Gateway Selection in 5G/Wi-Fi architecture: A fire emergency case study

Kaouther Ouali Boulila, Thi-Mai-Trang Nguyen, Mohammad Imran Syed, Anne Fladenmuller, Brigitte Kervella, Nicolas Peugnet

Gateway Selection in 5G/Wi-Fi architecture: A fire emergency case study

Kaouther Ouali*, Thi-Mai-Trang Nguyen*†, Mohammad Imran Syed*,
Anne Fladenmuller*, Brigitte Kervella*§, and Nicolas Peugnet*
*Sorbonne Université, CNRS, LIP6, Paris, France
†Université Sorbonne Paris Nord, L2TI, France
§UPJV, Université de Picardie Jules Verne, Amiens, France
{kaouther.ouali, mohammad-imran.syed, anne.fladenmuller, brigitte.kervella, nicolas.peugnet}@lip6.fr,
thimaitrang.nguyen@univ-paris13.fr

***Abstract*—Public safety and first-responder networks require technologies that outperform current emergency networks. 5G and Wi-Fi architectures seem to be promising solutions for those environments. However, seamless integration of these wireless networks needs to be well investigated to benefit from the advantages of both networks and to fulfill the firefighter requirements. For this purpose, we propose a gateway selection algorithm for firefighter interventions. Proximal Policy Optimization, a well-known reinforcement learning strategy, is used to define and train the agent. Simulation results demonstrate that the proposed framework outperforms the Host Network Association scheme defined in the Optimized Link State Routing Protocol in terms of network throughput and packet drop rate.**

***Index Terms*—5G, Wi-Fi, HNA, OLSR, Reinforcement learning, Proximal Policy Optimization, Gateway selection.**

I. Introduction

Firefighters often have to navigate unfamiliar locations in order to extinguish fires. This can put them at risk of getting lost or isolated. In some cases, they use the ANTARES (Adaptation Nationale des Transmissions Aux Risques Et aux Secours) systems to communicate. These systems only provide digital voice capabilities and limited data rates, suitable for sending short messages or status updates. However, obstacles such as smoke, fire, walls, and noise make communication difficult, which can result in a loss of contact between the firefighters and the commander who is coordinating the emergency operation at the control center (CC). In this context, the ENE5AI project falls. It consists of deploying a hybrid 5G/Wi-Fi platform to the firefighters of Paris (Brigade de Sapeurs-Pompiers de Paris (BSPP). In our scenario, a hybrid 5G/Wi-Fi wireless network, consisting of a 5G cellular network and a Wi-Fi mesh network, shows promise for improving communication during firefighting interventions.

A Wi-Fi mesh network (MN) can extend the 5G coverage to an indoor environment such as a burning building, whereas the 5G base station is used to connect the CC to the emergency network. We distinguish two types of node in a mesh network: mesh client (MC) and mesh router (MR). MRs provide wireless access to MCs and form the network's backbone. MCs are connected to MRs and communicate with the CC through 5G gateway (GW) nodes using multi-hop communication. Two types of traffic are generated. The first traffic type consists of data flows sent from a MC to another MC through the associated MRs like video streaming. In the second traffic type such as push-to-talk for public safety, the traffic has to be forwarded to the CC through the GW nodes.

Real experiments were carried out and allowed us to identify some problems. In fact, maintaining such a communication among the firefighters or between the firefighters and the commander requires finding a suitable gateway. Unfortunately, there are a lot of challenges in the schemes of selecting the best gateway. In fact, the live video streaming with high bandwidth demand can cause network congestion, particularly in presence of cross traffic. In addition, the environment is characterized as highly dynamic because of the building and channel conditions that very frequently results in connections and disconnections of the routers from the network, causing unstable network connections.

