Journal of
Sensor and Actuator Networks

MDPI

Article

A Multi-Agent Reinforcement Learning-Based Grant-Free Random Access Protocol for mMTC Massive MIMO Networks

Felipe Augusto Dutra Bueno [1], Alessandro Goedtel [2], Taufik Abrão [3] and José Carlos Marinello [2,*]

[1] Electrical and Computer Engineering, Faculty of Engineering, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada; felipeaugustodutrabueno@gmail.com
[2] Department of Electrical Engineering, Federal University of Technology-Parana, Av. Alberto Carazzai, 1640, Cornelio Procopio 86300 000, Brazil; agoedtel@utfpr.edu.br
[3] Department of Electrical Engineering, State University of Londrina, Rod. Celso Garcia Cid-PR445, Londrina 86057 970, Brazil; taufik@uel.br
* Correspondence: jcmarinello@utfpr.edu.br

Abstract: The expected huge number of connected devices in Internet of Things (IoT) applications characterizes the massive machine-type communication (mMTC) scenario, one prominent use case of beyond fifth-generation (B5G) systems. To meet mMTC connectivity requirements, *grant-free* (GF) random access (RA) protocols are seen as a promising solution due to the small amount of data that MTC devices usually transmit. In this paper, we propose a GF RA protocol based on a multi-agent reinforcement learning approach, applied to aid IoT devices in selecting the least congested RA pilots. The rewards obtained by the devices in collision cases resemble the congestion level of the chosen pilot. To enable the operation of the proposed method in a realistic B5G network scenario and aiming to reduce signaling overheads and centralized processing, the rewards in our proposed method are computed by the devices taking advantage of a large number of base station antennas. Numerical results demonstrate the superior performance of the proposed method in terms of latency, network throughput, and per-device throughput compared with other protocols.

Keywords: random access protocol; grant-free; beyond 5G; reinforcement learning; massive MIMO

check for updates

Citation: Bueno, F.A.D.; Goedtel, A.; Abrão, T.; Marinello, J.C. A Multi-Agent Reinforcement Learning-Based Grant-Free Random Access Protocol for mMTC Massive MIMO Networks. *J. Sens. Actuator Netw.* **2024**, *13*, 30. https://doi.org/10.3390/jsan13030030

Academic Editors: Elis Kulla and Donald Elmazi

Received: 23 March 2024
Revised: 22 April 2024
Accepted: 25 April 2024
Published: 30 April 2024

1. Introduction

Cellular Internet of Things (IoT) is an important research topic within beyond fifth-generation (B5G) networks [1]. Typical applications involve wireless sensor networks, smart cities, smart grids, smart factories, and connected vehicles [2,3]. The number of IoT devices has been explosively increasing recently, while most devices are low-power nodes whose batteries are expected to be usable for years. Furthermore, such IoT devices are usually distributed over a long range. Therefore, we can point out the main requirements of cellular IoT in B5G networks: massive connectivity, low power consumption, and broad coverage [1]. Exploiting new wireless technologies, such as massive multiple-input multiple-output (MIMO), intelligent reflecting surfaces, and others, is essential to achieve such goals.

Massive MIMO is already a successful technology [4]. Its fundamental idea is to equip base stations (BSs) with many antennas to serve single-antenna users scattered in the cell. It benefits from the fact that the effects of uncorrelated noise and fast fading disappear as the number of BS antennas grows to infinity, remaining only the inter-cellular interference that results from pilot contamination [5]. Massive MIMO systems usually employ a time-division duplex scheme that demands the transmission of only uplink (UL) pilot signals to acquire channel state information, since the downlink (DL) channel can be estimated by channel reciprocity [6]. However, while the number of devices is continually increasing, motivated by the wide spread of IoT applications, the number of resources the cellular network offers remains scarce. This fact gives rise to performance issues such as pilot collisions when two or more users choose the same pilot trying to access BS resources. Therefore, establishing an effective random-access (RA) policy is mandatory.

Several methods have been presented to enhance the traditional random access performance, such as access class barring (ACB), slotted access, and backoff [7]. Among several proposed solutions for the congestion problem resulting from massive access, there are, for instance, (**a**) [7], which investigates an efficient random access procedure based on ACB to decrease the access delay and the power consumption under wireless network congestion resulting from massive access; (**b**) the *strongest-user collision resolution* (SUCRe) protocol from [8]. The SUCRe protocol is a grant-based, four-step RA approach whose main idea consists of allowing just the strongest pilot competitor to access the BS resources each time. Numerical results indicate that the SUCRe protocol can solve about 90% of all collisions. However, the four-step procedure required for the BS to grant exclusive communication resources to devices could result in a performance bottleneck, being a source of excessive delay and signaling overhead. Therefore, it is not the ideal choice for massive machine-type communication (mMTC) systems, where accessing devices usually have small data packets to transmit sporadically. Other works such as [9–12] present proposals for optimization of the SUCRe protocol, showing promising results. However, they are also grant-based protocols that introduce extra complexity or overhead to the SUCRe protocol. B5G RA schemes should achieve high scalability under latency and reliability constraints to support new use cases. For this purpose, grant-free (GF) RA protocols have gained increasing interest, as they can drastically reduce control signaling for connection establishment [1].

Many GF RA protocols available in the literature are derived from the contention resolution diversity slotted ALOHA [13] and irregular repetition slotted ALOHA [14]. The idea behind such schemes is to repeat the transmission of data packets in several randomly chosen slots, including the indices of the chosen slots as side information [15,16]. Whenever a particular device chooses a non-colliding slot, its payload is successfully decoded, and its interference in the other slots is canceled through successive interference cancellation, increasing the occurrence of other non-colliding slots. Although achieving exceptional performances, these protocols have the *drawbacks* of requiring packet re-transmissions, a substantial overhead for side information signaling, increased complexity for successive interference cancellation evaluation, and the possibility of propagation errors in the decoding process.

RA optimization is challenging in ultra-dense mMTC networks while using machine learning tools is quite promising [17]. Among them, *reinforcement learning* (RL) has drawn much attention in the research community and has been demonstrated to assign RA slots to MTC devices effectively. RL is a machine learning technique that enables agents/devices to interact with the environment and learn efficient strategies that maximize long-term system performance. The most typical RL algorithm is the Q-learning (QL) algorithm, which can be implemented in a decentralized way at the user equipment even without an operating model of the environment.

Several recent works deal with the RA problem for IoT applications in low power wide area networks using LoRaWAN as access technology [18–20]. LoRaWAN networks use chirp spread spectrum modulation at the physical layer and pure the ALOHA method at the link layer [20], and can achieve low power and long-range communications. In [18], an adaptive algorithm is proposed to select the spreading factors of the nodes in a LoRaWAN scenario with multiple gateways, improving the throughput and packet delivery ratio of the network. Ref. [19] tackles the same problem by using a decentralized RL algorithm known as the multi-armed bandit, optimizing the spreading factors to minimize interference and maximize the energy efficiency of the end devices in the network. Similarly, Ref. [20] uses a deep RL algorithm optimizing the distribution of network resources such as spreading factor, transmission power, and channel, aiming to minimize the LoRaWAN energy transmission. Besides, several recent works have also applied RL algorithms for optimizing RA in non-orthogonal multiple access (NOMA) systems [21–23], in industrial edge-cloud networks [24], and in vehicle-to-everything (V2X) networks [25].

A different context is the use of cellular technology for IoT connectivity. As discussed in [26], using cellular technology for IoT access instead of wide area networks such as LoRaWAN presents significant benefits in terms of coverage. Since cellular IoT uses pre-existing mobile networks, an extensive coverage area is already in place. This allows one to manage device deployments in different locations. Network switching is also an advantage when using cellular IoT since the device will automatically connect to the network with the strongest signal in the area, ensuring constant and reliable connectivity. Given the above scenario, cellular IoT networks are widely disseminated worldwide and are expected to reach a number of 5.4 billion IoT connections by the end of 2028 [27].

In [28], an RL-based GF RA for pilot collision control is proposed in a cellular IoT scenario. The RL algorithm used in [28] is the Q-Learning, where each accessing device is an independent agent, and the rewards are simply $+1$ when it chooses a unique RA slot or -1 otherwise. A similar procedure is proposed in [29], in which the BS also sends $+1$ or -1 depending on the outcome of the device's transmission. However, in the case of collision, the actual reward computed by the device is -1 times the ratio of packets already transmitted by it, e.g., if the device has already transmitted 20% of its packets and collided in the current instant, its actual reward is -0.2. As an improvement, a collaborative QL RA scheme is proposed in [30], in which the negative rewards in case of collisions are proportional to the congestion level of the chosen RA slot. However, the protocol assumes that devices know the exact number of devices colliding by their chosen RA slot. The performance results are better than those obtained by the independent QL approach of [28]. Nonetheless, Refs. [28–30] do not assume a realistic system model and do not consider channel effects like multipath fading, path loss, thermal noise, and inter-cell interference (ICI), besides assuming that devices know the exact congestion levels in the case of [30].

