Decentralized Distribution Update Algorithm Based on Compressibility-controlled Wireless Sensor Network

Hehua Li and Chunling Wu
Institute of Information Security Technology, Chongqing College of Electronic Engineering, Chongqing 401331, China
Email: {lihehuaqce@yeah.net; wuh50@126.com}

Wei Wei
Xi’an University of technology, School of Computer Science and Engineering, Xi’an, China
E-mail: weiwei@xaut.edu.cn

Buddha Khoukhi
School of Computer Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON Canada K1S 5B6, Canada
E-mail: Budkhi@scs.carleton.ca

Abstract—Sensor network adopts the lossy compression techniques to collect long-term data, analyze the data tendency and the interested specific data model. In these applications, the sensor is established to collect large numbers of continuous data, and allow the access to the lossy and untimely data. In addition, the neighbor sensor data is correlated both in time and in space. Therefore, the data sensed by the sensor itself in the intermediate node and the data will be lossy compressed for prolonging the system operation lifetime. To study the optimal distributed problem of the bite-rate and the lossy degree. When the optimal distribution problem between the bit-rate and the lossy degree is discussed, how to make the optimal decision distributes the compressibility of all sensors in the satisfaction of the acceptable data distorted condition. For adopting the minimum transmittal bit-rate to collect the top-quality data. The optimal solution is introduced for the distribution problem, and the decentralized distribution algorithm is introduced in terms of the optimal solution. Compared with the average distribution strategy, the simulation result shows that the optimal solution and the decentralized actually can reduce large numbers of the network transmittal data volume.

Index Terms—Wireless Sensor Network, Slepian-Wolf Coding, Optimal Compressibility Configuration, Distributed Compressibility Configurations

I. INTRODUCTION

With the improvement of the technology in recent years, Internet of Things is called the third wave following the computer and the Internet in the world information industry. The experts in IT field think that Internet of Things can not only improve the efficiency of the economy and save cost, but also offer the powerful technology motivation for the recovery of the global economy. Internet of Things and the sensing network technology has been regarded as the critical strategy to revitalize the economy and establish the global competitive advantage in the world. The sensing technology is one of the critical technologies in the Internet of Things, and the Wireless Sensor Networks (WSNs) is widely applied in the medical treatment, disaster prevention and supervision, environmental polluted supervision, smart household, military purpose and other fields [1]. The paper emphasizes on long-term collecting the application environment of the mass continuous data. These systems mainly collect long-term data tendency, analyze the meaningful statistical data and find out the interested specific data model. Therefore, a small number of data is generally allowed to be distorted and to collect the untimely information. The sensors are established to collect large numbers of the continuous data in these systems, and the data in the neighbor sensors has the correlation both in time and in space [2][3]. In order to make the use of these correlations to reduce the data transmission quantity, the sensors located in the network transmittal intermediate nodes can compress those data with the use of lossy compression techniques after the data collected by the sensor itself and transmitted by the neighbor sensors is blended. The important two factor of influencing the lossy compression are bit-rate and lossy degree. The bit-rate shows that the transmittal digits needed by each number of the data, and the lossy degree is generally defined as the error between the restored data and the original data [4]. Such as Mean-Square Error, MSE.

The correlation between the bit-rate and the lossy degree can be described with the use of rate-distortion function. If the data’s correlation both in time and in space is high, the rate-distortion function will be rapidly descended with the increasing of the code’s bit-rate. Otherwise, if the data’s correlation both in time and in space is low, the rate-distortion function will be slowly
descended with the increasing of the code’s bit-rate. Therefore, the slope of rate-distortion function represents lossy degree to the changing degree of the increasing and decreasing bit-rate. When the function slope is steep, the sensors can use the fewer bit-rate to reach the appointed data quality. When the function slope is gradual, the sensors can use more bit-rate. The paper discusses the optimal distribution problem (ORDA) between the bit-rate and the lossy degree on the basis of the concept of the bit-rate and the lossy degree. How to make the optimal decision distributes the compressibility of all sensors in the satisfaction of the acceptable data distorted condition. The purpose lies in using the minimum transmittal data volume, collecting the top-quality data, prolonging the operation lifetime in the sensing network system (or the recharging time interval), reducing the system cost and promoting the commercial competitiveness [5] [6].

