An Adaptive DTN Routing Protocol Using a Q-Learning Framework for Archipelagic Emergency Networks

Agussalim[1*], Henni Endah Wahanani[2], and Andreas Nugroho Sihananto[3]
[1]Information Technology Department, Faculty of Computer Science,
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Surabaya, Indonesia 60294
[2,3]Informatics Department, Faculty of Computer Science,
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Surabaya, Indonesia 60294
Email: [1]agussalim.si@upnjatim.ac.id, [2]henniendah.if@upnjatim.ac.id,
[3]andreas.nugroho.jarkom@upnjatim.ac.id

***Abstract*—Natural disasters in archipelagic regions often disrupt communication networks, particularly in geographically isolated islands where terrestrial infrastructure is limited and highly vulnerable. Hence, adaptive, infrastructure-independent solutions are required to maintain connectivity during emergencies. The research proposes an adaptive routing protocol for Delay Tolerant Network (DTN), named Q-learning-based Forwarding Routing (QFR), designed to enhance data delivery performance in disaster scenarios characterized by intermittent connectivity and constrained resources. QFR employs a lightweight, tabular Q-learning framework to make intelligent forwarding decisions based on real-time state information, including buffer occupancy, encounter history, and local node density. The protocol further integrates adaptive replica control and priority-based scheduling mechanisms to regulate congestion and optimize bandwidth and buffer utilization. Performance evaluation is conducted using the ONE Simulator with realistic maritime mobility traces derived from vessel movement patterns around Madura Island, Indonesia, representing inter-island emergency communication conditions. The results indicate that QFR consistently outperforms benchmark protocols such as Epidemic and PRoPHETv2, particularly in maintaining a high delivery ratio under heavy traffic loads while keeping routing overhead moderate and latency stable. Time-series analysis further demonstrates QFR's ability to improve its performance over time as the agent learns. The key finding is that a lightweight, adaptive algorithm based on a tabular Q-learning framework provides a practical and effective solution for reliable communication in resource-constrained emergency networks, avoiding the computational complexity of deep reinforcement learning approaches.**

Received: June 26, 2025; received in revised form: Sep. 08, 2025; accepted: Sep. 08, 2025; available online: March 09, 2026.
*Corresponding Author

***Index Terms*—Delay Tolerant Network (DTN), Q-Learning, Disaster Response, Maritime Networks, Adaptive Routing**

I. Introduction

INDONESIA, with over 17,000 islands, faces significant challenges in maintaining reliable communication during natural disasters, particularly in remote maritime regions like Madura Island and its surrounding islets (Sapudi, Ra'as, and Poteran) [1]. These areas often lack robust telecommunication infrastructure, and disasters such as earthquakes and storms further exacerbate connectivity issues. For example, the 2018 Sapudi Island earthquake disrupted communications. It isolated the communities and hindered disaster response [2].

The geological setting further aggravates this situation. The Rembang–Madura–Kangean–Sakala (RMKS) Fault Zone contributes to recurring seismic activity in the region, including the moderate yet disruptive earthquakes recorded in 2018 (Madura with 4.3 Mw and Situbondo with 6.3 Mw) [3]. Even under normal conditions, signal coverage in maritime areas remains uneven, as indicated by independent measurements reported by nPerf [4]. These observations suggest that communication fragility in the region is not solely a disaster-induced phenomenon but also a structural issue tied to geography and infrastructure distribution. Consequently, networking approaches that assume continuous connectivity may not be appropriate for such environments.

Delay Tolerant Network (DTN) provides an alternative perspective by accepting intermittent connectivity as an operational constraint rather than a failure condition. Through the store–carry–forward mechanism, nodes temporarily buffer messages and forward them when encounter opportunities arise. In the Madura maritime setting, regularly operating vessels, like fishing boats and inter-island ferries, follow relatively predictable routes and schedules. These mobility patterns can be repurposed as opportunistic communication carriers, reducing reliance on centralized infrastructure and avoiding costly deployment of additional base stations [5].

DTN-based communication has been explored in disaster response [6], rural internet access [7], Internet of Things (IoT)-assisted healthcare [8], and remote monitoring applications [9]. However, determining which node should forward which message remains a persistent challenge. In maritime environments, node encounters are sparse, buffer capacity is limited, and energy resources are constrained. Traditional routing schemes, such as Epidemic [10], PRoPHET [11], MaxProp [12], and Spray-and-Wait [13], attempt to improve delivery through replication or probabilistic encounter estimation. While effective in certain scenarios, these approaches rely on fixed heuristics that may not adapt well to fluctuating traffic load and irregular contact patterns typical of island-based disaster conditions.

To overcome these limitations, the researchers propose an adaptive routing protocol for DTN that leverages Q-learning, a foundational, model-free reinforcement learning algorithm. The core idea behind Q-learning is to learn a policy that tells an agent which action to take under specific circumstances. It achieves this by learning a state-action value function (the 'Q-function') that estimates the expected future rewards for taking a specific action in a given state. In the context of DTN routing, each node acts as an agent that learns to make optimal forwarding decisions. It learns which encounters are more "valuable" for successful message delivery by continuously updating its Q-values based on feedback from the network environment. The research introduces the Q-learning-based Forwarding Routing (QFR) protocol, which applies Q-learning using a lightweight, tabular approach specifically designed for the dynamic, resource-constrained conditions of disaster-prone archipelagic environments.

A. Related Studies

DTN is communication architectures designed to operate under conditions of intermittent connectivity, long delays, and sparse infrastructure. Their store-carry-forward model enables data to be temporarily held at intermediate nodes until a suitable forwarding opportunity arises. It makes them highly suitable for scenarios such as disaster response, vehicular ad hoc networks, space communications, and rural connectivity [14].

In recent years, the integration of Machine Learning (ML) and Reinforcement Learning (RL) into DTN routing has attracted considerable attention. These approaches aim to move beyond static heuristics by enabling routing decisions to adapt to mobility patterns, network density, and resource availability. Early studies have explored predictive routing strategies using historical delivery data. For example, previous research has utilized delivery history within the IBR-DTN framework to estimate potential next-hop nodes [15], while another research has incorporated geographic prediction to capture the influence of node mobility in vehicular DTN environments [16]. Previous research has also further introduced reinforcement learning-based routing with topic-aware forwarding and congestion-awareness, emphasizing the prioritization of mission-critical data [17].

Subsequent research has investigated more adaptive and context-driven mechanisms. It has proposed Context-Adaptive Reinforcement Learning based routing (CARL-DTN), which combines Q-learning with fuzzy logic to regulate message replication based on contextual and density-related information. This approach demonstrates improved delivery performance and reduces overhead under varying conditions [18]. Another work has surveyed a broad range of ML techniques applied to DTN routing, highlighting the effectiveness of classifiers such as Naïve Bayes, Random Forest, and XGBoost [19]. In more specialized environments, previous research has explored deep reinforcement learning for congestion control in deep-space DTNs, emphasizing energy efficiency and transmission reliability [20].

Further developments have explored ensemble learning and adaptive decision models. Previous research has applied ensemble classifiers to refine routing strategies [21], while another study has proposed real-time ML-based adaptive routing in vehicular DTNs, achieving improvements in both delivery ratio and latency [22]. Research on High-Rate DTN (HDTN) has also incorporated neural networks and Bayesian models to enhance throughput and adaptability under high-load scenarios [23]. In addition, the Baton-relay protocol is introduced. It is a lightweight and privacy-aware routing strategy that leverages mobility context to improve buffer utilization and delivery efficiency [24].

While these approaches have advanced the field, they often present limitations in the context of

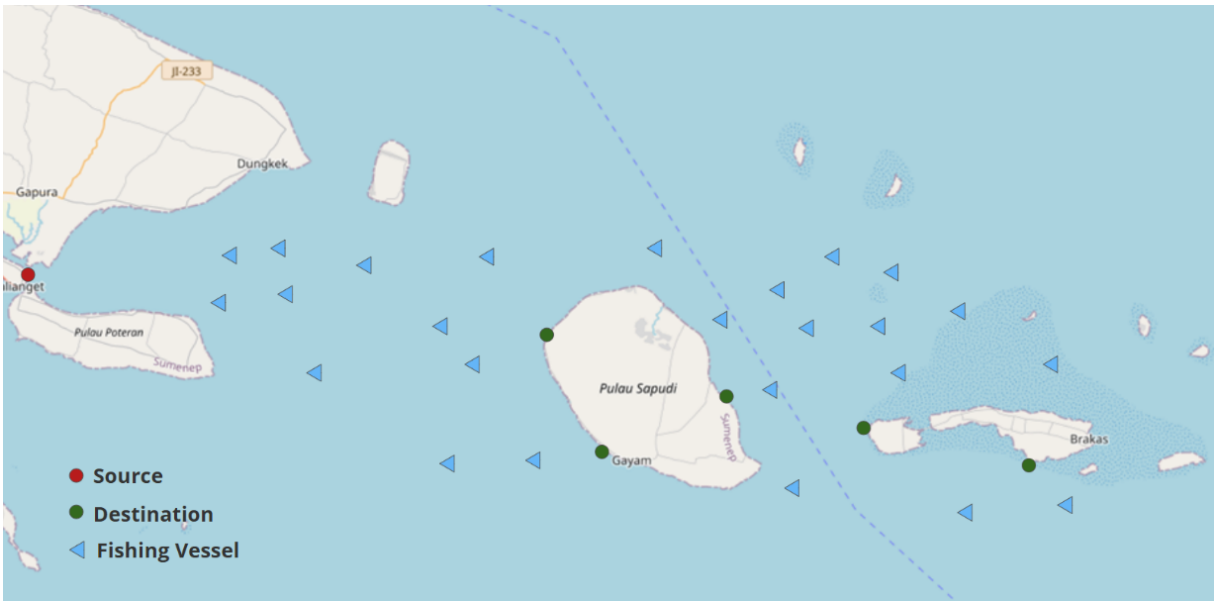

Fig. 1. System model of disaster response scenario.

