Challenges of Complex Data Processing in Real World Sensor Network Deployments

P. J. Marrón R. Sauter O. Saukh M. Gauger K. Rothermel
University of Stuttgart
Institute of Parallel and Distributed Systems (IPVS)
Universitätsstr. 38, D-70569 Stuttgart, Germany
{marron, sauter, saukh, gauger, rothermel}@informatik.uni-stuttgart.de

ABSTRACT

Long-term deployments of wireless sensor networks so far have been focused on the periodic gathering of simple sensor readings. However, technological advances allow working with more complex types of data that also create larger data volumes. In this paper we use a real world engineering application to identify a series of challenges related to complex data processing in sensor networks. We present a classification of these challenges and outline a set of preliminary solutions that we are currently in the process of developing for our motivating application.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software; C.2.4 [**Computer-Communication Networks**]: Distributed Systems; C.3 [**Computer Systems Organization**]: Special-Purpose and Application-Based Systems

Keywords

Sensor networks, data processing, challenges

1. INTRODUCTION

Since their establishment as a research area several years ago, sensor networks have been used in a wide variety of scenarios. However, in the early days and in most current deployments, applications deal mostly with environmental data such as temperature values, humidity, etc. which allow the system to infer the conditions of a large physical area.

Nowadays applications are becoming more and more complex and require the gathering, processing and forwarding of larger amounts of data. A prominent example of such applications is developed in engineering scenarios where normally the data unit is no longer a single temperature value but a complex signal represented as a series of discrete values over time. Moreover, these monitoring applications tend to combine data gathered by several sensors in order to decide whether or not an action has to be taken or a warning needs to be issued.

In the future, it is expected that applications will use even more complex data such as images from low-power cameras, so that the need for algorithms and architectures capable of dealing with large amounts of data using resource-constrained sensors and sensor networks will reach unprecedented heights.

This paper addresses some challenges and possible solutions associated with the gathering, in-network processing and forwarding of more complex data, such as time series, in wireless sensor networks. For this purpose, we concentrate on a specific application implemented as part of the European Project Sustainable Bridges [15] whose goal is to design a monitoring system based on wireless sensor technology that is able to detect structural defects autonomously.

The remainder of this paper is structured as follows: The next section deals with some related work and points out the differences to our approach. Section 3 provides more details about the Sustainable Bridges application needed to understand the challenges and solutions described in sections 4 and 5 respectively. Finally, section 6 concludes this paper and gives some insight about future work.

2. RELATED WORK

Several publications have already discussed challenges in the area of wireless sensor networks. Topics discussed in these challenge papers range from early visions of sensor networks and their specific challenges [2][5] through challenges identified based on research experiences [4] to challenges of specific subareas [1][3][10]. The need for collaborative processing of signals and information among the nodes is mentioned as an important challenge [3]. However, the main topic of this paper – the need for processing complex data in sensor networks and challenges related to it – has not been discussed in detail yet.

The importance of monitoring applications for wireless sensor networks has been discussed in several survey papers and is also demonstrated by classifications of sensor network applications [2] [12]. In order to reduce the amount of data to be stored in memory and transmitted over a network, different solutions have been proposed in different areas: digital signal processing, data compression and data aggregation.

The processing and filtering of analog measurements that have been converted to a digital signal is traditionally done with Digital Signal Processing (DSP) techniques [11]. Be-

