

<https://doi.org/10.31891/2219-9365-2026-86-38>

УДК 004.738.5

SIPKO Olena

Taras Shevchenko National University of Kyiv

<https://orcid.org/0000-0003-1385-119X>

e-mail: sipko.olena@knu.ua

ADAPTIVE REINFORCEMENT LEARNING-BASED ROUTING ALGORITHM FOR ENERGY-EFFICIENT IOT WIRELESS SENSOR NETWORKS

The article proposes an adaptive routing algorithm for improving energy efficiency in wireless sensor networks within Internet of Things environments. The proposed approach is based on reinforcement learning and enables sensor nodes to dynamically adapt routing decisions according to network conditions and residual node energy. The algorithm uses the Q-learning method to determine optimal transmission paths and minimize total energy consumption of the network. Simulation experiments were conducted using the NS-3 network simulation environment. The obtained results demonstrate that the proposed approach reduces energy consumption, increases network lifetime, and improves data transmission reliability in comparison with classical routing protocols such as LEACH and PEGASIS. The results confirm the effectiveness of reinforcement learning methods for adaptive routing in dynamic IoT environments.

Keywords: Internet of Things, wireless sensor networks, reinforcement learning, Q-learning, routing algorithm, energy efficiency

СІПКО Олена

Київський національний університет імені Тараса Шевченка

АДАПТИВНИЙ АЛГОРИТМ МАРШРУТИЗАЦІЇ НА ОСНОВІ НАВЧАННЯ ІЗ ПІДКРІПЛЕННЯМ ДЛЯ ЕНЕРГОЕФЕКТИВНИХ БЕЗДРОТОВИХ СЕНСОРНИХ МЕРЕЖ ІНТЕРНЕТУ РЕЧЕЙ

У статті пропонується адаптивний алгоритм маршрутизації для підвищення енергоефективності бездротових сенсорних мереж у середовищах Інтернету речей. Запропонований підхід базується на навчанні з підкріпленням і дозволяє сенсорним вузлам динамічно адаптувати рішення щодо маршрутизації відповідно до мережових умов та залишкової енергії вузла. Алгоритм використовує метод Q-навчання для визначення оптимальних шляхів передачі та мінімізації загального споживання енергії мережею.

Проведено аналіз існуючих протоколів енергоефективної маршрутизації, зокрема LEACH та PEGASIS, та виявлено їх обмеження в динамічних середовищах з нестабільною топологією та змінними рівнями заряду вузлів. Розроблений алгоритм усуває ці обмеження завдяки адаптивному механізму прийняття рішень на основі функції винагороди, яка враховує як енергоспоживання, так і надійність каналу зв'язку. Сенсорні вузли виступають агентами навчання, що взаємодіють із мережовим середовищем і поступово формують оптимальну стратегію маршрутизації без централізованого управління.

Моделювальні експерименти проводилися з використанням середовища моделювання мережі NS-3 для мережі з 100 вузлами та початковим рівнем енергії 2 Дж. Отримані результати демонструють, що запропонований підхід зменшує загальне енергоспоживання мережі приблизно на 25%, збільшує термін служби мережі приблизно на 30% та покращує надійність передачі даних шляхом зниження втрат пакетів і затримок передачі приблизно на 15% порівняно з класичними протоколами маршрутизації LEACH та PEGASIS.

Результати підтверджують ефективність методів навчання з підкріпленням для адаптивної маршрутизації в динамічних середовищах Інтернету речей. Запропонований метод може бути застосований у системах екологічного моніторингу, розумній інфраструктурі, промисловій автоматизації та сільськогосподарських мережах.

Ключові слова: Інтернет речей, бездротові сенсорні мережі, навчання з підкріпленням, Q-навчання, алгоритм маршрутизації, енергоефективність

Стаття надійшла до редакції / Received 07.04.2026

Прийнята до друку / Accepted 05.05.2026

Опубліковано / Published 31.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© SIPKO Olena

PROBLEM STATEMENT IN GENERAL TERMS

AND ITS CONNECTION WITH IMPORTANT SCIENTIFIC OR PRACTICAL TASKS

The Internet of Things (IoT) has become a key enabling technology for many application domains, including smart cities, healthcare systems, industrial automation, and environmental monitoring. Due to the rapid development of sensor technologies, communication protocols, and distributed computing infrastructures, IoT systems provide new opportunities for automation, efficient resource management, and improvement of quality of life. However, wireless sensor networks, which constitute the technological foundation of IoT environments, operate under strict energy constraints caused by the limited power capacity of sensor nodes [1-3].

