

A Novel Method for Data Calibration in Wireless Sensor Networks

Wei-min Qi

Institute for Interdisciplinary Research
Jianghan University
Wuhan, China
qwmin@126.com

*Jie Xiao

Wenhua College
Huazhong University of Science and Technology
Wuhan, China
Corresponding author
luckycrystalxj@163.com

Abstract—In order to provide efficient and suitable services for users in a ubiquitous computing environment, many kinds of context information technologies have been researched. Wireless sensor networks are among the most popular technologies providing such information. Therefore, it is very important to guarantee the reliability of sensor data gathered from wireless sensor networks. However there are several factors associated with faulty sensor readings which make sensor readings unreliable. The research put forward classifying faulty sensor readings into sensor faults and measurement errors, then propose a novel in-network data calibration algorithm which includes adaptive fault checking, measurement error elimination and data refinement. The proposed algorithm eliminates faulty readings as well as refines normal sensor readings and increase reliability. The simulation study shows that the in-network data calibration algorithm is highly reliable and its network overhead is very low compared to previous works.

Keywords-wireless sensor network; reliability; data calibration; refinement

I. INTRODUCTION

The advances on Micro Electro Mechanical Systems (MEMS) and wireless communications have made possible the foundation and subsequent development of the Wireless Sensor Networks (WSNs). WSNs consist of a group of sensors that communicate with each other through a radio channel. They work in a decentralized and self-organizing manner and do not rely on fixed network infrastructure. Each sensor can act as a relay to deliver information toward its destination, i.e. multi-hop manner [1]. Small sensor nodes sense various events such as variations in temperature, illumination, sound etc. They report sensed events to a base node (sink node) after converting analog signals to digital data. Application systems automatically utilize reported data to achieve their particular goals, such as environment/industrial/health monitoring and vehicle/enemy tracking [2]. Recent advances in System-On-Chip (SoC) and Micro-Electro Mechanical Systems (MEMS) enable the development of practical sensor nodes which are very small and powerful [3]. As a result, prototypes or practical sensor networks have emerged and numerous applications or services will use sensor networks in the near future.

To make sensor network technologies practical in the real-world, there remain several research issues which need to be resolved, such as efficient energy consumption for

prolonging the network lifetime, efficient key distribution algorithms for security, and minimizing sensor reading errors for reliability. In particular, minimization of sensor reading errors is inevitable, and it is a critical research issue.

The research focuses on a solution for correcting the sensor reading errors. There are two kinds of faulty sensor reading: sensor faults and measurement errors. Sensor fault means that a sensor module in a sensor node is malfunctioning. The node continuously reports inaccurate data. Measurement error means that a sensor node reports inaccurate data values because of its misalignment or its distance from an event.

The paper proposes an in-network data verification algorithm which eliminates faulty sensor readings (including sensor faults and measurement errors) as well as refines sensor readings. In-network means data verification is performed in distributed nodes, not a centralized sensor, and also means that verification is performed during data delivery. This guarantees scalability by adapting the method of local distribution, and reduces energy consumption and communication delays incurred by every node in schemes proposed in previous work.

II. RESEARCH STATUS

As the wireless networks becoming such a critical issue, there was substantial interest in fault-tolerant sensor fusion. Most research focused on distributed decision fusion, based on decisions from multiple sensors under conditions of sensor faults [4,5]. The traditional theory is still applied subsequent to wireless sensor networks becoming such a critical issue [6,7]. The application of the traditional approach to wireless sensor networks has many problems such as not considering network properties or communication overhead, and measurement errors of fault sensor readings does not considered.

The existing methods to deal with faulty sensor readings have several problems, which are as follows:

- (1) First, they didn't consider network properties such as energy consumption and communication delay.
- (2) Second, many researches only focused on sensor faults, the measurement errors of faulty sensor readings do not be considered.
- (3) Third, because their goal is decision correction, they only determine whether an event occurs or not.

The paper considers both sensor faults and measurement errors as faulty sensor readings, and we suggest a data value

correction scheme which is based on network properties to the greatest possible extent.

III. FAULT SENSOR READINGS

Faulty sensor readings in wireless sensor networks are discussed in this section. Prior to a detailed description of faulty sensor readings, we first show characteristics of normal sensor readings of a event without reading faults, according to the distance from the event source. The peak in the figure indicates the source of an event. This point is called CoS (Center of Stimulus) [8]. As the strength of an event is related to the distance, the data values of sensor readings are decreased, according to the distance from the peak.