In [3], the authors compare the most relevant IoT connectivity technologies, highlighting the main challenges and promising solutions. The existing technologies' main bottlenecks are high signaling overhead, wireless resource scarcity, and inefficient wireless resource usage. On the other hand, massive MIMO and machine learning tools are among the promising solutions to overcome those issues. Therefore, designing RA protocols leveraging massive MIMO technology in conjunction with machine learning tools for overcoming the bottleneck of current IoT technologies is quite relevant [3].

In this work, we propose a QL-based GF RA protocol specially designed for realistic mMTC B5G scenarios, leveraging massive MIMO technology to improve connectivity performance, avoiding pilot collisions. Our network scenarios consider realistic wireless propagation effects, including multipath fading, shadowing, path loss, thermal noise, and ICI. Besides, the devices compute the collaborative QL rewards in collision cases by estimating the congestion levels with *minimal signaling overhead*. Different figures of merit, including latency, network throughput, and per-device throughput reveal the proposed scheme's improved performance compared to others available in the literature while approaching the benchmark performance with perfect congestion level estimates. Besides, numerical results demonstrate the robustness of the proposed scheme regarding different system parameter variations, like the number of antennas or the number of packets each device has to send.

Contributions. The paper's contributions are threefold:

i. We propose an effective GF RA protocol designed for realistic mMTC B5G scenarios, leveraging the massive number of BS antennas to improve performance and employing QL to avoid pilot collisions.

ii. The proposed method employs an improved congestion level estimation at the devices with *minimal signaling overhead*, owing to massive MIMO propagation features.

iii. Extensive numerical results demonstrate the competitive performance achieved by our scheme, approaching the case with perfect congestion level estimates and showing robustness against system parameter changes.

Innovation to prior work. RL methods for channel access have already been investigated in the literature in different scenarios; however, we can highlight the following innovations of our work. The works in [19,20] proposed RL methods to optimize the spreading factors in LoRaWAN technology; differently, our focus is the RA problem of cellular IoT networks, which are known to provide extended coverage in comparison to low power wide area networks [26]. Besides, a remarkable benefit of addressing the B5G cellular IoT scenario is the opportunity to exploit massive MIMO, which has been seen as a promising technology to overcome the bottlenecks of current IoT connectivity technologies [3]. Furthermore, while [28–30] also proposed RL-based RA protocols for cellular IoT, they assumed a simple collision channel as a system model. Contrarily, we assume herein a much more realistic wireless system model for the communication environment, taking into account realistic effects like multipath fading, shadowing, path loss, thermal noise, and ICI while leveraging massive MIMO propagation features to achieve improved performance in such challenging mMTC use mode applications.

Organization. The remainder of this paper is organized as follows. Section 2 presents our adopted system model. Section 3 revisits the main QL-based GF RA protocols available in the literature. The proposed two-step QL GF RA massive MIMO protocol is diligently described in Section 4. Numerical results exploring the main metrics for analyzing the performance of random access networks are carried out in Section 5. The main conclusions and possible research directions are offered in Section 6.

2. System Model

The adopted scenario considers a cellular mMTC network where IoT devices employ a GF RA policy to transmit their packets contending for τ_p orthogonal pilot resources. We consider a time-division duplex scheme, where the wireless channels are assumed constant during an entire time slot. The BS is equipped with a massive number of BS antennas (M) localized at the center of the cell. Let $\mathcal{K}$ be the set of single-antenna devices in the cell, which decide to activate with probability P_a transmitting *payload data* together with a randomly selected *pilot sequence* to enable UL channel estimation at the BS side. We denote the τ_p mutually orthogonal pilot sequences as $\boldsymbol{s}_1, ..., \boldsymbol{s}_{\tau_p} \in \mathbb{C}^{\tau_p \times 1}$, such that each pilot has length τ_p and $\|\boldsymbol{s}_t\|^2 = \tau_p$, $\forall t \in [1, \tau_p]$. Besides, we denote the number of active devices as K_a, while each device has a number of L_k packets to transmit. Therefore, considering $\mathcal{S}_t \subset \mathcal{K}$ as the set of devices that want to transmit data selecting pilot t; its cardinality follows a binomial distribution:

$$|\mathcal{S}_t| \sim \mathcal{B}\left(K, \frac{P_a}{\tau_p}\right), \tag{1}$$

where $K = |\mathcal{K}|$ is the total number of devices in the cell. An illustrative representation of the adopted scenario is presented in Figure 1, while a detailed description of each protocol step is provided in Section 4.

The channel vector between BS and device k is denoted by $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$. The channel follows a complex Gaussian distribution $\mathbf{h}_k \sim \mathcal{CN}(0, \beta_k \mathbf{I}_M)$, where β_k is the large-scale fading coefficient, which follows an urban micro scenario [31]. So, the large-scale fading of the link between device k and the BS is

$$\beta_k = 10^{-\kappa \log(d_k) + \frac{g+\varphi}{10}}. \tag{2}$$

In this equation d_k is the distance between device k and the BS, $\kappa = 3.8$ is the path loss exponent, $\varphi \sim \mathcal{N}(0, \sigma_{\text{sf}}^2)$ is the shadow fading, with standard deviation $\sigma_{\text{sf}} = 10$ dB, and $g = -34.53$ dB is the path loss at the reference distance [31].

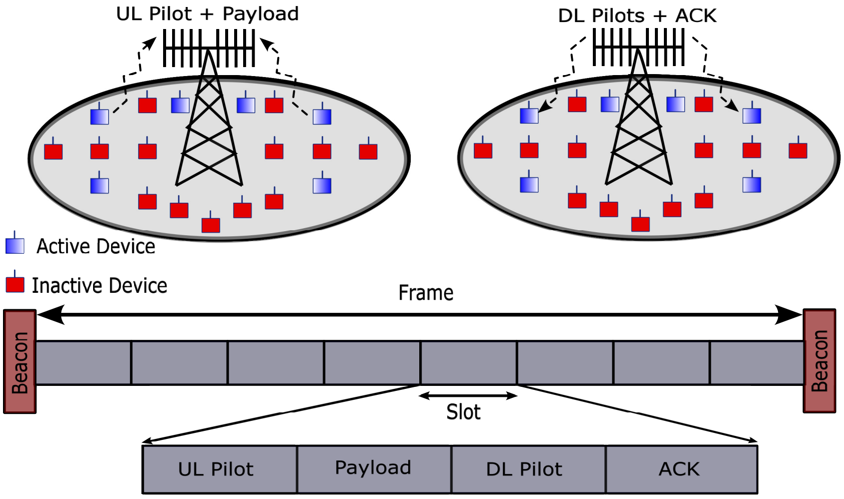

Figure 1. Illustrative representation of the adopted cellular mMTC network with IoT devices employing GF RA protocol for their activation and payload data transmission.

The devices transmit the payload data prepended with an RA pilot to enable channel estimation employing a GF RA protocol. If the transmitted payload data packet is successfully decoded at the BS, the device proceeds to transmit the next data packet. Otherwise, the device repeats the transmission of the same packet in the next frame, which increases *latency* and decreases the *per-device throughput* and the *network throughput*. A pilot collision occurs when two or more devices choose the same RA pilot in the above process and is one of the leading causes of packet losses. Therefore, designing strategies for decreasing the probabilities of pilot collisions is essential in mMTC networks, as described in the following sections.

3. Existing QL-Based GF RA Protocols

Q-learning can be used as a multi-agent RL method to aid IoT devices in selecting the least congested RA pilots in a decentralized way. In typical mMTC networks, active IoT devices cannot coordinate either with the BS or with each other for pilot selection; therefore, IoT devices act as individual learning agents, using their previous experience to enhance the probability of selecting exclusive RA pilots, minimizing the occurrence of pilot collisions. In this section, we revisit two previous works where this framework has been applied in a *simple collision channel*, i.e., assuming that the only communication impairment is *pilot collision*, while neglecting multipath fading, shadowing, path loss, thermal noise, and ICI. If the chosen RA pilot is exclusively selected by that device (no interference), it is assumed that the payload data packet transmitted by it is successfully decoded at the BS, and the device proceeds to transmit the next data packet. Otherwise, if a collision occurs, the device repeats the transmission of the same packet in the next frame. Thus, *reinforcement learning* techniques can be employed to guide the pilots' choice of devices towards the least congested ones, improving the connectivity performance of the network in the considered scenario.

The interaction between IoT devices and the environment can be modeled as a Markov Decision Process, where at each time step, a device can change its current state $x_t \in X$ to $x_{t+1} \in X$ by taking action $a_t \in A$ based on a transition probability function $f(x_t, a_t, x_{t+1})$ [30]. Depending on its state-action pair, the device is rewarded with $r_{t+1} \in R$ during the transition state. Besides, the expected return of a state-action pair is given by $Q^{\pi}(x,a) = \mathbb{E}\left[\sum_{j=0}^{J} \gamma^j r_{t+j+1} | x_t = x, a_t = a, \pi\right]$, where π is the established policy, $\gamma \in [0,1]$ is the discount factor and J is the length of one episode [30].

