A Federated Reinforcement Learning Framework for Incumbent Technologies in Beyond 5G Networks

Rashid Ali, Yousaf Bin Zikria, Sahil Garg, Ali Kashif Bashir, Mohammad S. Obaidat, and Hyung Seok Kim

Abstract

Incumbent wireless technologies for futuristic fifth generation (5G) and beyond 5G (B5G) networks, such as IEEE 802.11ax (WiFi), are vital to provide ubiquitous ultra-reliable and low-latency communication services with massively connected devices. Amalgamating WiFi networks with 5G/B5G networks has attracted strong researcher interest over the past two decades, because over 70 percent of mobile data traffic is generated by WiFi devices. However, WiFi channel resource scarcity for 5G/B5G is becoming ever more critical. One current problem regarding channel resource allocation is channel collision handling due to increased user densities. Reinforcement learning (RL) algorithms have recently helped develop prominent behaviorist learning techniques for resource allocation in 5G/B5G networks. An agent optimizes its behavior in an RL-based algorithm based on reward and accumulated value. However, densely deployed WiFi environments are distributed and dynamic, with frequent changes. Thus, relying on individual local estimations leads to higher error variance. Therefore, this article proposes a federated RL-based channel resource allocation framework for 5G/B5G networks, and suggests collaborating learning estimates for faster learning convergence. Experimental results verify that the proposed approach optimizes WiFi performance in terms of throughput by collaborative channel access parameter selection. This study also highlights six potential applications for the proposed framework.

Introduction

Fifth generation (5G) and beyond 5G (B5G) networks are promising to cooperate with future wireless local area networks (WLANs) as their incumbent technologies to provide inspiring services, such as ubiquitous ultra-reliable low-latency communication (URLLC) and high throughput services. Amalgamating WLAN (WiFi) networks with 5G/B5G networks has been an active research topic over the past two decades. Currently, over 70 percent of mobile data traffic is generated by WLAN networks [1], and given WiFi's favorable economics and high performance, it will remain an attractive choice for indoor and enterprise applications. Although cellular networks originated outdoors, we expect WiFi and 5G/B5G to coexist both indoors and outdoors. The IEEE Working Group (WG) recently launched an amendment to IEEE 802.11 WLANs, IEEE 802.11ax high-efficiency WLAN (HEW), that addresses massively connected device deployment scenarios, including train stations, sport stadiums, shopping malls, and so on. The International Telecommunication Union Radio Communication Sector has specified that 5G/B5G and future WiFi requirements, such as HEW, will potentially satisfy these requirements, allowing combination of environment features and device interactions with the environments to spontaneously manage channel resource allocation parameters at the medium access control (MAC) layer. WLAN devices, also referred to as stations (STAs), proficiently and dynamically manage wireless channel resources; for example, the MAC layer distributed coordination function utilizes a carrier sense multiple access with collision avoidance (CSMA/CA) mechanism to resolve collision issues in the network. Station performance generally relies on exploiting uncertainty due to system heterogeneity in terms of transmitted data variety. Therefore, it is imperative to examine effective and robust resource allocation schemes to accomplish HEW targeted objectives.

Reinforcement learning (RL) is the least demanding machine learning (ML) sub-field suitable for this challenge. RL is a difficult technique to realize, but essentially, it trains the model using cooperation and feedback (i.e., rewards for actions in an environment). Several learners perceive and interpret their environment, and then take actions and interact with it. An agent learns the environment and defines which actions generate the highest rewards. Subsequently, the agent performs various actions within its environment to change the environment's state, where each action generates a reward. The agent's policy maps environmental situations to actions (i.e., provides decision making). Value accumulates over time from rewards received implementing a specific policy. The Markov decision process (MDP) is a probabilistic model for successive decision problems' current states and actions performed to determine a probability distribution of prospective states. Finally, the agent's environmental view maps state-action pairs to probability distributions

Rashid Ali and Hyung Seok Kim (corresponding author) are with Sejong University; Yousaf Bin Zikria is with Yeungnam University; Sahil Garg is with the École de Technologie Supérieure; Ali Kashif Bashir is with Manchester Metropolitan University and National University of Science and Technology; Mohammad S. Obaidat is with the University of Sharjah and the University of Jordan, and also with the University of Science and Technology Beijing.

Digital Object Identifier:
10.1109/MNET.011.2000611

0890-8044/21/$25.00 © 2021 IEEE

over the states, creating the environment model. RL is a flourishing research technique for studying 5G/B5G system use cases, ranging from learning complex 5G/B5G networks with unfamiliar channel resource allocation parameters to deploying 5G/B5G networks [2]. It was inspired by behaviorist learning, where learners accomplish their goals by interacting with their environment, and employs explicit learning algorithms, such as Q-learning (QL), to solve MDP models [3].

Estimation parameters (e.g., Q-value estimates) are maintained and dynamically updated in RL as information comes to hand during learning. Excessive estimator variance during learning can be problematic, resulting in uneven or unstable learning, or even making effective learning impossible. Thus, few agents will fail to build decision policies using an RL algorithm with only local rewards available due to the trade-off between exploration and exploitation increasing error variance.

