UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Matheus Valente da Silva

A NOMA-BASED Q-LEARNING RANDOM ACCESS METHOD FOR MACHINE TYPE COMMUNICATIONS

Florianópolis
2020

Matheus Valente da Silva

A NOMA-BASED Q-LEARNING RANDOM ACCESS METHOD FOR MACHINE TYPE COMMUNICATIONS

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia Elétrica.
Orientador: Prof. Richard Demo Souza, Dr.
Coorientador: Prof. Hirley Alves, Dr.

Florianópolis
2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

da Silva, Matheus Valente
A NOMA-BASED Q-LEARNING RANDOM ACCESS METHOD FOR MACHINE TYPE COMMUNICATIONS / Matheus Valente da Silva ; orientador, Richard Demo Souza, coorientador, Hirley Alves, 2020.
48 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2020.

Inclui referências.

1. Engenharia Elétrica. 2. Internet of Things. 3. MTC. 4. NOMA. 5. Q-Learning. I. Demo Souza, Richard. II. Alves, Hirley. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Matheus Valente da Silva

A NOMA-BASED Q-LEARNING RANDOM ACCESS METHOD FOR MACHINE TYPE COMMUNICATIONS

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Paulo Henrique Valente Klaine, Dr.
University of Glasgow

Prof. Samuel Montejo Sanchez, Dr.
Universidad Tecnológica Metropolitana

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Engenharia Elétrica.

Documento assinado digitalmente
Telles Brunelli Lazzarin
Data: 07/12/2020 08:24:43-0300
CPF: 023.400.719-25

Coordenação do Programa de
Pós-Graduação

Documento assinado digitalmente
Richard Demo Souza
Data: 04/12/2020 19:56:15-0300
CPF: 004.267.379-89

Prof. Richard Demo Souza, Dr.
Orientador

Florianópolis, 2020.

Este trabalho é dedicado aos meus queridos pais.

ACKNOWLEDGEMENTS

Aos meus pais pelo total apoio e encorajamento durante este percurso.

Aos professores, em especial meu orientador Richard Souza, meu coorientador Hirley Alves e Taufik Abraão pela colaboração no artigo fruto deste trabalho.

À minha namorada por todo apoio.

Ao Programa de Exelência Acadêmica (PROEX), através da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de incetivo.

RESUMO

Machine Type Communications (MTC) é um dos principais casos de uso do 5G e tende a se tornar ainda mais relevante nas próximas gerações. Além disso, por conta da natureza ultra-densa das redes de massive MTC (mMTC), a otimização de métodos de acesso ao meio apresenta diversos desafios. Uma solução promissora é a utilização de métodos de aprendizagem de máquina, como aprendizagem por reforço, para alocar eficientemente recursos de rádio aos dispositivos MTC. Com isso em mente, neste trabalho é proposto um método distribuído baseado em Non-Orthogonal Multiple Access (NOMA) e *Q*-Learning para alocar dinamicamente os dispositivos MTC. Os resultados numéricos demonstram que o método proposto é capaz de melhorar muito o throughput da rede quando comparado a métodos de trabalhos recentes.

Palavras-chave: Internet das Coisas, Comunicações entre Máquinas, NOMA, Aprendizagem de Máquina.

RESUMO EXPANDIDO

Introdução

A implementação do 5G e demais redes de dispositivos móveis como a Internet da Coisas move o desenvolvimento de tecnologias de comunicação de máquinas. Essas redes e aplicações devem ser capazes de suportar uma quantidade enorme de dispositivos. A natureza ultra-densa das redes mMTC apresenta diversos desafio para a otimização de métodos de acesso ao meio. Uma das soluções promissoras senso estudadas é a utilização de algorítimos de aprendizagem de máquina para controle do tráfego de dispositivos. Entre esses algoritmos o *Q-Learning* se apresenta como um dos mais interessantes por sua capacidade de ser implementado de uma maneira distribuída. Além disso, para atender os requisitos também é necessário aumentar a eficiência espectral. Com isso em mente, neste trabalho é proposto um método distribuído baseado em *Non-Orthogonal Multiple Access* (NOMA) e *Q-Learning* para alocar dinamicamente os dispositivos MTC. Os resultados numéricos demonstram que o método proposto é capaz de melhorar muito o *throughput* da rede quando comparado a métodos de trabalhos recentes.

Objetivos

Esta dissertação tem como objetivo principal realizar o estudo dos efeitos da combinação de aprendizagem de máquina com NOMA para a melhoria de esquemas de acesso aleatório. Os objetivos secundários são: (i) propor um esquema que utilize *Q*-Learning e NOMA; (ii) analisar a melhoria no throughput da rede; (iii) analisar o impacto do esquema proposto em situações dinâmicas; (iv) analisar o efeito do controle de potência sobre a rede; (v) encontrar parâmetros para a otimização do algorítimo.

Metodologia

Inicialmente é apresentada uma revisão das tecnologias do estado da arte. Onde é apresentado o 5G, suas diferentes categorias e requisitos e a sua relação com as redes máquina-a-máquina. Além disso é feita uma revisão das tecnologias que integram as redes máquina-máquina atualmente. Seguindo com uma análise do 6G, suas tendências e seus desafios e como este trabalho irá focar em duas tecnologias cruciais para o sucesso do 6G que é NOMA e técnicas modernas de acesso aleatório.

Em seguida é apresentado o modelo do sistema a ser considerado durante o trabalho. O modelo este que assume um sistema onde todos os dispositivos rodam a mesma aplicação. Assim N dispositivos sincronizados são dispostos de forma uniformemente aleatória ao redor de uma estação rádio base. É definido também que o método de acesso ao meio é feito pro *grant-free Slotted Aloha*. Assim cada dispositivo transmite uma vez por *frame*.

Seguindo com proposta de trabalho é apresentado como NOMA é implementado no sistema. É considerado o uso de NOMA no *uplink*. Sendo feito o cancelamento sucessivo de interferência na estação rádio base. Assume-se também que os dispositivos são decodificados em ordem decrescente de potência recebia. Nota-se que para um melhor funcionamento do NOMA é necessária uma diversidade de potência recebida. Assim, o sistema permite que os dispositivos calculem uma potência de referência

a partir do seu conhecimento do canal e apliquem um desvio de Δ. Com isso os dispositivo podem transmitir em três níveis de potência $P_{ref} - \Delta, P_{ref}$ e $P_{ref} - \Delta$.

Feita a definição do modelo do sistema é apresentado o método proposto iniciando-se pela apresentação do algorítimo *Q-Learning*. Adapta-se o algorítimo para que ele seja aplicado de forma distribuída, assim cada dispositivo possui a sua própria *Q-Table*. Para isso os níveis de potência do diapositivos são considerados os estados enquanto a ação é considerada como transmitir em um determinado *slot*. Também é determinada a recompensa de +1 para transmissões com sucesso e de –1 para transmissões falhas.

Resultados e Discussão
O método proposto é então comparado com outros três esquemas: i) *Slotted Aloha*; *Slotted Aloha* com NOMA; e o método *Q-Learning* Colaborativo. Para analisar o comportamento médio do dos esquemas todas as curvas são o resultado de uma média de 30 simulações. Assim o método proposto apresenta o pico de *throughput*, calculado ao longo deste trabalho como o transmissões com sucesso por *slots*, em $N = 2,5K$ com um *throughput* de $1,6$. Assim com uma melhora significante sobre o *Q-Learning* Colaborativo que tem seu pico em $N = K$. Isto pode ser explicado pelo efeito coletivo de NOMA e o controle de potência que permitem uma melhor eficiência espectral.

Segue-se com uma análise do efeito de γ, coeficiente de desconto. Onde é apresentada uma análise sobre uma rede dinâmica. Nota-se que apesar de $\gamma = 0,5$ apresentar *throughputs* mais altos ele não é capaz de se recuperar de um *overload*. Justificando a escolha de $\gamma = 1$. Demonstra-se também que o método proposto consegue dar a possibilidade de todos os dispositivos serem decodificados. Alocando até três dispositivos por *slot*. Enquanto o método *Q-Learning* Colaborativo consegue alocar pelo menos um dipositivo por *slot* mas apresenta uma forte queda no *throughput* por não utilizar NOMA.

Além disso, também é analisado o efeito do controle de potência e NOMA sobre a potência de transmissão dos dispositivos. Nota-se que se NOMA não é utilizado em conjunto com o controle de potência há um aumento da potência de transmissão enquanto temos mais dispositivos que *slots* porque os dispositivos aprendem que uma potência maior significa uma maior probabilidade de sucesso em sua transmissão.

Também foram analisados outros parâmetros do algoritmo, como a política de exploração e a forma de inicialização da $Q-$ *Table*. Encontrou-se que os melhores resultados foram apresentados pela política de exploração *greedy* e com uma inicialização aleatória.

Considerações Finais
Esta dissertação introduz um novo método para acesso aleatório combinando a habilidade de medir incertezas do algorítimo *Q-Learning* e a eficiência espectral de NOMA. O método proposto permite também que os dispositivos reconheçam seus parceiros NOMA de maneira autônoma, sem a necessidade de um custo esquema de pareamento. Além disso, também aprendem qual é o melhor *slot* para transmitir. O método ainda requere uma complexidade mínima por parte do dispositivo. Sendo necessário

implementar apenas uma equação e memória o suficiente para uma tabela de tamanho dependente dos níveis de potência e *slots* do *frame*. Nota-se que o controle de potência além de fornecer a diversidade de potência recebida necessária para o bom funcionamento do NOMA ainda permite que os dispositivos utilizem uma menor potência de transmissão dada a redução da competição por *slot* fornecida pelo algorítimo.

Palavras-chave: Internet das Coisas, Comunicações entre Máquinas, NOMA, Aprendizagem de Máquina.

ABSTRACT

Machine Type Communications (MTC) is a main use case of 5G and beyond wireless networks. Moreover, due to the ultra-dense nature of massive MTC networks, Random Access (RA) optimization is very challenging. A promising solution is to use machine learning methods, such as reinforcement learning, to efficiently accommodate the MTC devices in RA slots. In this sense, we propose a distributed method based on Non-Orthogonal Multiple Access (NOMA) and *Q*-Learning to dynamically allocate RA slots to MTC devices. Numerical results show that the proposed method can significantly improve the network throughput when compared to recent work.

Keywords: Internet of Things, Machine Type Communications, NOMA, Machine Learning.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACB	Access Class Barring
BS	Base Station
CSI	Channel State Information
eMBB	enhanced Mobile BroadBand
HTC	Human-Type-Communication
IoT	Internet of Things
LoS	Line of Sight
LTE	Long-Term Evolution
MAC	Medium Access Control
MTC	Machine-Type-Communications
MTD	Machine-Type-Device
MDP	Markov Decision Process
mMTC	massive Machine-Type-Communications
NOMA	Non-Orthogonal Multiple Access
PDCCH	Physical Downlink Control CHannel
PHY	PHYsical
RAN	Radio Access Network
RA	Random Access
RACH	Random Access CHannel
SINR	Signal to Interference plus Noise Ratio
SA	Slotted Aloha
SIC	Successive Interference Cancellation
URLLC	Ultra Reliable Low Latency Communications

LIST OF SYMBOLS

B	Bandwidth
f_c	Carrier frequency
η	Path loss exponent
Δ	Power deviation
F	Noise Figure
P_t	Transmit power
r	Spectral efficiency
d_0	Reference distance
N	Number of devices
L	Number of Messages
K	Time-Slots
γ	Discount factor
α	Learning Rate
R	Reward

CONTENTS

1 INTRODUCTION

The deployment of 5G and beyond mobile networks, including the Internet of Things (IoT), is driving the development of advanced Machine-Type-Communications (MTC) networks (TULLBERG et al., 2016; AAZHANG et al., 2019; BI, 2019). These networks should be able to support new applications with a massive number of devices, such as those in smart cities and industry 4.0. According to Cisco, the number of MTC devices can be as large as 3.9 billions by 2022 (CISCO, 2019). With the surge of new devices many challenges arise with massive MTC (mMTC) networks, such as meeting diverse performance requirements and congestion in Radio Access Network (RAN). Moreover, mMTC networks suffer from inefficient Random Access (RA) procedures and resource allocation with current RA protocols performing poorly in ultra-dense networks (CLAZZER, 2019), and leading to the need of efficient RA schemes able to handle massive requests.

