Bounded-Error-Pruned Sensor Data Compression for Energy-Efficient IoT of Environmental Intelligence

Ray-I. Chang 1, Yu-Hsien Chu 1, Lien-Chen Wei 1 and Chia-Hui Wang 2,*

1 Department of Engineering Science and Ocean Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan; rayichang@ntu.edu.tw (R.-I.C.); d00525003@ntu.edu.tw (Y.-H.C.); d01525010@ntu.edu.tw (L.-C.W.)
2 Department of Computer Science and Information Engineering, Ming Chuan University, No. 5 Der-Ming Rd., Gwei Shan District, Taoyuan City 333, Taiwan
* Correspondence: wangch@mail.mcu.edu.tw; Tel.: +886-3-350-7001 (ext. 3421)

Received: 31 July 2020; Accepted: 14 September 2020; Published: 18 September 2020

Abstract: Emerging IoT (Internet of Things) technologies have enjoyed tremendous success in a variety of applications. Since sensors in IoT consume a lot of energy to transmit their data, data compression used to prolong system lifetime has become a hot research topic. In many real-world applications, such as IEI (IoT of environmental intelligence), the required sensing data usually have limited error tolerance according to the QoS (quality of sensor service) or QoD (quality of decision-making) required. However, the bounded-error-pruned sensor data can achieve higher data correlation for better compression without jeopardizing QoS/QoD. Moreover, the sensing data in widely spread sensors usually have strong temporal and spatial correlations. We thus propose a BESDC (bounded-error-pruned sensor data compression) scheme to achieve better leverage between the bounded error and compression ratio for sensor data. In this paper, our experiments on a sensor network of two-tier tree architecture consider four different environmental datasets including PM 2.5, CO, temperature and seismic wave with different scales of bounded errors. With the bounded errors required by the given IEI applications, our BESDC can reduce the total size of data transmission to minimize both energy consumption in sensor-tier devices and storage of fog-tier servers. The experimental results demonstrate that our BESDC can reduce transmission data by over 55% and save 50% energy consumption when assigning 1% of error tolerance within QoS2/QoD requirement. To the best of our knowledge, the proposed BESDC scheme can help other energy-efficient IoT schemes applying network topologies and routing protocols to further enhance energy-efficient IoT services.

Keywords: energy-efficient IoT; environmental intelligence; bounded-error sensor data; lossless compression

1. Introduction

In recent years, emerging IoT (Internet of Things) technologies have enjoyed tremendous success in varieties of applications [1–6]. They require sensor data to be timely, reliable and suitable for decision-making [7–9]. The quality of decision-making (QoD) depends on the quality of the information from sensor data. Although sensors are available for measuring many kinds of indicators, energy consumption remains one of the major limitations of IEI (IoT of environmental intelligence). Since sensors in IoT will consume considerable energy to transmit their data, data compression to prolong system lifetime has become a hot research topic. Different types of IEI applications can impose different quality of sensor service (QoS2) requirements. For example, an air quality monitoring application gathering air parameter measurements demands less stringent timing requirements than
a mobile robot navigation application. Koubâ, A. et al. [10] state that all IEI applications benefit from the QoS2 metrics of higher network throughput, lower message delay and longer system lifetime.

In many real-world applications, the required sensing data usually have limited error tolerance according to the QoD or QoS2 required. For example, in CWB (Central Weather Bureau) Taiwan, the bounded error of QoS2 is 0.1 degree Celsius for a temperature sensor [11]. As for a weather report, the bounded error of QoD is 1 degree Celsius. This motivates us to propose a system architecture for energy-efficient IEI, which is illustrated in Figure 1, to achieve better leverage between bounded error with preserved QoS2/QoD and compression ratio for different IEI applications. In this paper, the proposed BESDC (bounded-error-pruned sensor data compression) scheme can cost-efficiently prune the sensor data without QoS2/QoD deterioration in IEI applications, to prolong the sensor lifetime and save storage resource for IEI applications by reducing the size of transmitted sensor data.

The system architecture of an IEI application is illustrated in Figure 1. It indicates that many sensor nodes are deployed, and the sensor nodes are divided into two categories: end device (ED) and super node (SN). SNs are usually more powerful and faster than EDs in computing and communication resources and also much more expensive than EDs. In this paper, without loss of generality, we simply assume that each SN and its EDs are homogeneous and they are deployed in the monitored environment properly for IEI applications. The application-specific bounded error is given from the server (i.e., bounded error regulator) deployed on the fog-tier according to the required QoS2 or QoD for IEI applications.

In this paper, we apply a two-tier scheme to assign proportionally divided error bounds (pdEB) to the related sensor-tier devices (i.e., ED and SN) to guarantee the bounded error assignment (BEA) of pdEB from the bounded error regulator on the fog-tier. Then, BESDC determines whether the new sensing data should be compressed by comparing with not only the previous sensed data for temporal correlation but also the neighboring sensed data for spatial correlation, within the tolerated BEA. Besides, after bounded-error-pruning the sensor data within the required QoS2 or QoD, the dataset size for encoding can be reduced further to save resources of storage and transmission for sensor nodes and fog-computing machines. Ultimately, better sensor data compression can be further achieved in the bounded-error-pruned sensed values to save more energy consumption.
Experiments on four different environmental datasets, including PM 2.5, CO, temperature and seismic wave, are presented. Performance results show that the energy consumption of IEI sensor data transmission can be highly reduced even when a small bounded error is given. For example, when the error tolerance is assigned as 1%, BESDC can reduce transmission data by at least 55%. For a temperature sensor, 1% error tolerance of two-digit temperature value is around 0.5 degree Celsius in bounded error. Therefore, it is acceptable for applications to make a weather report with 1 degree Celsius bounded error in required QoS

The remainder of this paper is organized as follows. In Section 2, we survey the literature for IoT data compression including bounded error and unbounded error methods. In Section 3, we present BESDC and introduce the applied data, temporal and spatial correlation compression schemes. We use two-tier BEA for different devices running BESDC for sensor data. In Section 4, experiments and performance evaluation results of BESDC are shown. Finally, conclusions and future work are given in Section 5.

