Environmental-aware Reinforcement Learning-based Scheduler for Trustworthy 6G in the Factory Floor

Fjolla Ademaj-Berisha
Silicon Austria Labs GmbH
Linz, Austria

Roya Khanzadeh
Johannes Kepler University, JKU LIT SAL IWS Lab
Linz, Austria

Andreas Springer
Johannes Kepler University, JKU LIT SAL IWS Lab
Linz, Austria

Hans-Peter Bernhard
Silicon Austria Labs GmbH, Johannes Kepler University, Linz, Austria

ABSTRACT

6G networks will play a crucial role in Industry 4.0 advancing further smart manufacturing and trustworthy communication. This paper introduces a reinforcement learning scheduler that considers environmental knowledge designed to improve reliability and security aspects of trustworthiness by favoring to serve nodes when in line-of-sight. We evaluate the performance in a factory floor scenario for two different heights of remote-radio heads and various blockage densities. The results show that the proposed method outperforms traditional schedulers such as round robin and proportional fair in terms of reliability, availability and fairness enhancing thus reliability and security aspects of trustworthiness when the environment is prone to a mixture of line-of-sight and non-line-of-sight.

KEYWORDS

Trustworthiness, Scheduling, Reinforcement Learning, 6G, Factory floor, Reliability, Security;

ACM Reference Format:
Fjolla Ademaj-Berisha, Roya Khanzadeh, Andreas Springer, and Hans-Peter Bernhard. 2024. Environmental-aware Reinforcement Learning-based Scheduler for Trustworthy 6G in the Factory Floor. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24), November 18–22, 2024, Washington D.C., DC, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3636534.3696728

ACM MobiCom '24, November 18–22, 2024, Washington D.C., DC, USA

ACM ISBN 979-8-4007-0489-5/24/11
https://doi.org/10.1145/3636534.3696728

1 INTRODUCTION

The development of smart factories in Industry 4.0 (I4.0) with 5G has boosted productivity, efficiency, and flexibility. As 5G continues to be implemented and evolves, 6G is emerging, aiming to further enhance I4.0 with higher data rates, ultra-low latency, and advanced artificial intelligence (AI) integration, driving automation, trustworthy communication, and sustainable manufacturing. Trustworthiness is a key characteristic in the 6G vision, encompassing reliability, security, resilience, privacy, and safety. While these aspects are considered during design, the trustworthiness of 6G networks during operation largely depends on how operators deploy, configure, and optimize it [1]. Commonly, model-based algorithms for scheduling tasks have been designed and optimized to meet conventional network performance indicators (NPIs) such as data rate, throughput, error probability, and communication delay [8]. Advanced reinforcement learning (RL) methods have also been applied to further enhance these NPIs [10, 12].

Although these NPIs are crucial indicators, effective resource allocation and network scheduling can significantly enhance the network's trustworthiness, when optimized defining new metrics as trustworthiness indicators (TIs). In addition, with the recent advancements in environmental awareness thanks to integrated sensing and communication (ISAC) and multi-modal sensing [5, 11], it is important to consider this knowledge about the environment in addition to the traditional NPIs. Since environmental changes have a strong impact on the physical layer (PHY) parameters, we must take these PHY parameters into consideration when optimizing the network. The authors in [6] proposed to use the signal to interference ratio (SIR) as PHY layer characteristic to proportionally allocate resources within a cell to users who meet a certain SIR criteria for area service availability. However, due to the proportional fair (PF) strategy, this scheduler delivers symmetric performance across both available and non-available regions, offering no improvement in trustworthiness.

In this work, we propose an RL-based scheduling strategy to enhance the reliability and security aspects of trustworthiness, where an agent gradually learns to schedule available

resource blocks to maximize the defined rewards. We define two TIs, service reliability as an indicator for reliability, and service availability as an indicator for security. While service reliability ensures the correctness and consistency of services, the impact of the loss of service availability reflects the consequences of security breaches, i.e., a moving robot that doesn't receive any downlink messages for a long time might lose synchronization with the control system, potentially leading to erratic behaviour, collisions etc. We consider PHY parameters such as the state of being in line-of-sight (LOS) or non-line-of-sight (NLOS) for indicating regions where the service is reliable. We focus on indoor industrial environments, where blockages are typical and therefore transitions between LOS and NLOS are evident.