Satisfying the Bellman optimality equation, the Q-function can then be written as $Q^*(x,a) = \max_{\pi} Q^{\pi}(x,a)$. If a *greedy policy* $\pi(x) = \arg\max_a Q^{\pi}(x,a)$ is established for the

Q-function, then we have a QL algorithm that selects only the actions associated with the highest Q-value $Q(x, a)$ at each state, calculated iteratively as:

$$Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t) + \delta_t \Big[r_{t+1} + \gamma \max_a Q_t(x_{t+1}, a) - Q_t(x_t, a_t) \Big], \quad (3)$$

where δ_t is the learning rate at the tth time step. This model can then be applied for *decentralized pilot selection* as described in [28,30], in which each device has a Q-table of τ_p elements evaluating its experience in selecting the different pilots to transmit data. This Q-table can be updated according to:

$$Q_{t+1}(k, \ell) = Q_t(k, \ell) + \delta[R(k, \ell) - Q_t(k, \ell)], \quad (4)$$

where $Q_t(k, \ell)$ is related to the experience of the kth device in choosing the pilot $\ell = c(k)$, and $R(k, \ell)$ indicates the reward function for this choice. It is worth emphasizing the differences between Equations (3) and (4). The former represents the general framework for establishing a policy for selecting the changing-state actions of a Markov Decision Process under a greedy perspective, i.e., seeking to maximize the associated rewards. Applying this model to the specific problem of decentralized pilot selection in mMTC, the state represents the pilot the device has chosen in the current access attempt, and the action represents the pilot it chooses for the next access attempt. As the action determines univocally the state, the Q-table of a single device can be represented in a single dimension. However, the problem is multi-agent since we have K devices, and thus we have K unidimensional Q-tables, which can be better organized in a single Q-table with K lines. It is worth noting that updating each line occurs without considering the values of the other lines since there is no coordination between the devices. Finally, no discount factor is implemented in the mMTC random access context as proposed in [28,30], and thus we arrive at the Equation (4).

A simple way to compute the rewards $R(k, \ell)$ in (4) is proposed in [28] through an *independent* QL approach for mMTC (**iQmMTC**), where $R(k, \ell) = +1$ if the transmission succeeds, or $R(k, \ell) = -1$, otherwise. However, [30] has shown that better results can be achieved through a *collaborative* approach (**cQmMTC**), where the device is rewarded either with $R(k, \ell) = +1$ for a successful transmission or with $R(k, \ell) = -P_c(k)$ if the transmission fails, with the collaborative penalty function $P_c(k)$ being computed as

$$P_c(k) = \frac{1}{K_a} CL(k), \quad (5)$$

where K_a is the number of active devices in the cell and the congestion level $CL(k) = |\mathcal{S}_{c(k)}|$ is the *number of contenders* for pilot $c(k)$ [30], while $\mathcal{S}_{c(k)}$ is the set of devices choosing the same pilot $c(k)$.

In this framework, each device keeps its own $1 \times \tau_p$ Q-table. Initially, this Q-table is filled with zeros, and all τ_p RA pilots are equally likely to be chosen by it. Then, at each subsequent transmission attempt, the devices are rewarded with $R(k, \ell)$, and their Q-tables are updated according to (4). Once an individual device has updated its Q-table, it will only choose pilots among the ones with the highest Q-value. The process is repeated until the L packets are transmitted.

In the cQmMTC approach of [30], the congestion levels in case of collisions are computed as the number of contending devices, $|\mathcal{S}_{c(k)}|$. However, nothing is discussed about the feasibility of making this information available on the device's side. At first glance, it would require conceiving an estimator to be employed at the BS and then feedback the result to devices, which would spend significant signaling overhead. For comparison purposes, herein, we assume perfectly known the congestion level at the device's side for this scheme, referring to it as *genie* cQmMTC. Furthermore, in a practical network not all IoT devices are active at a given instant since they activate independently with certain probabilities in such a way that the actual number of active devices is not known by the BS. To avoid complex computations and excessive signaling overheads while simplifying the procedure, we propose a two-step QL-based GF RA protocol making use of a large number of BS antennas to allow a collaborative penalty function computation at the devices' side in case of collisions with minimal complexity and overhead, as described in the following.

4. Proposed QL GF RA Massive MIMO Protocol

Given the importance of designing pilot collision avoidance strategies for the investigated scenario and the unrealistic assumptions made by the available strategies described in Section 3, we propose in this section a two-step QL GF RA protocol operating in a realistic B5G system model, as illustrated in Figure 1. The evaluation in the realistic scenario allows us to leverage the massive number of antennas available at the BS to improve the operation of the proposed protocol.

When the device k transmits data, it randomly selects one of the τ_p pilot sequences and transmits it followed by its UL payload data packet $\mathbf{d}_k \in \mathbb{C}^{\tau_d \times 1}$, with a non-zero transmit power $\rho_k > 0$, where τ_d is the data length. We can denote the chosen pilot as $c(k) \in \{1, 2, \ldots, \tau_p\}$, and define the UL signal as $\mathbf{x}_k = [\mathbf{s}_{c(k)}^T, \mathbf{d}_k^T]^T \in \mathbb{C}^{(\tau_p+\tau_d)\times 1}$.

Thus, the BS receives the signal

$$\mathbf{Y} = [\mathbf{Y}_p, \mathbf{Y}_d] = \sum_{k\in\mathcal{K}} \sqrt{\rho_k}\mathbf{h}_k\mathbf{x}_k^T + \mathbf{N}, \tag{6}$$

where $\mathbf{Y} \in \mathbb{C}^{M\times(\tau_p+\tau_d)}$, $\mathbf{Y}_p \in \mathbb{C}^{M\times\tau_p}$, $\mathbf{Y}_d \in \mathbb{C}^{M\times\tau_d}$, and $\mathbf{N} \in \mathbb{C}^{M\times(\tau_p+\tau_d)}$ is the receiver noise with entries drawn from $\mathcal{CN}(0, \sigma^2)$. Besides, we have

$$\mathbf{Y}_p = \sum_{k\in\mathcal{K}} \sqrt{\rho_k}\mathbf{h}_k\mathbf{s}_{c(k)}^T + \mathbf{N}_p, \text{ and} \tag{7}$$

$$\mathbf{Y}_d = \sum_{k\in\mathcal{K}} \sqrt{\rho_k}\mathbf{h}_k\mathbf{d}_k^T + \mathbf{N}_d, \tag{8}$$

with $\mathbf{N} = [\mathbf{N}_p, \mathbf{N}_d]$. Hence, the BS correlates (7) with each pilot to generate channel estimates. For the case of an arbitrary pilot $\mathbf{s}_t$, with $t \in [1, \tau_p]$, it yields:

$$\mathbf{y}_t = \mathbf{Y}_p \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|} = \sum_{i\in\mathcal{S}_t} \sqrt{\rho_i\tau_p}\mathbf{h}_i + \mathbf{n}_t, \tag{9}$$

where $\mathbf{n}_t = \mathbf{N}_p \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|}$ is the effective receiver noise, so that $\mathbf{n}_t \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_M)$. As a result, the BS tries to decode the payloads in (8) using the channel estimates $\mathbf{y}_t$, evaluating:

$$\widehat{\mathbf{d}}_k^T = \frac{\mathbf{y}_t^H}{\sqrt{\tau_p}}\mathbf{Y}_d. \tag{10}$$

The signal-to-interference-plus-noise ratio (SINR) of $\widehat{\mathbf{d}}_k$ in (10) can be obtained following the SINR analysis of [32], adapting the results to our scenario, as follows:

$$\gamma_k^{\text{ul}} = \frac{M\rho_k^2\beta_k^2}{M\sum_{\substack{i\in\mathcal{S}_t\\ i\neq k}}\rho_i^2\beta_i^2 + \left[\sum_{i\in\mathcal{S}_t}\rho_i\beta_i + \frac{\sigma^2}{\tau_p}\right]\left[\sum_{j\in\mathcal{K}}\rho_j\beta_j + \sigma^2\right]}. \tag{11}$$

We assume that the decoding of $\widehat{\mathbf{d}}_k$ in (10) is always successful when k is the unique competitor for the pilot t (without pilot collisions). The BS responds with an ACK feedback message if the decoding of (10) is successful, together with the transmission of a precoded DL pilot signal $\mathbf{V} \in \mathbb{C}^{M\times\tau_p}$, with power q, according to:

$$\mathbf{V} = \sqrt{\frac{q}{\tau_p}}\sum_{t=1}^{\tau_p}\frac{\mathbf{y}_t^*}{||\mathbf{y}_t||}\mathbf{s}_t^T. \tag{12}$$

In (12), one can note that the BS uses the estimated channel of each pilot, $\mathbf{y}_t$, as a precoding vector to transmit such pilot, $\boldsymbol{s}_t$, in the DL. By normalizing each precoding vector and dividing by the total number of transmitted pilots, the BS ensures the average transmitted power is equal to q. The devices receive $\mathbf{z}_k \in \mathbb{C}^{\tau_p \times 1}$, $k \in \mathcal{S}_t$

$$\mathbf{z}_k^T = \mathbf{h}_k^T \mathbf{V} + \boldsymbol{\eta}_k^T, \tag{13}$$

where $\boldsymbol{\eta}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{\tau_p})$ is the noise. After correlating $\mathbf{z}_k$ with $\boldsymbol{s}_t$, the device calculates

$$z_k = \mathbf{z}_k^T \frac{\boldsymbol{s}_t^*}{||\boldsymbol{s}_t||} = \sqrt{q}\, \mathbf{h}_k^T \frac{\mathbf{y}_t^*}{||\mathbf{y}_t||} + \eta_k, \tag{14}$$

where $\eta_k \sim \mathcal{CN}(0, \sigma^2)$. It is important to highlight the different roles of UL and DL pilots. While the UL pilot transmissions allow the BS to acquire channel estimates in (9), which are then used to try to decode the payloads in (10), the DL pilot transmissions allow the UEs to compute z_k in (14), which are then used to evaluate how congested were their chosen pilots in the first step, updating their individual Q-table as described in the sequel.