Therefore, federated learning (FL) has recently emerged, training an ML algorithm on multiple local datasets contained in local devices without exchanging data samples. The general principle comprises training local models on local data samples and exchanging parameters (e.g., deep neural network weights) between these local models at some frequency to generate a global model. This article proposes a federated reinforcement learning (FRL) framework, extending FL techniques into RL models. FRL aims to learn a local Q-network policy for each agent by collaborating locally accumulated reward information among nearby agents, based on the assumption that all agents benefit from joining the federation in building decision policies.

Reinforcement Learning Distributed Setup

RL includes an MDP, where an agent takes activities in a stochastic environment over a sequence of time steps to maximize their value function collected from the environment. MDPs have been incorporated by many RL algorithms, such as Q-learning (QL) algorithms. Although QL algorithms were originally developed for single-learner tasks, several practical RL algorithms for 5G/B5G resource allocations involve multiple learners operating in a distributed fashion, such as autonomous resource provisioning and resource customization [4], autonomous resource slicing with dueling deep Q-networks [5], and dynamic resource reservation with deep RL algorithms [6]. The distributed setup for RL incorporates a central device, such as an access point (AP), which arranges learning procedures for all agents in the environment. Different agents intend to maximize accumulated reward using coordinated efforts in the environment.

Similar learning models have previously been considered for distributed supervised learning, such as FL. Collaboration among distributed learners in FL involves central devices exchanging data with all learners in the environment by gathering their rewards and local perceptions. However, this requires a central controller and learners to communicate often, and such frequent communication can become expensive and increase latency for most FL applications, including cloud-edge artificial intelligence (AI) frameworks, becoming a system performance bottleneck. Thus, improved communication-efficient approaches are essential for URLLC and high-throughput services under 5G/B5G systems.

Reinforcement learning (RL) is the least demanding machine learning (ML) sub-field suitable for this challenge. RL is a difficult technique to realize, but essentially it trains the model using cooperation and feedback (i.e., rewards for actions in an environment). Several learners perceive and interpret their environment, and then take actions and interact with it.

Distributed Setup of RL: Federated Reinforcement Learning

FRL, also known as collaborative RL, is one of the distributed setups of RL that learns the environment across numerous decentralized devices without sharing their actual data. This ML technique is very different from traditional centralized ML techniques, where all target data must be uploaded to a single server. FRL empowers numerous agents to construct typical strong ML models without sharing data and information. This decentralized technique tends to address security and privacy concerns by disseminating the information to distributed agents in the environment. FRL applications are already recognized by several next-generation technologies, such as 5G/B5G, the Internet of Things (IoT), and blockchain [5–10]. Therefore, this article extends FL techniques to RL models, creating an FRL model.

Article Contributions

Traditionally, WLAN CSMA/CA mechanisms use binary exponential backoff (BEB) to maintain a nearly collision-free environment. BEB uses a randomly selected backoff value from a contention window (CW) to contend for wireless channel resources. Initial and maximum CW are fixed by standardization, and an STA exponentially increases initial CW size each time it encounters a collision (collision in the WLAN is assumed if transmission acknowledgment is not received) until it reaches maximum CW. CW size is reset to its initial value once a packet is transmitted successfully. However, this blind increase/decrease of contention parameters degrades network performance; for example, resetting CW to its initial (minimum) size causes more collisions due to smaller CW for a more substantial number of contenders. Similarly, only a few STAs contend for channel resources when network density is minimal, and exponential CW increase for coincidental collisions causes unnecessary network delays. Thus, WLAN resources are drastically constrained due to shared channel resources, whereas applications become dynamically refined and diverse. An RL-based framework for channel resource allocation at the medium access control (MAC) layer in dense WLANs was proposed considering possible RL applications and features in WLANs [10]. The proposed RL-based framework utilized the QL algorithm to select optimal contention parameters for contending STAs, called an *intelligent* QL-based resource allocation (iQRA) mechanism. The authors showed that iQRA optimized BEB performance by utilizing channel-observation-based collision probabilities. In iQRA, each STA manages contention parameters based on observed collision probability, which was iterative-

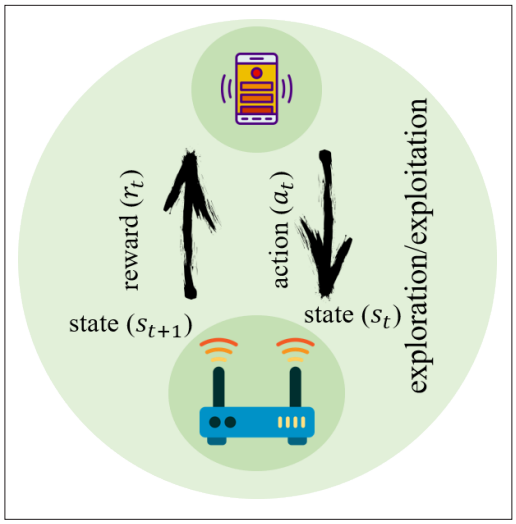

FIGURE 1. Example reinforced learning framework with core elements in a WLAN environment.

ly optimized using the QL algorithm.