2 ENVIRONMENTAL-AWARE KNOWLEDGE

Wireless communication performance is heavily influenced by the radio propagation environment. Recently, the notion of environmental awareness has become vital in 6G networks for high reliability, low latency, robust security, and massive connectivity, enhances the precision of channel knowledge prediction. This can be achieved through various sensing technologies like Radar, Camera, LiDAR, GPS, and ultra wide band (UWB). For instance, channel knowledge maps (CKM) provide location-specific channel information [11], while multi-modal sensing can predict channel conditions and detect objects by integrating multiple inputs [5].

Following the concept of CKM, this work focuses on the obstacle detection capability of the system, providing a binary spatial map of LOS and NLOS states of moving nodes in a given area, which we refer to as the environmental-aware map. It is important to note that LOS/NLOS is just one example of input knowledge. The spatial map can be generated using any other quality of service (QoS) parameters (e.g., [2, 7]), which can then be transformed into a binary decision by applying thresholds based on the required service bounds.

3 PROPOSED METHOD

We propose an RL-based scheduler which tries to allocate the resource blocks among K nodes in such a way that the overall network trustworthiness in terms of reliability and security is enhanced while maintaining fairness among the nodes. The proposed scheduler, illustrated in Fig. 1, utilizes environment-driven knowledge as the input data to create a resource allocation policy that maximizes the metrics defined for assessing the network's trustworthiness. In the proposed method, an RL agent observes the environment's state, denoted as $\mathbf{S}(t)$, and selects the action $a(t) \in 1, \ldots, K$ with the highest expected reward. This action influences the immediate rewards $r(t)$ received and alters the environment's state. With a single-user scheduling scenario, the agent's task is to choose one node to serve at each transmission time interval

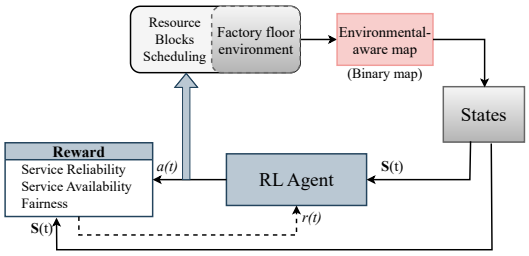


Figure 1: The proposed environmental-aware RL scheduler methodology.

(TTI). The environmental-aware knowledge gives details on the LOS/NLOS states of the kth moving node, $k = 1, \ldots, K$, at current TTI as $s(t, k)$. It also includes predictions for these states over the next T TTIs, as $s(t + i, k), i = 1, \ldots, T$, assuming the trajectories of the nodes are estimated for T upcoming TTIs. In addition, the number of TTIs that the kth node has been served within the previous L TTIs is tracked and stored as $h(k, t)$. Therefore, the state of the environment in our problem at TTI t is defined as $\mathbf{S}(t)$ with dimension $K \times (T + 2)$ and the kth row as

$$\mathbf{s}(k, t) = [h(k, t), s(t, k), s(t + 1, k), \ldots, s(t + T, k)]. \quad (1)$$

Due to the extensive number of states in our problem, the deep Q-network (DQN) algorithm with an online three-layer-neural network (NN), parameterized by θ is deployed to approximate the action-value function. Another NN with the same structure but a different parameter set θ^-, is used to approximate the target value. The online NN is updated at the ith iteration based on the difference between the estimated value and the target value as

$$L(\theta_i) = \mathbb{E}_{(s,a,r(t),s')}\left[\left(r(t) + \gamma \max_{a'} Q_{\theta^-}(s', a') - Q_{\theta_i}(s, a)\right)^2\right], \quad (2)$$

where s, a, $r(t)$, s' and γ are the current state, action, immediate reward, the next state and discount factor respectively. The immediate reward $r(t)$ is what the agent receives as feedback on how good or bad was the selected action in terms of the defined metrics. More details on how this reward function is designed in our problem will be given later in this section. In addition, the target network is also updated every number of episodes as $\theta^- \leftarrow \theta$.

At each TTI, the agent uses an ε-greedy policy: it selects the action with the maximum estimated value with probability $1 - \varepsilon$ or a random action with probability ε. The value of ε starts as $\epsilon_{\max}$ and decreases over time to shift from exploration to exploitation, with $\epsilon_{\min}$ set to limit further exploration.

