

Regional Density-aware Data Collection Using Unmanned Aerial Vehicle in Large-scale Wireless Sensor Networks

Eui-Jik Kim,¹ Hyo Hyun Choi,^{2*} and Jung-Hyok Kwon^{1**}

¹Department of Convergence Software, Hallym University,
1 Hallymdachak-gil, Chuncheon, Gangwon 24252, South Korea

²Department of Computer Science, Inha Technical College,
100 Inha-ro, Nam-gu, Incheon 22212, South Korea

(Received March 30, 2017; accepted January 23, 2018)

Keywords: data collection, mobile sink, regional density, unmanned aerial vehicle, wireless sensor network

This paper presents a regional density-aware data collection (RDDC) using unmanned aerial vehicles (UAVs) in large-scale wireless sensor networks (WSNs). The RDDC is designed to support fast data collection by adaptively adjusting the value of the minimum contention window (CW_{min}) of the sensor nodes depending on the regional node density of sensing fields. The operation of RDDC consists of the region-partitioning, node-scanning, and data-collecting phases. In the region-partitioning phase, the RDDC partitions the sensing field into multiple regions and assigns a region ID to each sensor node according to the region to which it belongs. In the node-scanning phase, the RDDC calculates the node density of each region. Finally, in the data-collecting phase, the RDDC assigns different CW_{min} values to each sensor node considering the regional node density of each region. The experimental simulation results show that the RDDC obtains higher aggregate throughput and shorter delay than the existing data collection method.

1. Introduction

Large-scale wireless sensor networks (WSNs) have been widely used for a variety of intelligent services such as environmental monitoring, smart cities, connected cars, smart farms, battlefield surveillance, and so on.⁽¹⁾ Most intelligent services use the data collected from the deployed sensor nodes to monitor changes in the surrounding environment. Therefore, data collection has been regarded as a fundamental function of large-scale WSNs and as one of the most important research topics for decades.⁽²⁾

So far, a number of studies have been conducted on data collection in large-scale WSNs. Typically, the sensor nodes send the data to a static sink using multihop communication for data collection.^(3–5) However, this approach significantly degrades network performance because the multihop relay and route selection causes long delays. Furthermore, it leads to a short network lifetime because the sensor nodes close to the sink tend to consume more energy because of the

*Corresponding author: e-mail: hchoi@inhac.ac.kr

**Corresponding author: e-mail: jhkwon@hallym.ac.kr

<http://dx.doi.org/10.18494/SAM.2018.1865>

frequent data reception and forwarding. To address the above problems, many studies on data collection have explored the use of multiple sinks.^(6,7) The use of multiple sinks can decrease the delay by selecting the sink having the minimum number of hops from among the sinks and increase the network lifetime through load balancing among the sensor nodes. However, it requires an appropriate sink deployment strategy to avoid redundant sink coverage. For efficient data collection, mobile sinks have been widely used for data collection.^(8,9) Unlike the existing static sink, a mobile sink moves toward the deployed sensor nodes along a given path for data collection. This approach uses only one mobile sink and one-hop communication. Therefore, it is a cost-effective solution for improving network performance characteristics such as throughput, delay, and network lifetime. However, the mobile sink uses ground transportation (i.e., a vehicle) for mobility. Thus, the path of the mobile sink is heavily limited by traffic and geographical conditions. In other words, the mobile sink may need to move a long distance to visit the sensor nodes for data collection.