The data collected in the sink end can be stored with the two-dimensional array (as shown in figure 1). The vertical axis represents the data produced by N sensors; the horizontal axis represents the data produced by each sensor in a certain point-in-time. Each data can be regarded as three-dimensional data (sensor ID, Value and Time). The time dimension is divided into the equidistant $\Delta t$, each $<s_i,t_i>$ represents the data value produced by the sensor $s_i$ in the time interval $t_i+\Delta t$. The possible application ranges of the discussed technology in the paper are as follows (without limitation):

- Medical treatment supervision
- Ultraviolet and polluted particle supervision in the city
- Traffic flow supervision
- Environmental disruption&debris flow and other disaster prevented supervision
- Video supervision
- Climate variability supervision
- The power consumed supervision in the smart grid
- The data collection needed by the scientific research

\[\text{Range-Sum Query 2: } \sum (S_t, S_{t+\Delta t}) \text{ [15 t 4]) \sum (S_t, S_{t+\Delta t}) \text{ [65 t 8]}\]

\[
\begin{array}{cccccccc}
S_1 & 20 & 21 & 19 & 20 & 18 & 19 & 20 & 19 \\
S_2 & 20 & 22 & 20 & 20 & 21 & 20 & 10 & 20 \\
S_3 & 21 & 22 & 19 & 21 & 19 & 21 & 20 & 22 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
S_n & 15 & 18 & 16 & 19 & 18 & 17 & 18 & 19
\end{array}
\]

Figure 1. Data query

II. RELATED RESEARCHES

Many technologies related to the distributed compression have been proposed [7] [8], such as Slepian-Wolf Coding, multi-source data convergence coding and routing technology, the shortest routing probability compression technology and so on. According to the exploration of the spatial relevant locations, the distributed compression technology can be divided into two types: the distributed compression technology based on Slepian-Wolf Theory and the distributed compression technology based on the explicit communication. Slepian-Wolf Coding is mainly applied to the wireless sensor network system whose environmental change is rather static. The system can establish the time and spatial relevance among each sensor’s data in advance for the environmental change is rather static. Therefore, the associated information’s efficiency in the established global data can maintain a longer time. In the period of time, the system can effectively code each sensor’s data without the explicit communication. Each sensor can be coded independently in the sensor network system framework which applies to the Slepian-Wolf coding, and the Short Path Tree (SPT) is the optimal transmittal routing. Reference [7] is based on Slepian-Wolf Theory, Gauss Random Data Source and multi-hop sensor network framework. The centralized optimal distribution strategy is proposed aiming at the decentralized lossy compression.

When the environmental change is static, the relevance among the sensor data can be established in advance. The theory and the experiment prove that Slepian-Wolf coding is the optimal coding method. However, the environmental change and the relevance among the global sensor data are not static in the real application environment. Therefore, Slepian-Wolf coding presents the bad effectiveness. The sensor can reach the decentralized compression through the explicit communication under the situation of the dynamic environmental change. Reference [9] proposes the data loss in the spatial and temporal support in-network and the sensor network system in the stored multi-resolution compression and query framework. Reference [10] introduces the technology by using the decentralized wavelet to convert the compression piecewise smooth data. Considering to the problem of the minimum communication cost, Reference [11] assumes a simplified data source model and then proposes a heuristic data transmission algorithm. Reference [12] considers the non-distortion compression technology, and then applies the non-distortion compression to choose the optimal transmittal routing problem on the basis on the sensor communication, that is NP-hard.

Compared with the Slepian-Wolf coding technology, the decentralized compression technology through the sensor information communication has the following superior features: 1. The relevant information among the global sensor data is not to be established in advance. 2. Sensor can practice the complex and high-compression efficient coding algorithm.3. Practice the simple detected errors. The optimal distribution problem of the compression bit-rate and the lossy degree through the sensor communication is complex, the paper mainly solve the optimal problem of the bit-rate and the lossy degree.
The optimal distribution formulation can be inferred through the mathematically unanalytical formula [13][14][15].