3.1 Defined metrics

The ultimate goal of the proposed scheduler is to enhance the reliability and security aspects of trustworthiness throughout the network based on two metrics, respectively: service

reliability and service availability, while maintaining fairness in the network. Considering only one node is served at each TTI, the service reliability metric is defined as

$$\mathbb{E}_t[\text{Rel}] = \mathbb{E}_t\Big[\sum_{k=1}^{K} z(t,k)\text{Rel}(t,k)\Big], \tag{3}$$

with $z(t,k) = 1$ if the kth node is served at TTI t, otherwise $z(t,k) = 0$, and $\text{Rel}(t,k)$ is the kth node reliability score at TTI t. We consider the state of LOS/NLOS as a binary parameter that can indicate service reliability at node k during TTI t as

$$\text{Rel}(t,k) = \begin{cases} 1 & \text{if } s(t,k) \equiv \text{LOS} \\ 0 & \text{if } s(t,k) \equiv \text{NLOS}, \end{cases} \tag{4}$$

Service availability is defined as a metric for assessing security, and is defined as

$$\mathbb{E}_t[\text{Avl}] = \mathbb{E}_t\Big[\frac{1}{K}\sum_{k=1}^{K} \text{Avl}(t,k)\Big], \tag{5}$$

with $\text{Avl}(t,k)$ as the kth node's availability at time t which is calculated as

$$\text{Avl}(t,k) = \begin{cases} 1 & \text{if } \text{TBA}(t,k) \le \zeta \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Here, $\text{TBA}(t,k)$ is time between access (TBA) and denotes the number of TTIs that have passed since the kth node last accessed the network, recorded at TTI t. Specifically, if the kth node is served at TTI t, then $\text{TBA}(t,k)$ is zero, otherwise it increases by one. In Eq. (6), ζ is the maximum number of allowed TTIs between accesses. If a node's TBA exceeds this threshold, its service is deemed unavailable.

Finally, the last metric is fairness and is calculated as follows [9]:

$$\mathbb{E}_t[\text{Fair}] = E_t\left[\frac{\left(\sum_{k=1}^{K}\sum_{\tau=t-L}^{t} z(\tau,k)\right)^2}{K\sum_{k=1}^{K}\left(\sum_{\tau=t-L}^{t} z(\tau,k)\right)^2}\right]. \tag{7}$$

3.2 Reward function

According to the defined evaluation metrics, the reward function is defined as follows

$$r(t) = c_1\widehat{\text{Rel}}(t) + c_2\widehat{\text{Avl}}(t) + c_3\widehat{\text{Fair}}(t), \tag{8}$$

where $\widehat{\text{Rel}(t)}$, $\widehat{\text{Avl}}(t)$ and $\widehat{\text{Fair}}(t)$ are respectively reliability, availability and fairness related rewards with c_1, c_2, c_3 as constants. If a node with a service reliability score of one is chosen at TTI t, the agent receives a positive reward of c_1. If no nodes have a service reliability score of one, no reward related to service reliability is given, i.e. $\widehat{\text{Rel}}(t) = 0$. However, if the agent selects a node with a service reliability score of zero while there exists at least one node with a reliability score of one, the agent receives a negative reward of $-c_1$.

To improve service availability, it is essential to reduce TBA for all nodes. Prioritizing nodes that are transitioning from LOS to NLOS based on their movement trajectories is key, as delaying their service can lower their future service reliability. Thus, the agent gains $c_2\widehat{\text{Avl}}(t) = \frac{c_2}{\sum_{i=1}^{T} s(t+i,a(t))}$ for choosing nodes with a reliability score of one that is likely to move to NLOS within the next T TTIs. Otherwise, the service availability-related reward would be zero. Finally, the fairness-related reward is the actual fairness score that the taken action archives as $\widehat{\text{Fair}}(t) = \text{Fair}(t)$.

