Improving Data Reliability via Exploiting Redundancy in Sensor Networks

Tony Sun, Ling-Jyh Chen, Chih-Chieh Han, Mario Gerla
UCLA Computer Science Department, Los Angeles, CA 90095, USA
{tonysun, cclljj, simonhan, gerla}@cs.ucla.edu

Abstract

Wireless sensor network applications interact with the physical world through analog sensors. However, decisions derived from flawed sensor measurements can adversely impact the correctness of the overall sensor network findings. To improve the reliability of decisions and minimize the impact of faulty sensor measurements, it is important to have a distributed scheme that enhances the trustworthiness of result by exploiting the presence of redundant data. In this paper, we present Confidence Weighted Voting (CWV), a distributed technique that can improve the underlying data reliability and fault tolerance of various sensor network applications. We examined CWV against Majority Voting (MV) and Distance Weighted Voting (DWV) techniques, and contrasted the level of data reliability of each approach in the prevalent presence of flawed sensors. The results show that CWV can consistently and substantially outperform the other two distributed schemes by providing as much as 49% more resiliency against faulty sensor measurements.

1. Introduction

With the rapid advances in wireless communication and embedded micro-sensing MEMS technologies, Wireless Sensor Network (WSN) has inspired many applications ideas that will fundamentally improve our understanding of our surrounding physical environment [2][5]. A set of applications such as hazardous environment exploration, environmental monitoring, military tracking and reconnaissance are just some key motivations for many recent research efforts in this area.

Since sensors are often deployed under inhospitable conditions, false readings from damaged sensors can adversely affect the correctness of sensor network findings. Therefore, to improve the reliability of network decisions, it is important for applications to have access to trustworthy data. Additionally, sensors are usually very limited in terms of computational power, memory, and energy. An energy conserving solution should achieve decisions via a distributed approach, thus avoiding the excessive amount of message overhead associated with the centralized technique. Other energy conserving schemes have also been studied by [16][18][19].

Realizing that multiple sensors monitoring the same location at the same time can ensure higher monitoring quality [6][19], and the fact that data from neighboring nodes can be used to distinguish the correctness of local data. It’s clear that redundant information can be utilized to improve the underlying reliability of local data. These highly localized results can be aggregated by methods such as [3][17] to provide higher data reliability to requesting applications such as event/target detection [1][3][4][11].

Minimizing the impact of faulty sensor measurements is related to the Byzantine problem [10]. Previous research used classification techniques such as neutral networks or Bayesian classifier [11] to accomplish better results. Other solutions such as [1][3] rely on higher level data collaboration schemes that aimed to accomplish better reliability without using redundant information from the network. However, these solutions often require excessive amount of states, memory, message overhead, or computational cost, and are consider unfitting for sensor network purposes.

In this paper, we present Confidence Weighted Voting (CWV), a simple distributed technique that improves the reliability of underlying data by exploiting redundant information. Since CWV uses neighboring data to discern the correctness of local data, it is capable of improving the baseline reliability of many applications such as [3][11][14]. We examined CWV against the classical Majority Voting (MV) [9] and Distance Weighted Voting (DWV) [9] techniques, and contrasted the level of data reliability of each approach in the prevalent presence of flawed sensors. We simulated the basic behaviors of CWV on top of k-cover deployment strategy (which guarantee redundancy when k>2), we also used an analytical model to prove the effectiveness of CWV over the other two schemes. Finally, we showed that CWV can outperform the other distributed voting schemes by providing as much as 49% more resiliency to sensor errors.

The rest of this paper is organized as follows. In section 2, we present the system model, the metrics used to evaluate our algorithms, and the k-coverage placement strategies used in our experiments. Section 3 elaborates on the details of Confidence Weighted Voting algorithm, and briefly describes the baselines algorithms to which we compare our work. Section 4 present simulation results and related analysis. We conclude the paper in Section 5.

2. System Model

In this section, we introduce the model of sensor network used in section 2.1. We then discuss the metrics used to evaluate the system performance in section 2.2. Lastly,
we described the $k$-coverage placement strategy used in our experiments in section 2.3.

2.1 Sensor Network Model

First, we assume that the sensor node knows its own location [7] and nodes are stationary. The nodes can also obtain their own location through location process described in [15]. For simplicity, we refer to the sensing area of a node as a circle with a nominal radius $r$ centered at the location of the node itself. With a set of sensors deployed in a region instructed to provide reliable discrete data, we are also assuming that an event can be detected by multiple sensors nodes due to our $k$-coverage sensor placement scheme described in section 2.3. The sensor performed event detection based on pre-established thresholds. We deploy the sensor nodes in a two-dimensional Euclidean plane. However, the technique can be extended to a three-dimensional space without much difficulty. Lastly, we assume that the nodes can directly communicate with the neighboring nodes within a radius larger than $2r$ ($r$ is nominal sensing radius). All the above are common assumptions for many sensor network applications.