However, the proposed mechanism forces every STA to optimize its contention parameters based on individually observed and accumulated Q-values from the value function. Wireless network environments are distributed and dynamic by nature, and change often. Thus, relying on individual local learning models (LLMs), such as RLs, leads to higher error variance, particularly when spectrum resources are shared among a massive number of connected networks and devices. Therefore, this article proposes an ML-aware architecture for next-generation dense WLANs and beyond, using an FRL model to optimize spectrum resource sharing. FRL is an ML technique where learners collaborate by interacting across multiple distributed edge devices holding LLMs without sharing actual data. This enables numerous learners to build a shared global learning model (GLM) based on their learning without sharing the local data, retaining critical data privacy and security. The FRL model infers distributed complex patterns from dense deployments, ensuring device minimum spectrum resource requirements. Therefore, we propose an FRL framework to optimize network contention parameters collectively and cooperatively in ultra-dense WLANs, and discuss six potential FRL framework applications for incumbent technologies in 5G/B5G networks.

Federated Reinforcement Learning Framework

This section describes the proposed FRL model for WiFi channel resource allocation.

Reinforcement Learning as a Framework

In RL-based frameworks for WiFi, an STA iteratively learns its action behaviors at a given time and maps them to prospective decisions to maximize reward, where the reward is a numerical response from the environment at a specific state. Typically, a learning STA in RL does not have any initial preference for actions to perform, but must discover optimal actions to achieve the best reward from the environment. The primary goal is to maximize the accumulated reward. Rewards for any action articulate how pleased the STA is in any specific state, and hence are the key motivation for changing the policy at any state. For example, the strategy of selecting low reward actions might be changed to choose other actions for a specific state in the future.

Another important RL technique is Q-value, that is, the aggregated reward collected over a long term. Reward for an action represents what is good in an immediate sense, whereas Q-value is the most important outcome when assessing decisions. For example, a state may reliably yield low reward, but have high Q-value if it is consistently trailed by various states that produce high rewards. Thus, we generally look for actions that yield high Q-value rather than high reward.

One RL challenge is the trade-off between exploration and exploitation. An RL-enabled STA must learn toward actions already attempted and check to be convincing in making a reward to get a significant accumulated reward. However, the STA needs to try actions that have not been picked previously, known as exploration, to find higher reward actions; as well as exploiting what has already been explored, under the constraint that the ultimate goal is to acquire the maximized accumulated reward, known as exploitation. Figure 1 shows an RL-based framework with core terminologies in a WLAN environment.

iQRA Mechanism

The RL-based iQRA mechanism assumes backoff stages as an available finite set of states, where a learning STA changes CW size by moving forward and backward in the states set. An action receives a reward at a given time in a specific state, with the objective to exploit its accumulated Q-value. One of the main iQRA objectives is to minimize WLAN channel collision probability. Therefore, the reward given by an action taken at a specific time is formulated as a function of channel collision probability [11], and then the STA observes its current state and takes an action. This action moves the STA to the next state. The iQRA mechanism aims to find an optimal policy that exploits accumulated Q-value, which is updated as

$$Q_t(s_t, a_t) = (1 - \alpha) \times Q_t(s_t, a_t) + \alpha \times \{r_t(s_t, a_t) + \beta \times max_{a'} Q_t(s', a')\}, \quad (1)$$

that is, the Q-value $Q_t(s_t, a_t)$ is updated iteratively after the STA performs action a_t at state s_t, where $\alpha \in [0, 1]$ is a learning-rate control parameter; $\beta \in [0, 1]$ is the discount factor to weight instant reward more aggressively than future reward; $r_t(s_t, a_t)$ is the current reward for action at state s_t; $r_t(s_t, a_t) + \beta \times max_{a'} Q_t(s', a')$ are learning estimates represented by ΔQ, where $max_{a'} Q_t(s', a')$ is the best estimated Q-value for the future state-action pair.

Thus, α directly effects an agent's learning, since $\alpha = 0$ means the Q-values are never updated, and hence nothing was learned by the agent, whereas $\alpha = 0.9$ means that learning can occur instantly or faster. Similarly, β models that future rewards are worth less than immediate rewards. Thus, the lower the value of beta is, the lesser will be the worth of future rewards, and vice versa for higher worth. An individually estimated Q-value at each STA may produce a higher error variance due to a dynamic and/or dense WLAN environment, leading to more severe estimations.