Gateway selection strategies often rely on fuzzy logic [1], queuing theory [2] or optimization techniques such as Genetic Algorithm [3] and Ant Colony Optimization [4]. The disadvantage of these solutions is that they select the gateway based on geographical parameters and high-level objectives, which can create bias for the algorithm. Other conflicting objectives such as the 5G channel quality must be taken into consideration. Indeed, the proposed paper presents a model to optimize the gateway selection based on the Channel Quality Indicator (CQI) and guaranteeing the load balancing in our 5G/Wi-Fi network. For this purpose, deep reinforcement learning over multiple competing objectives is used, named Multi-Objective Reinforcement Learning (MORL). To train the agent, we adopt the Proximal Policy Optimization (PPO) model, which performs comparably or better than state-of-the-art approaches while being much simpler to implement and tune. To the best of our knowledge, the previous studies of gateway selection use many factors such as speed, direction, and distance, ignoring other methods that bring all factors together or have a significant influence on selection like load balancing.

The rest of this paper is arranged as follows. Section II

This work was supported by the ENE5AI and IE6 projects funded by BPI-France, the DIM RFSI 5G-REISEP project funded by Ile-de-France Region and the 5G Hybrid Network funded by Sorbonne Paris Nord University.

reviews some related literature on gateway selection. Section III presents the proposed system model components in detail and discusses the reinforcement learning method used in the model. Section IV evaluates the presented solution with the Host Network Association (HNA) technique defined in the Optimized Link State Routing (OLSR) protocol. We present the conclusion and future work in Section V.

II. Related Work

Providing a stable 5G connection to mesh network infrastructure is a considerable challenge because of the wireless channel conditions in emergency scenarios. In this section, we provide an overview of gateway selection schemes for wireless networks to Internet connectivity. El Mouna Zhioua et al. presented a fuzzy gateway selection algorithm (FQGwS) based on signal strength, load, link connectivity duration, and QUality of Service (QoS) traffic classes to provide stable communication between the vehicles and Long-Term Evolution (LTE) infrastructure [1]. However, this kind of solution uses the reactive approach where vehicles exchange messages to find the best and appropriate gateway, thus causing a high amount of overhead. Kushwah et al. presented a gateway selection in wireless mesh network [5]. First, traffic in a mesh router is calculated based on the connecting degree and interface queue length. Then, according to this aggregated traffic, the candidate Internet gateways are selected where a high amount of traffic is generated. Finally, Internet gateways are chosen based on the reliability value, which is obtained for each candidate gateway using the path tracing method. Bozorgchenani et al. considered the problem of Internet gateway selection and reliability of routes in wireless mobile networks (WMN) [6]. The candidate gateway nodes are selected based on network traffic. The scheme has reduced delay and energy consumption is acceptable. However, there is neither optimization nor the use of machine learning in the selection procedure in WMN. A new routing protocol for mobile gateway selection (RT-MGwS) has adopted many parameters, namely robust parameters, like received signal strength (RSS), trust connection, the number of hops, and route lifetime, to establish a robust route protocol for mobile gateway discovery [7]. In [8, 9], the gateways are buses that are directly connected to the Internet. The authors employed reinforcement learning to find an appropriate gateway for the source vehicle. Thy tried to achieve two contradicting objectives. The first objective was to maximize the number of connected vehicles with the highest link connectivity duration (LCD). The second one is to increase the traffic volume routed by the gateways by minimizing the number of vehicles connected to the same gateway. A new gateway selection system by using multi-objective optimization (MOO) is introduced by Retal et al., which may be considered to be a way to find solutions that constitute a trade-off between the objectives [10]. In fact, the first objective aims to maximize the number of connected vehicles while the second one aims to perform a fair load distribution.

However, even though these kinds of solutions show good results in WMN, the methods used in VANET networks are not suitable in public safety scenarios.