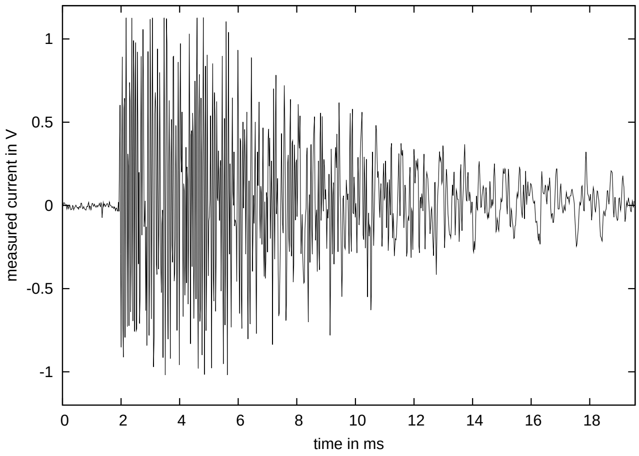

Figure 1: Example measurement of an acoustic emission emitted by a pencil lead break on concrete

sides filtering of data, one important application of DSP is data compression. However, the numerical operations required for DSP are often very complex which limits their applicability in wireless sensor networks.

Related to DSP mechanisms, there are different lossless as well as lossy data compression techniques used both for savings in the communication and storage of data [14]. However, efforts to exploit these techniques for wireless sensor networks have to find special solutions to deal with the resource constraints of the sensor nodes [8].

In-network processing and data aggregation are common and relatively well understood techniques to reduce the traffic load in wireless sensor networks [7][9]. However, there is a lack of techniques for processing and aggregating more complex data.

3. SUSTAINABLE BRIDGES

Traditional monitoring of civil structures, like bridges and buildings, has been performed either by visual inspection or by installing collections of sensors that communicate with each other using some kind of cable technology. In the first case, the interpretation and assessment of the structure is based on the experience of the expert, whereas the second method requires the expensive installation of kilometers of cables to cover the object subject to observation.

Therefore, one of the main goals in the Sustainable Bridges project is to develop and provide the necessary infrastructure and algorithms to allow for the cost-effective detection of structural defects and to better predict the remaining lifetime of these structures. The project is committed to the use of wireless sensor network technology in order to reduce the costs involved in the installation and maintenance of a cabled system. To achieve this, a number of sensor network challenges need to be solved.

Research is going on in the field of modal analysis and stress, strain and displacement measurement using the same wireless sensor network. In this paper we primarily concentrate on audible information (acoustic emission) to describe our challenges.

3.1 Acoustic Emission

Acoustic emission [6] belongs to the class of nondestructive material testing methods and exploits the fact that structural defects emit elastic waves which travel through the material and can be recorded by sensors on the surface. Theses effects are related – albeit on a different scale – to earthquakes caused by continental drift. Acoustic emission is therefore a passive method that captures spontaneous events.

Pencil lead breaks are used in laboratory tests to simulate defects in concrete structures. The generated emissions are similar enough to the ones created by real structural defects. For these tests sampling frequencies of 1 MHz are normally used and several thousands of samples are recorded. However, typical low end wireless sensor nodes are not able to handle such amounts of data. To solve this challenge, civil engineers and sensor network specialists must cooperate closely in developing new approaches to acoustic emission analysis. Civil engineers need to define lower bounds for the required quality and resolution of the data whereas sensor network specialists have to utilize the resources of the hardware as efficiently as possible. Combining the expertise of both fields, it should be possible to find solution approaches that build upon the special characteristics of sensor networks, like for example parallel processing or aggregation.

Fig. 1 shows an example of an acoustic emission signal recorded at 50 kHz by a real sensor node. The recording of the signal is triggered when the value crosses a certain predefined threshold. However, the analysis of acoustic emission signals also requires the data from a short time interval before the trigger. This so called pretrigger interval can be seen on the left side of the figure and has a length of 100 samples in this particular case.

3.2 Characteristics of Monitoring Applications

The Sustainable Bridges application is a typical example of a monitoring application that benefits from the deployment of sensor network technology. All sensor nodes are affixed to the structure, communicate in an ad-hoc fashion with their neighbors and are usually part of node clusters. The network is composed of temperature and humidity sensors, vibration sensors, material stress sensors, etc., located at specific points within and outside the bridge.

The purpose of these clusters of sensors is to analyze the data from the sensors and coordinate the decision process used for determining whether or not a structural defect has occurred using acoustic emission techniques. In cases showing a potential problem, enough sensor data needs to be forwarded to the base station so that the civil engineer in charge can make an informed decision.