III. IMPROVED DISTRIBUTION ALGORITHM

The wireless sensor network system framework discussed in the paper is shown in the figure 2. If there is a sink and N sensors in the wireless sensor network (\(s_i\)) represents the relations), \(X_i\)) represents the information collected by each sensor. The data captured by the environment from the sensor cannot be sent back to the sink one by one, while the data can be stored in the buffer. The data will be dealt with and then be transmitted through the lossy compression method at intervals for saving the power consumption. The sensors of the nodes located in the network can conduct further lossy compression to the received data and the retrieval data. After sink ending receives the compressed bit-stream, the data will be recovered. After sensor \(s_i\) conducts the lossy compression, \(R_i\) and \(D_i\) represent the bit-rate and the lossy degree. The power consumed by transmitting one-bite data to the father node in the network framework is represented by \(w\) (\(w\)) and the index power of the transmittal distance have an inverse relation [16].

\[
\begin{align*}
&\text{Objective function:} \\
&\min \sum_{i=1}^{N} w_i R_i \\
&\text{Subject to} \\
&H(D_1, \ldots, D_N) \leq D_i
\end{align*}
\]

The objective function in the formula (1) represents the total transmission cost consumed by transmitting all sensor’s data to the sink ending. Function in the constraint (2) represents that the lossy degree must be satisfied with all maximum allowable data lossy degree defined by users after the data in all sensors are recovered in the sink ending.

If the topology structure of the network transmit routing in the sensor is \(T\), \(h(i, j)\) represents the passed node numbers when sensor \(i\) and sensor \(j\) conduct the transmission in \(T\). Therefore, the data \(X_i\) in the sensor \(i\) along with the transmit routing in \(T\) is transmitted to sensor \(j\). The compressions are needed to be \(h(i, j)\) times. \(X^{(h(i,j))}_i\) represents the \(X_i\) recovery information in the sensor \(j\). MSE(mean-square error) in the \(X_i\) and \(X_i^{(h(i,j))}\) can be defined as follows:

\[
d(X_i^{(h(i,j))}, X_i^{(q)}) = E((X_i^{(h(i,j))} - X_i^{(q)})^2), \forall q \geq 0
\]

Considering the situation of collecting the data in the wireless sensor network, the node sensors in the network transmitted routing is in charge of the perceptive data by itself and the data received from other sensors, and then conducts the lossy compression. Through the intermediate node sensor, the lossy data will be transmitted to sink ending along with the network transmitted routing defined by \(T\). After the data is compressed through the intermediate node, the effect of transmitting the errors may be caused. Therefore, all sensor data’s lossy degree after being recovered from the sink ending can not be simply obtained by adding all lossy degrees caused by individual sensors. The lossy degree \(H(.)\) is counted as follows. \(H(.)\) can be shown:

\[
H(D_1, \ldots, D_N) = \sum_{i=1}^{N} d(X_i, X_i^{(h(i,j),\text{sink})})
\]

After the complex mathematical deduction, the estimation of lossy degree can be obtained when all sensor information is recovered in the sink ending:

\[
H(D_1, \ldots, D_N) = \sum_{i=1}^{N} D_i - 2N \rho
\]

\(\rho\) represents the relational factor between the error caused by the data \(X_i\) passes \(h(i, j)\) compression and the recovered data. The experimental result shows that the value of \(\rho\) and the variable degree are very small. The random \(\rho\) can be assumed, and the formula (4) can be changed as follows:

\[
H(D_1, \ldots, D_N) = \sum_{i=1}^{N} D_i - 2N \rho_i
\]

B. Optimal Mathematical Deduction

Considering the low-bit-rate coding, Mallat and Falzon [14] deducts that the bit-rate \(R_i\) and the lossy degree \(D_i\) in
the individual sensors can be estimated by the following deterministic rate-distortion model:

$$R_i = C_i \frac{1}{2^{\gamma_i}} D_i^{\frac{1}{2^{\gamma_i}}}$$  \hspace{1cm} (6)$$

The newly optimal distribution problem will be obtained by using the formula (5) and (6) to replace the formula (1) and (2).

$$\{D_i\}^{N}_{i=1} = \arg \min_{\{D_i\}_{i=1}^{N}} \sum_{i=1}^{N} w_i C_i \frac{1}{2^{\gamma_i}} D_i^{\frac{1}{2^{\gamma_i}}}$$ \hspace{1cm} (7)

Subject to

$$\sum_{i=1}^{N} D_i = 2N \rho \leq D_i$$ \hspace{1cm} (8)

Definition 1 proves the optimal solution of the newly distribution problem (7).

