Annals of Emerging Technologies in Computing (AETiC)
Vol. 6, No. 3, 2022

Research Article

A Novel Data Aggregation Mechanism using Reinforcement Learning for Cluster Heads in Wireless Multimedia Sensor Networks

Jia Uddin

Endicott College, Woosong University, South Korea
jia.uddin@wsu.ac.kr

Received: 27th March 2022; Accepted: 25th June 2022; Published: 1st July 2022

Abstract: Wireless multimedia sensor networks (WMSNs) are getting used in numerous applications nowadays. Many robust energy-efficient routing protocols have been proposed to handle multimedia traffic-intensive data like images and videos in WMSNs. It is a common trend in the literature to facilitate a WMSN with numerous sinks allowing cluster heads (CHs) to distribute the collected data to the adjacent sink node for delivery overhead mitigation. Using multiple sink nodes can be expensive and may incur high complexity in routing. There are many single-sink cluster-based routing protocols for WMSNs that lack in introducing optimal path selection among CHs. As a result, they suffer from transmission and queueing delay due to high data volume. To address these two conflicting issues, we propose a data aggregation mechanism based on reinforcement learning (RL) for CHs (RL-CH) in WMSN. The proposed method can be integrated to any of the cluster-based routing protocol for intelligent data transmission to sink node via cooperative CHs. Proposed RL-CH protocol performs better in terms of energy-efficiency, end-to-end delay, packet delivery ratio, and network lifetime. It gains 17.6% decrease in average end-to-end delay and 7.7% increase in PDR along with a network lifetime increased to 3.2% compared to the evolutionary game-based routing protocol which has been used as baseline.

Keywords: Cluster head; Routing protocols; Reinforcement learning; Wireless multimedia sensor networks.

1. Introduction

A wireless multimedia sensor network (WMSN) is a network that connects wireless sensor nodes fortified with multimedia devices able to retrieve audio, image and video along with the scalar data [1]. Some of the examples that incorporate the use of WMSNs are traffic management systems, border monitoring for critical military applications, controlling and monitoring industrial processes, etc. [1]. For WMSNs, a higher level of bandwidth is needed in comparison to the legacy wireless sensor networks (WSNs). Extensive processing is required due the high volume of data transmission and reception. This vastly increases energy consumption putting a negative impact on network lifetime. Again, multimedia sensor nodes are also different from the traditional ones in terms of the sensing region too. To acquire images/videos within a certain region, they have a field of views (FOVs). FOV is a directional view of the camera in a sensor node.

Cluster-based protocols are popular for WSNs where nodes are separated into clusters. A single cluster is simplified by a cluster head (CH) that gathers data from the other cluster members (CMs) within the cluster and conveys the data to the sink node. Usually, these protocols have a single sink node [2-4]. Typically, they lack a mechanism for data aggregation towards sink nodes from the CHs. Some of the cluster-based protocols in the literature [5-7] deploy multiple sink nodes too for increasing network performance. They can significantly improve some crucial performance metrics like packet delivery ratio, end-to-end delay, and network lifetime compared to protocols that adopt single sink nodes. To avoid failure in the sink node, we need to often establish multiple sink nodes in the network so that each single sensor node has double paths. Also, congestion is likely to happen in the sink node when many nodes are trying

Jia Uddin, "A Novel Data Aggregation Mechanism using Reinforcement Learning for Cluster Heads in Wireless Multimedia Sensor Networks", Annals of Emerging Technologies in Computing (AETiC), Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 69-78, Vol. 6, No. 3, 1st July 2022, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2022.03.006, Available: http://aetic.theiaer.org/archive/v6/v6n3/p6.html.

to connect to the one and only sink node simultaneously. A WMSN with single sink node supporting multi-hop data transmission and reception would often suffer due to energy hole problem which occurs when nodes closer to the sink node have a very fast decrease in the energy level since it has to perform most of the traffic aggregation because of its position. This can cause significant reduction of network lifetime. Therefore, to have better values in the key performance indicators (PDR, NL, etc.), multiple sink strategies are often proposed in the literature. This significant performance increase in the network due to multisink architecture incurs some crucial challenges to deal with. First, it is expensive to deploy multiple sink nodes and we need an automation algorithm for the placement of the sink nodes that adds more complexity. Moreover, there should be a mechanism in the protocol to choose the number of the sink nodes. To solve this contradicting issue of using single or multiple sink nodes for performance increase, in this study, we propose a single sink-oriented reinforcement learning (RL) based data aggregation mechanism among CHs (RL-CH) for WMSN that is built upon [7].