In terms of standardization, 3GPP is evolving 5G to improve mobility, while aggregating physical downlink control channel (PDCCH) enhancements, and addressing new MTC use cases. Enhancements in 3GPP Rel-16 and Rel-17 (3GPP, 2020; 5G AMERICAS, 2020) include: a) 2-step Random Access Channel (RACH) to reduce latency and signaling overhead; b) reliability improvements; c) power saving techniques; d) enhanced support for new use cases, including industrial IoT. However, there is still plenty of room for improvement both in terms of performance and efficiency.

The authors in (SHARMA; WANG, 2019b) present a comprehensive survey of the issues related with RAN congestion while introducing machine learning algorithms to improve RA for mMTC networks. Machine learning presents itself as a good alternative to solve the congestion problem as it is able to improve scheduling without a complex algorithm. Among other machine learning techniques, the reinforcement learning method known as *Q*-Learning stands out due to its capability of being implemented in a model-free and distributed manner (SUTTON; BARTO, 2018). The authors of (SHARMA; WANG, 2019a) propose a *Q*-Learning based method to address the RAN congestion using the number of collisions per slot as a reward. However, it needs substantial feedback from the base station (BS), besides the complexity of determining the number of colliding devices. Another example is (BELLO et al., 2018), which attempts to conciliate the traffic load of Human-Type-Communication (HTC) with MTC devices in the RA of a cellular network, making MTC devices learn which slot to access, reducing collisions and improving throughput.

The work in (JIHUN MOON; YUJIN LIM, 2017) utilizes *Q*-Learning at the BS to better adapt the barring factor in an Access Class Barring (ACB) scheme. Although this method reduces the load in the network, it is a reactive solution, while the current ever increasing number of devices calls for a proactive solution that tries to avoid

collisions rather than recovering from it at the base station. Another work that uses adaptive ACB is (ZHAO et al., 2018). They propose two algorithms to minimize delay or maximize throughput. However, one needs to know the number of devices in the network and the other requires the BS to know how many devices collided in a past slot. In (MOHAMMED et al., 2015), *Q*-Learning is used to select the best available BS in a Long-Term Evolution (LTE) network, using throughput and delay both as a QoS measurement and as the reward for the MTC devices. Whilst the MTC devices do select the best BS in this scheme, thus efficiently organizing RA within a cell, it does not deal with the growth in density of the mMTC networks and therefore overload is still a problem. In (HAN et al., 2019), Non-Orthogonal Multiple Access (NOMA) (SAITO et al., 2013) and *Q*-Learning are utilized in order to maximize energy efficiency in short packet communications. The method in (HAN et al., 2019) makes use of *Q*-Learning with the goal of pairing devices in sub channels. In order to maximize energy efficiency, they propose a power allocation scheme, finding the optimal transmit power for each device.

In this work, we propose the use of *Q*-Learning and NOMA, alongside with a power control scheme, to improve the throughput in mMTC networks. This work differs from (SHARMA; WANG, 2019b, 2019a; BELLO et al., 2018) because, besides using *Q*-Learning for slot allocation, we implement NOMA and consider the effect of path loss and fading. Unlike (SHARMA; WANG, 2019a), this method requires minimal feedback from the BS, a single bit per time slot, instead of the number of contending devices per time slot. Moreover, even though (JIHUN MOON; YUJIN LIM, 2017; MOHAMMED et al., 2015) use *Q*-Learning in a MTC network, neither use it to improve slot allocation. The first adapts a class barring factor, while the second selects the best BS for connection. Moreover, in (ZHAO et al., 2018) machine learning techniques are not used. Finally, compared to (HAN et al., 2019), we are looking at improving throughput rather than energy efficiency, although our proposed power control scheme prevents the excessive use of power.

The main contributions of this work are: i) evaluation of the beneficial impact in RA when combining *Q*-Learning with NOMA; ii) a RA scheme which improves the network throughput with limited transmit power and complexity. The average throughput is 2.52 times higher than the compared method for the case of 225 devices and 100 time-slots. The peak throughput is reached at 225 devices while the compared method reaches its peak at 100 devices. The increased number of the devices can be attributed to NOMA's spectral efficiency.

1.1 PUBLICATION

The work related in this dissertation yielded a publication in the IEEE Wireless Communications Letters:

M. V. da Silva, R. D. Souza, H. Alves and T. Abrão, “A NOMA-based *Q*-Learning Random Access Method for Machine Type Communications,” in *IEEE Wireless Communications Letters* , DOI: 10.1109/LWC.2020.3002691.

2 MACHINE TYPE COMMUNICATIONS

MTC is a fundamental building block of both mMTC and Ultra Reliable Low Latency Communications (URLLC), two of the three mains uses cases defined for 5G systems, the other being enhanced Mobile BroadBand (eMBB). The eMBB use case aims to achieve high date rates that address mainly human type communications such as high speed Internet, video calls and media streaming. URLLC focuses on providing an extremely reliable connection mainly for critical MTC applications, while mMTC intends to support a huge number of devices with limited radio and processing resources (OSSEIRAN et al., 2020). Thus, both URLLC and mMTC target mainly MTC applications but with different requirements. While the former needs high reliability to support applications like autonomous driving and real-time monitoring and control, the latter needs to support low-cost sensors and meters. One way to visualize the different 5G requirements is the diagram in Fig. 1.

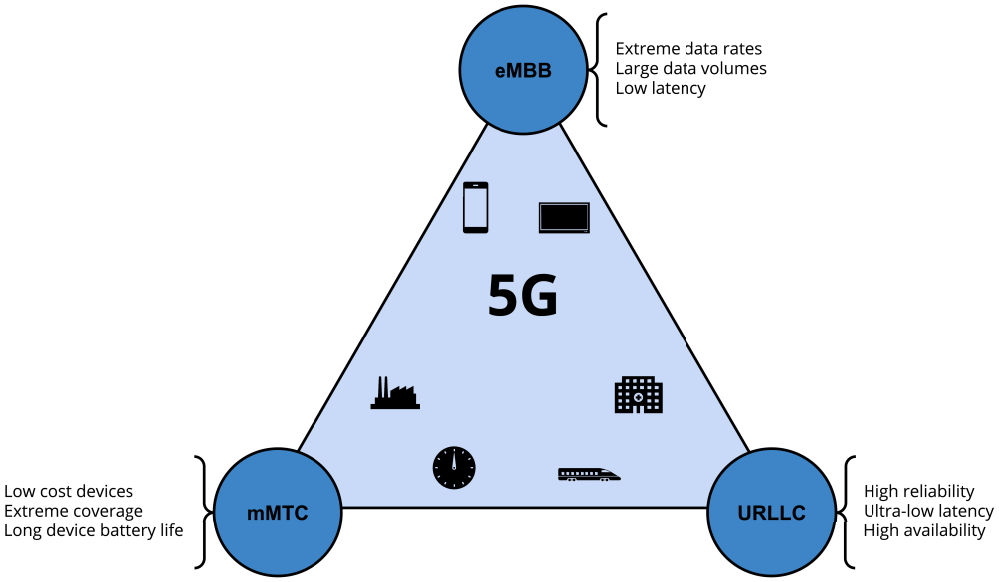


Figure 1 – Different requirements of the three main use cases planned for 5G: mMTC, URLLC and eMBB adapted from (OSSEIRAN et al., 2020)

Moreover, one of the biggest drivers of 5G is the IoT. According to Ericsson there are over one billion cellular IoT connections in 2020 alone (OSSEIRAN et al., 2020). The main IoT segment that better incorporates MTC is defined by Ericsson as massive IoT targeting a huge flow of low-complexity devices that do not need to communicate very often. As URLLC is not the focus of this work it will not be further discussed. However, recent developments of this use case are thoroughly discussed by the authors of (POPOVSKI et al., 2019).

2.1 CURRENT TECHNOLOGIES

IoT applications nowadays are being developed and deployed using several different standards and protocols. These protocols are not always originally designed for the IoT and, given the different set of requirements for each application, no protocol is able to fulfill such a wide range of specifications.

The most common IoT networks nowadays either use unlicensed spectrum through short-range technologies such as IEEE802.15.x and 802.11x or a proprietary technology such as SigFox or LoRA (BOCKELMANN et al., 2018), which shows the market need for new MTC solutions. In the following sub-sections we overview some of the most prominent technologies currently being used to implement IoT applications.

2.1.1 Bluetooth

The Bluetooth technology was invented by Ericsson in 1994 (LETHABY, 2017). The wireless standard for communication between mobiles phones and computers remains appealing due to its widespread implementation, specially to industrial applications such as data loggers and smart metering. Originally standardized as IEEE802.15.1, the Bluetooth link layer is now controlled by Bluetooth SIG. The original Bluetooth is now called the Classic Bluetooth, and the Bluetooth Low Energy (BLE) invented by Nokia was added to the standard Bluetooth 4.0 in 2010 making a shift towards the IoT applications (LETHABY, 2017). The Bluetooth 5, the most current BLE standard implemented in 2016, can reach data rates up to 2 Mbps and distances up to 750 meters (LETHABY, 2017).

2.1.2 IEEE802.15.4 based technologies

There are many technologies based on the IEEE802.15.4 standard. Such as Zigbee, Thread and 6LoWPAN, among others. This is the most prominent standard for low-power radio technologies and defines both the Physical (PHY) layer and the Medium Access Control (MAC) layer (PALATTELLA et al., 2013). The main characteristics of the 802.15.4 are listed by (MA; LUO, 2008) as:

1. Relatively low transmission rate: up to 250 kbps.
2. Low power consumption: small batteries should last several months.
3. Low cost: limited embedded processing power.
4. Short range: up to 100m.
5. Multiple device types: Full Function Device (FFD) and Reduced Function Device (RFD).

6. Two transmission modes: the beacon-enabled mode and the non-beacon-enabled mode.

7. Supports both mesh and star topologies.

The 6LoWPAN standard (P. THUBERT et al., 2017) defines an efficient adaptation layer between the 802.15.4 link layer and the TCP/IP protocol providing low-power devices direct access to the Internet. In fact, 6LoWPAN stands for IPv6 over Low Power Wireless Personal Area Networks. As the 6LoWPAN protocol is only an adaptation layer it offers every advantage of the 802.15.4 protocol with implementations for both ISM bands 2.4 GHz and 868/915 MHz. As most of the Internet infrastructure still relies on IPv4 most 6LoWPAN implementations come with an IPv4-IPv6 converter. One of the downsides of this protocol is the fact that as the 802.15.4 link layer has multiple modes the multiple 6LoWPAN solutions from different developers are not able to interact at the local level. One of most popular open-source implementations of this protocol is the Contiki OS (CONTIKI-OS. . . , 2020).

Trying to fill the gap left by 6LoWPAN the Thread Group (THREAD. . . , 2020) formed by Google, Samsung and other companies defined the Thread standard that connects low-power devices directly to the Internet and also ensures interoperability. Thread can reach data rates up to 250 kbps, operates in the 2.4 GHz ISM band and uses a mesh network implementation.

ZigBee is another standard based on 802.15.4. The standard is maintained by the ZigBee Alliance (ZIGBEE. . . , 2020), which consists of more than 400 companies. ZigBee is meant to be a complete solution providing interoperability among devices from different manufacturers. It operates in the 2.4 GHz ISM band and the network is designed with a mesh-topology. The main focus of ZigBee applications lies in smart homes and smart buildings.