4 SIMULATION RESULTS

We consider an indoor industrial environment and downlink communication in a 6G cellular network with R remote radio heads (RRHs) and K nodes. It is assumed that the base station (BS) unit can generate an environmental-aware map and can also predict it for the near future, based on the side information about the environment, including the node positions, movement directions etc. To generate environmental-aware maps, we consider the geometry-based stochastic channel model (GSCM) from 3GPP TR 38.901, tailored to an indoor factory scenario. This scenario focuses on factory halls of different sizes and varying clutter densities (e.g. machines, storage shelves, assembly lines etc.). This model allows for various geometrical factors, including the height of RRHs located either above the clutter (very high) or embedded within it (low height), as well as the height of the factory hall, and the dimension of clutter. In addition to the 3GPP TR 38.901 indoor factory scenario, we incorporated the spatial consistency adapted from [3, 4] into the model. This model adaptation is necessary for accurately reflecting spatial correlation, ensuring that closely located nodes in the factory floor experience similar channel propagation effects, due to shared scatterers in the environment.

4.1 Environmental-aware map for LOS/NLOS

We focus on one of the environment properties, the state of being in LOS/NLOS propagation, where for node k and TTI t, the LOS/NLOS state is obtained as

$$s(t,k) = \begin{cases} \text{LOS}, & \text{if } q(t,k) \le p_{\text{LOS}}(t,k) \\ \text{NLOS}, & \text{otherwise} \end{cases}, \tag{9}$$

with p_{LOS} denotes the probability of being in LOS which is a function of the link between nodes and the closest RRH considering both distance and height, and $q(t,k)$ are spatially correlated random variables for any t and k adapted from the methodology in [3]. Figure 2 illustrates one example of a binary environmental-aware map for an area of $300 \times 150\,\text{m}^2$ and a resolution of 0.5 m. We consider 18 RRHs, denoted in green circles, spaced on a square lattice with spacing D, located D/2 from the walls (see Tab. 1). White color denotes

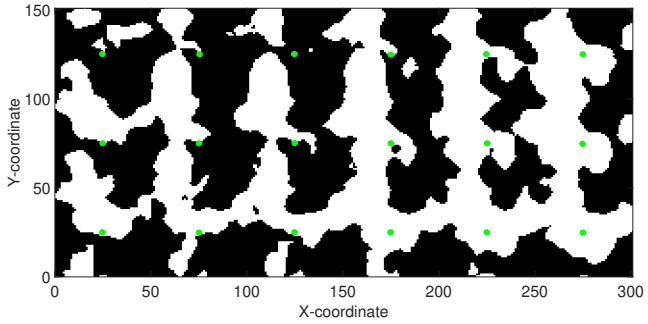


Figure 2: Environmental-aware map with LOS/NLOS states for the factory floor with clutter density of 35%**.**

LOS state and black color denotes NLOS state. The generated binary environmental-aware map is the input to the proposed environmental-aware scheduler. Note that the ratio between LOS/NLOS depends on the system parameters such as the clutter density, clutter size and height as well as RRH location, which are inferred in the statistics of the GSCM.

4.2 Scheduler evaluation

We evaluate the performance of the proposed scheduler with two well-studied scheduling methods in the literature, the round robin (RR) and PF schedulers [8], where RR schedules the nodes in the best fairness manner while PF tries to improve the overall throughput (which in our scenario results in the service reliability enhancement).

Table 1: Simulation parameters for the environment

Parameter	Value
Number of RRHs (R)	18
Distance between RRHs (D)	50 m
Ceiling height	10 m
Clutter height	5 m
Clutter density (μ)	$10\% - 80\%$
Clutter size	5 m
Node height	1.5 m
RRH height for low RRH	1.5 m
RRH height for high RRH	8 m
Number of nodes (K)	5
History length (L) [TTI]	10
Prediction length (T) [TTI]	10
$\epsilon_{\max}$, $\epsilon_{\min}$	1, 0.0001
c_1, c_2, c_3	10, 10, 5
Discount factor (γ)	0.95
Learning rate	0.1

Table 1 shows a summary of simulation parameters, where the parameters related to the geometry of the factory floor are taken from 3GPP TR 38.901. The deployed model is trained for 1000 iterations, while a new 300×150 m 2D map is generated independently for each iteration, with random node locations and movement trajectories initialization. Nodes move according to their initial trajectories, and all share the same traffic parameters and service requirements. We look at three aspects as defined in Sec. 3, reliability, availability and fairness. In addition, to provide a single metric for comparing the performance of the three methods, we use a cumulative evaluation index called Δ, adopted from [10], and defined as

$$\Delta = \Delta^{\text{Rel}} + \Delta^{\text{Avl}} + \Delta^{\text{Fair}} \tag{10}$$