Over the past few years, intelligent techniques like RL and game theory have been widely used for designing routing protocols for WSNs [8, 16]. However, these techniques are not yet well explored for WMSNs. In the RL, an agent learns what to do through steps by interacting with the environment. The agent yields reward while this learning process goes on. The primary goal of an agent is to discover an optimum policy by learning the environment through reward maximization. As mentioned earlier, RL-CH is built upon [7], which is an evolutionary game-based routing (EGR) protocol for WMSNs with several sink nodes. EGR has an intelligent mechanism to optimally select CHs using evolutionary games. Based on the FoVs, clusters are formed. It imposes a data redundancy avoidance algorithm too. The overlapped area shared by two closely placed sensor nodes is calculated and based on the calculated area it is concluded that if the captured data is redundant. The approach significantly reduces the number of redundant transmissions. In this study, we have replaced the multiple sink nodes in EGR with a single sink node and introduced a data aggregation method for CHs in the network based on RL. It outpaces EGR in terms of energy efficiency and network lifetime. RL-CH is especially suitable for networks that deploy thousands of sensor nodes for any kind of object tracking or surveillance networks.

The main contributions of this research can be summarized as follows:

- The inimitable and intelligent way of data aggregation method using RL has been proposed that can increase network lifetime and energy efficiency.
- The proposed scheme can be integrated to any other state-of-the-art clustering algorithm for introducing cooperation among CHs.

The proposed scheme can be used to facilitate single-sink architecture for WMSNs reducing the complexity of the network by alleviating the necessity of multiple sink nodes.

Rest of the paper is organized as follows: in Section 2, the related works are summarized and discussed in brief. In Section 3, the proposed RL-CH data aggregation method is presented in detail. In Section 4, the performance of the proposed RL-CH is evaluated via extensive simulation and compared with EGR protocol. Finally, the conclusion is in Section 5.

2. Related Works

A cluster-based protocol deployed with single a sink node for mobile target tracking application was proposed in [3]. It incorporated a nonlinear localization-oriented sensing model for WMSNs. The camera sensors were used to consider perspective projection and observation noises through dynamic collaboration.

An ant-based service-aware routing algorithm for WMSNs was proposed in [9]. This routing protocol is primarily based on the transmission between CHs and sink nodes. The CHs handover different classes of data by choosing appropriate paths for different QoS requirements. To utilize the network resources accurately, protocol proposed in [9] introduces a singularity value for the sink node to decrease transmission frequency and the number of control messages.

QoS aware multisink opportunistic routing (QMOR) was presented in [5] having a data redundancy avoidance algorithm to increase the network lifetime. The goal is to select and prioritize a forwarder list to achieve energy effective delivery of multimedia data under certain QoS requirements. QMOR has a multisink architecture that has been integrated with opportunistic routing. Even though QMOR shows performance rise in terms of energy efficiency, delay, and reliability, it suffers from overhearing in the network.

Alae and Barcelo-Ordinas proposed a cluster-based protocol for WMSNs that meets the FoV constraints [11]. Clusters were formed based on the overlapped FoVs of the two closely placed sensor nodes. If the overlapped area is wide, two sensors are in the same cluster. Energy is well conserved due to the avoidance of the redundant transmissions in the network. The proposed scheme does not associate its performance along with the other state-of-the-art protocols. Also, it does not consider crucial parameters like packet delivery ratio and end-to-end delay.

Recently, an evolutionary game theory-based routing protocol has been proposed in [7]. It has a cluster-based architecture with multiple sink nodes. Clusters are organized based on the overlapped FoVs of two closely placed sensor nodes. A data redundancy avoidance algorithm is also proposed so that the same copy of the data is not sent twice. The main contribution of the paper is implementing game theory to select the CHs intelligently using evolutionary games.

Nodes with higher residual energy were mostly selected as CHs in different rounds. It showed performance comparison with many of the state-of-the-art protocols and showed significant performance increase in terms of energy consumption, network lifetime, cluster formation time, packet delivery ratio, and average end-to-end delay.

Ding *et al.* provided a comprehensive survey and proposed a model based on machine learning to create an energy efficient green computing model [19]. An RL-based energy efficient routing protocol for WMSN was proposed in [20] that in particular used state-action-reward-state-action algorithm to achieve convergence. A cluster-based routing protocol was proposed in [21] that used a combination of metaheuristics and machine learning. It significantly improved network lifetime. Keerthika *et al.* proposed an RL-based optimized routing protocol that dealt with the energy hole problem and achieved significant improvement in energy performance [22]. A data aggregation sensitive algorithm for energy efficient routing based on a simple Q-learning algorithm was proposed in [23] that is not only data efficient but also suitable for longer network lifetime. A routing protocol adaptive to mobility and energy level was proposed in [24] that used reinforcement learning for such adaptation. Adhoc on-demand distance vector routing protocol centric routing protocol having Q-learning-based CH selection was proposed in [25]. The proposed protocol obtained better performance in terms of end-to-end delay, energy consumption and network lifetime.

The mentioned works in this section are either based on multiple sinks for multimedia sensor networks with no data aggregation mechanism among the CHs or single sink protocol for typical sensor networks not having capability of carrying multimedia data. Therefore, we intend to investigate how a data aggregation mechanism based on RL works against the multisink protocols.