2. Related Works

IEI applications need rapid information delivery from environmental transitions. They require a timely, reliable and suitable process in collecting diversified sensor data. As sensor data communication dominates the energy consumption in IoT, many compression mechanisms have been applied to prolong the IoT lifetime [12]. Data compression schemes applied for energy-saving in IoT can be divided into two categories: lossy and lossless. Lossy compression schemes usually give much better compression ratios but introduce some signal fidelity in required QoS/QoD lost from sensor data. Schemes in this category include JPEG [13–15], discrete cosine transform (DCT) [16], wavelet-based [17], vector quantization (VQ) [18], and multi-resolution spatial and temporal coding (MRCQ) [19]. As the errors between the compressed and original data are unbounded, these works are unable to support real-world IEI applications to make proper decisions for users.

Modern IEI applications are usually based on lossless compression schemes including Huffman coding (HC) [20], static HC (SHC) [21], adaptive HC (AHC) [22], modified AHC (MAHC) [23] and improved MAHC (IMAHC) [24]. HC refers to the use of a variable-length codebook where symbols with higher probabilities to appear are encoded with fewer bits. Given statistics information of source symbols, SHC is proposed to construct a codebook with an optimal prefix code. AHC separates encoded data into prefixes and suffixes based on the weight of the data. Prefix can be redefined as adaptive to have a better compression ratio. MAHC merges the codebook with the optimal prefix code of SHC and the codebook with dynamic suffix of AHC to have a better compression ratio. IMAHC applies dynamic codebook to adjust to rapid environmental changes for IoT. These lossless compression schemes maintain complete data integrity but have poor compression ratios compared with lossy compression.

Correlations considered in data compression can be spatial correlation, data correlation and temporal correlation. Spatial correlation suggests that the sensor nodes located closely in an intensive area usually have similar sensed data. Thus, we can apply the spatial correlation to compress transmission data. Spatial correlation compression schemes include minimum fusion Steiner tree (MFST) [25], adaptive fusion Steiner tree (AFST) [26], data fusion (DFuse) [27], context adaptive clustering (CAC) [28], CODA [29], clustered aggregation (CAG) [30], prediction-based monitoring (PREMON) [31], and data fusion using temporal and spatial correlations (DF-TS) [32]. Different from spatial correlation, data correlation uses information theory to encode high probability data with short length output. An example of a data correlation compression scheme is decentralized data fusion (DDF) [33]. Temporal correlation is based on the idea that sensor nodes usually have similar sensed data in a short period of time. Temporal correlation compression schemes include multi-level fusion synchronization (MFS) [34], temporal coherency-aware in-network aggregation (TiNA) [35], CAG, PREMON, and DF-TS.

Usually, IEI applications will tolerate limit errors for their required QoD and QoS. Bounded error can be considered valuable information to reduce the communication cost. The simplest way, as in the
TiNa scheme, is to use the bounded error between current and former data to decide whether or not to transmit the current data. In this paper, we use bounded error in advance for data compression to not only reduce the communication/storage cost for sensors and fog-computing machines but also cost-efficiently extend the sensors’ lifetime. A tolerable bounded error is helpful not only to improve the data compression ratio [36–39] for energy-efficient IoT applications but also to reduce the query response time for big sensor data [40,41].

In our previous work [37] for a preliminary wireless sensor network (WSN) topology, we first applied static bounded error in codebook for data correlation compression and outperformed DF-TS in sensor energy saving. Then, the dynamic bounded error in more realistic WSN topology was considered in [38] to further validate the energy efficiency by four different BEA in linear/non-linear proportions. Furthermore, the authors of [36] improved previous works described above by applying run length encoding (RLE) in data correlation compression and revising the spatial correlation compression method.

Different with [36–38], in a wireless body sensor network (WBSN), a recent article [39] first applied an error-bounded lossy compression scheme [42] in high performance computing (HPC) and then used a machine learning technique to rebuild the transmitted data with given bounded error on the edge node. However, for IEI sensor networks and different QoS/QoD required in their various applications, our proposed architecture applies overall lossless compression schemes in data, spatial and temporal correlations among bounded-error-pruned sensor data. Without the need to rebuild the lossy transmitted data in fog-computing machines, the total data size in transmitting and storing can be minimized for energy saving to prolong IoT lifetime for IEI applications in order to quickly access the bounded-error-pruned sensor data.