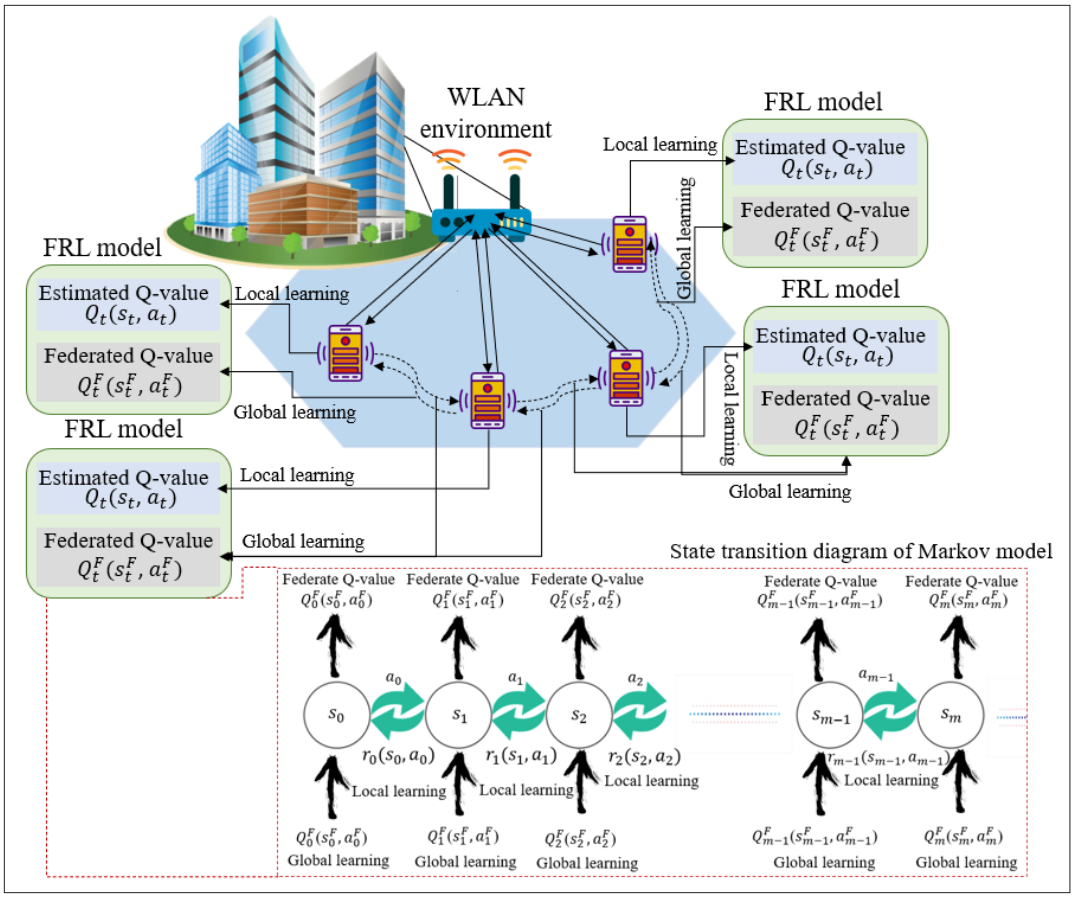

FIGURE 2. Proposed FRL framework for channel resource allocation in 5G/B5G WLANs (a state transition diagram of Markov decision process is also included at the bottom of the figure).

Therefore, we propose collaboration of the accumulated Q-value in the network with other STAs to reduce estimation error variance.

Proposed FRL Model

We combine RL and FL techniques to formulate the proposed FRL model. Combining RL and FL is performed considering local and global learning phases. Figure 2 shows the proposed FRL model, where we assume a WLAN AP as the WLAN environment's centralized device. A state transition diagram for the proposed MDP is also included (Fig. 2, bottom). The MDP defines that an STA moves from one state to another based on the performed action, while returning a reward back to the state. The STA accumulates the collected reward with the reward received from the global learning model, which is later federated within the network as a global Q-value. STAs around the AP have LLMs (e.g., RL-based iQRA) to optimize channel access parameters for transmission to and from the AP (estimated Q-value). In this LLM, an STA senses the channel for observation-based collision probability as formulated by iQRA. Figure 3 shows that under iQRA, competing STAs perform a BEB procedure for channel resources with a random backoff value after the channel is sensed idle for a distributed inter-frame space period. Discretized time slots during the BEB procedure are observed as either idle or busy. An STA in a WLAN environment formulates channel collision probability as the sum of total busy time slots divided by total number of time slots [11], which forms the reward for the STA to accumulate its Q-value further from Eq. 1. Since individual estimated Q-value in the proposed FRL model may suffer from large overestimation due to error variances, every STA integrates its locally updated value in an acknowledgment (ACK) packet, that is, a federated ACK (FACK) message, to collaborate with other active STAs in the WLAN (Fig. 3). This extra federated Q-value, referred to as double Q-learning (DQL) [12], is passed to the QL algorithm (Fig. 2), which is the GLM for the proposed FRL. Equation 1 shows that action a′ has maximum value for prospective state s′, depending on $Q_t(s_t,a_t)$. However, rather than using $\max_{a'}Q_t(s', a')$ to update Q-value, FRL uses the globally updated value received in FACK messages. Since FRL updates federated Q-value in the same WLAN environment, but with a different set of observations, it represents a fair Q-value estimate for this action. Both Q-value functions (LLM and GLM) must learn from separate sets of experiences in the same environment under the proposed FRL framework. Thus, an STA uses both Q-value functions to update its optimal Q-value for Eq. 1,

$$Q_t(s', a') = (1 - \alpha) \times Q_t(s', a') + \alpha \times \{r_t(s_t, a_t) + \beta \times Q_t^F(s_t^F, a_t^F)\}, \quad (2)$$

where $Q_t^F(s_t^F, a_t^F)$ represents a federated Q-value from other STAs in the wireless network.

