IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 January 2024, accepted 18 February 2024, date of publication 27 February 2024, date of current version 11 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3371169

RESEARCH ARTICLE

Deep Reinforcement Learning Driven Joint Dynamic TDD and RRC Connection Management Scheme in Massive IoT Networks

JAEEUN PARK[1], (Student Member, IEEE), JOOHYUNG LEE[2], (Senior Member, IEEE), DAEJIN KIM[3], (Member, IEEE), AND JUN KYUN CHOI[1], (Senior Member, IEEE)

[1]School of Information and Communication Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea
[2]School of Computing, Gachon University, Seongnam-si 13120, South Korea
[3]Samsung Electronics, Suwon 16677, South Korea

Corresponding author: Joohyung Lee (j17.lee@gachon.ac.kr)

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) funded by Korean Government [Ministry of Science and Information and Communications Technology (MSIT)] through the Development of Sub-Terahertz (THz) Band Wireless Transmission and Access Core Technology for 6G Terabit per Second (Tbps) Data Rate under Grant 2021-000269.

ABSTRACT To support dramatically increasing services from internet of thing (IoT) devices with the sporadic and fluctuated generation of short packet traffic, this paper investigates joint dynamic time division duplexing (TDD) and radio resource control (RRC) connection management in a single-cell massive IoT network. Specifically, under the grant-free transmission incurring packet collision, this study models the factors affecting the time resource utilization (TRU) and energy consumption of IoT devices as a comprehensive system utility and further formulates the problem as a decision-making process aiming for balancing the long-term average TRU and energy consumption. To address the formulated problem, based on the deep reinforcement learning framework, this paper designs an experience-driven joint dynamic TDD and RRC connection management scheme that intelligently i) determines the TDD configuration based on the most recent downlink (DL)/uplink (UL) traffic demands and ii) adjusts the RRC state of each IoT device to control the maximum number of transmitting IoT devices. Finally, trace-driven simulation results demonstrate that the proposed scheme outperforms existing benchmarks, such as *Static TDD* and *Dynamic TDD*, in terms of transmission success ratio difference (TSRD) (up to 89% reduction), time resource utilization (TRU) (up to 17× increase), and energy consumption (up to 70% reduction) of IoT devices.

INDEX TERMS Internet of Things, time division duplexing, grant-free transmission, radio resource control, deep reinforcement learning.

I. INTRODUCTION

With a dramatic upsurge of applications from internet of thing (IoT) devices with the sporadic and fluctuated generation of short packet traffic, exemplified by healthcare, smart home or autonomous vehicles [1], the development of massive IoT networks has been attracting significant attention from industry and academia [2], [3], [4]. Furthermore, the diverse array of applications associated with IoT devices necessitates attention to various requirements, encompassing power consumption, access delay, privacy, and security [5]. In alignment with this trajectory, the 3rd Generation Partnership Project (3GPP) has introduced both narrowband IoT (NB-IoT) and LTE-M under cellular networks to accommodate a substantial number of IoT devices while delivering tailored services [6]. Nevertheless, since the rapidly increasing number of IoT devices generates massive amounts of sporadic and fluctuating traffic under limited bandwidth, efficient

The associate editor coordinating the review of this manuscript and approving it for publication was Chin-Feng Lai.

radio resource management for achieving maximum resource utilization with the required latency constraints remains challenging [7].

The conventional LTE scheduling [8], based on grants triggered by IoT devices through the random access (RA) procedure, poses a potential limitation in reducing uplink (UL) transmission latency. This is due to the mandatory steps involved in requesting access grants within the RA procedure. Recently, an alternative approach has been proposed to address this limitation by eliminating the request-grant step in UL transmission [9]. This method, termed grant-free transmission, enables IoT devices to promptly transmit packets upon arrival over the shared frequency resource. Grant-free transmission presents distinct advantages, particularly in reducing the time spent on the request and grant process inherent in grant-based scheduling. This approach is especially appealing for IoT devices characterized by periodic and sporadic traffic patterns [10]. However, it is crucial to consider the impact of increased UL traffic density on overall throughput when employing contention-based grant-free transmission, as this may lead to a deterioration in performance.

Recent concerns have prompted extensive performance analyses and studies focused on enhancing grant-free transmission where the number of transmitting devices is a great issue. Moreover, the fluctuations in downlink (DL)/UL traffic generated by IoT devices increase the complexity of deploying grant-free transmission for massive IoT networks. It is noteworthy that while IoT devices in massive IoT networks typically produce more UL traffic than DL traffic, this proportion can vary depending on their applications [11]. Addressing this issue, dynamic time-division duplexing (TDD) serves as an efficient solution to manage DL/UL traffic variations by dynamically adjusting TDD configurations in real-time. In recent studies on dynamic TDD [12], the management of dynamic TDD has mainly concentrated on determining optimal TDD configurations from a physical layer perspective. This involves detailed modeling of cross-link interference (CLI) and forward-link interference (FLI) between base stations and devices. The objective of this management is to enhance the overall throughput of systems. However, existing studies, to the best of our knowledge, lack a comprehensive consideration of simultaneously controlling the number of transmitting IoT devices and adjusting relevant dynamic TDD configurations to support grant-free transmissions in massive IoT networks. This problem cannot be formulated as a convex problem. Additionally, conventional heuristic algorithms prove impractical in real-time applications due to their high complexity. In light of these challenges, experience-driven AI algorithms present a promising avenue for addressing the complexity of this problem in massive IoT networks. Consequently, there exists room for improvement in prior works, thus motivating the present research.

This paper introduces a novel approach to address the challenges in massive IoT networks through a deep reinforcement learning (DRL)-driven joint dynamic TDD and RRC connection management scheme. Specifically, we utilize TDD configuration selection and RRC connection control modules to effectively adapt to the sporadic and fluctuating traffic in the network. Each module's problem is formulated as a decision-making process, with the objective of striking a balance between the long-term average time resource utilization (TRU) and energy consumption, achieved by minimizing the transmission success ratio difference (TSRD). Within the DRL framework, our approach involves designing experience-driven agents that collaboratively address two key aspects: i) determining the optimal TDD configuration based on the most recent DL/UL traffic demands and ii) modifying the RRC state of each IoT device to control the maximum number of transmitting IoT devices. The contributions of this paper are detailed in the following.

- This research introduces a comprehensive joint dynamic TDD and RRC connection management scheme tailored for massive IoT networks. Operating collaboratively, the TDD configuration selection and RRC connection control modules effectively manage the challenges posed by sporadic and fluctuating short-packet traffic generation.
- The interrelated problems addressed by these modules are formulated using a DRL framework to minimize the TSRD, effectively balancing long-term TRU and energy consumption. Leveraging the DRL framework, our approach efficiently resolves non-convex problems by deploying jointly operating experience-driven agents.
- The scheme incorporates the estimation of the most recent DL/UL traffic demands, enabling the TDD configuration selection module to determine the optimal TDD configuration. Subsequently, guided by the chosen TDD configuration, the RRC connection control module manages the maximum number of transmitting IoT devices. This is achieved by dynamically adjusting their RRC states to facilitate grant-free transmission while mitigating packet collision.
- Trace-driven simulation results demonstrate the efficacy of the proposed scheme, achieving an impressive 89% reduction in TSRD and a remarkable 17-fold increase in TRU compared to existing benchmarks. Additionally, the proposed scheme significantly reduces IoT device energy consumption, achieving up to a 70% reduction. Notably, retransmission accounts for less than 25% of energy consumption, while the existing benchmarks surpass 90% in energy consumption.

The rest of this paper is structured as follows: Section II provides a comprehensive overview of the relevant literature. In Section III, the system model for massive IoT networks is presented, accompanied by a mathematical formulation of the problem pertaining to joint dynamic TDD and RRC connection management. Section IV elaborates the design of our DRL approach employed for solving the articulated problem. To assess the effectiveness of the proposed scheme, Section V presents trace-driven simulation results. The findings are

then discussed in terms of performance evaluation. Lastly, Section VI outlines potential avenues for future research within the scope of the present investigation, concluding with final remarks in Section VII.

II. RELATED WORK

For the efficient management of dynamic TDD in a massive IoT network, the performance of the resource scheduling algorithm is crucial. Grant-free transmission under OFDMA is a viable UL resource scheduling algorithm for IoT devices, yet it faces challenges such as packet collisions due to the fluctuations in UL demand. To address this, dynamic TDD and RRC connection control offer feasible solutions to facilitate grant-free transmission in a massive IoT network. This section provides a brief overview of existing research related to massive IoT networks, dynamic TDD, grant-free transmission, and RRC connection control.

A. MASSIVE IOT NETWORK

In the context of massive IoT networks, NB-IoT emerges as an alternative solution strategically deployed by cellular networks to cater to the connectivity, reliability, and quality of service (QoS) requirements for a vast number of IoT devices [13]. A specific focus in [14] delves into the DL resource allocation of NB-IoT, aiming to mitigate latency and user blockage within the expansive landscape of a massive IoT network. However, the predominant transmission mode for IoT devices, primarily comprising upstream communication, necessitates an optimized UL transmission strategy to meet the network's demanding requirements [11], [15].

For instance, Ahmed et al. [16] specifically addresses upstream QoS concerns in terms of communication delay, security risks, and message failure, based on the devices' requirements in QoS-Aware IoT-Based Vehicular Networks. Jiang et al. [17] scrutinize the collision probability of RA schemes designed for requesting UL grants, proposing a novel RA scheme in the context of a massive IoT environment to reduce average queueing and waiting delays for UL transmissions. Additionally, the work in [18] enhances the reliability of UL transmission in RACH procedure through real-time traffic value analysis. However, unlike conventional human-type communications (HTC), where straightforward granting and scheduling are necessary due to the huge data transmission within the limited bandwidth, the sporadic and short-packet data transmissions from massively deployed devices in the networks pose a unique challenge in terms of latency. Consequently, numerous studies concentrate on enabling grant-free transmissions for UL resource scheduling of IoT devices to reduce the time required for granting UL access [19].

Moreover, energy efficiency is a critical consideration in massive IoT networks due to the devices being powered by limited-energy batteries [20]. To address this in [21], IoT devices compress collected medical data before transmission to the cloud, impacting the consumption of IoT devices. The transmitted data are then restored at the cloud using machine learning techniques without degrading the quality of the original data. Lee et al. [22] propose an approach to adaptively control the transmission period of IoT devices, extending the lifespan of devices while maintaining high data quality by adjusting the transmission period concerning the fluctuation of monitored values at IoT devices.

Recent studies on efficient energy consumption of IoT devices revolve around the network access of uplink IoT traffic [23]. With the development of grant-free transmission supporting sporadic generation of short packets, Azari et al. [24] conduct a comparative analysis on grant-free and traditional grant-based transmission in terms of reliability and energy consumption. They investigate traffic load regimes at which IoT devices benefit from grant-free transmission in terms of energy efficiency. Additionally, efforts on further reducing the energy consumed from the repetitive transmissions in grant-free transmission by controlling the number of repetitions are being developed while satisfying the reliability requirements [25], [26].

Furthermore, addressing the significant DL/UL traffic volume imbalance is crucial in massive IoT networks. While conventional networks explore dynamic TDD management [27], [28], [29], [30], [31], the unique context of massive IoT networks creates a notable research gap, exacerbated by significant DL and UL traffic differences. This paper contributes by implementing a dynamic TDD scheme under grant-free transmission, an additional consideration that aims to facilitate transmission in massive IoT environments. The proposed approach optimizes resource utilization and enhances overall network efficiency, specifically addressing the challenge of achieving a more balanced DL/UL traffic flow in massive IoT environments.

B. DYNAMIC TDD

Table 1 presents brief descriptions of considerations addressed in existing works concerning dynamic TDD. Additionally, the table incorporates our proposed work, aiming to distinguish various considerations from previous studies that predominantly cater to conventional networks serving human-centric devices, to recent investigations on TDD networks supporting MTDs.

To date, researchers have been actively working on optimizing TDD configurations in HTC networks, [32], [33], [34], [35], [36], [37], [38]. In [32], efforts focus on maximizing the overall DL/UL weighted sum rate in a multi-cell environment. This involves adjusting power allocation, scheduling, and TDD configuration. Analytical solutions assume DL/UL resource demands are uniform, finding a Nash equilibrium between base stations' maximization issues. Meanwhile, [33], [34] employ a Poisson point process to model traffic generation, resulting in analytical solutions that enhance dynamic TDD optimization by mitigating interference between base stations and increase overall system throughput.

However, the TDD configurations in the aforementioned studies are determined based on packet arrival rates of all

TABLE 1. Comparisons of existing works in dynamic TDD.

Papers	Environment	Target device		Scope				Management		
		HTD	MTD	DL/UL fairness	Throughput	Energy	TRU*	TDD configuration	Power allocation	Access control
[32]	5G; Multi-cell	✓			✓			✓	✓	
[33]	4G; Multi-cell	✓			✓			✓	✓	✓
[34]	5G; Heterogeneous	✓			✓	✓		✓	✓	✓
[35]	5G; Single cell	✓		✓			✓	✓		
[36]	5G; Heterogeneous	✓		✓	✓			✓		
[37]	D2D; Heterogeneous	✓	✓		✓			✓		
[38]	5G; Multi-cell	✓		✓	✓			✓		
[40]	5G; Multi-cell		✓							✓
[41]	Massive MIMO; Cell-free	✓		✓	✓			✓	✓	
[42]	5G; Beamforming; multi-cell	✓	✓			✓			✓	
[43]	5G/6G; Beamforming; multi-cell	✓	✓			✓			✓	
This work	5G; Grant-free; Single cell		✓	✓		✓	✓	✓		✓

TRU*: Time resource utilization

users rather than the actual occurrence of DL/UL traffic demands at the base station. Ding et al. [35] explore two different schemes, one allocating DL/UL subframes proportional to packet arrival rates and the other proportional to the actual DL/UL traffic demands. Despite achieving the same average allocation of DL/UL subframes, the latter, considering actual traffic generation, exhibits higher resource utilization. However, this approach demands rapid adjustment of TDD configuration at the end of every TDD frame in response to DL/UL traffic variations. In response to this issue, [36] introduces a dynamic TDD control system, employing the DRL approach. The DRL agents at base stations operate within a coordinated single-leader multi-follower Stackelberg game framework to dynamically determine TDD configurations, adhering to the imposed time constraints.

