DEPENDABLE INFRASTRUCTURES AND DATA MANAGERS FOR SENSOR NETWORKS

Dr. Bhavani Thuraisingham

The MITRE Corporation
(at present on leave at the National Science Foundation)

ABSTRACT

This paper provides some directions for developing infrastructures and data managers for dependable sensor networks. By dependable systems we mean systems that adapt to the environment and are secure, fault tolerant and process data in real-time as needed. We start with a discussion of the need for dependable sensor information management and then provide an overview of dependable infrastructures and data managers for such networks. We also discuss some security issues for sensor information management. Finally some directions for further research are given.

1. INTRODUCTION

Sensor networks and sensor information management are critical technologies for many applications including process control, manufacturing and more recently for detecting and preventing terrorism. Sensors are embedded in a myriad of devices and locations including in shops, theatres, airports, railway stations, hospitals and many other public buildings. Data in the form of streams emanate from sensors and these streams have to be 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 include data tracking various individuals or data 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 under 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, airports and other public places. This video data may also be in the from of streams and will have to be mined to detect suspicious behavior.

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 find out whether similar attacks have occurred and what emergency response measures to take. In order 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 networks of 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 and information management considerations for sensor data management and mining. These include special data modeling, query 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 multi-sensor data fusion and sensor data mining.

This paper will discuss infrastructures and information/data managers for dependable sensor networks. By dependable we mean sensor networks that are secure, survivable, fault tolerant and can process data in real-time. There is quality of service tradeoffs that one needs to make to build dependable sensor networks. Our previous experience on 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, real-time processing and fault tolerant computing capabilities. The data managers have to process queries, execute transactions and manage data

Proceedings of the Ninth IEEE International Workshop on Object-Oriented Real-Time Dependable Systems (WORDS'03)
0-7695-2054-5/04 $20.00 © 2004 **IEEE**

that may be transient. Data may also be mined to extract information.

The sensor data/information manager will manage sensor/stream data. Various aspects such as data models and query strategies for sensor data processing need to be examined. The sensor data/information manager will be hosted as an application on the object-based infrastructure. Processing may be distributed among multiple sensor nodes and there has to be coordination between the nodes. Essentially we are proposing a distributed sensor information management system. One could also think of this system as a peer-to-peer system where the nodes are peers that have to work together to solve a problem.

The ideas in this paper are preliminary. It builds on the work we have carried out on real-time command and control systems based on objects and presented at several WORDS (Workshop on Object-Oriented Real-time Dependable Systems) and ISORC (International Symposium ob Real-time Computing) conferences between 1996 and 2002. Our goal is to use existing technologies and develop new technologies as needed to build an environment for sensor information management as well as host various applications. As we have stated, sensor information management and sensor data mining are becoming critical for many applications including preventing and detecting terrorism. Therefore, it is important to build dependable and survivable infrastructures and data managers for sensor networks. The organization of this paper is as follows. Dependable infrastructures will be discussed in Section 2. Dependable data managers will be discussed in Section 3. Some security considerations are discussed in Section 4. Directions are given in Section 5.

2. DEPENDABLE INFRASTCRTURES

For many applications including command and control, intelligence and process control, data is 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. That is, 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. For example 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. For example, how can we ensure that the aggregate data is valid even if the components that constitute the aggregate data may be corrupted? Some directions are given in [PERR03].

One possibility for developing dependable infrastructures and data managers for sensor networks 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 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 (see [BENS96], [GATE97]).

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 multi-sensor 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 after-thought. Essentially we are proposing a layered architecture for sensor information management. The infrastructure consists of middleware possibly based on object request brokers for sensors. The objects that constitute the middleware include objects for interprocess communication, memory management, and support for data fusion and aggregation. On top of the infrastructure we host a sensor data manager to be discussed in the next section. The data manager essentially manages sensor data, which may be in the form of streams. 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.