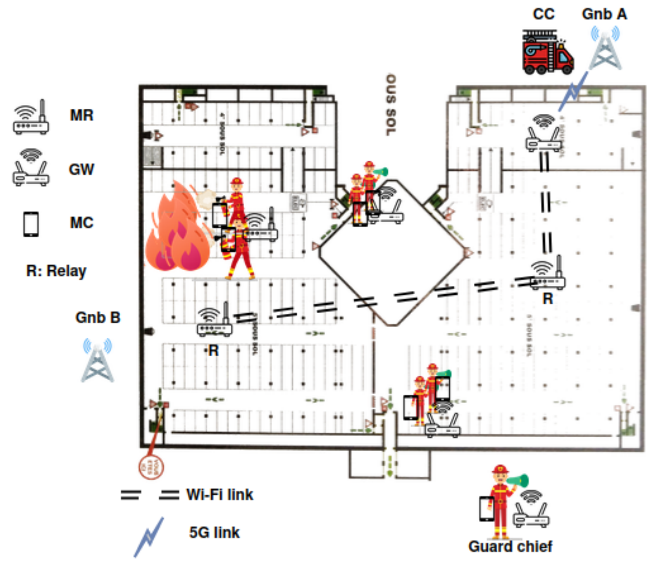


Fig. 1. System architecture

Following the suit of Alabbas et al. [8], we present a novel model to select the mobile gateways using deep reinforcement learning in firefighters intervention.

III. System Model

Our system model is a 5G/Wi-Fi hybrid network architecture. We focus on the case of an indoor fire scenario involving the Jussieu parking at Sorbonne University in flames as depicted in Fig. 1. Our solution is based on a set of mesh routers using standard IEEE 802.11 radio interfaces and Internet Protocol (IP)-based technology. These routers must be deployed inside and outside the parking in order to ensure full coverage of the area. Outdoor coverage can be provided by the 5G gNodeBs (gNB) placed at the firefighters' vehicle and connected to the control center. Both gNBs A and B are placed on the same level (level -2), while the fire breaks out at level -5. Indoor coverage can be provided by Ad Hoc nodes deployed by firefighters as they walk around the parking. We refer to these nodes as relays (R). Each firefighter carries a router and a smartphone with him. The idea is to build a mesh network with the relays, the carried routers, the gateways, and each firefighter's terminal equipment. Wi-Fi routers and gateways constitute the backbone of the mesh network, which ensures connectivity in the affected zone. The gateway selection solution is built using reinforcement learning. The main goal is to achieve two contradicting objectives by finding the best trade-off between them. The objectives are:

1) Choosing the GW with the highest CQI in order to route traffic with the best radio link conditions, which aims to maximize the throughput but compromises the load balancing.

2) Maximizing the Jain's index [11] illustrated in Eq. 1 by distributing traffic between multiple GWs to improve the system performances.

$$Jain\ Index = \frac{(\sum_{i=0}^{n} L_i)^2}{n \sum_{i=0}^{n} L_i^2} \quad (1)$$

where n is the number of gateways and L is the traffic load of each GW.

The CQI choice is based on the tests done in our lab to ensure a good quality of service and experience for the firefighters, while we choose load balancing to reduce the gateways' energy consumption, that have a limited battery, to prevent it from breaking down.

In the training phase, the RL agent starts to learn from the environment. The RL agent is able to find the best GW for each FT requesting to communicate when the training phase ends.

A. Deep Reinforcement Learning

Reinforcement Learning (RL) is a machine learning technique that mimics human interaction with the environment to learn new skills [12]. The main components of the RL system are the agent and the environment. It can be framed as a Markov decision process. The concepts state (S), action (A) and reward (R) represent the interaction of the agent with its environment. At each time (t), the agent observes the environment state (s_t) and takes action (a_t) from the set of available actions, causing a state transition to a new state (s_{t+1}). The agent obtains a reward (r_t) that indicates whether the decision taken is correct or wrong. The mapping between the action a and the state s is modeled by the policy $\pi(a, s)$ reflecting the interaction of the agent with its environment. The policy π represents the action probability as follows:

$$\pi(a|s) = P(a = a_t | s = s_t) \quad (2)$$

The agent looks for the optimal policy $\pi^*(a, s)$ by maximizing the cumulative discounted reward for each s $\in$ S and a $\in$ A following eq.3.

$$\pi^*(a|s) = \underset{\pi(a|s)}{\arg\max} \sum_{t=t_0}^{t_{end}} \gamma^{t-t_0} r_t \quad (3)$$

where $\gamma \in (0, 1)$ is the discount factor and t is the time horizon. Algorithms for policy optimization fall into two categories: value-based algorithms and policy-based algorithms, which have better convergence and are more appropriate for large action spaces.

