The Virtual Sensor Concept

Separating Sensor Software from the Hardware

Jarrold Trevathan, Wayne Read, Abdul Sattar, Simon Schmidtke and Tony Sharp

Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, Australia

Substation33, Logan Australia

j.trevathan@griffith.edu.au

Abstract—Aquatic environmental sensors are expensive which limits the ability to undertake widescale remote monitoring. In most instances, too much logic and functionality are contained within the sensor device itself. This approach is expensive and computationally restrictive. This paper presents a new architectural paradigm for remotely deployed sensors whereby the calibration logic (and other functionality) is separated from the physical sensor hardware. A sensor device is only responsible for taking raw unprocessed sensor readings, which are transmitted back to a central server. All processing occurs on the server where computational capability and sophistication are unbounded. This approach allows for significant flexibility in terms of dynamic calibration adjustments to be applied to sensor data (e.g., in the case of sensor fouling or device decay) and for statistical quality assurance/data production algorithms to be applied. We present an example of how this paradigm can be adopted in this context to a turbidity sensor.

Keywords—turbidity sensor; virtual sensor concept; dynamic calibration; Internet of Things; statistical regression;

I. INTRODUCTION

Existing remote water quality monitoring platforms are prohibitively expensive for most applications [1-5]. Much of this has to do with the expense of the individual sensors. While these sensors are highly capable and sophisticated devices, the actual design approach taken in conjunction with the larger sensor network delivery system, results in severely restrictive proprietary systems. This design paradigm has existed for many decades now despite continual improvements in sensor and *Internet of Things* (IoT) [6] technologies.

Drawing parallels between remote aquatic environmental monitoring systems and the *personal computer* (PC) market [7, 8], we can see that PCs have continued to become more powerful, less expensive and more user friendly. One may ask the question, “why hasn’t the remote aquatic environmental monitoring market followed the same trend?”. This divergence has occurred as PCs adopted dramatic shifts in architectural paradigms for virtual memory, *what you see is what you get* (WYSIWYG), plug and play and user-friendliness. Unfortunately, for the main part environmental monitoring platforms have not followed suit and the architectural paradigms remain stagnant [9, 10].

In an aquatic monitoring platform, the financial costs of the sensors are threefold: 1) the *cost of the sensor* (the initial outlay); 2) the *cost if the sensor is lost/damaged*; and 3) the *cost of the data if it is lost/corrupted*. As the sensor software is integrated with the hardware, the calibration implementations are primitive and inflexible (typically due to

computational constraints). There is no control over how the data is converted from raw values into meaningful data. Instead the end-user gets a series of processed data. In addition, functions such as the calibration coefficients and the sampling rate (for example) cannot be remotely altered.

This paper presents a new architectural paradigm for remotely deployed sensors whereby the software (i.e., the calibration logic and other functionality) is separated from the physical sensor hardware. A sensor device is only responsible for taking raw unprocessed sensor readings (i.e., voltages), which are transmitted back to a central server. This reduces the financial costs of the sensor device (in the aforementioned context). The server undertakes processing in an environment where computational capability and sophistication are unbounded. Such an architecture allows for significant flexibility in terms of dynamic calibration adjustments to be applied to sensor data (e.g., in the case of sensor fouling or device decay) and for statistical quality assurance/data production algorithms to be used. We demonstrate how this paradigm can be applied to a practical turbidity sensor design [11].

This paper is organised as follows: Section 2 outlines the problem motivation. Section 3 presents a new paradigm for separating sensor software from hardware. Section 4 illustrates how the paradigm can be used on a turbidity sensor; and Section 5 provides some concluding remarks and avenues for future work.

II. PROBLEM MOTIVATION

A. Remote Aquatic Environmental Monitoring

Fig. 1. The Aquatic Environmental Monitoring Network

Figure 1 illustrates the example aquatic environmental monitoring network that is the motivating case study for this paper (see [4, 5]). A series of remote monitoring buoys are

Fig. 1. The Virtual Sensor Concept

deployed in a water body. Each buoy contains a series of sensors (i.e., light, temperature, turbidity, salinity, pH and dissolved oxygen). Every 15 minutes a buoy takes samples and sends this data back to the end user via the Internet (3G GSM (*Global System for Mobile Communications*) network).