2.1.3 Wi-Fi

The Wi-Fi technology (WI-FI. . . , 2020) is based on the IEEE802.11 standard and was designed as a wireless replacement for the Ethernet 802.3. The Wi-Fi can be considered ubiquitous as it is deployed in most offices, schools and other business. The advantage of being widely deployed led IoT applications to use Wi-Fi without needing additional infrastructure and custom gateways. The Wi-Fi embraces different protocols from IEEE802.11, with 802.11ah, known as Wi-Fi Ha-Low, being the one designed for low power devices for applications such as home automation. This protocol uses the 900MHz band and provides a longer battery life (LETHABY, 2017).

2.1.4 LPWANs

Low-Power Wide Area Networks (LPWANs), unlike the technologies above, focus on long range radio communication, from 10km to 30km in rural areas and 1km to 5km in urban areas (MEKKI et al., 2019). Two of the most well established LPWAN solutions are SigFox and LoRA.

SigFox (SIGFOX. . . , 2020) works as a network provider, its network utilizes a ultra-narrow band modulation in the 868/915 MHz ISM band. It can provide a range up to 30km in rural areas. SigFox is able to achieve such a long range by using an extremely low data rate up to 100 bps.

LoRa in its turn is a physical layer protocol that uses Chirp Spread Spectrum (CSS) modulation spreading a narrow-band signal over a wider channel, resulting in a noise and interference resilient communication (LORA, 2015). LoRaWAN is a LoRa based MAC layer protocol maintained by the LoRa Alliance (LORA. . . , 2020). LoRaWAN defines different device classes to better fulfill the different requirements for distinct IoT applications.

2.1.5 Cellular

Cellular networks, even though they were not initially designed for IoT applications, are very attractive to the IoT market due to their worldwide availability. The downside of using cellular networks is that they tend to be cost and energy inefficient. With that in mind, the 3GPP defined two standards that work as an extension of LTE for IoT applications (T-MOBILE, 2019). The Long Term Evolution for Machines (LTE-M) was defined in Release 12 in 2015 and the NB-IoT in Release 13 in 2016. While the first uses a significantly higher bandwidth and reaches higher data rates the last one is less complex and cheaper, making it extremely attractive to IoT applications (ZAIDI et al., 2019).

A simplified comparison of the technologies described above can be seen in Table 1.

Table 1 – Current wireless IoT technologies adapted from (LETHABY, 2017) and (T-MOBILE, 2019).

	Maximum data rate	**Range**	**Power consumption**	**Topology**
Wi-Fi	72 Mbps	100 m	Up to 1 Year	Star
BLE/Bluetooth 5	2 Mbps	750 m	Up to years	Point-to-point/Mesh
Thread	250 Kbps	100 m	Up to years	Mesh and Star
ZigBee	250 Kbps	130m LoS	Up to years	Mesh and Star
SigFox	100 bps	30 km	Up to 10 Years	Star
LoRaWAN	20 kbps	20 km	Up to 10 Years	Star of Stars
NB-IoT	250 Kbps	10 km	Up 5 to 10 Years	Star
LTE-M	1 Mbps	10 km	Up to 10 Years	Star

2.2 5G AND MTC REQUIREMENTS

The vision for the 5G intended to include a wide range of diverse applications with its main requirements being lower latency, higher data rates, greater reliability and increased security (OSSEIRAN et al., 2020). However, as previously explained, 5G was divided into three categories in order to meet these divergent demands. The sheer number of mMTC devices connected calls for a paradigm shift in device connectivity and management. That being said, to fulfill the vision of billions of wireless connected devices, 5G defines the connection density as 1,000,000 devices per squared kilometer with a minimum QoS value (OSSEIRAN et al., 2020). The other key requirement for mMTC applications is the energy efficiency, as in many deployments devices are meant to be in areas difficult to access. Therefore, the devices should not only be energy efficient while communicating, but also consume very little energy when there are no transmissions (OSSEIRAN et al., 2020).

Besides massive connectivity, industrial applications usually require low latency and high reliability. Both of which are key driving requirements of 5G. According to Ericsson (OSSEIRAN et al., 2020), the main requirements for URLLC and mMTC are the ones defined in Table 2 below.

Table 2 – 5G Requirements for URLLC and mMTC (OSSEIRAN et al., 2020)

Requirement	**Value**
User plane latency	1ms
Control plane latency	10 ms
Connection latency	1,000,000 devices per km^2
Reliability	99.999% success rate
Mobility interruption time	0 ms
Battery life	10 years

2.3 6G: CHALLENGES AND TRENDS FOR MTC

Following the trend of a new generation of cellular communications every decade, by 2030 we should have 5G with world wide deployment. Such fact raises the question, what direction will 6G take? According to (MAHMOOD et al., 2020), the following six trends will form society in the next ten years and therefore drive the development of new technologies:

1. Autonomous mobility: MTC and Artificial Intelligence will play a big role in autonomous driving.
2. Connected living: With 6G cities and homes should be fully connected.
3. Factories of the future: Industry 5.0 should be more interactive with real time control and monitoring.

4. Digital reality as frontier technology: The augmented, virtual and mixed reality should play a fundamental role in man and machine interaction.

5. Towards a 'zero' world: Zero-energy and zero-touch paradigms call for improvements on MTC devices and networks.

6. Data as the new oil: Considering the amount of devices connected the data collected will be extremely important as the age of information arises.

The drivers and use cases above point towards a new set of requirements. The already stringent requirements of the 5G in latency, reliability, connection density, low cost and low energy will become even more rigorous. The connection density should be up to 100 devices per $1m^3$ in order to keep up with Industrial IoT and the arrival of Industry 5.0. Real-time monitoring and control will require an end-to-end latency of $1ms$. When thinking about energy efficiency, 6G aims at zero energy, making devices fully sustainable. Besides the requirements inherited from 5G becoming more strict, 6G has its own set of requirements to fulfill its vision.

As discussed before in this chapter, MTC is divided into two 5G use cases, URLLC (or critical MTC) and mMTC. With 6G applications becoming even more diverse, the authors of (MAHMOOD et al., 2020) propose a new set of categories for 6G MTC:

1. Dependable cMTC: ultra-reliability and low latency with security.

2. Broadband cMTC: high data rate with high reliability and low latency.

3. Scalable cMTC: massive connectivity with high reliability and low latency.

4. Globally-scalable mMTC: supporting ultra-wide coverage.

5. Zero-energy mMTC: Energy-efficient radios with extremely long battery life.

As we can see, 6G envisioned service classes mostly involve combining 5G service characteristics which highly depend on improvements at the PHY and MAC layers for massive connectivity. At the PHY Layer, the authors of (MAHMOOD et al., 2020) point out that we should look into non-orthogonal solutions, CSI-free/limited schemes and coding for short packets. NOMA has the potential to improve resource sharing while CSI-free can become more interesting in 6G. The likelihood of operating with a strong line of sight increases with denser networks and statistical beam-forming relying on channel statistics can operate with near-optimum performance eliminating the need for CSI acquisition (MAHMOOD et al., 2020). Coding for short packets becomes increasingly important as the coding schemes for 5G, low-density-parity-check and polar codes, are not well optimized for short packets. Medium Access challenges include the need for a modern random access design as scheduling a huge amount of devices

rapidly becomes impractical and random access schemes like ALOHA have serious throughput limitations. In this work we will be exploring two of the above technologies seen as fundamental for 6G: NOMA and modern RA.

3 SYSTEM MODEL

Assuming a single communication system in which all devices run the same application, we consider a setup with N synchronized devices distributed in a circular cell around a common Base Station (BS), as shown in Fig. 2. All devices transmit at the same power P_t, frequency f_c, and rate, each one having L data packets ready for transmission. Medium access is based on grant free Slotted Aloha (SA), where each device transmits in one of K time-slots within a frame, which is illustrated in Fig. 3. There is no restriction on the quantity of devices per time-slot thus several devices can be allocated to the same time-slot. After each frame the BS sends a group feedback using one bit per time-slot, informing if the transmissions were successful or not. This control message is also used to synchronize the devices. As usual, we assume that the BS acquires CSI by means of pilots within a header contained in each transmission from the devices. Moreover, assuming a quasi-static scenario, the devices can estimate the statistics of their channels using the common control message and apply channel inversion to reach a reference average power at the BS that assures a given outage probability.

The message from the m-th device, $m \in \{1, 2, \cdots, M\}$, $M \leq N$, transmitting in the k-th time slot, $k \in \{1, 2, \cdots, K\}$, is considered to be successfully decoded if the Signal to Interference plus Noise Ratio (SINR) at the BS is larger than the threshold from Shannon's capacity, so that (GOLDSMITH, 2005)

$$\text{SINR}_{m,k} \geq 2^r - 1, \tag{1}$$

where $\text{SINR}_{m,k}$ is the SINR for the m-th device transmitting in the k-th time slot, and r is the spectral efficiency in bits/s/Hz. As we consider a quasi-static non-LoS scenario, the asymptotic outage probability is a meaningful performance metric even in the finite blocklength regime (MARY et al., 2016).

3.1 NOMA

We consider the use of NOMA in the uplink, with SIC at the BS to decode colliding packets (SAITO et al., 2013). The signal received by the BS in the k-th time-slot, in the t-th frame, is

$$y_k(t) = \sum_{m=1}^{M} x_{m,k}(t) + n_k(t), \tag{2}$$

where $x_{m,k}(t)$ is the attenuated signal received at the BS from the m-th device in time-slot k, with instantaneous power $P_{m,k}$, while $n_k(t)$ is the additive white Gaussian noise.

The BS then performs SIC on the overall received signal y_k, starting from the strongest to the weakest user. Without loss of generality we assume that the users are

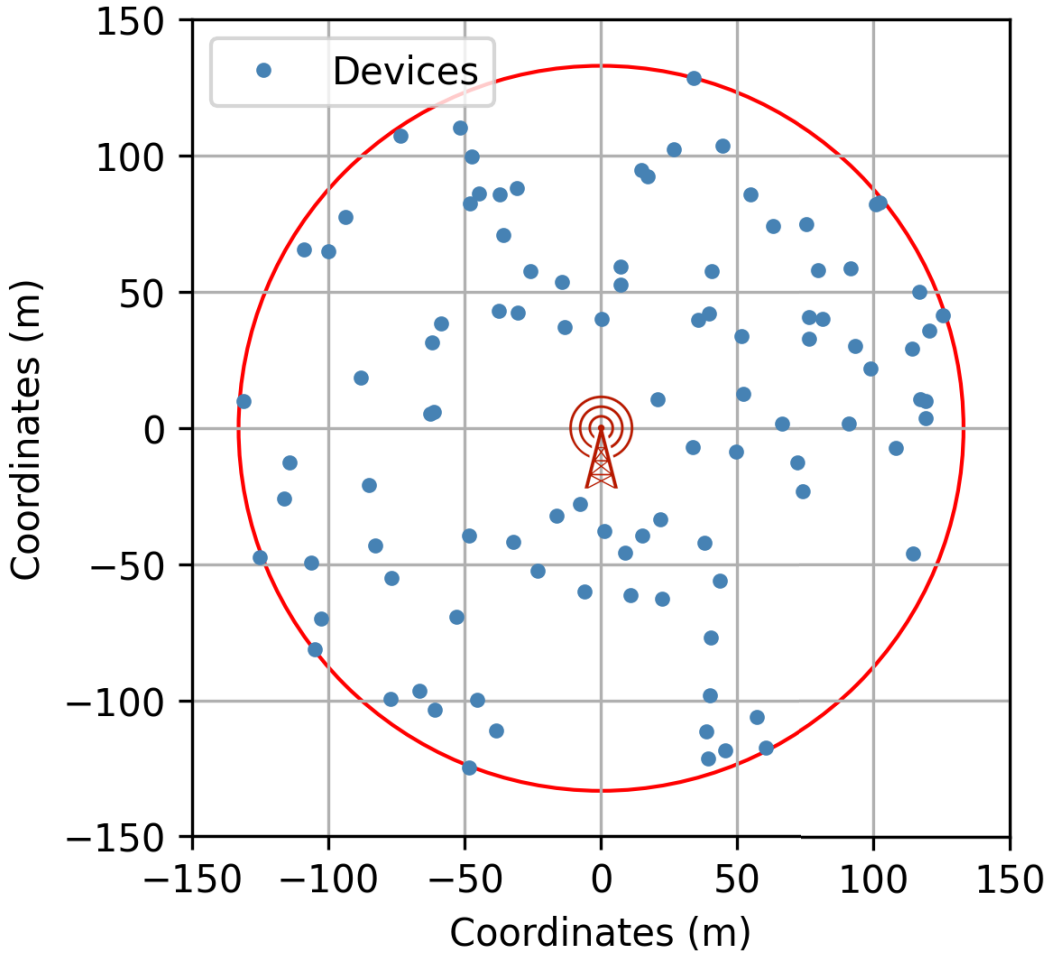


Figure 2 – Device disposition

ordered in decreasing received power from $m = 1$ to $m = M$. Then, the SINR for the m-th device after SIC becomes