One of the critical challenges in the deployment of large-scale IoT networks is efficient energy management in order to extend the operational lifetime of the network. Traditional routing protocols such as LEACH and PEGASIS are based on static or semi-static communication structures that do not always consider dynamic changes in network

topology or variations in node energy levels [2]. In practical scenarios, sensor nodes may fail, deplete their energy resources, or change their communication conditions, which complicates the maintenance of optimal routing paths [4-6].

Recent research indicates that machine learning techniques, particularly reinforcement learning (RL), can significantly improve routing adaptation in dynamic network environments [4, 6-8]. Reinforcement learning allows network nodes to autonomously make routing decisions based on local information about the network state and historical transmission data [10-11]. In particular, the Q-learning algorithm enables nodes to learn optimal routing strategies by maximizing cumulative rewards, thereby reducing energy consumption and improving communication efficiency [4, 10].

In this work, an adaptive routing algorithm based on reinforcement learning is proposed for IoT wireless sensor networks. The main objectives of the research are as follows:

- analysis of existing energy-efficient routing approaches in IoT networks;
- development of an adaptive routing algorithm based on the Q-learning method;
- simulation of the proposed algorithm in the NS-3 network simulation environment;
- evaluation of the effectiveness of the proposed method in comparison with classical routing protocols such as LEACH and PEGASIS.

The obtained results can be applied to optimize the operation of IoT systems in various domains, including environmental monitoring, smart buildings, logistics systems, and agricultural monitoring networks. The proposed method demonstrates the potential of reinforcement learning techniques for improving the autonomy and energy efficiency of IoT networks.

ANALYSIS OF RESEARCH AND PUBLICATIONS

Energy efficiency is one of the key challenges in the design and operation of wireless sensor networks used in Internet of Things environments. Since sensor nodes usually operate on limited battery resources, efficient routing strategies play a crucial role in extending the operational lifetime of the network.

A large number of routing protocols have been proposed to address the problem of energy consumption in wireless sensor networks. Among the most widely used approaches are hierarchical routing protocols such as LEACH (Low-Energy Adaptive Clustering Hierarchy) and PEGASIS (Power-Efficient Gathering in Sensor Information Systems). These protocols aim to reduce communication overhead by organizing sensor nodes into clusters or chains, thereby minimizing the number of direct transmissions to the base station [1,2].

The LEACH protocol uses a cluster-based architecture in which cluster heads are periodically selected to aggregate and forward data from member nodes to the base station. This approach reduces communication costs and balances energy consumption among nodes. However, the random selection of cluster heads may lead to uneven energy distribution and reduced efficiency in dynamic network environments.

The PEGASIS protocol improves the LEACH approach by organizing nodes into a communication chain where each node communicates only with its nearest neighbor. Data is gradually aggregated along the chain and transmitted to the base station by a designated leader node. Although PEGASIS can reduce communication energy costs, it still relies on relatively static routing structures and does not fully adapt to changes in network topology or node energy levels.

In recent years, researchers have increasingly explored the use of machine learning techniques to improve routing efficiency in wireless sensor networks. In particular, reinforcement learning has attracted significant attention due to its ability to support adaptive decision-making in dynamic environments.

Reinforcement learning-based routing methods allow sensor nodes to learn optimal communication strategies through interaction with the network environment. Instead of relying on predefined routing rules, nodes continuously update their routing policies based on observed rewards and network feedback. Such approaches enable better adaptation to topology changes, node failures, and variations in energy availability [3].

Despite the promising results reported in recent studies, many reinforcement learning-based routing methods focus primarily on optimizing routing paths without explicitly considering the residual energy of sensor nodes or the long-term balance of network energy consumption.

Therefore, there remains a need for adaptive routing algorithms that simultaneously consider communication efficiency and energy balancing in wireless sensor networks.