Collaborative penalty function computation. Let $\alpha_t = \sum_{i \in \mathcal{S}_t} \rho_i \beta_i \tau_p$ be the *sum of average channel gains of the devices* in $\mathcal{S}_t$ seen at the BS according to (9), then an asymptotically *error-free estimator* for α_t is proposed in a similar scenario (in [8], the $\hat{\alpha}_{t,k}$ estimate is computed as part of the four-steps grant-based handshake procedure of the SUCRe protocol and used to let the devices decide whether they should retransmit the chosen pilot or not, depending on whether it is the strongest contender. Differently, herein we employ a GF RA protocol, in which the devices transmit the payload data together with an RA pilot to enable channel estimation and data decoding in a reduced number of steps) in [8] as follows:

$$\hat{\alpha}_{t,k} = \max\left(\left[\frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)} \right]^2 \frac{q \rho_k \beta_k^2 \tau_p}{[\Re(z_k)]^2} - \sigma^2,\ \rho_k \beta_k \tau_p \right), \tag{15}$$

$\Re(\cdot)$ is the real part and $\Gamma(\cdot)$ is the complete Gamma function.

Given the estimate $\hat{\alpha}_{t,k}$ in (15), reminding that $\alpha_{c(k)} = \sum_{i \in \mathcal{S}_{c(k)}} \rho_i \beta_i \tau_p$, and since the device k knows its average channel gain β_k, it can compute a measure of how congested is its chosen pilot as follows

$$\widehat{\phi}_k = \frac{\hat{\alpha}_{t,k}}{\rho_k \beta_k \tau_p}, \qquad 1 \leq \widehat{\phi}_k < \infty. \tag{16}$$

One can note that as long as $\widehat{\phi}_k$ approaches 1, it indicates it is likely that no other device has chosen the pilot $c(k)$. On the other hand, as $\widehat{\phi}_k$ increases, it indicates it is likely that many other devices chose the same pilot $c(k)$. Therefore, $\widehat{\phi}_k$ can be seen as a *rough estimate* of $CL(k)$; hence, the penalty function $P_c(k)$ in (5) can be computed approximately as

$$P_c(k) \approx \frac{\widehat{\phi}_k}{\widehat{K}_a}, \tag{17}$$

in which $\widehat{K}_a$ is an estimate of the number of active devices in the cell. We propose to employ a simple estimator, which computes $\widehat{K}_a$ as the expected number of active devices, supposing no pilot collision occurs. In this way, one can compute:

$$\widehat{K}_a = K \cdot P_a \cdot \mathbb{E}[L_k]. \tag{18}$$

The penalty function $P_c(k)$ in (17) can be used in a realistic massive MIMO scenario, allowing a practical implementation of the cQmMTC approach in a GF RA protocol.

Operationalizing RL for GF RA pilot collision mitigation. We can summarize how RL is made specific to operate in the system under consideration in the proposed protocol as follows:

(***i***) the devices choose their RA pilots along the transmissions of the L_k packets according to their own Q-table;

(***ii***) based on the outcome of each transmission, its Q-table is updated following (4), while computing the rewards as:

$$R(k,\ell) = \begin{cases} \quad +1, & \text{if the transmission succeeds} \\ -P_c(k) = -\frac{\widehat{\phi}_k}{\widehat{K}_a}, & \text{in case of pilot collision,} \end{cases} \tag{19}$$

with $\widehat{\phi}_k$ and $\widehat{K}_a$ being computed as (16) and (18), respectively.

5. Numerical Results

In this section, we evaluate the performance of the proposed QL-based GF RA protocol in terms of: (***i***) average latency, considered herein as the total number of attempts, A_k, the device makes to transmit its L_k packets, (***ii***) average network throughput, defined as the ratio between the number of successfully transmitted packets (without collisions) at certain time step and the number of available pilots τ_p, and (***iii***) the per-device throughput, considered as the ratio of the total number of successfully sent packets by each device, L_k, to the total number of attempts, A_k, that the device has to make to send them, in such a way that $L_k \leq A_k$. For the simulations, we consider a massive MIMO BS equipped with M antennas at the center of a hexagonal cell with a radius of 250 m, surrounded by six neighboring hexagonal cells with the same radius. Each neighboring cell has a fixed number of $K_{\text{ici}} = 400$ active interfering devices. The simulation parameters are set as $\rho = 27$ dBm, $q = 40$ dBm, $\tau_p = 400$, and $\delta = 0.1$. It is worth noting that the choice of the parameters is based on [29,30] in the case of τ_p and δ, while being based on [8,31] in the case of ρ and q. Furthermore, concerning the number of active interfering devices in the neighboring cells K_{ici}, it is made equal to τ_p similarly as in [8]. Table 1 summarizes the numerical parameters adopted in our simulations. With respect to L and P_a, we investigate three different scenarios in this section: (***i***) $L_k = L, \forall k \in \mathcal{K}$, and $P_a = 1$, such that $K_a = K$; (***ii***) random L_k, and $P_a = 1$, such that $K_a = K$; (***iii***) random L_k, $\forall k \in \mathcal{K}$, and $P_a = 0.1\%$, such that $K_a \leq K$ is also random. One can see that while scenario (***i***) results equal to that evaluated in [29,30], scenarios (***ii***) and (***iii***) become gradually more realistic and challenging with a random number of packets and active devices.

Table 1. Numerical Parameters for Simulation Settings.

Parameter	Value	Description
M	100	Number of BS antennas in the center and neighboring cells
τ_p	400	Number of available RA pilot sequences
q	40 dBm	Transmit power of the BS
ρ	27 dBm	Transmit power of the UEs
σ^2	−98.65 dBm	Noise variance
f_c	3 GHz	Carrier frequency
δ	0.1	Learning rate
K_{ici}	400	Number of active devices in each neighboring cell
R	250 m	Radius of the cells
σ_{sf}	10 dB	Shadow fading standard deviation
κ	3.8	Path loss exponent
g	−34.53 dB	Path loss at the reference distance
	10,000	Number of Monte-Carlo realizations
	27 dBm	Transmit power of UEs in adjacent cells
	6	Number of neighboring cells

We investigate four protocols: the (**a**) *baseline* scheme, which is equivalent to the slotted ALOHA protocol, with the devices choosing the pilots uniformly at random; (**b**) the *iQmMTC* approach of [28]; (**c**) the *cQmMTC* approach of [30], assuming that the actual values of $|\mathcal{S}_t|$ and K_a are perfectly known at the devices' side, like if a genie could inform this to them; and (**d**) our proposed two-step QL GF RA protocol leveraging the *massive* MIMO propagation features to efficiently compute the negative rewards of the QL framework at the devices' side, which we denote as *mQmMTC*. Besides, for this last one, we investigate its performance in the scenarios with and without ICI (it is worth noting that for the baseline, iQmMTC, and cQmMTC in the adopted scenario, only pilot collisions degrade their connectivity performance; therefore, ICI does not matter to them).

5.1. Fixed L_k and $P_a = 1$

In this subsection, we take $L_k = L$, $\forall k \in \mathcal{K}$, and $P_a = 1$, such that $K_a = K$. Although these simplifying assumptions usually do not hold in practice, they are useful in unveiling the full potential of the investigated methods and in evaluating the performance losses when not assuming them, which is carried out in the following subsections. The performance results presented in Figures 2 and 3 have been generated with 10,000 Monte-Carlo realizations. Each realization is a frame or a time step in the QL framework, in which each device can transmit only one pilot and the payload packet. The number of antennas is kept fixed at $M = 100$, and the number of devices varies from 25 to 800 in steps of 25.