$$\text{SINR}_{m,k} = \frac{P_{m,k}}{\sum_{j=m+1}^{M} P_{j,k} + \overline{P}_n}, \tag{3}$$

where $P_{m,k} = h_{m,k}^2 \overline{P}_{m,k}$ is the instantaneous received power from the m-th device in the k-th time slot, $h_{m,k}$ is Rayleigh fading, which is independent and identically distributed in time and space, while $\overline{P}_{m,k}$ is the average received power, which is modelled considering log-distance path loss (GOLDSMITH, 2005),

$$\overline{P}_{m,k} = \overline{P}_{m,k}(d_0) - 10\eta \log_{10}\left(\frac{d_{m,k}}{d_0}\right), \tag{4}$$

where $d_{m,k}$ is the distance from that device to the BS, d_0 is the reference distance, $\overline{P}_{m,k}(d_0)$ is calculated using the Friis equation, while η is the path loss exponent. Finally, $\overline{P}_n = FN_0B$ denotes the noise power, where N_0 is the noise power spectral density, B is the bandwidth, and F is the noise figure.

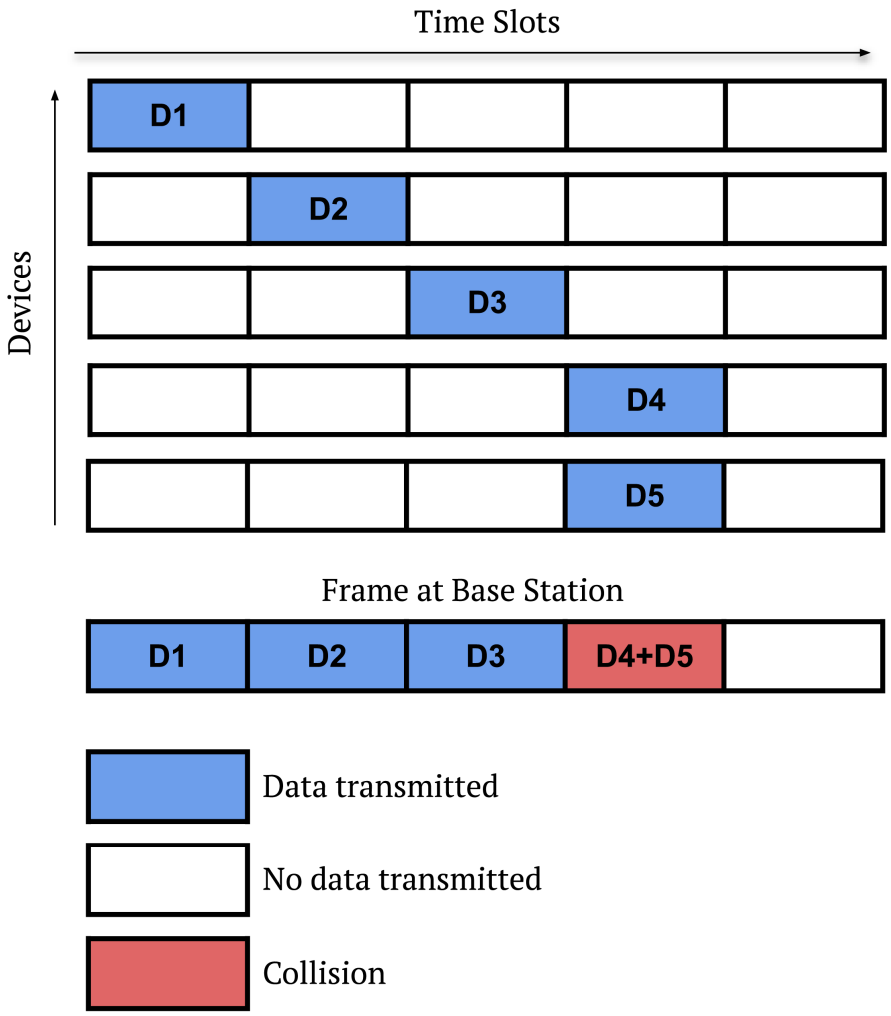


Figure 3 – Frame example

It is important to mention that (SHARMA; WANG, 2019a), which is the most relevant reference for this work as it is used as base for comparison, considers a hard collision model, in which the transmissions fail if a collision happens in a given time-slot, whatever the SINR is. The hard model limits the performance since, in many cases, it is possible to decode the strongest user in a collision (BJÖRNSON et al., 2017). Moreover, the introduction of the NOMA strategy above allows us to potentially decode all colliding users, greatly impacting the overall throughput. As mentioned previously, the devices can apply channel inversion to reach a certain average power at the BS. However, NOMA does not work well if devices yield the same power at the BS. In order to add the needed power diversity for NOMA to work properly, we let the devices deviate $\pm\Delta$ from a reference power, which in turn is calculated so that $P_{\text{ref}} - \Delta$ reaches a target outage probability. Meaning that each device's P_{ref} has its own value. Thus, devices have three options of transmit power: $P_{ref} - \Delta$, P_{ref}, and $P_{\text{ref}} + \Delta$. An appropriate Δ can increase NOMA efficiency as we no longer rely on the devices position to create the diversity for NOMA pairing. This can be observed on the 4, here we can clearly see

that there is no correlation between the device's position and it's chosen power level.

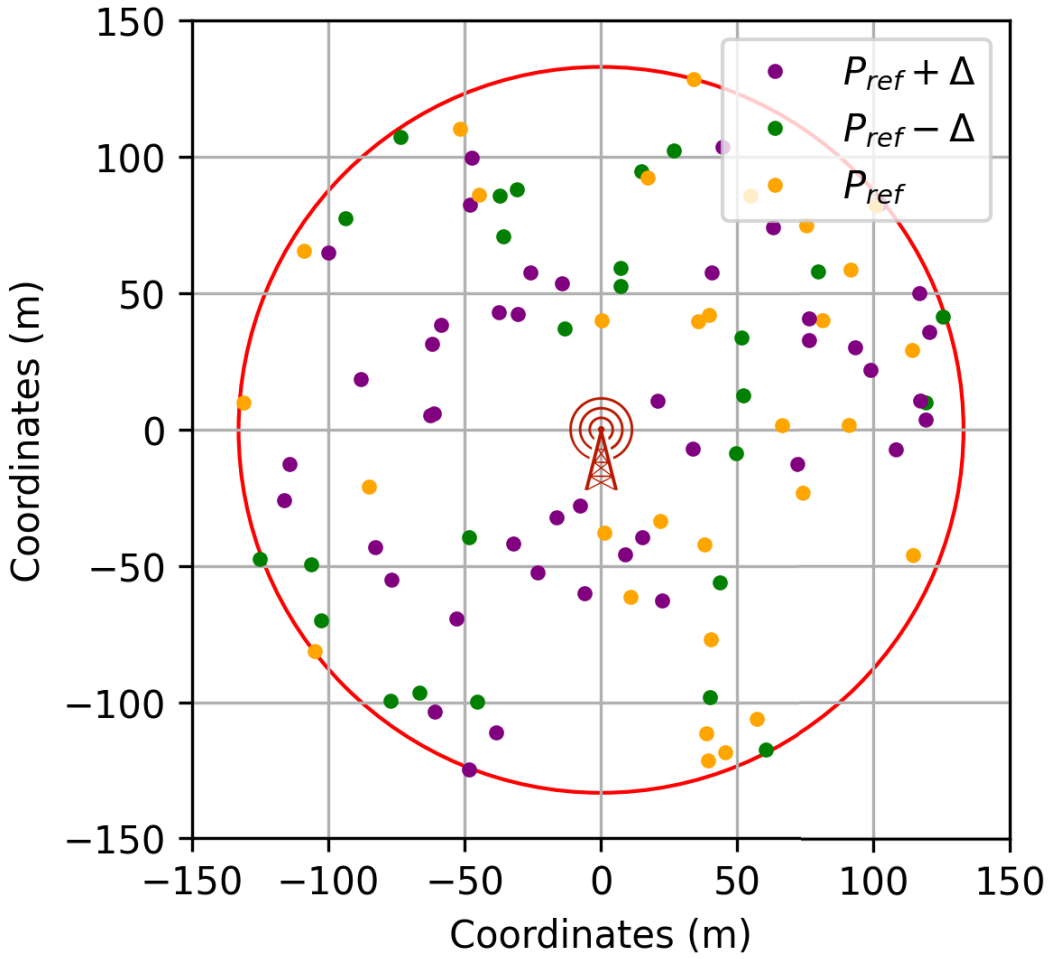


Figure 4 – Device disposition and Power level

4 PROPOSED METHOD

This work proposes a *Q*-Learning based method to optimize slot allocation taking advantage of NOMA spectral efficiency, making it possible for two or more devices to transmit at the same time-slot. This method allows the MTC devices to autonomously find NOMA partners and their dedicated time-slot while also preventing the excessive use of transmit power.

4.1 *Q*-LEARNING

The use of reinforcement learning has great potential in MTC networks (SHARMA; WANG, 2019a; BELLO et al., 2018; JIHUN MOON; YUJIN LIM, 2017; MOHAMMED et al., 2015; HAN et al., 2019), specially the widely adopted *Q*-Learning algorithm, because it is model-free and can be implemented in a distributed fashion. By modeling the RA in an MTC network as a Markov Decision Process (MDP) allows us to use *Q*-Learning. In an MDP the agent interacts with the environment in a sequential manner, selecting actions based on the state of the environment. The agent gets a reward based on its action and moves to the next state (SUTTON; BARTO, 2018).

The *Q*-Learning algorithm formulates this agent-environment relationship with an action-value function, the *Q*-table. At each time step u, while in a state S_u, an agent performs an action A_u trying to maximize its action-value function. The *Q*-value update rule can be defined as (SUTTON; BARTO, 2018)

$$Q(S_u, A_u) \leftarrow (1-\alpha)\, Q(S_u, A_u) + \alpha \left(R_{u+1} + \gamma \max_a Q(S_{u+1}, a) \right), \tag{5}$$

where $\alpha \in [0, 1]$ is the learning rate, $\gamma \in [0, 1]$ is the discount factor quantifying the importance of future rewards ($\gamma = 0$ values only immediate rewards while a higher γ would aim at a better long-term reward), and R is the reward.

We can apply the *Q*-Learning algorithm to the system model by considering that the agents are the MTC devices, the environment is the network, while the state-action pair is the combination of the transmit power and the time-slot, with every device having its own *Q*-Table. The simplest way to implement the *Q*-Learning algorithm is to apply a greedy policy, this way the device always chooses the time-slot and transmit power pair with the highest *Q*-value. Moreover, the greedy policy also presented the best results during the simulation campaign when compared to ϵ-greedy policies. In this work, similar to (SHARMA; WANG, 2019a), the reward is defined as the following:

$$R = \begin{cases} +1, \text{ successful transmission} \\ -1, \text{ failed transmission} \end{cases}. \tag{6}$$

The work in (SHARMA; WANG, 2019a) also proposes an alternative reward using a congestion level that improves the performance of the method proposed therein.