In this context, the development of an energy-aware routing approach based on reinforcement learning represents an important research direction for improving the performance and sustainability of IoT networks.

REINFORCEMENT LEARNING FOR ADAPTIVE ROUTING IN IoT NETWORKS

Reinforcement Learning is a machine learning paradigm based on the concept of learning through interaction with an environment. In this approach, an intelligent agent learns to make decisions by performing actions and receiving feedback in the form of rewards or penalties. The objective of the agent is to maximize the cumulative reward obtained over time, which leads to the formation of an optimal decision-making strategy.

In wireless sensor networks, reinforcement learning can be applied to routing optimization by allowing sensor nodes to adaptively select communication paths according to current network conditions. Instead of relying on predefined routing rules, nodes continuously adjust their behavior based on feedback from the network environment. This adaptive mechanism makes reinforcement learning particularly suitable for dynamic IoT systems where network topology, communication conditions, and energy availability may change over time.

The reinforcement learning framework includes several fundamental components:

- Agent – a sensor node participating in the routing process.
- Environment – the wireless sensor network topology and communication conditions.
- State – parameters describing the current network condition, such as residual energy of nodes, communication delays, or link quality.
- Action – the selection of the next node for data forwarding.
- Reward function – a quantitative measure that evaluates the effectiveness of a routing decision based on factors such as energy consumption and communication reliability.

During the learning process, the agent observes the current state of the network and selects an action according to its routing policy. After performing the action, the environment provides a reward that reflects the quality of the chosen decision. Based on this feedback, the agent updates its policy in order to increase the probability of selecting more effective actions in the future.

Through repeated interactions with the environment, the agent gradually learns routing strategies that minimize energy consumption while maintaining reliable data transmission. As a result, reinforcement learning enables decentralized and adaptive routing decisions in wireless sensor networks.

The use of reinforcement learning methods in IoT environments provides several important advantages. First, it allows routing strategies to adapt dynamically to changes in network topology and node energy levels. Second, it reduces the need for centralized control and predefined routing rules. Finally, reinforcement learning enables more efficient utilization of network resources by balancing the communication load among sensor nodes.

These properties make reinforcement learning a promising approach for improving the energy efficiency and adaptability of wireless sensor networks in Internet of Things applications.

Exploration-exploitation strategy in reinforcement learning

One of the fundamental challenges in reinforcement learning is the exploration–exploitation trade-off. The agent must decide whether to exploit previously learned knowledge by selecting the action with the highest expected reward or to explore alternative actions that may potentially lead to better outcomes.

In the context of wireless sensor network routing, exploitation corresponds to selecting the routing path that has previously demonstrated the highest efficiency in terms of energy consumption and communication reliability. Exploration, on the other hand, involves testing alternative routing paths that may provide improved performance under changing network conditions.

Balancing these two strategies is essential for achieving optimal routing performance. Excessive exploitation may cause the algorithm to converge prematurely to a suboptimal routing policy, while excessive exploration may increase communication overhead and energy consumption.

A commonly used approach for managing this trade-off is the ϵ -greedy strategy. In this method, the agent selects the action with the highest Q-value with probability $1 - \epsilon$, while with probability ϵ it selects a random action in order to explore alternative possibilities.

This strategy allows the routing algorithm to gradually shift from exploration to exploitation as the learning process progresses. In wireless sensor networks, such an approach enables nodes to discover efficient routing paths while adapting to changes in network topology and energy distribution.

Q-LEARNING-BASED ROUTING ALGORITHM

In this study, the Q-learning algorithm is applied to optimize routing decisions in IoT wireless sensor networks. Q-learning is a model-free reinforcement learning method that enables an agent to learn optimal actions through interaction with the environment without requiring a prior model of the system.

The algorithm is based on estimating a Q-function, which represents the expected cumulative reward obtained by performing a specific action in a given state. These values are stored in a structure known as a Q-table, where each element $Q(s, a)$ indicates the expected utility of performing action a in state s .