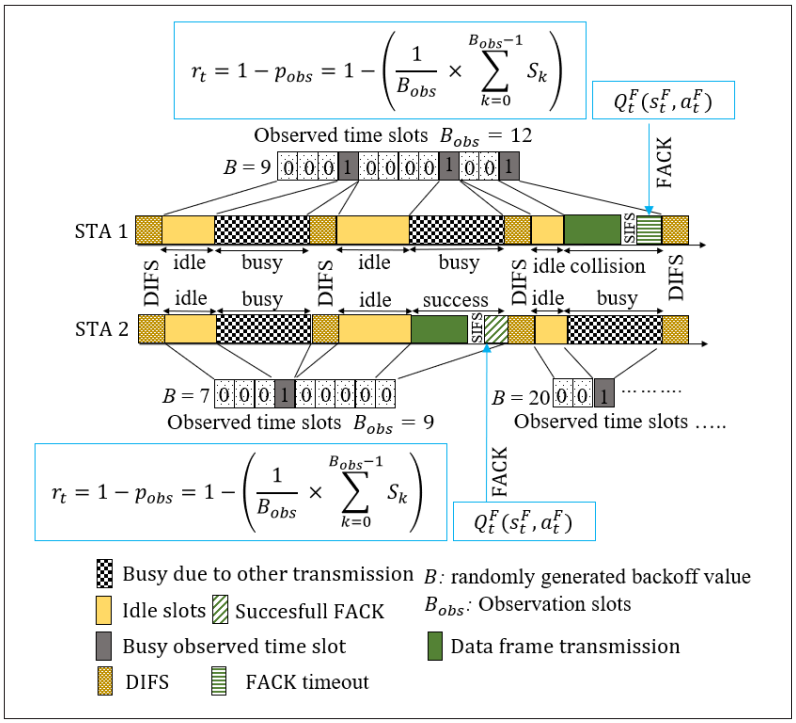

FIGURE 3. FACK collaboration method in the CSMA/CA module.

Performance Evaluation

We simulated the proposed FRL and RL-based channel access mechanisms using the ns-3 network simulator, with an IEEE 802.11ax HEW model for dense WLANs. Table 1 shows some important simulation parameters. Figure 4a compares Q-value learning estimate (ΔQ) convergence comparison between an RL (QL) and FRL-based (DQL) algorithm for 1000 iterations. FRL converges considerably faster than RL, suggesting significant performance enhancement. Rapid convergence helps STAs learn their WLAN environment swiftly and hence optimize their resource allocations. The reason for this faster convergence is that the RL mechanism depends on an individual/local estimator (Q-value accumulator), whereas the FRL mechanism depends on the double estimator (individual/local and federated/global).

Figure 4b compares the proposed FRL-based MAC layer channel access mechanism (FiQRA) in terms of throughput with respect to number of STAs with iQRA. The proposed FRL-based mechanism exhibits considerably improved throughput compared to the RL-based scheme. Consequently, FRL is more effective at learning the wireless network.

Substantial Impact

Limitations and challenges posed by current ML frameworks are always ignored, including believing and optimization based on individual learning. The proposed FRL framework expands wireless network capabilities by implementing a distributed and coordinated spectrum resource, providing a more dynamic and robust technique. This opens new approaches beyond conventional RL-enabled frameworks to overcome spectrum resource sharing challenges. Higher throughput verifies that shifting wireless networks from RL- to FRL-based mechanisms has potential for 5G and B5G networks.

Parameter type	Value
Frequency	5 GHz
Channel bandwidth	160 MHz
Data rate (MCS11)	1201 Mb/s
Payload size	1472 bytes
Transmission range	10 m
CW_{min}	32
CW_{max}	1024
Simulation time	500 s
Propagation loss	LogDistancePropagation
Mobility	ConstantPositionMobility
Rate-adaptation	ConstantRateWifiManager
Error-rate	NistErrorRateModel

TABLE 1. MAC/PHY layer simulation parameters for performance evaluation.

Potential Applications

This section presents some potential applications for the proposed FRL model.

FRL-Based iQRA Mechanism

Under iQRA, STAs optimize their contention parameters based on their individually estimated Q-value. The QL algorithm tends to perform poorly in dynamic and stochastic WLAN environments, mainly from substantial Q-value overestimation due to using $max_{a'}Q_t(s', a')$ in the QL algorithm. Therefore, we propose to use an FRL-based iQRA (FiQRA) mechanism where an STA competing for the channel resources is expected to learn the WLAN environment faster than an STA in an iQRA mechanism, as shown in Fig, 4a. Faster convergence allows the FRL-based MAC layer channel access mechanism to choose optimal channel access parameters (e.g., CW) faster than the iQRA mechanism.

Figure 4b shows that throughput for FRL-based MAC protocol is independent of the number of contending STAs in the WLAN. This is justified since system throughput depends on channel access parameters, such as backoff CW, and FiQRA optimizes CW based on federated information from other STAs within the WLAN.

Shared Information Network Estimation Method

Performance for a WLAN environment strongly relies on the number of contending STAs simultaneously trying to access channel resources. However, the actual number of active STAs cannot be retrieved even in the presence of an AP. Estimating the number of active devices raises many implications for an STA in a WLAN environment. The BEB CW depends on the number of STAs to maximize WLAN system performance [13]; hence, we propose to utilize FRL-based shared information network estimation (SINE). SINE uses a FACK message to collaborate an estimated number of STAs.