4. CHALLENGES OF DATA PROCESSING

Dealing with data in sensor networks – simple sensor readings as well as complex signals – poses a set of challenges that can be classified into three main categories: Data gathering and storage, data analysis and processing, and data forwarding. Each of these challenge areas is described in detail in the following sections.

4.1 Data Gathering and Storage

For simple sensor readings, data gathering does not pose special problems to sensor nodes. However, the requirements for the acquisition of complex signals cannot be underestimated. On the one hand, *sampling rates of up to 200 kHz* for our acoustic emission application example require algorithms and implementations that are fine-tuned to the respective hardware. The sampling resolution, on the other hand, strictly depends on the hardware properties and can either be achieved by the hardware itself or not.

Since it is not possible to give real-time guarantees using low-end sensor hardware, the algorithms in the data analysis and processing stages must consider that *samples might have been lost*.

Another characteristic of our application which also applies to many other scenarios is the *event based nature of the application task*. Although many current information retrieval methods for wireless sensor networks sample data periodically, there are also many scenarios containing unpredictable events in signals that have to be detected exactly at the moment they occur. A purely software-based solution for real time event recognition is, therefore, not feasible on low end sensor nodes. A combined software and hardware solution is needed. The hardware part can be as simple as a trigger, potentially operating on a narrow frequency range determined by appropriate sensor thresholds, leaving the actual classification of a signal to the software. We use this approach for our acoustic emission application. However, there are two problems that need to be solved: Since the sensor nodes often are in power saving mode when a relevant event occurs, the *wakeup time tends to be too long* to allow the recording of the complete signal. The second problem is the *difficulty in capturing the pretrigger interval* of the event needed to analyze the acoustic emission.

The example of detecting acoustic emission events shows the necessity of interdisplinary research: Observations have shown that in many cases multiple acoustic emission events occur in rapid succession. This allows the first event signal to be used as a trigger to wake up the node and start continuous sampling to catch following events, thus completely including the pretrigger interval.

High requirements for the storage of data is another problem not found when only dealing with simple sensor readings. The amount of main memory required for storage of application data limits the choice of suitable sensor hardware, because the potential savings achievable by using a more efficient software solution have hard limits: The most-advanced version of our acoustic emission application uses four channels sampling with a combined frequency of 200 kHz. The lower bound for the signal length has yet to be determined but may be as high as 512 samples, thus occupying 4096 bytes of RAM which is the total amount available in the MICA2 motes.

Another problem is the *necessity of long term storage*. Here we also need very efficient solutions saving on the data size and the number of access operations, because writing to the flash memory of the sensor nodes is a very energy-intensive operation. This requirement is tightly coupled to the other two categories of our classification because it shares the requirement to reduce data size with the data forwarding problem category and both need support from data analysis algorithms.

4.2 Data Analysis / Data Processing

One way of classifying data analysis and data processing algorithms is based on whether they process signals in real-time during the sampling process or whether they operate on the complete data set after it has been recorded. The former approach allows more sophisticated ways of restricting the analysis effort to the relevant parts of the incoming signal and also reduces the storage requirements. However, it is not feasible for high frequency sampling and many algorithms cannot work on data streams.

Data processing algorithms that work on a recorded signal have a much broader spectrum of applications. For instance, it is possible to use ad-hoc or statically configured clusters for advanced processing of the data. This assembly of data from different nodes can be used to reduce errors or to fulfill application requirements. In our acoustic emissions example, one localization scheme requires multiple nodes working together on their collected data. *Distributed analysis of data* requires a cluster management customized to application specific requirements and also time synchronization among the nodes.

Apart from special solutions tailored to specific application requirements, a set of *generic algorithms* should be able to support the demands of applications or provide a foundation to build solutions on. Possible problem fields for such generic approaches might be *sensor calibration* and *data comparison*.

