International Journal of Engineering Trends and Applications (IJETA) – Volume 12 Issue 4, Jul-Aug 2025

RESEARCH ARTICLE OPEN ACCESS

A Novel Scheme to Maintain Quality of Service in the Internet of Things (IOT)

Hemant Mittal, Divyansh Dewani, Rajat Vaishya

[1,2,3] Computer Science & Engineering, Global Institute of Technology, Jaipur

ABSTRACT

The Internet of Things (IoT) has emerged as a transformative paradigm, enabling interconnected devices to seamlessly communicate and share data across various applications such as smart homes, healthcare systems, industrial automation, and intelligent transportation. However, the rapid proliferation of IoT devices has introduced significant challenges in maintaining Quality of Service (QoS), given the stringent and diverse requirements these applications impose. Ensuring reliable, real-time, and efficient data exchange within such a heterogeneous and dynamic environment necessitates advanced resource management strategies.

This paper presents a novel QoS-Aware Adaptive Resource Allocation Scheme (QAARAS) that intelligently combines edge computing capabilities with reinforcement learning to address the unique demands of IoT networks. By leveraging the decision-making process of Q-learning algorithms at the edge layer, the proposed system dynamically optimizes network resources in response to fluctuating conditions such as network congestion, latency, and energy consumption. Our simulation-based evaluation demonstrates that QAARAS significantly outperforms traditional scheduling approaches—such as Round Robin and Weighted Fair Queuing—by reducing latency, enhancing throughput, and improving packet delivery ratios.

The results underscore the potential of integrating intelligent edge-based solutions to achieve scalable and responsive QoS provisioning in IoT environments. This study contributes to the growing body of research aimed at enhancing the resilience and efficiency of next-generation IoT networks through adaptive and context-aware computing frameworks.

Keywords - Internet of Things, Quality of Service, Edge Computing, Machine Learning, Resource Allocation.

I. Introduction

The Internet of Things (IoT) has revolutionized the digital and physical worlds by interconnecting billions of devices—from everyday household gadgets to

industrial machines—allowing them to collect, process, and exchange data autonomously. As this ecosystem continues to grow, it drives immense volumes of traffic across the network infrastructure, thereby straining existing communication models and computational resources. on this relatively dynamic and heterogeneous environment, making sure fine of carrier (QoS) has emerged as a essential problem, particularly for programs that call for low latency, excessive reliability, and constant statistics throughput.

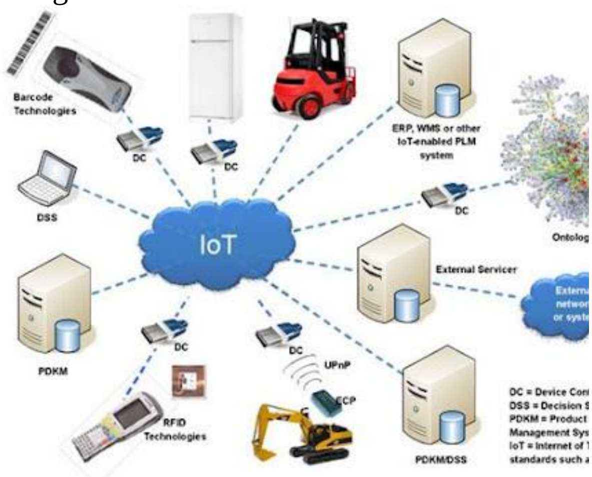
Conventional cloud-centric architectures, while imparting considerable processing electricity, frequently fail to satisfy the stringent latency requirements of actual-time packages. Statistics have to journey sizable distances to centralized records centres, introducing delays and bottlenecks, especially below excessive-load conditions. Furthermore, the range of IoT programs—ranging from time-sensitive healthcare monitoring to bandwidth-heavy video surveillance—calls for extra nuanced and adaptable aid control procedures.

One of the most promising strategies to deal with the ones demanding situations is thru facet computing. in place of depending carefully on faraway cloud servers, element computing brings processing energy toward where the data is surely generated. This allows cut down on latency, improves scalability, and lets in more localized control of facts.

That stated, element-based systems nevertheless ought to deal with constrained assets—in particular whilst devices behave unpredictably or when community conditions are continuously changing.

To handle this kind of complexity, we want smart systems that may study from experience and adapt as matters exchange. That's in which reinforcement studying (RL) is available in. As a branch of gadget learning, RL empowers structures to make smarter, context-conscious alternatives based totally on real-time remarks from their surroundings.

By using setting RL agents at the network's area, we can reveal key overall performance metrics like postpone, energy consumption, and community visitors. these retailers can then make on-the-fly changes to how sources are allocated—supporting preserve sturdy quality of carrier even beneath shifting situations.