However, it requires the BS to detect how many devices collided in each time slot. The method in this paper requires only an acknowledgement bit per time slot, informing the success or not of the transmissions (irrespective of their number), which is much simpler in practice.

4.2 NOVEL RA METHOD: COMBINING *Q*-LEARNING AND NOMA

For each device, the *Q*-table for every possible (transmit power, time-slot) pair is randomly initialized following a uniform distribution between -1 and 1. This initialization adds an extra degree of randomness, improving throughput over an all 0's initialization. Then, the devices choose the (power, time-slot) pair with the highest *Q*-value. Next, the devices transmit their messages and the BS tries to recover them making use of SIC. At the end of the frame, the BS sends a single feedback message with one bit per time-slot, informing if the messages were successfully decoded or not. With this feedback each device updates its *Q*-value and proceeds to the next transmission. This process continues for several iterations (or frames), until it eventually converges[1].

Algorithm 1 SIC-based Distributed Q-learning RA Method

Require: *Q*-Table random initialized between -1 and 1
for Every frame **do**
  for Every device **do**
    Select the (power, time-slot) with the highest *Q* value
    if More than one slot with the highest value **then**
      Choose randomly among them
    end if
  end for
  BS uses SIC, (3), to recover the transmitted messages
  BS broadcasts feedback message
  for Every device **do**
    Update *Q*-value for (power, time-slot) pair using (5)
  end for
end for

The proposed method is summarized in **Algorithm 1**. Note that it adds minimal complexity at the device, requiring memory for storing one *Q*-Table with the number of power levels times the number of time-slots, in this case 3×100 slots, and the computational resources (calculation and memory) for (5). At the BS the increased complexity with respect to the method in (SHARMA; WANG, 2019a) is the SIC decoding, which is non-negligible. However, it is not unrealistic to assume that the BS has more processing power than the devices, while, it is very unlikely that many devices successfully share

[1] The convergence of *Q*-Learning is well known (SUTTON; BARTO, 2018; KAR et al., 2013), however the convergence of a multi-agent distributed *Q*-Learning needs further investigation, which is outside the scope of this work. Nevertheless, in our extensive simulation campaign the proposed method always converged.

the same time-slot, reducing the SIC complexity. Moreover, note that the *Q*-Learning implementation in **Algorithm 1** is distributed, each device updates its own *Q*-Table, which in turn influences their choice of (power, time-slot) in the next frame and the whole environment output. Implementing a centralized *Q*-Learning algorithm in the BS would be much more complex, requiring the BS to be aware of every device, storing all *Q*-Tables and making it more difficult to deploy new nodes. Also, implementing the *Q*-Learning at the BS would require extensive feedback as the BS would have to inform every device of its time-slot and power. Compared to related works that use *Q*-Learning, our method only requires the extra storage for the three power levels.

5 RESULTS

We investigate the performance of the proposed method by means of computer simulations, considering the setup in Chapter 3, with the parameters in Table 3, unless stated otherwise. In order to get the average behaviour, the curves presented here are the result of 30 simulation runs. The proposed method is compared to three schemes. The first two are: i) SA, and ii) SA with NOMA. In SA the devices randomly choose the time-slot within a frame, without any feedback from the BS. In SA with NOMA the BS applies SIC decoding in order to try to recover some of the colliding packets. The third method comes from (SHARMA; WANG, 2019a): iii) Collaborative *Q*-Learning. In this method *Q*-Learning is used to allocate devices to slots, as discussed in Chapter 3, but without NOMA. However, a different reward is employed, returning the congestion level of each time-slot, requiring the knowledge of how many devices collided in each time-slot. Note that all of the methods above do not use power control transmitting at a fixed P_t = 10 dBm. Finally the proposed method is presented using two different discount factors.

Table 3 – Simulation Parameters

Parameter	**Value**
Bandwidth B	100 kHz
Carrier frequency f_c	915 MHz
Cell radius	133 m
Path loss exponent η	3
Power Deviation Δ	7.78 dB
Noise figure F	6 dB
Noise PSD N_0	-174 dBm/Hz
Outage Probability	0.01
Transmit power P_t	10 dBm
Spectral efficiency r	2 bits/s/Hz
Reference distance d_0	1 m
Devices N	25-300
Messages L	100
Simulation Runs	30
Time-slots K	100
Learning rate α	0.1
Discount factor γ	0.5 and 1

First, we look at the throughput, the number of successful transmissions over the number of time-slots; a metric of how efficiently the frame is being utilized. It is important to note that this is a worst case scenario simulation where every device is transmitting in every frame. However, the proposed method does not require a transmission every frame. The device is free to move into sleep mode whenever necessary. Fig. 5 shows that the addition of NOMA considerably improves the throughput of SA. More-

over, SA with NOMA is able to outperform Collaborative *Q*-Learning from (SHARMA; WANG, 2019a) when the number of devices is relatively large. However, the proposed *Q*-Learning method with NOMA outperforms all the other strategies, while requiring a very reduced feedback (one bit per time slot), which is much simpler in practice than the reward used in Collaborative *Q*-Learning (SHARMA; WANG, 2019a). Note that the peak performance occurs when $N = K$ for Collaborative *Q*-Learning, but with the proposed method it is obtained for $N = 2.25K$. This behaviour can be attributed to the joint effort of NOMA and Power Control which increases spectral efficiency, allowing for more successful transmissions per time-slot. After $N = 225$ performance falls as we have more devices per slot and the fading becomes more relevant. Another interesting takeaway is that while the proposed method with $\gamma = 0.5$ performs slightly better when $N < 150$, when $\gamma = 1.0$ the performance is drastically better for $N > 150$. This can be due to the fact that γ takes future rewards into consideration. For a smaller N the devices are able to find their time-slot faster, making future rewards less important, while for a larger N the devices can take longer finding their pairs and slots making the role of γ crucial.

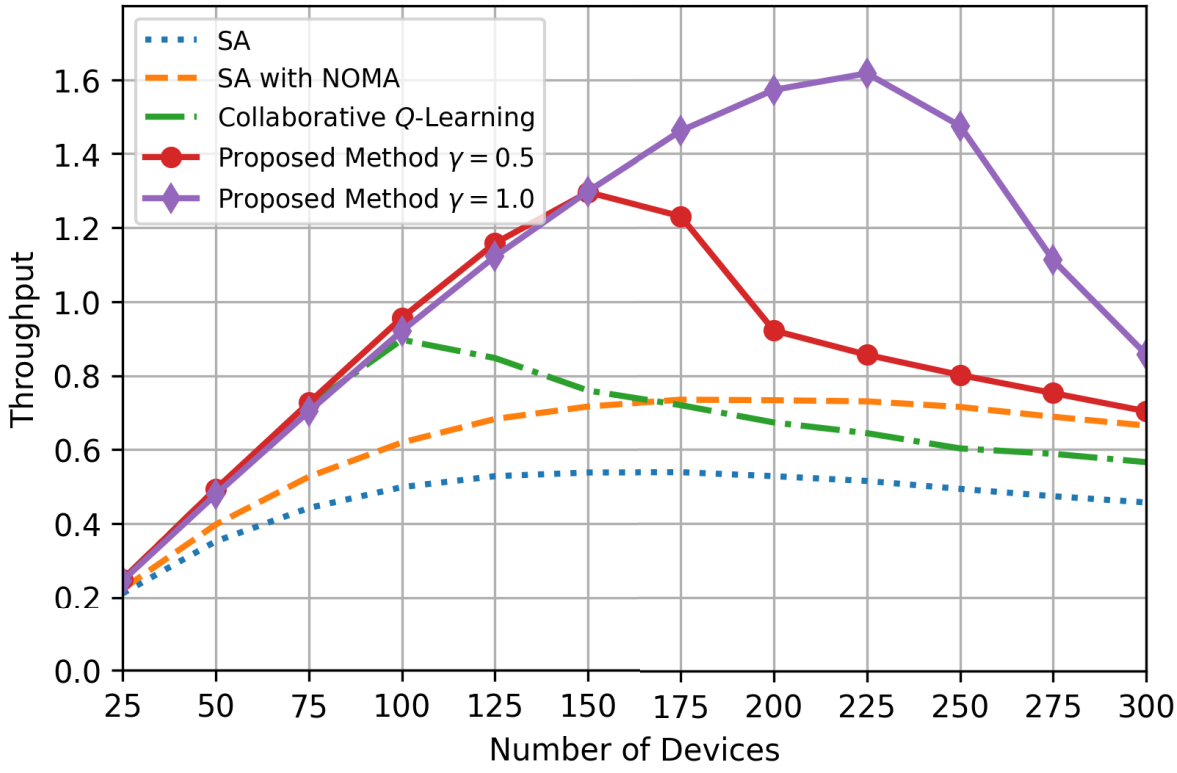


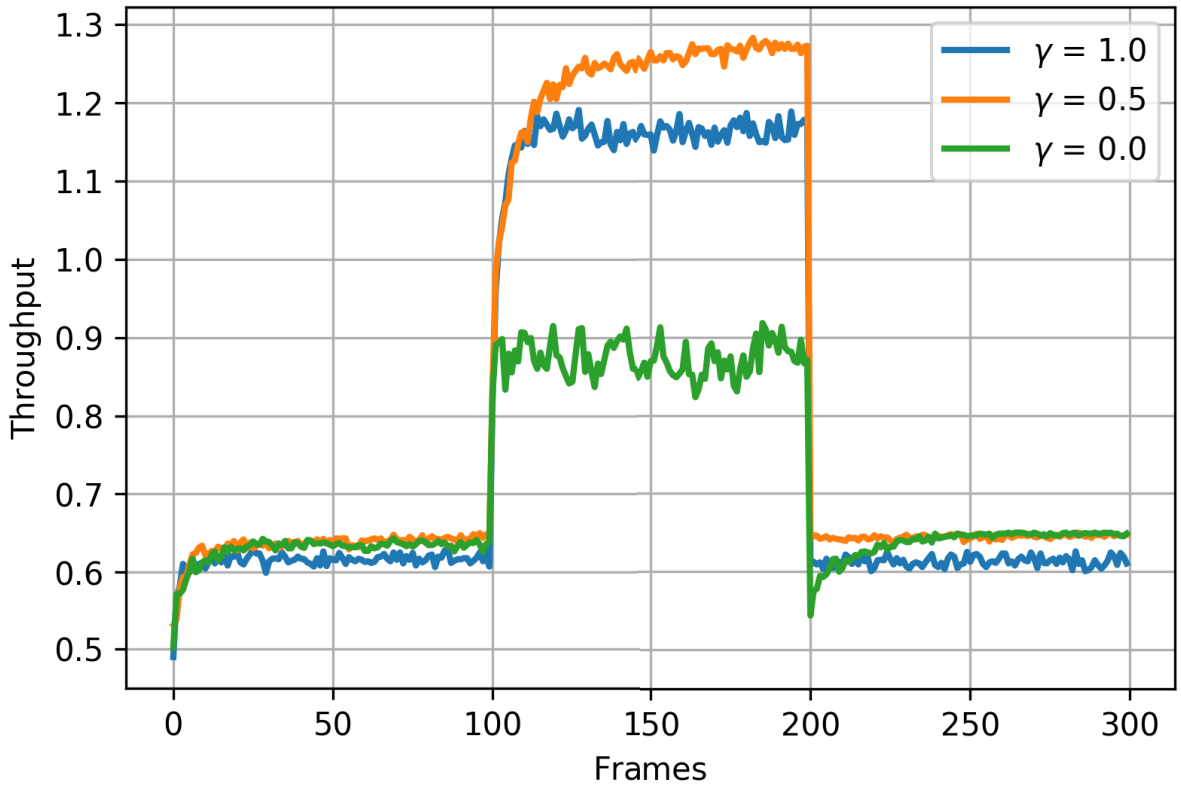
Figure 5 – Throughput versus number of devices for different RA methods and the proposed scheme.