archipelagic disaster response. For example, deep RL methods, while powerful, typically demand significant computational resources, making them less suitable for the low-power devices expected in an emergency network. Protocols like CARL-DTN may rely on complex metrics such as social context. It can be impractical to compute in sparse maritime networks where interactions are unpredictable. The novelty of the proposed QFR protocol lies in its lightweight RL framework, specifically tailored for maritime disaster scenarios with minimal infrastructure. QFR distinguishes itself from more complex models, such as CARL-DTN, by using a lean state-action model that intentionally avoids complex social or geographic metrics. Instead, its state is defined by immediately available, low-cost data, namely buffer occupancy, encounter history, and local node density, making it highly efficient for resource-constrained environments. This focus on simplicity and practicality is further reflected in its use of a standard Q-learning algorithm with a simple, binary action space (forward or not forward), which minimizes computational overhead. In contrast to deep RL approaches that require more intensive processing, this design choice ensures QFR is practical for deployment on basic IoT devices and mobile nodes. Furthermore, the protocol’s maritime-specific adaptation is evident in its reward system and adaptive replica management, which are explicitly designed to balance delivery reliability with resource conservation in the dynamics of maritime disaster networks, particularly in sparse, unpredictable environments.

II. Research Method

A. System Model

The proposed communication system is modelled as a dynamic network composed of mobile nodes, specifically fishing vessels acting as message carriers (i.e., routers) within a DTN. These mobile nodes facilitate data exchange among islands lacking permanent Internet infrastructure by employing a store-carry-forward mechanism inherent to DTN. As illustrated in Fig. 1, the study area encompasses Madura Island and several smaller surrounding islands in East Java, Indonesia, namely Poteran, Sapudi, and Brakas Islands. These islands are geographically dispersed and particularly vulnerable to communication isolation during natural disasters, such as earthquakes, high tides, and tropical storms. To address this challenge, the researchers integrate IoT sensors deployed on each island to monitor environmental parameters in real time, including sea level, wind speed, and other disaster-related indicators.

The sensor data collected from these IoT devices is transmitted via the DTN infrastructure, where fishing boats serve as opportunistic mobile relays (data mules). These vessels, which naturally traverse inter-island routes as part of their daily operations, are leveraged to carry and forward data between isolated islands and central emergency coordination hubs. By exploiting the semi-regular movement patterns of these boats, the system aims to ensure reliable, energy-efficient data

delivery even in the absence of a fixed communication infrastructure. This architecture provides a scalable, context-aware solution for enabling disaster-resilient communication in remote archipelagic regions.

Formally, the DTNs is represented as a time-varying graph $G(t)$ defined as:

$$G(t) = (V, E(t)). \tag{1}$$

In a network in Eq. (1), V denotes the set of nodes, comprising both mobile nodes (i.e., fishing vessels) and stationary nodes (i.e., island-based IoT gateways). Meanwhile, $E(t)$ represents the set of active communication links at time t, which are inherently time-dependent due to node mobility.

Given the intermittency of inter-node connectivity, the set of edges $E(t)$ is defined dynamically as:

$$E\,(t) - \left\{ \begin{matrix} (u,v)\, u, v \epsilon V, \\ u \neq v, and\, u, v\, in\, range\, on\, time\, t \end{matrix} \right\}. \tag{2}$$

In Eq. (2), each node $v \in V$ is equipped with a finite buffer capacity $B_{\max}$ to temporarily store messages under the store-carry-forward paradigm. Messages are associated with a predefined Time-to-Live (TTL), which they are discarded if not successfully delivered. The message size and the remaining buffer capacity significantly influence forwarding decisions and overall network performance, thereby requiring efficient scheduling and buffer management strategies in resource-constrained environments. The u represents a source node within the set of vertices V. This node may correspond to a mobile node (e.g., a fishing vessel) or a stationary node (e.g., an island-based IoT gateway). Meanwhile, v represents a neighboring (receiving) node within the set of vertices V that has the potential to establish a communication link with node u at time t.

B. Adaptive Reinforcement Learning-based Routing Protocol

The RL provides a framework for developing adaptive behavior in dynamic environments. However, many advanced RL approaches, particularly those based on deep neural networks, require substantial computational resources and are unsuitable for resource-constrained DTN nodes. Therefore, tabular Q-learning is adopted as a lightweight and efficient alternative. The proposed protocol utilizes the Q-learning framework to dynamically optimize routing decisions based on experiential feedback while maintaining low computational overhead.

C. Q-Learning Algorithm

The Q-learning algorithm iteratively updates a state-action value function $Q(s, a)$ stored in a lookup table (Q-table). This function estimates the expected cumulative reward obtained by taking action a in state s. The update mechanism is driven by the Temporal Difference (TD) error, defined as the difference between the current estimate and the newly observed reward. The update rule is expressed as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \Big[r + \gamma \max_{a'} Q(s',a') \; Q(s,a) \Big]. \tag{3}$$

In Eq. (3), s is real-time network state representing perceivable attributes of the surrounding environment and a is a chosen action (forwarding decision). Then, r is immediate reward obtained after executing action a. It also has s' as subsequent state after action execution, $\alpha \in [0, 1]$ as learning rate controlling the magnitude of updates, and $\gamma \in [0, 1]$ as discount factor determining the importance of future rewards.

Here, the term $[\delta = r + \gamma \max_{a'} Q(s', a') Q(s, a)]$ represents the TD error. The parameter α, also known as the learning rate, determines the extent to which this error contributes to updating the Q-value. This iterative process enables the agent to refine its value estimates even in the absence of an explicit environmental model and eventually converge toward an optimal forwarding policy.

In Q-learning, learning is fundamentally error-driven, where the "error" is quantified using the TD formulation. Specifically, the term $[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$ captures the discrepancy between the current estimate of the state-action value $Q(s, a)$ and the newly observed estimate based on the immediate reward r and the predicted future value. The learning rate α regulates how much the previous Q-value is adjusted in response to this discrepancy.

Through repeated updates, the Q-learning process performs an incremental optimization analogous to gradient descent. It aims to minimize the TD error and progressively improve the accuracy of value estimation. Consequently, the agent is able to learn an effective forwarding policy and converge toward the optimal solution over time.

D. State Representation

The state s is defined as a feature vector that integrates relevant networking and node conditions to facilitate forwarding decisions. Formally, the state is represented as:

$$s = (M, E, D). \tag{4}$$

In Eq. (4), M denotes the number of messages currently stored in a node, representing the buffer status

and indicating the level of network congestion. The parameter E corresponds to the historical encounter count with a prospective forwarding node and serves as a proxy for delivery predictability. Finally, D represents the local node density, defined as the number of neighboring nodes available for forwarding in the surrounding area. By jointly incorporating buffer occupancy, contact history, and local topology, the proposed state representation enables the routing agent to make informed forwarding decisions under dynamic and resource-constrained network conditions.

E. Action Space

To optimize the decision-making process at each node, the protocol restricts the action space to a binary decision set. Specifically, the agent can either forward the message to a neighboring node or retain it. Formally, the action space is defined as:

$$a \in \{0, 1\}. \tag{5}$$

In Eq. (5), $a = 1$ denotes forwarding the message to a neighboring node, and $a = 0$ denotes not forwarding the message. This deliberate simplification of the decision structure reduces computational complexity and accelerates the learning process of the RL agent. Consequently, the agent can efficiently learn the value of forwarding decisions and progressively improve overall message delivery performance.

F. Reward Design

The reward function is designed to encourage successful message delivery while penalizing failures and excessive network congestion. A positive reward is assigned when a message is successfully forwarded and received, whereas negative rewards are imposed for omission and unsuccessful transmissions. Formally, the reward function is defined as:

$$r = \begin{cases} +5, & \text{if a message forwarded and received,} \\ -3, & \text{if forwarding fails,} \\ -0.2D, & \text{if } D > 5. \end{cases} \tag{6}$$

In Eq. (6), D denotes the local node density. The density-dependent penalty is introduced to mitigate congestion in highly dense network regions, which may lead to increased collisions, packet loss, and degraded overall network performance. By incorporating both delivery success and congestion awareness, the reward mechanism enables the routing agent to learn a balanced forwarding strategy. Consequently, the system converges toward an equilibrium state in which resource consumption is minimized while maintaining effective multi-hop connectivity and delivery performance.