The use of an unmanned aerial vehicle (UAV) as a sink can overcome the limitations of the existing mobile sink, since the movement of a UAV is free from the traffic and geographical conditions. Therefore, in recent years, the use of UAVs has been actively studied in the field of data collection. Say *et al.* propose a priority-based data-gathering framework in UAV-assisted WSNs for faster and more reliable data collection, longer network lifetime, and real-time data transmission.⁽¹⁰⁾ To this end, the framework employs a priority-based contention window adjustment scheme (PCWAS) that adaptively assigns the minimum contention window (CW_{min}) to the deployed sensor nodes depending on the transmission priority. However, the PCWAS may suffer from frequent collisions since it assigns CW_{min} to sensor nodes without considering node density. Okcu and Soy Turk propose a received signal strength indicator (RSSI)-based hybrid and energy-efficient distributed clustering (rHEED) algorithm for UAV-integrated WSNs, which is designed to construct well-balanced clusters by considering both the RSSI values of the data received from the UAV and the energy levels of sensor nodes.⁽¹¹⁾ In this research, the UAV collects the data only from the cluster heads (CHs) that periodically receive the data from neighboring sensor nodes. This data collection can mitigate the collision problem; however, it may result in unpredictable delays since CH selection has high complexity and incurs considerable overheads.

In this paper, we propose a regional density-aware data collection (RDDC) using UAVs in large-scale WSNs, which provides fast data collection by adaptively adjusting the value of CW_{min} of the sensor nodes depending on the regional node density. The RDDC not only mitigates collisions among sensor nodes in dense regions but also reduces unnecessary waiting time for channel access in sparse regions. Consequently, it can achieve high throughput and low delay performance. The operation of the RDDC consists of three phases: 1) region partitioning, 2) node scanning, and 3) data collection. In the region-partitioning phase, the RDDC partitions the sensing field into multiple regions by calculating the data collection coverage, and then assigns a region ID to each sensor node according to the region to which it belongs. In the node scanning phase, the RDDC calculates the regional node density of each collection region. Finally, in the data collection phase, the RDDC assigns different CW_{min} values to each sensor node considering the node density of each region and notifies each value. Subsequently, the

UAV starts to collect data from the sensor nodes. To verify the superiority of the RDDC, we conduct an experimental simulation under various scenarios. The results show that, on average, the RDDC achieves a 9.3% lower delay and a 3.4% higher aggregate throughput than the existing UAV-based data collection method.

The rest of this paper is organized as follows. In Sect. 2, the design and operation of RDDC are described in detail. The experimental simulation and results are presented in Sect. 3. We conclude this paper in Sect. 4.

2. Design of RDDC

The RDDC adaptively adjusts the value of CW_{min} of the sensor nodes depending on the regional node density to quickly collect data in large-scale WSNs. The operation of the RDDC consists of region-partitioning, node-scanning, and data-collecting phases. The regional node density is a key parameter of the RDDC operation, which is determined by the number of sensor nodes in the partitioned region. In the following subsections, we describe the design and operation of the RDDC in detail.

2.1 Region-partitioning phase

In the region-partitioning phase, the RDDC divides the sensing field into multiple regions and calculates their areas. In the RDDC, a “region” is determined by the maximum area (i.e., data collection coverage) that a UAV can collect data from at a particular location; thus, it can be calculated using the altitude of the UAV (h) and the maximum transmission range of the sensor node (tr). To simplify the data collection coverage calculation, we assume that the UAV and sensor nodes use an omnidirectional antenna and a fixed transmission power. Figure 1 shows an example of the data collection coverage: the data collection coverage has a circular shape, so its size (c) is easily calculated using Eq. (1).

$$c = \pi r^2 = \pi((tr)^2 - h^2) \quad (1)$$

Figure 2 shows an example of region partitioning. In Fig. 2, the RDDC partitions the entire sensing field of the WSN into multiple regions based on the size of the data collection coverage. The grid division method is used to create multiple square regions of identical size.⁽¹²⁾ Each

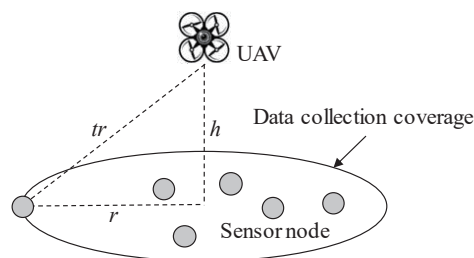


Fig. 1. Example of the data collection coverage.

region maintains a unique region ID, and each sensor node is assigned a region ID according to the region to which it belongs.

