

Uncoordinated Spectrum Access for Massive Connectivity in IoT Applications

Joseph Doumit, Marie-Josepha Youssef, Charbel Abdel Nour, Joumana Farah, Catherine Douillard

Abstract

Boosted by the Internet of Things (IoT), future cellular communication networks will be able to support new services with high social impact. Corresponding applications are expected to span almost every domain, especially the most critical ones, such as the medical field or smart factories. However, their adoption is conditioned by strict compliance with stringent quality of service (QoS) requirements for the underlying network. In particular, the achieved throughput and latency are of utmost importance for efficient monitoring with multiple sensors in a variety of application areas. In this paper, we propose an intelligent self-configurable system structure, achieved through resource allocation techniques using machine learning, and capable of flexibly fulfilling such requirements. In addition, special care was taken to minimize the transmit power of the IoT devices while efficiently supporting different application-dependent traffic types, an essential prerequisite for any wide adoption of such systems with massive network access.

INTRODUCTION

Internet of Things (IoT) devices are expected to account for the largest part of mobile traffic in the fifth generation (5G) and beyond 5G (B5G) networks [1]. Following a trend already started with 5G, the sixth-generation (6G) is expected to go beyond people-to-people communication to connect devices, sensors, vehicles, computing resources, and robotic agents [2]. Therefore, 6G will require technologies enabling the support of a massive deployment of IoT devices with different requirements in terms of data rate, latency, and reliability. To do so, various techniques can be smartly combined to pave the way for scalable and self-configurable massive IoT systems.

To increase system capacity for massive systems, the concept of non-orthogonal multiple access (NOMA) has been introduced to support more devices than the number of available orthogonal resources in time and frequency. NOMA allows multiple devices to share the same time and frequency resources via power domain or code domain multiplexing [3]. Regardless of the adopted NOMA technique, it increases the receiver complexity, required to recover the non-orthogonal signals. In this paper, we focus on power-domain NOMA in the uplink and will be denoted by NOMA in the following for brevity.

In addition to NOMA, relaying can be useful in massive communication settings. Indeed, it can reduce network congestion at the base station (BS) level, extend cellular coverage and increase achieved rates thanks to shorter point-to-point transmissions. Mobile relays that can be deployed quickly and efficiently, such as drones, can also restore communications in damaged network infrastructures [4]. To achieve high data rates, reduce latency and increase spectral efficiency, full-duplex relays, simultaneously transmitting to the BS and receiving from IoT devices on the same frequency resource,

can be implemented. However, full-duplex relays suffer from self-interference, which must be mitigated [4].

On a different note, the development of scalable networks having variable numbers of devices and traffic types is essential for unleashing the potential of IoT. Uncoordinated or grant-free spectrum access strategies can be suited to this context. Indeed, since such strategies do not require any coordination at the level of the nodes in the network, signaling overhead and communication latency can be significantly reduced, compared to conventional access schemes [5]. Moreover, to reduce planning and maintenance costs, such networks can be equipped with self-organizing capabilities, enabling efficient resource sharing through power and spectrum allocation in a distributed manner. To that end, techniques such as reinforcement learning (RL) [6] can be particularly appealing.

Hence, by using NOMA in a grant-free communication setting with relays, systems supporting scalable and self-configurable massive access can be efficiently devised. However, such complex systems need to be thoroughly designed in order to meet the quality of service (QoS) requirements for IoT applications. The design of such a system represents the main focus of this work.

Prerequisites and System Components

1) *NOMA*: For an efficient uplink NOMA, the multiplexed devices are allocated different power levels according to their channel conditions and are directly superimposed after classic channel coding and modulation [7]. Figure 1 illustrates the principle of NOMA for the uplink. At the receiver, multi-user detection algorithms such as successive interference cancellation (SIC) are used to cancel multi-user interference and separate the different signals. SIC exploits the signal-to-interference-plus-noise ratio (SINR) difference between users to recover the transmitted signals. To ensure successful SIC, the signals of devices transmitting on the same channel must be received with different power levels. Therefore, if two or more devices transmitting on the same channel choose the same power level, a collision occurs, leading to a failed signal

Joseph Doumit, Charbel Abdel Nour and Catherine Douillard are with IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France. Marie-Josepha Youssef is with Capital Fund Management, Paris, France. Joumana Farah is with INSA Rennes, France.