3. Bounded-Error Sensor Data Compression (BESDC)

In an IEI application, the application-specific bounded error is given from a fog-computing machine called the bounded error regulator according to the QoS or QoD required by IEI decision-makers. Then, a two-tier BEA scheme is applied based on the current topology of the deployed IoT to assign the error bounds of pdEB to the fog-tier machines like sink and the sensor-tier devices of SN and ED. With the given error bounds, EDs will use temporal and data correlations to compress the receiving sensor data and send the compressed sensor data to the corresponding SN. With the incoming compressed data between different neighboring EDs, SN will use spatial and data correlation compression to compress incoming sensor data further. Then, SN will send out the compressed data to the corresponding sink and the fog-tier sink will forward these compressed data to the fog-tier servers. Finally, the corresponding servers will store and decompress the compressed sensor data according to the pdEBs. Then, they will save the decompressed bounded-error-pruned data to the database for later query from users.

3.1. Basic Processing Flow

The flow chart of our proposed BESDC scheme is illustrated in Figure 2. It can be simply divided into two phases: offline phase and online phase. The offline phase is processed only while a sensor node is newly deployed and needs to generate its initial codebook. In two-tier BEA, a bounded error for the fog-tier sink is obtained from the application-specific bounded error of the user-required QoS or QoD. We assigned the fog-tier bounded error $\tau_{total}$ to the sink by proportionally dividing the BEA, as shown in Figure 3. The sink will calculate the sensor-tier bounded errors $\tau_{1}$ to $\tau_{9}$ by Equation (1) with the proportional weight $w_{i}$ according to the status, capability, etc., of sensor $i$. This value $w_{i}$ can be intelligently assigned from the fog-tier server (i.e., bounded error regulator, BER) or just equally assigned $w_{i} = 1/n$ via the fog-tier sink. After the bounded error $\tau_{i}$ is assigned to the sensor node $i$, we try to evaluate the data correlation of sensed data for constructing the initial codebook. Usually, the servers can spend a time period evaluating the data correlation of sensed data for constructing
the initial codebook. In order to reduce the offline phase latency for constructing the initial codebook, different sensors can be assigned the same initial codebook by default or a fog-tier server like BER.

\[ \tau_i = \tau_{total} w_j \]  

\[ \sum_{j=1}^{n} w_j = 1 \]  

\[ S_{ED} \equiv \langle s_1, s_2, \ldots, s_n \rangle \]  

\[ \tilde{S}_{ED} \equiv \langle \tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n \rangle \]  

\[ \mid \tilde{S}_{ED} \mid \leq |S_{ED}| \leq n \]  

\[ \max_{1 \leq j \leq n} \mid \tilde{s}_i - s_j \mid \leq \tau_i \]  

\[ QoS^2(\tilde{S}_{ED}) = QoS^2(S_{ED}) \]  

\[ QoD(\tilde{S}_{ED}) = QoD(S_{ED}) \]  

In the online phase, the new sensed data of ED, which are represented as in Equation (3), should be compressed or not by comparing with former sensed data (i.e., temporal correlation between sensor value \( s_{j+1} \) and \( s_j \)) and be encoded through the codebook (i.e., data correlation in sensor data similarity, as shown in Equation (5)) within the bounded error constraint \( \tau_i \) (as shown in the given criteria in Equation (6)). In Equation (3), \( S_{ED} \) is defined as a sequence of ED’s sensor values of length \( n \). \( \tilde{S}_{ED} \) is defined as a sequence of bounded-error-pruned sensor values from the sequence \( S_{ED} \), as the bounded error \( \tau_i \) given in Equation (4). \( \mid \tilde{S}_{ED} \mid \) and \( |S_{ED}| \) are the numbers of different values represented in sequences \( \tilde{S}_{ED} \) and \( S_{ED} \) in Equation (5), respectively.

Figure 2. The processing flow of our proposed BESDC scheme.
we try to count the frequency of same data value appearance. Due to the neighboring sensed data (i.e., spatial correlation) and be encoded through the codebook (i.e., data correlation) within the given bounded error. Finally, these compressed data after decompression and report the results for IIEI applications.

3.2. EB-RLE, EB-HC and Spatial Correlation Compression

RLE shortens a representation of a long sequence of a symbol by the sequence length and the symbol itself. For example, “rrrrrrrr” can be encoded as “8,r”. Notably, real sensor data usually have no long running lengths of the same value. Traditional RLE seems unsuitable for sensor data compression. However, in our EB-RLE, if a long sequence of data is within a certain bounded error range, it can be encoded by RLE. In this paper, we proposed a greedy algorithm to find the “longest sequence” of sensor data which can be represented as the same value within the given bounded error.

Figure 4 shows an example with a sequence of 10 sensor data (22.2, 22.6, 22.8, 23.1, 23.5, 23.7, 24.2, 23.6, 22.8) and the given bounded error 0.5. By using EB-RLE, we at first get the longest running sequence (22.2, 22.6, 22.8, 23.1) in the left gray-shaded area on Figure 4 within the given bounded error 0.5. They can be encoded as “4,22.7”. Then, we continue to get the longest running sequence (23.5, 23.7, 24.2, 23.6, 23.3) within the given bounded error 0.5. They can be encoded as “5,23.8”. The final EB-RLE result is an RLE sequence (“4,22.7”, “5,23.8”, “1,23.3”).