Furthermore, the integration of traffic demand estimation under the DRL framework is explored by [37] and [38] to enhance TDD configuration adjustment. Addressing user mobility in a heterogeneous network, Tang et al. [37] predict variations in DL/UL traffic demands for users moving between adjacent cells. Simultaneously, Esswie et al. [38] estimate UL traffic demand based on the retransmission probability for the Hybrid Automatic Repeat and Request (HARQ) scheme. Consequently, in a massive Internet of Things (IoT) network, the considerations of traffic demand estimation and the DRL framework are imperative for efficient DL/UL resource management and practical deployment.

While previous studies primarily focus on TDD deployment within human-type communication networks, several studies introduced in [39] shift their attention to TDD deployment within massive IoT networks where UL transmissions are dominated by the short and sporadic traffic generated by IoT devices. For instance, in [40], the study emphasizes the characteristics of IoT traffic when dynamically determining TDD configurations. This involves leveraging actual IoT traffic data collected at each access point, where DRL-based agents benefit from a training dataset composed of this real IoT data.

In contrast, Fukue et al. [41] highlight the energy constraints of IoT devices as they address the trade-off between system throughput and DL/UL traffic fairness in TDD management of IoT networks. Lee et al. [42], [43] conduct a comprehensive study on the coexistence of HTC and MTC, exploring power allocation in beamforming under dynamic TDD management. Various constraints, such as energy consumption and QoS, are considered.

While contemporary studies on TDD management in massive IoT networks actively account for characteristics and requirements posed by IoT devices, it is crucial not to overlook potential applications anticipated within the IoT network itself. These applications may differ from the conventional system network primarily focused on providing HTC services [44]. Therefore, for the comprehensive management of TDD in massive IoT networks, it is essential to consider the influence of network-specific transmission types, including grant-free transmission [45].

C. GRANT-FREE TRANSMISSION AND RRC CONNECTION CONTROL

Numerous studies have contributed to the understanding of grant-free transmission and its implications on reliability and efficiency. Jacobsen et al. [46] conducted an evaluation in a large urban macro network scenario, highlighting the superior performance of grant-free transmission in terms of lower latency and successful packet delivery compared to grant-based schemes. Berardinelli et al. [47] delved into the outage probability of different variations of grant-free OFDMA transmission. Notably, the blind scheme, known as K-repetition [48], demonstrated the lowest outage probability with sufficient frequency bandwidth, yet exhibited performance degradation with increasing packet arrival rates due to frequent collisions. In this regard, Kim et al. [49] proposed an analytical framework for grant-free multiple access protocols in the IoT scenario with sporadic traffic. Their results underscored that, with the same multi-packet reception capability and resource blocks, a higher multi-packet reception capability showed a greater influence on increasing throughput and reducing latency by alleviating the packet collision issues in the grant-free transmission.

Furthermore, to support a higher number of devices in massive IoT networks, the implementation of grant-free transmission in non-orthogonal multiple access (NOMA)

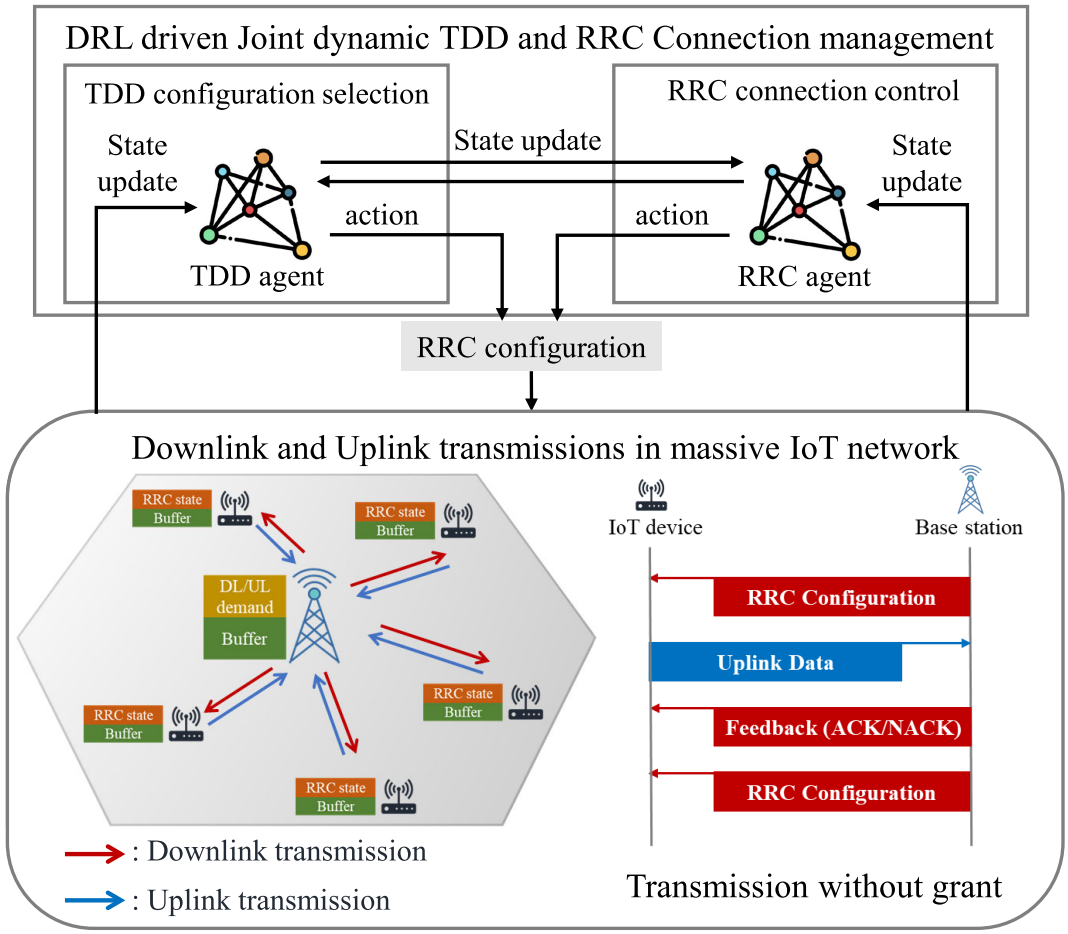


FIGURE 1. Proposed system model in a single-cell massive IoT network environment.

[50], [51] has been explored. From a theoretical perspective, NOMA gains higher capacity than OFDMA, by overloading non-orthogonal resource allocation. In the context of grant-free NOMA systems, advancements were made with dynamic compressive sensing [52], where block error rates in grant-free NOMA systems are lower compared to conventional CS. Additionally, Zhang et al. [53] introduced a DRL-based grant-free NOMA algorithm for subchannel and discrete power level selection, achieving highly successful access. However, computational resource constraints poses high challenges in applying NOMA to massive IoT networks.

On the other hand, to handle the challenge posed by a substantial number of devices in implementing grant-free transmission under OFDMA within a massive IoT network, the management of RRC connection control emerges as a practical solution [54]. For instance, Han et al. [55] introduce an RRC assignment method aimed at controlling random access collisions in the context of massive machine-type communication. In the study by Leyva et al. [56], a novel access class barring approach based on RRC signaling is proposed to enhance network access performance during periods of high congestion in machine-type communications. Addressing energy consumption concerns, the regulation of device transmissions is achieved by selectively defining RRC states based on their traffic generation, thereby enhancing energy efficiency [57]. In light of these considerations, this paper also leverages RRC connection control to facilitate grant-free transmission based on OFDMA in massive IoT networks, offering increased successful transmissions and reduced energy consumption aligned with the demands of massive IoT deployments.

III. SYSTEM MODEL

A. NETWORK MODEL

The proposed system model, depicted in Fig. 1, delineates our approach within a single-cell massive IoT network under TDD mode for DL/UL transmissions. A proportional fair scheduling algorithm, specifically discussed in [38] and [36], is adopted for DL resource scheduling to fairly transmit DL packets across a myriad of IoT devices, without bias towards any specific device [58]. Conversely, for UL resource scheduling, the system employs grant-free transmission, specifically Transmission Without Grant

(TWG) of Type1, as defined in New Radio (NR) Release 17 [59]. This strategy accommodates the sporadic generation of UL packets, benefiting from reduced contention for resource blocks among a moderate number of IoT devices [60].

In the proposed scheme, depicted in the bottom right of Fig. 1, the base station transmits a RRC configuration signal to IoT devices. This signal encapsulates TDD configurations and RRC states of each device, signifying either RRC IDLE or RRC CONNECTED. For simplicity, we denote RRC states in this paper as IDLE or CONNECTED. CONNECTED IoT devices transmit UL data based on the predetermined TDD configuration structure. Upon a successful UL transmission, the base station issues a feedback signal, acknowledging the successful transmission. This process repeats in subsequent periods, with new RRC configuration signals issued to IoT devices.

Our approach integrates TDD configuration selection and RRC connection control modules to manage fluctuating DL/UL traffic demands in a single-cell massive IoT environment. Specifically, DRL-driven TDD and RRC agents jointly determine TDD configurations and RRC states, respectively, as illustrated in Fig. 1. The agents update their states through information exchange and observations within the massive IoT network. The TDD agent determines the TDD configuration index, representing the ratio of DL/UL subframes in the determined configuration. Simultaneously, the RRC agent adjusts the number of CONNECTED IoT devices and randomly defines their RRC states. These parameters are then included in the RRC configuration signal transmitted to IoT devices.

Each period comprises 10 time slots with a 1ms duration, allowing for the allocation of DL or UL subframes. The total number of DL/UL subframes in each period consistently equals 10. The RRC configuration signal is dynamically modified in every period through joint management by the TDD and RRC agents, aiming to address sporadic DL/UL traffic generation. The objective is to minimize the difference in the successful transmission ratio of DL/UL traffic and enhance the number of successful UL transmissions. This adaptive approach underscores the efficiency and adaptability of our proposed system model in addressing the unique challenges of a single-cell massive IoT network.

B. TRAFFIC DEMAND MODEL

The system's traffic demand model, a pivotal element of our proposed approach, quantifies the DL/UL traffic demands by expressing the total number of DL/UL packets slated for transmission. In alignment with the sporadic packet generation characteristic of IoT devices [47], [49], [61], we model the packet generation process with a Poisson arrival rate. The probabilities of DL/UL packet arrivals, denoted as q_{DL} and q_{UL}, can be computed, considering that up to one DL/UL packet can be generated for each IoT device during each period T_{period}:

$$q_{DL} = 1 - e^{-\lambda_{DL} T_{period}} \tag{1}$$

and

$$q_{UL} = 1 - e^{-\lambda_{UL} T_{period}}, \tag{2}$$

where λ_{DL} and λ_{UL} are the DL/UL packet arrival rates, respectively. Packets are generated at the commencement of each period and transmitted to the base station or IoT devices based on their resource scheduling algorithms. To simplify the model, we assume that packets not successfully transmitted in one period are reserved until the next period for retransmission [62].

The DL/UL traffic demands at period t, represented as $Q_{DL}(t)$ and $Q_{UL}(t)$, respectively, can be expressed as the sum of packets reserved from the previous period and the packets generated at the start of period t. Specifically, for $Q_{DL}(t)$, the value cannot surpass the base station's buffer size, denoted as $Q_{DL,\max}$, such as $Q_{DL}(t) \leq Q_{DL,\max}$ for $t \in T$, where T signifies a set of periods. Moreover, considering the limited buffer size of IoT devices [63], each device can store up to one packet, ensuring that the maximum UL traffic demand does not exceed the number of UL IoT devices, i.e., $Q_{UL}(t) \leq N_{UL}$ for $t \in T$, where N_{UL} is the number of UL IoT devices.

Dynamic TDD, as motivated by [26], seeks to align DL/UL subframe resources with variations in traffic demands. Transmission Success Ratios (TSRs) for DL/UL at period t, denoted as $\Omega_{DL}(t)$ and $\Omega_{UL}(t)$, respectively, are defined as follows:

$$\Omega_{DL}(t) = \frac{P_{DL}(t)}{Q_{DL}(t)} \tag{3}$$

and

$$\Omega_{UL}(t) = \frac{P_{UL}(t)}{Q_{UL}(t)} \tag{4}$$

where $P_{DL}(t)$ and $P_{UL}(t)$ represent the number of successfully transmitted DL/UL packets during period t. Consequently, $\Omega_{DL}(t)$ and $\Omega_{UL}(t)$ delineate the ratios of successfully transmitted packets to the total demands at period t. These values guide the adjustment of DL/UL subframe resource allocation through TDD configuration selection. For instance, a lower $\Omega_{DL}(t)$ compared to $\Omega_{UL}(t)$ can prompt the selection of a TDD configuration with more DL subframe resources to increase $\Omega_{DL}(t+1)$ in the subsequent period.

C. DOWNLINK AND UPLINK TRANSMISSION

The DL resource scheduling within our proposed system model is underpinned by a user-prioritized proportional fair algorithm, as detailed in [64]. The scheduling prioritization coefficient for each DL is expressed as:

$$P = \frac{D^{\alpha}}{R^{\beta}} \tag{5}$$

Here, α and β are set as 1 in [64], and D and R denote the current achievable data rate and the historical average data rate for the corresponding IoT device, respectively. The data rate for each IoT device in single-cell DL transmission is modeled using the Signal-to-Noise Ratio (SNR), impacted by the physical location of the device in relation to the base station and the noise parameter. The prioritization coefficient is influenced by the noise parameter of additive white Gaussian noise [36], [37], [38], considering that data rates are subject to various random processes of environmental noise. Consequently, DL resource scheduling prioritizes devices with lower Gaussian noise values, assuming that the scheduled DL packets are successfully transmitted, given the scheduling occurs under stabilized throughput conditions for devices [65].