One class of processing algorithms that is strongly related to both data storage and forwarding are *compression algorithms*. In the optimal case, the signal is completely analyzed on the sensor node and the information to be stored and transmitted can be reduced to some meta-information like a classification and a timestamp. In many cases, however, this is not possible due to the resource constraints of the sensor node hardware. In other cases, doing the analysis in the sensor network is more expensive than transmitting the complete data to a central base station. Algorithms for data compression are also candidates for the development of generic solution approaches.

A second field in the area of digital signal processing for which many algorithms already exist is the *reduction of signal noise*. Noise reduction and filtering algorithms have to be selected based on their efficiency on sensor node hardware and the requirements of the application. We expect this to require substantial development effort for adaptation and optimization.

4.3 Data Forwarding

There are two extremes considering the amount of data that must be forwarded between nodes: Sending the complete data or sending only analysis results. Existing routing protocols can be used if only the results of the data analysis must be sent. However, if signal data must be sent to the base station it might be necessary to develop *new data forwarding solutions*. The extreme case is when the complete analysis is done at the base station and all collected data must be sent there. A simple example can be used to illustrate the dimensions we are dealing with here. If events are to be detected in a small network of 30 nodes with an average distance of 2 hops to the base station and on average one event is detected every day then about 240 KB of data have to be transmitted not including packet overhead or retransmissions. If you compare this to an hourly retrieval of five simple sensor readings, which results in about 7 KB of data, you get a feeling for the costs of collecting complex event data.

A lot of research has been conducted to develop *efficient data aggregation systems* for simple sensor readings. There is a clear need for corresponding algorithms working on com-

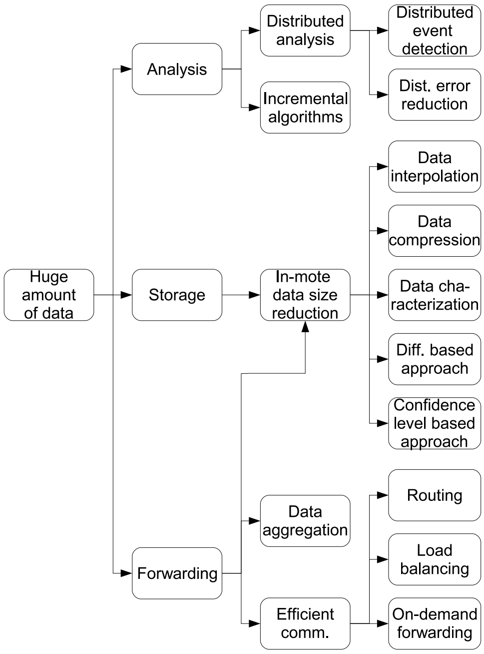

Figure 2: Solution spectrum for handling huge amounts of data

plex signals. Attributes for classifying these algorithms primarily include data size reduction and the impact on error rates. It is even possible to reduce these error rates by combining measurements from different sensor nodes.

Analogous to data processing, it is possible to use either application specific or *generic algorithms* for data forwarding. Additionally, data processing and data forwarding algorithms may be combined to achieve optimal system behavior with reduced development effort.

Efficient routing protocols and high quality link estimation capabilities have to provide the foundation for higher level aggregation functions. The requirements regarding reliability are tightly coupled to the rest of the system.

The abilities of the server can have a big impact on the performance of the overall system. If its data processing and visualization algorithms are able to *cope with partially distorted or incomplete signals* the requirements on the performance of all sensor network algorithms – data gathering, processing and forwarding – can be reduced thereby saving resources and development costs.

5. PRELIMINARY SOLUTIONS FOR SUSTAINABLE BRIDGES

All three classes of challenges have in common the need for both generic but customizable as well as application-specific solutions. Only a solid base of well-understood methods allows solutions for the unique domain of wireless sensor networks that combine acceptable development effort with an adequate level of performance.

In Fig. 2 we present an overview of the spectrum of possible solution approaches we envision for the main problem of coping with huge amounts of data in sensor networks.