Definition: The proposed optimal solution of the bit-rate and lossy degree distributed problem in the paper is as follows:

$$R^{opt} = \alpha_i (\lambda^*)^{\frac{1}{2^{\gamma_i}}}, i = 1, \ldots, N$$

$$D^{opt}_i = \frac{\alpha_i w_i}{2^{\gamma_i} - 1} (\lambda^*)^{\frac{1}{2^{\gamma_i}}}, i = 1, \ldots, N$$ \hspace{1cm} (9)

where

$$\alpha_i = \frac{C_i (2^{\gamma_i} - 1)}{w_i}$$

$$\lambda^* = f(\lambda)$$ unique root.

$$f(\lambda) = \sum_{i=1}^{N} \frac{a_i w_i}{2^{\gamma_i} - 1} \lambda^{\frac{1}{2^{\gamma_i}}} - 2N \rho - D_i = 0$$ \hspace{1cm} (10)

Therefore, it can be given the lossy degree whose configurations are large, and the transmittal data numbers can be reduced. When the values of $\alpha_i$ and $\gamma_i$ in the sensors are decreasing, the relevance of the collected data is higher, and the small number of the bit-rate can reach to the low lossy degree. Therefore, it can be given the lossy degree whose configurations are small.

As to the limitation of the pages in the paper, the deducted procedures are omitted. \( f(\lambda) \) in the formula (10) is the convex function which is increasing. Therefore, the unique solution can be obtained through the simple bisection algorithm. Formula (9) shows that the formula of the optimal solution is related to the parameters of $\alpha_i$ and $\gamma_i$. When the values of $\alpha_i$ and $\gamma_i$ in the sensors are increasing, the relevance of the collected data is lower.

C. Decentralized Distribution Algorithm

The paper also extends the optimal distribution solution in the formula (9), and then proposes the decentralized distribution algorithm. If the value of $\gamma_i$ can be replaced by the average value (the assumption is based on the observed result whose variance of the $\gamma_i$ value is small), namely, $\gamma_i = \bar{\gamma}$.

$$\bar{\gamma} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i$$

The cost of the assumption is that the estimation of the formula (6) will become inaccurate. What’s worse, the optimal solution (9) will be influenced. The optimal solution (9) will be deducted into the decentralized distribution theology solution on the basis of the assumption. The experimental result shows that the assumption just can produce the iny error. From the observation of the formula (9), $\alpha_i$ is the maximum parameter of influencing the optimal solution. $D_i^{\text{gross}}$ represents the accumulated lossy degree of the data source $X_i$. $D_i^{\text{gross}}$ is defined as the total lossy degree of the individual sensors whose sensor $i$ is the sub-transmition routing tree $T_i$ of the root node.

$$D_i^{\text{gross}} = D_i + \sum_{j \in I_i} D_j$$

$A_i$ represents the total $\alpha_0$ parameter in the sensor $i$ itself and all sensors belonging to $T$ sub-tree.

$$A_i = a_i w_i + \sum_{j \in I_i} a_i w_j$$

$I_i$ represents the immediate children in the sensor $i$ belonging to $T$ sub-tree. Theorem proves the solution of the decentralized distribution problem.

Theorem 2: Assuming $\forall i, \gamma_i = \bar{\gamma}$, the decentralized distribution solution in the formula (9) is as follows:

$$D_i = \frac{a_i w_i}{A_i} D_i^{\text{gross}}$$ \hspace{1cm} (11)

$$D_i^{\text{gross}} = \frac{A_i}{A_i} D_i^{\text{gross}}, \forall j \in I_i$$ \hspace{1cm} (12)

According to formula (11) and formula (12), the collected lossy degree $D_i^{\text{gross}}$ in the sensor $i$ can be obtained by its father node. The $D$ can be directly obtained through the information of $D_i^{\text{gross}}$.