3. RL-CH protocol

The goal of the proposed RL-CH data aggregation mechanism is to facilitate the cluster-based protocols for WMSN with an optimal method for data transmission from the CHs in a single sink architecture. If a CH is very far away from the sink node, a long-distance single hop transmission from the CH to the sink node may result in significant energy consumption. Also, we do not know the shortest and the most convenient route among the CHs to the sink node. RL algorithms are smart choices [12-14] for this kind of scenario for optimal route finding. In this section, we first discuss the network architecture of EGR. After that, we put an effort to elaborate the implication of RL-CH which is built upon EGR. EGR has recently been published that has put tight efforts to use game theory for the first time to elect CHs optimally in a multi-sink architecture claiming to outsmart many of the state-of-the-art protocols in terms of crucial performance metrics like end-to-end delay, cluster formation time, packet delivery ratio, network lifetime, and energy consumption.

3.1. EGR Network Architecture

EGR has a network model consisting of 100 multimedia sensor nodes that are deployed randomly. As a clustering protocol, it has three kinds of nodes. They are CH, CM, and sink nodes. CHs have the responsibility to communicate with the sink node. They gather data from the members in the cluster. In EGR, CHs do not participate in data collection while acting as CHs. CMs collect from the network and transfer to the CHs. They cannot send data to the sink node directly. They have to depend upon CHs. CMs strive for the FoV of the object and based on the remaining energy, one is selected as a winner. Figure 1 shows the network architecture of EGR. The red dots are the CMs and the small black circles represent the CHs. The quadrilateral in Figure 1 represents the sink node in the network. We can see all the CMs transmitting to CHs and some of the CHs transmitting to the sink node.

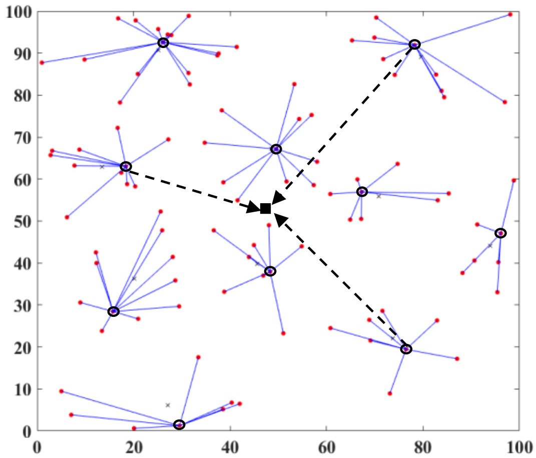

Figure 1. Network architecture for EGR

Sensor nodes form clusters in EGR based on the overlapped FoVs. One node acts as CH and others as members in each cluster. The process of EGR is divided into two phases. They are: "setup" phase, and "steady-state" phase. In the setup phase, clusters are designed, CHs are elected judiciously using evolutionary games, CHs advertise to CMs in the cluster, and CMs send join requests to the CHs. In the steady state phase, sink node broadcasts to the CHs, CMs use data redundancy avoidance algorithm to transmit data to the CHs. CHs finally transmit data to the sink node. Readers can prompt [7] for further mathematical details about the protocol.

When the nodes get divided into clusters based on the FoV constraints in EGR, nodes in each cluster must be divided into two distinct classes. One is nodes with high energy levels, and the another is nodes with low energy levels. Nodes in different clusters must send an initial message to the sink node. The message holds the node ID and residual energy of a node. Based on the residual energy level, the sink node fixes a threshold energy value (median). This energy threshold is adopted to separate the nodes in distinct energy classes. Finally, through game theoretic computation, the nodes either decide to be CHs or CMs.

3.2. RL-CH Data Aggregation Mechanism

EGR has a hierarchical network architecture featuring clustering among the sensor nodes where there are CHs and CMs. Since our interest lies in data aggregation of the CHs, we consider the nodes acting as CHs a graph, $G = (V, E)$, where each CH node is a vertex (V_i) and each edge (e_{ij}) is a bidirectional communication channel considering two nodes for example V_i and V_j.

RL has been used to solve many routing problems over the past few years. For RL-CH, we have used Q-learning [8, 17, 18] to accommodate multiple CHs in the model. Q-learning is a model free RL algorithm where an agent interacts with the environment and learns the optimal policy by yielding some scalar rewards. Q-values to each possible agent action is assigned that represents how good an action is. When an agent selects and performs one action, it gets a reward based on how good the action was. This reward is utilized to update the Q-value. Over some episodic learning, the agent gets to learn real action values.

3.2.1. Agent states

In our proposed Q-learning based routing model, the state of an agent incorporates all the CHs including the destination sink node. Let us assume that $N_{CH,D}$ represents the CHs and the destination sink node together. We can define the state as a tuple as a tuple {$N_{CH,D}$, $h_D^{N_{CH}}$}, where $N_{CH,D}$ represents a CH node including the goal state sink node, and $h_D^{N_{CH}}$ is the routing information through all the neighboring nodes N_{CH} to the destination sink node, *D*.