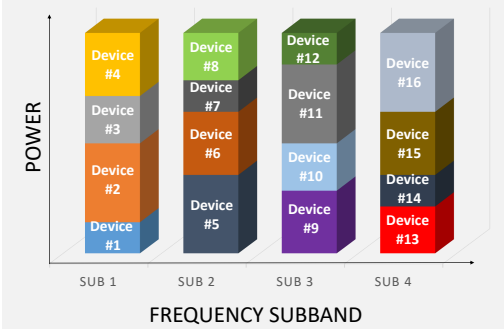


Fig. 1: Illustration example of uplink NOMA principle with four power levels per subband. The devices' signals are superimposed in the power domain on the frequency resources at each time slot. Each device is associated with a power level on one subband and the height of each box represents the value of the received power.

decoding [8]. In the case of a collision, devices with signals having weaker received power levels than the colliding signals cannot be recovered either. These penalizing repercussions can impact upper-level signals if their target SINR is not met. Therefore, signal collisions should be minimized.

2) *RL*: RL is a framework where one or multiple learning agents interact with an environment in order to achieve a certain goal, usually represented by the maximization of a numerical reward. Knowing the environment status, the learning agent chooses its action based on a given strategy. Depending on the chosen action, it receives a reward and moves to a new state. This cycle is then repeated. That said, actions that achieve the highest rewards are not known beforehand. This motivates a two-phase approach consisting of exploration and exploitation phases [6]. The exploration phase intends to progressively adapt and define a rewards mapping based on the actions undertaken. This mapping depends on the average of previous rewards obtained during a sliding window and the strategy of the agent. Then, the exploitation phase is the phase where the agent continuously chooses the action having the highest reward for future plays. Many studies resort to the multi-armed bandits (MAB), a RL framework applied in a single-state environment [9], to solve the problem of channel allocation in ad-hoc networks and treat the exploration and exploitation phases separately to learn the channels [10] and to manage the strategy for edge users [11].

3) *Relaying and RL*: One of the main drawbacks of the MAB framework resides in the necessity to perform an exploration phase for each variation of one of the essential system parameters, e.g., the number of active devices at the relay level or the transmission conditions including the channel model. To circumvent the need for an exploration phase when the number of active devices varies, we propose to introduce relays equipped with queuing capabilities to help prevent or at least reduce data loss in case of system congestion, hence efficiently managing sporadic traffic. This can be of particular interest since it

enables a higher level of flexibility and scalability for the MAB framework in managing the underlying network congestion. However, since it requires memory buffers, queuing tends to increase the transmission link latency, which can considerably hinder the performance of some IoT applications. To this effect, we propose to introduce specific latency-related design constraints for the scheduling algorithm deployed at the relays. To evaluate the efficiency of the proposed algorithms, the age of information (AoI) is investigated at the receivers. AoI is a metric used to measure the information freshness when it reaches the destination [12]. It is defined as the time elapsed since the latest useful piece of information, that reached its intended destination, has been generated at its source. AoI is particularly relevant in applications in which information is time-varying and where the timeliness of information is crucial, such as vehicular monitoring systems or industrial sensor networks [13]. The target framework accounts for the AoI constraints, depending on the type of considered traffic.