In order to further evaluate the performance and convergence of the proposed method, next we consider a dynamic operation setup in three stages: i) first, $N/2$ devices send $L/2$ messages, ii) then, the second half of the devices join the network and

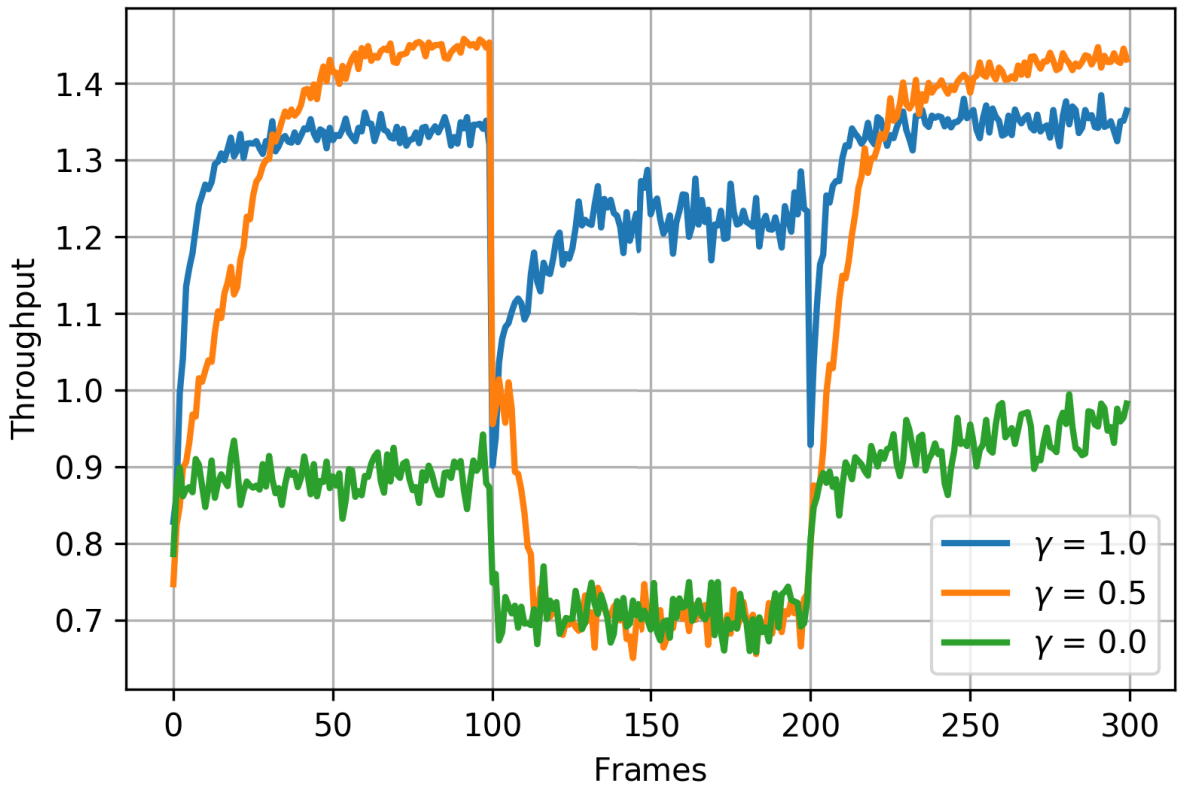
therefore in this stage N devices transmit $L/2$ messages, iii) lastly, in the third stage, as the first $N/2$ devices already transmitted their L messages, only the second half of $N/2$ devices transmit their final $L/2$ messages. Moreover, we consider two cases: $N = 130$ and $L = 200$, so that $N/2 < K$, and $N = 300$ and $L = 200$, so that $N/2 > K$. Finally, in such dynamic operation mode we can better investigate the effect of the discount factor γ in the performance of the proposed algorithm, so that we consider $\gamma \in \{0, 0.5, 1.0\}$.

As can be seen in Fig. 6-a, when $N/2 < K$, in the first stage the value of γ does not make a difference, a consequence of having more time-slots than transmitting devices. At the second stage the network is overloaded, with more devices than time-slots. In this case it is possible to see the positive effect of a larger γ, leading to faster convergence. Finally, at the third stage the network is under-loaded, so that throughput decreases but again the choice of γ does not impact significantly. In Fig. 6-b $N/2 > K$, so that the network already starts with more devices than available time-slots. In this situation the advantage of a large γ is evident, converging faster and to larger values of throughput for the three stages.

In order to better understand the behaviour of the curves in Fig. 6, recall that the discount factor γ prioritizes future rewards by softening the penalty when a collision happens, as the reward is added to the maximum Q-value weighted by γ. The curves in Fig. 6 provide a better insight on how the proposed method with $\gamma = 0.5$ is able to slightly outperform $\gamma = 1.0$ for a smaller quantity of devices. The positive effects of a smaller penalty when a collision happens can be noticed in the middle stage in Fig. 6-b, as the network suddenly becomes overloaded. A smaller γ can dismiss potential slots too quickly after collisions, making them unlikely to be utilized, resulting in a drop in the maximum system capacity and consequently in throughput, while a larger γ is able to recover the network faster and better allocate the devices resulting in a larger throughput. Moreover, we also investigated the impact of the learning rate α and found out that it is not very significant. Therefore, as $\alpha = 0.1$ has been used in (SHARMA; WANG, 2019a), we then used it in all methods.



(a) $N/2 < K$ with $N = 130$ and $K = 100$.

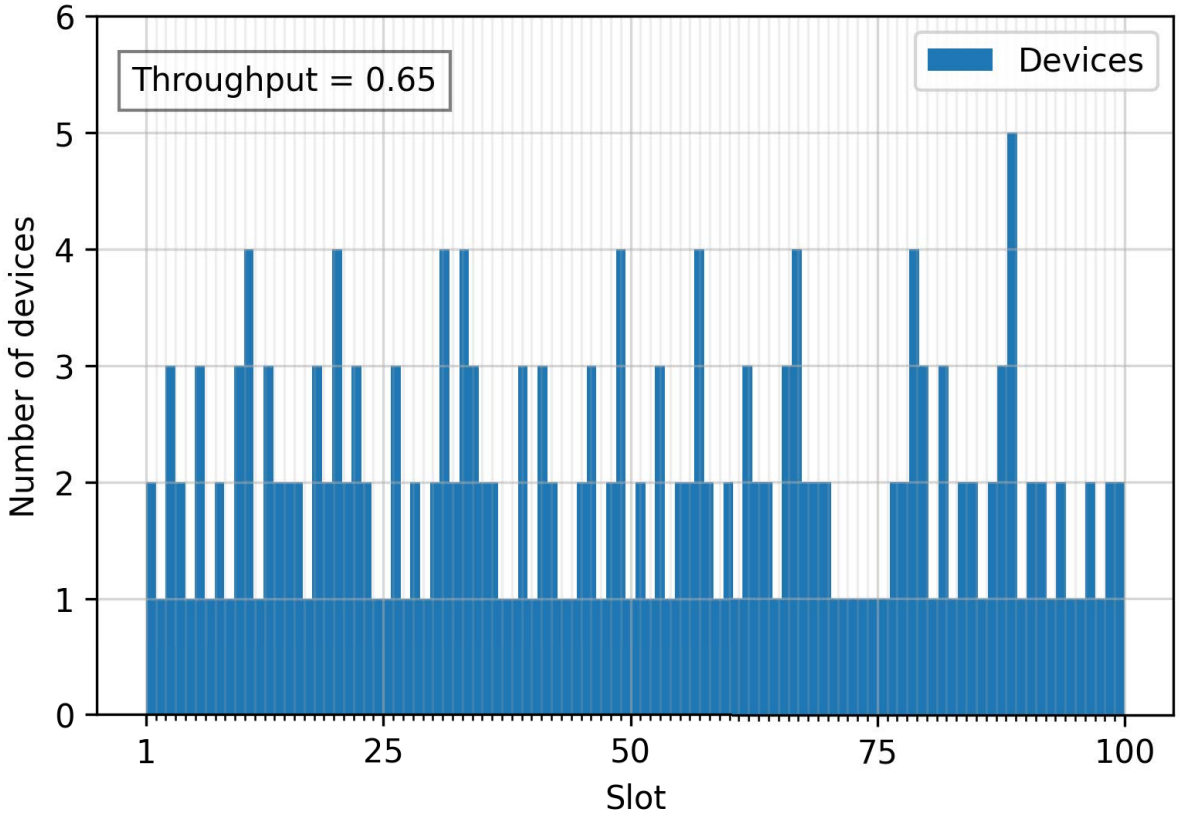


(b) $N/2 > K$ with $N = 300$ and $K = 100$.

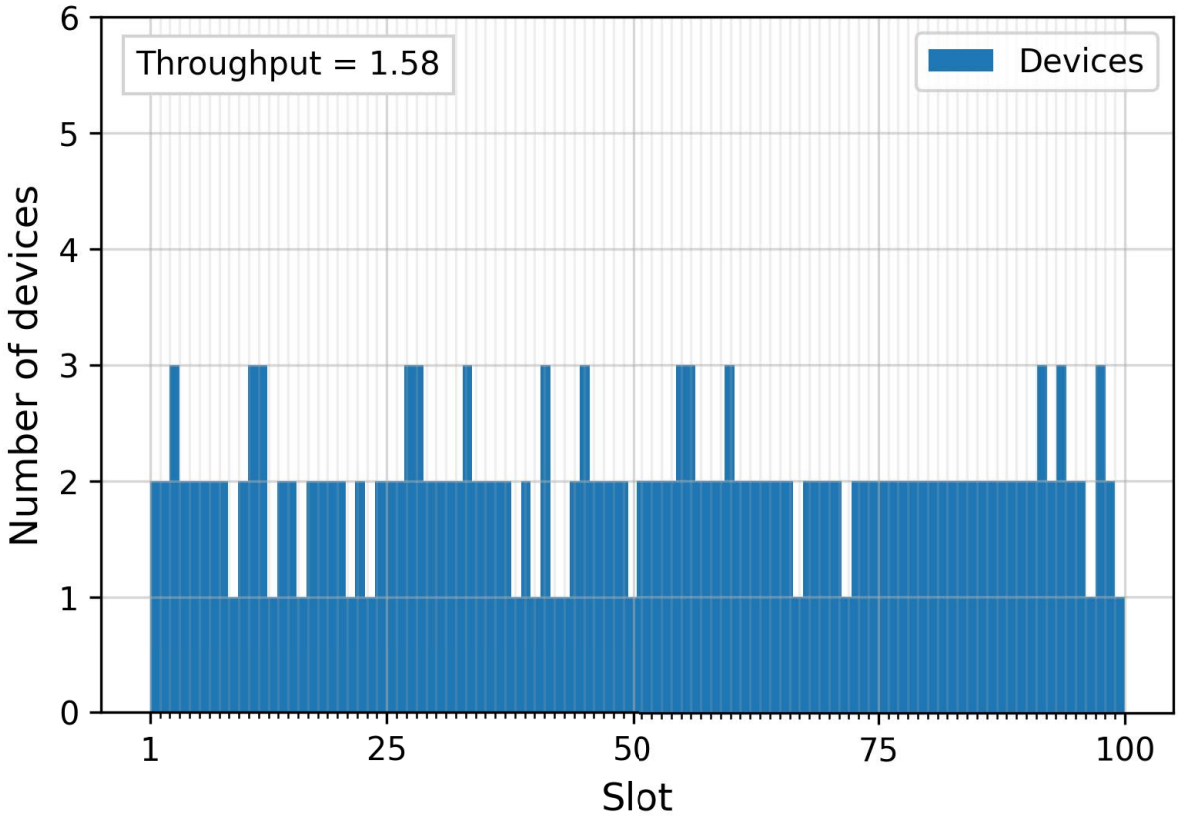
Figure 6 – Convergence analysis as a function of the discount factor *gamma*, for a dynamic network, with a varying number of nodes.