2.2 Node-scanning phase

In the node-scanning phase, the UAV calculates the regional node density of each region; it moves along the center of each region and receives a hello message from the sensor node. In Fig. 2, the bold line indicates the moving path of the UAV, which follows the order of the region ID values. The UAV counts the number of hello messages received from the sensor nodes with identical region IDs to calculate the regional node density of the corresponding region. Note that the regional node density of each region is proportional to the number of nodes that belong to the corresponding region because all regions have the same size.

The node-scanning phase apparently incurs unnecessary overhead during data collection. However, it is conducted only once at the beginning of the data collection process or when the deployment of the sensor nodes changes. Because the locations of the sensor nodes rarely change in large-scale WSNs, this scanning overhead can be ignored over long data collection periods.

2.3 Data-collecting phase

In the data-collecting phase, the UAV collects data from the sensor nodes using the same path as in the node-scanning phase. At the beginning of the data-collecting phase, the UAV assigns different CW_{min} values to each sensor node considering the regional node density of each region and records each value. The purpose of the CW_{min} adjustment is to mitigate collisions among the sensor nodes in dense regions and reduce unnecessary waiting time for channel access in sparse regions. Accordingly, the CW_{min} adjustment of the RDDC effectively

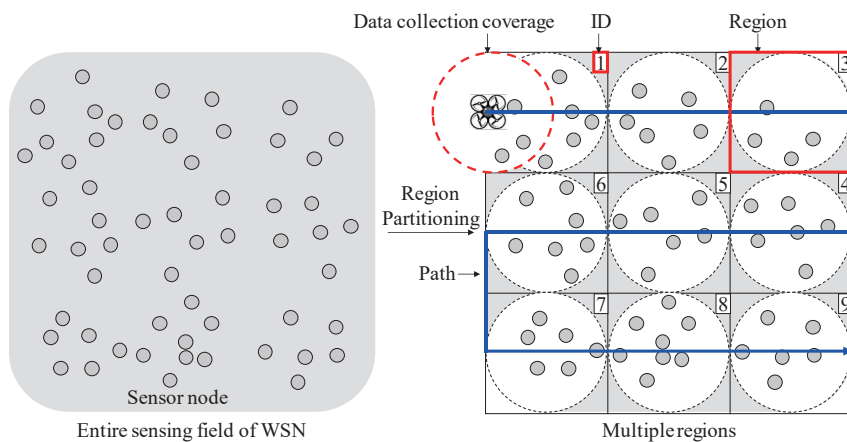


Fig. 2. (Color online) Operational example of region partitioning.

maximizes the aggregate throughput. We assume that each sensor node operates under saturated conditions, and that the communication between the UAV and the sensor nodes is performed with carrier sense multiple access with collision avoidance (CSMA/CA). We further assume that the UAV can collect data from the sensor nodes in all regions. Thus, the aggregate throughput in the i -th region (Th_i) is shown by ⁽¹³⁾

$$Th_i = \frac{n_i \tau_i (1 - \tau_i)^{n_i - 1} E[P]}{(1 - \tau_i)^{n_i} \sigma + n_i \tau_i (1 - \tau_i)^{n_i - 1} T_S + (\tau_i^{n_i} - n_i \tau_i (1 - \tau_i)^{n_i - 1}) T_C}, \quad (2)$$

where n_i is the number of sensor nodes in the i -th region, $E[P]$ is the expectation of a transmitted payload size, σ is the duration of an empty slot time, T_S is the average time of a successful transmission, T_C is the average time of a collision, and τ_i is the transmission probability at a random time slot in the i -th region. The transmission probability τ_i is expressed as

$$\tau_i = \frac{4(1 - \tau_i)^{n_i - 1} - 2}{(2(1 - \tau_i)^{n_i - 1} - 1)(W_i + 1) + (1 - (1 - \tau_i)^{n_i - 1})W_i(1 - (2 - 2(1 - \tau_i)^{n_i - 1})^m)}, \quad (3)$$

where W_i is the CW_{min} value of the i -th region and m is the maximum backoff stage. Using Eqs. (2) and (3), we can derive the function $Th_i(W_i)$, which defines the relationship between the aggregate throughput and CW_{min} . Then, we can obtain the adjusted CW_{min} ($W_{i,new}$) that maximizes the aggregate throughput of the i -th region using Eq. (4).