In terms of UL resource scheduling, we implement K-repetition grant-free transmission from [47] in the TWG framework. Here, K denotes the number of repetitive transmissions of a single packet for each device, corresponding to the number of UL subframes in one TDD frame. Upon UL packet generation, IoT devices in the CONNECTED state transmit packets to the base station using K-repetition grant-free transmission. Successful UL transmission is acknowledged if the base station correctly receives any of the repetitive transmissions of a single packet, whereas unsuccessful transmission occurs when all transmissions fail due to packet collision. Although high multi-packet reception capability at the base station, as detailed in [49], can alleviate packet collisions, it cannot fully overcome them. Consequently, we deploy grant-free OFDMA transmission, where the multi-packet reception capability is conservatively set, ensuring that collisions between any number of packets induce packet failure at the base station. This strategic deployment enhances the robustness of UL transmission in the face of packet collisions, contributing to the reliability of our proposed system model.

D. TDD CONFIGURATION SELECTION

The primary objective of TDD configuration selection is to align the DL/UL TSRs as defined in (3) and (4). This alignment, aimed at reducing the disparity between DL/UL TSRs, enhances fairness in delivering services to DL/UL IoT devices, especially in the face of fluctuating traffic patterns. In massive IoT networks, the base station can measure both the number of successfully transmitted DL/UL packets and the DL traffic demand. However, unlike the UL resource scheduling outlined in [36], the actual amount of UL traffic demand is treated as an unknown factor due to practical considerations.

In [38], the estimation of UL traffic demand involves multiplying the reciprocal of the average block error rate by the number of retransmitted packets. Similarly, an estimation based on the success probability distribution of UL transmission is considered, where the success probability hinges on the failure probability due to packet collision. For example, with N_c shared resource blocks capable of carrying a single packet each and N_T simultaneously transmitting devices, the probabilities of packet collision, $p_c\left(N_T, k\right)$, and failure due to collision, $p_f\left(N_T, k\right)$, are calculated as:

$$p_c\left(N_T, k\right) = \binom{N_T}{k} \frac{\left(N_c - 1\right)^{N_T - 1 - k}}{N_c^{N_T - 1}}, \tag{6}$$

and

$$p_f\left(N_T, k\right) = p_c\left(N_T, k\right) 1\left(k < \delta\right), \tag{7}$$

where δ refers to the multi-packet reception capability of a base station, set to 2 to indicate that any collision with another packet results in packet failure [66]. The total failure probability, $p_f\left(N_T\right)$, is obtained by summing $p_f\left(N_T, k\right)$ over k from 0 to $N_T - 1$, which can be expressed as

$$p_f\left(N_T\right) = \sum_{k=0}^{N_T - 1} p_f\left(N_T, k\right). \tag{8}$$

Concerning K-repetition grant-free transmission, where the failure probability at each time slot in a TDD frame is independent, the probability of transmission failure for a single packet at every time slot in a TDD frame is computed by squaring $p_f\left(N_T\right)$ and multiplying it by the number of UL subframes in a TDD frame, f_{UL}. Taking this into consideration, the success probability is given by $1 - p_f\left(N_T\right)^{f_{UL}}$. This success probability is identically independently distributed (IID) for each IoT device, allowing the expected number of successfully transmitted packets to be calculated by multiplying the number of transmitting devices with the success probability.

The number of transmitting IoT devices is approximated to find the value at which the expected number of successfully transmitted packets is closest to the actual number. Based on the estimated number of transmitting IoT devices, $\tilde{N}_T$, the estimated amount of UL traffic demand, $\tilde{Q}_{UL}$, is computed as

$$\tilde{N}_T = \arg\min_N \left| N_a - N\left(1 - p_f\left(N\right)^{f_{UL}}\right) \right| \tag{9}$$

and

$$\tilde{Q}_{UL} = \frac{\tilde{N}_T}{N_s} N_{UL} \tag{10}$$

where N_a is the number of successfully received UL packets and N_s is that of the CONNECTED IoT devices. The solution in (9) is determined by an exhaustive search method, with computational overhead reduced by skipping rounds where the number of successfully received packets exceeds that of transmitting devices. As the solution represents the estimated number of transmitting IoT devices from randomly selected devices, the UL traffic demand can be estimated in proportion to the total number of UL IoT devices.

By adjusting the TDD configuration index at period t, c_t, the number of successful DL/UL transmissions, $P_{UL}\left(t\right)$ and $P_{DL}\left(t\right)$, can be controlled to reduce the difference of DL/UL TSRs. Therefore, based on the estimated value of $\tilde{Q}_{UL}$ and

measured values of Q_{DL}, the optimization problem of TDD configuration selection is formulated as **Problem 1**:

$$\begin{aligned} \underset{C}{\text{minimize}} \quad & \sum_{t\in T} \left| \Omega_{DL}(t) - \tilde{\Omega}_{UL}(t) \right| \\ \text{subject to} \quad & Q_{DL}(t) \le Q_{DL,\max} \\ & \tilde{Q}_{DL}(t) \le Q_{UL,\max}, \end{aligned}$$

over a set of TDD configuration indices, $C = \{c_1, c_2, \cdots, c_T\}$, where $\tilde{\Omega}_{UL}(t)$ is defined as $\frac{P_{UL}(t)}{\tilde{Q}_{UL}(t)}$ so that the objective function approximates the summation of the difference between DL/UL TSRs over a set of periods, T. To achieve high fairness in terms of DL/UL transmission in massive IoT networks, it is necessary to minimize the summation with respect to the varying amount of DL/UL demands. The constraints indicate the maximum amount of DL/UL demands that can be accommodated, such as $Q_{DL,\max}$ and $Q_{UL,\max}$, respectively.

E. RRC CONNECTION CONTROL

In the context of the determined TDD configuration from the TDD configuration selection, the focus of RRC connection control is to maximize the number of successful UL packets through the adjustment of the RRC state for each IoT device. Defining the RRC state for each IoT device facilitates control over the maximum number of UL transmissions, thereby mitigating packet collisions and increasing the overall success of UL transmissions. The RRC states and the success of UL transmissions for IoT devices are denoted as a connection vector, $\boldsymbol{\alpha}(t)$, and an observation vector, $\boldsymbol{\beta}(t)$, respectively:

$$\boldsymbol{\alpha}(t) = \left(\alpha_1(t), \alpha_2(t), \cdots, \alpha_{N_{UL}}(t)\right),$$

$$\alpha_i(t) = \begin{cases} 1, & \text{if RRC state of } i\text{-th device is CONNECTED} \\ 0, & \text{if RRC state of } i\text{-th device is IDLE} \end{cases}$$

$$\boldsymbol{\beta}(t) = \left(\beta_1(t), \beta_2(t), \cdots, \beta_{N_{UL}}(t)\right),$$

$$\beta_i(t) = \begin{cases} 1, & \text{if packet from } i\text{-th device is successful} \\ 0, & \text{otherwise.} \end{cases}$$

RRC connection control aims to determine the number of CONNECTED IoT devices at a given period t, denoted as $N_s(t)$. Subsequently, the CONNECTED IoT devices are randomly selected from N_{UL} devices, ensuring that the sum of all elements in $\boldsymbol{\alpha}(t)$ equals $N_s(t)$. Once a packet arrives, devices in the CONNECTED state transmit their packets to the base station, and the transmission results are reflected in $\boldsymbol{\beta}(t)$.

Due to the unknown packet arrival rate of each IoT device at the base station and the variability in the expected number of successful packets based on the number of transmitting IoT devices and UL subframes, RRC connection control adjusts the number of CONNECTED IoT devices at each period t, $N_s(t)$, to effectively control the number of successfully transmitted packets denoted by $\boldsymbol{\beta}(t)$. Thus, the optimization problem for RRC connection control is formulated as **Problem 2**:

$$\begin{aligned} \underset{N}{\text{maximize}} \quad & \sum_{t\in T} \sum_{i=1}^{N_{UL}} \beta_i(t) \\ \text{subject to} \quad & N_s(t) = \sum_{i=1}^{N_{UL}} \alpha_i(t). \end{aligned}$$

over a set of numbers of CONNECTED IoT devices, $N = \{N_s(1), N_s(2), \cdots, N_s(T)\}$. The objective function in **Problem 2** represents the number of successful UL packets at each period t as the summation of all elements of $\boldsymbol{\beta}(t)$, and it is accumulated over the set of periods, T. The constraint ensures that the number of randomly selected CONNECTED IoT devices at period t equals $N_s(t)$.

IV. DRL-BASED ALGORITHM DESIGN

In addressing both **Problem 1** and **Problem 2**, decisions regarding TDD configuration selection and RRC connection control must be made at each period, where choices made in every period influence the subsequent periods. Notably, the search space for **Problem 1** grows in proportion to the base station's buffer size and the quantity of IoT devices, while **Problem 2**'s search space exhibits exponential growth concerning the number of IoT devices. Consequently, the presence of a large number of IoT devices in massive IoT networks introduces challenges in finding solutions within this expansive search space.

To tackle this challenge, we propose the utilization of a DRL algorithm. Specifically, our design involves training the TDD and RRC agents using the DRL algorithm to dynamically determine TDD configurations at the base station and manage RRC states for UL IoT devices, respectively. In this section, we articulate the definition of states, actions, and rewards for both the TDD and RRC agents. Furthermore, we elucidate the application of the proximal policy optimization (PPO) algorithm, the method through which these agents undergo training. This strategic approach aims to enhance the adaptability and decision-making capabilities of the TDD and RRC agents within the context of large-scale IoT networks.

A. TDD AGENT DESIGN

To facilitate TDD configuration selection, the TDD agent engages in determining each configuration at the onset of every period, immediately following the conclusion of the preceding TDD frame. Guided by the latest DL/UL traffic demands, the TDD agent makes decisions regarding the number of DL/UL subframes in the upcoming TDD frame. Consequently, the base station observes the count of successfully transmitted DL/UL packets, enabling the computation of the current TSRs for DL/UL. In the following, we elaborate on the TDD agent's design, specifically focusing on the definition of its state space, action space, and reward function:

- State space: When selecting the TDD configuration at period t, consideration is given to information regarding the current DL/UL traffic demands, influenced by prior TDD configuration selections. The state of the TDD agent is represented as follows:

$$s_t^{\varsigma} = \left\{ s_{t,1}^{\varsigma}, s_{t,2}^{\varsigma}, s_{t,3}^{\varsigma} \right\} \quad \text{where}$$
$$s_{t,1}^{\varsigma} = \text{TDD configuration index at period } t-1,$$
$$s_{t,2}^{\varsigma} = \frac{Q_{DL}(t)}{Q_{DL}(t) + \tilde{Q}_{UL}(t-1)},$$
$$s_{t,3}^{\varsigma} = \frac{\tilde{Q}_{UL}(t-1)}{Q_{DL}(t) + \tilde{Q}_{UL}(t-1)}.$$

  where $s_{t,2}^{\varsigma}$ and $s_{t,3}^{\varsigma}$ represent the demand ratios of the most recent DL/UL transmissions, respectively.
- Action space: The TDD agent, in its role of selecting random or specific ratios of DL and UL subframes to either explore or maximize returns, defines the number of DL and UL subframes within a range of 1 to 9, which sums up to 10. Given this finite set, the selection of DL and UL subframe ratios is expressed as:

$$a_t^{\varsigma} = \text{TDD configuration index at period } t,$$

  where different TDD configuration indices indicate distinct ratios of DL and UL subframes.
- Reward function: Under the state, s_t^{ς}, when the action a_t^{ς} is implemented, the TDD agent receives a reward r_t^{ς}. Given the objective of minimizing the disparity between TSRs of DL/UL, the reward is formulated as:

$$r_t^{\varsigma} = -C_{const} \left| \frac{P_{DL}(t)}{Q_{DL}(t)} - \frac{P_{UL}(t)}{\tilde{Q}_{UL}(t)} \right|.$$

  The inclusion of the negative sign is intended for leveraging a reinforcement learning algorithm that maximizes return. Additionally, C_{const} is multiplied to offset the small value of the ratio difference.

B. RRC AGENT DESIGN

Considering the success probability detailed in section III-D, the number of successfully received packets hinges on both the quantity of transmitting IoT devices and the number of UL subframes in the TDD configuration. Notably, the packet arrival rate of each IoT device remains undisclosed to the base station. Consequently, the RRC agent seeks to optimize the number of successful UL packets by acquiring insights into the correlation between the number of CONNECTED IoT devices and successful UL transmissions. The following outlines the state space, action space, and reward function for the RRC agent:

- State space: Leveraging historical data encompassing the count of CONNECTED IoT devices and the number of successfully transmitted packets, the RRC agent deduces the optimal number of CONNECTED devices to maximize successful UL transmissions within a given TDD configuration. Accordingly, the state of the RRC agent is articulated as:

$$s_t^{\xi} = \left\{ s_{t,1}^{\xi}, s_{t,2}^{\xi}, s_{t,3}^{\xi} \right\} \quad \text{where}$$
$$s_{t,1}^{\xi} = \text{TDD configuration index at period } t,$$
$$s_{t,2}^{\xi} = \frac{\sum_{i=1}^{N_{UL}} \alpha_i (t-1)}{N_{UL}},$$
$$s_{t,3}^{\xi} = \frac{\sum_{i=1}^{N_{UL}} \beta_i (t-1)}{N_{UL}}.$$

  where $s_{t,2}^{\xi}$ and $s_{t,3}^{\xi}$ denote the ratios of CONNECTED IoT devices and successful UL packets, respectively, concerning the total number of UL IoT devices at period t.
- Action space: The RRC agent, in choosing either a random or a specific number of CONNECTED IoT devices to explore or maximize returns, respectively, defines its action as:

$$a_t^{\xi} = N_s(t).$$

  Based on the determined number of CONNECTED IoT devices, the values of each element in the connection vector, $\boldsymbol{\alpha}(t)$, are determined as explained in section III-E.
- Reward function: When the action a_t^{ξ} is implemented under the state, s_t^{ξ},, the RRC agent receives a reward r_t^{ξ}. Aimed at maximizing the number of successful UL packets, the reward is expressed as:

$$r_t^{\xi}(t) = \sum_{i=1}^{N_{UL}} \beta_i(t).$$

C. DRL TRAINING METHODOLOGY

In this study, we employ the Actor-Critic method [67], [68] for training both the TDD and RRC agent, widely applied across diverse domains for dynamic decision-making processes. Among the policy optimization algorithms within the Actor-Critic framework [69], [70], [71], [72] such as DPG, TRPO, A2C, and PPO, we opt for the PPO algorithm to maximize the agents' objectives. PPO stands out due to its ease of implementation, encompassing hyperparameter tuning and data sampling, and its ability to provide stable updates as the policy gradually deviates from the previous one. This stability enables the RRC agent to consistently enhance its return across various TDD configurations, enabling the TDD agent to effectively maximize its return [73]. Consequently, PPO is employed to train both the TDD and RRC agents in optimizing their policies.

