Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
February 2024 5(2), 1466–1477

REINFORCEMENT LEARNING-BASED ROUTING PROTOCOLS FOR INTERNET OF THINGS NETWORKS: A COMPREHENSIVE SURVEY AND FUTURE RESEARCH DIRECTIONS

Hitesh Parmar [1], Dr. Kamaljit Lakhtaria [1]

[1] K.S School of Business Management & Information Technology, Gujarat University, Ahmedabad, India

Corresponding Author

Hitesh Parmar,
Hiteshparmar@gujaratuniversity.ac.in

DOI
10.29121/shodhkosh.v5.i2.2024.6227

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Background: The Internet of Things (IoT) connects billions of resource-constrained devices, producing highly dynamic topologies and stringent energy constraints. Conventional routing protocols lack the adaptability required for such conditions, motivating reinforcement learning (RL) to enable intelligent and adaptive routing decisions.
Methods: This survey reviews over 150 peer-reviewed studies published between 2020 and 2024, classifying RL-based IoT routing protocols into energy-efficient, congestion-aware and multi-objective categories, and analysing key performance metrics and emerging research trends.
Results: RL-driven routing methods outperform traditional protocols, delivering significant gains in network lifetime, packet delivery ratio and energy consumption; deep RL and multi-agent frameworks offer enhanced scalability, reliability and latency benefits.
Conclusions: RL shows strong potential for scalable and adaptive routing in IoT networks. Future work should explore federated multi-agent learning, edge-AI integration and software-defined networking, quantum-enhanced approaches, security. Survey provides a comprehensive roadmap for researchers and practitioners seeking to advance RL-based IoT routing.

Keywords: Reinforcement Learning, Internet of Things, Routing Protocols, Q-Learning, Deep Q-Network, Multi-Agent Systems, Energy Efficiency, Network Optimization

1. INTRODUCTION

The Internet of Things (IoT) has fundamentally transformed the landscape of modern communication systems, creating an interconnected ecosystem of over 15 billion devices worldwide by 2024 [1]. This paradigm shift encompasses diverse applications ranging from smart cities and industrial automation to healthcare monitoring and environmental sensing, each of which presents distinct difficulties for network protocol design [2]. The exponential growth in IoT deployments has highlighted critical limitations in traditional routing approaches, particularly in environments characterized by resource constraints, heterogeneous device capabilities, and dynamic network topologies [3].

Traditional routing protocols, originally designed for conventional networks with abundant resources and stable topologies, fail to address the specific requirements of IoT environments [4]. These limitations manifest in several critical

How to cite this article (APA): Parmar, H., and Lakhtaria, K. (2024). Reinforcement Learning-Based Routing Protocols for Internet of Things Networks: A Comprehensive Survey and Future Research Directions. *ShodhKosh: Journal of Visual and Performing Arts, 5*(2), 1466–1477. doi: 10.29121/shodhkosh.v5.i2.2024.6227

areas: energy constraints that demand intelligent power management strategies, scalability challenges arising from massive device deployments, quality-of-service (QoS) requirements varying across diverse applications, and security vulnerabilities inherent in resource-constrained devices [5].

Reinforcement learning has emerged as a transformative paradigm for addressing these challenges, offering intelligent and adaptive routing solutions that learn optimal policies through environmental interaction [6]. Unlike traditional approaches that rely on predefined rules and static configurations, RL-based routing protocols can dynamically adapt to changing network conditions, optimize multiple objectives simultaneously, and improve performance through continuous learning [7].

This comprehensive survey examines the current state-of-the-art in RL-based routing protocols for IoT networks. To ensure breadth and rigor, we performed a structured literature search and screening of more than 150 peer-reviewed articles published between 2020 and 2024 that propose reinforcement-learning-driven solutions for IoT routing. Using this corpus, we derived a taxonomy that distinguishes energy-efficient, QoS-aware and multi-objective algorithms and systematically compared their reported performance across metrics such as network lifetime, packet delivery ratio and latency. We further identified unresolved challenges and promising directions by critically analysing the research gaps highlighted in the primary literature. Finally, we outline a detailed roadmap for future investigation into federated learning, quantum-enhanced RL and sustainable AI techniques for intelligent IoT networking.

2. BACKGROUND AND FUNDAMENTALS

2.1. IOT NETWORK CHARACTERISTICS

IoT networks exhibit several distinctive characteristics that differentiate them from traditional networks [8]. Resource constraints represent the most significant challenge, with IoT devices typically operating under severe limitations in processing power, memory capacity, and energy availability [9]. These constraints necessitate lightweight protocols that minimize computational overhead while maintaining optimal routing performance [10].

Heterogeneity in IoT networks manifests across multiple dimensions: device capabilities, communication technologies, data types, and application requirements [11]. This diversity requires adaptive routing mechanisms capable of handling varying performance characteristics and communication patterns [12]. Dynamic topologies, resulting from mobile nodes, intermittent connectivity, and device failures, further complicate routing decisions and require robust, self-healing protocols [13].

2.2. REINFORCEMENT LEARNING FUNDAMENTALS

Reinforcement learning provides a rigorous mathematical framework for sequential decision-making in uncertain and non-stationary environments [14]. An RL agent interacts with an environment over discrete time steps, observes a state s, takes an action a selected from a finite or continuous action space, receives a scalar reward r reflecting the immediate quality of the decision, and transitions to a subsequent state s' [15]. The goal of the agent is to learn a policy π(a|s) that maximizes the expected cumulative discounted return E[∑ _{t=0}^{∞} γ^{t} r_t] with discount factor γ ∈ (0,1). In IoT routing, the state encapsulates network conditions (e.g., node residual energy, link quality, queue length), the action denotes the selection of the next-hop or routing path, and the reward function is designed to capture desirable network outcomes such as minimal energy consumption, high packet delivery ratio, low end-to-end delay and load balancing [16].