3.3. EB-HC

In conventional HC, as the length of code depends on the frequency of the data which appear, we try to count the frequency of same data value appearance. Different from HC, EB-HC can merge all data within the given bounded error range into one code. In this paper, we proposed a greedy algorithm to find the “largest group” of sensor data which can be represented by the same value within the given bounded error. Considering the same example of sensor data sequence (22.2, 22.6, 22.8, 23.1, 23.5, 23.7, 24.2, 23.6, 23.3, 22.8) in Figure 4, our EB-HC algorithm sorts the upper bounds and the lower bounds of the bounded error bars to obtain scan lines EB1 to EB17. For each scan line, the related counter represents its “bump-to-bar” counting result to indicate the number of sensor data that can be represented by one single value. We have the related counters EBi = 1, 2, 3, 4, 5, 5, 5, 7, 7, 7, 6, 5, 5,
5, 4, 2, 1 for \( i = 1 \) to 17. By greedy, we select EB8 to represent seven data including \( \{22.8, 23.1, 23.5, 23.7, 23.6, 23.3, 22.8\} \). These seven data can be represented by the same value 23.2. Then, \( \{22.2, 22.6\} \) can be represented by 22.7. The final result is a bit sequence 00111101111 with the codebook ("1,23.2", "01,22.7", "00,24.7").

For spatial correlation compression, SN processes multiple encoded data from neighboring EDs to further reduce data transmission to the fog-tier sink to extend the SN lifetime and save storage in fog-computing machines. While SN receives packets from neighboring EDs, SN verifies whether each sensor data encoding part (SDEP) in the payloads of two packets from similar geography regions is identical. For our proposed spatial correlation compression, if two packets share exactly the same SDEP, BESDC will use _S_NoOpt to replace one of the SDEPs. _S_NoOpt is encoded as “1” and the other SDEPs will be labeled by adding a “0” in front of them before they are encoded. Figure 5 illustrates two packets sent from two neighboring EDs and the packet format after they are compressed. The second SDEPs of these two packets are identical, so one of them will be replaced by _S_NoOpt. Then, this spatial correlation encoding scheme can further compress the incoming sensor data from different EDs to reduce the data transmission from SN to the fog-tier sink for more energy saving for SN.

**Figure 4.** An example of EB-RLE and EB-HC under bounded error 0.5.

**Figure 5.** Spatial correlation encoding for sensor packets from 2 neighboring EDs.
4. Experimental Results and Performance Evaluations

In this section, the compression ratio and energy consumption of IoT in BESDC for IEI applications are evaluated with four different datasets. Different scales of bounded errors are applied in these datasets. These environmental datasets are temperature, PM 2.5 and CO of air quality monitoring from [43] in Taiwan and seismic wave [44] from the Taiwan Chi-Chi earthquake which happened on 21 September 1999. Table 1 shows the features of correlation in each dataset.

| Type            | Correlation (Temporal, Spatial) | Samples | Frequency | Date Collected |
|-----------------|---------------------------------|---------|-----------|----------------|
| 1 Temperature   | High, High                      | 8760    | 1 sample/h. | 2018           |
| 2 PM 2.5        | High, Low                       | 8760    | 1 sample/h. | 2018           |
| 3 CO            | High, Low                       | 8760    | 1 sample/h. | 2018           |
| 4 Seismic Wave  | Low, Low                        | 10,000  | 10 samples/s. | 21 September 1999 |

This experiment with performance evaluation is divided into four parts. The first part introduces the processing in two-tier BEA for this experiment. Then, the second part introduces the calculation of compression ratio and energy consumption in performance evaluations from experiments. The third part evaluates the performance improvement from both temporal and data correlation compression of EB-RLE and EB-HC. The fourth part evaluates the performance improvement from spatial correlation compression of SN in BESDC scheme. The performance improvement is focused on the sensor data compression and energy saving ratios as compared with the original sensor data without compression.

4.1. Two-Tier BEA

This 2-tier BEA of BESDC assigns corresponding bounded errors to each sink from IEI BEA server. Without loss of generality, τ_{total} is set to the value described in Equation (9), where SD_{Max} is the maximal value among all sensor data collected from n devices (including SN and ED) under the sink and \rho_j is the jth inserted bounded error (0 < \rho_j < 1). In our experiment, the inserted proportional weight \omega_i for each sensor device i is simply 1/n and \rho_j is 0.25%, 0.5%, 0.75% and 1.0%. Then, the bounded error \tau_i for each sensor device i can be rewritten as Equation (10).

\[
\tau_{total} = n \times SD_{Max} \times \rho_j
\]  

\[
\tau_i = SD_{Max} \times \rho_j
\]  

4.2. Calculations of Compression and Energy Consumption Ratios for BESDC

The compression performance of BESDC is simply calculated by comparing the compressed sensor data size with the original sensor data size from the above-mentioned four different environmental datasets. The calculation is listed in Equation (11).

\[
\text{compression ratio} = \frac{\text{BESDC compressed data size}}{\text{Original data size}}
\]  

To evaluate the energy consumption, the parameters shown in Table 2 are adopted from [45]. According to these parameters from CC2420 transceiver and ARM7TDMI data-sheets, energy consumed in data compression is calculated according to the number of different machine instructions executed in the proposed BESDC algorithms on ED and SN via the assembly code output from Visual Studio 2010, including different instructions of ADD (addition), MUL (multiply), CMP (comparison) and SHF (shift). As well, energy consumed in data transmission (transmit and receive) is calculated according to
the number of bits compressed by BESDC in data, temporal and spatial correlations. In our experiments for BESDC performance, the energy consumption ratio is calculated as defined in Equation (12).