The TDD agent comprises an Actor network, $\pi^{\varsigma}\left(a_t^{\varsigma} \mid s_t^{\varsigma}; \theta_a^{\varsigma}\right)$, and a Critic network, $V^{\varsigma}\left(s_t^{\varsigma}; \theta_v^{\varsigma}\right)$, representing the policy network with parameters θ_a^{ς} and an estimated value function with parameters θ_v^{ς}, respectively. Similarly, for the RRC agent, its Actor network, $\pi^{\xi}\left(a_t^{\xi} \mid s_t^{\xi}; \theta_a^{\xi}\right)$, and Critic network, $V^{\xi}\left(s_t^{\xi}; \theta_v^{\xi}\right)$, mirror

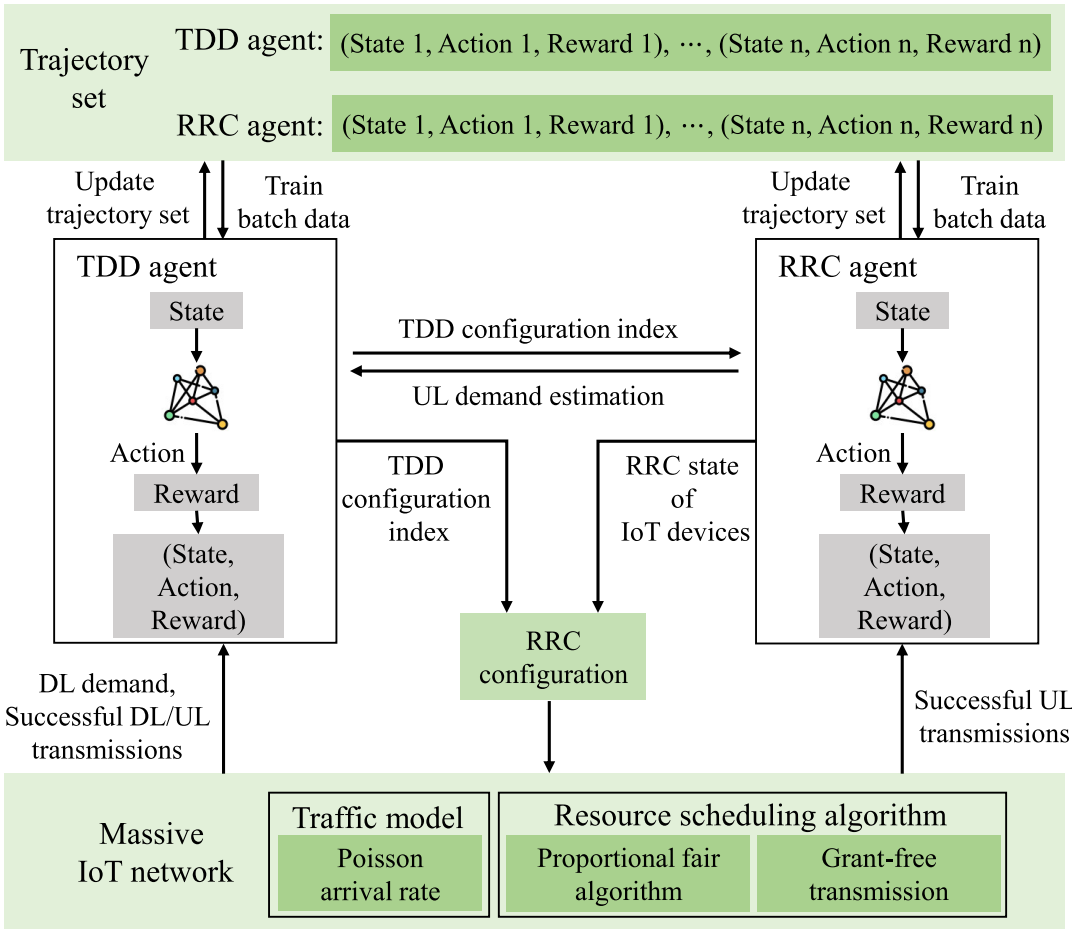


FIGURE 2. The framework of DRL driven joint management of TDD agent and RRC agent.

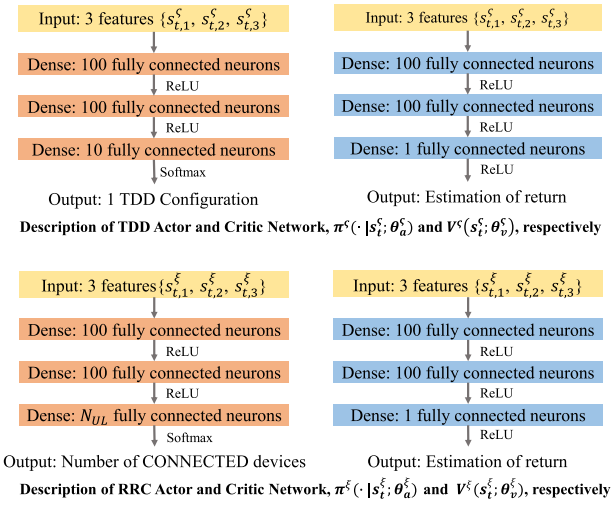


FIGURE 3. Structure of TDD agent and RRC agent.

those of the TDD agent. An overview of the training process for both agents within the DRL framework in a massive IoT network is presented in Fig. 2. The parameter settings of the TDD agent and RRC agent are depicted in Fig. 3 with their detailed training procedure outlined in Algorithm 1.

The training initiates with the initialization of parameters for the Actor and Critic networks for both the TDD and RRC agents. Concurrently, trajectory sets, $\mathcal{D}^{\varsigma}$ and $\mathcal{D}^{\xi}$, are initialized to store training data (lines 1-2 in Algorithm 1). At each period's inception, the base station observes DL/UL traffic demands, $Q_{DL}(t)$ and $\tilde{Q}_{UL}(t-1)$, respectively. Subsequently, the TDD policy network, $\pi^{\varsigma}\left(\cdot|s^{\varsigma}_t;\theta^{\varsigma}_a\right)$, determines the action, a^{ς}_t, based on the state, s^{ς}_t (lines 5-7 in Algorithm 1). The determined TDD configuration index, a^{ς}_t, is then fed into the RRC policy network, $\pi^{\xi}\left(\cdot|s^{\xi}_t;\theta^{\xi}_a\right)$, to obtain the action, a^{ξ}_t. The value of a^{ξ}_t determines the connection vector, $\boldsymbol{\alpha}(t)$ (lines 8-10 in Algorithm 1). At the period's conclusion, the base station observes the number of successfully transmitted DL packets, $P_{DL}(t)$, and UL packets, $P_{UL}(t)$. It estimates the UL traffic demand, $\tilde{Q}_{UL}(t)$, and computes rewards, r^{ς}_t and r^{ξ}_t, storing experiences, $\left(s^{\varsigma}_t, a^{\varsigma}_t, r^{\varsigma}_t\right)$ and $\left(s^{\xi}_t, a^{\xi}_t, r^{\xi}_t\right)$ in trajectory sets, $\mathcal{D}^{\varsigma}$ and $\mathcal{D}^{\xi}$, respectively (lines 11-14 in Algorithm 1).

At the end of each episode, a batch of trajectories, composed of $\left(s^{\xi}_t, a^{\xi}_t, s^{\xi}_{t+1}, r^{\xi}_t\right)$, is randomly sampled from the trajectory set, $\mathcal{D}^{\xi}$, to serve as a training set. These trajectories update the parameters, θ^{ξ}_a and θ^{ξ}_v, of the RRC

Algorithm 1 DRL Training Procedure

1: Randomly initialize TDD agent with $\theta^{\varsigma}_{a,0}$ and $\theta^{\varsigma}_{v,0}$, and RRC agent with $\theta^{\xi}_{a,0}$ and $\theta^{\xi}_{v,0}$
2: Initialize trajectory sets $\mathcal{D}^{\varsigma}$ and $\mathcal{D}^{\xi}$
3: **for** each episode $\in \{1, 2, \cdots, I\}$ **do**
4: **for** each period $t \in \{1, 2, \cdots, J\}$ **do**
5: Observe traffic demands $Q_{DL}(t)$ and $\tilde{Q}_{UL}(t-1)$
6: $s^{\varsigma}_t \leftarrow \left\{s^{\varsigma}_{t,1}, s^{\varsigma}_{t,2}, s^{\varsigma}_{t,3}\right\}$
7: Get action a^{ς}_t based on $\pi^{\varsigma}\left(\cdot|s^{\varsigma}_t; \theta^{\varsigma}_a\right)$
8: $s^{\xi}_t \leftarrow \left\{s^{\xi}_{t,1}, s^{\xi}_{t,2}, s^{\xi}_{t,3}\right\}$
9: Get action a^{ξ}_t based on $\pi^{\xi}\left(\cdot|s^{\xi}_t; \theta^{\xi}_a\right)$
10: Get connection vector $\boldsymbol{\alpha}(t)$
11: Observe transmission success $P_{DL}(t)$ and $P_{UL}(t)$
12: Compute UL traffic demand estimate $\tilde{Q}_{UL}(t)$
13: Compute reward r^{ς}_t and r^{ξ}_t
14: Store $\left(s^{\varsigma}_t, a^{\varsigma}_t, r^{\varsigma}_t\right)$ in D^{ς}, and $\left(s^{\xi}_t, a^{\xi}_t, r^{\xi}_t\right)$ in D^{ξ}
15: **end for**
16: **for** $m \in \left\{1, 2, \cdots, M^{\xi}\right\}$ **do**
17: Sample a batch of trajectories consisting of $\left(s^{\xi}_t, a^{\xi}_t, s^{\xi}_{t+1}, r^{\xi}_t\right)$ from $\mathcal{D}^{\xi}$
18: Update θ^{ξ}_a and θ^{ξ}_v using PPO
19: **end for**
20: Clear trajectory set $\mathcal{D}^{\xi}$
21: **if** episode$\%R == 0$ **then**
22: **for** $m \in \{1, 2, \cdots, M^{\varsigma}\}$ **do**
23: Sample a batch of trajectories consisting of $\left(s^{\varsigma}_t, a^{\varsigma}_t, s^{\varsigma}_{t+1}, r^{\varsigma}_t\right)$ from $\mathcal{D}^{\varsigma}$
24: Update θ^{ς}_a and θ^{ς}_v using PPO
25: **end for**
26: Clear trajectory set $\mathcal{D}^{\varsigma}$
27: **end if**
28: **end for**

agent using PPO. This process repeats M^{ξ} times, after which $\mathcal{D}^{\xi}$ is cleared for the subsequent episode (lines 16-19 in Algorithm 1). Similarly, at every *R-th* episode's conclusion, the parameters of the TDD agent, θ^{ς}_a and θ^{ς}_v, are updated using a batch of $\left(s^{\varsigma}_t, a^{\varsigma}_t, s^{\varsigma}_{t+1}, r^{\varsigma}_t\right)$ sampled from the trajectory set, $\mathcal{D}^{\varsigma}$. After M^{ς} rounds of update, $\mathcal{D}^{\varsigma}$ is cleared for the next round of updates (lines 21-26 in Algorithm 1). Notably, by setting R to a value larger than 1, the parameters of the RRC agent undergo more frequent updates compared to those of the TDD agent. This strategic approach enables the RRC agent to maximize successful UL transmissions across varied TDD configurations and intermittent UL traffic occurrences, leveraging the TDD agent to balance DL/UL TSRs.

V. PERFORMANCE EVALUATIONS

In this section, we conduct a comprehensive evaluation of the proposed algorithm, comparing its performance against three distinct schemes: *Static TDD*, *Dynamic TDD*, and *MinMax*, within a simulated DL/UL traffic generation environment. To assess the efficacy of dynamically allocating DL/UL subframes in response to actual traffic patterns, we contrast our proposed algorithm with *Static TDD*, where the allocation of DL/UL subframes remains fixed and is proportionate to the average DL/UL traffic generation [35].

Furthermore, for an in-depth performance evaluation of RRC connection control in the context of grant-free OFDM transmission within our proposed algorithm, we introduce *Dynamic TDD*. This scheme dynamically manages TDD configurations to ensure a balanced DL/UL TSRs. Additionally, we validate the performance of the proposed scheme in comparison to *MinMax*, which adopts the approaches proposed in [38] and [41]. The method in these studies aims to maximize the overall system throughput in allocating resources to each device and simultaneously select TDD configurations that minimize the DL/UL traffic imbalance.