The calculation of the whole decentralized distribution starts from sink ending, and adopts the top-down manner to calculate the sensors of the node located in the end. The procedure of the whole calculation is shown in the figure 4. Each sensor will calculate itself and the sensor $i$ located in the sub-tree in the system initiation phase. When the system presents the object lossy degree $D_t$, sink adopts formula (2) to calculate the collected lossy degree $D_i^{\text{gross}}$ of all sub-nodes related to the sink communication, and then transmits $D_i^{\text{gross}}$ to all sub-nodes (step1 to step5). After the sub-nodes receive $D_i^{\text{gross}}$, its lossy degree distribution $D_i^{\text{opt}}$ can be calculated through the formula (11) (Rican be calculated through the formula 6), and then adopts the formula (12) to calculate $D_i^{\text{gross}}$ of the sub-nodes (step6 to step11). Later, $D_i^{\text{gross}}$ can be transmitted to the sub-nodes. The whole calculating process starts from sink. The recursive calculation can be conducted until the ending nodes in the $T$ (step12 to step13). According to the calculation of the figure 4, each node in the transmitting tree $T$ just...
conductions once distributed calculation. Therefore, the complexing degree is $O(N)$.

**Algorithm 1 DRDA Scheme**

Require: $D_{\text{prev}}$
Ensure: $D_j$
1: if Node $i$ is the sink then
2: for $j \in I_i$ do
3: \quad Compute $D_j$ using (12).
4: \quad end for
5: Send $(D_{\text{prev}})_{i \in I_i}$ to immediate child node $j$.
6: else if Node $i$ is an internal node then
7: Compute $D_i$ by (11).
8: for $j \in I_i$ do
9: \quad Compute $D_j$ using (12).
10: \quad end for
11: Send $(D_{\text{prev}})_{i \in I_i}$ to immediate child node $j$.
12: else if Node $i$ is a leaf sensor node then
13: $D_i = D_{\text{prev}}$.
14: end if

Figure 4. Decentralized distribution algorithm

In order to reduce the transmitted information during the updating of the sensors and the distribution, the paper also introduces a decentralized distributed update algorithm (as shown in the figure 5). The formula (11) and the formula (12) can observe that $\frac{1}{N} \sum_{k=1}^{N} a_k w_k$ can be obtained when the $D_j$ is calculated through the following ratio.

**Algorithm 2 Distributed Distortion Update**

1: Let $UC(i)$ be the set of nodes in $T_i$ that transmit update messages.
2: if Node $i$ is a leaf sensor node then
3: if $|A_i^{\text{new}} - A_i^{\text{old}}| > \epsilon \cdot |A_i^{\text{old}}|$ then
4: \quad Update $A_i^{\text{old}}$.
5: \quad Send an update request to the parent node.
6: \quad end if
7: else if Node $i$ is an internal node then
8: Compute $A_i^{\text{new}} = A_i^{\text{old}} - \left(\sum_{UC(i) \in T_i} \alpha_i w_k + \alpha_i w_k \right)$.
9: if $|A_i^{\text{new}} - A_i^{\text{old}}| > \epsilon \cdot |A_i^{\text{old}}|^2$ then
10: \quad Update $A_i^{\text{old}}$ and $\alpha_i^{\text{old}}$.
11: \quad Send an update request to the parent node.
12: \quad end else if $UC(i) \neq \emptyset$ then
13: \quad end if
14: Execute Algorithm 1 to recompute a new distortion allocation for nodes in $T_i$.
15: end if
16: end if

Figure 5. Decentralized distribute update algorithm

Therefore, sensor $i$ can decide whether a group of new distribution can be requested to be calculated again according to the error size of $|A_i^{\text{new}} - A_i^{\text{old}}|$. When the error is small, the calculating movement of the distributed update be stopped in the node $i$ so that the transmitted information consumed by continuing to be updated can be reduced. When sensor $i$ is the ending node, and is satisfied with the $|A_i^{\text{new}} - A_i^{\text{old}}| > \epsilon \cdot |A_i^{\text{old}}|$ (step2 to step6), the update information including the new parameter $A_i^{\text{new}}$ will be given to the father node.