\[
\text{energy consumption ratio} = \frac{\text{Energy consumption with compression}}{\text{Energy consumption without compression}} \quad (12)
\]

The energy consumption function \( e_{TX}(n) \) in joule units (i.e., J) for ED or SN to transmit \( n \) bits can be calculated as defined in Equation (13). For example, to calculate the energy consumption to transmit 1 bit using the applied parameters in Table 2, we can have the energy consumption value of \( e_{TX}(1) = I_{TX} \times V_{TX} \times T_{TX} \times 1 = 17.4 \text{ mA} \times 3.3 \text{ V} \times 32 \text{ µs/bit} \times 1 \text{ bit} = 1.83744 \times 10^{-3} \text{ J} \) (i.e., \( T_{TX} \) indicates the time for ED or SN to send out 1 bit to network). The energy consumption \( e_{RV}(m) \) for ED or SN to receive \( m \) bits can be calculated as defined in Equation (14). Thus, we can have \( e_{RV}(1) = I_{RV} \times V_{RV} \times T_{RV} \times 1 = 2.08032 \times 10^{-3} \text{ J} \). Then, the energy consumption function \( e_{COM}(n, P_{sensor}) \) for \( n \)-bit compression via the output of assembly code from data compression program \( P_{sensor} \) running on sensors, such as ED or SN in proposed BESDC, can be calculated as defined in Equation (15).

In Equation (15), \( I_{add}^{P_{sensor}}, I_{mul}^{P_{sensor}}, I_{cmp}^{P_{sensor}} \) and \( I_{shf}^{P_{sensor}} \) are indicated as the respective numbers for different assembly instructions of ADD, MUL, CMP and SHF counted in running program \( P_{sensor} \). Then, \( \varepsilon_{add}, \varepsilon_{mul}, \varepsilon_{cmp} \) and \( \varepsilon_{shf} \) were and were the measured energy consumption values, respectively, in experience from [45] for each instruction of ADD, MUL, CMP and SHF in a program of encoding 1 bit. Therefore, for the sensor node of ED running EB-RLE with EB-HC in program \( P_{ED} \), we can have the ED total energy consumption function \( e_{ED}^{ED}(n) \) as shown in Equation (16), for compressing \( n \)-bit data and transmitting compressed \( \tilde{n} \)-bit data to network. In Equation (17), we can have the total energy consumption function \( e_{SN}^{SN}(m) \) for SN running spatial correlation compression with EB-HC in program \( P_{SN} \) to receive \( m \)-bit data from ED and compress them to \( \tilde{m} \)-bit data, finally sending the compressed \( \tilde{m} \)-bit data to fog-computing machines.

\[
e_{TX}(n) = I_{TX} \times V_{TX} \times T_{TX} \times n \quad (13)
\]
\[
e_{RV}(m) = I_{RV} \times V_{RV} \times T_{RV} \times m \quad (14)
\]
\[
e_{COM}(n, P_{sensor}) = n \times \left[ I_{add}^{P_{sensor}} \varepsilon_{add} + I_{mul}^{P_{sensor}} \varepsilon_{mul} + I_{cmp}^{P_{sensor}} \varepsilon_{cmp} + I_{shf}^{P_{sensor}} \varepsilon_{shf} \right] \quad (15)
\]
\[
e_{ED}^{ED}(n) = e_{COM}(n, P_{ED}) + e_{TX}(\tilde{n}) \quad (16)
\]
\[
e_{SN}^{SN}(m) = e_{RV}(m) + e_{COM}(m, P_{SN}) + e_{TX}(\tilde{m}) \quad (17)
\]

4.3. Performance Improvement from Temporal and Data Correlation Compression

In Figure 6, the compression performance of EB-RLE (temporal correlation compression) and EB-HC (data correlation compress) in BESDC shows the compression ratio for four different environmental datasets of PM 2.5, temperature, CO and seismic wave on four different bounded errors of 0.25%, 0.5%, 0.75% and 1%. As shown in Figure 6, the seismic wave dataset has worse EB-RLE/EB-HC compression ratio than others, since high-frequency seismic waves usually preserve low temporal and data correlations in sensor data values.
Since the energy consumption of the sensor node is dominated by data transmission, the computing of EB-RLE and EB-HC in BESDC will not affect too much the overall energy consumption in sensor nodes (i.e., EDs). Thus, the energy consumption ratio performance as shown in Figure 7 has a similar trend to Figure 6. This indicates that the higher data compression ratios preserve the lower energy consumption to substantially extend the IoT lifetime for IEI applications. It is known that the larger the bounded error given, the better compression and energy consumption ratios which should be preserved. However, the compression and energy consumption ratios of the PM 2.5 dataset with a bounded error of 0.75% are almost the same as the bounded error of 0.5%. This is because the experimental dataset of PM 2.5 preserves much the same value ranges as the bounded errors given in 0.5% and 0.75%.