Correspondingly, in our application of MinMax, we define the RRC states of all UL devices to maximize the number of received packets for each TDD configuration. Subsequently, we select the TDD configuration that minimizes the difference between DL/UL TSRs. It is noteworthy that, while the overall amount of UL traffic demand is presumed to be known at the base station in both *Static TDD* and *Dynamic TDD*, the UL traffic demand of each device is known in *MinMax*, forming a basis for our comparative analyses.

TABLE 2. Simulation parameters and values.

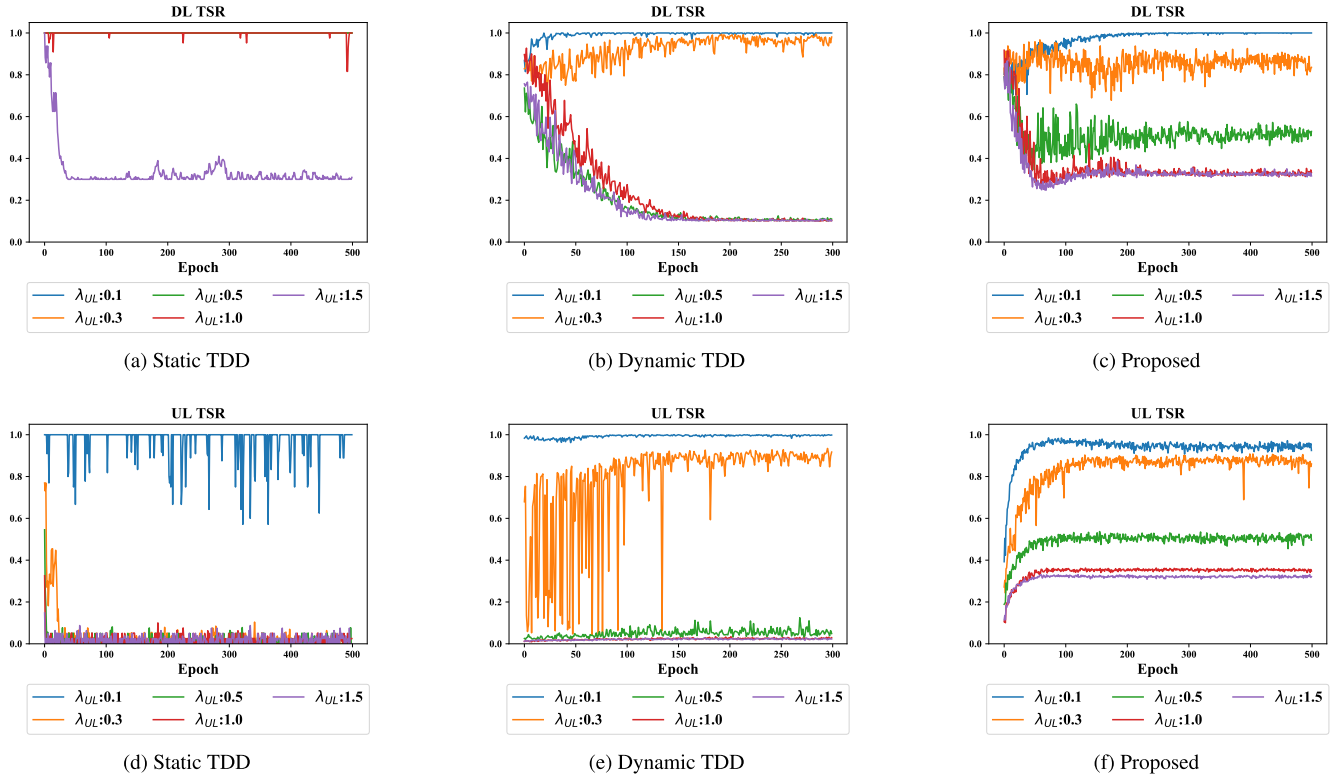
Parameter	Value
Duration of subframe	1 ms
Duration of period	10ms
Number of DL resource blocks	10
Number of UL resource blocks	10
Maximum packet reception capability	1
DL scheduling	Proportional fair algorithm
UL Scheduling	Grant-free transmission
Traffic model	Poisson arrival process
Number of Devices	80
DL packet arrival rate per device	0.5 packet/sec
UL packet arrival rate per device	0.1∼1.5 packet/sec
Transmission power	0.45 W
Receiving power	0.15 W
Sleeping power	0.00012 W

TABLE 3. Metric acronyms.

Acronym	Description
TSR	Traffic success ratio
DL/UL TSR	Downlink/uplink traffic success ratio
TSRD	Traffic success ratio difference
TRU	Time resource utilization
DL/UL TRU	Downlink/uplink time resource utilization

A. SIMULATION SETTINGS

Based on the transmission records of IoT devices obtained from a prominent Finnish mobile network operator [11], the average DL/UL throughputs for each IoT device were



(a) Static TDD (b) Dynamic TDD (c) Proposed

(d) Static TDD (e) Dynamic TDD (f) Proposed

FIGURE 4. Training progress with respect to DL ratio (top) and UL ratio (bottom).

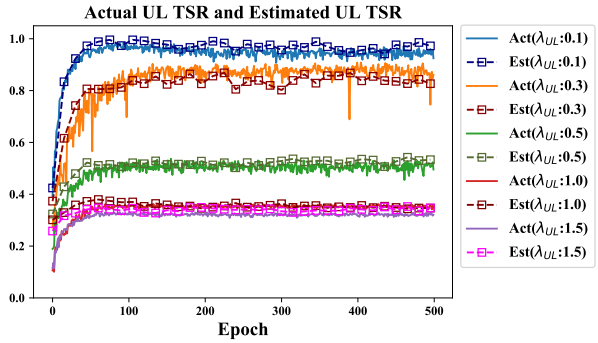


FIGURE 5. Actual and estimated UL ratio of the proposed algorithm.

observed to vary between 4 to 20 and 12 to 41 *bytes*/*sec*, respectively, over a 2-year period from September 2016 to August 2018. Considering the typical packet size of IoT devices, as reported by Nokia Networks [74], to be less than 32 *bytes*, our experimental setup encompasses a range of UL packet arrival rates from 0.1 to 1.5 packets/sec. Concurrently, we maintain a constant DL packet arrival rate of 0.5 packets/sec for each IoT device. The overall simulation parameters and values used for this study are summarized in Table 2. Additionally, the training parameters I, J, M^{ς}, M^{ξ} and R in Algorithm 1 are consistently set to 500, 100, 50, 150 and 10, respectively, across all variations in UL packet arrival rates, ensuring a standardized foundation for our experimentation.

B. PERFORMANCE INDICATORS

The primary performance assessments in this study are centered around TSRD and TRU, as defined by Ding et al. [35]. TSRD, representing the absolute difference between DL and UL TSRs, is categorized into actual TSRD and estimated TSRD. Actual TSRD, denoted as Actual TSRD (t), is the absolute difference between DL TSR, $\Omega_{DL}(t)$, and UL TSR, $\Omega_{UL}(t)$. Conversely, estimated TSRD, labeled as Estimated TSRD (t), quantifies the difference between DL TSR, $\Omega_{DL}(t)$, and estimated UL TSR, $\tilde{\Omega}_{UL}(t)$. Notably, the estimated TSRD incorporates the use of the estimated UL TSR.

The TRU metric, which signifies the multiplication of subframe allocation ratios with their respective utilization, is defined at each period t. Given the unique UL resource scheduling in this paper, we adjust the computation of TRU to account for the distinctions in DL and UL resource scheduling in comparison to the definition provided by Ding et al. [35]. The acronyms associated with these metrics are consolidated in Table 3, and the following elaborates on each metric:

- TSRD: TSRD (t) expresses the absolute difference between DL and UL TSRs at period t. It comprises two components: Actual TSRD and Estimated TSRD, calculated as follows:

$$\text{Actual TSRD}\,(t) = |\Omega_{DL}(t) - \Omega_{UL}(t)| \tag{11}$$

and

$$\text{Estimated TSRD}\,(t) = \left|\Omega_{DL}\,(t) - \tilde{\Omega}_{UL}\,(t)\right|. \quad (12)$$

- TRU: TRU (t) is the product of the subframe allocation ratio and its utilization at period t. The utilization is computed by the ratio of the number of successfully transmitted packets to the maximum number of successfully transmittable packets, given the number of resource blocks and the number of allocated subframes. Two distinct TRUs are defined: DL TRU and UL TRU, representing TRUs for DL and UL transmissions, respectively:

$$\text{DL TRU}\,(t) = \frac{f_{DL}\,(t)}{T_{period}} \frac{P_{DL}\,(t)}{P_{DL,\max}\,(N_c, f_{DL}\,(t))} \quad (13)$$

and

$$\text{UL TRU}\,(t) = \frac{f_{UL}\,(t)}{T_{period}} \frac{P_{UL}\,(t)}{P_{UL,\max}\,(N_c, f_{UL}\,(t))}. \quad (14)$$

Here, $P_{DL,\max}\,(N_c, f_{DL}\,(t))$ and $P_{UL,\max}\,(N_c, f_{UL}\,(t))$ represent the maximum numbers of successfully transmittable DL and UL packets, respectively, considering the given number of resource blocks and the duration of DL and UL subframes at period t as $N_c, f_{DL}\,(t)$, and $f_{UL}\,(t)$.

C. TRAINING RESULT

In Fig. 4, we present a comparative analysis of the DL/UL TSRs across three schemes during the training process. Notably, while *Static TDD* requires no explicit training, its results are included for comparison with schemes that undergo training. The training processes are conducted under various settings of λ_{UL} ranging from 0.1 to 1.5. For clarity, the results are showcased for values of λ_{UL}, specifically 0.1, 0.3, 0.5, 1, and 1.5.

In Fig.4 (a) and Fig.4 (d), representing *Static TDD*, the DL/UL TSRs exhibit a rough balance only at $\lambda_{UL} = 0.1$, where low packet generation leads to the successful transmission of all generated packets. However, with increasing λ_{UL}, UL TSR approaches zero while DL TSR remains relatively high, even at $\lambda_{UL} = 1.5$, which is three times higher than λ_{DL}. Consequently, *Static TDD* encounters challenges in maintaining DL/UL TSR balance.

For *Dynamic TDD*, as shown in Fig.4 (b) and Fig.4 (e), DL/UL TSRs tend to sustain similar values as λ_{UL} increases compared to *Static TDD*. At low λ_{UL}, DL/UL TSRs approach unity. With rising λ_{UL}, UL TSR experiences a rapid decline due to an increasing number of packet collisions, resulting in a reduced count of successful UL transmissions. Consequently, the number of DL transmissions decreases to align with the diminished UL TSR by allocating a reduced number of DL subframes.

Contrastingly, the proposed algorithm, depicted in Fig.4 (c) and Fig.4 (f), showcases a less drastic decrease in DL/UL TSRs as λ_{UL} increases. Particularly, at λ_{UL} three times higher than λ_{DL}, DL/UL TSRs converge at a substantially higher value compared to *Dynamic TDD*. The RRC connection control, optimizing the RRC states of IoT devices to maximize successful UL transmissions, contributes to achieving higher DL/UL TSRs compared to previous cases.

In addition, Fig. 5 presents additional results illustrating UL TSR based on actual and estimated UL traffic demand, represented by solid and square lines, respectively. The estimated value closely aligns with the actual UL TSR, indicating the effectiveness of the proposed algorithm in approximating real-time UL traffic demand.

D. PERFORMANCE COMPARISON FOR TSRD

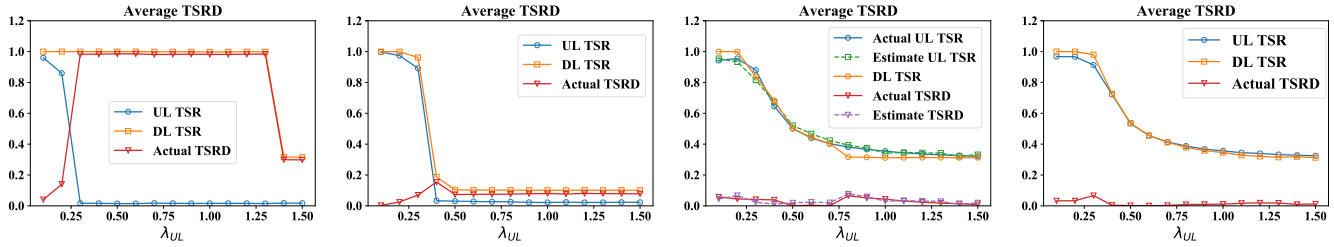
TSRD measures the fairness in managing DL/UL traffic demands, which is defined as the difference between DL/UL TSRs. By minimizing TSRD, the primary objective of this study, the network can properly provide unbiased service toward both DL/UL transmissions. In Fig.6, the performance of each scheme is evaluated in terms of TSRD across a range of λ_{UL} values from 0.1 to 1.5. UL TSR, shown in Fig.6 (a), Fig.6 (b) and Fig.6 (d), is calculated based on the actual amount of UL traffic demand, denoted as the actual UL TSR in Fig.6 (c). Additionally, the estimated UL TSR in Fig. 6 (c) is computed based on the estimated amount of UL traffic demand. Actual TSRD and estimated TSRD are obtained from the actual DL/UL TSRs and estimated UL TSR, following (11) and (12), respectively.

For *Static TDD* in Fig. 6 (a), nearly all DL/UL traffic demands are successfully transmitted when λ_{UL} is 0.1, resulting in DL/UL TSRs close to one and minimal TSRD. However, with increasing λ_{UL}, the number of successful UL packets rapidly declines due to packet collisions. Consequently, TSRD approaches DL TSR, and for λ_{UL} exceeding 1.3, TSRD decreases due to reduced DL TSR with fewer DL subframe allocations while UL TSR remains low.

Dynamic TDD in Fig. 6 (b) adjusts TDD configurations to minimize the difference in DL/UL TSRs for increasing λ_{UL}. Consequently, TSRD remains at a low value over increasing λ_{UL}. However, frequent packet collisions in UL transmission lead to low DL/UL TSRs.