G. Exploration–Exploitation Trade-off

To balance exploration and exploitation during routing decisions, an ϵ-greedy strategy is employed for action selection. The exploration probability decreases over time according to:

$$\epsilon_{t+1} = \max\left(\epsilon_{\min}, \epsilon_t \times \delta\right). \tag{7}$$

In Eq. (7), $\epsilon_{\min}$ denotes the minimum exploration rate, and $\delta \in (0, 1)$ represents the decay factor. This mechanism allows the agent to initially explore the action space through random decisions and gradually shift toward exploiting learned optimal forwarding policies as training progresses.

H. Adaptive Replica Management

To improve the efficiency of network resource utilization, the proposed protocol employs a dynamic replica management mechanism that adjusts the maximum number of message replicas according to the local node density D. The replica limit is adapted as follows:

$$R_{\max}(D) = \begin{cases} 3, & D > 10, \\ 5, & 5 < D \leq 10, \\ 10, & D \leq 5. \end{cases} \tag{8}$$

In Eq. (8), $R_{\max}$ denotes the maximum allowable number of message replicas. In high-density scenarios ($D > 10$), the system restricts replication to three copies to mitigate bandwidth contention and buffer saturation. Adaptive replication control based on contextual network conditions has been shown to improve efficiency in DTN environments [18]. For moderate-density conditions ($5 < D \leq 10$), up to five replicas are permitted to balance delivery reliability and transmission overhead. In sparse networks ($D \leq 5$), the protocol allows up to ten replicas to compensate for infrequent encounters and intermittent connectivity. The numerical thresholds adopted in the research are designed and evaluated empirically through simulation.

I. Message Scheduling and Buffer Management

Message scheduling focuses on the order of relay activities based on whether messages are likely to be delivered within their assigned TTL. A priority metric is defined as Eq. (9). Messages with higher priority values, indicating longer traversal relative to

the remaining TTL, are transmitted first to reduce the risk of expiration.

$$Priority(m) = \frac{HopCount(m)}{TTL(m)}. \quad (9)$$

For buffer management, a drop-oldest policy is adopted. When the buffer reaches its capacity, the message with the smallest remaining TTL (i.e., closest to expiration) is discarded to accommodate incoming messages with higher delivery potential. This strategy ensures efficient utilization of limited storage resources while maintaining delivery performance under constrained network conditions.

III. Results and Discussion

To evaluate the proposed protocol, the ONE Simulator [25], a widely accepted and standard tool for evaluating DTN routing protocols, is used. Its primary advantage provides a controlled, repeatable environment, which is essential for isolating the performance of the routing algorithm from external, uncontrollable variables. While a real-world deployment is more complex, this simulation environment is appropriate for a rigorous comparative analysis of protocol logic.

Given the disaster-response focus of the research, the simulation scenarios are designed to closely mirror actual maritime operations around Madura Island and its neighboring islands, particularly Sapudi and Ra'as, using the VesselFinder application (https://www.vesselfinder.com/), as shown in Fig. 2. Moreover, for slight fishing vessel movements that are not detected by VellesFinder, the movement patterns are derived from structured interviews with local fishermen, reflecting real-world navigational behavior. These patterns are implemented using the Map-Based Movement model provided by the ONE Simulator. Geographic data are integrated using actual maps of the study area to ensure spatial accuracy in vessel routes, as illustrated in Fig. 3.

The simulation is run for 24 hours, a timeframe selected to be sufficient to capture the typical daily operational cycles of the fishing vessels as mobile nodes. This duration allows for a significant number of opportunistic encounters, providing enough data for the QFR protocol to demonstrate its adaptive capabilities and for performance differences between protocols to become evident. The network traffic is generated by sending messages between nodes across the islands, simulating the need for inter-island communication during an emergency. The randomness in the simulation arises not from client routing, but from the inherently opportunistic and unpredictable nature of node encounters due to mobility, which is the core challenge in DTN.

TABLE I
Simulation Parameters.

Parameter	Value
Simulation Duration	24 Hours
Number of Fishing Vessels	50 Nodes
Wi-Fi Transmission Range	300–500 m
Wi-Fi Transmission Speed	563 kbps
Message TTL	720 s
Vessel Buffer Size	10 MB
Generated Messages	1000–5000
Warm-Up Duration	1800 s
Routing Protocols	Epidemic, QFR, PRoPHET V2, SNHD, SNW
L Value (SNHH & SNW)	9 messages
Learning Rate (α)	0.1
Discount Factor (γ)	0.9
Exploration Rate (ϵ)	0.3

Note: Time-to-Live (TTL), Q-learning-based Forwarding Routing (QFR), Spray and Hop Distance (SNHD), and Spray-and-Wait (SNW).

Table I summarizes the simulation parameters. Their configuration is intended to approximate realistic operational conditions typically encountered in maritime disaster-response scenarios, where bandwidth is limited, storage capacity is constrained, and node availability is highly intermittent. Such constraints are deliberately incorporated to ensure that the evaluation reflects practical deployment challenges rather than idealized network assumptions.

The configuration of the Q-learning mechanism focuses on three principal parameters: the learning rate (α), the discount factor (γ), and the exploration rate (ϵ). These parameters are not selected arbitrarily but are refined through iterative experimentation to achieve stable learning behavior under dynamic network conditions. The learning rate (α) is set to 0.1 to prevent abrupt policy shifts and support gradual adaptation during early training. The discount factor (γ) is set to 0.9 to emphasize long-term delivery outcomes, which are particularly relevant in sparse DTN environments where successful forwarding often depends on delayed opportunities.

To maintain a balance between exploration and exploitation, the exploration rate (ϵ) is initially set to 0.3 and then reduced by a factor of 0.995. This configuration encourages broader exploration during the initial phase of learning while allowing the routing policy to stabilize as the agent accumulates experience. Such a strategy is essential in DTN environments, where routing decisions must continuously adapt to changing mobility patterns and encounter dynamics. The sensitivity of the proposed routing mechanism to variations in these parameters is further analyzed in the discussion section.

Next, performance evaluation is conducted using key metrics that reflect both the efficiency and reliability

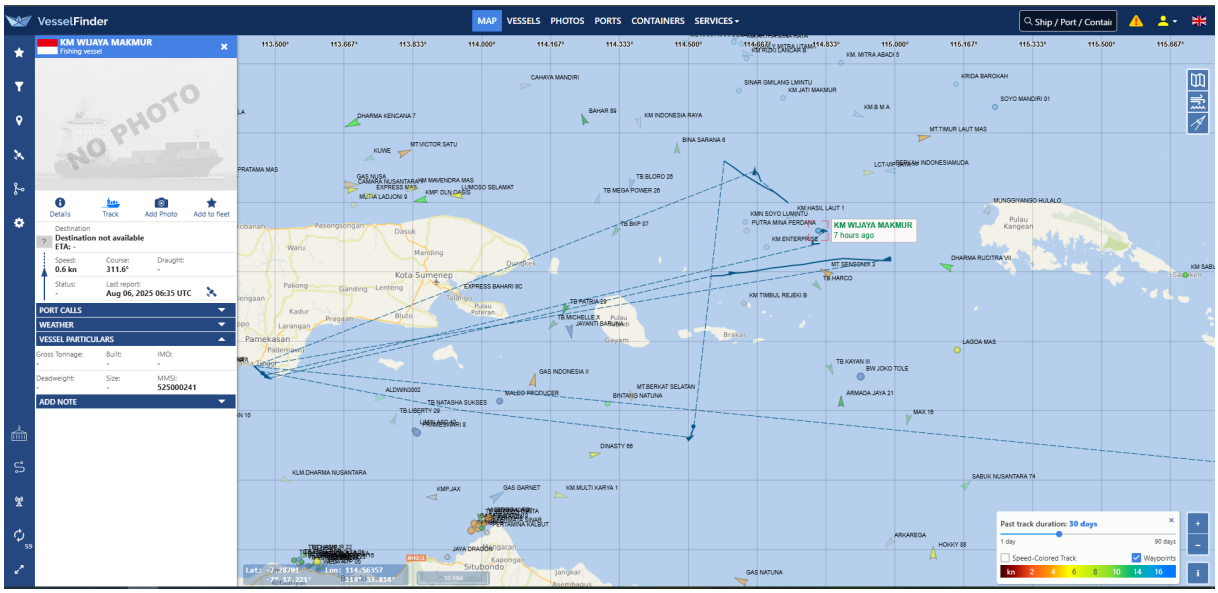

Fig. 2. Fishing vessel movement based on Vesselfinder application (https://www.vesselfinder.com/).