4.4. Performance Improvement from Spatial Correlation Compression

BESDC spatial correlation compression can help the SN device to reduce the sensor data transmission to the sink after collecting compressed sensor data (data correlation and temporal
correlation compression) from its neighboring ED to the sink, comparing their similarity and compressing (spatial correlation). In the experiments for the compression and energy consumption performance in spatial correlation compression in BESDC, we use three different environmental datasets of PM 2.5, temperature and CO on two neighboring EDs (both in Taichung City, Taiwan) and two faraway EDs (in Northern and Southern Taiwan) to demonstrate the BESDC performance in spatial correlation compression and energy consumption on SN. Besides, we add the experimental results simulated for other energy-efficient IoT works [32] (i.e., DF-TS temperature as shown in Figures 8–11) from our previous work [37]) to further compare the performance improvement from the proposed BESDC.

![Figure 8](image1.png)

**Figure 8.** Data compression ratio in SN with neighboring EDs by spatial correlation compression.

![Figure 9](image2.png)

**Figure 9.** Energy consumption ratio in SN with neighboring EDs by spatial correlation compression.
As shown in Figures 8 and 9, the performance results of data compression and energy consumption for SN with two neighboring EDs using the BESDC spatial correction compression scheme show that the compression ratio and energy consumption ratio can be improved to extend the SN lifetime for IEI applications.

While considering spatial correlation compression for the SN with two faraway EDs, their data compression and energy consumption ratios on SN are shown in Figures 10 and 11, respectively. Compared with the performance results in Figures 8 and 9 (i.e., SN with two neighboring EDs), they have worse data compression and energy consumption ratios since two EDs are far away and have low spatial correlations in the experimental datasets.
However, in the experimental results of SN with two neighboring EDs, the compression and energy consumption ratios of the temperature dataset in bounded error of 0.75% are almost the same as the bounded error of 0.5% (i.e., BESDC temperature in Figures 8 and 9). This is the same reason (as mentioned in the previous subsection) that the experimental dataset of temperature preserves much the same value ranges as the bounded errors given in 0.5% and 0.75%. Then, compared to the quantitative results from DF-TS temperature in Figures 8–11, the compression and energy-consumption ratios from BESDC temperature show better improvement than DF-TS temperature. Therefore, according to the overall performance results, BESDC can comprehensively improve the performance of data compression and energy consumption for the SN to extend the IoT lifetime of IEI applications.

5. Conclusions and Future Work

In this paper, BESDC is proposed to reduce the energy consumption for extending the IoT lifetime with a given bounded error. Two-tier BEA is applied to assign the proper bounded error to EDs and SNs. Then, EB-RLE and EB-HC are used to improve temporal and data correlation (codebook) compression in sensor data within given bounded error. The spatial correlation compression is applied for SN to compress the bounded error sensor data difference among neighboring sensed data received from different EDs. Thus, the total size of transmission data in sensor-tier devices and fog-tier machines can be minimized to achieve energy and storage saving within the given bounded errors.

The experiments for four different environmental datasets in Taiwan, including forest temperature, PM2.5 and CO of air quality and seismic wave, demonstrate that our BESDC can reduce transmission data by around 28% to 68% and save around 20% to 72% energy consumption for ED when assigning 0.25% to 1% of error tolerance, respectively. For SN using the same bounded error range, spatial correlation compression can reduce transmission data by around 34% to 73% and save around 45% to 78% energy consumption, given that the SN connects to two neighboring EDs. Moreover, for SN with two faraway EDs, the transmission data can be reduced by around 28% to 72% and the energy consumption can be saved from 38% to 77%.

Since the proposed BESDC scheme is implemented between perception and network layers of IoT, we believe it can help other energy-efficient IoT schemes applying higher layer approaches, such as network topologies and routing protocols (e.g., [25,26,46]), to reduce their payload overhead for further enhancing their energy-efficient IoT services, without affecting the QoS/QoD required.

The future work will consider improving the proposed BESDC for cost-effective prediction of the lost sensor data via learning from the data received previously. We will also consider applying the smart contract technique in fast big sensor data query [40,41] with different bounded errors for users on differentiated service levels, since the smart contract using blockchain technology can preserve the resilience and transparency for big sensor data access at different levels of privacy controls.