We have solved all of the general sensor network challenges around data management, transmission, congestion, etc. These issues are not the focus of this paper. Instead we are interested here on how to reduce the costs of the sensors (purchase, replacement, data loss) while dealing with sensor calibration, fouling, interference issues and data quality assurance/production. Unfortunately, these issues cannot be easily addressed with the existing architectural paradigms of aquatic environmental monitoring platforms.

B. Aquatic Environmental Sensor Calibration

There are two aspects to calibration: 1) Calibration of a sensor in laboratory conditions against another device with a known accuracy; and 2) Validation of the sensor's readings out in the field [12, 13]. Once sensors are deployed in the field, they will start fouling. This is largely due to factors such as algal growth, barnacles, suspended sediment, and animal excrement. This fouling impacts the sensor readings over time [14-16]. As such, there is a need to periodically maintain the buoys through wiping the sensors approximately once a week (depending on the environment).

Figure 2A illustrates the traditional configuration for sensor devices and the calibration approach. A sensor device takes an environmental reading via its sensor. The calibration equation is then applied to the sensor reading based on the calibration data. The output is the calibrated sensor reading. The sensor reading, calibration data and calibrated sensor reading are stored on the sensor device. The sensor device is also responsible for the computations involved in applying the calibration data. However, this approach is expensive and the accuracy and flexibility of the calibration implementation is limited when computational power is an issue.

In a remote aquatic environmental monitoring application, it would be beneficial if dynamic calibration adjustments could be made to prolong sensor life in light of fouling and device decay. However, this would only be possible by rethinking the architectural paradigm.

III. THE VIRTUAL SENSOR CONCEPT

In a traditional sensor cluster, there is no strict demarcation between the hardware and the software. In most cases, individual sensor heads are fully calibrated, autonomous components of the cluster attached to a communications chip;

individual sensor heads may have communication software hardwired into them as well. Calibration, data protocols including sensor reading estimation are all determined prior to the construction of the sensor cluster and hardwired in – in fact, prior to the manufacture of the sensor head.

We propose the *Virtual Sensor Concept* (VSC), where we separate the hardware from the software, in much the same way that they are separated in a computer (see Figure 2B). For example, the physical memory of a computer might be 16 Gigabytes, whereas the virtual memory could be up to a Terabyte of addressable memory. Logically, the virtual memory can be segmented, threaded processes can be run, data protected from execution and so on. The physical memory is just a linear array of addressable words that store binary number – the operating system software takes care of the logical structure of the memory.

We propose the same basic concept here applied to sensor clusters, indeed individual sensor heads. The sensor heads are stripped of all calibration and communication software and return the basic data, a stream of bytes to the communication chip, typically a basic processor (such as an Arduino) which then transmits this raw data back to a web server. No processing or any other manipulation of the data occurs on the sensor cluster; any transformation to useable data occurs on the web server.

As an example, we will consider a turbidity sensor. A typical sensor head returns a set of (voltage) readings at each point in time. On a traditional sensor head, this set of readings is converted into a turbidity estimate and then this estimate is then sent to the communication chip for transmission or storage on the memory of the sensor cluster. Note that the method used to estimate the turbidity from the raw data is fixed and cannot be changed; any additional information from the raw data is lost at this point.

A. Dynamic remote calibration adjustment

The VSC separates the calibration calculation from the sensor. A sensor device is only responsible for taking raw unprocessed sensor readings (i.e., voltages). The sensor reading is then transmitted via the Internet to a central web server – no actual calibration coefficients are applied to the data prior to transmission. The mathematics/processing is done on the server, which has more computational capability. This approach allows for arbitrary function approximation to be used in the calibration approach. Given enough calibration data, the calibration curve for the sensor can be fitted continuously over the range where an expensive calibration device (or calibration standard) is accurate [17].

If the fouling rate of a water body at a given location is known, a dynamic calibration adjustment can be applied to a sensor's data. That is, by using a mathematical technique that approximates the fouling growth rate, the interference of the fouling can be taken into account and the sensor's data adjusted to remove the error. This method is only possible due to the system using the VSC by separating the calibration from the physical sensor. By applying dynamic calibration adjustments, sensors can stay out in the field longer and can increase the amount of time between maintenance cycles – thereby reducing labour costs.