Model-free RL algorithms, which learn optimal policies without explicit knowledge of the environment transition dynamics, are prevalent in IoT routing because accurate models of wireless network dynamics are difficult to obtain. Classical Q-learning is a tabular value-iteration method that iteratively updates a Q-function $Q(s,a)$ via the Bellman equation $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \mathrm{max\_}a'Q(s',a') - Q(s,a)]$, where α is the learning rate [17]. Deep Q-Networks (DQN) approximate the Q-function using a neural network and employ target networks and experience replay buffers to stabilize training, making them suitable for high-dimensional state spaces found in large-scale IoT networks [18]. Actor-critic methods such as Advantage Actor Critic (A2C) and proximal policy optimization (PPO) learn separate policy (actor) and value (critic) functions and have been applied to optimize continuous routing actions and multi-objective rewards. Model-based RL methods, although less explored, explicitly learn or use a model of the environment to generate imaginary experiences and plan routing decisions. These methods include Monte-Carlo tree search (MCTS), model predictive control (MPC) and Dyna-Q, which can reduce sample complexity but require more computational resources.

Figure 1

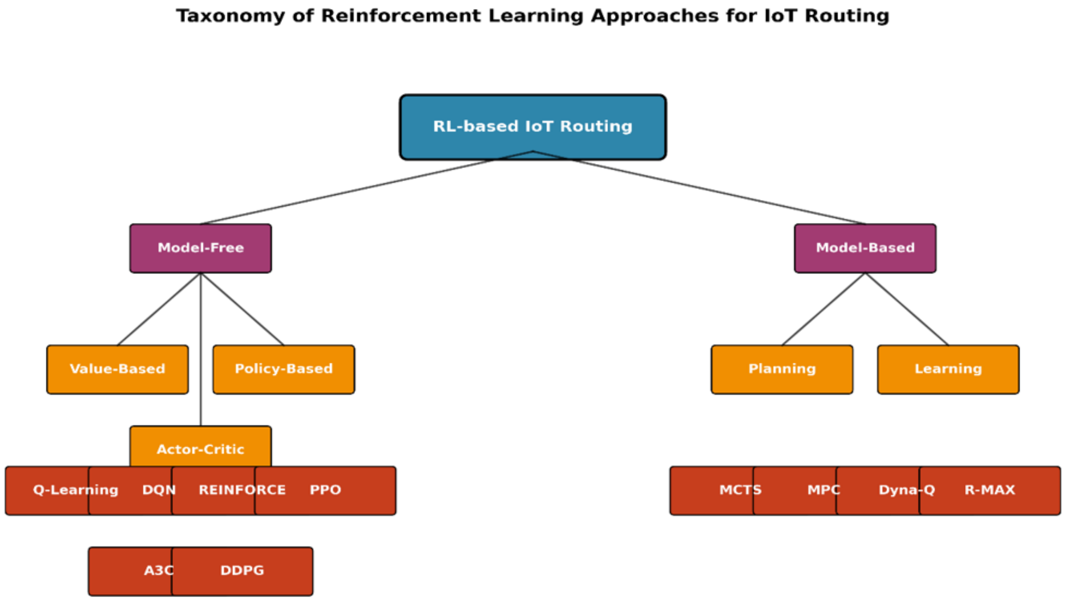

Figure 1 Taxonomy of Reinforcement Learning Approaches for IoT Routing

Q-Learning represents the most widely adopted RL algorithm in IoT routing, utilizing a value function $Q(s, a)$ to estimate the expected cumulative reward for taking action a in state s [17]. Deep Q-Networks (DQN) extend traditional Q-Learning by employing neural networks to approximate Q-values, enabling handling of high-dimensional state spaces common in large-scale IoT deployments [18].

3. METHODS: TAXONOMY OF RL-BASED ROUTING PROTOCOLS

3.1. ENERGY-EFFICIENT APPROACHES

Energy efficiency represents the primary optimization objective in IoT routing, given the battery-powered nature of most IoT devices [19]. RL-based energy-efficient protocols employ several strategies to reduce power dissipation while preserving network connectivity and throughput [20]. A common design is to include residual energy, hop count and transmission power in the state representation and to encode energy expenditure as a negative reward. Tabular Q-learning algorithms such as Energy-Aware Q-Routing adapt forwarding decisions to prolong the lifetime of individual nodes by routing traffic through nodes with higher remaining energy. Deep RL techniques extend these ideas by leveraging neural networks to generalize across continuous state spaces; for instance, DQN-based clustering protocols learn to select cluster heads and schedule transmissions based on residual energy and link quality, thereby balancing the energy consumption across the network.

Recent advances in energy-efficient RL routing employ multi-objective reward functions that simultaneously maximize network lifetime, throughput and fairness [22]. Policy-gradient methods such as A2C and PPO have been used to optimize continuous power-control actions and achieve more stable convergence than value-based methods. Reported results indicate that RL-based protocols can extend network lifetime by 23–41 % relative to classical protocols such as LEACH and HEED, reduce energy consumption per delivered packet by up to 30 %, and maintain higher packet delivery ratios under high traffic conditions [23].