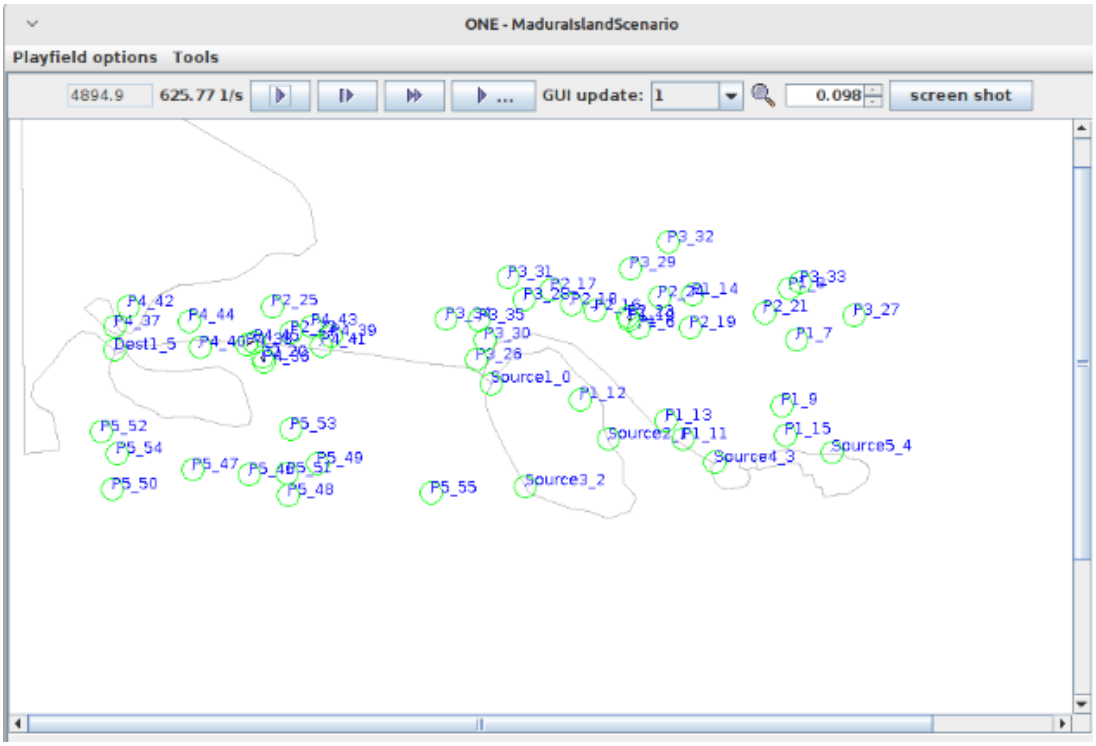

Fig. 3. Implementation of Madura Island Map into the ONE Simulator.

of the routing protocol in the context of emergency communications. The proposed QFR protocol is compared against widely adopted DTN routing schemes, including Epidemic, PRoPHET V2, Spray-and-Wait, and Spray and Hop Distance (SNHD) [26]. The following performance indicators are used to examine routing behavior.

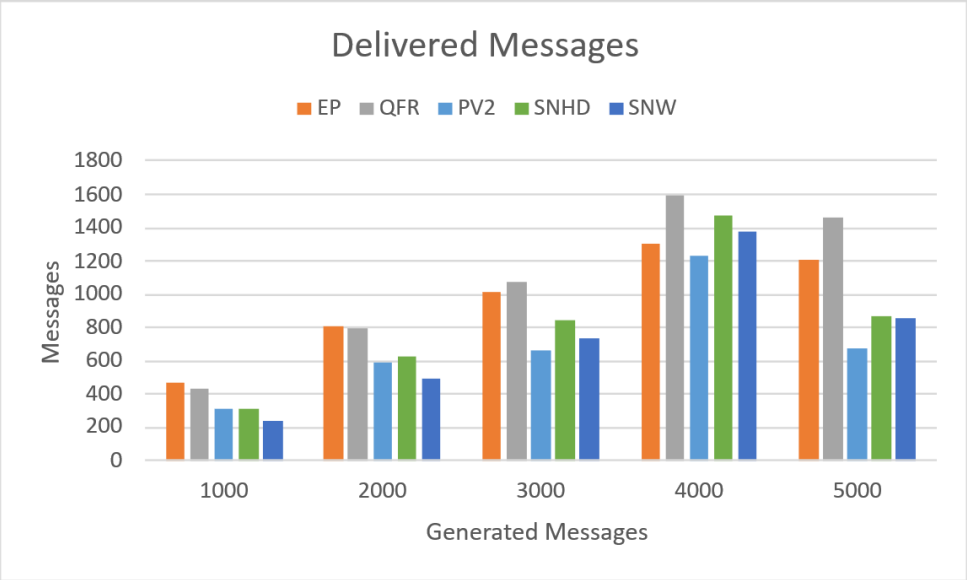

Fig. 4. Message delivery ratio under varying traffic loads. Note: Epidemic (EP), PRoPHETv2 (PV2), Q-learning-based Forwarding Routing (QFR), Spray-and-Wait (SNW), and Spray and Hop Distance (SNHD).

A. *Delivery Ratio*

The ability to deliver messages despite network challenges is a fundamental performance indicator in DTN environments. The simulation results, as depicted in Fig. 4, clearly indicate that QFR demonstrates a remarkable capability for maintaining a high delivery ratio, especially when network congestion intensifies. While protocols like Epidemic show initial promise in low-traffic scenarios, they experience a catastrophic performance collapse under heavy loads. In contrast, QFR not only sustains its performance but consistently outperforms all benchmark protocols in high-traffic environments, establishing itself as a highly reliable and scalable solution.

The primary reason for QFR’s superior performance lies in its intelligent and adaptive forwarding strategy, which is powered by the Q-learning algorithm. Unlike the indiscriminate “blind flooding” approach of Epidemic routing, QFR enables each node to learn from its interactions with other nodes. By analyzing the state vector in Eq. (4), a node can make an informed judgment about whether forwarding a message to a neighbor is a high-value action. The agent learns to prioritize nodes with a strong historical encounter rate (E), as these nodes are statistically more likely to carry the message closer to its destination. This experience-based decision-making process ensures that network resources are used purposefully.

Furthermore, QFR’s design incorporates a crucial congestion avoidance mechanism that is vital for scalability. The protocol’s reward function is engineered to actively discourage behaviours that lead to network saturation. When a node detects that the local density (D) of neighbouring nodes is high, a penalty is applied for forwarding actions, teaching the agent to hold back its messages to prevent overwhelming the network. It stands in direct opposition to the Epidemic protocol, which continues to flood messages regardless of network conditions, leading to severe buffer overflows and widespread message loss. QFR’s ability to recognize and adapt to network congestion is therefore a key factor in its robust delivery performance under stress.

B. *Overhead Ratio*

The overhead ratio provides critical insight into the efficiency and resourcefulness of a protocol, quantifying the amount of redundant network traffic generated for each successful delivery. The results in Fig. 5 highlight a clear distinction in efficiency among the protocols. Epidemic and PRoPHETv2 are shown to be highly inefficient, generating an enormous amount of overhead that grows with network traffic. In contrast, QFR maintains a controlled and moderate level of overhead, proving to be significantly more efficient than its flooding-based counterparts.

QFR’s efficiency is an emergent property of two core design principles: a reward system that incentivizes resource conservation and an explicit mechanism for replica control. The Q-learning agent’s fundamental objective is to maximize its cumulative reward, which is only granted upon the final, successful delivery of a message. There is no reward for simply relaying a

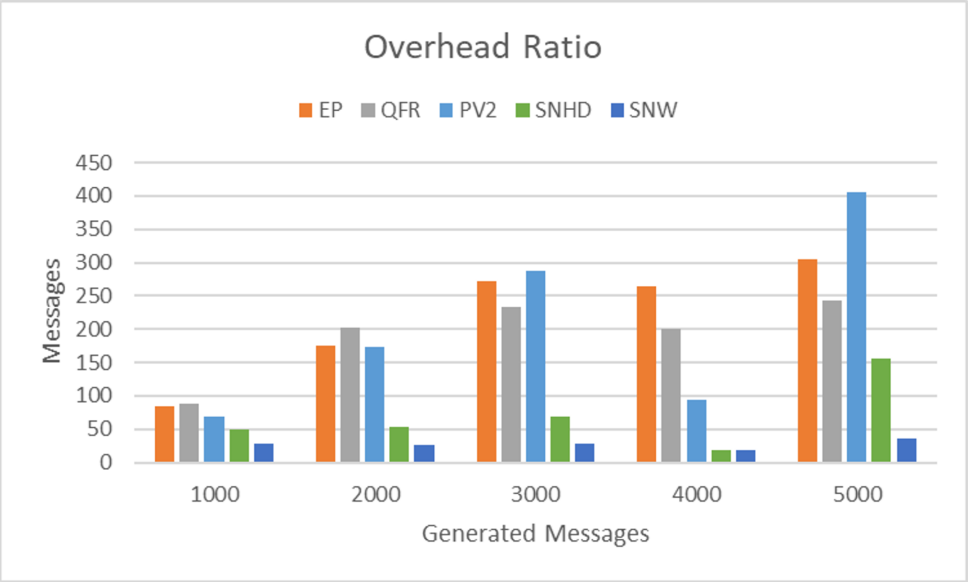

Fig. 5. Overhead ratio for each protocol under varying traffic loads. Note: Epidemic (EP), PRoPHETv2 (PV2), Q-learning-based Forwarding Routing (QFR), Spray-and-Wait (SNW), and Spray and Hop Distance (SNHD).

message multiple times. Consequently, the agent learns over time that indiscriminate, redundant transmissions are wasteful actions that consume resources without contributing to its goal. This learned behavior naturally suppresses the tendency to flood the network.