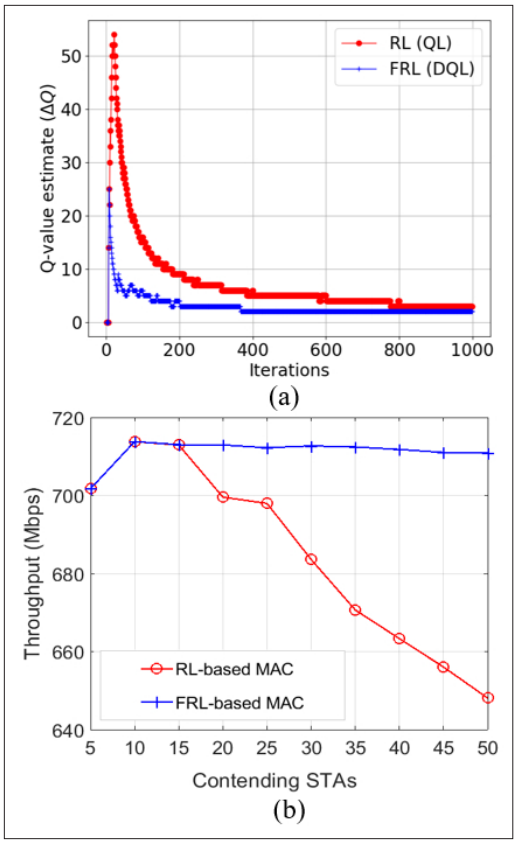

FIGURE 4. a) Q-value estimate (ΔQ convergence for FRL (DQL) and RL (QL) algorithms (n = 25); b) performance evaluation for the proposed FRL-based MAC layer channel access mechanism with the RL-based MAC layer channel access mechanism in terms of throughput.

FRL at the Edge

Edge computing has received considerable research attention due to its capacity to broaden cloud computing efficiencies to the network's edge with low latency. Various low-latency 5G/B5G applications utilize edge computing features, including autonomous driving cars, augmented reality (AR), remote surveillance, and tactile Internet (Fig. 5). However, connected edge user devices have rigorous computational resource constraints. One approach is to utilize a cloud network to furnish edge devices with on-demand computing resources, but characteristic deferral related to end-to-end communications with a cloud server can generate intolerable latency. Therefore, it is essential to utilize FRL at the edge (FEdge) to empower insightful applications.

Conventional RL-based frameworks use centralized learning data, which requires transferring information from numerous geographically distributed devices to a central device. In contrast, collaborative and edge deployed RL techniques, such as FRL, cope with environment data privacy challenges in RL mechanisms. FRL preserves data privacy in the network by avoiding centralized or single point learning. FEdge includes wireless devices within its environment and calculates the global learning model at the edge of the network.

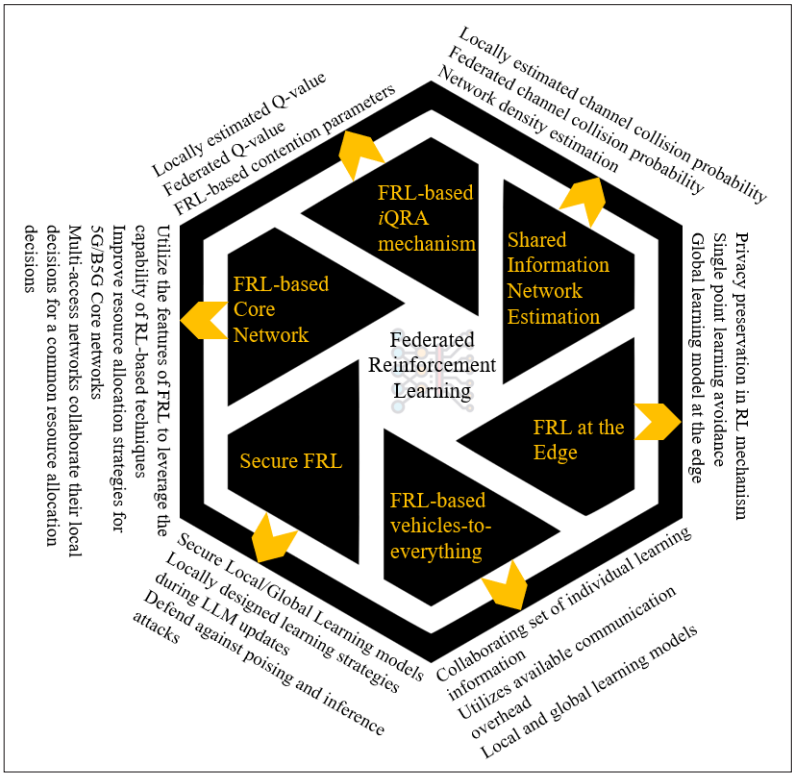

FIGURE 5. Potential FRL-based framework applications.

FRL-Based Vehicle-to-Everything Communications

Vehicle-to-everything (V2X) communication over 5G/B5G networks is a key enabler for self-driving autonomous cars. V2X communication also helps improve transportation frameworks' collaborative task handling [14]. However, autonomous services performance in self-driving cars, including instant navigation, collision evasion, and collaborative task handling, strongly depends on URLLC, with target end-to-end (E2E) latency < 1 ms. Most current related work focuses on improving V2X network expected latency to achieve this target, utilizing probabilistic controls to sustain queuing delays at the end devices.