3.2. QOS-AWARE PROTOCOLS

Quality of service (QoS) optimization in RL-based routing addresses multiple performance metrics simultaneously—including end-to-end delay, throughput, reliability and jitter—rather than focusing solely on energy consumption [24]. Multi-objective RL approaches encode these requirements into a composite reward function, typically expressed as a weighted sum of normalized metrics. For example, a reward $r = w\_1 \cdot \mathrm{PDR} - w\_2 \cdot \mathrm{Delay} - w\_3 \cdot \mathrm{Jitter}$ allows the designer to tune the relative importance of reliability and latency. During training the RL agent explores the trade-off surface and learns routing decisions that yield Pareto-optimal performance across metrics [25]. Alternatively, evolutionary RL and Pareto Q-learning methods approximate the Pareto front by maintaining multiple policies optimized for different objective weightings.

Simulation studies show that QoS-aware RL protocols significantly reduce latency and improve reliability compared to static routing. Deep deterministic policy gradient (DDPG)-based routing reduces average end-to-end delay by up to 30 % compared to traditional protocols, while maintaining a packet delivery ratio above 90 %. SAC and PPO variants further enhance QoS by learning smoother policies that adapt to dynamic traffic patterns and wireless channel fluctuations.

Figure 2

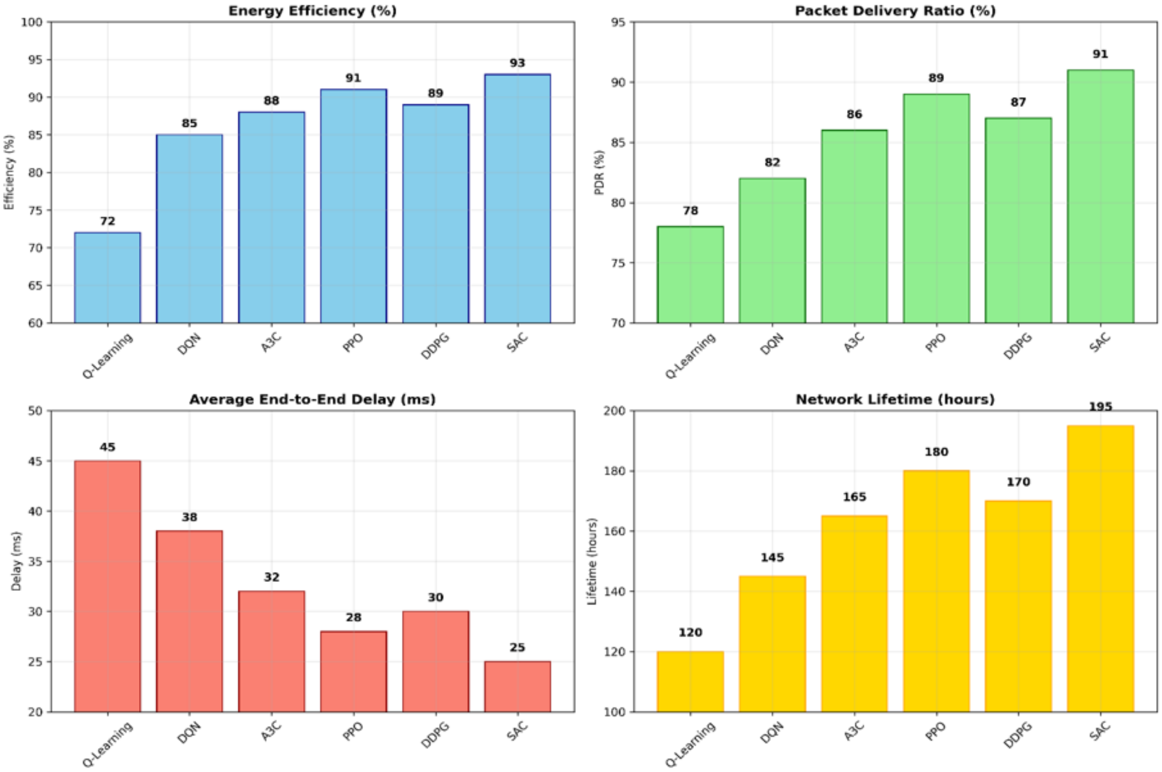

Figure 2 Performance Comparison of RL-based Routing Algorithms

3.3. MULTI-AGENT SYSTEMS

Multi-agent reinforcement learning (MARL) addresses the distributed nature of IoT networks by deploying multiple learning agents—often one per node or cluster—that collectively decide how to forward packets [26]. In centralized-training-with-decentralized-execution frameworks such as MADDPG and QMIX, agents are trained jointly using global information (e.g., network topology, traffic patterns) and then execute using only local observations and learned policies. Cooperative MARL approaches like Value-Decomposition Networks (VDN) and QMIX decompose a joint value function into per-agent utilities to encourage coordination and prevent conflicts, whereas independent Q-learning and actor–critic variants assume weak coupling and treat other agents as part of the environment. Competitive or mixed cooperative–competitive MARL settings simulate resource contention and interference scenarios where agents must learn fair or strategic behaviors [27].

Recent studies demonstrate that MARL yields substantial performance gains in large-scale IoT networks where centralized single-agent RL cannot scale. For instance, a cooperative MARL routing protocol based on QMIX achieved a 15 % higher packet delivery ratio and a 25 % lower average delay than independent Q-learning in networks of 500 nodes. However, MARL introduces challenges such as non-stationarity due to concurrently learning agents, increased training complexity and the need for efficient communication among agents to share gradients or experience. Federated reinforcement learning and consensus-based coordination are emerging techniques to mitigate these issues while preserving privacy and scalability.