This learned resourcefulness is powerfully augmented by the adaptive replica management algorithm, which acts as a hard backstop against excessive transmissions. This mechanism dynamically adjusts the maximum number of message copies based on real-time local node density. For example, in a densely populated area where it is $D > 10$, the protocol strictly limits a message to only three replicas. This explicit, context-aware rule operates in tandem with the learned policy to ensure that overhead is kept low. Therefore, it prevents the network-clogging behavior that renders flooding-based protocols inefficient and impractical in resource-constrained environments.

C. Average Latency

Average latency measures the end-to-end delay of messages, a critical factor in time-sensitive emergency communications. The simulation results in Fig. 6 reveal that QFR offers a balanced and stable latency performance. While a protocol like Spray-and-Wait can occasionally achieve lower latency due to its direct delivery strategy, QFR consistently outperforms the high-latency Epidemic and PRoPHETv2 protocols, demonstrating greater stability across varying traffic loads.

QFR’s moderate latency is the result of an intelligent trade-off between delivery speed and delivery reliability. Instead of forwarding a message to the very first node it encounters, the QFR agent may decide to wait for a more optimal opportunity. It uses its learned Q-values to assess whether a current neighbor is a truly reliable forwarder (e.g., one with a high historical encounter value). Holding a message slightly longer to find a better carrier may incrementally increase its latency. However, this strategy substantially increases its overall probability of reaching the destination, avoiding the much longer delays associated with inefficient, multi-hop paths.

To ensure that this calculated waiting does not result in excessive delays, QFR integrates a priority-based message scheduling system. This mechanism prioritizes messages within the buffer based on the ratio of their Hop Count to their remaining TTL, formulated in Eq. (9). Messages that are older or have already travelled far are given higher priority for transmission. This proactive scheduling prevents urgent messages from being delayed by newer ones, ensuring a smooth flow of data and effectively managing the trade-off between cautious forwarding and timely delivery.

D. Average Buffer Time

The average buffer time indicates how long messages reside in the memory of intermediate nodes, a key indicator of buffer utilization and network fluidity. As shown in Fig. 7, protocols that rely on a “wait” strategy, such as Spray-and-Wait, exhibit extremely

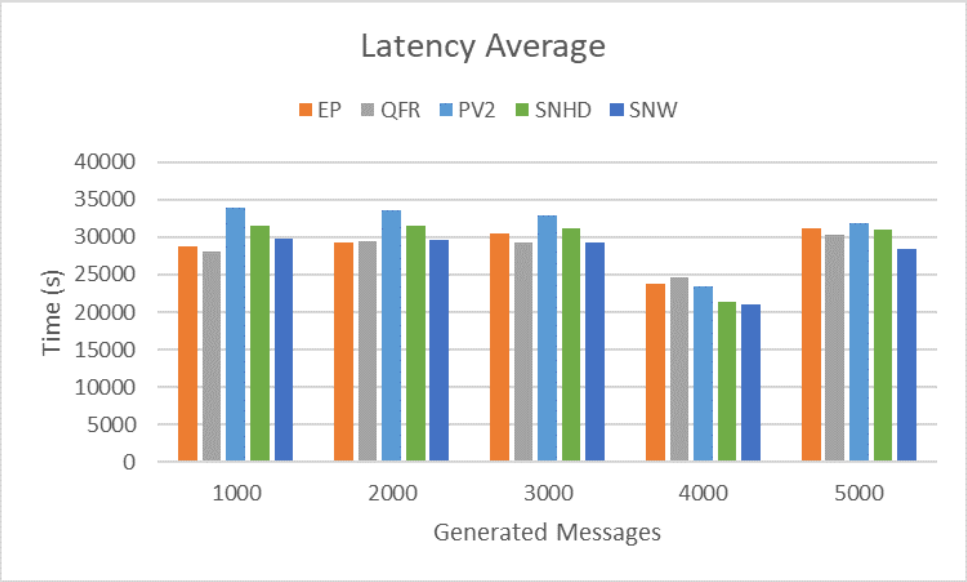

Fig. 6. Average message latency under varying traffic loads. Note: Epidemic (EP), PRoPHETv2 (PV2), Q-learning-based Forwarding Routing (QFR), Spray-and-Wait (SNW), and Spray and Hop Distance (SNHD).

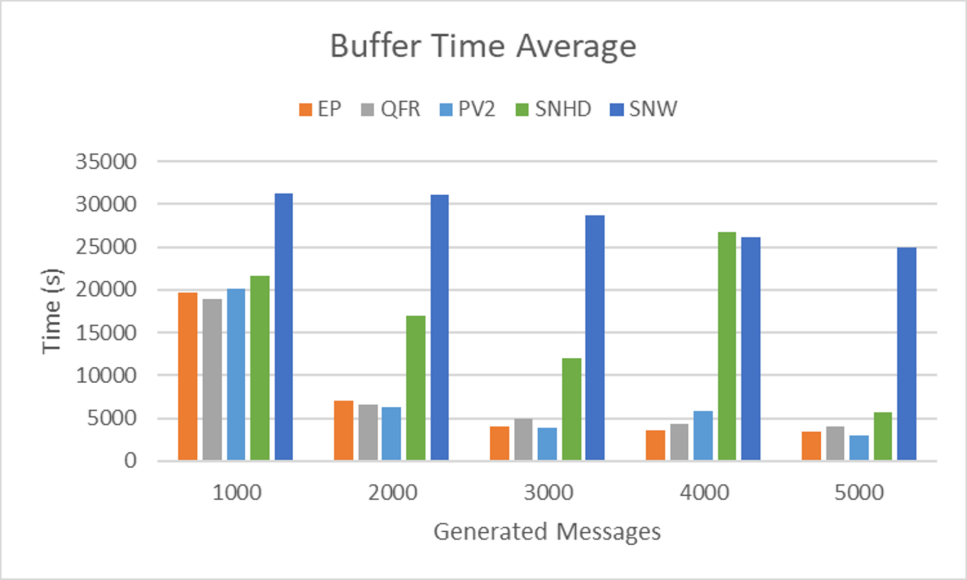

Fig. 7. Average buffer time under varying traffic loads. Note: Epidemic (EP), PRoPHETv2 (PV2), Q-learning-based Forwarding Routing (QFR), Spray-and-Wait (SNW), and Spray and Hop Distance (SNHD).

long buffer times as they hold messages until they meet the final destination. Conversely, QFR consistently demonstrates one of the lowest average buffer times, particularly in high-traffic scenarios, highlighting its efficiency in managing this critical resource.

QFR’s ability to maintain low buffer times is a direct consequence of its dynamic and efficient message handling. The Q-learning agent is incentivized to find effective forwarding opportunities quickly. It learns to avoid accepting messages that it is unlikely to deliver and actively seeks out reliable nodes to pass messages to. This behavior creates a continuous and efficient “flow” of data through the network, preventing messages from becoming stagnant in any single node’s buffer for extended periods.