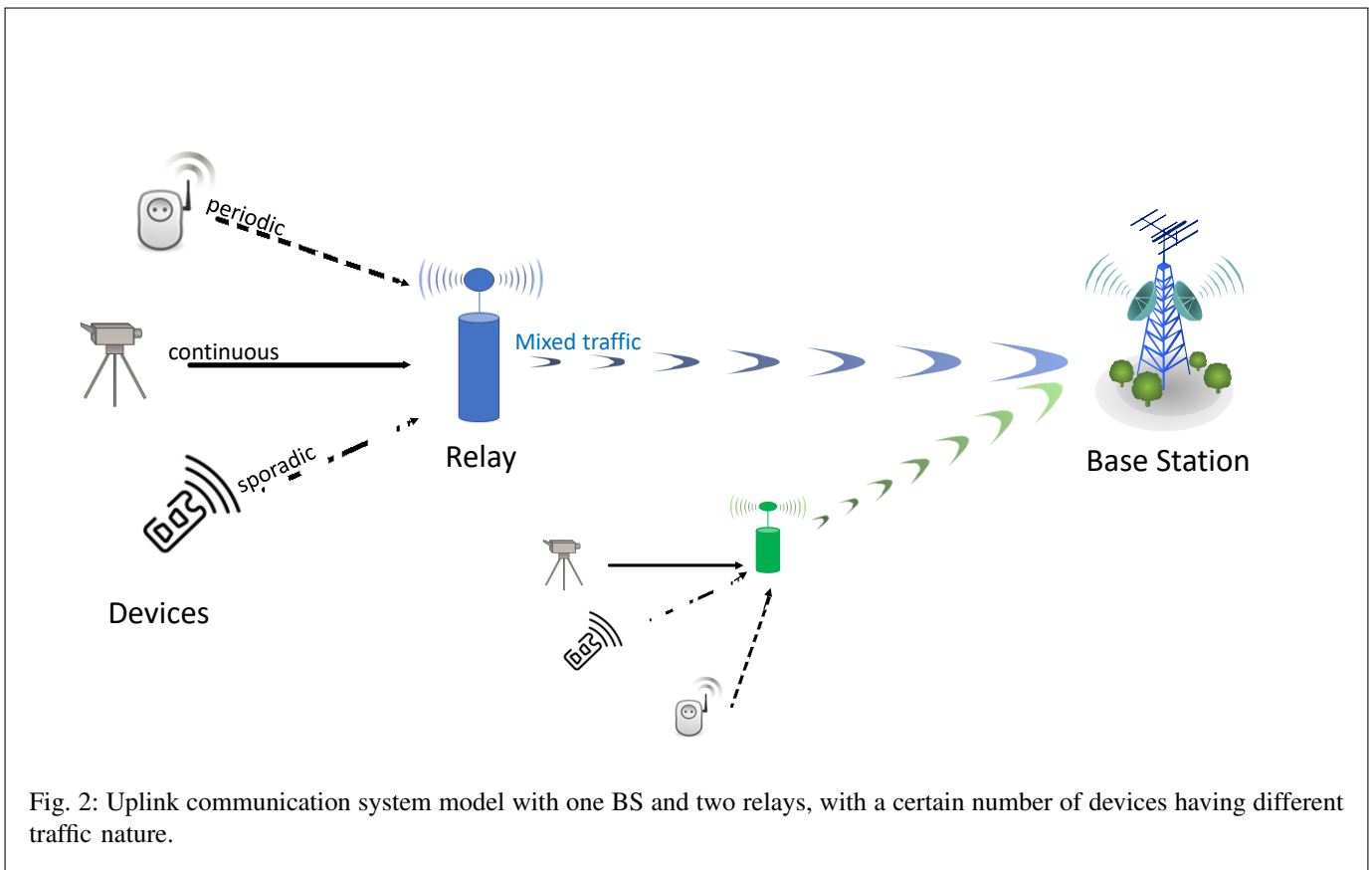
Contributions

Previous research has not considered the utilization of full-duplex relays in an uncoordinated uplink NOMA system with transmit power limitations and AoI restrictions for devices with heterogeneous traffic types. Thus, in this work, we investigate such a system. The focus is on enabling IoT devices to autonomously organize their communications with available relays in an uncoordinated spectrum access environment. The system includes a mixture of traffic types, including continuous, sporadic, and periodic transmissions. Furthermore, each IoT device operates under a limited power budget. We propose a multi-step algorithm for uplink massive IoT communication systems capable of supporting different traffic types while leveraging new intelligent self-configurable network architectures. The main contribution of this article concerns the proposal of efficient methods to:

- enable a flexible and scalable communication network that handles devices having multiple traffic types with latency and power constraints while reducing the data loss and the congestion at the level of the relays in a distributed communication system,
- allocate the available spectrum and power resources to the devices in an uncoordinated manner while maximizing achieved rates and minimizing transmit power,
- reduce the AoI for the IoT devices having latency constraints by using the MAB framework.