3.2.2. Actions

An action for our RL model is the possible routing decision for a data packet. That is, the decision of forwarding the data packet to the best possible neighboring CH. We define the action set, $A = \{A_1, A_2, \dots, A_n\}$. Each action, A_i consists of two sub elements/actions, $\{N_{CH_i}, D\}$. Here, N_{CH_i} is the intended next hop for the destination *D*.

3.2.3. Q-Values

We mentioned earlier that Q-values represent how good an action is. For our model, it presents an approximate cost of the route. The challenge pops out while deciding the initial Q-values. It is common to randomly initialize the values, but we have used a more sophisticated way. Hop count from the source to goal state has been used. To calculate the value of a sub action, following equation is used:

$$Q(A_i) = \sum hops_D^{CH_i} \tag{1}$$

where $hops_D^{CH_i}$ are the number of hops to reach the destination sink node. This estimation of the values is the upper bound of the actual value. It is assumed that no links were shared after the next hop. Therefore, it is normal that the Q-values will decrease during learning period and the best possible actions will be denoted by smaller Q-values. The Q-value of a complete action *A*, with sub actions $A_1, A_2, \dots, A_n$, can be calculated as:

$$Q(A_i) = [\sum_{i=1}^{n} Q(A_i)] - (n - 1) \tag{2}$$

where, *n* is the number of sub actions. Based on the action taken, an agent receives a reward from the environment. For our model, each CH to which a data packet is transmitted, sends the reward as response for evaluating the goodness of the sub action. Q-value is updated using the following equation:

$$Q_N(A_i) = Q_P(A_i) + \alpha(R(A_i) - Q_P(A_i)) \tag{3}$$

where $Q_N(A_i)$ is the updated Q-value, $Q_P(A_i)$ is the old Q-value, R is the reward function, and α is the learning rate.

3.2.4. Reward function

In our reward function, the node picks its lowest Q-value for the destination and adds the cost of the action. The cost of the action is '1' because every time there is one hop transmission in the network when the data is getting forwarded in steps. Also, each node piggybacks the residual energy level of the previously traversed node in every round. If the energy level is greater than a certain threshold it is subtracted or added otherwise in the reward function. It is formulated as follows:

$$R(A_i) = H_a + \min Q(A) \pm R_E \quad (4)$$

where H_a is the action cost, and R_E is the residual energy level. Many other crucial parameters like received signal strength, link quality, etc. can be added to the reward function depending on the necessity and different types of case scenario.

A learning agent must choose an action from the available ones in the list. It is assumed that neighboring nodes can overhear when a data packet is transmitted. The broadcast model of RL-CH permits to piggyback information for the neighboring nodes. An example of such event is providing reward values based on the action taken. The reward value is calculated using the reward function presented in equation (4). Figure 2 presents an RL-based model of the CHs based on EGR protocol There are altogether 10 CHs, each marked with different alphabets. The sink node is named as *goal state* and the source node is $A(CH_1)$. From A, there can be many different routes to the sink node. They have been addressed as different colors in the diagram. As the data packet will traverse the network, it will learn about its optimal route in terms of hop count and residual energy based on the reward function. For the following figure (Figure 2), at the initial stage, *Route 3* will be the best route since the energy level will be same at the beginning. As the simulation time increases, it may change based on the remaining energy of the nodes.

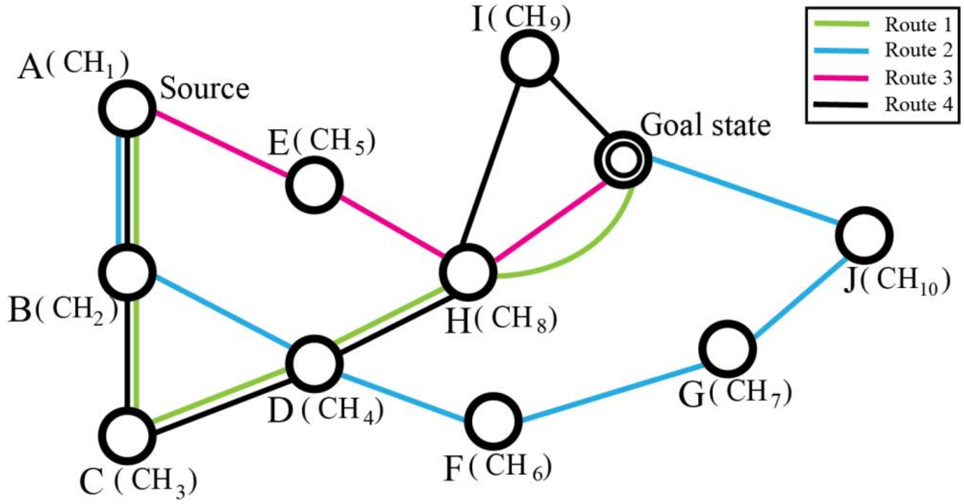

Figure 2. Network diagram of the CHs considering node A as source

In our RL algorithm, we want to maintain a balance between the exploration and exploitation strategies. This epsilon greedy policy is usually used to balance the exploration and exploitation. A random number between 0 to 1 is generated based on which we select a threshold. If the generated random number is larger than the threshold (for epsilon greedy process it is referred as the term, ε). We define every transmission time interval (TTI) as an episode in our designed RL system. The data aggregation algorithm is as follows:

Algorithm 1 Q-learning-based traffic steering

1. **Initialize:** Network parameters
2. **for** *TTI* = 1 to *T* **do**
3. **for** every V_i, Vj, and e_{ij} **do**
4. **if** ($rand \leq \varepsilon$)
5. choose action randomly
6. **else**
7. select A_i using greedy policy
8. Nodes are selected for the data packets for the next hop
9. Reward calculation based on eqn. (4)
10. Updating state
11. Updating Q-values by eqn. (3)
12. **end for**
13. **end for**

4. Performance Evaluation

In this section of the paper, the performance of the proposed RL-CH data aggregation mechanism built upon EGR is evaluated via simulation. Obtained results for RL-CH were compared with EGR protocol having no data aggregation scheme. The experimental environment for WMSN built using WISE-Mnet++ [15] has been used to conduct simulation.

4.1. Simulation Environment

The quality of any routing protocol depends on some crucial performance metrics. We have considered two performance metrics for this evaluation. One is average residual energy, and another is network lifetime. They are defined as follows:

- Average residual energy is the residual of a node after a particular amount of simulation time.
- Time when at least half of the nodes are alive is network lifetime.

We have deployed 100 sensor nodes randomly in a 100m×100m network space. There is a moving target in the environment. Nodes compete for FoVs of the object and images are captured by a winner. Sensing range of the node is 30m. We have simulation parameters that are identical to EGR. Simulation environment for RL-CH is displayed by Table 1 [7]. In Table 2, workload parameters of the simulation are summarized [7].

Table 1. Simulation parameters

Parameter	Magnitude
Area if the network	100 × 100 m
Sensor nodes in the area	100
Data rate	2 Mbps
Size of the image	176 × 144
Rate (frame)	30 fps
Offset angle	60°
Radius for sensing	30 m
Energy level when simulation started	2 J

Table 2. Workload parameters

Parameter	Value
Traffic model	Poisson distribution (48 packets/60 seconds)
Packet size	22 bytes
Rate of transmission	2 Mbps
Range for transmission	15 m

4.2. Simulation Results and Discussions

Figure 3 illustrates the reduction of the average residual energy in RL-CH embedded EGR and EGR. EGR has higher energy efficiency compared to many other existing protocols because of the data redundancy avoidance algorithm and best CH election process that uses game theory. Using RL-CH in EGR makes it more energy efficient because the number of transmissions is reduced. RL-CH allows CHs to add their data each time they participate in routing being a part of the optimized route. As mentioned in our proposed protocol, we consider the energy level of the nodes in our state space. Highest rewards are obtained whenever the data is aggregated via the nodes with more energy. Therefore, the system learns the optimal nodes via policies and energy is conserved. From Figure 3, we can see that EGR with RL-CH has depleted energy slower than typical EGR protocol.

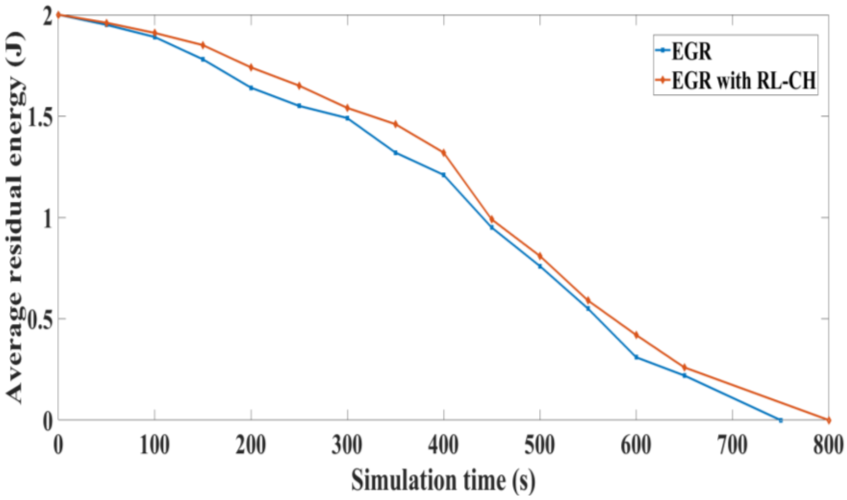

Figure 3. Average residual energy.