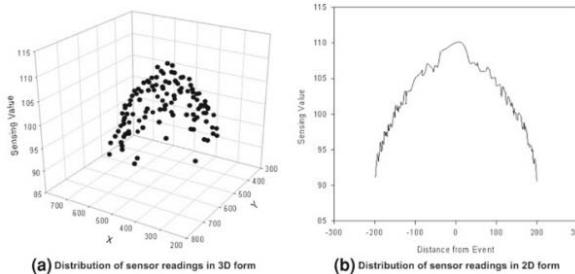


Figure 1. Sensor readings without faults

For example, the temperature increases as the sensor moves closer to the heater. And, the intensity of illumination rises exponentially, according to the distance from the light. Also, the vibration at the source is stronger than at any other location in the event region. Fig. 1b is a two-dimensional graph which indicates the signal strength according to the distance from the source. Fig. 2 consists of graphs of sensor readings including faults. Here, GT means the ground truth, GT_{\max} is maximum value of the ground truth and GT_{\min} is the minimum value of the ground truth. We consider the median of GT (GT_{median}) as the real value of the event. The graphs show both sensor faults and measurement errors. Then we classify the probability of a faulty sensor reading as follows.

$$P = \alpha + \beta + \gamma \quad (1)$$

In Eq.1, α means the probability of a sensor fault, β and γ mean the probability of a measurement error. We can distinguish probability α when the sensor readings remain identical, even though the sensor readings of neighbor nodes may change. An event cannot affect the value of a sensor fault. The probability of β is for a measurement error higher than GT_{\max} or lower than GT_{\min} . As β is related to the change of an event, it refers to measurement errors.

The probability of γ indicates a measurement error between GT_{\max} and GT_{\min} .

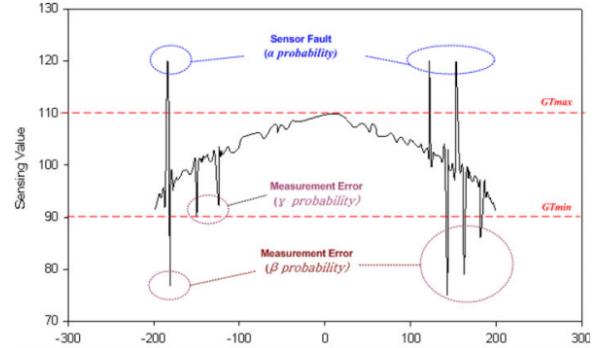


Figure 2. Faulty sensor readings and classification

IV. FAULT SENSOR READINGS ELIMINATION

Through means of reducing α and β , sensor faults and measurement errors of faulty sensor readings will under control. The paper applies three steps to achieve this.

A. Adaptive Sensor Fault Checking

The faulty sensor readings can be classified into sensor faults and measurement errors[9]. To eliminate sensor faults, we consider the variation of sensor readings with sensor faults. If a sensor module such as a thermometer is malfunctioning, the data values derived from the sensor module are nearly always identical. The research developed an adaptive sensor fault check algorithm to reduce sensor fault probability α . The principle is that a node regards its readings as due to a sensor fault if a node reports nearly identical values over a long period. In self-checking, each sensor node checks variable k , which is a consecutive count of constant data values[10,11]. Based on the value of k each node calculates the probability of a sensor fault(P_{fault}). We set the probability of a sensor fault to increase exponentially according to increments of value k .

$$P_{\text{fault}} = P_{\text{fault_prev}} + c \times k^2 \quad (2)$$

Eq.2 shows the calculation of the sensor fault probability based on k . Here c is the rate of change. While sensor data values are constant, P_{fault} increases according to increments of k exponentially.

Fig.3 is the graph of P_{fault} where $c=0.0005$. There are three phases in the graph, according to the sensor count: 0-150, 150-250, 250-450. The first and second phases correspond to no-event or continuous-event, then the probabilities fall to 0 before approaching 100% because the sensor data values change. In the third phase, the probability approaches 100%, so the node sends a neighbor confirmation message. If there is an acknowledgement message, the node concludes that its reading is due to a

sensor fault. By this means we eliminate the sensor fault and reduce α probability.

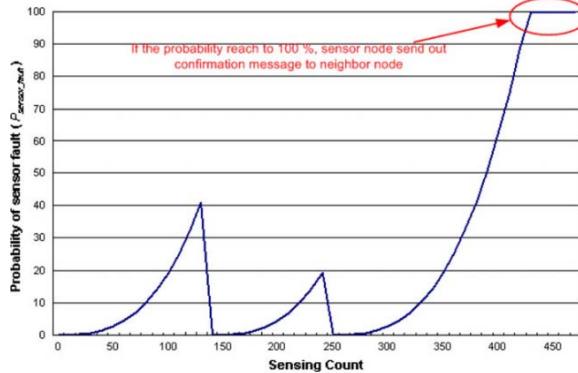


Figure 3. Variation of sensor fault probability

B. Data Gathering in an Event Group

In order to eliminating the measurement errors, we proposed data gathering method which gathers all data at a node called an event header node. We use the average and standard deviation to eliminate measurement errors and reduce the probability of β . The research should meet follow aspects: (1) an event group composition algorithm which distinguishes whether or not a node sensed an event, (2) an event header node selection algorithm, (3) a path setup mechanism which delivers the data to the event header node.