SYSTEM MODEL

In the following, we consider an uplink communication system with a BS, full-duplex relays uniformly deployed over the cell and a certain number of IoT devices, as shown in Fig. 2. Depending on the targeted application, the IoT devices generate different traffic types: continuous, periodic and sporadic. Devices with continuous traffic type transmit their signals non-stop, such as cameras for video surveillance applications. Devices with a non-continuous traffic type are mainly sensors. A periodic traffic type is defined by the periodicity and the duration of the signal transmissions while



devices with sporadic traffic are assigned a probability of access to the spectrum. Moreover, it is assumed that the totality of a sporadic-type message can be transmitted during one time slot. The relevance of these three types of traffic can be illustrated with an application example in an e-health context: the remote monitoring of a medical act. One can imagine that the intervention is filmed by cameras that send a continuous stream, whereas specific sensors send periodic information about the patient (e.g., blood pressure) and vital sign changes are remotely managed by sensors sending sporadic control signals.

We also consider that, due to inconvenient radio conditions between the BS and the IoT devices with respect to the available transmit power, the devices cannot transmit their signals directly to the BS. Hence relays are introduced to forward the messages to the BS. A two-hop communication network is, therefore, set up with two types of links: links from IoT devices to relays (relay uplinks) and links from relays to BS (backhaul links). We assume that all transmission links undergo large-scale fading, due to path loss and shadowing, and the corresponding channel gain is assumed to be perfectly estimated for all links.

As mentioned earlier, spectrum access is fully uncoordinated: IoT devices must therefore arrange communications without exchanging information with each other and without any coordination at the level of the BS or the relays. To do so, each IoT device is allowed to communicate with one relay over one frequency subband.

For relay uplinks, multiple access is provided through NOMA. At the relay, a limited number of received power levels is allowed and SIC decoding is performed in the decreasing order of power levels. The transmitting devices are assumed to have an exact knowledge of their channel gains with the relays as well as the received power levels at the relays in order to control their transmit power level. The number of received power levels at the relays is the result of a trade-off. On one hand, IoT devices are power-limited. On the other hand, there is an interest in having as many power levels as possible in order to satisfy the requirements in terms of spectral efficiency and massive connectivity. Let L be the number of received power levels available per channel at the relay. We also consider that continuous traffic has priority over periodic and sporadic traffic. For this purpose, the $L - 1$ highest power levels are dedicated to devices delivering continuous traffic and the L -th weakest level is only available for non-continuous (sporadic/periodic) traffic. This NOMA model ensures that the minimum rate required for continuous-traffic devices is guaranteed even if multiple collisions occur on the lowest power level. The maximum number of handled collisions depends on the average number of active devices generating non-continuous traffic.

As already mentioned, each relay operates in full-duplex mode. It receives the NOMA superimposed signals from the IoT devices, decodes these signals using SIC, and stores them in the relay memory buffer. The contents of the buffer are then transmitted to the BS on the same frequency band over

a number of time slots that depends on the QoS requirements of the IoT devices, the number of correctly decoded signals at the relay, the backhaul channel conditions and the power budget of the relay.

For backhaul links, each relay is assigned a frequency subband, and a conventional orthogonal multiple access is implemented. Thereby, backhaul links do not suffer from inter-relay interference. Moreover, since the relays operate in a full-duplex mode, the same subbands are used in the uplink and backhaul links.

With the proposed system model, the overall instantaneous AoI of an IoT device is computed as the summation of the AoIs for the relay uplinks and backhaul links that carry the data sent by the device. For the relay uplink, the AoI depends on the number of collisions the message undergoes before being successfully decoded at the relay. It can be reduced by optimizing the subband and power level allocations and by minimizing the number of signal collisions. For the backhaul link, AoI is directly related to the congestion level of the buffer. It is measured as the time that a message has to be stored in the buffer before being sent to the BS.