Otherwise, the updating movement of the sub-tree $T_i$ whose root is sensor $i$ will be stopped in the node $i$ and cannot continue to transmitting the information. Later, node $i$ just need to conduct the decentralized distribution algorithm in the figure 4 and update the distribution of all sensors in the sub-tree $T_i$. Each node just needs to conduct once redistributed calculation under the worst situation. Therefore, the complexing degree in the figure 5 algorithm is $O(N)$.

IV. SIMULATING EXPERIMENTAL RESULTS

A. Simulating Experimental Setting

The simulating experiment is implemented through the consumed-power model proposed in the reference [17]. The object application should be set up as the air-conditioning monitoring system in the supermarket or library. The sensor network topology of the multi-hop is produced by the following methods: sink is put in the center of the network, and 30 sensors are put in the 200-mx200-m rectangle region randomly. The maximum transmitting distance of each sensor is set up as 40m. The real transmitting range can reduce the consumed transmitting power according to the dynamic distance adjustment between each sensor and its father node. The transmitting routing tree of the data is nearly the shortest routing tree whose maximum height is 3. The data in the experimental result is the average value obtained from many groups of network topology experimental result. One of a group of the example is as shown in the figure 6. The figure 6 is used as the interpretation figure of the experimental result.

Figure 6. The network topology model in the stimulation experiment

The experiment adopts the temperature data set provided by Live from Earth and Mars project, University of Washington. The data set has been widely used as the simulation experiment of the related researches in the sensor network. The collected interval of each temperature data in the data set is one minute. The data set is cut into many sub-data set whose distance is equal, and then the data collected from each sensor is simulated. The neighbor sub-data set is defined as the neighbor sensor’s data, and then the spatial relevance of the data is simulated. Figure 7 is the representative data section of LEM data. The higher peak of the information value represents daytime, and the higher valley of the information value represents night. The temporary storage length of each sensor is set up as 64 minutes. Each sensor adopts Embedded zero-tree wavelet (EZW) coding to execute once lossy compression for the data
every 64 minutes. The transmitting packing length of updating each lossy degree is 32bits in the experiment. The experimental results show that the average result is obtained every 50 times.

B. Performance Comparison

Figure 8 is the relational graph between the object lossy degree and $\varepsilon$. Apparently, $\rho$ is the small negative value and the small variance, and then $\rho$ is decreasing with the increasing of $D_t$. Figure 9 is the relational graph between the consumed power and $D_t$. ORDA in the figure 9 is the optimal distribution. DRDA represents the decentralized distribution and URDA represents the average distribution strategy. The figure 9 shows that ORDA and DRDA are superior to the average distribution. For example, when $D_t$ is 2, ORDA improves 16 times of the URDA’s transmitting power. When $D_t$ is 4, ORDA improves 13 times of the URDA’s transmitting power. When $D_t$ is smaller, the advantages of ORDA and DRDA are more apparent. Figure 10 is the relational figure between the transmitting volume and $\varepsilon$ in the decentralized distributed situation. Apparently, the relevance between the transmitting volume and $\varepsilon$ is not very high. $\varepsilon=0.3$ is the nice choice in the experimental examples. Figure 11 is the relational figure among the whole transmitting power, the data transmitting power and the distributed updating communication consumed power. Apparently, $\varepsilon$ has not big influence to the whole transmitting power.

V. CONCLUSION

Aiming at the strategy of the optimal distribution and the decentralized distribution proposed by the multi-hop wireless sensor network, the experimental results show that the transmitting power consumed by the two proposed distributed algorithms is lower than the average distribution. When the system has the following characteristics: 1. The density of the sensor is higher; 2. The distribution of the sensor is not rule; 3. The data of sensor presents a larger time and spatial relevance, the proposed methods are more effective.

ACKNOWLEDGMENT

This work is supported by the Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.2013JK1139).

REFERENCES