As part of our work on evolvable real-time command and control systems we examined real-time processing for object request brokers and we were instrumental in establishing a special interest group, which later became a task force within the Object Management Group (OMG). Subsequently commercial real-time object request brokers were developed. The question is, do we need special purpose object request brokers for sensor networks? Our challenge now is to develop object request brokers for sensor networks. We need to examine special features for object request brokers for managing sensor data. We may need to start a working group or a special interest group within OMG to investigate these issues and subsequently specify standards for object request brokers to manage sensor data.

Another aspect that is important is end-to-end dependability. This includes security, real-time processing and fault tolerance for not only all of the components such as infrastructures, data managers, networks, applications and operating systems; we also need to ensure the dependability of the composition of the entire system. This will be a major challenge even if we consider security, real-time processing and fault tolerant computing individually. Integrating all of them will be a complex task and will have to address many research issues and challenges.

3. DEPENDABLE DATA MANAGERS

Various research efforts are under way to develop sensor data management systems. These include the efforts described in [STAN03], [DOBR02], and [CARN03] (see for example the work at Stanford University, Cornell University MIT, University of California Berkeley, Brown University and Brandies 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 data. 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 queries. We need special query strategies for processing such continuous queries. 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 (Structured Query Language) are being extended with constructs for querying stream data. Research efforts are also under way for extending XML (eXtensible Markup Language) 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.

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

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 trade-off studies between developing distributed sensor data management systems and aggregating the data and managing at a central location.

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.

Information management includes extracting information and knowledge from data as well as managing data warehouses, mining the data, and visualizing the data. Lot of 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 which data to store and which 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 [KANN02]). 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, 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.

Our approach to building a real-time data manager for the command and control systems is to host it as an application object on top of the infrastructure, which is essentially an object request broker. The data manager is also based on an object model tailored for real-time applications. The data manager manages the track information. We also developed special concurrency control algorithms for real-time data processing. We need to examine this research and see if we can take a similar approach for a sensor network. Essentially can we develop an object model for sensor information management? How do we host such a data manager on top of the infrastructure developed for sensor networks? Furthermore, how do we ensur5e security and dependability? As we have stated integrating security, fault tolerant computing and real-time processing is a major challenge. In the next section we will examine some security issues for sensor information management.

4. SECURITY ISSUES

As we have stated in Section 1, we need dependable infrastructures and data managers that are flexible, secure, fault tolerant and can process data in real-time. That is, we need flexible policies that can adapt to the environment. In this section we discuss some aspects of security.

Much work has been carried out on securing data management systems (see [FERR00]). The early work was on access control and later researchers focused on multilevel secure database management systems. More recently research is focusing 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 [THUR02] 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 air port security officer 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 responses 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, [THUR03a]). There is also research on privacy preserving data mining and the techniques have to be examined for sensor data mining [GEHR03].

Finally we need to examine security policies for sensor data. These security policies may be dynamic and therefore 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. One of the 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. Some preliminary directions are given in [THUR03b] and [THUR03c].

As we have stated in section 3, we need end-to-end security. That is, the infrastructures, data managers, applications, networks and operating systems have to be secure. OMG has developed standards for securing object request brokers. We need to take advantage of such developments. However, we need to examine security issues specific to object request brokers for sensor information management. We also need to examine composability of the various secure components. For example, are the interfaces between the various components satisfying the security properties? How can we integrate or compose the security policies of the various components. There is little work reported on securing large-scale information systems. Now, we not only have to examine security for such systems, we also need to examine security for such systems that manage and process sensor data. In addition, we need not only secure systems but also systems that are dependable and survivable. As stated in section 3, developing such systems will be a major challenge.

One approach is to develop various features incrementally for data managers and middleware. However this often means that at the end some features are left out. For example, if we build components and examine only security and later on examine real-time processing, then it may be difficult to incorporate real-time processing into the policy. This means that we need to examine all of the features simultaneously. This means security engineering has to be integrated with software engineering.

5. SUMMARY AND DIRECTIONS

This paper has provided some directions in developing infrastructures and data managers for dependable sensor networks. By dependable systems we mean systems that adapt to the environment and are secure, fault tolerant and process data in real-time as needed. We started with a discussion of the need for dependable sensor information management and then provided an overview of dependable infrastructures and data managers for such networks. We also discussed some security issues for sensor information management.