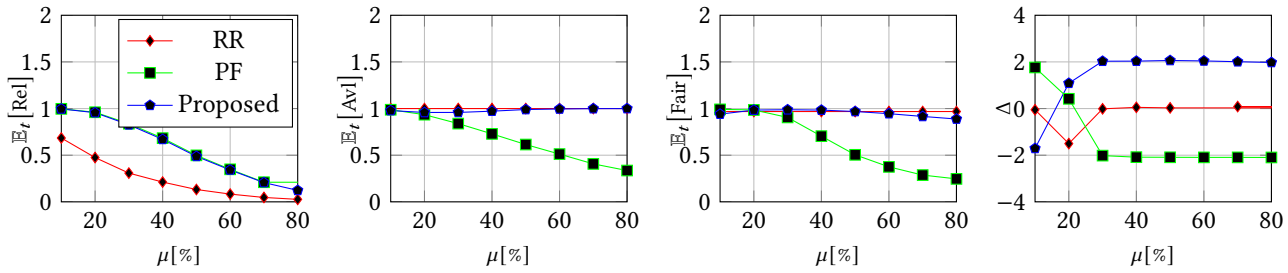
where

$$\Delta^{j} = \frac{y_{i,j} - \bar{y}_j}{\nu_j} \tag{11}$$

with $y_{i,j}$ representing the specific value of algorithm i in terms of jth evaluation metric, including service reliability, service availability and fairness; $\bar{y}_j$ and ν_j are respectively the average and standard deviation values of the jth metric across all algorithms. Figure 3 compares the achieved performance when RRHs are clutter embedded with a height of 1.5 m. The results are evaluated over the clutter density $\mu = \{10\%, 20\%, ..., 80\%\}$. As expected, PF and the proposed method significantly outperform RR in terms of service reliability, whereas in terms of service availability, RR is superior closely followed by the proposed algorithms. Only for clutter density $\mu \leq 50\%$, RR slightly outperforms our proposed method, revealing that the proposed method is advantageous when there are more blockages on the factory floor. Considering fairness, the proposed and RR algorithms perform similarly and noticeably better than PF. When comparing all three aspects, defined by Δ, the proposed algorithm is superior for all cases when $\mu \geq 20\%$, denoting environments with an increase of NLOS areas. Figure 4 shows the same comparison when RRHs have a height of 8 m. Similarly, PF and the proposed method significantly outperform RR in terms of service reliability, however achieving higher reliability as compared to the low RRH height scenario. In terms of both service availability and fairness, the proposed algorithm and RR perform closely, surpassing PF. When comparing all three aspects, defined by the Δ, the proposed algorithm is superior for all cases when $\mu \geq 32\%$, making the proposed method ideal for blockage-prone factory floors. As the algorithm is designed to share resources and prioritize depending on the prediction and history of LOS/NLOS states, it does perform poorly only when most nodes are in LOS most of the time - which is not realistic for factory floor evironments.

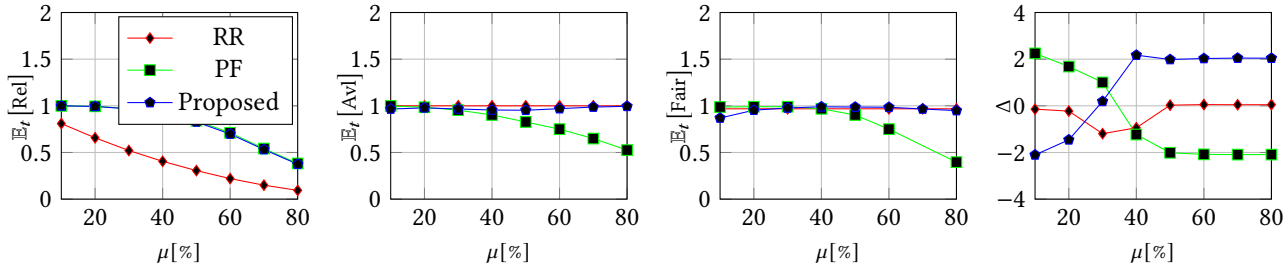
5 CONCLUSION

In this work, we presented a scheduling strategy based on RL combined with environmental awareness. Our scenario of interest is an indoor factory floor where we consider current and near-future environmental knowledge in the form of a binary environmental-aware map for indicating LOS/NLOS areas. The RL-based scheduler is designed to enhance two



(a) service reliability **(b) service availability** **(c) fairness** **(d) cumulative index**

Figure 3: Evaluation results for the scenario with low RRH height (1.5 m) and clutter density $\mu = \{10\%, 20\%, ..., 80\%\}$.