Network lifetime is also another crucial performance metric for any routing protocol. Since WMSNs deal with huge chunks of data, due the high data volume in the network is necessary to maintain a fair amount of network lifetime. RL-CH with EGR protocol performs better than typical EGR. Figure 4 presents the network lifetime graph for both protocols. We can see from the figure that EGR with RL-CH has a higher network lifetime. There are two reasons behind such performance increase. As we could see from Figure 3 that we have a performance increase in terms of average residual energy, it eventually leads to longer network lifetime since these two parameters are highly related. Other than this, aggregating CHs are chosen based on the energy level. When nodes closer to the sink node suffer from energy decline due to the energy hole problem, they are not chosen for data forwarding based on our algorithm.

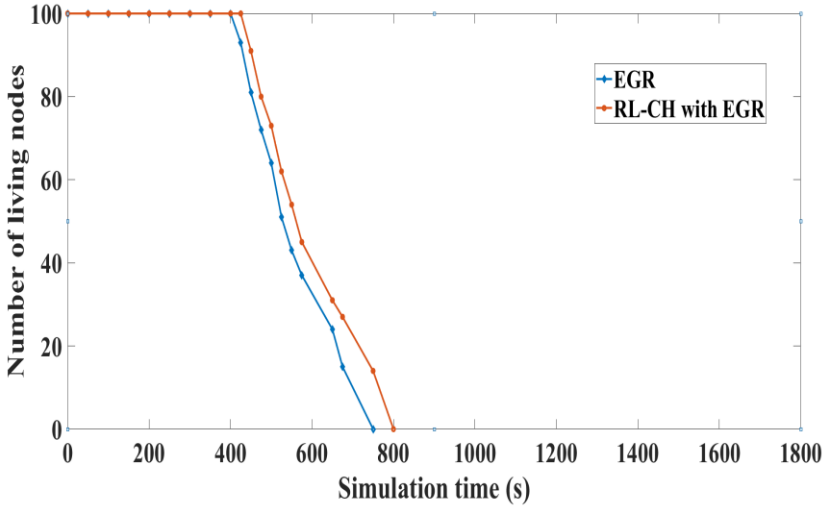

Figure 4. Number of living nodes showing network lifetime.

The next parameter that we considered for proving effectiveness of the proposed routing mechanism is packet delivery ratio. It is the ratio of the packets efficiently delivered and packets that were dropped. From Figure 5, we can see that the proposed RL-CH with EGR performs better than the typical EGR multi sink protocol. Since data is processed by the nodes with higher energy in the proposed method, the packets are more unlikely to suffer from dropping compared to the EGR baseline.

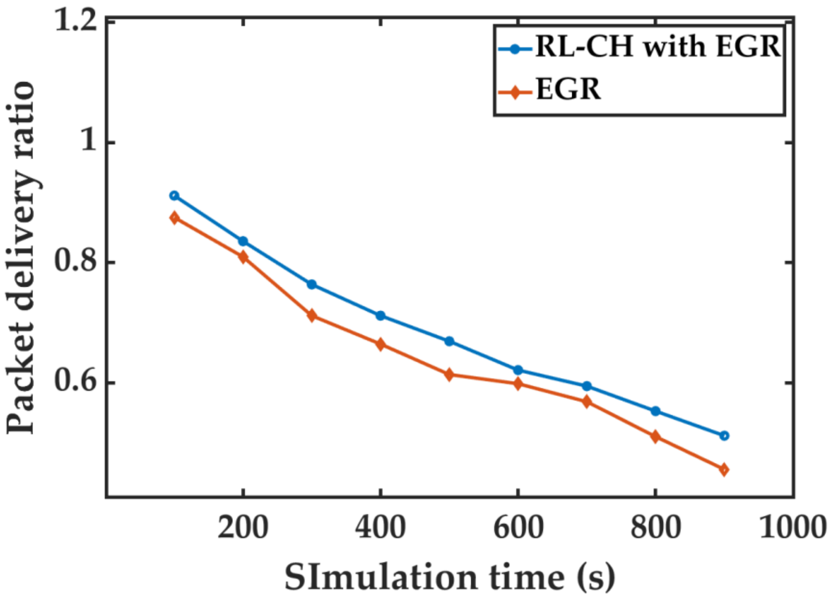

Figure 5. Comparison of the packet delivery ratio.

Lastly, the proposed RL-based protocol is compared considering the end-to-end delay too. From Figure 6, we can vividly visualize that when we introduced RL with single sink architecture, the performance was better for delay parameters.