Figure 2 shows average latency *versus* K results. The proximity of the proposed mQmMTC results for both scenarios with and without interference with the ideal cQmMTC protocol is noteworthy. Also, both the mQmMTC and the ideal cQmMTC results are below the baseline for any value of K. Also, they are below the independent QmMTC for $K > 400$, which corroborates the cQmMTC superiority presented in [30]. Similarly, Figure 3 reveals the average network throughput *vs* K results where the mQmMTC is very close to the ideal cQmMTC performance for both the scenarios with and without interference while being consistently superior to the baseline results for any number of devices K. Furthermore, compared with iQmMTC, the obtained performances are very similar in the region of $K \leq \tau_p = 400$ devices, while the performance obtained by the proposed mQmMTC approach becomes remarkably superior for a higher number of devices. For example, with $K = 600$ active devices, while baseline and iQmMTC achieve network throughputs of $\approx$0.34 and 0.36, respectively, our mQmMTC protocol achieves a network throughput of $\approx$0.49, an improvement of $\approx$44% and $\approx$36%, respectively.

The performance results presented in Figures 4 and 5 are generated with 80,000 Monte-Carlo realizations. The number of devices is fixed at $K = 400$, and the number of available pilots is also fixed at $\tau_p = 400$. The number of BS antennas varies in the range $M \in [1, 100]$, both in the home cell as well as in the neighboring cells. Figure 4 depicts the average latency with an increasing number of BS antennas M, while Figure 5 reveals the behavior of the average network throughput vs. M. Both figures of merit for the proposed mQmMTC approach improve with the increasing number of antennas M since the reward computations in (17) benefit from the large number of BS antennas (favorable propagation effect). Besides, in both cases the presence of ICI slightly deteriorates the congestion level estimates in (16) and the reward computations, leading to a small performance degradation. Despite that, the scenarios with and without ICI perform similarly, achieving improved connectivity performances by benefiting from the many BS antennas. The results of the ideal cQmMTC are also included in the figures as a lower bound (avg. latency) and upper bound (avg. throughput), respectively. The average percentual degradations of the results of mQmMTC regarding the ideal cQmMTC are also shown in both figures for $M = 30$ and $M = 100$. One can see that the improvement caused by increasing M from $M = 10$ to $M = 100$ is not as significant as when increasing M from $M = 1$ to $M = 10$. Therefore, we can conclude that our proposed mQmMTC RA protocol can achieve improved connectivity performance even with a small number of BS antennas. Indeed, $M \approx 30$ antennas at the BS are revealed to be sufficient to attain reliable congestion level estimation $\widehat{\varphi}_k$ when a maximum acceptable degradation level of 3.5% is considered.

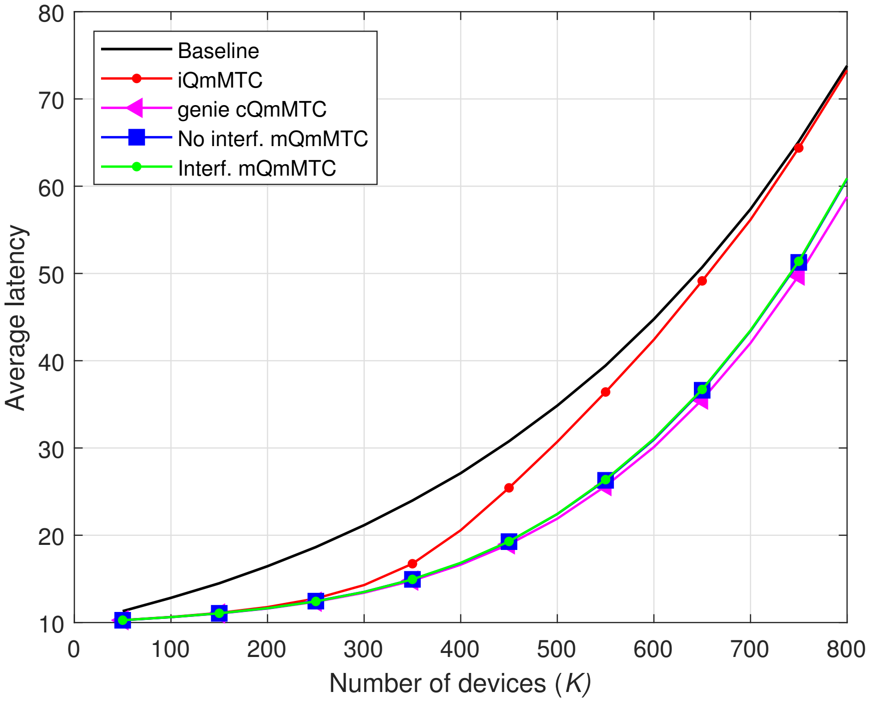

Figure 2. Average latency × K, for $L = 10$ packets, and $M = 100$.

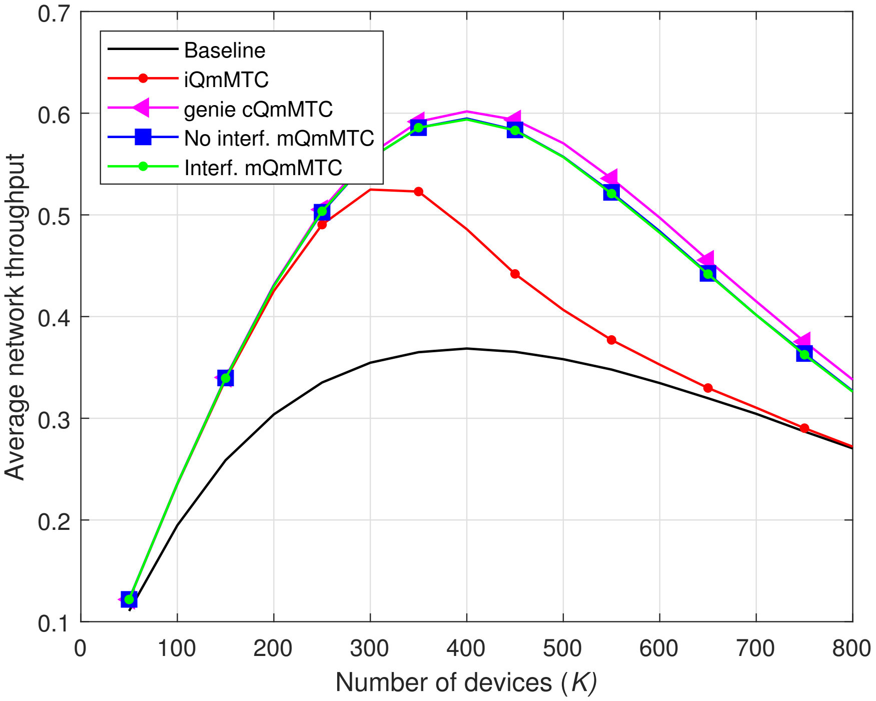

Figure 3. Average network throughput × K, for $L = 10$ packets, and $M = 100$.

Figures 6 and 7 present, respectively, the curves of the latency and network throughput *versus* L. The results shown in both figures are generated with 10,000 Monte-Carlo realizations. The number of active devices is kept fixed at $K = 600$, and the number of antennas is also fixed at $M = 100$. These results demonstrate the superiority of the proposed methods even when each device has a small number of packets to send ($L \leq 10$). In fact, a minimum number of $L = 2$ is enough for the proposed methods (mQmMTC without and with ICI) to produce a result superior to the baseline and the iQmMTC methods while approximating the cQmMTC method.

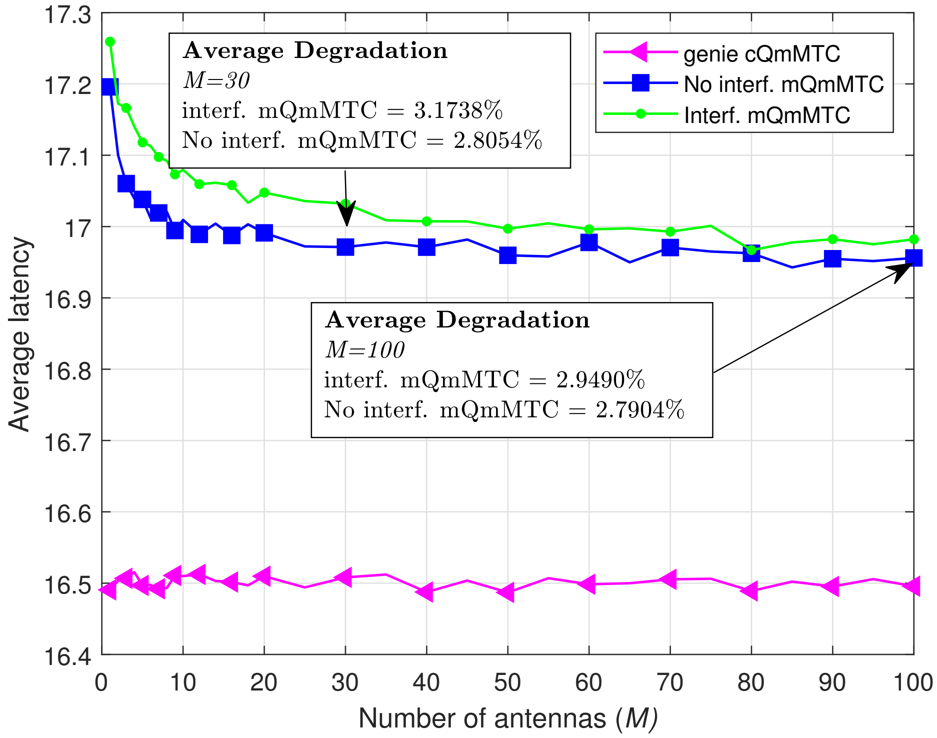

Figure 4. Average latency × M, for $L = 10$ packets, and $K = 400$.