Proximal Policy Optimization (PPO) [13] algorithm is an actor-critic method combining both the policy-based and the value-based algorithms that help to stabilize the training with two neural networks. The first one (the actor) controls how the agent behaves. It takes the state s as entries and outputs the policy $\pi(a, s)$. The second one (the critic) optimizes $V(s)$ that measures how good the action a taken is. PPO uses the advantage estimate $A(s, a)$ to reduce the variance, expressed as follows:

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t) \quad (4)$$

$$Q(s_t, a_t) = r_t + \sum_{i=1}^{T-1} \gamma^i r_{t+i} + \gamma^{t+T} V(s_{t+T}) \quad (5)$$

$Q(s, a)$ represents the cumulative discounted reward where for the state s_t the action a_t is taken. $V(s)$ represents the baseline, which is the method used to update the policy to choose better actions. PPO makes use of the Trust Region Policy Optimization (TRPO) technique to make sure that the updated policy never deviates from the original, increasing its stability and reliability. Wherefore, we use the PPO algorithm to train our agent in the gateway selection phase.

B. Gateway Selection

1) Observation State Definition: It represents the relationship between the GW and the gNBs, with which it is associated, in terms of channel quality. The state is expressed as follows:

$$S = \begin{pmatrix} CQI_1 & CQI_2 & \dots & CQI_n \\ L_1 & L_2 & \dots & L_n \end{pmatrix}$$

- n is the number of gateways.
- CQI_i denotes the channel quality indicator of the gateway i and the gNBs, with which it is associated.
- L_i denotes the i^{th} GW throughput.

2) Action Definition: In the proposed algorithm, the set of actions represents the number of GWs.
$A = \{a_1, a_2, \dots, a_n\}$, where a_i represents the selection of GW_i.

3) Reward Definition: The reward is assigned based on two metrics: the first metric is the 5G channel quality, whereas the second is the amount of traffic routed by each GW. Setting the reward function with many objectives needs to apply Multi-Objective Reinforcement Learning (MORL) which aims to find a trade-off solution for multiple objectives. The reward function is expressed as:

$$R = w_1 \frac{CQI_i}{15} + w_2\ Jain\ Index_i \quad (6)$$

where CQI_i is the channel quality of the GW_i, while $Jain\ Index_i$ represents the Jain index when the traffic volume is routed by GW_i. w_1 and w_2 are weights of the multi-objective function and defined as follows:

$$w_1, \quad w_2 \in R+ \quad (7)$$

$$w_1 + w_2 = 1 \quad (8)$$

To find the best weight, we choose the same weight value to give the same importance for both objectives, so no one is neglected over the other. The CQI is normalized to not bias the agent.

The reward value is equal or close to 1 when the action is valid. Else, the reward gets close to zero.

The zero reward is applied when:

1) The agent selects a GW which has a poor CQI.
2) The agent selects an overloaded GW, thus ignoring the network load balancing.

C. Agent State Parameters

In the proposed system, the gateway selection algorithm is built in a centralized agent server. PPO is employed to maximize the GWs selection return. The reward r is a multi-objective reward in which the agent tends to find a GW for a mesh client with the maximum CQI and Jain load index. The episodic environment is considered during agent training so that each episode consists of 100 steps. Our environment is variable as it depends on dynamic traffic and moving nodes, in which it is difficult to determine the future return. In such case, it is appropriate for the RL agent to maximize the current reward rather than the cumulative discount reward received over the future. This is carried out by adopting γ = 0. To prove our choice, we trained the agent with different γ values. We found that the lower the gamma value is, the faster the agent learns and higher performance we get (Fig. 2). The figure below shows the learning performance when using a neural network with fully connected input and an output layer. We see that after around 50 episodes, the agent is able to perfectly predict the best gateway from the current observation.