2.2 Performance Metrics

The performance of the algorithms can be measured in terms of data reliability against varying level of faulty sensors. Each node is given a failure probability, which defines how likely the sensor will report an incorrect value. The behavior of these faulty sensors is assumed to be arbitrary. Error occurrences are assumed to be uniformly distributed. Reliability of the network is then measured by how likely we can achieve the correct representation of the environment given our deployment strategy, algorithm, and sensor failure rate. We used an event/target scenario to test our algorithm [3]. Since sensors need to combine their sensed values to reach a representative decision for the region in question, and results computed based on these incorrect measurements can radically impact the correctness of sensor network findings. The network will likely contain some faulty sensors, while we need to arrive at a correct decision regardless of the distortion from the flawed sensors.

2.3 $k$-coverage Placement Strategy

Several coverage models [8][12][13] have been proposed for different application scenarios. In this paper, we assume that a point $p$ is monitored if their Euclidian distance to a sensor is less than the sensing range of $r$. The coverage configuration problem bares close resemblance to the Art Gallery Problem, which deals with determining the number of observers necessary to cover an art gallery room such that every point in the art gallery is monitored by at least one observer. This problem is optimally solved in a 2D plane, but in shown to be NP-hard when extended to a 3D space. Based on the coverage model, an area is having a coverage degree of $k$ (i.e., being $k$-covered) if every location inside $A$ is covered by at least $k$ nodes. Practically speaking, a network with higher degree of coverage can achieve higher sensing accuracy and be more robust against sensing failures.

In this paper, we used a close approximation of the $k$-coverage scheme. The details of our implementation are summarized in Table 1. Since random deployment and $k \times 3$ scenarios can be roughly approximately by a combination of basic $k$-coverage cases, we only used three basic $k$-coverage cases to reveal the fundamental properties of our algorithms.

3. Distributed Voting Algorithms

In this section, we present the algorithm for MV in section 3.1, DWV in section 3.2, and CWV in section 3.3. All three distributed algorithms shares the same characteristics in their simplicity, speed, scalability, and low message overhead.

3.1. Majority Voting Algorithm

To realize a distributed Majority Voting (MV) scheme, sensor readings are first gather from neighboring sensor nodes, and local decisions are achieved based on the majority opinion of the collected data. For instance, the decision for area $A$ in Figure 1 is reached through majority voting on result gathered from sensor 1, 2, and 3 (Since $A$ is covered by 3 sensors). Similarly, the decision reached in area $B$ came from majority voting on result reported by sensor 1 and 3. Whenever a tied for majority occurs, the final decision is randomly chosen.

Suppose the number of deployed sensors in the investigating area is $m$ and the possible report value of each sensor is an integer from $1$ to $n$, the Majority Voting scheme can be formulated by the following equations:

\[
MV(x, y) = \max \sum_{j=1}^{m} \delta_j C_{ij}(x, y), i = 1, 2, \ldots, n
\]

\[
\delta_j = \begin{cases} 
0; & \text{the report value from sensor } j \text{ is not } i \\
1; & \text{the report value from sensor } j \text{ is } i 
\end{cases}
\]

\[
C_{ij}(x, y) = \begin{cases} 
0; & \text{point } (x, y) \text{ is not covered by sensor } j \\
1; & \text{point } (x, y) \text{ is covered by sensor } j 
\end{cases}
\]

3.2. Distance Weighted Voting Algorithm
Distance Weighted Voting (DWV) is a weighted variant of MV. DWV is motivated by the assumption that the sensor nearest to the point in question has the most accurate data. Therefore, data closes to the point in question bares more weight in terms of decision making. Suppose $d_{j,(x,y)}$ is the distance from point $(x,y)$ to sensor $j$, the number of deployed sensors is $m$, and the possible report value from each sensor is an integer from 1 to $n$. DWV can be formulized by the following equation:

$$DWV(x,y) = \max_i \sum_j \frac{1}{d_{j,(x,y)}} \delta_i C_{j,(x,y)}, i=1,2,...n$$

where $\delta_i$ and $C_{j,(x,y)}$ shares the same definition as in MV.