However, although probabilistic control approaches may improve network reliability on shorter queue length devices, they fail to control exceptional actions for large queue length devices with low channel access probability. Therefore, a few end devices (vehicles) in V2X networks may encounter volatile delays, degrading network performance. The major issue with probabilistic methods is the lack of sufficient information samples that include rare extraordinary events [15]. In V2X communication, roadside units aid vehicles in gathering information samples over the network, but this incurs additional communication overhead.

As discussed above, FRL enables learning models to enable sharing of individual/local learning information with other devices within available communication overheads (e.g., FACK messages). The FRL framework also does not require synchronization among other devices in a V2X network due to implementing the QL algorithm locally. Thus, a vehicle can learn and explore the environment with the help of locally available information even if connectivity between the vehi-

Future work will implement the proposed FRL model for these applications and contribute to the research community. We encourage researchers from institutions and industry to consider the proposed FRL model for potential research and practical applications.

cle and roadside unit(s) fail.

Secure FRL

One of the critical 5G/B5G challenges while fulfilling URLLC requirements is to maintain data security and privacy. Traditional ML techniques need at least some private information to be analyzed at the central device to implement ML model training. Hence, centralized training models may cause potential security threats and privacy leakages within the network. In contrast, FL models for wireless networks enable trustworthy application scenarios for network devices. However, critical security and privacy challenges occur due to training data leakages. A brutal security attack is possible during GLM training, and user inference attacks can occur at any time, where rich semantic information may be breached at intermediate agents.

We propose FRL to overcome poising and user inference attacks in the networks. The FRL model uses value-based LLM training models in an automated and anonymous manner for each device in the network. LLMs have more robust defenses against user inference attacks using locally designed learning strategies during LLM updates. The proposed secure FRL framework creates a decentralized and trustless data store to eliminate dependency impacts on agents in the conventional FL model.

FRL-Based 5G/B5G Core Network

An essential improvement to the core network (CN) architecture is to be able to handle the intimidating 5G/B5G network services for URLLC requirements. A CN is the heart of the network and transmits multi-radio access in a network that requires seamless network-wide service experience. 3GPP [2] describes a CN architecture, called New Radio, to enable services transmission over wired/fixed networks, wireless networks, and converged access networks. A CN with strategically located servers at the edges and cloud is critical to support URLLC requirements. Effective network control and resource management mechanisms are required to enhance current deployed converged access network performance to support URLLC requirements.

RL has been considered previously as a promising solution to achieve intelligent and optimized network control and resource allocation. However, the major issue with RL techniques is that every RL-enabled device must learn the optimal decision through exploration. Thus, optimal actions are approached at the expense of excessive time in performing sub-optimal actions. Selecting sub-optimal actions during exploration could negatively affect overall performance. Therefore, we propose to utilize FRL to leverage RL capabilities and hence improve resource allocation strategies for 5G/B5G CN, while simultaneously achieving URLLC services through FRL techniques. The proposed FRL-based 5G CN enables multi-access networks to collaborate on their LLM-based decisions for common (GLM-based) resource allocation decisions. Multiple networks share their individual exploration to speed up learning optimal strategies, hence benefiting all participating converged networks.

Conclusion

The reinforcement learning (RL) behaviorist learning technique has been employed recently to enable machine intelligence in 5G and B5G wireless communication networks. RL is an emerging ML technique in numerous active 5G/B5G research areas. It employs explicit RL algorithms to learn unfamiliar wireless network environments and resource allocations in ultra-dense WLANs, for example, Q-learning to solve Markov decision process models.

However, wireless networks are dynamic environments that continuously change, and RL algorithms rely on individual estimations. Excessive estimator variance during learning may increase error variance. Therefore, we propose a federated RL (FRL) model for ML-enabled resource allocations in ultra-dense 5G/B5G wireless networks, such as IEEE 802.11ax WLANs, to overcome wireless channel collision issues. Experimental results verified that the proposed FRL model was superior to non-federated RL.

We also highlight six potential applications for the proposed FRL model in 5G/B5G: the FiQRA mechanism, SINE method, FEdge, FRL-based V2X communication, secure FRL, and FRL-based 5G/B5G core networks.

Future work will implement the proposed FRL model for these applications and contribute to the research community. We encourage researchers from institutions and industry to consider the proposed FRL model for potential research and practical applications.

Acknowledgment

Rashid Ali and Yousaf Bin Zikria are equally contributing first authors. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No.2019R1A4A1023746, No. 2019R1F1A1060799).

References

[1] M. Dano, "How Much Cellular and Wi-Fi Data Are Smartphone Users Consuming, and with Which Apps?" *FierceWireless*, Jan. 24, 2017; https://www.fiercewireless.com/wireless/howmuch-cellular-and-wi-fi-data-are-smartphone-users-consuming-andwhich- apps-verizon-0, accessed: Jan. 20, 2020.

[2] 3GPP, "Study on New Radio (NR) Access Technology (TR 38.912, TSG RAN)," Rel. 14, Mar. 2017; http: //www.3gpp.org/DynaReport/23799.htm.

[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 1998.

[4] G. Sun *et al.*, "Dynamic Reservation and Deep Reinforcement Learning Based Autonomous Resource Slicing for Virtualized Radio Access Networks," *IEEE Access*, vol. 7, 2019, pp. 45,758–72. DOI: 10.1109/ACCESS. 2019.2909670.