Author Contributions: Conceptualization, R.-I.C.; methodology, R.-I.C.; software, Y.-H.C., L.-C.W.; validation, R.-I.C., C.-H.W.; formal analysis, R.-I.C., Y.-H.C., C.-H.W.; investigation, Y.-H.C., L.-C.W., C.-H.W.; data curation, R.-I.C., Y.-H.C.; writing—original draft preparation, Y.-H.C., L.-C.W., C.-H.W.; writing—review and editing, R.-I.C., Y.-H.C., C.-H.W.; visualization, Y.-H.C., L.-C.W., C.-H.W.; supervision, R.-I.C.; project administration, R.-I.C.; funding acquisition, R.-I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MOST Taiwan, grant numbers 108-2410-H-002-230-MY2 and 109-3116-F-002-002-CC1 and the APC was funded by 108-2410-H-002-230-MY2 and 109-3116-F-002-002-CC1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alloulah, M.; Huang, H. Future Millimeter-Wave Indoor Systems: A Blueprint for Joint Communication and Sensing. *Computer* 2019, 52, 16–24. [CrossRef]
2. Savazzi, S.; Sigg, S.; Vicentini, F.; Kianoush, S.; Findling, R. On the Use of Stray Wireless Signals for Sensing: A Look Beyond 5G for the Next Generation of Industry. *Computer* 2019, 52, 25–36. [CrossRef]
3. Satyanarayanan, M.; Davies, N. Augmenting Cognition Through Edge Computing. *Computer* **2019**, *52*, 37–46. [CrossRef]
4. Gao, N.; Shao, W.; Salim, F.D. Predicting Personality Traits from Physical Activity Intensity. *Computer* **2019**, *52*, 47–56. [CrossRef]
5. Tariq, N.; Asim, M.; Al-Obeidat, F.; Farooqi, M.Z.; Baker, T.; Hammoud, M.; Ghaifir, I. The Security of Big Data in Fog-Enabled IoT Applications Including Blockchain: A Survey. *Sensors* **2019**, *19*, 1788. [CrossRef]
6. Gill, S.S.; Tuli, S.; Xu, M.; Singh, J.; Singh, K.V.; Lindsay, D.; Tuli, S.; Smirnova, D.; Singh, M.; Jain, U.; et al. Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet Things* **2019**, *8*, 100118. [CrossRef]
7. Nabeeh, N.A.; Abdel-Basset, M.; El Ghareeb, H.A.; Aboelfetouh, A. Neutrosophic Multi-Criteria Decision Making Approach for IoT-Based Enterprises. *IEEE Access* **2019**, *7*, 59559–59574. [CrossRef]
8. Fang, S.; Da Xu, L.; Zhu, Y.; Ahati, J.; Pei, H.; Yan, J.; Liu, Z. An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1596–1605. [CrossRef]
9. Li, N.; Sun, M.; Bi, Z.; Su, Z.; Wang, C. A new methodology to support group decision-making for IoT-based emergency response systems. *Inf. Syst. Front.* **2013**, *16*, 953–977. [CrossRef]
10. Koubaa, A.; Severino, R.; Alves, M.; Tovar, E. Improving Quality-of-Service in Wireless Sensor Networks by Mitigating “Hidden-Node Collisions”. *IEEE Trans. Ind. Inform.* **2009**, *5*, 299–313. [CrossRef]
11. Wang, Y.J. Study on Physical Air Temperature Measurement. Available online: http://photino.cwb.gov.tw/rdcweb/lib/cd/07mb/MB/PDF/40/No.1/01.pdf (accessed on 31 October 2019).
12. Chang, R.-I.; Li, M.-H.; Chuang, P.; Lin, J.-W. Bounded-error Data Compression and Aggregation in Wireless Sensor Networks. In *Smart Sensors Networks*; Academic Press: Cambridge, MA, USA, 2017; pp. 143–157.
13. Bilgin, A.; Marcellin, M.W.; Altbach, I. Compression of electrocardiogram signals using JPEG2000. *IEEE Trans. Consum. Electron.* **2003**, *49*, 833–840. [CrossRef]
14. Liu, H.S.; Chuang, C.C.; Lin, C.C.; Chang, R.I.; Wang, C.H.; Hsieh, C.C. Data compression for energy efficient communication on ubiquitous sensor networks. *J. Appl. Sci. Eng.* **2011**, *14*, 245–254.
15. Lin, C.C.; Chuang, C.C.; Chiang, C.W.; Chang, R.I. A novel data compression method using improved JPEG-LS in wireless sensor networks. In Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT) IEEE, Phoenix Park, Korea, 7–10 February 2010; Volume 1, pp. 346–351.
16. Bendifallah, A.; Benzid, R.; Boulemden, M. Improved ECG compression method using discrete cosine transform. *Electron. Lett.* **2011**, *47*, 87. [CrossRef]
17. Chen, J.; Ma, J.; Zhang, Y.; Shi, X. A Wavelet-Based ECG Compression Algorithm Using Golomb Codes. In Proceedings of the 2006 International Conference on Communications, Circuits and Systems, Guilin, China, 25–28 June 2006; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA; Volume 1, pp. 130–133.
18. Fang, J.; Li, H. Hyperplane-Based Vector Quantization for Distributed Estimation in Wireless Sensor Networks. *IEEE Trans. Inf. Theory* **2009**, *55*, 5682–5699. [CrossRef]
19. Wang, Y.-C.; Hsieh, Y.-Y.; Tseng, Y.-C. Multiresolution Spatial and Temporal Coding in a Wireless Sensor Network for Long-Term Monitoring Applications. *IEEE Trans. Comput.* **2009**, *58*, 827–838. [CrossRef]
20. Kawahara, M.; Chiu, Y.-J.; Berger, T. High-speed software implementation of Huffman coding. In Proceedings of the DCC’98 Data Compression Conference (Cat. No. 98TH100225), Snowbird, UT, USA, 30 March–1 April 1998; p. 553.
21. Marcelloni, F.; Vecchio, M. A Simple Algorithm for Data Compression in Wireless Sensor Networks. *IEEE Commun. Lett.* **2008**, *12*, 411–413. [CrossRef]
22. Javed, M.Y.; Nadeem, A. Data compression through adaptive Huffman coding schemes. In Proceedings of the 2000 TENCON Proceedings Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119), Kuala Lumpur, Malaysia, 24–27 September 2000; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA.