3.3. Confidence Weighted Voting Algorithm

Like DWV, Confidence Weighted Voting (CWV) is another weighted variant of MV. Yet, instead of granting the nearest sensors higher weights, CWV gives higher weights to those sensors that are more likely to be correct (i.e. with higher confidence of correctness). The confidence value of each sensor can be determined in a distributed manner by comparing its sensing results with its sensing neighbors that share overlapping coverage area. The confidence value of sensor $i$, $conf(i)$ is then defined as:

$$conf(i) = \frac{\sum \Delta_{i,j} A_{i,j}}{\sum A_{i,j}}$$

$$\Delta_{i,j} = \begin{cases} 0; \text{if sensor } i \text{ and } j \text{ report different results} \\ 1; \text{if sensor } i \text{ and } j \text{ report the same result} \end{cases}$$

$$A_{i,j} = \begin{cases} 0; \text{if the coverage of sensor } i \text{ and } j \text{ is not overlapped} \\ 1; \text{if the coverage of sensor } i \text{ and } j \text{ is overlapped} \end{cases}$$

and CWV is then formulized as:

$$CWV(x,y) = \max_i \sum_j conf(j) \delta_i C_{j,(x,y)}, i=1,2,...n$$

where $\delta_i$ and $C_{j,(x,y)}$ shares the same definition as in MV.

4. Simulation Results

In this section, we evaluated the reliability of the three distributed voting algorithms described in section 3 according to the metrics presented in section 2.2. The robustness of the algorithms is assessed against varying degree of sensor failure rate and $k$-cover strategy. We used Monte Carlo simulations in section 4.1 to contrast the reliability of the three schemes, and we used an analytical model to prove the effectiveness of CWV against MV in section 4.2.

4.1. Reliability of Different Voting Algorithms

Figure 2 illustrates the reliabilities of different distributed voting algorithms under different degree of coverage and sensor error rate. The reliability of the three schemes clearly decreases as the sensor error rate increases, and reliability increases as the degree of sensor coverage increases. It is obvious that reliability increases with data redundancy. In particularly, when the sensor error rate is at 40%, MV improved 7% in reliability when degree of coverage increased from 1 to 2, additionally, when the degree of coverage increased from 2 to 3, MV experience another 17% in improvement. For CWV, it gained 33% in improvement in reliability when degree of coverage increased from 1 to 2, it also experience another 10% in improvement when degree of coverage increase from 2 to 3. This indicates that CWV can better utilize the added redundancy and achieved higher reliability. On the other hand, DWV scheme improves very little from the increase in degree of coverage. This is partly due to the fact that DWV rely heavily on the nearest neighbor’s result; therefore it is more likely to be biased when its nearest neighbor’s data is incorrect.
is at 40%, CWV outperforms MV by 7%, 34%, and 28% when the degree of coverage is 1, 2, and 3 respectively.

From the simulation results, it is clear that higher degree of coverage can achieve better data reliability. However, since high degree of coverage usually requires more sensor nodes and deployment cost. The design tradeoff between reliability and degree of coverage should be considered when deploying such a technique. From a communication overhead perspective, CWV algorithm incurs roughly twice the amount of overhead as would MV; therefore, reliability trade-off with communication overhead should also be considered when a distributed voting algorithm is to be used.

Notice that when majority of the sensors are reporting incorrect values (sensor error rate greater than 0.5), none of the schemes are expected to provide acceptable reliability in these scenarios. Therefore, discussion on those cases is not very meaningful.

4.2. Analysis

In this section, we present an analytical model for the Majority Voting scheme, and used the modeling result to discuss the reliability issue associated with different degrees of coverage and sensor error rates. For simplicity, we use a 1-cover placement strategy discussed in section 2.3, and the knowledge that k-cover can be roughly achieved by overlapping k 1-cover placements on the investigating rectangle area.

The analytical model of 1-cover placement can be derived by dividing the investigating rectangular area into several smaller equilateral triangles with side length equals to r, which is the same r as the a sensor’s sensing radius, this is also illustrated in Figure 3. Furthermore, in each equilateral triangle, the gray-color area is covered by exactly one sensor, and the white-color area is covered by two sensors. The overall system reliability can then be approximated by modeling the reliability of one equilateral triangle area; this is assuming that the width and length of the investigating area is much greater than sensors’ sensing range.

Suppose that 1-covered area within the equilateral triangle (gray-color area) is $A_1$, the 2-covered area within the equilateral triangle (white-color area) is $A_2$, and let the sensor error rate to be $e$. The system reliability $R_i$ (reliability of 1-covered area) can be modeled as:

$$R_i = 1 - \frac{A_1 e - A_2 e (1 - e)}{A_1 + A_2}$$

Since $A_1 = \left(\frac{\sqrt{3}}{2} \frac{\pi}{6}\right)^2$ and $A_2 = \left(\frac{\pi}{6} - \frac{\sqrt{3}}{4}\right)^2$ we get

$$R_i = 1 - e^{\left(2\pi - 3\sqrt{3}\right) + \left(2\pi - 3\sqrt{3}\right)}$$

In order to model k=n, n 1-covered placements are overlapped on the same investigating rectangle area. Error is reported in this model when either 1) the majority of the covered sensors are erroneous, or 2) half of the sensors are faulty and the random decision outputs the incorrect information. Therefore, the overall system reliability can be modeled with two cases:

Case 1: when n is odd

$$R_n = 1 - \sum_{i=1}^{n} \left(1 - R_i \times R_i^{-n/2}\right)$$

Case 2: when n is even

$$R_n = 1 - \sum_{i=1}^{n} \left(1 - R_i \times R_i^{-n/2}\right) - \frac{n}{n/2} \left(1 - R_i \times R_i^{-n/2}\right)$$

Based the analytical model above, Figure 4 depicts the relation of reliability against different sensor error rates. From the graph, we observed decreasing marginal gain in reliability as degree of sensor coverage increases. This is further evidence that placement strategy and reliability requirement is a design tradeoff that need to be considered before deployment.