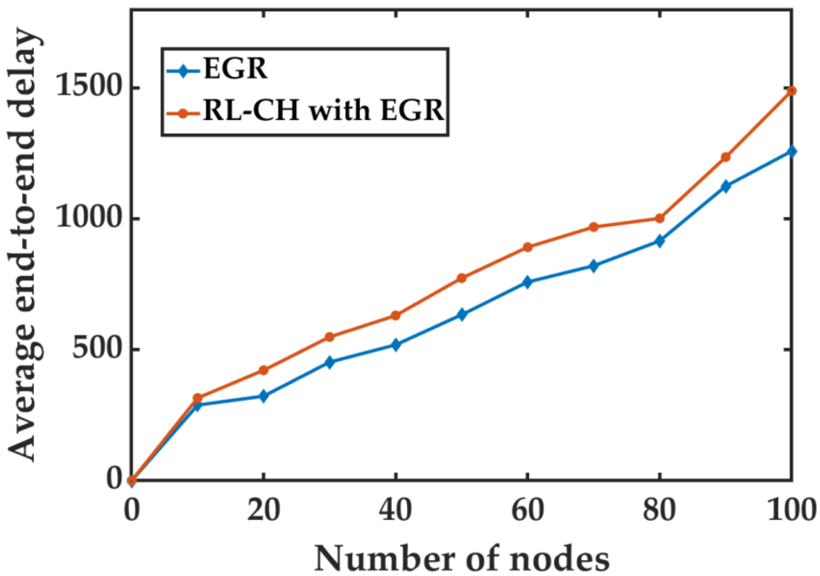

Figure 6. Comparison of the average end-to-end delay

5. Conclusion

WMSNs are very traffic intensive networks where the data volume is very high. Using multiple sinks can be costly and can increase the complexity of the routing algorithm. In this study, we have adopted EGR, a recently proposed multi-sink cluster-based routing protocol that claims to be more energy efficient compared to many of the existing state-of-the-art protocols to implement an intelligent data aggregation method among CHs within a single sink architecture. RL-CH learns the optimal policy by traversing the whole network using episodic learning procedure through state-action-reward chronology. Proposed RL-CH data aggregation scheme can outsmart EGR in terms of average residual energy, network lifetime, average end-to-end delay, and packet delivery ratio. It gains 17.6% decrease in average end-to-end delay and 7.7% increase in PDR along with a network lifetime increased to 3.2% compared to EGR. RL-CH can be implemented to improvise any other cluster-based routing protocols for WMSNs for better performance. In the future, we can consider advanced RL algorithms like deep Q-network (DQN) or hierarchical RL. The considered state space proposed in this paper is limited to energy level only. Increasing the state space with other different network parameters can make the aggregation strategy more efficient where DQN can be a good strategy.

Acknowledgement

This research is funded by Woosong University Academic Research in 2022.

References

[1] Islam T. Almalkawi, Manel Guerrero Zapata, Jamal N. Al-Karaki and Julain Morillo-Pozo, "Wireless Multimedia Sensor Networks: Current Trends and Future Directions", *Sensors*, Vol. 10, No. 7, pp. 6662-6717, 2010, DOI: 10.3390/s100706662.

[2] Heewook Shin, Sangman Moh and Ilyong Chung, "Clustering with One-Time Setup for Reduced Energy Consumption and Prolonged Lifetime in Wireless Sensor Networks", *International Journal of Distributed Sensor Networks*, Vol. 9, No. 2, p. 301869, 2013, DOI: 10.1155/2013/301869.

[3] Liang Liu, Xi Zhang and Huandong Ma, "Dynamic Node Collaboration for Mobile Target Tracking in Wireless Camera Sensor Networks", in the Proceedingsa of *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, Rio de Janeiro, Brazil, 19–25 April, 2009, DOI: 10.1109/infcom.2009.5062032.

[4] Wendi B. Heinzelman, Anantha P. Chandrakasan and Hari Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 660-670, 2002, DOI: 10.1109/twc.2002.804190.

[5] Hang Shen, Guangwei Bai, Zhenmin Tang and Lu Zhao, "QMOR: QoS-Aware Multi-sink Opportunistic Routing for Wireless Multimedia Sensor Networks", *Wireless Personal Communications*, Vol. 75, No. 2, pp. 1307-1330, 2013, DOI: 10.1007/s11277-013-1425-0.

[6] Md. Razzaque, Muhammad Mahbub Alam, Md Mamun-Or-Rashid and Choong Seon Hong, "Multi-Constrained QoS Geographic Routing for Heterogeneous Traffic in Sensor Networks", in *Proceeding of the 5th IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, USA, 3–6 January, 2008, DOI: 10.1109/ccnc08.2007.42.

[7] Md Arafat Habib and Ssangman Moh, "Robust Evolutionary-Game-Based Routing for Wireless Multimedia Sensor Networks", *Sensors*, Vol. 19, No. 16, p. 3544, 2019, DOI: 10.3390/s19163544.

[8] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, UK: MIT Press, 1998, Available: https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf.

[9] Yan Sun, Huadong Ma, Liang Liu and Yu'e Zheng, "ASAR: An ant-based service-aware routing algorithm for multimedia sensor networks", *Frontiers of Electrical and Electronic Engineering in China*, Vol. 3, No. 1, pp. 25-33, 2008, DOI: 10.1007/s11460-008-0013-7.