QFR’s proactive buffer management policies further support this fluid routing. The protocol employs a drop-oldest policy, where a message with the smallest

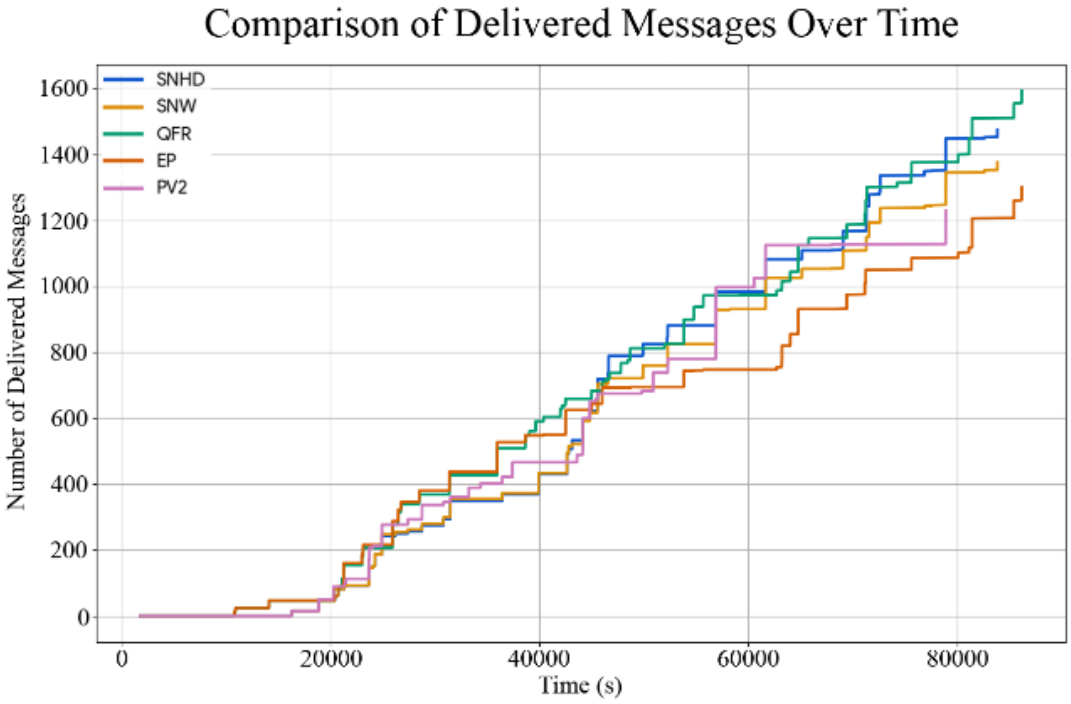

Fig. 8. Performance of routing protocols on delivered messages over time. Note: Epidemic (EP), PRoPHETv2 (PV2), Q-learning-based Forwarding Routing (QFR), Spray-and-Wait (SNW), and Spray and Hop Distance (SNHD).

remaining TTL is discarded to make space when the buffer is full. It is combined with the priority scheduling mechanism, which works to transmit high-priority messages out of the buffer as quickly as possible. Together, these strategies ensure that buffer space is utilized efficiently, turnover is high, and the risk of message loss due to buffer overflow, a critical problem in other protocols, is significantly minimized.

E. Time-Series Performance Analysis

Figure 8 shows a comparison of the cumulative number of messages delivered as a function of simulation time over 24 hours (86,400 seconds), with a total of 4,000 generated messages. This visualization dynamically illustrates how each protocol behaves and adapts throughout the entire scenario. From the plotted curves, it is evident that while several protocols show initial strength, QFR and SNHD consistently emerge as the most resilient and effective protocols by the end of the simulation, with QFR ultimately delivering the highest number of messages.

The QFR curve (teal line) reveals the hallmark characteristics of a learning-based algorithm. In the initial stages of the simulation (i.e., before 20,000 seconds), QFR’s performance does not significantly differ from that of the other protocols. It can be attributed to the exploration phase of the Q-learning algorithm, where the agent is actively trying various actions, including random ones, to gather data and “learn” about the network environment. However, an apparent acceleration in QFR’s performance curve occurs as time progresses, particularly in the latter half of the simulation (after approximately 50,000 seconds). This increasing rate of delivery is visual proof that the learning process has been successful. The Q-table has been populated with accurate values, and the agent begins to shift into the exploitation phase, consistently making intelligent forwarding decisions based on its accumulated knowledge. This ability to improve its effectiveness over time is what distinguishes QFR as a truly adaptive system.

QFR’s dynamics become even more apparent when contrasted with the other protocols. The Epidemic protocol (orange line), for instance, shows a powerful initial performance due to its aggressive message replication. However, its curve is marked by several long periods of stagnation (flat lines), which represent moments when the network has become saturated, leading to buffer overflows and delivery failures. This highlights the weakness of a non-adaptive strategy that cannot handle network congestion. Similarly, the performance of PRoPHETv2 (pink line), while adaptive, appears to fade over time, suggesting its probabilistic model may not adapt quickly enough to the highly

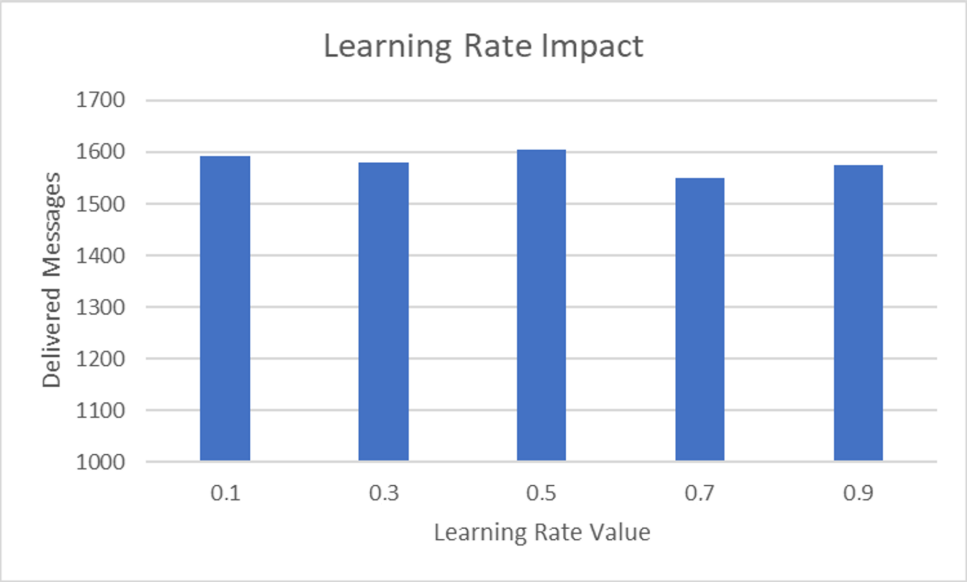

Fig. 9. Learning rate evaluation of Q-learning-based Forwarding Routing (QFR) routing.

dynamic node encounters in this scenario. Therefore, this graph not only confirms QFR's superiority in total deliveries but, more importantly, proves that its strength lies in its ability to learn and dynamically optimize its strategy, making it a uniquely robust solution for long-duration and unpredictable network conditions.

F. Learning Rate (α) Evaluation

A simulation study with a constant workload of 4,000 messages is designed to explore the QFR protocol's sensitivity in learning rate (α) variations. Five distinct learning rates are chosen to represent the range of cautious to aggressive learning: 0.1, 0.3, 0.5, 0.7, and 0.9. The primary performance metric, like before, is the number of messages delivered.

Figure 9 summarizes the performance results, which demonstrate the apparent sensitivity of the QFR protocol's performance to learning rate. In particular, the peak delivery performance of QFR is observed with a learning rate of α= 0.5, which results in slightly less than 1,600 messages being delivered. This result indicates that a balanced rate enables a Q-learning agent to successfully balance exploration of new forwarding strategies with exploitation of previously effective ones. In a rapidly changing network environment, such adaptation of routing policies can significantly improve the reliability of message delivery.

Learning rates of α= 0.1 and α= 0.3 also perform strongly, yielding a comparable number of messages, albeit slightly lower than α= 0.5. This result suggests that more conservative learning rates can still allow for robust performance, as the agent steadily updates its Q-values, which in turn facilitate learning and lead to consistent and steadfast forward decisions. The small gap in performance between lower learning rates and α= 0.5 supports the QFR protocol's versatility and effectiveness across a broader spectrum of conservative to moderate learning behaviors.

In comparison, increased learning rates of α= 0.7 and α= 0.9 result in a sharp decrease in performance. At these rates, the number of messages delivered decreases. This result suggests that exceedingly high learning rates may result in unstable Q-value updates. With high learning rates, the agent tends to "overreward" recent interactions, resulting in highly volatile policies that make effective generalization impossible. In this case, the delivery outcomes will be more variable and less reliable as the agent attempts to use unstable forwarding strategies.

The assessment verifies that the QFR protocol performs adequately across a range of learning rates. Even so, its efficiency can be enhanced, sometimes dramatically, by adjusting the α parameter. Of the values tested, α= 0.5 is the best option. This learning rate maintains an equilibrium between the rigidity of learning and the flexibility required to forward messages in environments such as DTN, which has sporadic connections and minimal infrastructure. Hence, α= 0.5 is the optimal choice for these challenging conditions.