Figure 3: Analysis of 1-coverage placement

Figure 4: Reliability of MV with different coverage degrees

Note that the analytical model in this section is based on the simplified assumption that allows modeling k-cover placement by overlapping k 1-covered placements. If a better placement technique is used (e.g. through combination of 2-cover and $k$-cover placement method discussed in section 2.3), it is possible to obtain better reliabilities than our analytical model, although the difference between the real reliability and the modeling one should be moderately small.
at least 9 using MV; whereas to provide 80% reliability with 0.4 sensor error rate, the coverage degree must be larger than 17 if MV is used. From this figure, it is obvious that sensor deployment cost can easily reach unacceptable level if MV scheme is used.

However, recalling the simulation results depicted in Figure 2, CWV can easily achieve 95% reliability with 3-covered placement at 0.4 sensor error rates. It is evident that although redundancy in coverage can improve data reliability for MV scheme, a well-designed voting strategy (e.g. CWV) can achieve even better reliability at a much lower cost. As a result, CWV indeed outperforms MV in terms of effectiveness.

Figure 5: Reliability of MV with different coverage degree when sensor error rate is fixed

5. Conclusions
To improve the reliability of decisions and minimize the impact of faulty sensor measurements, it is important to have a distributed scheme that enhances the trustworthiness of result by exploiting the presence of redundant data. In this paper, we present Confidence Weighted Voting (CWV), a simple distributed technique that improves the reliability of underlying data by exploiting redundant information. Since CWV uses neighboring data to discern the correctness of local data, it is capable of improving the baseline reliability of many applications. We examined CWV against the MV and DWV techniques, and contrasted the level of data reliability of each approach in the prevalent presence of flawed sensors. We simulated the basic behaviors of CWV via Monte Carlo simulations, and created an analytical model to prove the effectiveness of CWV over the other two schemes. Our results showed that CWV can consistently outperform the other distributed voting schemes by providing as much as 49% more resiliency to sensor errors.

6. References
Table 1: Details on $k$-Cover placement strategy

<table>
<thead>
<tr>
<th>Coordinate System</th>
<th>1-coverage</th>
<th>2-coverage</th>
<th>3-coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of rows</td>
<td>$\left\lceil \frac{2l}{\sqrt{3}r} \right\rceil + 1$</td>
<td>$\left\lceil \frac{l+3r}{2} \right\rceil, \left\lceil \frac{l+r}{2} \right\rceil, \left\lceil \frac{l}{3r} \right\rceil, \left\lceil \frac{l-r}{3r} \right\rceil$</td>
<td>$\left\lceil \frac{2l}{\sqrt{3}r} \right\rceil + 1$</td>
</tr>
<tr>
<td>Number of columns</td>
<td>Odd row: $\left\lceil \frac{w-r}{3r} \right\rceil$</td>
<td>Even row: $\left\lceil \frac{w}{3r} \right\rceil + 1$</td>
<td>Odd: $\left\lceil \frac{w}{r} \right\rceil + 1$</td>
</tr>
<tr>
<td>Total number of sensors</td>
<td>$\left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w-r}{3r} \right) \times \left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w}{3r} \right) \times \left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w}{3r} \right) \times \left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w}{3r} \right)$</td>
<td>$\left( \frac{l+3r}{2} \right) \times \left( \frac{l+r}{2} \right) \times \left( \frac{l}{3r} \right) \times \left( \frac{l-r}{3r} \right) \times \left( \frac{w}{3r} \right) \times \left( \frac{w}{3r} \right) \times \left( \frac{w}{3r} \right)$</td>
<td>$\left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w}{3r} \right) \times \left( \frac{2l}{\sqrt{3}r} \right) \times \left( \frac{w}{3r} \right)$</td>
</tr>
</tbody>
</table>

Sensor Coordinates

Odd row:

Even row:

$\left( \frac{r+3m}{3}, \frac{\sqrt{3}r}{2}, \frac{\sqrt{3}mr}{2} \right)$ if $m = 0, 1, 2, \ldots, \left\lceil \frac{2l}{\sqrt{3}r} \right\rceil - 1$ and $n = 0, 1, 2, \ldots, \left\lceil \frac{w-r}{3r} \right\rceil - 1$.