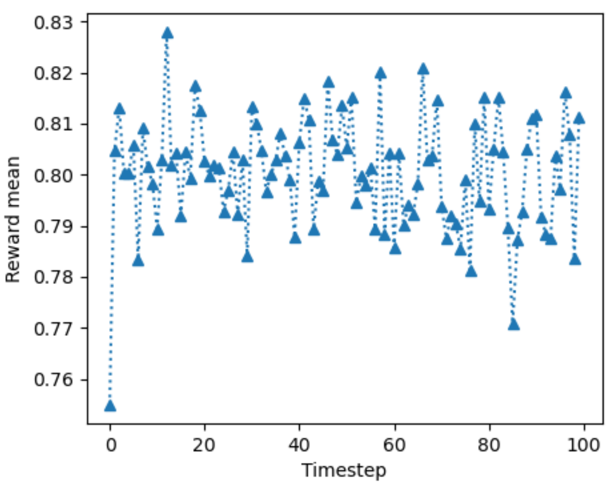


Fig. 2. Training the agent with $\gamma = 0$

D. Extended OLSR

According to the OLSR protocol, HNA messages are sent only by a node that has a 5G interface. The purpose is to provide connectivity from the OLSR network (Wi-Fi) to a 5G network. The gateways send HNA messages each HNA_INTERVAL containing a list of addresses of the associated network and its network mask (netmask). So the gateways located in 5G networks construct tuples, where each tuple contains in particular a A_time field which specifies the time at which this tuple expires and hence must be removed.

Upon receiving an HNA message by an MR, the Host Network Association Set is updated with the information of the received HNA message and the routing table is updated accordingly [14].

In order to support the deep RL decision in OLSR, we need to extend existing HNA messages to carry out the selected gateway for the network nodes. This task has been designed and implemented in a generic field, referred to as $Override$ for OLSR. This field takes only two values ($True$, $False$). When an action is taken by the agent, only the selected gateway sends HNA messages ($Override = True$) to the Wi-Fi network. When an HNA message is received, the association set ignores the existing tuples and stores the new one. The routing table is therefore updated appropriately.

The format of the extended HNA messages is illustrated in Fig. 3. This field is included in the data field in the generic OLSR packet format specified in [14].