5.1 Data Gathering and Storage

As described previously, one distinguishing feature of data gathering solutions relates to the problem of deciding which functionality should be implemented in hardware and which in software. We opted for a *strong focus on the software implementation* in order to combine maximum flexibility with low hardware costs. We use a custom-built sensor board which provides a high pass filter and an anti-aliasing filter as well as an amplifier. The sensor board contributes to the goal of saving energy by providing a hardware trigger with a configurable threshold that allows the sensor node to stay in a low power state for most of the time.

Lossless data compression is only of limited use when working with complex data signals. *Lossy compression*, on the other hand, has shown in the last years its usefulness for compressing image, audio and video data. It remains an open question whether existing algorithms provide sufficient accuracy for applications like acoustic emission or whether new solutions must be developed. Furthermore, resource constraints of wireless sensor networks prohibit the use of most advanced algorithms. An optimal solution would both be adaptive, considering the actual data to control the compression level, and configurable by allowing the application to specify different accuracy requirements. These requirements could be determined by preliminary data analysis or specified by the user.

Another approach, perhaps best named *data characterization*, allows the application to store only certain features of a recorded signal. Statistics provides a whole range of processing techniques to characterize data series. Very simple examples are the average or the standard deviation of a signal. However, collecting such a simple set of metrics is certainly not sufficient for acoustic emission and most other applications working on complex signals. A more difficult example involves the transformation from the time to the frequency domain which allows for a considerable reduction of storage requirements.

Storing only the differences between a recorded signal and the expected values is another method for reducing the data size in the network. In the case of acoustic emission, the basic assumption that allows to apply this optimization is the high level of correlation among waves emitted by structural defects. Thus, after the data analysis has shown that a recorded signal is indeed an acoustic emission, storing only the differences to an expected signal promises high compression rates. The approach can also be combined with data characterization to store, for example, only the deviation of the spectrum of a recorded signal from a predetermined reference pattern.

5.2 Data Analysis / Data Processing

Analysis algorithms that work in an efficient way directly on streaming data seem to be the solution for overcoming storage constraints. They are particularly needed when a hardware trigger – a simple threshold or a more advanced solution – cannot be used because the event detection is too complex. However, *continuous sampling and data processing* are not yet feasible in long-running sensor network deployments without an external power source due to their high-energy consumption. Nevertheless, they can be used in combination with a *method that independently determines intervals of high event probability.*

An intrinsic characteristic of wireless sensor network algorithms is network wide data processing and in particular *data processing in clusters.* A straightforward approach to reduce the number of false positives is *majority voting* within a cluster. For the Sustainable Bridges project, the use of predefined clusters is particularly simple and beneficial as the network topology is static and engineers manually determine the locations of all nodes. A new quality of data processing is constituted by the calculation of the location of defects using multiple measurements and time differences within a cluster.

Iterative approaches for data analysis algorithms allow high accuracy in combination with low energy usage. If computationally inexpensive algorithms can be used to identify most recorded signals as irrelevant, more energy can be spent on further analysing relevant portions of the data. This is especially important for advanced cluster based processing algorithms that require wireless communication and are therefore much more energy-intensive than local computations.

Confidence estimations are another useful approach for reducing the energy consumption of the network as a whole. If the result of the analysis of a signal is clear, there is no need to store or forward the signal itself. However, if the first level of analysis do not produce a clear answer, storing and forwarding the signal is necessary to allow for more advanced analysis algorithms at the base station to decide on whether or not the signal represents a structural defect.

5.3 Data Forwarding

Like for storing data, new methods for data forwarding are mainly important when the complete (or only slightly filtered) signal must be transmitted to a base station.

With an increasing amount of data to be transmitted, link quality estimations gain importance for the routing of data. We developed a *new routing metric GEM* [13] that allows a better route selection and whose applicability is not limited to the sustainable bridges project.