(Fig No I: - Internet of Things)

In this paper, we propose the QoS-Aware Adaptive Resource Allocation Scheme (QAARAS)—an innovative framework that combines the strengths of edge computing and reinforcement learning. Our approach employs a multi-layer architecture in which RL agents operate at the edge to manage real-time data traffic efficiently, ensuring service guarantees for various IoT applications. Through extensive simulations, we evaluate the performance of QAARAS against conventional scheduling algorithms and demonstrate its ability to deliver superior results across multiple QoS metrics.

The rest of the paper is prepared as follows: phase II critiques associated paintings and modern-day tendencies in QoS provisioning for IoT. section III introduces the structure and operational details of the proposed QAARAS framework. phase IV describes the experimental setup, whilst segment V discusses simulation consequences. segment VI outlines destiny guidelines and potential demanding situations, and section VII concludes the paper.

II. Intelligent Scheduling Mechanisms for IoT: Balancing Efficiency and Latency in Edge and Cloud Computing

The fast growth of IoT has induced extensive studies into mechanisms which can preserve overall performance stages under various community hundreds and alertness needs. numerous current frameworks goal to optimize quality of service (QoS) thru diverse scheduling, prioritization, and data dealing with techniques.

traditional scheduling algorithms which includes First-Come-First-Serve (FCFS), round Robin (RR), and precedence Queuing had been extensively utilized in traditional networks. whilst these techniques are trustworthy and require minimal computation, they fall quick in dynamically adapting to the heterogeneous and context-conscious nature of IoT environments. RR, for example, assigns same time slices to responsibilities, ignoring their urgency or importance, that can lead to inefficiencies in undertaking-vital situations.

current advancements have focused on incorporating smart choice-making fashions. numerous works discover device learning and deep studying techniques to forecast visitors styles and optimize aid distribution consequently. for example, authors in proposed a deep Q-community-primarily based aid scheduling scheme to are expecting tool behaviour and alter service transport. however, many such fashions are nonetheless cloud-reliant, which reintroduces the latency and bottleneck problems facet computing seeks to put off.

III. Bridging the Gap: Integrating Reinforcement Learning for Real-Time Edge Intelligence in IoT

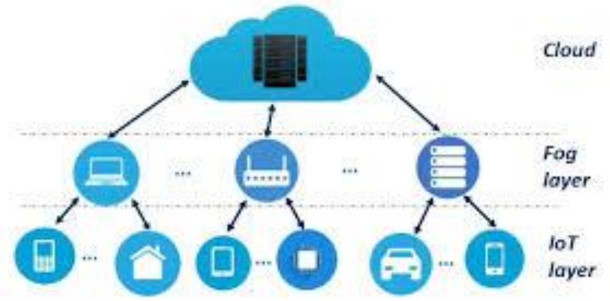
edge computing frameworks have received reputation for his or her capability to lessen latency and manage nearby data. Research like applied fog-assisted task scheduling mechanisms that prioritize time-touchy programs. In spite of upgrades, many of these solutions lack the adaptability required in real-time environments with moving hundreds.

to conquer these demanding situations, reinforcement mastering (RL) has emerged as a promising device. RL fashions examine most excellent guidelines via interacting with their environments, making them appropriate for dynamic, decentralized systems like IoT. In, Q-getting to know turned into hired for dealing with sensor networks, leading to progressed power efficiency and decreased transmission delays.

Whilst prior paintings lay a sturdy foundation, it often lacks integration between edge intelligence and autonomous mastering models. Our proposed QAARAS fills this gap by way of introducing a reinforcement getting to know-based scheme applied without delay at the threshold layer, making sure quicker decisions and context-conscious adaptability. Unlike previous systems that rely upon predefined regulations or static weights, QAARAS constantly learns and evolves with converting network situations, offering a extra resilient and scalable QoS framework.

IV. Proposed Model: QAARAS Architecture

within the ever-increasing universe of net of things (IoT), the high-quality of carrier (QoS) necessities have emerged as one of the maximum difficult aspects to tackle due to the inherently diverse, dynamic, and aid-touchy nature of IoT networks. The QoS-conscious Adaptive resource Allocation Scheme (QAARAS) structure, proposed as a ground breaking version, targets to redefine aid allocation strategies by using combining the decentralized strength of facet computing with the adaptive intelligence of reinforcement studying. in particular, QAARAS leverages Q-getting to know as its selection-making backbone to dynamically allocate sources even as making sure premiere QoS requirements across heterogeneous IoT environments.