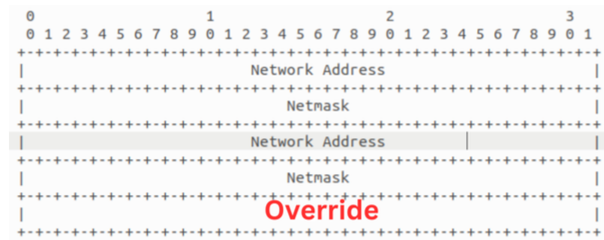


Fig. 3. The modified HNA message format

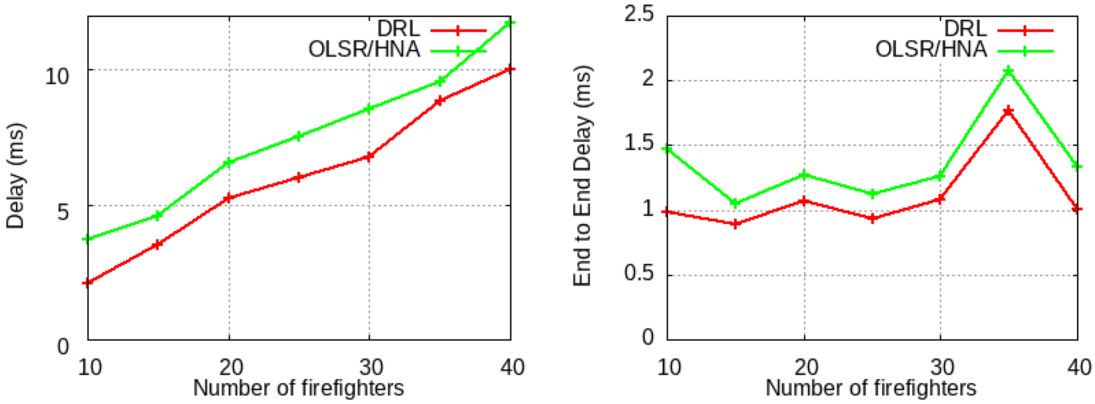
IV. Results

This section presents the performance evaluation of the proposed algorithm. We build our testing environment in Network Simulator NS-3 [15]. The DRL is implemented by using the Python programming language. The stable-Baseline3 library [16] is used to implement and train the RL agent. The openGym [17] is used to interact with NS-3.

Our reference scenario describes the concept of firefighter(s) broadcasting the video and audio of their intervention. The push-to-talk (PTT) communication [18] takes place via a client/server setup, which means that, an MC sends signals to its access point MR, and talks with the other MCs via a PTT server through the 5G/Wi-Fi network. The video streaming flow is routed via the Wi-Fi network. The scenario is illustrated in Fig. 1. The guard chief carries a gateway and moves randomly around the parking for surveillance. The other firefighters which carry an MR advance step by step to the fire to put it out. For comparison purposes, the OLSR/HNA mechanism is simulated. Therefore, it is adopted as a benchmark. The entire simulation parameters are listed in Table I.

In Fig. 5, the packet loss ratio for the PTT application is depicted. Our solution has better performance in increasing the number of firefighters in comparison with the OLSR protocol. Poor air-interface signal quality may increase the packet error rate, which results in more packet retransmissions and segmentation. As a result, the number of lost packets increases.

Fig. IV presents the delay of the video streaming and the PTT applications. The DRL outperforms the OLSR/HNA



(a) PTT delay

(b) Video delay

Fig. 4. End-to-End delay (ms)

TABLE I
PARAMETERS SETTING

Parameter	Setting
5G frequency	3.149 GHz (n78 band)
Wi-Fi frequency	5 GHz
gNB downlink power	0 dB
GW transmission power	23 dB
Firefighter average velocity	0.5 m/s
Scheduling algorithm	Round Robin
Simulation time	175 s
Number of GWs (n)	4
Number of MRs (m)	9
Area	100 m x 100 m x 10 m
PTT message size	60 Kbyte
Video message size	1500 Kbyte
HNA_INTERVAL	5s [14]

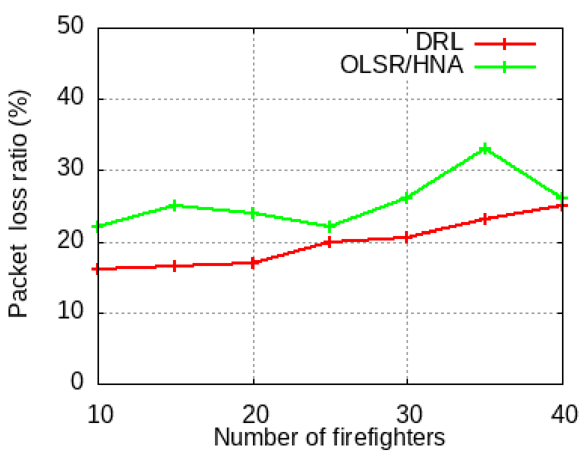


Fig. 5. PPT Packet loss ratio (%)

mechanism. The load balancing is a suitable solution to decrease the end-to-end delay. It reduces the strain on each GW and makes them more efficient, speeding up performance and reducing latency. Besides, when selecting a GW with poor CQI, the block error rate (BLER) over the air interface is high. Therefore, multiple retransmissions over the physical layer are required before data is successfully transmitted, prolonging transmission latency. The delay exceeds 2s, which is logical for the OLSR due to its high convergence time in such a dense network.