$$W_{i,new} = \arg \max_{W_i} (Th(W_i)), \quad W_i = 2^j - 1, \quad j \in [0, m] \quad (4)$$

3. Performance Evaluation

To evaluate the performance of the RDDC, we conducted an experimental simulation using MATLAB. The performance of the RDDC is compared with that of the existing data collection method to verify the superiority of the RDDC. In the following subsections, the simulation scenario, setting, and results are described in detail.

3.1 Simulation scenario and setting

We performed the simulation with various scenarios. To investigate the effect of the altitude of the UAV on the delay and aggregate throughput, the simulation was conducted with two different altitudes of UAV: 50 and 70 m. In addition, the number of sensor nodes was set to change from 100 to 1000 in each simulation. With this setting, we investigated the changes in the performance characteristics (i.e., delay and aggregate throughput) of the RDDC with the increase in the number of sensor nodes.

In the simulation, the sensor nodes are randomly deployed in the entire sensing field of $1000 \times 1000 \text{ m}^2$. The UAV and sensor nodes communicate with one another using the IEEE 802.11n medium access control (MAC)/physical (PHY) layers at 130 Mbps. The transmission range of sensor nodes is set to 100 m. Thus, the side length of the region is 173.2 m when the altitude of the UAV is 50 m and becomes 142.8 m when the altitude of the UAV changes to 70 m. The detailed simulation parameters are listed in Table 1.

3.2 Simulation results

In Figs. 3 and 4, the UAV altitude was set to 50 m. Figure 3 shows the variation of the delay with the increase in the number of sensor nodes. Overall, the RDDC achieves a shorter delay than the existing data collection method because the RDDC adaptively adjusts the CW_{min} value of the sensor nodes depending on the regional node density of each region, whereas the existing data collection method uses a fixed value of CW_{min} (i.e., 31). In particular, when the number of sensor nodes is 100, the delay of RDDC is 21.7% shorter than that of the existing data collection

Table 1
Simulation parameters.

Parameter	Value	Parameter	Value
MAC/PHY model	IEEE 802.11n	PHY header	34 B
Data rate	130 Mbps	MAC header	16 B
Size of sensing field	$1000 \times 1000 \text{ m}^2$	Payload	5120 B
Number of sensor nodes	100–1000	CW_{min}	0–1023
Transmission range	100 m	CW_{max}	1023
Altitude of UAV	50 m, 70 m	SIFS ¹⁾	10 μs
Slot time	20 μs	DIFS ²⁾	50 μs

¹⁾Short interframe space

²⁾Distributed coordination function interframe space

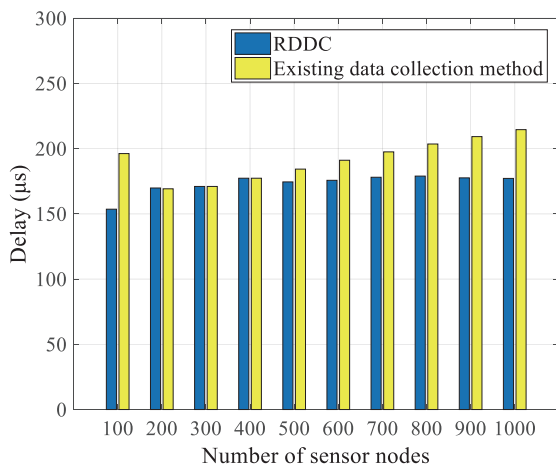


Fig. 3. (Color online) Delay (UAV altitude = 50 m).