The ideas presented in this paper are preliminary. There is still a lot to be done. Security research for sensor database management is just beginning. We need to develop security policies and architectures for sensor information management. We also need to develop appropriate infrastructures possibly based on objects. We need to examine object request brokers and investigate ways to support sensor networks. We need end-to-end dependability. That is, while each object or component may be dependable, we also need to investigate the composability of the dependable objects. Depending on the level of assurance that is required of the system, we may need to investigate formal method and testing methods for sensor network. This paper has provided some preliminary directions.

The web is having a major impact on many technologies including for sensors, databases and middleware. That is, sensors are connected via the web and databases are accessed via the web. Therefore, we need to examine web security techniques as well as web data management techniques for secure sensor information management on the web. Other research areas include peer-to-peer sensor information management where collections of sensors act as peers and share information with each other. Trust management also needs consideration for sensor networks. In summary, the security technologies include access control based on user roles and credentials, data mining for detecting and preventing intrusions, web security, trust management as well as encryption. Finally for all of the areas we need to integrate security, fault tolerant computing and real-time processing to build dependable sensor information systems and networks.

ACKNOWLEDGEMENT

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

DISCLAIMER

The views and conclusions expressed in this paper are those of the author and do not reflect the policies or procedures of the National Science Foundation, the MITRE Corporation or of the US Government.

REFERENCES

[BENS96] E. Bensley, M. Gates, P. Krupp, A. Shafer, M. Squadrito, B. Thuraisingham and T. Wheeler, Object-Oriented Approach to Designing an Infrastructure and Data Manager for Real-time Command and Control System, Proceedings IEEE Words, 1996.

[CARN03] D. Carney, U. Çetintemel, A. Rasin, S. Zdonik, M. Cherniack, and M. Stonebraker, Operator Scheduling in a Data Stream Manager. Proceedings of the 29th International Conference on Very Large Data Bases Berlin, Germany, 2003.

[DOBR02] A. Dobra, M. Garofalakis, J. Gehrke, and R. Rastogi. "Processing Complex Aggregate Queries over Data Streams", Proceedings of the 2002 ACM Sigmod International Conference on Management of Data, Madison, WI, 2002.

[FERR00] Ferrari E., and B. Thuraisingham, Database Security, Artech House 2000 (editors: M. Piattini and O. Diaz)

[GATE97] M. Gates, P. Krupp, J. Maurer, A. Shafer, M. Squadrito, B. Thuraisingham and T. Wheeler, Object-Oriented Approach to Implementing an Infrastructure and Data Manager for Real-time Command and Control System, Proceedings IEEE Words, 1997.

[GEHR03] J. Gehrke, Special Issue on Data Mining and Privacy, SIGKDD Explorations, January 2003.

[KANN02] K. Rajgopal Kannan, S. Sarangi, S Ray and S. Sitharama Iyengar, Minimal Sensor Integrity in Sensor Grids, Proceedings of the International Conference on Parallel Processing, 2002.

[PERR03] A. Perrig, SIA: Secure Information Aggregation in Sensor Networks, CMU Report, 2003.

[STAN03] Stanford Sensor Data Management Group, STREAM: The Stanford Stream Data Manager, IEEE Data Engineering Bulletin, 2003.

[THUR99] B. Thuraisingham and J. Mauer, Survivability of Real-time Command and Control Systems, IEEE Transactions on Knowledge and Data Engineering, January 1999.

[THUR02] B. Thuraisingham, Data and Applications Security: Developments and Directions, Proceedings of the IEEE COMPSAC Conference, Oxford, UK, August 2002.

[THUR03a] B. Thuraisingham, Web Data Mining and Applications in Business Intelligence and Counter-terrorism, CRC Press, 2003.

[THUR03b] B. Thuraisingham, Secure Sensor Information Management: Issues and Directions, Accepted in IEEE Signal, 2003

[THUR03c] B. Thuraisingham, Security and Privacy Issues for Sensor Database Systems, Accepted in Sensor Letters, 2003.