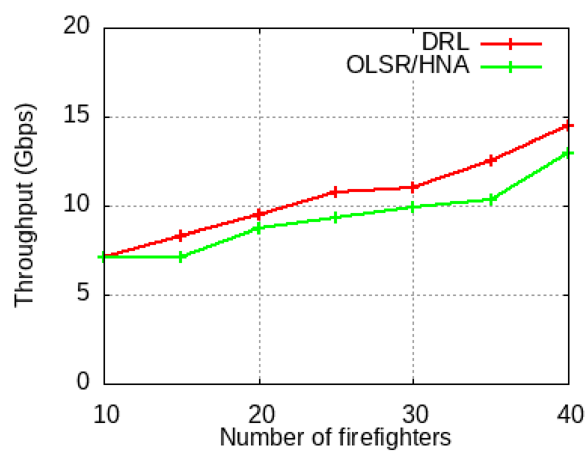


Fig. 6. Average PTT throughput (Mbps)

Fig. 6 presents the average throughput for the PTT application. When the DRL algorithm selects the GW with the high CQI, it is expected for the gNB to send data with large transport block size per TTI, which is equivalent to increase in the throughput. Fig. 7 shows the Jain index which takes into consideration the load balancing among the GWs; therefore, the DRL solution exhibits more equitable distribution in comparison with the HNA mechanism. The reliance on the OLSR protocol in selecting the GW causes inequality and a wide variation in the distribution of the load over the GWs.

Fig. 8 presents the number of sent HNA messages N^{HNA}_{DRL} and N^{HNA}_{OLSR} sent respectively by the OLSR/HNA protocol and the DRL solution by increasing the number of routers (GW and MR). Hence, we can deduce the following formulas:

$$N^{HNA}_{OLSR} = n * (m + n - 1) \tag{9}$$

$$N^{HNA}_{DRL} = m + n - 1 \tag{10}$$

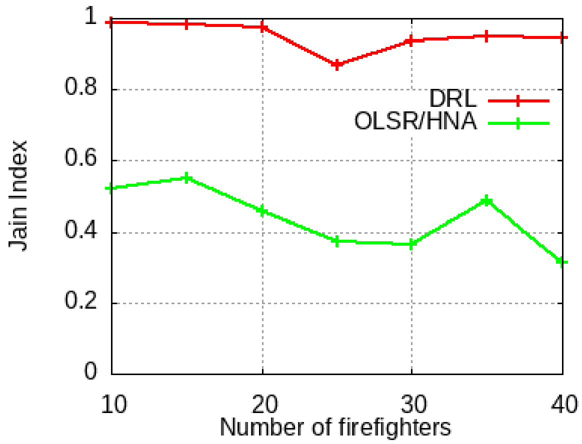


Fig. 7. Jain Index

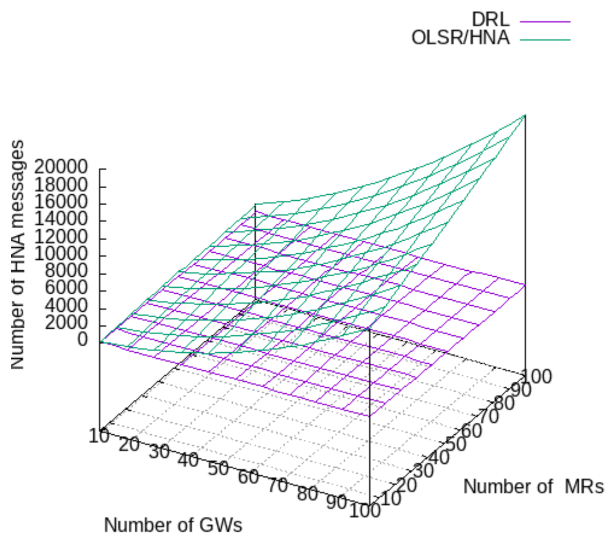


Fig. 8. Number of exchanged HNA messages

Our solution reduces the total number of exchanged HNA messages by n for each HNA_INTERVAL. In contrast, the number of HNA messages with the OLSR protocol is high and affected drastically by increasing the number of routers, as depicted in Fig. 8. This high increase will lead to network congestion and a high energy consumption. That means the OLSR protocol is impractical to be applied in the gateway selection system, which needs to be executed in real-time and in huge networks, especially in public safety scenario where routers have limited batteries.