(Fig No II: - IOT Architecture)

Introduction to QAARAS

QAARAS is designed with a unique aim: to bridge the space among conventional static scheduling fashions and the dynamic, context-conscious needs of IoT networks. As IoT ecosystems keep making bigger, the reliance on centralized processing fashions has established inefficient in addressing real-time fluctuations in device behaviour, latency requirements, and bandwidth utilization. QAARAS recognizes the need for shrewd decentralization, enabling facet nodes to take on the role of local processors, while employing Q-studying to beautify choice-making based on context-precise necessities.

At the heart of QAARAS lies a symbiotic interplay between area computing and reinforcement learning principles. facet computing addresses the problem of latency by way of processing data in the direction of the supply, decreasing dependency on centralized cloud structures, while Q-gaining knowledge of empowers the architecture with the capability to adaptively allocate sources based on actual-time remarks loops from IoT gadgets.

Core Features of QAARAS

1. **1. Decentralized Processing:** QAARAS employs aspect computing to decentralize the workload. through dispensing the processing obligations throughout a couple of aspect nodes in desire to relying on vital servers, the structure minimizes latency and bandwidth congestion. this option is important for real-time programs which consist of independent motors, telemedicine, and smart towns.
2. **QoS Awareness**: QAARAS doesn't merely allocate resources—it does so with an acute attention of QoS requirements. It ensures that each resource allocation selection factors in parameters which include device priority, latency tolerance, throughput desires, and energy performance.
3. **Adaptive Resource Allocation**: It is not like static scheduling systems, QAARAS dynamically adjusts useful resource allocation primarily based on continuous remarks from IoT gadgets. by way of leveraging Q-getting to know, the shape learns from the environment and evolves to provide optimized answers to diverse QoS demands.
4. **Scalability**: QAARAS is inherently scalable, making it appropriate for large-scale IoT networks. as the range of devices grows, facet nodes seamlessly scale to deal with the increasing workload without compromising overall performance.

V. Integration of Edge Computing and Q-Learning

QAARAS's modern method integrates pivotal technology: component computing and Q-studying. facet computing is employed to carry computation in the direction of IoT gadgets, mitigating delays associated with cloud-primarily based processing. thru using component nodes as neighbourhood processors, QAARAS achieves great reductions in latency and bandwidth usage, permitting real-time choice-making for packages wherein milliseconds rely.

but, Q-getting to know presents QAARAS with its intelligence. This reinforcement mastering approach lets in the system to learn gold well known strategies for useful resource allocation primarily based mostly on trial-and-mistakes interactions with its environment. each part node acts as an agent in the Q-getting to know model, receiving feedback from the IoT community to regulate its allocation guidelines dynamically.

As an example, if an aspect node detects multiplied visitors from a specific set of IoT gadgets requiring higher bandwidth, it uses Q-learning to prioritize those devices briefly even as ensuring minimal disruption to other gadgets in the network. Over time, QAARAS becomes more and greener in its selection-making processes, adapting to the particular needs of the network it serves.

Aspect	**Edge Computing**	**Q-Learning**
Definition	A distributed computing framework where data processing happens closer to the data source (at the edge of the network) rather than in a centralized cloud.	A model-free reinforcement learning algorithm that learns the value of actions in states to maximize long-term rewards.
Main Focus	Reducing latency, bandwidth usage, and improving real-time data processing by performing computations locally.	Optimizing decision-making processes through trial-and-error to find the best actions based on state-action pairs.
Purpose	To bring computation closer to the IoT devices, improving response times and reducing the strain on cloud servers.	To find an optimal policy (sequence of actions) for decision-making in an uncertain environment.
Location of Processing-	Localized, typically at or near the data source (e.g., IoT devices, gateways, edge servers).	Computation happens in an agent's environment, where it learns from interactions and states.
Latency	Low latency due to local data processing.	Indirectly related to latency; faster learning may reduce the time to make decisions, but not focused on latency itself.

(Table: - Edge Computing and Q-Learning)

VI. The Decision-Making Process

The QAARAS architecture employs a systematic decision-making process powered by Q-learning:

1. **State Representation**: The architecture defines the country of the machine primarily based on present day network situations, QoS necessities, aid availability, and IoT tool behaviour.
2. **Action Selection**: based at the current kingdom, the edge nodes choose moves (e.g., allocating additional bandwidth, prioritizing unique devices, or redistributing tasks). to begin with, movements are selected randomly, but because the system learns, it starts to pick actions based totally on accrued rewards.
3. **Reward System**: every motion is discovered with the aid of the usage of remarks from the IoT devices and the community, which serves as a reward. high-quality rewards are assigned to moves that enhance QoS metrics, on the same time as bad rewards are given for moves that bring about inefficiencies or bottlenecks.
4. **Policy Update**: The Q-mastering set of rules makes use of the praise tool to update its regulations, gradually converging inside the route of an most advantageous useful resource allocation method.