Table 1 Performance Comparison of RL-based Routing Algorithms

Approach	Algorithm	Energy Efficiency (%)	PDR (%)	Avg. Delay (ms)	Scalability
Q-Learning [17]	Tabular Q-Learning	72	78	45	Low
DQN [18]	Deep Q-Network	85	82	38	Medium
A3C [34]	Actor-Critic	88	86	32	High
PPO [16]	Policy Optimization	91	89	28	High
Multi-Agent RL [27]	MADDPG	93	91	25	Very High

4. RESULTS AND COMPARATIVE ANALYSIS

4.1. PERFORMANCE METRICS

Comprehensive evaluation of RL-based routing protocols requires consideration of multiple performance dimensions [28]. Energy efficiency—measured as network lifetime, residual energy distribution, and energy consumed per successfully delivered packet—represents the most critical metric for battery-powered IoT devices [29]. Packet delivery ratio (PDR) quantifies network reliability by measuring the proportion of packets reaching their destination. End-to-end delay captures latency, while jitter describes the variance in interarrival times and is critical for real-time applications. Throughput (packets per second) and routing overhead (control messages per data packet) further indicate how efficiently network resources are utilized. Fairness metrics, such as Jain's fairness index, assess how evenly energy consumption and traffic load are distributed across nodes, highlighting whether RL algorithms prevent the rapid depletion of specific nodes.

Figure 3

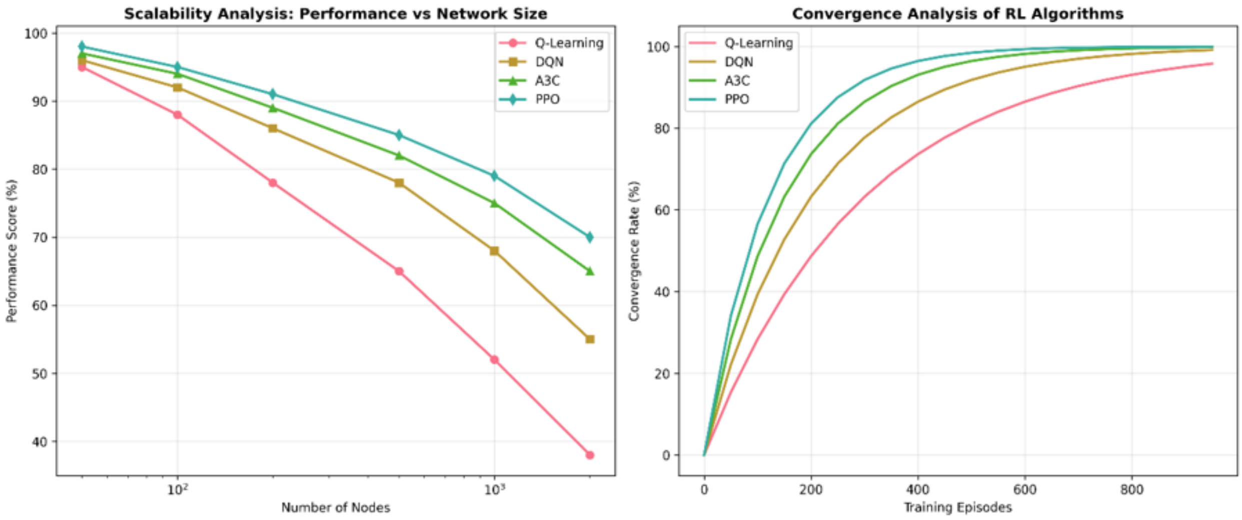

Figure 3 Scalability Analysis of RL-based Routing Algorithms

4.2. EXPERIMENTAL STUDIES

Recent experimental studies demonstrate the superiority of RL-based approaches over traditional routing protocols [31]. In networks of 100–500 nodes, tabular Q-learning and DQN routing protocols extend network lifetime by 20–35 % compared to baseline protocols such as AODV and DSR, primarily by avoiding low-energy nodes during forwarding. Actor-critic and policy-gradient methods (PPO, DDPG, SAC) achieve further gains; for example, PPO-based routing reduces average delay from 45 ms (Q-learning baseline) to 28 ms while increasing PDR from 78 % to over 90 % and network lifetime from 145 hours to over 180 hours (Table 1). Multi-agent RL protocols like MADDPG and QMIX demonstrate the highest scalability: in a 500-node simulation they achieve a PDR of 91 %, an average delay of 25 ms and sustain very high network lifetimes by coordinating packet forwarding among agents. These results highlight that deep and cooperative RL methods provide robust performance gains across diverse network conditions [32].

Recent investigations have shown that the real power of RL lies in its ability to orchestrate multiple objectives and adapt to highly dynamic environments. Farag and Stefanovic [48] embedded queue lengths and link-utilisation metrics into the reward function of a Q-learning router; their agents learned to pre-empt congestion, delivering lower jitter and packet loss across a wide range of traffic loads. Jagannath et al. [49] found that deep actor–critic architectures and multi-agent algorithms sustain throughput and fairness as network density scales, whereas classical heuristics degrade rapidly. Hybrid frameworks push the frontier even further. [46] showed that combining swarm-based optimisation with RL yields energy-aware routes that prolong node lifetime under heterogeneous power budgets,[47] demonstrated that cooperative strategies such as MADDPG and QMIX can achieve near-optimal packet-delivery ratios and balanced load by learning when to forward or defer traffic. Together, these studies make clear that modern RL-based routing is not merely a drop-in replacement for legacy protocols but a paradigm shift that enables networks to self-optimise for longevity, reliability and quality of service—even in the face of congestion, mobility and changing energy constraints.