B. Statistical data quality assurance/production techniques

Data quality can be assessed by its state of completeness, validity, consistency, timeliness and accuracy. More accurate and useful information can be extracted from good quality data sets. However, erroneous and missing data can arise from hardware failures, power outage and communication breakdown. Sensors do not automatically ensure their data quality or flag missing/suspicious data. Instead, this needs to be examined and handled by end users.

Many approaches have been proposed to handle the missing data such as: Simple averaging; Drawing a straight line between start and end point of the data gap or using 5 data points outside the range of the gap to predict the curvature of the time series and using a best-fit curved line to fill the gap; Local distribution mechanisms that uses the data from neighbour nodes to identify anomalous measurements and evaluate missing data; and using high performing sensors as a pinpoint for adjusting other sensors' readings. However, there are several limitations in the aforementioned mechanisms. The simple averaging method leads to erroneous values as environmental phenomena vary widely over space and time. The use of data from neighbour nodes or high performing sensors is only effective when there are harmonious relationships among the sensors. Therefore, this approach is not applicable when sensors are deployed far from each other. Deploying sensors with short distances can result in a low-density coverage area.

Using the VSC, the raw data from each of the sensors is available on the server. The raw data can be used to statistically examine the correlation between the different data sets. When data is missing from one sensor, it can be estimated using the other sensor data when the correlation between them is significant. Although this could be done with the calibrated data, using the raw data removes any errors (however small or large) introduced during the calibration process, strengthening any underlying correlations and reducing any errors in the statistical modelling process [18]. The VSC can use statistical models to quality assure the entire data set by examining the variation across the sensors (see Figure 2C).

IV. EXAMPLE OF THE VIRTUAL SENSOR CONCEPT

Let us present an example of how the VSC can be used to dynamically recalibrate a turbidity sensor and compare. The turbidity sensor head that we have developed (see [11]) takes thirty voltage readings that are used to estimate the turbidity at a particular time. In a contemporary sensor, the voltage estimate would typically be a point estimate or a simple statistic like the mean. This value would then be used by the (fixed) calibration equations to provide an estimate of the turbidity. Once the turbidity is estimated any other information contained in the thirty data points is lost.

Using the VSC, these thirty readings are relayed back to the web server and an almost arbitrary set of statistics including (but not limited to) the mean, median, mode, weighted mean can be used to estimate the voltage. The method used to estimate the voltage can be changed arbitrarily (and retrospectively, assuming the original raw data has been stored). The calibration equations can now be used to determine the turbidity estimate.

During the initial calibration process, the calibration equation coefficients can be determined for a clean sensor head as well as for a set of simulated fouled sensor heads. This set of coefficients can then be stored and used by the web server as the sensor head becomes fouled.

The 30 point distribution can be used to aid in determining the appropriate set of calibration equation coefficients. Statistically, the variation and distribution of the voltages at each point in time can be utilised to select the calibration coefficients dynamically. The method used to determine a point estimate can also be changed (e.g., from the mode to the mean). Outliers can also be removed and the point at which the sensor cluster needs maintenance can be determined.

This example only uses the data from a turbidity sensor. When there is a cluster of sensors, multivariate regression analysis and cross correlation analysis can be used to help select and validate the calibration equation coefficients to be used for the current level of fouling. In addition, the variance and distribution at discrete points in time can be compared with the overall variances to predict the rate of aggregation of fouling and predict the accuracy of the turbidity estimates. This can also be used to determine when the sensor cluster needs cleaning and maintenance.

V. CONCLUSIONS

This paper presented the Virtual Sensor Concept – a new architectural paradigm for remotely deployed sensors whereby a sensor's software is separated from the hardware. The VSC has been applied to a turbidity sensor developed by the authors [11] and this has been used to exemplify the approach. The calibration logic and other functionality is removed from the physical sensor device to reduce costs (i.e., purchase cost, replacement cost and data loss cost). By stripping out all but the primitive components the sensor head is a fraction of the cost of a contemporary sensor – as little as one tenth of the cost. When the sensor head is lost or accidentally destroyed, replacing it is not a big financial issue.