The main features of this system model are summarized in Table I.

TABLE I: System features

Feature	Type or nature
Type of system	Uplink with uncoordinated spectrum access and NOMA
Traffic	Mixed: continuous, periodic and sporadic
Links	Relay uplink: IoT device \rightarrow relay Backhaul link: relay \rightarrow BS
Relay mode	Full duplex
Channel state information	Perfectly estimated at the receiver
Optimization problem	Subband/power level pair allocation
Objective function	Traffic type-dependent device satisfaction with best possible spectral efficiency

THE RESOURCE ALLOCATION PROBLEM

A major problem to be solved in uncoordinated access systems is the allocation of communication resources to the different devices of the system, especially frequency subbands and power levels, taking the specificity of their traffic type into account. The QoS requirements are different for devices generating continuous traffic and those transmitting data in periodic or sporadic mode. The proposed resource allocation approach is formulated using the MAB framework, and solved with one of the most common algorithms called the upper confidence bound (UCB) algorithm [14]. In the MAB framework, agents are called players and executing an action consists in

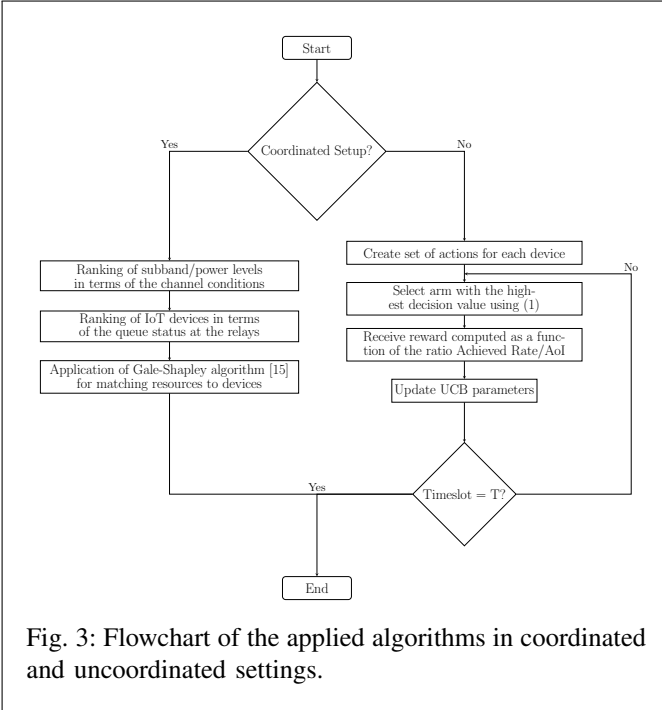
playing an arm. In the UCB algorithm, one has first to define the set of players, the set of arms and the rewards. In our system, the players are the IoT devices. Each device has to transmit on one frequency subband/power level combination. Hence, the set of arms is the set of available combinations of frequency subbands/power levels or, equivalently, relays/power levels. The nature of the rewards depends on the type of devices. Since devices generating continuous traffic are mainly dedicated to the more rate-hungry applications, their reward is the rate assigned to the device. For non-continuous traffic, we focus on applications requiring relatively low throughput but strict latency. In this case, the reward also includes the AoI. In addition, other constraints have to be taken into account, such as the power budget of the devices and relays. Thus, the algorithms are proposed in uncoordinated communication systems to achieve system stability and minimize the AoI. The allocation of the resources for continuous traffic-type devices guarantees system stability with no continuous overflow, whereas the allocation for non-continuous traffic-type devices is designed to meet the AoI constraints.

CONTINUOUS TRAFFIC: SYSTEM STABILITY

The first objective is to set up a stable system, where the relay buffers do not suffer from persistent overflows. To do so, the rate requirements of devices generating continuous traffic should not exceed the capacity of the backhaul link.

The proposed solution is based on the UCB approach, where the **IoT devices with continuous traffic** form the set of **players**. In practice, since each IoT device has a specific power budget and because of possible bad radio conditions, some power levels may not be available for all devices. Therefore, each device forms its own set of available **arms** from the $L - 1$ **highest power levels** according to its power budget and the radio channel conditions to the relays. In our system, the **reward** received by a device when it plays an arm is the **rate achieved** by the device, given by Shannon's formula as developed in [8].