In the proposed algorithm, RRC connection control aims to maximize the number of successful UL transmissions under various TDD configurations. Consequently, while the actual UL TSR and DL TSR gradually decrease for increasing λ_{UL} in Fig. 6 (c), they remain higher than those of *Dynamic TDD*. Moreover, the TSRD of the proposed algorithm maintains low values for increasing λ_{UL} compared to *Dynamic TDD* and *Static TDD*.

In Fig. 6 (d), itMinMax demonstrates the similar performance to the proposed algorithm. This similarity arises from the fact that both schemes aim to maximize the number of successful UL packets by dynamically adjusting the RRC states while selecting the TDD configuration to alleviate the DL/UL traffic imbalance. *MinMax*, benefiting from precise knowledge of DL/UL traffic demand at the base station, achieves higher DL/UL TSRs with sufficiently low TSRD in



(a) Static TDD (b) Dynamic TDD (c) Proposed (d) MinMax

FIGURE 6. Average TSRD for different schemes.

TABLE 4. Average TSRD comparison.

Metric	Scheme	λ_{UL}														
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
DL TSR	Static TDD	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.998	0.999	0.999	0.317	0.315
	Dynamic TDD	1.000	0.999	0.962	0.187	0.104	0.102	0.102	0.102	0.101	0.101	0.100	0.102	0.101	0.102	0.101
	Proposed	1.000	1.000	0.839	0.684	0.501	0.443	0.403	0.316	0.315	0.311	0.312	0.313	0.313	0.312	0.313
	MinMax	1.000	0.999	0.980	0.726	0.536	0.456	0.412	0.379	0.358	0.345	0.328	0.321	0.315	0.319	0.312
UL TSR	Static TDD	0.960	0.860	0.017	0.016	0.014	0.015	0.018	0.016	0.016	0.016	0.016	0.016	0.015	0.017	0.017
	Dynamic TDD	0.997	0.974	0.892	0.033	0.030	0.028	0.027	0.026	0.023	0.022	0.024	0.022	0.022	0.023	0.023
	Proposed	0.942	0.955	0.880	0.645	0.501	0.439	0.403	0.381	0.366	0.354	0.344	0.336	0.330	0.326	0.321
	MinMax	0.967	0.966	0.913	0.723	0.534	0.455	0.414	0.387	0.368	0.356	0.345	0.340	0.333	0.328	0.325
Average TSRD	Static TDD	0.040	0.140	0.983	0.984	0.986	0.985	0.981	0.983	0.983	0.982	0.982	0.983	0.984	0.300	0.298
	Dynamic TDD	0.003	0.025	0.070	0.154	0.074	0.074	0.075	0.076	0.078	0.079	0.076	0.080	0.079	0.079	0.078
	Proposed	0.058	0.045	0.053	0.039	0.000	0.000	0.000	0.065	0.051	0.043	0.032	0.023	0.017	0.014	0.008
	MinMax	0.032	0.033	0.067	0.004	0.002	0.000	0.003	0.008	0.009	0.010	0.017	0.019	0.018	0.010	0.012

most cases compared to the proposed scheme. The proposed algorithm relies on the estimated traffic demand and historical training data for its operation. Nevertheless, even without precise information on the UL traffic demand of each device, the proposed scheme still exhibits desirable performance when compared to *Static TDD* and *Dynamic TDD* in the context of UL traffic demand variations. In conclusion, the proposed scheme achieves low TSRD which indicates its capability of balancing the DL/UL traffic demands for the different sets of UL traffic demands

Table 4 provides specific DL/UL TSRs and TSRD values for each scheme in Fig. 6, calculated based on the actual amount of UL traffic demand. While TSRDs for all schemes are close to zero when traffic generation is low, *Static TDD* performs the worst, struggling to maintain low TSRD for increasing λ_{UL}. In contrast, *Dynamic TDD*, the proposed algorithm, and *MinMax* consistently achieve TSRD values below 0.1 in many cases. Notably, *Dynamic TDD* experiences a gradual increase in TSRD along with λ_{UL}, while the proposed algorithm outperforms it by achieving an 89% reduction in TSRD when λ_{UL} is 1.5. Furthermore, the DL/UL TSRs of the proposed algorithm surpass those of *Dynamic TDD* by 3 and 16 times, respectively. Moreover, even without the UL demand information of each device, the proposed algorithm maintains high UL TSR with less than 10% gap from that of *MinMax* for various sets of UL traffic demand. These results underscore the proposed algorithm's capability to stably achieve low TSRD while accommodating the gradual decline of DL/UL TSRs for increasing λ_{UL}.

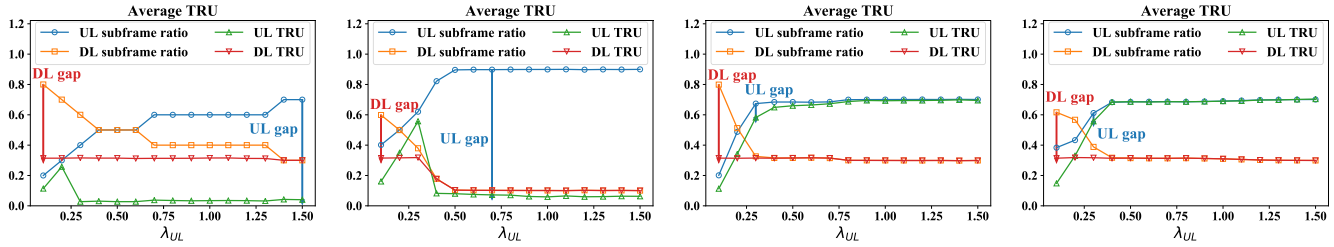
E. PERFORMANCE COMPARISON FOR TRU

In grant-free transmission, the devices suffer from high transmission failure which severely degrades their resource utilization for high UL traffic demand. In this regard, TRU provides crucial information indicating the sufficiency of resources utilized by transmitting devices. Fig.7 presents the average TRU for each scheme across a range of λ_{UL} values from 0.1 to 1.5. The DL/UL subframe ratios in Fig.7 represent the average proportions of their subframes in the TDD frame, ensuring their sum equals one. DL TRU and UL TRU, computed according to (13) and (14), indicate the extent of successfully utilized resources, considering their DL/UL subframe ratios, respectively. The gap between the subframe ratio and TRU signifies the amount of wasted resources, as indicated by red and blue arrows in each figure.

For *Static TDD* in Fig. 7 (a), the UL gap widens with increasing λ_{UL} due to more frequent packet collisions, resulting in low utilization of UL resources. Concurrently, the allocated amount of DL resources decreases for higher λ_{UL}, reducing the DL gap as the relative amount of utilized DL resources increases despite fewer allocated DL resources.

In the case of *Dynamic TDD*, as shown in Fig. 7 (b), the UL gap expands with growing UL traffic demand. Additionally, the gap increases more significantly due to the rapid decline in the number of successful UL transmissions. Conversely, DL resources are sufficiently utilized when λ_{UL} is above 0.3 as the DL subframe is significantly less allocated for transmission.

For the proposed algorithm illustrated in Fig. 7 (c), DL/UL resources are under-utilized at low λ_{UL} as all demands



(a) Static TDD (b) Dynamic TDD (c) Proposed (d) MinMax

FIGURE 7. Average TRU for different schemes.

TABLE 5. Average TRU comparison.

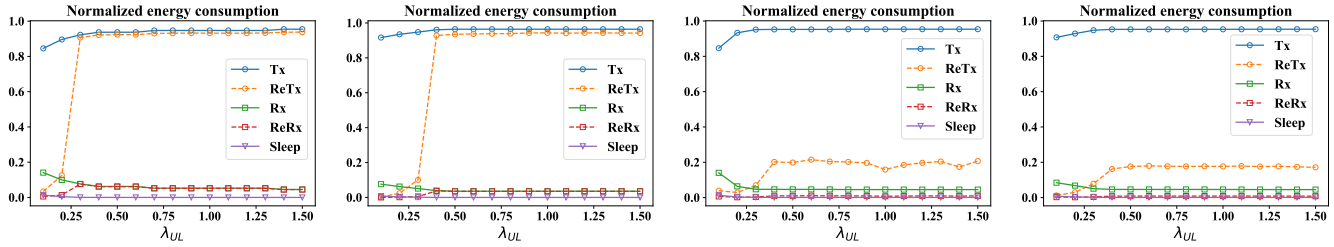
Scheme	Metric	λ_{UL}														
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Static TDD	DL subframe ratio	0.8	0.7	0.5	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.3
	DL TRU	0.315	0.314	0.316	0.314	0.314	0.312	0.313	0.313	0.315	0.315	0.315	0.312	0.312	0.300	0.300
	DL utilization(%)	39.2	44.9	52.6	62.9	62.9	62.4	78.3	78.3	78.4	78.9	78.9	78.1	78.0	100	100
	UL subframe ratio	0.2	0.3	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7
	UL TRU	0.114	0.257	0.027	0.031	0.027	0.027	0.037	0.035	0.033	0.034	0.035	0.034	0.032	0.042	0.040
	UL utilization(%)	56.8	85.8	6.8	6.2	5.5	5.5	6.3	5.8	5.5	5.7	5.8	5.7	5.3	6.0	5.7
Dynamic TDD	DL subframe ratio	0.600	0.499	0.379	0.178	0.104	0.102	0.102	0.101	0.101	0.101	0.100	0.102	0.101	0.101	0.100
	DL TRU	0.315	0.316	0.318	0.177	0.103	0.102	0.101	0.101	0.101	0.101	0.100	0.102	0.101	0.101	0.100
	DL utilization(%)	52.6	63.3	83.8	99.5	99.4	100	99.4	100	100	100	100	100	100	100	100
	UL subframe ratio	0.401	0.501	0.620	0.822	0.896	0.898	0.898	0.898	0.899	0.899	0.900	0.897	0.899	0.899	0.900
	UL TRU	0.160	0.349	0.557	0.082	0.080	0.075	0.072	0.070	0.062	0.059	0.065	0.060	0.061	0.063	0.063
	UL utilization(%)	40.0	69.8	89.8	10.0	8.9	8.9	8.0	7.8	6.8	6.6	7.2	6.6	6.7	7.0	7.0
Proposed	DL subframe ratio	0.799	0.512	0.326	0.316	0.316	0.317	0.315	0.300	0.300	0.299	0.299	0.299	0.299	0.298	0.299
	DL TRU	0.314	0.314	0.315	0.314	0.315	0.316	0.315	0.300	0.299	0.299	0.298	0.299	0.299	0.297	0.299
	DL utilization(%)	39.4	61.2	96.5	99.4	99.8	99.8	100	100	99.9	99.9	99.8	100	100	99.8	100
	UL subframe ratio	0.201	0.488	0.674	0.684	0.684	0.683	0.685	0.700	0.700	0.700	0.700	0.701	0.700	0.702	0.701
	UL TRU	0.113	0.342	0.582	0.649	0.660	0.665	0.673	0.687	0.694	0.693	0.694	0.694	0.696	0.699	0.695
	UL utilization(%)	56.1	70.1	86.3	94.8	96.4	97.3	98.2	98.2	99.2	98.9	99.0	99.0	99.3	99.5	99.1
MinMax	DL subframe ratio	0.616	0.567	0.388	0.317	0.315	0.313	0.313	0.314	0.312	0.310	0.306	0.302	0.300	0.300	0.299
	DL TRU	0.314	0.318	0.317	0.313	0.315	0.313	0.313	0.314	0.312	0.310	0.306	0.302	0.300	0.300	0.299
	DL utilization(%)	50.9	56.1	81.6	98.8	100	100	100	100	100	100	100	100	100	100	100
	UL subframe ratio	0.384	0.433	0.612	0.693	0.685	0.687	0.687	0.686	0.688	0.690	0.694	0.698	0.700	0.700	0.701
	UL TRU	0.148	0.330	0.557	0.683	0.685	0.684	0.686	0.686	0.687	0.690	0.691	0.698	0.700	0.700	0.701
	UL utilization(%)	38.6	76.2	91.1	100	100	99.9	99.9	99.6	99.9	100	99.6	100	100	100	100

can be processed within the allocated resources. However, with increasing UL traffic demand, the proposed algorithm effectively utilizes DL/UL resources, surpassing the utilization levels achieved by the previous schemes. Similarly, in Fig. 7 (d), itMinMax fully utilizes the determined set of DL/UL resources at the base station to maximize the number of successfully transmitted packets in both directions as UL traffic demand grows. In comparison, it processes a higher number of UL traffic demands under a similar amount of UL resources, indicated by higher UL TSR in Table 4. However, such a slight increase can be achieved at the cost of information exchange of UL traffic demand between the base station and devices, which is not available in grant-free transmissions.

Table 5 provides specific TRU values for each scheme, along with DL/UL utilization expressed as a percentage. DL utilization tends to increase with λ_{UL} for all schemes. However, UL utilization for *Static TDD* and *Dynamic TDD* drops below 10% with increasing UL traffic demand due to frequent packet collisions. In contrast, the proposed algorithm maintains steady UL utilization, remaining above 95% for increasing λ_{UL}. Consequently, the UL TRU of the proposed algorithm exceeds that of *Static TDD* and *Dynamic TDD* by 17 and 11 times, respectively, when λ_{UL} is 1.5. These results indicate that the proposed algorithm enables efficient utilization of UL resources, with over 90% of resources wasted by *Static TDD* and *Dynamic TDD* for λ_{UL} above 0.2 and 0.3, respectively. Moreover, the feasibility of the proposed algorithm in terms of resource utilization can be validated as it closely follows that of *MinMax* in high UL traffic demand. To conclude, the proposed scheme benefits from more efficient utilization of determined UL resources by leveraging RRC connection control than the *Static TDD* and *Dynamic TDD*, while maintaining a higher amount of allocated UL resources to flexibly adapt to the increasing UL traffic demand. In addition, compared to *MinMax* where the UL traffic demand is fully known, the proposed scheme achieves sufficient resource utilization in grant-free transmission from the estimated amount of traffic demands.