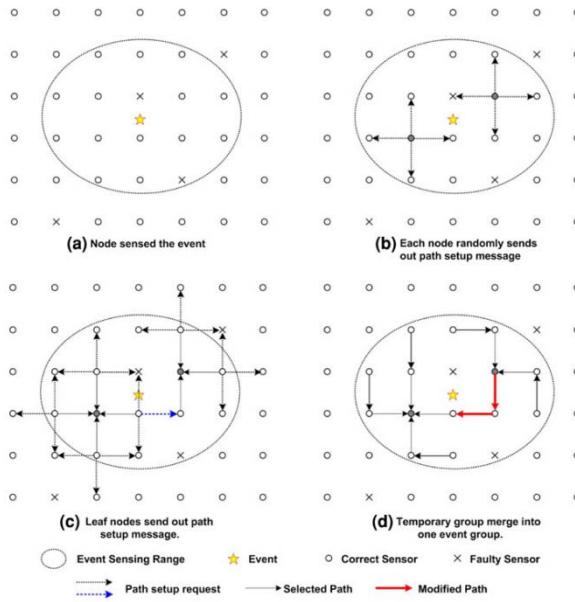


Figure 4. Data gathering

Fig. 4 shows the functions of these methods. First, an event occurs and nodes sense the event (Fig. 4a). Then sensor nodes classify into two groups, an event sensor group and a non-sensor group. Each node can identify its group

using P_{fault} , if neighbor nodes have identical values of P_{fault} , this implies that they sensed an identical event. Therefore, we composed an event group by identifying P_{fault} values for each.

Subsequent to an event, a node which identifies itself as an event-sensor node randomly sends a path setup message (Fig. 4b). The neighbor nodes which received the path setup message set the path to the requestor only if it is an event-sensor node, rather than a sensor-fault node. Then it resends the path setup message to its neighbor nodes. This process is repeated until all event-sensor nodes have set a path (Fig. 4c). There could be several partitions in the event group because a node randomly initiates a path setup. While repeating path setup requests, some requests reach other partitions. In this case the receiver modifies its path to the requestor. This implies that one partition becomes a sub-tree of the requestor partition (Fig. 4d). Finally, one path setup initiator becomes an event header node during path setup. Subsequent to path setup, the data is transmitted to the event header node along with the selected path.

C. Measurement Error Elimination

After gathering the data values, the event header node performs the process of eliminating measurement errors that are definitely due to faulty sensor readings. The basic principle is that when data values differ greatly to others, these are considered as measurement errors. We use the average and standard deviation of the gathered data values, including measurement errors.

It is sufficient to filter measurement errors that are definitely due to faulty sensor readings. There are extremely small measurement errors within the standard deviation.

V. SIMULATION RESULT

Table I shows simulation parameters. Simulation performed 500s in each scenario. Fig. 5 and Fig. 6 show the accuracy of the proposed algorithms according to the probability of faulty sensor readings. We estimate the accuracy of sensor readings in terms of two perspectives; decision correction and value correction.

TABLE I. SIMULATION PARAMETERS

Setting Name	Parameters
Number of nodes	1,000
Transmission range	40m
Simulation area size	1,000m \times 1,000m
Initial energy	10,000J
Sending energy	10mJ
Receiving energy	1.5mJ
Simulation time	500s

While using adaptive fault check algorithm and measurement error elimination, the accuracy is increased as

shown in Fig. 5. If we do not use an algorithm, where the probability of a faulty sensor reading is 60%, the accuracy of decision correction is only 44%. This implies that only 44% of sensor nodes report correct data values and this results in inaccurate sensor values. As shown in Fig. 6 the accuracy of value correction is 0 %. If we use an adaptive fault check algorithm, we eliminate sensor faults. The figures imply that more than 50% of sensor nodes report correct values. Therefore, the accuracy of value correction is increased to 40%. This accuracy is still low because of measurement errors.