During the learning process, the agent continuously updates Q-values based on feedback from the environment. The update rule for the Q-function is defined as:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)], \quad (1)$$

where:

$Q(s, a)$ – current value of the Q-function for state s and action a ;

α – learning rate coefficient;

r – reward obtained after performing action;

γ – discount factor determining the importance of future rewards;

$\max_{a'} Q(s', a')$ – next state of the system.

The Q-learning algorithm gradually updates Q-values during repeated interactions between the agent and the network environment. As the number of iterations increases, the routing strategy converges toward an optimal policy that maximizes the cumulative reward.

In the context of IoT wireless sensor networks, the reward function is designed to consider both energy consumption and communication performance. This allows the routing process to balance network load and prevent excessive energy depletion of individual sensor nodes.

The operation of the proposed routing system consists of several main stages:

- 1) Network initialization. Sensor nodes are deployed within the network area, and initial energy levels are assigned to each node.
- 2) Agent learning. Sensor nodes act as agents and learn optimal routing strategies using the Q-learning algorithm.
- 3) Route evaluation. After each data transmission, the Q-function is updated according to the obtained reward and the observed network state.
- 4) Route adaptation. Routing paths are dynamically adjusted in response to changes in network topology and energy levels of nodes.

Through repeated iterations, the algorithm gradually improves routing decisions and enables sensor nodes to select energy-efficient communication paths. This adaptive routing strategy contributes to extending network lifetime and improving the reliability of data transmission in IoT environments.

Reward function design

An important component of reinforcement learning-based routing is the design of the reward function, which determines how the agent evaluates the effectiveness of its actions. The reward function guides the learning process by encouraging actions that improve network performance and penalizing inefficient routing decisions.

In wireless sensor networks, the reward function should reflect several key factors influencing the efficiency of network operation. In the proposed routing algorithm, the reward is determined based on two primary parameters: energy consumption of sensor nodes and communication quality.

When a node selects the next hop for data transmission, the reward value is calculated according to the residual energy of the selected node and the quality of the communication link. Routing decisions that lead to lower energy consumption and stable communication links receive higher reward values. Conversely, actions that result in excessive energy usage or unreliable communication are penalized.

This approach allows the routing algorithm to distribute network traffic more evenly among sensor nodes and prevent the rapid depletion of individual nodes. As a result, the overall energy balance of the network is improved, which contributes to extending the operational lifetime of the wireless sensor network.

The use of an energy-aware reward function also enables the algorithm to adapt to changes in network conditions. When nodes experience different energy levels or communication conditions change, the reinforcement learning process automatically adjusts routing decisions in order to maintain efficient network operation.

SIMULATION AND EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed routing algorithm, a series of simulation experiments were conducted using the NS-3 network simulation environment. NS-3 is a widely used discrete-event network simulator that allows researchers to model complex communication systems and analyze the behavior of wireless sensor networks under different operating conditions. The use of this simulator makes it possible to evaluate the performance of routing algorithms before their implementation in real IoT infrastructures.

The simulated system represents a wireless sensor network deployed for a typical Internet of Things monitoring scenario. Sensor nodes are randomly distributed within the network area and communicate with a base station using multi-hop communication. Each node periodically transmits sensed data to the base station, and routing decisions are determined by the selected routing algorithm.

The main simulation parameters used in the experiments are summarized below:

- number of sensor nodes – 100;
- communication radius – 50 m;
- initial energy of each node – 2 J;
- routing protocols used for comparison – LEACH and PEGASIS.

The performance of the proposed Q-learning-based routing algorithm was evaluated using several key performance indicators. These indicators include total energy consumption of the network, network lifetime, and communication reliability measured through packet loss and transmission delays.

The overall comparative performance of the considered routing approaches is presented in Figure 1. This figure shows the relative performance indicators of the classical routing protocols LEACH and PEGASIS in comparison with the proposed Q-learning-based routing algorithm.