G. Discount Factor (γ) Evaluation

To further understand the behavior of the QFR protocol, an evaluation is conducted to analyze its sensitivity

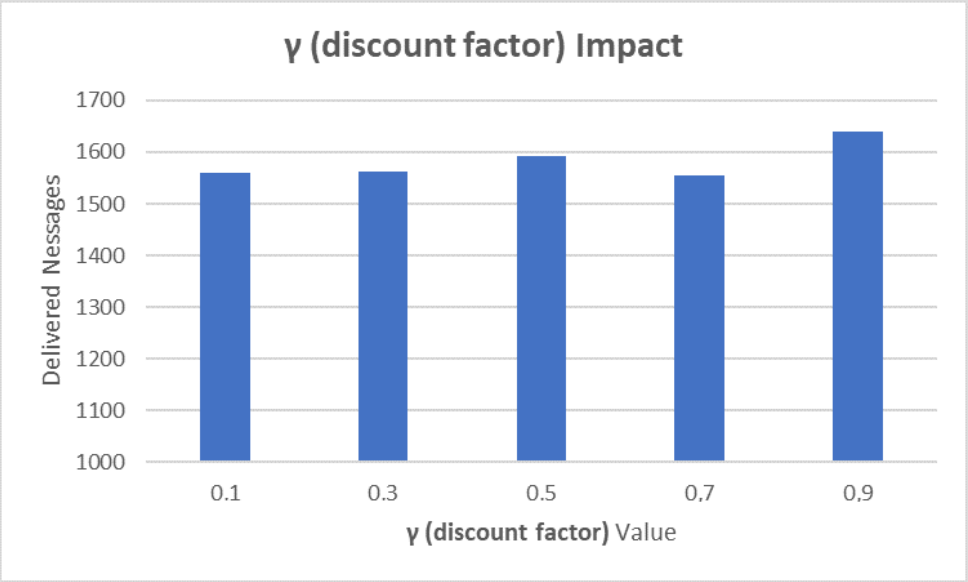

Fig. 10. Discount factor evaluation of Q-learning-based Forwarding Routing (QFR).

to the discount factor, γ, as shown in Fig. 10. This parameter is fundamental to the Q-learning algorithm, as it determines the extent to which future rewards are valued in comparison to immediate ones. The experiment is performed under a fixed network load of 4,000 messages and with the previously optimized learning rate (α= 0.5) to isolate the impact of γ. The graph displays the total number of delivered messages for five distinct γ values ranging from 0.1 to 0.9. The results reveal two key findings. First, the QFR protocol demonstrates considerable robustness, performing well across the entire spectrum of tested values. Second, a clear trend emerges, indicating that higher values of γ yield superior performance, with the peak delivery ratio achieved at γ= 0.9.

The observed trend, where performance improves as γ approaches 1, can be attributed to the nature of the DTN environment. A low discount factor (e.g., γ= 0.1) makes the learning agent "myopic," causing it to heavily prioritize immediate rewards over long-term gains. In a DTN context, it translates to a strategy of forwarding messages to any available node as quickly as possible, without sufficient consideration for that node's actual utility in the delivery chain. Conversely, a high discount factor (e.g., γ= 0.9) makes the agent "farsighted". It learns that the significant, cumulative reward of a successful future delivery is more valuable than any small, immediate benefit from a premature hand-off. It encourages the agent to adopt a more patient and strategic policy, such as carrying a message for a longer duration until it encounters a highly reliable relay node, which is a crucial strategy for success in sparse and disconnected networks.

Beyond the optimal performance at γ= 0.9, the graph also highlights the protocol's impressive robustness. Even at the lowest discount factor of γ= 0.1, QFR delivers over 1,550 messages. The high number indicates that the protocol does not suffer a catastrophic failure even with a sub-optimal parameter setting. This resilience suggests that other mechanisms within the QFR framework, such as the adaptive replica management and the use of encounter history in the state representation, provide a strong baseline performance that mitigates the adverse effects of a myopic policy. This robustness is a significant practical advantage. It implies that QFR can be deployed effectively in real-world scenarios without requiring exhaustive and precise parameter tuning, making it a more reliable and flexible solution for dynamic emergency networks.

H. Exploration Decay (δ) Evaluation

Figure 11 illustrates the evaluation of the exploration decay factor, δ, on the performance of the QFR protocol. This parameter governs the speed at which the agent transitions from an initial phase of exploration (taking random actions to learn) to a phase of exploitation (using the best-known strategies to maximize rewards). The experiment is run with a fixed network load and the previously determined optimal values for the learning rate (α= 0.5) and discount factor (γ= 0.9). The graph presents the total delivered messages across five different δ values, from a relatively fast decay (0.9) to a very slow decay (0.999). The most striking takeaway from these results is the remarkable

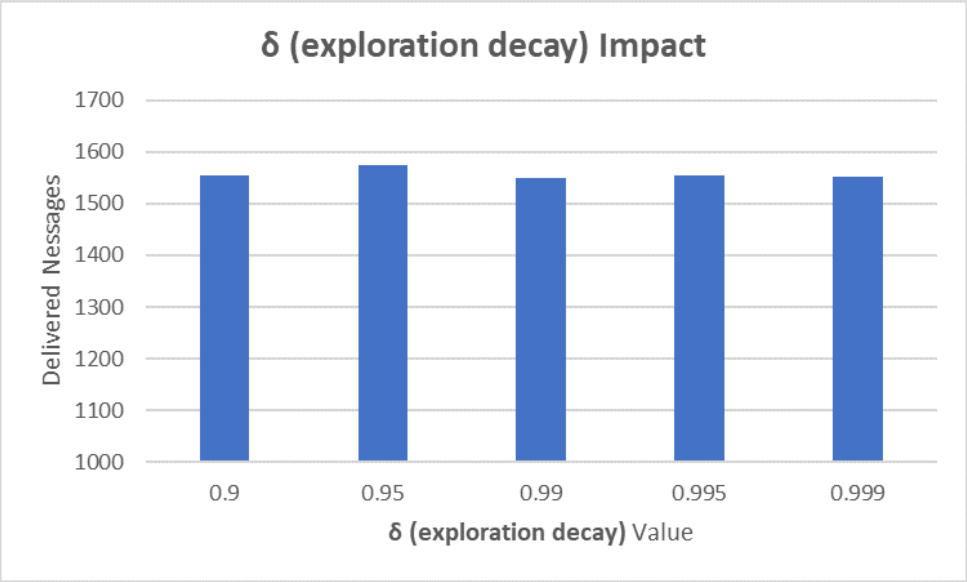

Fig. 11. Exploration decay evaluation of Q-learning-based Forwarding Routing (QFR).

robustness of the QFR protocol. At the same time, a slight peak performance is observed at δ= 0.95. The overall performance remains highly consistent and effective across the entire range of tested values.

The subtle variations in performance reveal the nuances of the exploration-exploitation trade-off within this specific DTN environment. The peak at δ= 0.95 suggests that this value strikes an optimal balance for the 24-hour simulation period. It provides a sufficiently long initial exploration phase for the agent to discover a diverse set of effective forwarding routes within the dynamic network. Subsequently, the exploration rate (ε) decays at a moderate pace, allowing the agent to shift decisively towards exploiting its accumulated knowledge for the majority of the simulation, thereby maximizing the number of successful deliveries. The slightly lower performance at very slow decay rates (δ= 0.99 and 0.999) indicates that prolonged exploration, while helpful in adapting to volatility, introduces a degree of inefficiency by causing the agent to continue making random, sub-optimal decisions even after a good policy has been found.

Ultimately, the most significant finding from this analysis is the protocol's high degree of stability and its low sensitivity to the precise tuning of the decay parameter. The fact that the difference between the best and worst outcomes is minimal (less than 2%) demonstrates the robustness of the overall QFR framework. This result implies that the core learning mechanism, driven by the state representation and reward function, is powerful enough to guide the agent towards an effective policy regardless of the specific exploration schedule. It is a highly desirable characteristic for a real-world protocol, as it suggests that QFR does not require meticulous and fragile parameter tuning to function effectively. This inherent stability increases its practical value, making it a reliable and deployable solution for unpredictable emergency communication scenarios.

I. Computational Overhead Analysis

A critical factor for protocols in resource-constrained environments is their computational overhead. QFR is intentionally designed to be lightweight. The protocol's core logic relies on a tabular Q-learning approach, where the primary computational task is updating a Q-table. The complexity of this operation is directly tied to the size of the state-action space.

To ensure feasibility on low-power devices, the state representation is kept deliberately simple and discrete. For instance, the number of messages and encounters is capped during state mapping to limit the total number of states. The action space is a simple binary choice (forward/do not forward), further constraining complexity. As a result, the Q-table remains small and manageable, and the update calculation, as shown, is computationally inexpensive, involving only basic arithmetic operations. This low overhead makes QFR significantly more practical for the TTGO LoRa32 V2.1 ESP32, as implemented in [27], compared with CARL-DTN, which requires the computational power for neural network forward passes.

IV. Conclusion

The research develops QFR, a Q-learning-based forwarding protocol for DTN in disaster-prone archipelagic regions. It adapts its routing decisions by leveraging buffer occupancy, encounter history, and network density to overcome intermittent connectivity. Simulations confirm its reliability for emergency communications, showing that QFR achieves a superior delivery ratio of up to 1602 messages at α= 0.5, with moderate overhead and stable latency. However, these findings are based on simulations, which cannot fully capture the complexities of a real-world deployment, thus presenting several avenues for future work. To build upon these promising results, the future efforts will focus on addressing real-world constraints, such as hardware limitations and energy consumption, through a real-world implementation using TTGO LoRa32 V2.1 ESP32 devices with DTN7. It will involve developing energy-aware routing logic, for instance, by modifying the reward function to penalize energy-intensive transmissions, thereby making the protocol sustainable on battery-powered nodes.