[10] Md. Abdul Hamid, Muhammad Mahbub Alam and Choong Seon Hong, "Design of a QoS-Aware Routing Mechanism for Wireless Multimedia Sensor Networks", in *Proceeding of the IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, New Orleans, LA, USA, 2008, DOI: 10.1109/glocom.2008.ecp.159.

[11] Mohammad Alaei and Jose M. Barcelo-Ordinas, "A Method for Clustering and Cooperation in Wireless Multimedia Sensor Networks", *Sensors*, Vol. 10, No. 4, pp. 3145-3169, 2010, DOI: 10.3390/s100403145.

[12] Anna Förster and Amy L. Murphy, "Froms: A failure tolerant and mobility enabled multicast routing paradigm with reinforcement learning for WSNs", *Ad Hoc Networks*, Vol. 9, No. 5, pp. 940-965, 2011, DOI: 10.1016/j.adhoc.2010.11.006.

[13] Xuedong Liang, Ilangko Balasingham and Sang-Seon Byun, "A multi-agent reinforcement learning based routing protocol for wireless sensor networks", in *Proceeding of the IEEE International Symposium on Wireless Communication Systems*, Reykjavik, Iceland, May 2008, DOI: 10.1109/iswcs.2008.4726117.

[14] Irina Gerasimov and Robert Simon, "A bandwidth-reservation mechanism for on-demand ad hoc path finding", in *Proceedings of the 35th Annual Simulation Symposium*, Massachusetts Ave., NW Washington, DC, United States, April 2002, DOI: 10.1109/simsym.2002.1000079.

[15] Christian Nastasi and Andrea Cavallaro, "WiSE-MNet: an experimental environment for wireless multimedia sensor networks", *Sensor Signal Processing for Defence (SSPD 2011)*, pp. 11, 2011, DOI: 10.1049/ic.2011.0171.

[16] Md Arafat Habib and Sangman Moh, "Game theory-based Routing for Wireless Sensor Networks: A Comparative Survey", *Applied Sciences*, Vol. 9, No. 14, p. 2896, 2019, Available: 10.3390/app9142896.

[17] Arafat Habib and Muhidul Islam Khan, "Reinforcement learning based autonomic virtual machine management in clouds", in *Proceeding of the 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 1083-1088, Dhaka, Bangladesh, 2016, DOI: 10.1109/ICIEV.2016.7760166.

[18] Arafat Habib, Muhidul Islam Khan and Jia Uddin, "Optimal route selection in complex multi-stage supply chain networks using SARSA(λ)", in *Proceeding of the 19th International Conference on Computer and Information Technology (ICCIT)*, 2016, pp. 170-175, Dhaka, Bangladesh, DOI: 10.1109/ICCITECHN.2016.7860190.

[19] Ding, Qianao, Rongbo Zhu, Hao Liu and Maode Ma, "An Overview of Machine Learning-Based Energy-Efficient Routing Algorithms in Wireless Sensor Networks", *Electronics*, Vol. 10, No. 13: 1539, 2021, DOI: 10.3390/electronics10131539.

[20] Upasna Joshi and Rajiv Kumar, "Reinforcement learning based energy efficient protocol for wireless multimedia sensor networks", *Multimed Tools Appl.* 81, 2827–2840 (2022), DOI: 10.1007/s11042-021-11387-w.

[21] Hojjatollah Esmaeili, Behrouz Minaei Bidgoli and Vesal Hakami, "CMML: Combined metaheuristic-machine learning for adaptable routing in clustered wireless sensor networks", *Applied Soft Computing*, Vol. 118, p. 108477, 2022, DOI: 10.1016/j.asoc.2022.108477.

[22] A. Keerthika, V. Berlin Hency, "Reinforcement-Learning based energy efficient optimized routing protocol for WSN", *Peer-to-Peer Networking and Applications,* Vol. 15, pp. 1685–1704, 2022, DOI: 10.1007/s12083-022-01315-6.

[23] Wan-Kyu Yun and Sang-Jo Yoo, "Q-Learning-Based Data-Aggregation-Aware Energy-Efficient Routing Protocol for Wireless Sensor Networks", in *IEEE Access*, Vol. 9, pp. 10737-10750, 2021, DOI: 10.1109/ACCESS.2021.3051360.

[24] Vially Kazadi Mutombo, Seungyeon Lee, Jusuk Lee and Jiman Hong, "EER-RL: Energy-Efficient Routing Based on Reinforcement Learning", *Mobile Information Systems*, Vol. 2021, pp. 1-12, 2021, DOI: 10.1155/2021/5589145.

© 2022 by the author(s). Published by Annals of Emerging Technologies in Computing (AETiC), under the terms and conditions of the Creative Commons Attribution (CC BY) license which can be accessed at http://creativecommons.org/licenses/by/4.0.