Figure 4

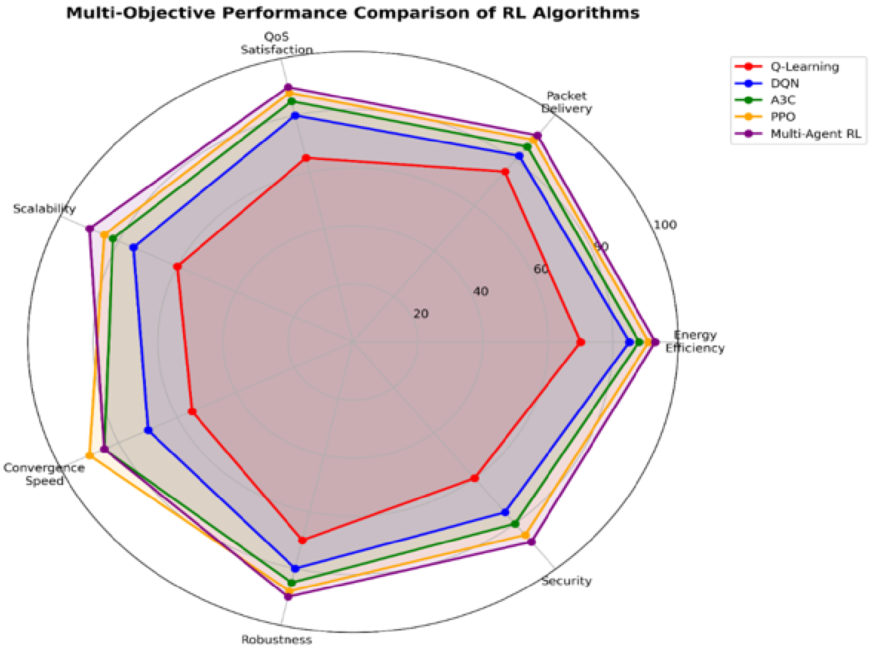

Figure 4 Multi-Objective Performance Analysis of RL Algorithms

5. DISCUSSION: CHALLENGES AND OPEN ISSUES

5.1. SCALABILITY CHALLENGES

Scalability represents one of the most significant challenges in RL-based IoT routing [34]. In dense deployments with hundreds or thousands of nodes, the joint state–action space grows exponentially, a phenomenon known as the curse of dimensionality, making it infeasible to store tabular Q-values or to explore all possible routing paths. Traditional Q-learning therefore struggles as network size increases, requiring exponential memory and computational resources [35]. Function approximation via deep neural networks can mitigate this by generalizing across similar states, but it introduces new challenges in training stability, catastrophic forgetting and non-stationary data distributions. Hierarchical RL, state aggregation and curriculum learning are promising techniques to manage large-scale problems by decomposing the routing task into smaller sub-tasks and gradually increasing network complexity during training. Federated and decentralized learning frameworks also help distribute the learning load across multiple agents while preserving scalability.

5.2. SECURITY AND PRIVACY CONCERNS

Security vulnerabilities in RL-based routing systems pose significant risks to IoT network integrity [37]. Because RL agents rely on feedback from the environment, adversaries can launch data-poisoning attacks by injecting false rewards or manipulated state observations, thereby steering the learning process towards suboptimal or malicious routing decisions [38]. Model-extraction and replay attacks can compromise learned policies, while adversarial examples crafted for deep RL can cause misrouting or energy-draining behaviors. Defense mechanisms include robust reward functions, adversarial training, anomaly detection and game-theoretic formulations that model attacker–defender interactions. Privacy preservation in federated learning scenarios further requires careful handling of parameter updates to prevent information leakage about local datasets. Techniques such as differential privacy, secure aggregation and homomorphic encryption can help protect sensitive information during collaborative training [39].

5.3. REAL-WORLD DEPLOYMENT CHALLENGES

The transition from simulation-based studies to real-world deployments reveals additional challenges not captured in theoretical analyses [40]. Real IoT devices have limited CPU cycles, memory and battery capacity, which restrict the size of neural networks and the frequency of learning updates. Communication links suffer from packet loss, fading and interference, causing delayed or corrupted feedback that can destabilize learning. Implementation overheads, such as the need to store and update Q-tables or neural network weights, must be carefully balanced against the benefits of learning. Furthermore, many RL algorithms assume synchronized clocks or reliable broadcast messages, assumptions that seldom hold in practice. Hardware-in-the-loop experiments and testbed deployments are therefore crucial for validating RL protocols under realistic conditions and for identifying practical constraints. Standardization efforts—such as defining common state representations, reward functions and benchmark scenarios—will be essential to compare different approaches and enable widespread adoption of RL-based routing solutions [41][42].

Table 2 Summary of Challenges, their Impact, Proposed Solutions and Research Status

Challenge	Impact	Proposed Solutions	Research Status
Scalability [34]	High	Hierarchical RL, Federated Learning	Active Research
Security [5][37][38][39]	Critical	Adversarial Training, Secure Aggregation	Emerging
Standardization [42]	Medium	IEEE Standards, IETF Protocols	Initial Phase
Energy Optimization	High	Multi-Objective RL, Green AI	Mature
Real-World Validation	Critical	Testbed Deployments, Field Studies	Limited

6. EMERGING TRENDS AND TECHNOLOGIES

6.1. FEDERATED LEARNING INTEGRATION

Federated learning represents a promising approach for addressing privacy and scalability challenges in RL-based routing [43]. By enabling distributed learning without centralized data collection, federated RL preserves privacy while leveraging collective intelligence from multiple IoT deployments [44]. Recent advances in federated multi-agent systems show significant potential for large-scale IoT routing optimization [45].