Next, we illustrate how the devices are allocated to time slots in two cases, Collaborative *Q*-Learning (SHARMA; WANG, 2019a) and the proposed method, considering $N = 200$, while showing only the allocation at the last frame. In Fig. 7-a we can see that, even though Collaborative *Q*-Learning succeeds in allocating at least one device to every time slot, this success is not reflected in the throughput as we have more devices than slots. Moreover, note that after allocating one device per slot the other devices are poorly distributed and interfere with each other, what could end up preventing successful transmissions.

Then, we have the proposed method in Fig. 7-b, where the algorithm is able to allocate devices exploiting all time-slots, while also taking advantage of NOMA. Note that almost every slot is allocated to two devices providing a good distribution and there are no more than 3 devices per slot. Therefore, the algorithm was able to distribute the resources in such a way that every device has the possibility of being decoded, considering the three available power levels.



(a) Collaborative *Q*-Learning



(b) Proposed Method

Figure 7 – Allocation of devices to time slots for the proposed method.

Moreover, the power control in the Q-learning based RA method leads to important power savings when compared to the case without power control. The average transmit power when using power control is well below the average when power control is deactivated, in which every device had the same transmit power as shown in Fig. 8. Note that the transmit power when there is no power control is used as an upper limit for the power control case, and the lower limit is the case where the closest device to the base station is allocated to transmit at $P_{ref} - \Delta$. It is important to note here that when NOMA is not enabled the average power raises along with the number of devices until $N = 125$. That is due to the fact that when SIC is not performed at the BS the devices learn that using a higher power will increase their chances of having a successful transmission. However, after the number of devices becomes significantly larger than the number of slots, this strategy no longer works and the devices become better distributed among the three power modes, lowering the overall transmit power.

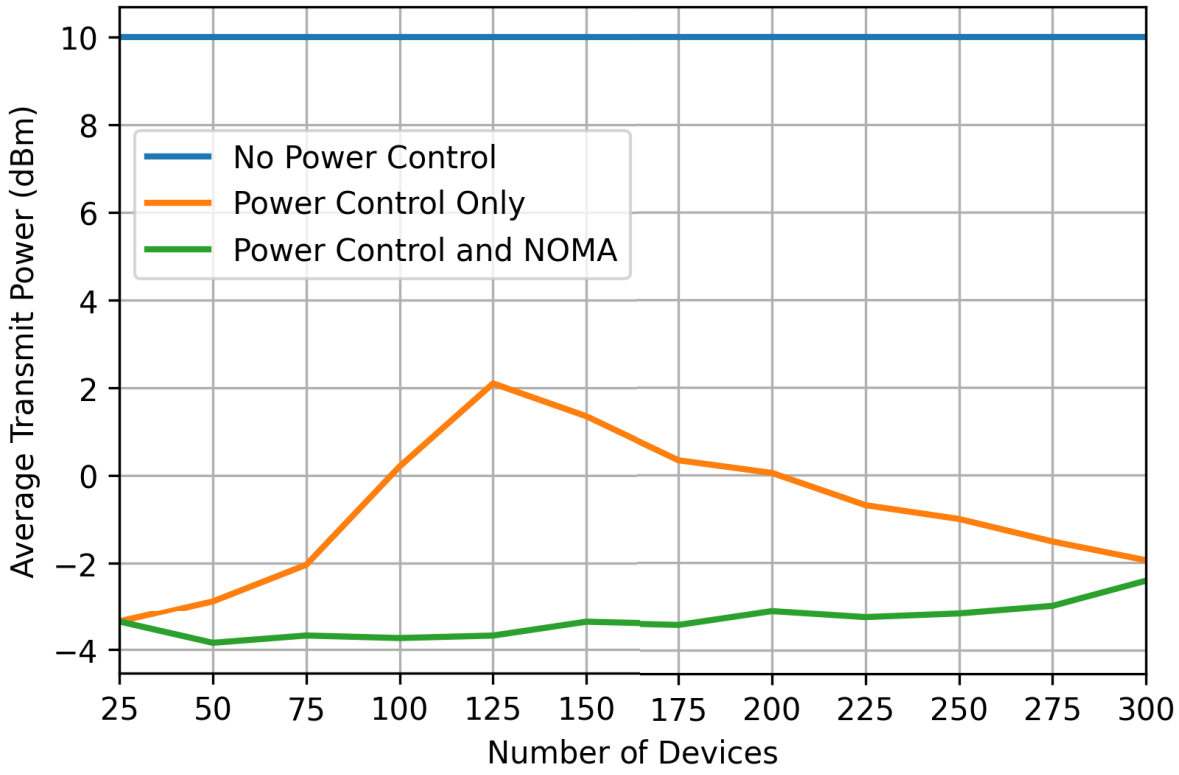


Figure 8 – Average Transmit Power versus number of devices

When working with *Q*-Learning, ϵ-greedy policy can often achieve better results, presenting a good trade-off between exploration and exploitation. However, it is not the case for our method and scenario as shown in Fig. 9. The main reason behind the utilization of a random search scheme is that a greedy search can oftentimes lead the algorithm to a sub-optimal solution. The simplicity of our method has to be kept in mind, as we only have the action of transmitting on a selected slot, the device has little to gain further exploring in search for other slots. Nevertheless, exploration happens as

devices can disrupt one another and fading could hinder the possibility of a successful transmission even in a slot where the devices already converged. For those reasons it is important to keep potential slots from being discarded. Even though our algorithm performs a greedy search it does not lack in exploration. As pointed out earlier, the discount factor used prevents the algorithm from converging to a sub-optimal solution.

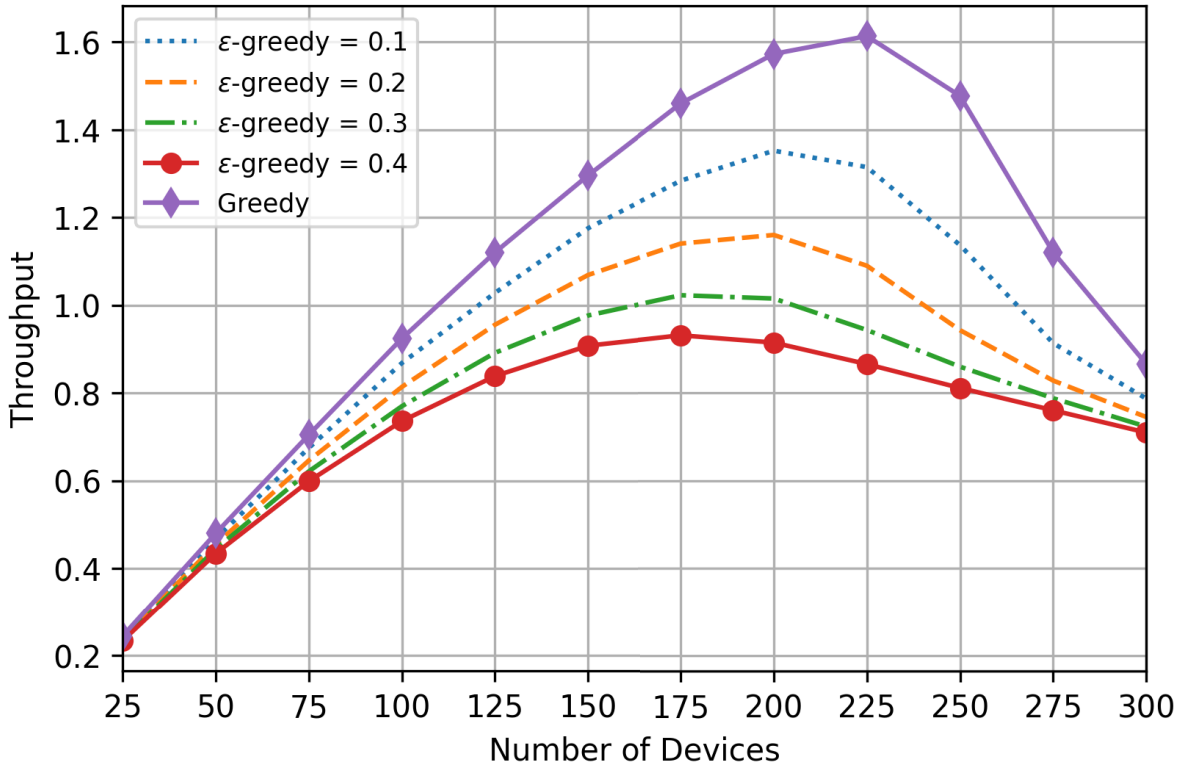


Figure 9 – Throughput versus number of devices for greedy and ϵ-greedy policies.

Furthermore, another key parameter in *Q*-Learning is the initialization scheme. Initializing the Q-Values to 0 is quite common in the related work. Taking into account that the end result for the Q-Values tend to be negative, initializing them to 0 is actually considered an optimistic initialization (SUTTON; BARTO, 2018) and can motivate exploration. Nevertheless, adding another degree of randomness can be beneficial in avoiding early collisions. That being said, the random initialization has been chosen because it slightly improves the overall throughput when the number of devices becomes very large as shown in Fig. 10.Note that the values from the random initialization are chosen from a uniform distribution over an [–1, 1] interval.

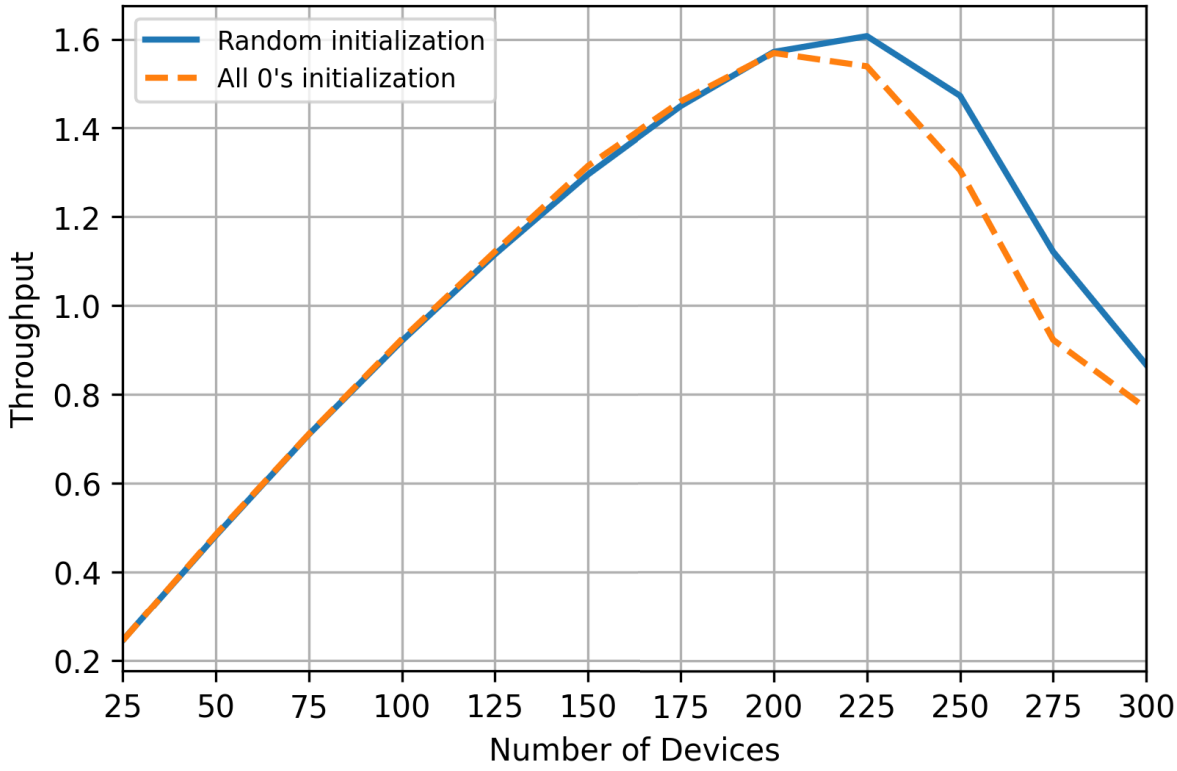


Figure 10 – Throughput versus number of devices for All 0's initialization and uniformly random initialization.