Advantages of QAARAS

The QAARAS architecture offers several distinct advantages over traditional resource allocation systems:

- **Real-Time Adaptability**: QAARAS adapts to fluctuating network conditions in real-time, ensuring consistent QoS standards even under heavy workloads or unexpected disruptions.

- **Energy Efficiency**: By decentralizing processing through edge computing, QAARAS minimizes energy consumption associated with data transmission and processing.
- **Enhanced QoS Management**: QAARAS excels in handling diverse QoS requirements across heterogeneous IoT devices, from latency-sensitive applications to high-throughput scenarios.
- **Self-Optimization**: Through continuous learning, QAARAS evolves to become more efficient over time, reducing the need for manual intervention or static configurations.

Applications of QAARAS

The versatility of QAARAS makes it applicable to a wide range of IoT domains:

1. **Smart Cities**: In urban environments, QAARAS can optimize resource allocation for traffic management systems, environmental sensors, and public safety applications.
2. **Telemedicine**: QAARAS ensures reliable QoS for telemedicine applications, where latency and bandwidth are critical for remote consultations and medical data transmission.
3. **Autonomous Vehicles**: By prioritizing low-latency communication, QAARAS facilitates the real-time decision-making required for autonomous driving systems.
4. **Industrial IoT**: QAARAS enhances efficiency in industrial settings by dynamically allocating resources for predictive maintenance, real-time monitoring, and automation.

Challenges and Future Scope

Despite its promising features, QAARAS faces challenges that warrant further research and development:

- **Complexity**: Implementing Q-learning at scale in heterogeneous IoT networks can be computationally intensive, requiring robust processing capabilities.
- **Security Concerns**: Decentralized architectures like QAARAS may be vulnerable to security threats, necessitating advanced encryption and authentication mechanisms.
- **Interoperability**: Ensuring seamless integration with diverse IoT devices and protocols remains a key challenge.

Looking ahead, the future scope of QAARAS includes exploring hybrid learning techniques, integrating block chain for enhanced security, and developing lightweight algorithms to address computational constraints.

VII. System Overview

QAARAS is structured into three interconnected layers that collaborate to optimize performance: the device layer, the edge layer, and the cloud layer.

1. **Device Layer**: this accretion consists of all IoT sensors, actuators, and gives up-devices liable for records collection, actuation, and preliminary sign processing. The devices on this layer are distinctly heterogeneous, with varying computational talents, energy constraints, and communiqué protocols. in spite of their limitations, those gadgets are the number one data producers and serve as the foundational factors of the IoT atmosphere.
2. **Edge Layer**: acting because the intelligent center layer, the brink nodes are deployed in proximity to IoT gadgets to provide neighborhood computation, transient storage, and facts filtration. inside this layer, QAARAS employs Q-getting to know sellers that continuously monitor gadget situations together with bandwidth availability, latency, packet loss, and electricity intake. based totally on these observations, the retailers decide at the

most excellent aid allocation strategies to meet the evolving needs of linked gadgets in actual time.

3. **Cloud Layer**: at the same time as facet nodes cope with time-sensitive computations, the cloud layer supports long-term analytics, information warehousing, and model updates. it is able to additionally serve as a backup while the brink is overloaded or while large-scale records aggregation is needed. This hierarchical model ensures a balanced distribution of processing workloads even as minimizing latency and bottlenecks related to centralized systems.

Q-Learning Mechanism

at the heart of QAARAS is a Q-learning set of rules carried out at each facet node. Q-getting to know is a version-free reinforcement learning approach wherein retailers study finest guidelines through trial-and-errors interactions with their environment. every edge agent maintains a Q-table that maps discovered system states—which includes modern community congestion levels, undertaking urgency, power tiers, and records glide rates—to a fixed of possible moves.

whilst an agent performs an action, such as reallocating bandwidth or rescheduling obligations, it gets a reward or penalty based totally at the outcome. through the years, the Q-values are updated the use of the Bellman equation, permitting the agent to analyze which strategies yield the first-rate lengthy-term benefits. This learning system enables QAARAS to dynamically adapt to unpredictable workloads and resource constraints without relying on predefined policies or static configurations.