V. Conclusion

We present, in this article, a gateway selection algorithm with the aim of finding the best mobile gateway for firefighters in need of 5G access. For this purpose, an DRL algorithm is adopted. DRL uses two objectives to optimize the gateway selection problem; the 5G channel quality indicator and the network load balancing. To determine a trade-off between the two contradicting objectives and to implement and train the agent, the weighted sum method and the proximal policy optimization strategy are used, respectively. Compared with the existing OLSR/HNA mechanism for gateway selection, the simulation results show that the proposed approach is effective in terms of reducing the delay and packet loss, distributing the traffic among gateways, and decreasing the network overhead. In the future, we plan to add other parameters to the state space like the network topology to be able to select the best routing path rather than the OLSR protocol.

References

[1] G. El Mouna Zhioua, N. Tabbane, H. Labiod, and S. Tabbane, "A fuzzy multi-metric QoS-balancing gateway selection algorithm in a clustered VANET to LTE advanced hybrid cellular network," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 2, pp. 804–817, 2014.

[2] R. Kushwah, S. Tapaswi, and A. Kumar, "Multipath delay analysis using queuing theory for gateway selection in hybrid MANET," *Wireless Personal Communications*, vol. 111, pp. 9–32, 2020.

[3] R. Kushwah, S. Tapaswi, and A. Kumar, "A novel technique for gateway selection in hybrid MANET using genetic algorithm," *Wireless Personal Communications*, vol. 126, no. 2, pp. 1273–1299, 2022.

[4] I. Jabri, T. Mekki, A. Rachedi, and M. B. Jemaa, "Vehicular fog gateways selection on the internet of vehicles: A fuzzy logic with ant colony optimization based approach," *Ad Hoc Networks*, vol. 91, p. 101879, 2019.

[5] R. Kushwah, "A novel traffic aware reliable gateway selection in wireless mesh network," *Cluster Computing*, pp. 1–15, 2023.

[6] A. Bozorgchenani and M. Jahanshahi, "A novel reliability and traffic aware gateway selection scheme in wireless mesh networks," *Wireless Personal Communications*, vol. 96, pp. 6111–6128, 2017.

[7] B. Sharef, R. Alsaqour, M. Alawi, M. Abdelhaq, and E. Sundararajan, "Robust and trust dynamic mobile gateway selection in heterogeneous VANET-UMTS network," *Vehicular communications*, vol. 12, pp. 75–87, 2018.

[8] H. Alabbas and Á. Huszák, "Reinforcement learning based gateway selection in vanets," *International journal of electrical and computer engineering systems*, vol. 13, no. 3, pp. 195–202, 2022.

[9] H. Alabbas and Á. Huszák, "New gateway selection algorithm based on multi-objective integer programming and reinforcement learning," *Infocommunications Journal*, vol. 14, no. 4, pp. 4–10, 2022.

[10] S. Retal and A. Idrissi, "A multi-objective optimization system for mobile gateways selection in vehicular Ad-Hoc networks," *Computers & Electrical Engineering*, vol. 73, pp. 289–303, 2019.

[11] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, vol. 21, 1984.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[14] T. Clausen and P. Jacquet, "Optimized link state routing protocol (OLSR)," Tech. Rep., 2003.

[15] G. Carneiro, "NS-3: Network simulator 3," in *UTM lab meeting April*, vol. 20, no. 1, 2010, pp. 4–5.

[16] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021.

[17] P. Gawłowicz and A. Zubow, "ns-3 meets OpenAI Gym: The Playground for Machine Learning in Networking Research," in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, November 2019.

[18] W. Garey, T. R. Henderson, Y. Sun, R. Rouil, and S. Gamboa, "Modeling MCPTT and user behavior in ns-3." in *SIMULTECH*, 2021, pp. 30–41.