6 CONCLUSION

We introduced a novel method for RA combining the *Q*-Learning ability to measure uncertainties and NOMA spectral efficiency. The proposed method enables MTC devices to automatically choose their time-slots and transmit power for improving throughput. Moreover, the method requires minimal additional complexity at the device-side, as only a simple equation has to be implemented and $3 \times K$ numerical values are stored while also preventing the device from using an unnecessary amount of transmit power when compared to methods without power control. From the BS it requires very limited feedback, one bit per time slot. Simulations showed that using a larger discount factor presents the best performance when operating with a large number of devices, converging faster for a higher throughput and better handling network overload in a dynamic scenario. Furthermore, the proposed method provides significant gains in performance over other solutions.

6.1 FUTURE WORKS

Following the work done in this dissertation we can expand this research in a variety of ways, with the two main topics being the application of machine learning to communication systems and further investigating the utilization of NOMA through SIC in different models. The work done in (ZHOU et al., 2020) also uses *Q*-Learning to improve throughput. But unlike our work it focuses on assigning secondary users to subcarrier bands. However, the most interesting idea proposed by the authors that we could adapt and investigate its benefits to our model is that the devices are able to exchange information before certain frames and update the whole Q-Table. This could improve our model reducing collisions and providing an earlier convergence. Of course, this exchange of information means a trade-off between reducing collisions and the computational cost and would have to be thoroughly investigated. Besides that, the addition of NOMA along the lines of our model could also be an interesting research topic.

Another step in the direction of machine learning and random access would be looking into implementing Deep *Q*-Learning to our model. The authors of (ZHANG et al., 2020) have proposed a Deep Reinforcement Learning grant-free NOMA algorithm looking to improve throughput. This would however greatly increase the complexity and computational cost of our method. Perhaps low-end devices can not handle the storage and computational demands of neural networks. This would have to be deeply investigated.

When looking at the other side of this work we have yet another interesting path forward. Adapting our system model with conjunction of the fast model proposed by (HAJIZADEH et al., 2019) looking to add the capture effect and SIC. Furthermore,

the method here proposed can be adapted to a myriad of different system models to analyse the impact and interaction of NOMA and *Q*-Learning.

REFERENCES

3GPP. **(Release 16). Technical Specification Group Services and System Aspects**. [S.l.], Mar. 2020. Available from: www.3gpp.org/ftp/Specs/archive/21%5C_series/21.916.

5G AMERICAS. **The 5G Evolution: 3GPP Releases 16-17**. [S.l.], Jan. 2020. Available from: www.5gamericas.org/wp-content/uploads/2020/01/5G-Evolution-3GPP-R16-R17-FINAL.pdf.

AAZHANG, Behnaam et al. **Key drivers and research challenges for 6G ubiquitous wireless intelligence (white paper)**. [S.l.: s.n.], 2019. ISBN 978-952-62-2353-7. Available from: http://urn.fi/urn:isbn:9789526223544.

BELLO, L. M.; MITCHELL, P. D.; GRACE, D. Intelligent RACH Access Techniques to Support M2M Traffic in Cellular Networks. **IEEE Transactions on Vehicular Technology**, v. 67, n. 9, p. 8905–8918, Sept. 2018. ISSN 1939-9359. DOI: 10.1109/TVT.2018.2852952.

BI, Q. Ten Trends in the Cellular Industry and an Outlook on 6G. **IEEE Communications Magazine**, v. 57, n. 12, p. 31–36, Dec. 2019. ISSN 1558-1896. DOI: 10.1109/MCOM.001.1900315.

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 Transactions on Wireless Communications**, v. 16, n. 4, p. 2220–2234, Apr. 2017. ISSN 1558-2248. DOI: 10.1109/TWC.2017.2660489.

BOCKELMANN, C. et al. Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks. **IEEE Access**, v. 6, p. 28969–28992, 2018.

CISCO. **Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022**. [S.l.], Feb. 2019.

CLAZZER, Federico. From 5G to 6G: Has the Time for Modern Random Access Come? In: 6G Wireless Summit Event Levi Finland. [S.l.: s.n.], 2019.

CONTIKI-OS. [S.l.: s.n.]. Available from: http://www.contiki-os.org/. Visited on: 24 Nov. 2020.

GOLDSMITH, Andrea. **Wireless Communications**. USA: Cambridge University Press, 2005. ISBN 0521837162.

HAJIZADEH, H.; NABI, M.; TAVAKOLI, R.; GOOSSENS, K. A Scalable and Fast Model for Performance Analysis of IEEE 802.15.4 TSCH Networks. In: 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). [S.l.: s.n.], 2019. P. 1–7.

HAN, S.; XU, X.; LIU, Z.; XIAO, P.; MOESSNER, K.; TAO, X.; ZHANG, P. Energy-Efficient Short Packet Communications for Uplink NOMA-Based Massive MTC Networks. **IEEE Transactions on Vehicular Technology**, v. 68, n. 12, p. 12066–12078, Dec. 2019. ISSN 1939-9359. DOI: 10.1109/TVT.2019.2948761.

JIHUN MOON; YUJIN LIM. Access control of MTC devices using reinforcement learning approach. In: 2017 International Conference on Information Networking (ICOIN). [S.l.: s.n.], Jan. 2017. P. 641–643. DOI: 10.1109/ICOIN.2017.7899576.

KAR, S.; MOURA, J. M. F.; POOR, H. V. Distributed reinforcement learning in multi-agent networks. In: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). [S.l.: s.n.], Dec. 2013. P. 296–299. DOI: 10.1109/CAMSAP.2013.6714066.

LETHABY, Nick. **Wireless connectivity for the Internet of Things: One size does not fit all**. [S.l.], Oct. 2017. Available from: https://www.ti.com/lit/wp/swry010a/swry010a.pdf.

LORA. **What is LoRaWAN**. [S.l.], Mar. 2015.

LORA Alliance. [S.l.: s.n.]. Available from: https://lora-alliance.org/. Visited on: 24 Nov. 2020.

MA, X.; LUO, W. The Analysis of 6LowPAN Technology. In: 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application. [S.l.: s.n.], 2008. P. 963–966.

MAHMOOD, Nurul H.; LÓPEZ, Onel; PARK, O. **White Paper on Critical and Massive Machine Type Communications Towards 6G**. [S.l.], June 2020.

MARY, P.; GORCE, J.; UNSAL, A.; POOR, H. V. Finite Blocklength Information Theory: What Is the Practical Impact on Wireless Communications? In: 2016 IEEE Globecom Workshops (GC Wkshps). [S.l.: s.n.], 2016. P. 1–6.

MEKKI, Kais; BAJIC, Eddy; CHAXEL, Frederic; MEYER, Fernand. A comparative study of LPWAN technologies for large-scale IoT deployment. **ICT Express**, v. 5, n. 1, p. 1–7, 2019. ISSN 2405-9595. DOI: https://doi.org/10.1016/j.icte.2017.12.005. Available from: http://www.sciencedirect.com/science/article/pii/S2405959517302953.

MOHAMMED, A. H.; KHWAJA, A. S.; ANPALAGAN, A.; WOUNGANG, I. Base Station Selection in M2M Communication Using Q-Learning Algorithm in LTE-A Networks. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications. [S.l.: s.n.], Mar. 2015. P. 17–22. DOI: 10.1109/AINA.2015.160.

OSSEIRAN, A.; PARKVALL, S.; PERSSON, P.; ZAIDI, A.; MAGNUSSON, S.; BALACHANDRAN, K. **5G Wireless Access: An Overview**. [S.l.], Apr. 2020.

P. THUBERT, Ed.; BORMANN, C.; TOUTAIN, L.; CRAGIE, R. **IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN)**. [S.l.], Apr. 2017. Available from: https://tools.ietf.org/html/rfc8138.

PALATTELLA, M. R.; ACCETTURA, N.; VILAJOSANA, X.; WATTEYNE, T.; GRIECO, L. A.; BOGGIA, G.; DOHLER, M. Standardized Protocol Stack for the Internet of (Important) Things. **IEEE Communications Surveys Tutorials**, v. 15, n. 3, p. 1389–1406, 2013.

POPOVSKI, P.; STEFANOVIĆ, Č.; NIELSEN, J. J.; DE CARVALHO, E.; ANGJELICHINOSKI, M.; TRILLINGSGAARD, K. F.; BANA, A. Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC). **IEEE Transactions on Communications**, v. 67, n. 8, p. 5783–5801, 2019.

SAITO, Y.; KISHIYAMA, Y.; BENJEBBOUR, A.; NAKAMURA, T.; LI, A.; HIGUCHI, K. Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In: IEEE Vehicular Technology Conference (VTC). [S.l.: s.n.], 2013. P. 1–5. DOI: 10.1109/VTCSpring.2013.6692652.

SHARMA, S. K.; WANG, X. Collaborative Distributed Q-Learning for RACH Congestion Minimization in Cellular IoT Networks. **IEEE Communications Letters**, v. 23, n. 4, p. 600–603, Apr. 2019a. ISSN 2373-7891. DOI: 10.1109/LCOMM.2019.2896929.

SHARMA, S. K.; WANG, X. Towards Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions. **IEEE Communications Surveys Tutorials**, p. 1–1, 2019b. ISSN 2373-745X. DOI: 10.1109/COMST.2019.2916177.

SIGFOX. [S.l.: s.n.], June 2020. Available from: https://www.sigfox.com/en.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement learning: an introduction**. [S.l.]: The MIT Press, 2018.

T-MOBILE. **The Game Changer for the Internet of Things**. [S.l.], Mar. 2019.

THREAD. Thread. Available from: https://threadgroup.org/. Visited on: 24 Nov. 2020.

TULLBERG, H.; POPOVSKI, P.; LI, Z.; UUSITALO, M. A.; HOGLUND, A.; BULAKCI, O.; FALLGREN, M.; MONSERRAT, J. F. The METIS 5G System Concept: Meeting the 5G Requirements. v. 54, n. 12, p. 132–139, Dec. 2016. ISSN 0163-6804. DOI: 10.1109/MCOM.2016.1500799CM.

WI-FI Alliance. [S.l.: s.n.]. Available from: https://www.wi-fi.org/. Visited on: 24 Nov. 2020.

ZAIDI, A.; HOGAN, M.; KUHLINS, C. **Cellular IoT Evolution for Industry Digitalization**. [S.l.], Jan. 2019.

ZHANG, J.; TAO, X.; WU, H.; ZHANG, N.; ZHANG, X. Deep Reinforcement Learning for Throughput Improvement of the Uplink Grant-Free NOMA System. **IEEE Internet of Things Journal**, v. 7, n. 7, p. 6369–6379, 2020.

ZHAO, L.; XU, X.; ZHU, K.; HAN, S.; TAO, X. QoS-based Dynamic Allocation and Adaptive ACB Mechanism for RAN Overload Avoidance in MTC. In: 2018 IEEE Global Communications Conference. [S.l.: s.n.], 2018. P. 1–6.

ZHOU, Y.; ZHOU, F.; WU, Y.; HU, R. Q.; WANG, Y. Subcarrier Assignment Schemes Based on Q-Learning in Wideband Cognitive Radio Networks. **IEEE Transactions on Vehicular Technology**, v. 69, n. 1, p. 1168–1172, 2020.

ZIGBEE Alliance. [S.l.: s.n.], June 2020. Available from: `https://zigbeealliance.org/`.