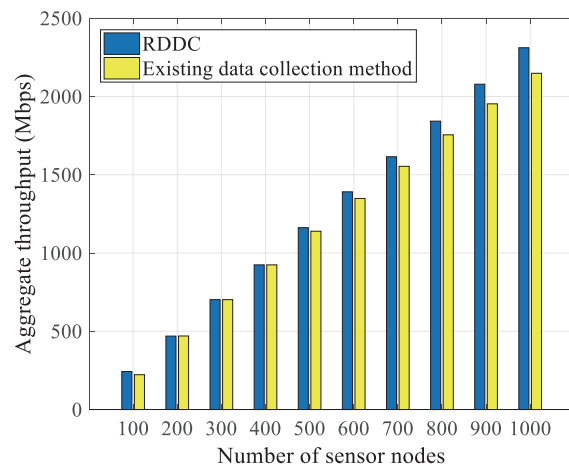


Fig. 4. (Color online) Aggregate throughput (UAV altitude = 50 m).

method because a small number of sensor nodes are placed in the region. In other words, the fixed value of CW_{min} in the existing data collection results in a long waiting time for channel access. The RDDC can reduce the collisions among the sensor nodes by increasing the value of CW_{min} when the number of sensor nodes increases. Therefore, the delay of the RDDC does not significantly change when the number of sensor nodes exceeds 500, whereas the delay of the existing data collection method lengthens because the number of collisions among the sensor nodes increases.

Figure 4 shows the variation of the aggregate throughput when the number of sensor nodes increases. In Fig. 4, the aggregate throughput increases when the number of sensor nodes increases, since the UAV receives data from a larger number of sensor nodes. The RDDC can reduce the waiting time for channel access in a sparse region and mitigate the collisions among the sensor nodes in a dense region. Therefore, the RDDC obtains a higher aggregate throughput than the existing data collection method. In this simulation, the RDDC achieves a 4.3% higher aggregate throughput than the existing data collection method on average. In particular, when the number of sensor nodes is 1000, the RDDC achieves a 7.6% higher aggregate throughput than the existing data collection method because the RDDC significantly reduces the number of collisions by assigning a larger value of CW_{min} to the sensor nodes.

Figures 5 and 6 show the delay and aggregate throughput when the altitude of the UAV is 70 m. The high altitude of the UAV causes the small size of the region. Therefore, the number of regions in the sensing field increases, and the number of sensor nodes in each region decreases. Thus, at a UAV altitude of 70 m, the RDDC achieves a 2.2% shorter delay and a 35.4% lower aggregate throughput than at a UAV altitude of 50 m. Moreover, in Figs. 5 and 6, the RDDC achieves a 9.2% shorter delay and a 2.6% higher aggregate throughput than the existing data collection method because of the CW_{min} adjustment.

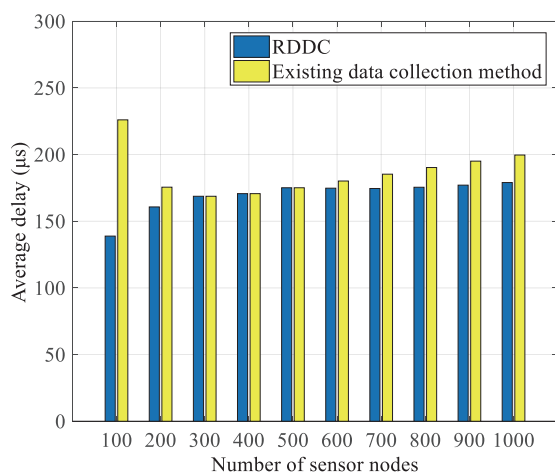


Fig. 5. (Color online) Delay (UAV altitude = 70 m).