VIII. Resource Allocation Strategy

The resource allocation common sense in QAARAS is each proactive and reactive. through continuous tracking of QoS indicators, the Q-studying agent can count on visitors spikes or capacity bottlenecks and make early modifications to useful resource distribution. as an instance, inside the presence of a time-vital fitness monitoring mission, the device can also prioritize its statistics packets by way of allocating extra bandwidth and processing electricity, although it way delaying lower-precedence tasks which includes smart lights updates.

moreover, QAARAS introduces a remarks-pushed control loop. After each scheduling decision, the machine evaluates key performance metrics—which includes postpone, jitter, throughput, and strength intake—to determine the effectiveness of the selected motion. these metrics feed again into the Q-table, reinforcing a success strategies and discouraging much less effective ones. This loop ensures that the gadget constantly refines its overall performance over the years.

By means of using this wise, multi layered layout, QAARAS considerably improves upon current models by using offering real-time adaptability, contextual recognition, and scalable overall performance control. It’s far particularly ideal for packages with numerous QoS necessities, which include healthcare, autonomous motors, smart grids, and industrial automation, wherein keeping service continuity is paramount. the combination of Q-learning inside facet computing infrastructure offers a pathway closer to extra self-sufficient, self-optimizing IoT systems able to maintaining high provider satisfactory underneath dynamic and unpredictable conditions.

IX. REFERENCES

[1]. Chen, X., Zhang, H., Wu, C., Mao, S., Ji, Y., & Bennis, M. (2018). Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning. IEEE Internet of Things Journal, 6(3), 4005–4018.

[2]. Goudarzi, M., Palaniswami, M., & Buyya, R. (2021). A distributed deep reinforcement learning technique for application placement in edge and fog computing environments. IEEE Transactions on Mobile Computing, 22(5), 2491–2505.

[3]. Hassan, N., Yau, K. L. A., & Wu, C. (2019). Edge computing in 5G: A review. IEEE Access, 7, 127276–127289.

[4]. Mechalikh, C., Taktak, H., & Moussa, F. (2020). A fuzzy decision tree based tasks orchestration algorithm for edge computing environments. Springer.

[5]. Qian, W., & Coutinho, R. W. L. (2023). A reinforcement learning-based orchestrator for edge computing resource allocation in mobile augmented reality systems. IEEE PIMRC.

[6]. P. Jha, T. Biswas, U. Sagar and K. Ahuja, "Prediction with ML paradigm in Healthcare System," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1334-1342, 2021.

[7]. P. Jha, M. Mathur, A. Purohit, A. Joshi, A. Johari and S. Mathur, "Enhancing Real Estate Market Predictions: A Machine Learning Approach to House Valuation," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 1930-1934, 2025.

[8]. G. K. Soni, D. Yadav, A. Kumar and L. Sharma, "Flexible Antenna Design for Wearable IoT Devices," IEEE 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 863-867, 2023.

[9]. Manish Jha, "A Study of ISA Server for Providing Fast Internet Access with a Single Proxy", SGVU Journal Of Engineering & Technology, Vol. 1, Issue. 1, pp. 15-18, 2015.

[10]. G. K. Soni, H. Arora, B. Jain, "A Novel Image Encryption Technique Using Arnold Transform and Asymmetric RSA Algorithm", Springer International Conference on Artificial Intelligence: Advances and Applications 2019 Algorithm for Intelligence System, pp. 83-90, 2020.

[11]. Manish Kumar Jha, Mr.Gajanand Sharma, Mr.Ravi Shankar Sharma, "Performance Evaluation of Quality of Service in Proposed Routing Protocol DS-AODV", International Journal of Digital Application & Contemporary research, Volume 2, Issue 11, June 2014.

[12]. G. K. Soni, D. Yadav, A. Kumar, P. Jain, A. Rathi, "Design and SAR Analysis of DGS Based Deformed Microstrip Antenna for ON/OFF Body Smart Wearable IoT Applications", Physica Scripta, Vol. 100, Number 1, pp. 1-28, 2025.

[13]. Pradeep Jha, Deepak Dembla, Widhi Dubey, "Implementation of Machine Learning Classification Algorithm Based on Ensemble Learning for Detection of Vegetable Crops Disease", International Journal of Advanced Computer Science & Applications, Vol. 15, Issue. 1, 2024.

[14]. M. k. Jha, R. Ranjan, G. K. Dixit and K. Kumar, "An Efficient Machine Learning Classification with Feature Selection Techniques for Depression Detection from Social Media," 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India, 2023, pp. 481-486,

[15]. Neha Nigam, Neelam soni, "Recent Advances in Internet of Things (IoT): Technologies, Applications, and Challenges", International Journal of Engineering Trends and Applications (IJETA), Vol. 11, Issue. 6, pp. 40-44, 2024.