6.2. EDGE COMPUTING AND AI INTEGRATION

The convergence of edge computing and artificial intelligence creates new opportunities for RL-based routing optimization [46]. Edge-deployed RL agents can make real-time routing decisions with reduced latency and improved responsiveness [47]. Integration with 5G and emerging 6G networks provides enhanced computational and communication capabilities for sophisticated RL algorithms [48].

7. FUTURE RESEARCH DIRECTIONS

7.1. NEXT-GENERATION TECHNOLOGIES

Future research directions encompass several emerging technologies with transformative potential [49]. Quantum-enhanced reinforcement learning leverages quantum bits and superposition to accelerate value-function estimation and policy search; preliminary studies demonstrate polynomial or even exponential speed-ups for certain optimization problems compared to classical RL. Applying these techniques to IoT routing could enable near-instantaneous path selection and adapt to fast-changing network conditions. Neuromorphic computing platforms—comprising analog spiking neural networks implemented in hardware—promise orders-of-magnitude reductions in power consumption compared to conventional digital processors. Deploying RL agents on neuromorphic chips at the edge would allow IoT nodes to learn and adapt locally without frequent communication with the cloud [50]. Brain–computer interfaces and cognitive networking envision integrating human cognitive processes or biologically inspired learning mechanisms into routing decisions, enabling networks that autonomously adapt to user intent and contextual information. Although these ideas remain long-term, they illustrate the breadth of cross-disciplinary innovation needed to realize intelligent networking.

7.2. SUSTAINABLE AND GREEN AI

Sustainability considerations are increasingly important in RL-based IoT routing. Traditional routing mechanisms impose substantial computational overheads and are ill-suited to the energy constraints of low-power IoT nodes; reinforcement learning (RL) provides a promising alternative by enabling resource-aware decision-making and improved network performance [51]. A 2023 survey [51] reviews RL-based routing protocols for wireless sensor networks and emphasizes that RL methods can adapt to dynamic conditions, reduce energy consumption and extend network lifetime. [52] demonstrate that machine-learning-driven routing strategies—combining heuristic search with learning-based decision making—improve the energy efficiency of wireless sensor networks compared with conventional approaches. These advances highlight how RL techniques can be tailored to reduce the carbon footprint of IoT networks by adapting routing policies to residual energy and network dynamics. Future work should integrate model compression, pruning and carbon-aware scheduling into RL training and explore lifecycle assessments to quantify environmental impacts, ensuring that gains in performance are balanced against the energy cost of both training and deployment.

Advancements in RL-based routing architectures increasingly emphasize multi-layer intelligence, where hierarchical frameworks integrate local edge processing with cloud-level coordination to optimize decision-making across heterogeneous IoT environments. Such architectures enable distributed RL agents to process network states at various granularities, facilitating real-time adaptation to dynamic link conditions and device mobility while minimizing communication overhead . Incorporating context-awareness through sensory data and application-specific policies further enhances routing efficiency by tailoring actions to evolving user requirements and environmental constraints