(a) service reliability **(b) service availability** **(c) fairness** **(d) cumulative index**

Figure 4: Evaluation results for the scenario with high RRH height (8 m) and clutter density $\mu = \{10\%, 20\%, ..., 80\%\}$.

aspects of trustworthiness in 6G networks, reliability and security. Our results demonstrated the approach's effectiveness over conventional methods in terms of defined metrics.

ACKNOWLEDGMENTS

This work was supported by the European Unions Horizon Europe project DETERMINISTIC6G under grant agreement No. 101096504. It has been also in part supported by the "University SAL Labs" initiative of Silicon Austria Labs (SAL) and its Austrian partner universities for applied fundamental research for electronic-based systems.

REFERENCES

[1] 2023. 6G Trustworthiness Considerations. *Next Generation Mobile Networks Alliance* (2023).

[2] Fjolla Ademaj and Hans-Peter Bernhard. 2022. Quality-of-Service-Based Minimal Latency Routing for Wireless Networks. *IEEE Transactions on Industrial Informatics* 18, 3 (2022), 1811–1822. https://doi.org/10.1109/TII.2021.3071596

[3] Fjolla Ademaj, Martin K. Mueller, Stefan Schwarz, and Markus Rupp. 2017. Modeling of Spatially Correlated Geometry-Based Stochastic Channels. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. 1–6. https://doi.org/10.1109/VTCFall.2017.8287884

[4] Fjolla Ademaj, Stefan Schwarz, Taulant Berisha, and Markus Rupp. 2019. A Spatial Consistency Model for Geometry-Based Stochastic Channels. *IEEE Access* 7 (2019), 183414–183427. https://doi.org/10.1109/ACCESS.2019.2958154

[5] Gouranga Charan, Tawfik Osman, Andrew Hredzak, Ngwe Thawdar, and Ahmed Alkhateeb. 2022. Vision-Position Multi-Modal Beam Prediction Using Real Millimeter Wave Datasets. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. 2727–2731. https://doi.org/10.1109/WCNC51071.2022.9771835

[6] Mustafa Emara, Miltiades C Filippou, and Ingolf Karls. 2018. Availability and reliability of wireless links in 5G systems: A space-time approach. In *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 1–6.

[7] Damir Hamidovic, Armin Hadziaganovic, Raheeb Muzaffar, and Hans-Peter Bernhard. 2023. 5G Campus Network Factory Floor Measurements with Varying Channel and QoS Flow Priorities. In *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*. 1–6. https://doi.org/10.1109/IECON51785.2023.10311715

[8] Asmae Mamane, Mohammed Fattah, Mohammed El Ghazi, Moulhime El Bekkali, Younes Balboul, and Said Mazer. 2022. Scheduling algorithms for 5G networks and beyond: Classification and survey. *IEEE Access* 10 (2022), 51643–51661.

[9] Doan Perdana, Aji Nur Sanyoto, and Yoseph Gustommy Bisono. 2019. Performance evaluation and comparison of scheduling algorithms on 5G networks using network simulator. *International Journal of Computers Communications & Control* 14, 4 (2019), 530–539.

[10] Jian Xiong, Hengrui Hu, Peng Cheng, Can Yang, Zhiping Shi, and Lin Gui. 2022. Wireless Resource Scheduling for High Mobility Scenarios: A Combined Traffic and Channel Quality Prediction Approach. *IEEE Transactions on Broadcasting* 68, 3 (2022), 712–722. https://doi.org/10.1109/TBC.2022.3141609

[11] Yong Zeng and Xiaoli Xu. 2021. Toward Environment-Aware 6G Communications via Channel Knowledge Map. *IEEE Wireless Communications* 28, 3 (2021), 84–91. https://doi.org/10.1109/MWC.001.2000327

[12] Yibo Zhou, Fengxiao Tang, Yuichi Kawamoto, and Nei Kato. 2019. Reinforcement learning-based radio resource control in 5G vehicular network. *IEEE Wireless Communications Letters* 9, 5 (2019), 611–614.