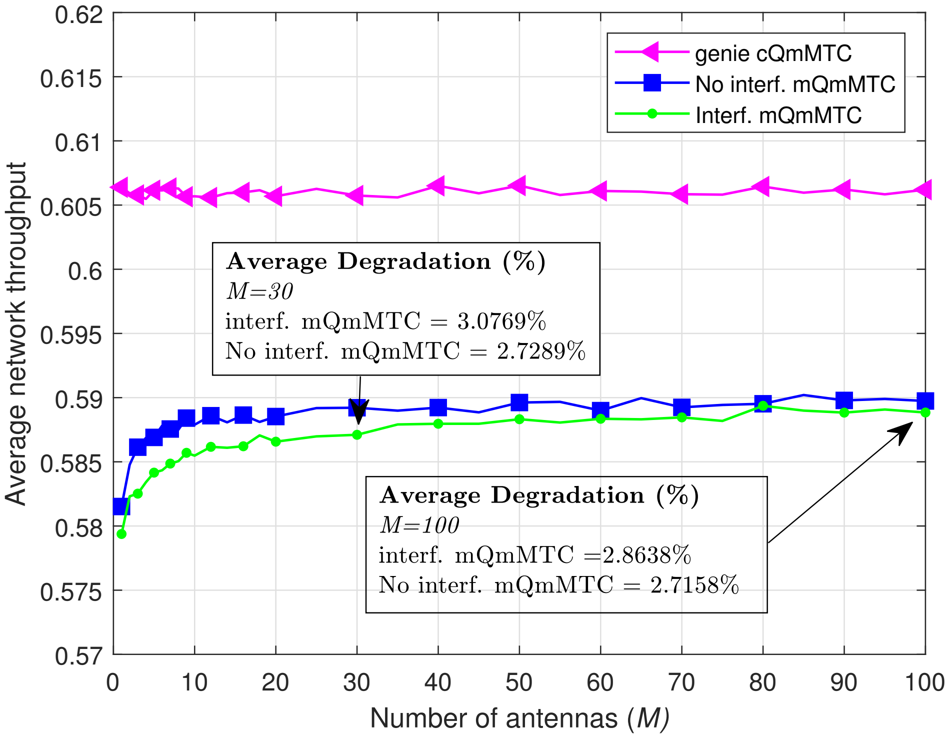

Figure 5. Average network throughput × M, for $L = 10$ packets, and $K = 400$.

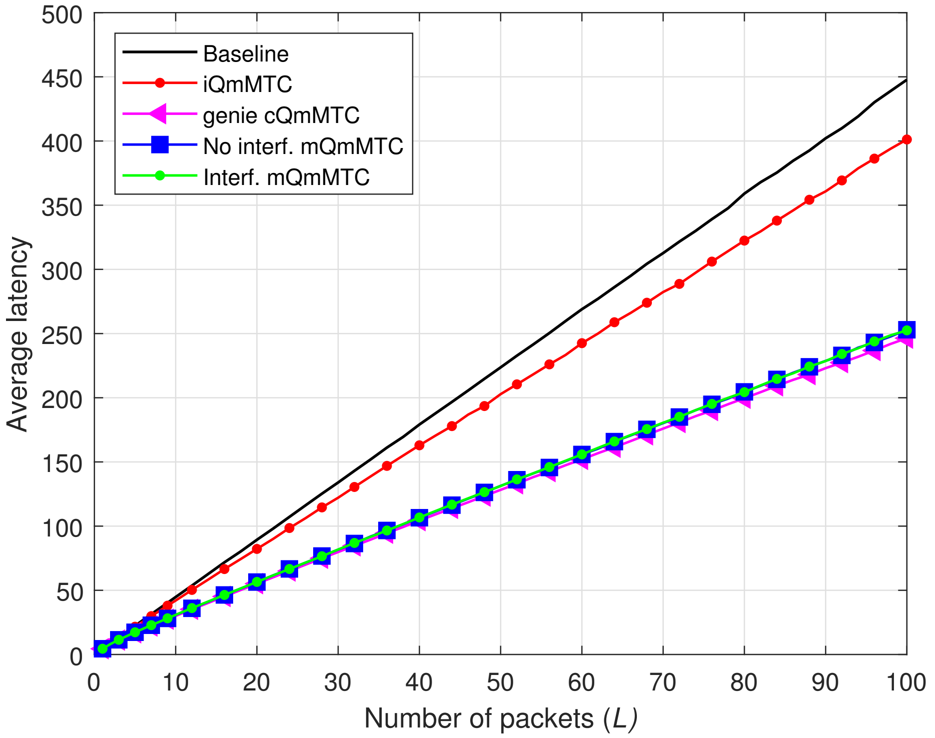

Figure 6. Average latency × L, for $K = 600$, and $M = 100$.

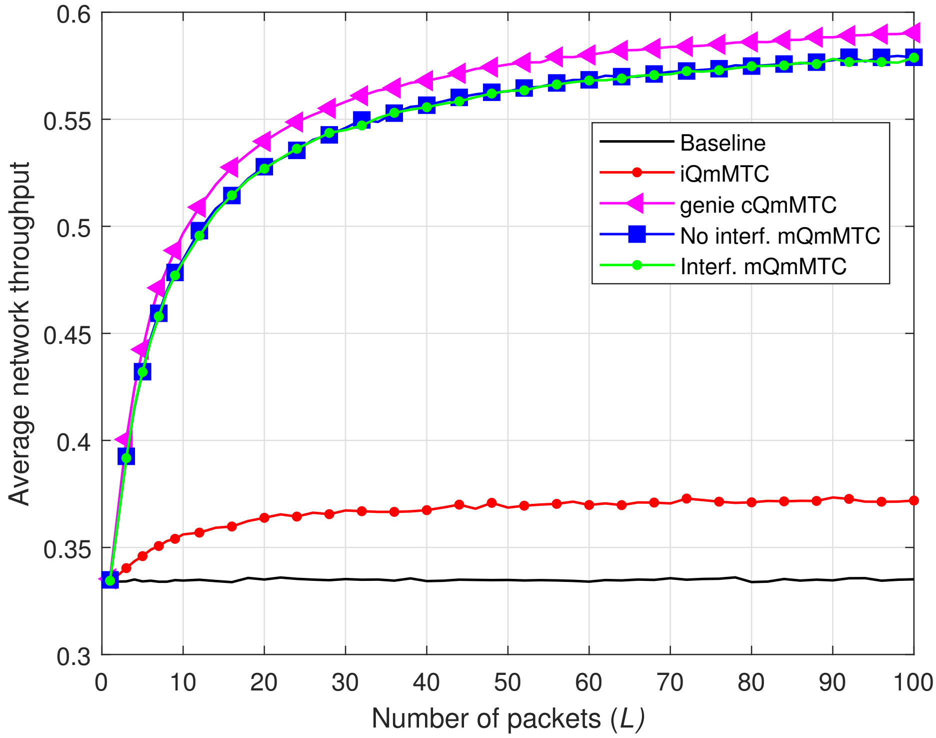

Figure 7. Average network throughput × L, for $K = 600$, and $M = 100$.

5.2. Random L_k and $P_a = 1$

In this subsection, we evaluate the scenario when the number of packets L_k sent by each device is random and follows a discrete uniform distribution as $L_k \sim \mathcal{U}[1, 10]$, while we still maintain $P_a = 1$ such that $K = K_a$. Figures 8 and 9 show, respectively, the performance results of latency and network throughput. One can note that the superiority of the results achieved by the proposed methods over the results achieved by the iQmMTC and baseline methods are preserved. Indeed, the curves in Figures 8 and 9 practically present the same shapes as the ones in Figures 2 and 3, respectively, but with a little performance degradation due to the reduction in the average number of transmitted packets, which limits the learning capability of the QL algorithm in seeking less congested pilots.