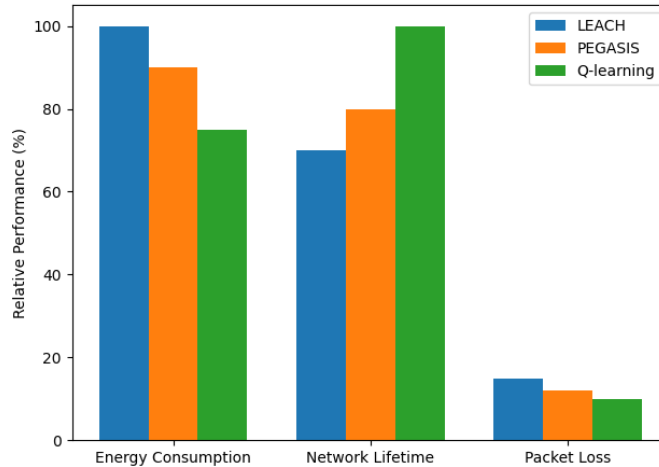


Fig. 1. Comparison of simulation results for LEACH, PEGASIS, and Q-learning-based routing algorithm

As shown in Figure 1, the proposed routing approach demonstrates improved performance across several key metrics. In particular, the Q-learning-based routing algorithm provides lower energy consumption compared with the classical routing methods. This improvement is achieved due to the adaptive routing strategy, which allows sensor nodes to dynamically select communication paths according to the current network conditions and the residual energy of neighboring nodes.

Another important indicator of wireless sensor network performance is the network lifetime, which is defined as the time during which a significant portion of the sensor nodes remain operational. The dynamics of active sensor nodes during the simulation process are illustrated in Figure 2.

As can be observed in Figure 2, the Q-learning-based routing algorithm maintains a higher number of active nodes throughout the simulation period compared with the LEACH and PEGASIS protocols. This behavior indicates that the proposed algorithm distributes communication load more evenly across the network, preventing excessive energy depletion of individual nodes.

As a result, the operational lifetime of the wireless sensor network increases significantly. The simulation results indicate that the proposed routing method extends the network lifetime by approximately 30% compared with traditional routing protocols.

In addition to improving network lifetime, the proposed routing algorithm also enhances the reliability of data transmission. The reinforcement learning mechanism allows nodes to adapt their routing decisions based on network feedback, which reduces the probability of selecting unstable communication links.

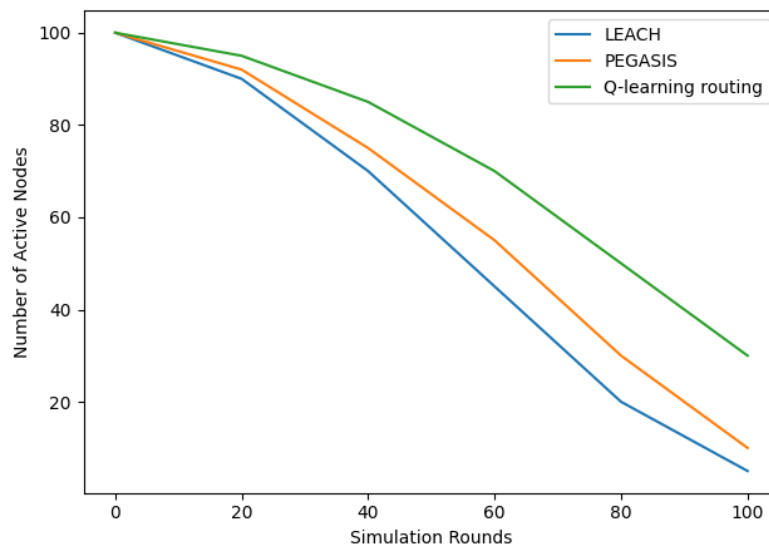


Fig. 2. Network lifetime comparison for LEACH, PEGASIS and Q-learning-based routing algorithm

The experimental results demonstrate that the total energy consumption of the network is reduced by approximately 25% compared with the LEACH protocol. Furthermore, packet loss and transmission delays are reduced by approximately 15%, which contributes to improved overall communication quality.

The obtained results confirm that the proposed Q-learning-based routing approach significantly improves the efficiency of wireless sensor networks operating in Internet of Things environments. By dynamically adapting routing decisions according to network conditions and energy distribution among nodes, the proposed method enables more balanced resource utilization and increases the stability of data transmission.

Overall, the simulation experiments demonstrate that reinforcement learning techniques provide an effective mechanism for improving the performance of IoT wireless sensor networks. The proposed routing algorithm shows clear advantages over classical routing protocols and can be considered a promising approach for energy-efficient and adaptive communication in large-scale IoT systems.