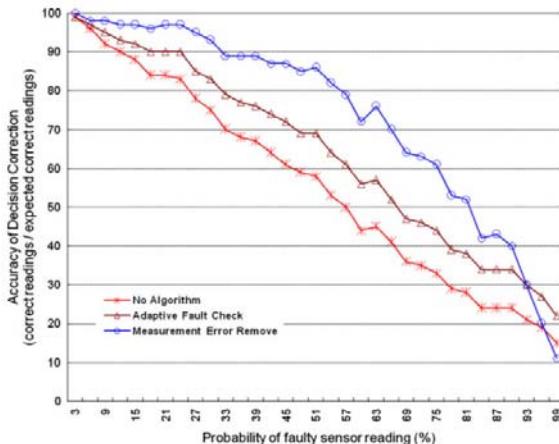


Figure 5. Accuracy of sensor readings at the start of decision correction

As shown in Fig. 6, if we do not use an algorithm, the accuracy of value correction is decreased to 15% in the case of a 50% probability of faulty sensor readings. However, if we apply an adaptive fault check algorithm, the accuracy is higher than 50% in the case of 50% probability of faulty sensor readings.

Moreover, if we use measurement error elimination, the accuracy is higher than 80%, even though the probability of a faulty sensor reading is 70%. This shows that the proposed algorithm can eliminate sensor faults and measurement errors effectively.

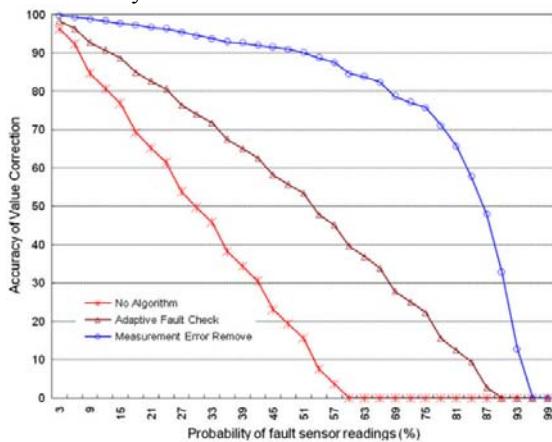


Figure 6. Accuracy of sensor readings at the start of value correction

VI. CONCLUSION

The paper proposed an in-network data verification scheme that enhances the reliability of sensor readings in cases of faulty sensor readings. First we classify faulty sensor readings into sensor faults and measurement errors. Then an adaptive fault check algorithm for sensor faults was applied to measure error elimination and data refinement. We effectively eliminate sensor faults and measurement errors. Simulation results show that sensor faults and measurement errors were eliminated effectively, and network overhead was reduced by a maximum of 70% compared to previous schemes, in the case of 500 event group nodes and 10% neighbor nodes. The scheme increases the accuracy of decision correction and value correction by a maximum of 30% and 80% respectively. The scheme improves the reliability of sensor readings and enables more practical sensor networks.

ACKNOWLEDGMENT

This work is supported by Science and Technology Research Key Project of Hubei province Education Department (Project Numbers: D20123405).

REFERENCES

- [1] Pottie, G. J., & Kaiser, W. J. "Wireless sensor networks". Communications of the ACM, 43(5), 2000, pp. 51–58.
- [2] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. "Wireless sensor networks". A Survey Computer Networks, 38 (2), 2002, pp. 393–422.
- [3] Park, C ., & Chou, P. H. Eco: "Ultra-wearable and expandable wireless sensor platform". In Proceedings of the third international workshop on body sensor networks, April 2006.
- [4] Thomopoulos, S. C . A., Viswanathan, R ., & Bougoulias, D. K. "Optimal distributed decision fusion". IEEE Transaction on Aerospace and Electronic Systems, 25 (5), 1989, pp. 761–764.
- [5] Tsitsiklis, J., & Athans, M. "On the complexity of distributed decision problems". IEEE Transactions on Automatic Control, 30(5), 1985, pp. 440–446.
- [6] Krishnamachari, B., & Iyengar, S. "Efficient and fault-tolerant feature extraction in wireless sensor networks". IPSN 2003, Lecture Note in Computer Science, 26(34), 2003, pp. 488–501.
- [7] Krishnamachari, B., & Iyengar, S. "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks". IEEE Transactions on Computers, 53(3), 2004, pp. 241–250.
- [8] Ye, F., Zhong, G., Lu, S., & Zhang, L. "GRAdient broadcast: A robust data delivery protocol for large scale sensor networks". ACM Wireless Networks, 11(2), 2005, pp. 285–298.
- [9] Luo, X., Dong, M., & Huang, Y. "On distributed fault-tolerant detection in wireless sensor networks". IEEE Transactions on Computers, 55 (1), 2006, pp. 58–70.
- [10] Bokareva, T., Bulusu, N., & Jha, S. "SASHA: Towards a self-healing hybrid sensor network architecture". In Proceedings of the 2nd IEEE international workshop on embedded networked sensors (EmNets-II), Sydney, Australia, May 2005.
- [11] Ding, M., Chen, D., Xing, K., & Cheng, X. "Localized fault-tolerant event boundary detection in sensor networks". In Proceedings of the 24th annual joint conference of the IEEE computer and communications societies (INFOCOM 2005), Vol. 2, Mar 2005. pp. 902–913.