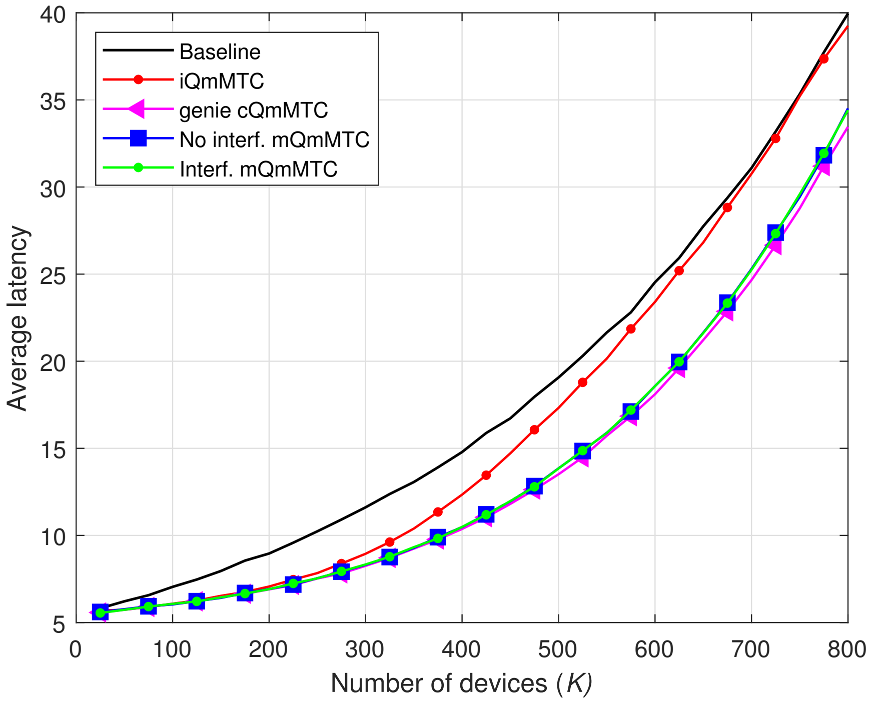

Figure 8. Average latency × K, for $L_k \sim \mathcal{U}(1,10)$, and $M = 100$.

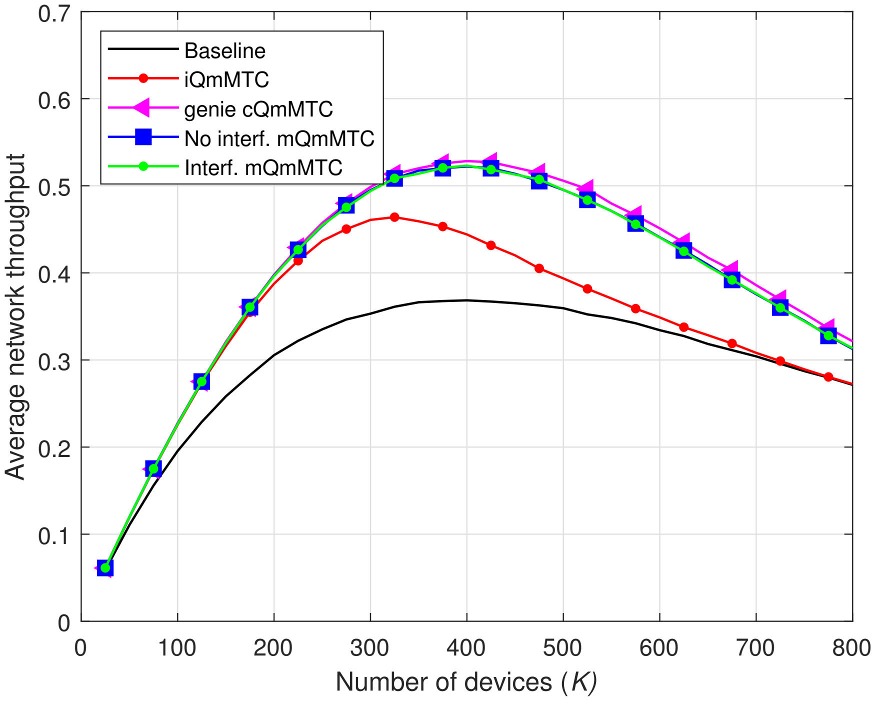

Figure 9. Average network throughput × K, for $L_k \sim \mathcal{U}(1,10)$, and $M = 100$.

5.3. Random L_k and $P_a = 0.1\%$

We consider in this subsection a random number of packets sent by each device following $L_k \sim \mathcal{U}[1,10]$, and a random number of devices being activated at each frame following a binomial distribution with an activation probability of $P_a = 0.1\%$, such that $K_a \leq K$. The number of available RA pilots is also reduced to $\tau_p = 40$ in order to keep the simulation time short. The results presented in Figures 10 and 11 are generated with 64000 Monte-Carlo realizations. Figure 10 presents the average per-device throughput and Figure 11 the average network throughput. The reward in the QL framework of the mQmMTC GF RA protocol is calculated using (19) while assuming that $\mathbb{E}[L_k] = 5.5$ is known at the devices' side. In terms of both performance metrics, our proposed mQmMTC protocol remains quite close to that of the ideal cQmMTC protocol, while always superior to that of Baseline and iQmMTC. While the per-device throughput of Baseline drops below 0.5 for $K \approx 2400$, this happens with $K \approx 2800$ for iQmMTC, with $K \approx 3100$ for mQmMTC, and with $K \approx 3400$ for the ideal cQmMTC. Similarly, the network throughput falling point in Figure 11 occurs with $K \approx 2300$ for the Baseline, with $K \approx 2400$ for iQmMTC, with $K \approx 2800$ for mQmMTC, and with $K \approx 3100$ for the ideal cQmMTC. These results corroborate the feasibility of the proposed mQmMTC GF RA protocol to address the challenges of massive machine-type communications in the framework of next-generation massive multiple access systems.

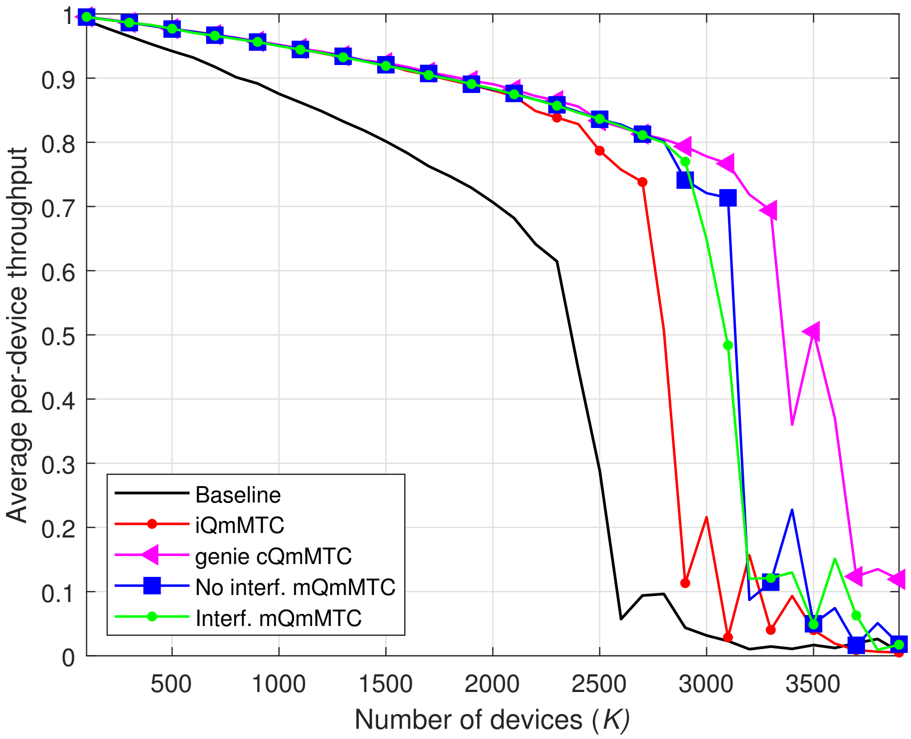

Figure 10. Average per-device throughput × K, for $L_k \sim \mathcal{U}(1,10)$, and $M = 100$.

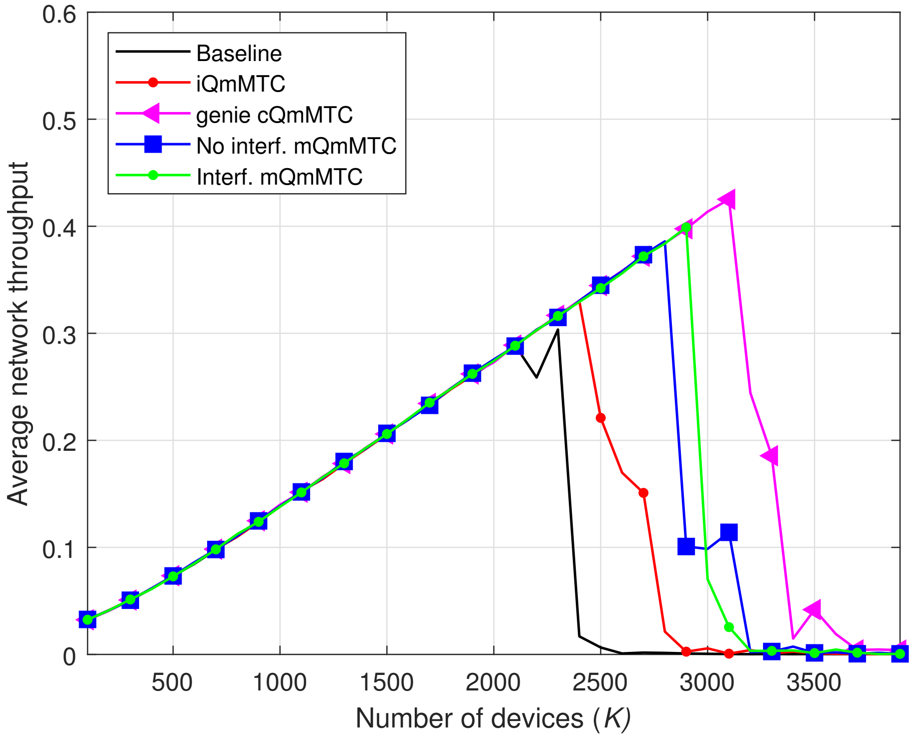

Figure 11. Average network throughput × K, for $L_k \sim \mathcal{U}(1{,}10)$, and $M = 100$.