CONCLUSIONS FROM THIS STUDY AND PROSPECTS FOR FURTHER RESEARCH IN THIS DIRECTION

This paper presents an adaptive routing approach for improving the energy efficiency of wireless sensor networks in Internet of Things environments. The proposed method is based on reinforcement learning and utilizes the Q-learning algorithm to dynamically select routing paths according to the current network state and the residual energy of sensor nodes. The study includes an analysis of existing energy-efficient routing protocols used in wireless sensor networks, such as LEACH and PEGASIS. The limitations of these approaches in dynamic network environments motivated the development of an adaptive routing strategy capable of responding to changes in network topology and energy distribution.

The proposed Q-learning-based routing algorithm allows sensor nodes to learn optimal routing decisions through interaction with the network environment. The reward function takes into account both energy consumption and communication reliability, which enables balanced resource utilization and prevents premature energy depletion of individual nodes. Simulation experiments were conducted using the NS-3 network simulator in order to evaluate the effectiveness of the proposed method. The obtained results demonstrate that the Q-learning-based routing algorithm significantly improves the performance of wireless sensor networks compared with traditional routing protocols.

In particular, the experimental results show that the proposed approach reduces overall network energy consumption by approximately 25%, increases network lifetime by approximately 30%, and improves communication reliability by reducing packet loss and transmission delays by approximately 15%. The obtained results confirm that reinforcement learning techniques represent a promising direction for improving the efficiency and adaptability of IoT wireless sensor networks. The proposed method can be applied in various IoT applications, including environmental monitoring, smart infrastructure, industrial automation, and agricultural systems.

Future research may focus on the application of deep reinforcement learning methods, such as Deep Q-Networks, as well as integration of the proposed routing approach with emerging communication technologies including 5G networks and edge computing architectures.

References

1. Atzori L., Iera A., Morabito G. The Internet of Things: A survey. *Computer Networks*. 2010. Vol. 54(15). P. 2787–2805. DOI: 10.1016/j.comnet.2010.05.010
2. Akyildiz I. F., Vuran M. C. *Wireless Sensor Networks*. Wiley, 2010. DOI: 10.1002/9780470515181
3. Rahman, T., Ahmed, R., & Khan, M. (2023). Deep reinforcement learning-based routing in wireless sensor networks. *Neural Computing and Applications*, 35(7), 5117–5134.
4. Sutton R. S., Barto A. G. *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press, 2018.
5. Al-Turjman F., Zahmatkesh H. An overview of security and privacy in smart cities' IoT communications. *Future Generation Computer Systems*. 2019. Vol. 89. P. 653–663. DOI: 10.1016/j.future.2018.01.060
6. Wang J., Gao Y., Liu W., Sangaiyah A. K., Kim H. Energy efficient routing algorithm with mobile sink support for wireless sensor networks. *Sensors*. 2019. Vol. 19(7). P. 1496. DOI: 10.3390/s19071496
7. Tang F., Kawamoto Y., Kato N., Liu J. Future intelligent and secure vehicular network toward 6G: Machine-learning approaches. *Proceedings of the IEEE*. 2020. Vol. 108(2). P. 292–307. DOI: 10.1109/JPROC.2019.2954592
8. Chen Y., Zhao Q. On the lifetime of wireless sensor networks. *IEEE Communications Letters*. 2020. Vol. 24(3). P. 638–642. DOI: 10.1109/LCOMM.2019.2963053
9. Nguyen T. T., Nguyen N. D., Nahavandi S. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Transactions on Cybernetics*. 2020. Vol. 9. P. 3826–3839. DOI: 10.1109/TCYB.2020.2977374
10. Chavan A., Akhare S., Bannagare M. Energy-efficient routing in IoT networks using reinforcement learning. *International journal of innovative research in technology*. 2025. Vol. 12 P. 3526–3528.
11. Ullah, Z.; Al-Turjman, F.; Mostarda, L.; Gagliardi, R. Applications of artificial intelligence and machine learning in smart cities. *Comput. Commun.* 2020, 154, 313–323.