A sensor device using the VSC architecture is only responsible for taking raw unprocessed sensor readings (i.e., voltages) – a back-end server undertakes all processing, calculation and quality assurance. This approach allows for almost arbitrary flexibility in converting raw data to processed information. In addition, the sensor can be re-calibrated dynamically and remotely, an impossibility for a contemporary sensor. Finally, the VSC architecture allows for statistical quality assurance/data production algorithms to be applied easily and accurately. Future work involves applying the VSC across a complete suite of sensors and applying dynamic calibration adjustments to sensors while deployed.

ACKNOWLEDGMENT

This paper is supported in part by the Australian Research Council Linkage Project LP190101083.

REFERENCES

- [1] L.A. Seders, C.A. Shea, M.D. Lemmon, P.A. Maurice and J.W. Talley, "LakeNet: An integrated sensor network for environmental sensing in Lakes", *Environmental Engineering Science*, vol 24, pp. 183–191, 2007.
- [2] C. Albaladejo, F. Soto, R. Torres, P. Sánchez, J.A. López, "A low-cost sensor buoy system for monitoring shallow marine environments", *Sensors*, vol. 12, pp. 9613–9634, 2012.
- [3] G. Lockridge, "Development of a low-cost Arduino-based sonde for coastal applications", *Sensors*, vol. 16, pp. 528, 2016.
- [4] J. Trevathan, R. Johnstone, T. Chiffings, I. Atkinson, N. Bergmann, W. Read, and T. Stevens, "SEMAT—The next generation of inexpensive marine environmental monitoring and measurement systems", *Sensors*, 12, pp. 9711–9748, 2012.
- [5] J. Trevathan and R. Johnstone, "Smart Environmental Monitoring and Assessment Technologies (SEMAT)—A New Paradigm for Low-Cost, Remote Aquatic Environmental Monitoring", *Sensors*, vol. 18, no. 7, pp. 2248, 2018.
- [6] S. Madakam and V. Lake, "Internet of Things (IoT): A literature review", *Journal of Computer and Communications*, 3(05), 164, 2015.
- [7] F. Malerba, R. Nelson, L. Orsenigo, and S. Winter, "History-friendly models of industry evolution: the computer industry", *Industrial and corporate change*, 8(1), pp. 3–40, 1999.
- [8] E. Esposito and M. Mastroianni, "Technological evolution of personal computers and market implications", *Technological Forecasting and Social Change*, 59(3), pp. 235–254, 1998.
- [9] G. Gigan and I. Atkinson, "Sensor Abstraction Layer: a unique software interface to effectively manage sensor networks", In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 479–484, 2007.
- [10] J. Trevathan and T. Myers, "A unique solution for designing low-Cost, heterogeneous sensor networks using a middleware integration platform", *International Journal of Computer, Information, Systems and Control Engineering*, 8(4), pp. 53–62, 2014.
- [11] J. Trevathan, W. Read and S. Schmidtke, "Towards the Development of an Affordable and Practical Light Attenuation Turbidity Sensor for Remote Near Real-Time Aquatic Monitoring", *Sensors*, 20(7), 1993, 2020.
- [12] V. Bychkovskiy, S. Megerian, D. Estrin and M. Potkonjak, "A collaborative approach to in-place sensor calibration", In *Information processing in sensor networks*, pp. 301–316, 2003.
- [13] K. Whitehouse and D. Culler, "Calibration as parameter estimation in sensor networks", In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pp. 59–67, 2002.
- [14] P. Ridd and P. Larcombe, "Biofouling control for optical backscatter suspended sediment sensors", *Marine Geology*, 116(3–4), pp. 255–258, 1994.
- [15] D.V. Manov, G.C. Chang and T.D. Dickey, "Methods for reducing biofouling of moored optical sensors", *Journal of Atmospheric and Oceanic Technology*, 21(6), pp. 958–968, 2004.
- [16] L. Delauney, C. Compere and M. Lehaitre, "Biofouling protection for marine environmental sensors", *Ocean Science*, 6(2), pp. 503–511, 2010.
- [17] L. Brieman and J.H. Friedman, "Predicting multivariate responses in multiple linear regression", *Journal of the Royal Statistical Society*, V 59 (1), pp 3–54, 1997.
- [18] A.J. Izenman. "Multivariate Regression. Modern Multivariate Statistical Techniques", Springer Texts in Statistics. 2013.