[5] D. Loghin *et al.*, "The Disruptions of 5G on Data-Driven Technologies and Applications," *IEEE Trans. Knowledge and Data Engineering*, vol. 32, June 2020, pp. 1179–98. DOI: 10.1109/TKDE.2020.2967670.

[6] S. Savazzi, M. Nicoli, and V. Rampa, "Federated Learning with Cooperating Devices: A Consensus Approach for Massive IoT Networks," *IEEE Internet of Things J.*, vol. 7, no. 5, May 2020, pp. 4641–54. DOI: 10.1109/JIOT.2020.2964162.

[7] Y. Lu *et al.*, "Differentially Private Asynchronous Federated

Learning for Mobile Edge Computing in Urban Informatics," *IEEE Trans. Industrial Informatics*, vol. 16, no. 3, Mar. 2020, pp. 2134–43. DOI: 10.1109/TII.2019.2942179.
[8] J. Kang *et al.*, "Incentive Design for Efficient Federated Learning in Mobile Networks: A Contract Theory Approach," *Proc. 2019 IEEE VTS Asia Pacific Wireless Communications Symp.*, Singapore, 2019, pp. 1–5. DOI: 10.1109/VTS-APWCS.2019.8851649.
[9] M. Yao *et al.*, "Artificial Intelligence Defined 5G Radio Access Networks," *IEEE Commun. Mag.*, vol. 57, no. 3, Mar. 2019, pp. 14–20. DOI: 10.1109/MCOM.2019.1800629.
[10] Z. Yu *et al.*, "Federated Learning-Based Proactive Content Caching in Edge Computing," *Proc. 2018 IEEE GLOBECOM*, Abu Dhabi, UAE, 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647616.
[11] R. Ali *et al.*, "Deep Reinforcement Learning Paradigm for Performance Optimization of Channel Observation-Based MAC Protocols in Dense WLANs," *IEEE Access*, vol. 7, Dec. 2018, pp. 3500–11. DOI: 10.1109/ACCESS.2018.2886216.
[12] H. V. Hasselt *et al.*, "Deep Reinforcement Learning with Double Q-Learning," *Proc. Assn. Advancement of Artificial Intelligence*, 2016, pp. 2094–2100.
[13] G. Bianchi and I. Tinnirello, "Kalman Filter Estimation of the Number of Competing Terminals in an IEEE 802.11 Network," *Proc. IEEE INFOCOM 2003,* San Francisco, CA, 2003, vol. 2, pp. 844–52. DOI: 10.1109/INFCOM.2003.1208922.
[14] S. Garg *et al.*, "Edge Computing-Based Security Framework for Big Data Analytics in VANETs," *IEEE Network*, vol. 33, no. 2, Mar./Apr. 2019, pp. 72–81. DOI: 10.1109/MNET.2019.1800239.
[15] S. Garg *et al.*, "SDN-Based Secure and Privacy-Preserving Scheme for Vehicular Networks: A 5G Perspective," *IEEE Trans. Vehic. Tech.*, vol. 68, no. 9, Sept. 2019, pp. 8421–34. DOI: 10.1109/TVT.2019.2917776.

Biographies

Rashid Ali [S'17, M'20] is currently an assistant professor at the School of Intelligent Mechatronics Engineering, Sejong University, Seoul, Korea. He received his Ph.D. degree (2019) in information and communication engineering from the Department of Information and Communication Engineering, Yeungnam University, Korea. His research interests include next-generation wireless networks, 5G and beyond, and reinforcement learning.

Yousaf Bin Zikria [SM' 17] is currently working as an assistant professor in the Department of Information and Communication Engineering, College of Engineering, Yeungnam University, South Korea. He received a Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University in 2016. He authored more than 80 peer-reviewed journals, conferences, patents, and book chapters.

Sahil Garg [S'15, M'18] received his Ph.D. degree from the Thapar Institute of Engineering and Technology, Patiala, India, in 2018. He is currently a postdoctoral research fellow at École de Technologie Supérieure, Université du Québec, Montréal, Canada. He has many research contributions in the area of machine learning, big data analytics, security and privacy, the Internet of Things, and cloud computing.

Ali Kashif Bashir [M'15, SM'16] is a senior lecturer/associate professor at Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. He has published 180+ articles in the fields of wireless networks, distributed systems, network/cyber security, network function virtualization, machine learning, and so on. He is serving as the Editor-in-Chief of the *IEEE Future Directions Newsletter*.

Mohammad S. Obaidat [S'85, M'86, SM'91, F'05] received his Ph.D. degree in computer engineering in computer science from The Ohio State University, Columbus. He has published more than 1000 refereed technical articles, about half of them journal articles, over 70 books, and about 70 book chapters. He is Editor-in-Chief of three scholarly journals and an Editor of many other international journals.

Hyung Seok Kim [S'02, M'06] received his B.S. degree from the School of Electrical Engineering, Seoul National University, South Korea, and his MS and Ph.D. degrees from the School of Electrical Engineering and Computer Engineering, Seoul National University, respectively. He is currently a faculty member with the Department of Information and Communication Engineering, Sejong University. His research interests include wireless communication networks.