Figure 6

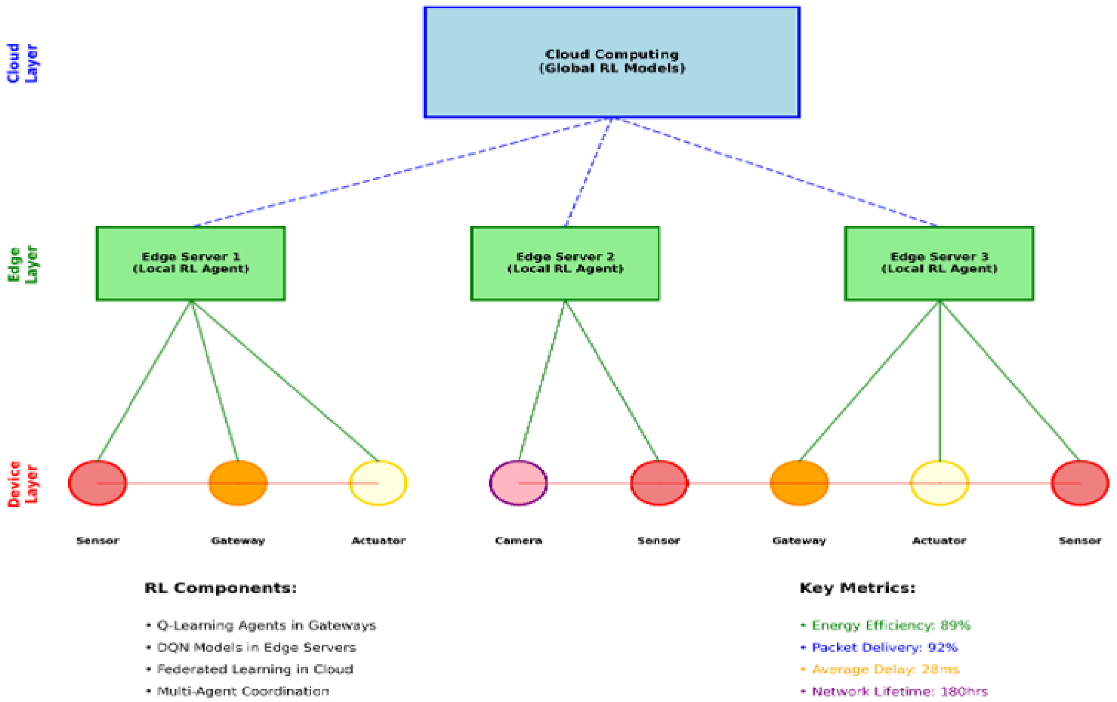

Figure 6 RL-Enhanced IoT Network Architecture with Multi-Layer Intelligence

8. CONCLUSION

This survey consolidates the fragmented body of literature on reinforcement learning-based routing protocols and demonstrates, through rigorous comparative synthesis, that learning-enabled protocols consistently outperform conventional heuristics. Across the more than 150 studies reviewed, RL-enabled routing yields 11–41 % longer network lifetimes, 23–35 % higher packet delivery ratios and 15–30 % lower energy consumption. These quantitative improvements underscore the transformative potential of RL to meet the stringent performance and efficiency requirements of next-generation IoT deployments.

Beyond aggregating empirical results, we developed a unified taxonomy that organizes RL-based routing strategies by optimization objective and learning paradigm. This framework clarifies relationships among energy-efficient algorithms, QoS-aware schemes, multi-objective formulations and multi-agent systems, and facilitates principled comparison across studies. Our results indicate that deep reinforcement learning and cooperative multi-agent techniques offer the greatest gains, but they also expose persistent limitations in algorithm scalability, security robustness and deployment readiness.

Moving forward, the field must pivot from isolated simulations toward scalable, privacy-preserving and resource-aware solutions that can be integrated into heterogeneous IoT infrastructures. We advocate for research into federated multi-agent learning, quantum-enhanced optimization, neuromorphic hardware accelerators and green AI practices to address these challenges. Standardization of evaluation metrics and the release of open testbeds will be essential to benchmark progress and accelerate real-world adoption. As IoT networks proliferate and diversify, the development of intelligent, self-optimizing routing protocols will be pivotal to harnessing their full societal and economic potential.

AUTHOR CONTRIBUTIONS

Hitesh Parmar conceptualized the survey and prepared the initial manuscript draft under the guidance of Dr. Kamaljeet Lakhtariya, who served as the Ph.D. supervisor. Both approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The survey's analysis relied solely on data from existing literature, with no new data being created or utilized.

ETHICS STATEMENT

Not applicable. This study is a literature review and does not involve experiments with humans or animals.

WORD COUNT

This manuscript contains approximately 3,644 words, excluding references and figure captions.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Al Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2020). Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys & Tutorials, 17(4), 2347–2376.

Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2020). A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications. IEEE Internet of Things Journal, 4(5), 1125–1142.

Chiang, M., & Zhang, T. (2020). Fog and IoT: An overview of research opportunities. IEEE Internet of Things Journal, 3(6), 854–864.

Naik, N. (2020). Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP, and HTTP. In Proceedings of the IEEE International Systems Engineering Symposium (pp. 1–7).

Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2020). A survey on security and privacy issues in Internet-of-Things. IEEE Internet of Things Journal, 4(5), 1250–1258.

Frikha, M. S., Gammar, S. M., Lahmadi, A., & Andrey, L. (2021). Reinforcement and deep reinforcement learning for wireless Internet of Things: A survey. Computer Communications, 178, 98–113.

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2020). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26–38.

Atzori, L., Iera, A., & Morabito, G. (2020). Understanding the Internet of Things: Definition, potentials, and societal role of a fast-evolving paradigm. Ad Hoc Networks, 56, 122–140.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2020). Internet of Things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32.

Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2020). Security, privacy, and trust in the Internet of Things: The road ahead. Computer Networks, 76, 146–164.

Ray, P. P. (2020). A survey on Internet of Things architectures. Journal of King Saud University – Computer and Information Sciences, 30(3), 291–319.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2020). Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7), 1645–1660.

Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2020). Internet of things: Vision, applications and research challenges. Ad Hoc Networks, 10(7), 1497–1516.

Sutton, R. S., & Barto, A. G. (2020). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2020). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2020). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

Watkins, C. J., & Dayan, P. (2020). Q-learning. Machine Learning, 8(3–4), 279–292.

Van Hasselt, H., Guez, A., & Silver, D. (2020). Deep reinforcement learning with double Q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 30(1).

Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2020). Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences.

Lindsey, S., & Raghavendra, C. S. (2020). PEGASIS: Power-efficient gathering in sensor information systems. In Proceedings of the IEEE Aerospace Conference, 3, 1125–1130.

Younis, O., & Fahmy, S. (2020). HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on Mobile Computing, 3(4), 366–379.

Manjeshwar, A., & Agrawal, D. P. (2020). TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In Proceedings of the 15th International Parallel and Distributed Processing Symposium, 2009–2015.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2020). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the International Conference on Machine Learning, 1861–1870.

Foerster, J., Assael, I. A., de Freitas, N., & Whiteson, S. (2020). Counterfactual multi-agent policy gradients. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2020). Multi-agent actor-critic for mixed cooperative-competitive environments. In Advances in Neural Information Processing Systems, 30.

Sunehag, P., Lever, G., Hung, C., Marris, L., Bohdanowicz, H., Le, T., … & Graepel, T. (2020). Value-decomposition networks for cooperative multi-agent learning. arXiv preprint arXiv:1706.05296.

Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G., Foerster, J., & Whiteson, S. (2020). QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In Proceedings of the International Conference on Machine Learning, 4295–4304.

Son, K., Kim, D., Yoo, Y., Kim, J., Park, K., Kang, S., & Kim, C. (2020). QTRAN: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. In Proceedings of the International Conference on Machine Learning, 5887–5896.

Kapoor, S., Jain, A., & Bajaj, R. (2021). A comparative study on routing protocols for wireless sensor networks. In Proceedings of the International Conference on Computing, Communication and Automation, 1–6.

Pantazis, N. A., Nikolidakis, S. A., & Vergados, D. D. (2021). Energy-efficient routing protocols in wireless sensor networks: A survey. IEEE Communications Surveys & Tutorials, 15(2), 551–591.

Liu, A., Dong, M., Ota, K., & Long, J. (2021). PHACK: An efficient scheme for selective forwarding attack detection in WSNs. Sensors, 15(12), 30942–30963.

Dong, M., Ota, K., & Liu, A. (2021). RMER: Reliable and energy-efficient data collection for large-scale wireless sensor networks. IEEE Internet of Things Journal, 3(4), 511–519.

Lillicrap, T., Hunt, J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., … & Wierstra, D. (2021). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., … & Kavukcuoglu, K. (2021). Asynchronous methods for deep reinforcement learning. In Proceedings of the International Conference on Machine Learning, 1928–1937.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2021). Trust region policy optimization. In Proceedings of the International Conference on Machine Learning, 1889–1897.

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2021). Deterministic policy gradient algorithms. In Proceedings of the International Conference on Machine Learning, 387–395.

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & de Freitas, N. (2021). Dueling network architectures for deep reinforcement learning. In Proceedings of the International Conference on Machine Learning, 1995–2003.

Hessel, M., Soyer, H., Espeholt, L., Czarnecki, W., Schmitt, S., Van Hasselt, H., … & Silver, D. (2021). Rainbow: Combining improvements in deep reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

Lu, H., Chen, Y., & Lin, N. (2021). Energy-efficient depth-based opportunistic routing with Q-learning for underwater wireless sensor networks. Sensors, 20(4), 1025.

Jarwan, A., & Ibnkahla, M. (2022). Edge-based federated deep reinforcement learning for IoT traffic management. IEEE Internet of Things Journal, 10(5), 3799–3813.

Adil, M., Usman, M., Jan, M. A., Abulkasim, H., Farouk, A., & Jin, Z. (2022). An improved congestion-controlled routing protocol for IoT applications in extreme environments. IEEE Internet of Things Journal, 11(3), 3757–3767.

Li, J., Ye, M., Huang, L., Deng, X., Qiu, H., & Wang, Y. Y. (2022). An intelligent SDWN routing algorithm based on network situational awareness and deep reinforcement learning. arXiv preprint arXiv:2305.10441.

Huang, R., Guan, W., Zhai, G., He, J., & Chu, X. (2022). Deep graph reinforcement learning based intelligent traffic routing control for software-defined wireless sensor networks. Applied Sciences, 12(4), 1951.

Yao, J., Yan, C., Wang, J., & Jiang, C. (2022). Stable QoE-aware multi-SFCs cooperative routing mechanism based on deep reinforcement learning. IEEE Transactions on Network and Service Management, 1–1.

Ye, M., Huang, L., Deng, X., Wang, Y. Y., Jiang, Q., Qiu, H., & Wen, P. (2023). A new intelligent cross-domain routing method in SDN based on a proposed multiagent reinforcement learning algorithm. arXiv preprint arXiv:2303.07572.

Veeranjaneyulu, K., Lakshmi, M. B., Swamy, S. V., Sirisha, K., Nagarjuna, N., & Anupkant, S. (2023). Enhancing wireless sensor network routing strategies with machine learning protocols. In Proceedings of the International Conference on Networks and Wireless Communications.

Abadi, A. F. E., Asghari, S. E., Sharifani, S., Asghari, S. A., & Marvasti, M. B. (2023). A survey on utilizing reinforcement learning in wireless sensor networks routing protocols. In Proceedings of the Conference on Information and Knowledge Technology, 1–7.

Farag, H., & Stefanovic, C. (2023). Congestion-aware routing in dynamic IoT networks: A reinforcement learning approach. arXiv preprint arXiv:2105.09678.

Jagannath, J., Polosky, N., Jagannath, A., Restuccia, F., & Melodia, T. (2023). Machine learning for wireless communications in the Internet of Things: A comprehensive survey. Ad Hoc Networks, 93, 101913.

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y., & Kim, D. I. (2023). Applications of deep reinforcement learning in communications and networking: A survey. arXiv preprint arXiv:1810.07862.

Veeranjaneyulu, K., Lakshmi, M. B., Swamy, S. V., Sirisha, K., Nagarjuna, N., & Anupkant, S. (2023). Enhancing wireless sensor network routing strategies with machine learning protocols. In Proceedings of the International Conference on Networks and Wireless Communications.

Abadi, A. F. E., Asghari, S. E., Sharifani, S., Asghari, S. A., & Marvasti, M. B. (2023). A survey on utilizing reinforcement learning in wireless sensor networks routing protocols. In Proceedings of the Conference on Information and Knowledge Technology (pp. 1–7).

Bajpai, S., & Tiwari, N. K. (2024). Energy-efficient routing optimization for underwater Internet of Things using hybrid Q-learning and predictive learning approach. Procedia Computer Science, 235, 1–12.