Route selection based on energy dispersion (load balancing) is another very important attribute of the forwarding algorithm to prevent the premature energy depletion of single nodes which would reduce the accuracy of the network as a whole as all nodes should not only provide routing services but also participate in the sensing process.

While using the previously mentioned methods is gaining importance due to the increasing data volume, there is also the need for completely *new aggregation algorithms* that are able to work on time series. The behavior of these algorithms should not only adapt to application requirements but also to the actual signal data. While parameters like the maximum error ratio are rather well understood for statistical and digital signal processing methods, wireless sensor networks also need to integrate energy consumption and ultimately network lifetime as configuration properties.

Yet another approach is *on-demand forwarding* allowing the server – based on more complete knowledge and advanced algorithms – to determine the required accuracy for selected parts of the signal or even allowing the engineer to request individual parts of the data for manual inspection.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have shown, using the Sustainable Bridges application as a running example, the kinds of issues and challenges associated with the gathering, processing and forwarding of complex data items such as time series. We have also proposed some preliminary solutions that, when fully implemented for the Sustainable Bridges project, promise interesting results.

Regarding future work, there is still a lot of work to do. We will continue the implementation of the proposed solutions and will perform real world experiments on existing bridges as soon as a complete system is available. We also envision that the solutions proposed for this particular example can be used on other monitoring applications that use similar types of data.

7. REFERENCES

[1] I. F. Akyildiz and I. H. Kasimoglu. Wireless Sensor and Actor Networks: Research Challenges. *Ad Hoc Networks*, 2(4):351–367, 2004.

[2] I. F. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. Cayirci. A Survey on Sensor Networks. *IEEE Communications Magazine*, 40(8):102–114, 2002.

[3] H. Chan and A. Perrig. Security and Privacy in Sensor Networks. *IEEE Computer*, 36(10):103–105, 2003.

[4] C.-Y. Chong and S. P. Kumar. Sensor Networks: Evolution, Opportunities, and Challenges. *Proc. of the IEEE*, 91:1247–1256, 2003.

[5] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. In *Proc. of the 5th ACM/IEEE Intl. Conf. on Mobile Computing and Networking*, 1999.

[6] C. Grosse. Monitoring of Large Structures using Acoustic Emission Techniques. In *In Proc. of the Symposium "Nietdestructief onderzoek in de bouwsector"*, pages 15–24, 2003.

[7] J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan. Building Efficient Wireless Sensor Networks with Low-Level Naming. In *Proc. of the 18th ACM Symposium on Operating Systems Principles*, 2001.

[8] N. Kimura and S. Latifi. A Survey on Data Compression in Wireless Sensor Networks. In *Proc. of the Intl. Symposium on Information Technology: Coding and Computing*, April 2005.

[9] B. Krishnamachari, D. Estrin, and S. B. Wicker. The Impact of Data Aggregation in Wireless Sensor Networks. In *Proc. of the 22nd Intl. Conference on Distributed Computing Systems*, 2002.

[10] K. Lorincz, D. J. Malan, T. R. F. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton. Sensor Networks for Emergency Response: Challenges and Opportunities. *IEEE Pervasive Computing*, 3(4):16–23, 2004.

[11] P. A. Lynn and W. Fuerst. *Introductory Digital Signal Processing with Computer Applications*. John Wiley & Sons, Inc., New York, NY, USA, 1989.

[12] K. Römer and F. Mattern. The Design Space of Wireless Sensor Networks. *IEEE Wireless Communications*, 11(6):54–61, Dec. 2004.

[13] O. Saukh, P. J. Marrón, A. Lachenmann, M. Gauger, D. Minder, and K. Rothermel. Generic Routing Metric and Policies for WSNs. In *Proc. of the 3rd European Workshop on Wireless Sensor Networks*, Feb. 2006.

[14] K. Sayood. *Introduction to Data Compression, Second Edition*. Morgan Kaufmann Publishers, 2000.

[15] Sustainable bridges web site. http://www.sustainablebridges.net.