23. Tharini, C.; Ranjan, P.V. Design of Modified Adaptive Huffman Data Compression Algorithm for Wireless Sensor Network. *J. Comput. Sci.* **2009**, *5*, 466–470. [CrossRef]
24. Huang, N.Y.; Lin, C.C.; Chuang, C.C.; Chang, R.I. Lossless compression for low correlation data in wireless sensor networks. In Proceedings of the Workshop on Wireless Ad Hoc and Sensor Network, Wisla, Poland, 18–20 October 2010.
25. Luo, H.; Liu, Y.; Das, S.K. Routing Correlated Data with Fusion Cost in Wireless Sensor Networks. *IEEE Trans. Mob. Comput.* 2006, 5, 1620–1632. [CrossRef]
26. Luo, H.; Luo, J.; Liu, Y.; Das, S.K. Adaptive Data Fusion for Energy Efficient Routing in Wireless Sensor Networks. *IEEE Trans. Comput.* 2006, 55, 1286–1299. [CrossRef]
27. Kumar, R.; Wolenetz, M.; Agarwalla, B. DFuse: A framework for distributed data fusion. In Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, Los Angeles, CA, USA, 5–7 November 2003; pp. 114–125.
28. Jin, G.-Y.; Park, M.-S. CAC: Context Adaptive Clustering for Efficient Data Aggregation in Wireless Sensor Networks; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2006; Volume 3976, pp. 1132–1137.
29. Lee, S.; Yoo, J.; Chung, T. Distance-based energy efficient clustering for wireless sensor networks. In Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks LCN-04, Washington, DC, USA, 16–18 November 2004; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA.
30. Yoon, S.; Shahabi, C. The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks. *ACM Trans. Sens. Netw.* 2007, 3, 3. [CrossRef]
31. Goel, S.; Imielinski, T. Prediction-based monitoring in sensor networks. *ACM SIGCOMM Comput. Commun. Rev.* 2001, 31, 82–98. [CrossRef]
32. Choi, K.; Kim, M.-H.; Chae, K.; Park, J.-I.; Joo, S.-S. An efficient data fusion and assurance mechanism using temporal and spatial correlations for home automation networks. *IEEE Trans. Consum. Electron.* 2009, 55, 1330–1336. [CrossRef]
33. Makarenko, A.; Whyte, H.D. Decentralized data fusion and control in active sensor networks. In Proceedings of the International Conference on Information Fusion, Stockholm, Sweden, 28 June–1 July 2004.
34. Yuan, W.; Krishnamurthy, S.V.; Tripathi, S.K. Synchronization of multiple levels of data fusion in wireless sensor networks. In Proceedings of the Global Telecommunications Conference, Dallas, TX, USA, 29 November–3 December 2004.
35. Sharaf, M.A.; Beaver, J.; Labrinidis, A.; Chrysanthis, P.K. TINA: A scheme for temporal coherency-aware in-network aggregation. In Proceedings of the 3rd ACM International Workshop on Data Engineering for Wireless and Mobile Access, San Diego, CA, USA, 19 September 2003; pp. 69–76.
36. Lin, C.-L.; Tsai, J.-H.; Chu, Y.-H.; Chang, R.-I. Concept of bounded error to improve wireless sensor network data compression. In Proceedings of the IEEE SENSORS 2014, Valencia, Spain, 2–5 November 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1240–1243.
37. Li, M.-H.; Lin, C.-C.; Chang, R.-I. Error-Bounded Data Compression Using Data, Temporal and Spatial Correlations in Wireless Sensor Networks. In Proceedings of the 2010 International Conference on Multimedia Information Networking and Security, Nanjing, China, 4–6 November 2010; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA; pp. 111–115.
38. Chen, Y.-H.; Huang, N.-Y.; Chu, Y.-H.; Li, M.-H.; Chang, R.-I.; Wang, C.-H. Dynamic bounded-error data compression and aggregation in wireless sensor network. In Proceedings of the 2012 IEEE Sensors, Taipei, Taiwan, 28–31 October 2012; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA; pp. 1–4.
39. Azar, J.; Makhoul, A.; Barhamgi, M.; Couturier, R. An energy efficient IoT data compression approach for edge machine learning. *Futur. Gener. Comput. Syst.* 2019, 96, 168–175. [CrossRef]
40. Agarwal, S.; Mozafari, B.; Panda, A.; Milner, H.; Madden, S.; Stoica, I. BlinkDB: Queries with bounded-errors and bounded response times on very large data. In *EuroSys’13: Proceedings of the 8th ACM European Conference on Computer Systems*, Prague, Czech Republic, 14–17 April 2013; ACM: New York, NY, USA, 2003; pp. 29–42.
41. Hellerstein, J.M.; Haas, P.J.; Wang, H.J. Online aggregation. *ACM SIGMOD Rec.* 1997, 26, 171–182. [CrossRef]
42. Di, S.; Cappello, F. Fast Error-Bounded Lossy HPC Data Compression with SZ. In Proceedings of the 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Chicago, IL, USA, 23–27 May 2016; pp. 730–739.
43. Taiwan Air Quality Monitoring Network. Available online: https://taqm.epa.gov.tw/taqm/en/ (accessed on 1 February 2019).
44. U.S. Geological Survey: Earthquake Hazards Program. Available online: http://earthquake.usgs.gov/ (accessed on 1 February 2019).
45. Liang, Y.; Peng, W. Minimizing energy consumptions in wireless sensor networks via two-modal transmission. *ACM SIGCOMM Comput. Commun. Rev.* 2010, 40, 12–18. [CrossRef]

46. Habib, A.; Moh, S. Game theory-based Routing for Wireless Sensor Networks: A Comparative Survey. *Appl. Sci.* 2019, 9, 2896. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).