Furthermore, future work will focus on mitigating dynamic environmental factors, such as severe weather or signal interference, which are not previously modeled, by incorporating environmental data into the state representation to enable more robust decision-making. A more thorough investigation into the impact of the discount factor (γ) and exploration decay (δ) is also planned to enhance the protocol's adaptability across a broader range of disaster scenarios. In addition, the learning rate (α) will be explored as an adaptive parameter, allowing it to change dynamically in response to network stability. Through these directions, QFR is expected to evolve from a robust simulation framework into a field-tested solution for enhancing disaster resilience in isolated archipelagic communities.

Acknowledgement

This research was supported by a grant from the LPPM of Universitas Pembangunan Nasional "Veteran" Jawa Timur in 2025 under Grant No. SPP/104/UN.63.8/LT/V/2025. The authors gratefully acknowledge the financial support provided by Universitas Pembangunan Nasional "Veteran" Jawa Timur for the completion of this research.

Author Contribution

Designed the Q-learning-based routing framework and overall research methodology, A.; Implemented the QFR protocol and conducted simulations using the ONE Simulator, A.; Drafted the manuscript, prepared figures, and interpreted the experimental results, A.; Prepared and configured maritime mobility traces and simulation parameters, A.; Contributed to the research design and validation framework, H. E. W.; Reviewed and revised the manuscript for technical accuracy and clarity, H. E. W.; Provided academic supervision and methodological guidance, H. E. W.; Provided guidance on simulation environment setup and evaluation metrics, A. N. S.; Assisted in performance evaluation and result verification, A. N. S.; Contributed to manuscript revision and critical feedback, A. N. S.; and Provided research supervision and strategic direction, A. N. S.

Data Availability

The data that support the findings of this study are available from the corresponding author, Agussalim, upon reasonable request. The data are not publicly available because they are part of an ongoing research project and will be used in future publications.

References

[1] I. Desportes, W. Wicaksono, and M. Voss, "Disaster cultures—Indonesia and its tsunami warning system," 2024. [Online]. Available: https://hal.science/hal-04829233v1/file/DisasterCultures_Indonesia_KFS_WP32.pdf

[2] E. T. Paripurno, D. N. Yalesrie, A. A. Al-Kudus, Y. N. Maharani, A. R. B. Nugroho, N. E. Nugroho, J. Purwanta, G. A. Pratama, G. Mahojwala, and W. Putra, "The influence of geological conditions for the level of building damages a preliminary study on the impact of the Bawean Island earthquake, East Java, Indonesia," vol. 1486, pp. 1–10, 2025.

[3] F. Muttaqy, A. D. Nugraha, N. T. Puspito, D. P. Sahara, Z. Zulfakriza, S. Rohadi, and P. Supendi, "Double-difference earthquake relocation using waveform cross-correlation in Central and East Java, Indonesia," *Geoscience Letters*, vol. 10, no. 1, pp. 1–16, 2023.

[4] nPerf, "Telkomsel's 3G/4G/5G coverage map in Indonesia." [Online]. Available: https://www.nperf.com/en/map/ID/-/5119.Telkomsel/signal?ll=-7.069548170712799&lg=114.6361541748047&zoom=10

[5] A. Förster, J. Dede, A. Könsgen, K. Kuladinithi, V. Kuppusamy, A. Timm-Giel, A. Udugama, and A. Willig, "A beginner's guide to infrastructureless networking concepts," *IET Networks*, vol. 13, no. 1, pp. 66–110, 2024.

[6] E. Rosas, O. Andrade, and N. Hidalgo, "Effective communication for message prioritization in DTN for disaster scenarios," *Peer-to-Peer Networking*

and Applications, vol. 16, no. 1, pp. 368–382, 2023.

[7] S. Perumal, V. Raman, G. N. Samy, B. Shanmugam, K. Kisenasamy, and S. Ponnan, "Comprehensive literature review on Delay Tolerant Network (DTN) framework for improving the efficiency of internet connection in rural regions of Malaysia," *International Journal of System Assurance Engineering and Management*, vol. 13, no. Suppl 1, pp. 764–777, 2022.

[8] M. Jesús-Azabal, J. Berrocal-Olmeda, J. García-Alonso, and J. Galán-Jiménez, "A self-sustainable DTN solution for isolation monitoring in remote areas," in *International Workshop on Gerontechnology*. Évora, Portugal: Springer, Oct. 5–6, 2020, pp. 57–68.

[9] E. Yaacoub, K. Abualsaud, T. Khattab, and A. Chehab, "Secure transmission of IoT mHealth patient monitoring data from remote areas using DTN," *IEEE Network*, vol. 34, no. 5, pp. 226–231, 2020.

[10] G. Koukis, K. Safouri, and V. Tsaoussidis, "All about Delay-Tolerant Networking (DTN) contributions to Future Internet," *Future Internet*, vol. 16, no. 4, pp. 1–21, 2024.

[11] S. H. Park, S. Cho, and J. R. Lee, "Energy-efficient probabilistic routing algorithm for Internet of Things," *Journal of Applied Mathematics*, vol. 2014, no. 1, pp. 1–7, 2014.

[12] A. Lohachab and A. Jangra, "Opportunistic Internet of Things (IoT): Demystifying the effective possibilities of opportunisitc networks towards IoT," in *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. Noida, India: IEEE, March 7–8, 2019, pp. 1100–1105.

[13] L. Wu, S. Cao, Y. Chen, J. Cui, and Y. Chang, "An adaptive multiple spray-and-wait routing algorithm based on social circles in delay tolerant networks," *Computer Networks*, vol. 189, 2021.

[14] Y. Yahara, A. Kato, M. Takai, and S. Ishihara, "On interactions between evacuation behavior and information dissemination via heterogeneous DTN," *Journal of Information Processing*, vol. 30, pp. 120–129, 2022.

[15] H. Liang, Y. Shang, and S. Wang, "[retracted] study on DTN routing protocol of vehicle ad hoc network based on machine learning," *Wireless Communications and Mobile Computing*, vol. 2021, 2021.

[16] S. Datta and S. K. Madria, "Prioritized content determination and dissemination using reinforcement learning in DTNs," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 20–32, 2021.

[17] S. Bajpai and A. Chauhan, "Evolution of machine learning techniques for optimizing delay tolerant routing," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. Greater Noida, India: IEEE, Dec. 16–17, 2022, pp. 294–299.

[18] F. Y. Yesuf and M. Prathap, "CARL-DTN: Context adaptive reinforcement learning based routing algorithm in delay tolerant network," 2021. [Online]. Available: https://arxiv.org/abs/2105.00544

[19] P. G. Buzzi, D. Selva, and M. S. Net, "Autonomous delay tolerant network management using reinforcement learning," *Journal of Aerospace Information Systems*, vol. 18, no. 7, pp. 404–416, 2021.

[20] S. C. K. Tekouabou, Y. Maleh, and A. Nayyar, "Towards to intelligent routing for DTN protocols using machine learning techniques," *Simulation Modelling Practice and Theory*, vol. 117, 2022.

[21] S. Liu, H. Shen, B. L. Smith, and V. Fessmann, "Machine learning based intelligent routing for VDTNs," in *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*. Honolulu, HI, USA: IEEE, July 24–27, 2023, pp. 1–10.

[22] L. Yang, J. A. Fraire, K. Zhao, R. Wang, W. Li, and H. Yang, "Optimizing deep-space DTN congestion control via deep reinforcement learning," *Computer Networks*, vol. 255, 2024.

[23] M. A. Salam, A. F. M. S. Saif, P. H. Katroju, and R. Kassouf-Short, "A constructive analysis on machine learning integration in High Delay Tolerant Networking (HDTN)," in *2024 7th International Conference on Advanced Communication Technologies and Networking (CommNet)*. Rabat, Morocco: IEEE, Dec. 4–6, 2024, pp. 1–6.

[24] J. Koteich, N. Mitton, and R. Wolhuter, "Mobility context aware routing protocol in dtn," in *2025 International Conference on Information Networking (ICOIN)*. Chiang Mai, Thailand: IEEE, Jan. 15–17, 2025, pp. 12–17.

[25] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. Rome, Italy: Association for Computing Machinery, March 2–6, 2009, pp. 1–10.

[26] Agussalim and M. Tsuru, "Spray and hop distance routing protocol in multiple-island DTN scenarios," in *Proceedings of the 11th Interna-*

tional Conference on Future Internet Technologies. Nanjing, China: Association for Computing Machinery, June 15–17, 2016, pp. 49–55.

[27] J. Höchst, L. Baumgärtner, F. Kuntke, A. Penning, A. Sterz, M. Sommer, and B. Freisleben, “Mobile device-to-device communication for crisis scenarios using low-cost LoRa modems,” in *Disaster management and information technology: Professional response and recovery management in the age of disasters*. Springer, 2023, pp. 235–268.