F. PERFORMANCE COMPARISON FOR ENERGY CONSUMPTION

The energy consumption of IoT devices is a critical concern given their limited energy capacity, particularly in grant-free transmission where a single packet is repeatedly transmitted based on the number of UL subframes, constituting a significant portion of the overall energy consumption.



(a) Static TDD (b) Dynamic TDD (c) Proposed (d) MinMax

FIGURE 8. Normalized energy consumption for different schemes.

TABLE 6. Energy consumption comparison.

Scheme	Metric	lambda														
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Static TDD	Proportion of ReTx/Tx	4.0	14.0	98.3	98.3	98.5	98.5	98.2	98.3	98.4	98.4	98.4	98.4	98.5	98.2	98.3
	Average energy consumption	0.105	0.307	1.859	2.322	2.345	2.354	2.797	2.813	2.817	2.824	2.826	2.831	2.833	3.280	3.282
Dynamic TDD	Proportion of ReTx/Tx	0.3	2.6	10.6	96.6	97.0	98.2	97.3	97.4	97.7	97.8	97.6	97.8	97.8	97.7	97.7
	Average energy consumption	0.185	0.446	0.837	3.603	3.989	4.049	4.082	4.103	4.129	4.140	4.147	4.150	4.159	4.162	4.169
Proposed	Proportion of ReTx/Tx	4.6	3.1	7.3	21.2	20.8	22.6	21.4	21.2	20.6	16.6	19.4	20.6	21.4	18.2	21.7
	Average energy consumption	0.105	0.435	0.731	1.142	1.156	1.190	1.188	1.219	1.222	1.161	1.203	1.222	1.237	1.195	1.239
MinMax	Proportion of ReTx/Tx	1.369	3.213	8.323	16.990	18.478	18.859	18.622	18.544	18.594	18.538	18.759	18.413	18.553	18.246	17.993
	Average energy consumption	0.171	0.412	0.816	1.144	1.169	1.171	1.171	1.169	1.173	1.178	1.186	1.193	1.197	1.198	1.199

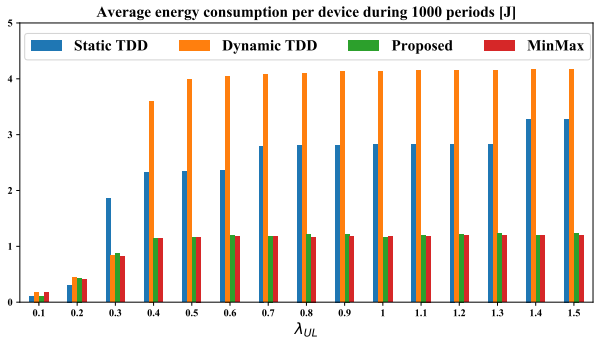


FIGURE 9. Average energy consumption.

Moreover, a large portion of UL transmission can result in transmission failure when the network encounters an increasing number of transmitting devices. As a result, the energy efficiency of transmitting devices can deteriorate rapidly in grant-free transmission. In this experiment, we calculated the energy consumption for UL IoT devices in the simulated environment, considering three operational modes for IoT devices: transmission (Tx), receiving (Rx), and sleep, with energy usage rates of 0.45, 0.15, and 0.00012 W, respectively [75]. The energy computation also included transmission (ReTx) and receiving (ReRx) energy resulting from retransmissions. Fig.8 displays the normalized energy consumption of each mode for each scheme. Meanwhile, Fig.9 presents the average energy consumption of IoT devices during 1000 periods for each scheme.

For *Static TDD*, retransmissions become more frequent as λ_{UL} increases, leading to ReTx dominating the energy consumption by Tx for λ_{UL} over 0.2, indicating the failure of almost all UL transmissions in Fig. 8 (a). Similarly, in the case of *Dynamic TDD* from Fig. 8 (b), frequent packet collisions for increasing λ_{UL} result in ReTx contributing significantly to the energy consumption by Tx. Additionally, due to the high number of allocated UL subframes, the average energy consumption of IoT devices rises rapidly from repetitive transmissions, compared to *Static TDD* in Fig. 9.

In contrast, the proposed scheme from Fig. 8 (c) exhibits fewer retransmissions since the number of transmitting IoT devices is regulated by RRC connection control. Consequently, the energy consumption of ReTx constitutes a smaller portion of the energy consumption by Tx compared to the previous cases. On the other hand, *MinMax* consistently demonstrates the small and steady portion of retransmissions among the compared schemes for varying λ_{UL}. Moreover, according to Fig. 9, its average energy consumption gradually increases and remains at low values compared to *Static TDD* and *Dynamic TDD* as λ_{UL} grows.

Nevertheless, the proposed scheme supports behavior similar to *MinMax* by effectively regulating the number of transmitting devices, with slight variations due to imperfect estimation of UL traffic demand. In conclusion, the proposed algorithm allocates fewer UL subframes, under the regulated UL transmissions through RRC connection control, resulting in a low average energy consumption of IoT devices according to Fig. 9. This allocation strategy reduces the proportion occupied by retransmissions.

Table 6 summarizes the energy consumption of each scheme as depicted in Fig. 8 and Fig. 9. For ease of comparison, the table presents the proportion of ReTx in Tx, as transmission accounts for the majority of energy consumption. For *Static TDD* and *Dynamic TDD*, the proportion of ReTx in Tx exceeds 90% as UL traffic demand increases. However, this value remains below 25% for the proposed algorithm, indicating a reduction of roughly 70% in overall retransmissions. Moreover, by adjusting the number of transmitting devices, the average energy consumption

of the proposed algorithm is reduced by up to 62% and 70%, compared to that of *Static TDD* and *Dynamic TDD*, respectively.

VI. DISCUSSION

Facilitating the sporadic and short packet transmissions of IoT devices is achievable through grant-free transmission. Given the expectation of UL transmission consuming a significant portion of overall communication compared to DL transmission, the deployment of dynamic TDD becomes essential to maintain fairness between DL and UL traffic demand. However, solely adjusting the TDD configuration can lead to a severe degradation in performance, especially with the increasing UL traffic demand and transmission failures in grant-free transmission. To address this challenge, we incorporate RRC connection control, defining the RRC state for each device to regulate the total number of UL transmissions. The joint operation of dynamic TDD and RRC connection control enhances the number of successful transmissions for the determined TDD configuration. This not only maximizes resource utilization but also improves energy efficiency significantly.

Regardless, there are important issues to be addressed regarding the deployment of the proposed method in networks. To balance the DL/UL traffic, their traffic demand needs to be acquired at the base station. Compared to the DL traffic demand, the UL traffic demand is unknown to the base station. In response, we implemented a statistical method to estimate the UL traffic demand based on the modeling of collision dynamics of grant-free transmission. However, the accuracy of the estimation can fluctuate depending on the considerations comprehended in the dynamics. Regardless, with ongoing research on elaborating the dynamics of grant-free transmissions [76], we can mitigate the estimation errors by simultaneously adopting the modeling to improve the performance of the proposed method.

Moreover, the computational constraint is an important aspect to be considered for the application of deep neural networks in wireless communication networks. The proposed method requires the management of two deep neural networks, TDD agent and RRC agent, at each base station to determine the DL/UL configuration and RRC state of each device. While such implementation imposes a high burden on the base station due to its limited amount of resources consumed by both agents, lightweight neural networks are concurrently developed to foster the application on wireless networks to reduce the computational cost with satisfying performance compared to its counterparts with deeper layers [77], [78].

On the other hand, this research unfolds several promising avenues for future exploration. Firstly, while our current investigation centers on TDD management within a single-cell environment, there exists ample potential for expansion into a multi-cell setting, a scenario more reflective of real-world conditions given the increasing densification of base stations to meet 5G requirements. The extension of our work to TDD management in a massive IoT network across multiple base stations is crucial, especially considering the impact of CLI and FLI between these base stations, which significantly influence DL/UL transmissions of IoT devices and contribute to packet collisions between UL devices. Hence, future research could explore the adaptation of our single-cell environment scenario to a multi-cell environment.

Secondly, our proposed approach can be further enhanced to accommodate the diverse requirements of individual IoT devices. Different applications may have varying latency and throughput prerequisites, and IoT devices with distinct packet sizes might demand transmission fairness with the base station. In future investigations, incorporating such diverse requirements alongside DL/UL traffic demand can be explored. This could involve redefining DL/UL traffic demand by integrating device-specific requirements with an overall count of DL/UL packets for effective TDD management in a massive IoT network.

Lastly, energy consumption continues to be a critical concern for IoT devices. While our proposed method succeeds in reducing average energy consumption by constraining the number of transmitting devices, thereby mitigating the need for frequent retransmissions and averting a surge in average energy consumption during high UL traffic demand, energy expenditure in grant-free transmission is also contingent on repetitive transmissions. In future endeavors, the current work can be extended by modeling the trade-off between the number of successful UL transmissions and the frequency of repetitive transmissions, thereby delving into TDD management from the perspective of IoT device energy consumption.

VII. CONCLUSION

In this study, we explored a novel approach to joint dynamic TDD and RRC connection management in the face of DL/UL traffic variations within a single-cell massive IoT network. Our focus involved formulating an optimization problem for the TDD configuration selection module, aiming to minimize the disparity between DL/UL TSRs across diverse traffic demand scenarios. Additionally, we proposed an optimization problem for RRC connection control, seeking to maximize the number of successful UL transmissions by dynamically adjusting the RRC state of each device. Leveraging deep reinforcement learning, we employed TDD and RRC agents to address these optimization problems. The synergy between the agents, facilitated by information sharing, including TDD configuration and estimated UL traffic demand, enables them to take strategic actions to optimize their respective objective functions.

We evaluated the performance of our proposed method, considering metrics such as TSRD, TRU, and energy consumption, comparing it against the benchmarks of *Static TDD* and *Dynamic TDD*. Simulation results revealed that our method significantly reduced TSRD compared to *Dynamic TDD*, achieving an 89% reduction, accompanied by 3 and 16 times higher DL/UL TSRs, respectively. Notably, the UL

TRU for our algorithm demonstrated efficient utilization, exceeding 90%, resulting in 17 and 11 times higher UL TRU compared to *Dynamic TDD* and *Static TDD*, respectively. Furthermore, our method substantially reduced average energy consumption by 70% and 62% for *Dynamic TDD* and *Static TDD*, respectively. Notably, retransmission accounted for less than 25% of consumed energy in our proposed method, contrasting with over 90% for the comparative schemes. Thus, our approach not only surpasses existing benchmarks in managing DL/UL traffic imbalances but also achieves a greater number of successful transmissions with enhanced resource utilization and reduced energy consumption for IoT devices.

Looking ahead, our future work aims to extend this study to dynamic TDD management in a multi-cell massive IoT network environment, where CLI and FLI impact the success probability of DL/UL transmissions in grant-free transmission scenarios. Additionally, considering QoS for each device introduces an additional layer of complexity, as prioritizing DL/UL transmissions based on individual device requirements becomes imperative in the dynamic TDD management paradigm.

REFERENCES

[1] Q. Qi, X. Chen, C. Zhong, and Z. Zhang, "Integration of energy, computation and communication in 6G cellular Internet of Things," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1333–1337, Jun. 2020.

[2] S. Mumtaz, A. Alsohaily, Z. Pang, A. Rayes, K. F. Tsang, and J. Rodriguez, "Massive Internet of Things for industrial applications: Addressing wireless IIoT connectivity challenges and ecosystem fragmentation," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 28–33, Mar. 2017.

[3] A. Mukherjee, P. Goswami, M. A. Khan, L. Manman, L. Yang, and P. Pillai, "Energy-efficient resource allocation strategy in massive IoT for industrial 6G applications," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5194–5201, Apr. 2021.

[4] M. R. Mahmood, M. A. Matin, P. Sarigiannidis, and S. K. Goudos, "A comprehensive review on artificial intelligence/machine learning algorithms for empowering the future IoT toward 6G era," *IEEE Access*, vol. 10, pp. 87535–87562, 2022.

[5] I. Ahmad, A. Ahmed, and S. Abdullah, "Detection and prevention COVID-19 patients using IoT and blockchain technology," in *Proc. Int. Conf. Eng. Softw. Modern Challenges*. Springer, 2021, pp. 91–100.

[6] 3GPP, *Release 14 Description: Summary of Rel-14 Work Items*, Version 14.0.0. Standard TR 38.214, 3rd Gener. Partnership Project, Mar. 2017.

[7] A. Chowdhury and S. A. Raut, "A survey study on Internet of Things resource management," *J. Netw. Comput. Appl.*, vol. 120, pp. 42–60, Oct. 2018.

[8] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures*, Version 14.2.0 Standard TS 36.213, 3rd Gener. Partnership Project, Mar. 2017.

[9] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, and T. H. Jacobsen, "Uplink grant-free access solutions for URLLC services in 5G new radio," in *Proc. 16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2019, pp. 607–612.

[10] 3GPP, *Semi-Persistent Scheduling for 5G New Radio URLLC*, TSG Radio Access Network (RAN) Standard R1-167309, 3rd Gener. Partnership Project, Aug. 2016.

[11] B. Finley and A. Vesselkov, "Cellular IoT traffic characterization and evolution," in *Proc. IEEE 5th World Forum Internet Things (WF-IoT)*, Apr. 2019, pp. 622–627.

[12] H. Kim, J. Kim, and D. Hong, "Dynamic TDD systems for 5G and beyond: A survey of cross-link interference mitigation," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2315–2348, 4th Quart., 2020.

[13] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 11891–11915, Aug. 2021.

[14] S. Narayanan, D. Tsolkas, N. Passas, and L. Merakos, "NB-IoT: A candidate technology for massive IoT in the 5G era," in *Proc. IEEE 23rd Int. Workshop Comput. Aided Model. Design Commun. Links Netw. (CAMAD)*, Sep. 2018, pp. 1–6.