6. Final Remarks

In this work, we have applied the collaborative, distributed, and decentralized QL-based GF RA protocol to a massive MIMO scenario for pilot collision control, assuming realistic wireless propagation effects, such as multipath fading, shadowing, path loss, thermal noise, and ICI. As devices cannot know the exact number of pilot contenders without incurring excessive complexity and signaling overhead, our proposed approach takes advantage of the massive number of BS antennas to allow the devices to compute the QL rewards in a simplified way. We have also shown that our proposed approach is robust regarding the number of packets to transmit, which can be as small as 10 or even random, following a discrete uniform distribution, and regarding the number of active devices, which can be randomly activated following a binomial distribution. Our proposed method is also robust regarding the number of antenna variations and does not require more than ≈30 antennas at the BS to produce significantly improved performance, very close to the ideal (*genie*) cQmMTC protocol of [30]. Possible research directions involve the extension of the protocol to non-terrestrial networks or leveraging other technologies like reconfigurable intelligent surfaces and cell-free massive MIMO.

Author Contributions: F.A.D.B.: Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing—original draft, Investigation, Validation. A.G.: Conceptualization, Methodology, Writing—review and editing, Visualization, Investigation, Supervision. T.A.: Writing—review and editing, Visualization, Investigation, Supervision, Project administration, Funding acquisition. J.C.M.: Conceptualization, Methodology, Software, Data curation, Writing—review and editing, Visualization, Investigation, Supervision, Project administration, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grants: 310681/2019-7 and 314618/2023-6, and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES, Grant ID: 88881.883505/2023-01).

Data Availability Statement: The raw data will be made available by the authors on request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chen, X.; Ng, D.W.K.; Yu, W.; Larsson, E.G.; Al-Dhahir, N.; Schober, R. Massive Access for 5G and Beyond. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 615–637. [CrossRef]
2. Choi, J.; Ding, N.; Le, N.-P.; Ding, Z. Grant-Free Random Access in Machine-Type Communication: Approaches and Challenges. *IEEE Wirel. Commun.* **2022**, *29*, 151–158. [CrossRef]
3. Ding, J.; Nemati, M.; Ranaweera, C.; Choi, J. IoT Connectivity Technologies and Applications: A Survey. *IEEE Access* **2020**, *8*, 67646–67673. [CrossRef]
4. Björnson, E.; Sanguinetti, L.; Wymeersch, H.; Hoydis, J.; Marzetta, T.L. Massive MIMO is a Reality-What is Next?: Five Promising Research Directions for Antenna Arrays. *Digit. Signal Process.* **2019**, *94*, 3–20. [CrossRef]
5. Marzetta, T.L. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [CrossRef]
6. Björnson, E.; Larsson, E.G.; Marzetta, T.L. Massive MIMO: Ten Myths and One Critical Question. *IEEE Commun. Mag.* **2016**, *54*, 114–123. [CrossRef]
7. Lee, M.; Kim, Y.; Piao, Y.; Lee, T.J. Recycling Random Access Opportunities with Secondary Access Class Barring. *IEEE Trans. Mob. Comput.* **2020**, *19*, 2189–2201. [CrossRef]
8. Björnson, E.; de Carvalho, E.; Sørensen, J.H.; Larsson, E.G.; Popovski, P. A Random Access Protocol for Pilot Allocation in Crowded Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 2220–2234. [CrossRef]
9. Han, H.; Fang, L.; Lu, W.; Chi, K.; Zhai, W.; Zhao, J. A Novel Grant-Based Pilot Access Scheme for Crowded Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2021**, *70*, 11111–11115. [CrossRef]
10. Marinello, J.C.; Abrão, T. Collision Resolution Protocol via Soft Decision Retransmission Criterion. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4094–4097. [CrossRef]
11. Han, H.; Li, Y.; Guo, X. A Graph-Based Random Access Protocol for Crowded Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 7348–7361. [CrossRef]
12. Marinello, J.C.; Abrão, T.; Souza, R.D.; de Carvalho, E.; Popovski, P. Achieving Fair Random Access Performance in Massive MIMO Crowded Machine-Type Networks. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 503–507. [CrossRef]
13. Casini, E.; De Gaudenzi, R.; Del Rio Herrero, O. Contention Resolution Diversity Slotted ALOHA (CRDSA): An Enhanced Random Access Scheme for Satellite Access Packet Networks. *IEEE Trans. Wirel. Commun.* **2007**, *6*, 1408–1419. [CrossRef]
14. Liva, G. Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA. *IEEE Trans. Commun.* **2011**, *59*, 477–487. [CrossRef]
15. Valentini, L.; Chiani, M.; Paolini, E. Massive Grant-Free Access with Massive MIMO and Spatially Coupled Replicas. *IEEE Trans. Commun.* **2022**, *70*, 7337–7350. [CrossRef]
16. Valentini, L.; Chiani, M.; Paolini, E. Interference Cancellation Algorithms for Grant-Free Multiple Access with Massive MIMO. *IEEE Trans. Commun.* **2023**, *71*, 4665–4677. [CrossRef]
17. Sharma, S.K.; Wang, X. Toward Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 426–471. [CrossRef]
18. Loubany, A.; Lahoud, S.; El Chall, R. Adaptive algorithm for spreading factor selection in LoRaWAN networks with multiple gateways. *Comput. Netw.* **2020**, *182*, 107491. [CrossRef]
19. Askhedkar, A.R.; Chaudhari, B.S. Multi-Armed Bandit Algorithm Policy for LoRa Network Performance Enhancement. *J. Sens. Actuator Netw.* **2023**, *12*, 38. [CrossRef]
20. Park, G.; Lee, W.; Joe, I. Network resource optimization with reinforcement learning for low power wide area networks. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 176. [CrossRef]
21. da Silva, M.V.; Montejo-Sánchez, S.; Souza, R.D.; Alves, H.; Abrão, T. D2D Assisted Q-Learning Random Access for NOMA-Based MTC Networks. *IEEE Access* **2022**, *10*, 30694–30706. [CrossRef]
22. Ko, Y.; Choi, J. Reinforcement Learning for NOMA-ALOHA Under Fading. *IEEE Trans. Commun.* **2022**, *70*, 6861–6873. [CrossRef]
23. Jeong, Y.J.; Yu, S.; Lee, J.W. DRL-Based Resource Allocation for NOMA-Enabled D2D Communications Underlay Cellular Networks. *IEEE Access* **2023**, *11*, 140270–140286. [CrossRef]
24. Yang, C.; Wang, Y.; Lan, S.; Zhu, L. Multi-agent Reinforcement Learning based Distributed Channel Access for Industrial Edge-Cloud Web 3.0. *IEEE Trans. Netw. Sci. Eng.* 2024, *early access*. [CrossRef]
25. Lee, I.; Kim, D.K. Decentralized Multi-Agent DQN-Based Resource Allocation for Heterogeneous Traffic in V2X Communications. *IEEE Access* **2024**, *12*, 3070–3084. [CrossRef]
26. Vaezi, M.; Azari, A.; Khosravirad, S.R.; Shirvanimoghaddam, M.; Azari, M.M.; Chasaki, D.; Popovski, P. Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 1117–1174. [CrossRef]
27. Ericsson. *Ericsson Mobility Report*; Technical Report; Ericsson: Stockholm, Sweden, 2023.

28. Bello, L.M.; Mitchell, P.; Grace, D. Application of Q-Learning for RACH Access to Support M2M Traffic over a Cellular Network. In Proceedings of the European Wireless 2014; 20th European Wireless Conference, Barcelona, Spain, 14–16 May 2014; pp. 1–6.
29. Silva, G.M.F.; Abrão, T. Throughput and latency in the distributed Q-learning random access mMTC networks. *Comput. Netw.* **2022**, *206*, 108787. [CrossRef]
30. Sharma, S.K.; Wang, X. Collaborative Distributed Q-Learning for RACH Congestion Minimization in Cellular IoT Networks. *IEEE Commun. Lett.* **2019**, *23*, 600–603. [CrossRef]
31. 3GPP. Spatial channel model for Multiple Input Multiple Output (MIMO) simulations. Technical Report (TR) 25.996, 3rd Generation Partnership Project (3GPP), 2018. Version 15.0.0. Available online: https://portal.etsi.org/webapp/workprogram/Report_WorkItem.asp?WKI_ID=55817 (accessed on 22 March 2024).
32. Marinello, J.C.; Brante, G.; Souza, R.D.; Abrão, T. Exploring the Non-Overlapping Visibility Regions in XL-MIMO Random Access and Scheduling. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6597–6610. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.