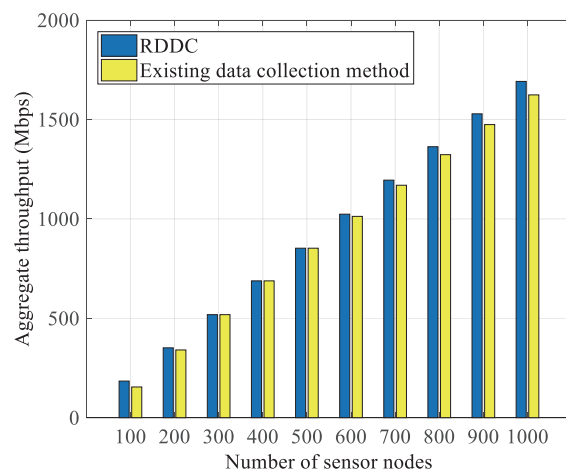


Fig. 6. (Color online) Aggregate throughput (UAV altitude = 70 m).

4. Conclusions

In this paper, we propose a RDDC using UAVs in large-scale WSNs. To support fast data collection by the UAV, the RDDC adaptively adjusts the value of CW_{min} of the sensor nodes according to the regional node density. The operation of the RDDC consists of three phases. In the region-partitioning phase, the RDDC divides the sensing field into multiple regions and assigns a region ID to each sensor node. In the node-scanning phase, the RDDC calculates the regional node density of each region. Finally, in the data-collecting phase, the RDDC assigns different CW_{min} values to each sensor node considering the regional node density of each region. To evaluate the performance of the RDDC, an experimental simulation was conducted using a sensing field of $1000 \times 1000 \text{ m}^2$. The results show that the RDDC achieves a 9.4% shorter delay and a 4.3% higher aggregate throughput than the existing data collection method when the altitude of the UAV is 50 m. When the altitude of the UAV is 70 m, the delay and aggregate throughput of RDDC are improved by 9.2 and 2.6%, respectively.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01061155). This research was also supported by Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2017R1D1A1B03031055).

References

- 1 N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han: IEEE Commun. Surv. Tutorials **18** (2016) 2546.
- 2 A. W. Khan, A. H. Abdullah, M. H. Anisi, and J. I. Bangash: Sensors **14** (2014) 2510.
- 3 M. Dong, K. Ota, and A. Liu: IEEE Internet Things J. **3** (2016) 511.
- 4 J. Yin, Y. Yang, and L. Wang: Sensors **16** (2016) 462.
- 5 T. Meng, F. Wu, Z. Yang, G. Chen, and A. V. Vasilakos: IEEE Trans. Comput. **65** (2016) 244.
- 6 S. Chen, M. Coolbeth, H. Dinh, Y. A. Kim, and B. Wang: Proc. WASA: Int. Conf. Wireless Algorithms, Systems, and Applications (Springer, 2009) 284.
- 7 Y. K. Sia, H. G. Goh, S. Y. Liew, and M. L. Gan: Proc. 12th IEEE Int. Conf. Fuzzy Systems and Knowledge Discovery (FSKD) (IEEE, 2015) 15.
- 8 S. Yang, U. Adeel, Y. Tahir, and J. A. McCann: IEEE Trans. Mob. Comput. **16** (2017) 1420.
- 9 C. Wang, S. Guo, and Y. Yang: IEEE Trans. Mob. Comput. **15** (2016) 2969.
- 10 S. Say, H. Inata, J. Liu, and S. Shimamoto: IEEE Sens. J. **16** (2016) 5785.
- 11 H. Okcu and M. Soyuturk: Int. J. Ad Hoc Ubiquitous Comput. **15** (2014) 106.
- 12 C. Wang, F. Ma, J. Yan, D. De, and S. K. Das: Int. J. Distrib. Sens. Networks **11** (2015) 1.
- 13 G. Bianchi: IEEE J. Sel. Areas Commun. **18** (2000) 535.