance cameras and various embedded devices. Effective techniques for sensor information management will be critical for protecting our critical infrastructures as well as for our national security.

Various research programs are under way at funding agencies such as the National Science Foundation and the Defense Advanced Research Projects Agency on sensor information management and sensor networks (see for example, [11] and [12]). There are many challenges and many opportunities, and with various research programs we can expect much progress to be made.



▲ 10. Object for integration.

Acknowledgments

I thank NSF and MITRE Corporation for their support to continue my work on information management, data and applications security, data mining, and counter-terrorism.

Bhavani Thuraisingham is the program director for Data and Applications Security at the National Science Foundation (NSF) in Arlington, Virginia, and also manages the information management focus area for NSF's Information Technology Research, on leave from the MITRE Corporation. She received the M.Sc. degree from the University of Bristol and the Ph.D. degree from the University of Wales, both in the United Kingdom. She is an IEEE Fellow.

References

- [1] B. Thuraisingham, *Web Data Mining and Applications in Business Intelligence and Counter-Terrorism*. Boca Raton, FL: CRC Press, 2003.
- [2] Stanford Sensor Data Management Group, "STREAM: The Stanford stream data manager," *IEEE Data Eng. Bull.*, 2003.
- [3] A. Dobra, M. Garofalakis, J. Gehrke, and R. Rastogi, "Processing complex aggregate queries over data streams," in *Proc. 2002 ACM Sigmod Int. Conf. Management of Data*, Madison, WI, 2002.
- [4] D. Carney, U. Çetintemel, A. Rasin, S. Zdonik, M. Cherniack, and M. Stonebraker, "Operator scheduling in a data stream manager," in *Proc. 29th Int. Conf. Very Large Data Bases Berlin, Germany*, 2003.
- [5] K. Rajgopal Kannan, S. Sarangi, S. Ray, and S. Sitharama Iyengar, "Minimal sensor integrity in sensor grids," in *Proc. Int. Conf. Parallel Processing*, 2002.
- [6] B. Thuraisingham, "Data and applications security: Developments and directions," in *Proc. IEEE COMPSAC Conf.*, Oxford, UK, Aug. 2002.
- [7] B. Thuraisingham, "Data mining, national security, privacy and civil liberties," *SIGKDD Explorations*, Jan. 2003.
- [8] J. Gehrke, "Special issue on data mining and privacy," *SIGKDD Explorations*, Jan. 2003.
- [9] R. Brooks and S.S. Iyengar, "Robust distributed computing and sensing algorithm," *IEEE Comput.*, June 1996.
- [10] B. Thuraisingham and J. Mauer, "Survivability of real-time command and control systems," *IEEE Trans. Knowledge Data Eng.* Jan. 1999.
- [11] National Science Foundation, <http://www.cise.nsf.gov/fndg/pubs/display2.cfm?pubid=5623&div=anir>
- [12] Defense Advanced Research Projects Agency, <http://dtsn.darpa.mil/ixo/programdetail.asp?progid=42>