[15] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1960–1973, Dec. 2013.

[16] A. Ahmed, S. Abdullah, S. Iftikhar, I. Ahmad, S. Ajmal, and Q. Hussain, "A novel blockchain based secured and QoS aware IoT vehicular network in edge cloud computing," *IEEE Access*, vol. 10, pp. 77707–77722, 2022.

[17] N. Jiang, Y. Deng, A. Nallanathan, X. Kang, and T. Q. S. Quek, "Analyzing random access collisions in massive IoT networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6853–6870, Oct. 2018.

[18] N. Jiang, Y. Deng, A. Nallanathan, and J. Yuan, "A decoupled learning strategy for massive access optimization in cellular IoT networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 668–685, Mar. 2021.

[19] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 3rd Quart., 2020.

[20] T. Sanislav, G. D. Mois, S. Zeadally, and S. C. Folea, "Energy harvesting techniques for Internet of Things (IoT)," *IEEE Access*, vol. 9, pp. 39530–39549, 2021.

[21] J. Azar, A. Makhoul, M. Barhamgi, and R. Couturier, "An energy efficient IoT data compression approach for edge machine learning," *Future Gener. Comput. Syst.*, vol. 96, pp. 168–175, Jul. 2019.

[22] G. H. Lee, H. Park, J. W. Jang, J. Han, and J. K. Choi, "PPO-based autonomous transmission period control system in IoT edge computing," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21705–21720, Jul. 2023.

[23] J. Choi, J. Ding, N.-P. Le, and Z. Ding, "Grant-free random access in machine-type communication: Approaches and challenges," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 151–158, Feb. 2022.

[24] A. Azari, C. Stefanovic, P. Popovski, and C. Cavdar, "Energy-efficient and reliable IoT access without radio resource reservation," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 2, pp. 908–920, Jun. 2021.

[25] A. Dataesatu, K. Sanada, H. Hatano, K. Mori, and P. Boonsrimuang, "Adaptive K-repetition transmission with site diversity reception for energy-efficient grant-free URLLC in 5G NR," *IEICE Trans. Commun.*, vol. 107, no. 1, pp. 74–84, 2024.

[26] L. Zhao, S. Yang, X. Chi, W. Chen, and S. Ma, "Achieving energy-efficient uplink URLLC with MIMO-aided grant-free access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1407–1420, Feb. 2022.

[27] S. Chen, S. Sun, Y. Wang, G. Xiao, and R. Tamrakar, "A comprehensive survey of TDD-based mobile communication systems from TD-SCDMA 3G to TD-LTE(A) 4G and 5G directions," *China Commun.*, vol. 12, no. 2, pp. 40–60, Feb. 2015.

[28] A. S. Hamza, S. S. Khalifa, H. S. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1642–1670, 4th Quart., 2013.

[29] A. Daeinabi, K. Sandrasegaran, and X. Zhu, "Survey of intercell interference mitigation techniques in LTE downlink networks," in *Proc. Australas. Telecommun. Netw. Appl. Conf. (ATNAC)*, Nov. 2012, pp. 1–6.

[30] R. Kwan and C. Leung, "A survey of scheduling and interference mitigation in LTE," *J. Electr. Comput. Eng.*, vol. 2010, pp. 1–10, Jan. 2010.

[31] E. Pateromichelakis, M. Shariat, A. U. Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3GPP LTE/LTE—A," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 701–717, 2nd Quart., 2013.

[32] C. K. Sheemar, L. Badia, and S. Tomasin, "Game-theoretic mode scheduling for dynamic TDD in 5G systems," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2425–2429, Jul. 2021.

[33] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "Joint optimization of user association and dynamic TDD for ultra-dense networks," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 2681–2689.

[34] Y. Ramamoorthi and A. Kumar, "Dynamic time division duplexing for downlink/uplink decoupled millimeter wave-based cellular networks," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1441–1445, Aug. 2019.

[35] M. Ding, D. Lopez Perez, G. Mao, and Z. Lin, "What is the true value of dynamic TDD: A MAC layer perspective," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–7.

[36] V. D. Tuong, N.-N. Dao, W. Noh, and S. Cho, "Deep reinforcement learning-based hierarchical time division duplexing control for dense wireless and mobile networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7135–7150, Nov. 2021.

[37] F. Tang, Y. Zhou, and N. Kato, "Deep reinforcement learning for dynamic uplink/downlink resource allocation in high mobility 5G HetNet," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2773–2782, Dec. 2020.

[38] A. A. Esswie, K. I. Pedersen, and P. E. Mogensen, "Online radio pattern optimization based on dual reinforcement-learning approach for 5G URLLC networks," *IEEE Access*, vol. 8, pp. 132922–132936, 2020.

[39] S. Wu, H. Guo, J. Xu, S. Zhu, and H. Wang, "In-band full duplex wireless communications and networking for IoT devices: Progress, challenges and opportunities," *Future Gener. Comput. Syst.*, vol. 92, pp. 705–714, Mar. 2019.

[40] H. Zhang, T. Zhou, T. Xu, and H. Hu, "Remote interference discrimination testbed employing AI ensemble algorithms for 6G TDD networks," *Sensors*, vol. 23, no. 4, p. 2264, Feb. 2023.

[41] S. Fukue, H. Iimori, G. T. F. D. Abreu, and K. Ishibashi, "Joint access configuration and beamforming for cell-free massive MIMO systems with dynamic TDD," *IEEE Access*, vol. 10, pp. 40130–40149, 2022.

[42] C.-H. Lee, R. Y. Chang, C.-T. Lin, and S.-M. Cheng, "Beamforming and power allocation in dynamic TDD networks supporting machine-type communication," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[43] C.-H. Lee, R. Y. Chang, S.-M. Cheng, C.-H. Lin, and C.-H. Hsiao, "Joint beamforming and power allocation for M2M/H2H co-existence in green dynamic TDD networks: Low-complexity optimal designs," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4799–4815, Mar. 2022.

[44] A. Mahmood, L. Beltramelli, S. Fakhrul Abedin, S. Zeb, N. I. Mowla, S. A. Hassan, E. Sisinni, and M. Gidlund, "Industrial IoT in 5G-and-beyond networks: Vision, architecture, and design trends," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4122–4137, Jun. 2022.

[45] M. Gharbieh, H. ElSawy, M. Emara, H.-C. Yang, and M.-S. Alouini, "Grant-free opportunistic uplink transmission in wireless-powered IoT: A spatio-temporal model," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 991–1006, Feb. 2021.

[46] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovacs, and T. K. Madsen, "System level analysis of uplink grant-free transmission for URLLC," in *Proc. IEEE Globecom Workshops*, Dec. 2017, pp. 1–6.

[47] G. Berardinelli, N. Huda Mahmood, R. Abreu, T. Jacobsen, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Reliability analysis of uplink grant-free transmission over shared resources," *IEEE Access*, vol. 6, pp. 23602–23611, 2018.

[48] 3GPP, *NR: Physical Layer Procedures for Data*, Version 15.4.0. Standard TS 38.214, 3rd Gener. Partnership Project, Mar. 2020.

[49] T. Kim and B. C. Jung, "Performance analysis of grant-free multiple access for supporting sporadic traffic in massive IoT networks," *IEEE Access*, vol. 7, pp. 166648–166656, 2019.

[50] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[51] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[52] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic compressive sensing-based multi-user detection for uplink grant-free NOMA," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2320–2323, Nov. 2016.

[53] J. Zhang, X. Tao, H. Wu, N. Zhang, and X. Zhang, "Deep reinforcement learning for throughput improvement of the uplink grant-free NOMA system," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6369–6379, Jul. 2020.

[54] K. Chatzikokolakis, A. Kaloxylos, P. Spapis, N. Alonistioti, C. Zhou, J. Eichinger, and Ö. Bulakci, "On the way to massive access in 5G: Challenges and solutions for massive machine communications," in *Proc. Int. Conf. Cognit. Radio Oriented Wireless Netw.*, 2015, pp. 708–717.

[55] B. Han and H. D. Schotten, "Grouping-based random access collision control for massive machine-type communication," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–7.

[56] I. Leyva-Mayorga, M. A. Rodriguez-Hernandez, V. Pla, J. Martinez-Bauset, and L. Tello-Oquendo, "Adaptive access class barring for efficient mMTC," *Comput. Netw.*, vol. 149, pp. 252–264, Feb. 2019.

[57] S. Ryoo, J. Jung, and R. Ahn, "Energy efficiency enhancement with RRC connection control for 5G new RAT," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[58] I. Budhiraja, N. Kumar, and S. Tyagi, "Deep-reinforcement-learning-based proportional fair scheduling control scheme for underlay D2D communication," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3143–3156, Mar. 2021.

[59] 3GPP, *NR: Medium Access Control (MAC) Protocol Specification*, Standard TS 38.321, 3rd Gener. Partnership Project, Oct. 2022.

[60] H. Seo, J.-P. Hong, and W. Choi, "Low latency random access for sporadic MTC devices in Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5108–5118, Jun. 2019.

[61] E. Park, T. Kim, and Y. Han, "An efficient time-shifted random access scheme for cellular-based IoT networks," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 522–525, Mar. 2019.

[62] S. E. Elayoubi, P. Brown, M. Deghel, and A. Galindo-Serrano, "Radio resource allocation and retransmission schemes for URLLC over 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 896–904, Apr. 2019.

[63] C. Ren, X. Lyu, W. Ni, H. Tian, W. Song, and R. P. Liu, "Distributed online optimization of fog computing for Internet of Things under finite device buffers," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5434–5448, Jun. 2020.

[64] J. Yang, Z. Yifan, W. Ying, and Z. Ping, "Average rate updating mechanism in proportional fair scheduler for HDR," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2004, pp. 3464–3466.

[65] T. Kolding, "QoS-aware proportional fair packet scheduling with required activity detection," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.

[66] Y. J. Zhang, P. X. Zheng, and S. C. Liew, "How does multiple-packet reception capability scale the performance of wireless local area networks?" *IEEE Trans. Mobile Comput.*, vol. 8, no. 7, pp. 923–935, Jul. 2009.

[67] V. Konda and J. Tsitsiklis, "Actor-critic algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, 1999, pp. 1–7.

[68] G. Dulac-Arnold, R. Evans, H. van Hasselt, P. Sunehag, T. Lillicrap, J. Hunt, T. Mann, T. Weber, T. Degris, and B. Coppin, "Deep reinforcement learning in large discrete action spaces," 2015, *arXiv:1512.07679*.

[69] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 387–395.

[70] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, Lille, France, 2015, pp. 1889–1897.

[71] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.

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

[73] J. Zhang, S. Guo, Z. Qu, D. Zeng, Y. Zhan, Q. Liu, and R. Akerkar, "Adaptive federated learning on non-IID data with resource constraint," *IEEE Trans. Comput.*, vol. 71, no. 7, pp. 1655–1667, Jul. 2022.

[74] H. Malik, M. M. Alam, Y. L. Moullec, and A. Kuusik, "NarrowBand-IoT performance analysis for healthcare applications," *Proc. Comput. Sci.*, vol. 130, pp. 1077–1083, Jan. 2018.

[75] S. Lee, H. Choi, T. Kim, H.-S. Park, and J. K. Choi, "A novel energy-conscious access point (eAP) system with cross-layer design in Wi-Fi networks for reliable IoT services," *IEEE Access*, vol. 10, pp. 61228–61248, 2022.

[76] H. Zhou, Y. Deng, L. Feltrin, and A. Höglund, "Analyzing novel grant-based and grant-free access schemes for small data transmission," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2805–2819, Apr. 2022.

[77] Z. Cao, W.-T. Shih, J. Guo, C.-K. Wen, and S. Jin, "Lightweight convolutional neural networks for CSI feedback in massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2624–2628, Aug. 2021.

[78] R. Zhao, Z. Li, Z. Xue, T. Ohtsuki, and G. Gui, "A novel approach based on lightweight deep neural network for network intrusion detection," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.

JAEEUN PARK (Student Member, IEEE) received the B.S. degree in electrical engineering from Hanyang University, Seoul, South Korea, in 2019, and the M.S. degree in electronic engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2021, where he is currently pursuing the Ph.D. degree. He is also with the Media Network Laboratory, KAIST, under the supervision of Dr. Junkyun Choi. His current research interests include resource allocation, optimization, and machine learning.

JOOHYUNG LEE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008, 2010, and 2014, respectively. He is currently an Associate Professor with the School of Computing, Gachon University, and also a Visiting Fellow with the Department of Electrical and Computer Engineering, Princeton University. From 2012 to 2013, he was a Visiting Researcher with the Information Engineering Group, Department of Electronic Engineering, City University of Hong Kong, Hong Kong. From 2014 to 2017, he was a Senior Engineer with Samsung Electronics. He has contributed several articles to the International Telecommunication Union Telecommunication (ITU-T) and Third Generation Partnership Project (3GPP). His current research interests include federated learning, 6G networks, cloud/edge computing, and network economics.

DAEJIN KIM (Member, IEEE) received the Ph.D. degree in electrical engineering from KAIST, in 2020. He is currently a Staff Engineer with Samsung Electronics. His research interests include networks, information theory, coding theory, game theory, and machine learning.

JUN KYUN CHOI (Senior Member, IEEE) received the B.Sc. (Eng.) degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1982, and the M.Sc. (Eng.) and Ph.D. degrees in electronics engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1985 and 1988, respectively. From June 1986 to December 1997, he was with the Electronics and Telecommunication Research Institute (ETRI). In January 1998, he joined Information and Communications University (ICU), Daejeon, South Korea, as a Professor. In 2009, he moved to KAIST as a Professor. He is an Executive Member of the Institute of Electronics Engineers of Korea (IEEK), a member of Editorial Board of Korea Information Processing Society (KIPS), and a Life Member of Korea Institute of Communication Science (KICS).