In case two or more devices play the same arm, a collision occurs, resulting in a zero reward for devices with non-decodable signals. The main challenge is then the choice of the decision function that makes a device choose a frequency subband/power level pair. Every IoT device with continuous traffic aims to achieve a transmission rate at least equal to the requested one while spending the least amount of transmit power. Thus, two parameters should be considered in the design of the decision function: the rate achieved by the device and its transmit power which is the chosen received power level inversely proportional to the square of the channel gain. The minimum value of the transmit power is reached when the device selects the radio channel with the highest gain, i.e., the best radio channel. However, if each device chooses its best arm with respect to its channel gain, system convergence cannot always be guaranteed. Hence, IoT devices should choose their subband/power level pair to strike a compromise between the transmit power and the achieved rate. In practice, the trade-off is obtained via a weight β , whose value depends



on system requirements and which can be adjusted by the system operator:

$$decision = (1 - \beta) \times Reward - \beta \times Power. \quad (1)$$

In addition, in order to prevent the overflow of buffers at the relay level due to a mismatch between the aggregated continuous traffic rate and the capacity of the backhaul link and, therefore, to guarantee the stability of the system, a specific procedure has been implemented. When buffering occurs and the memory contents reach a specified threshold, a certain number of IoT devices with continuous traffic type are invited to transmit to a less congested relay to avoid any overflow. However, since the system is uncoordinated, no information is shared between the devices and the involved relay. Therefore, the devices need to learn alone that they need to choose another relay. To do so, when the relay has reached its filling threshold, a zero reward (instead of the achieved rate) is assigned to all the connected devices for a limited period, called *timeout period*. Its duration is a function of the system parameters and the number of connected devices. During the timeout period, the IoT devices connected to the relay are invited to run a new exploration phase and to discover new suitable arms. By playing new arms associated with less congested relays, the devices reach a new stable state of the system able to satisfy the required constraints. Once reached, a new exploitation phase can begin.

NON-CONTINUOUS TRAFFIC: AOI MINIMIZATION

The proposed solution is also based on the UCB approach, where the **IoT devices with non-continuous traffic** types form the set of **players**.

The set of **arms** is different from that of devices with continuous traffic because only the L -th and **weakest power**

level of each subband is available for IoT devices with non-continuous traffic. However, some subbands may not be available for some devices since periodic and sporadic transmissions need to efficiently share the available resources. In fact, collision management, channel quality and available device power budget may introduce restrictions related to subband access.

As a major shift from the case of continuous traffic devices, the **reward** for non-continuous traffic devices also accounts for the AoI. In practice, we have implemented a reward function **proportional to the achieved rate** and **inversely proportional to the AoI**. Hence, if a relay is congested, the AoI of connected devices tends to increase over successive time slots. Then, the corresponding reward decreases during these time slots and the device is encouraged to play new arms associated with less congested relays.

Significant improvement can still be achieved by devising a reward dependent upon a predicted AoI for the next time slot instead of the measured current value. In fact, with a deterministic access periodicity for continuous and periodic traffic devices, this prediction aims at estimating the access probability of devices with sporadic traffic. When coupled with the filling status of the relay, estimating the number of non-continuous devices connected in the next time slot allows the relay to predict whether it can avoid congestion or whether it should encourage the connected devices with sporadic traffic to connect to other relays. This incentive is achieved by controlling the value of the reward. The algorithms developed throughout the article are summarized in Fig. 3.

RESOURCE ALLOCATION USING COORDINATED MATCHING

In order to assess the efficiency of the proposed uncoordinated access approach, a coordinated allocation technique based on matching theory is introduced to benchmark the performance of the proposed algorithm. While lacking the self-configuration capability of the proposed approach, this reference technique is feasible in practice at the cost of comparatively high signaling overhead.

Matching theory is a mathematical framework in economics [15] that describes the formation of mutually beneficial relationships. In particular, it is commonly used to solve assignment problems between two disjoint sets of players. The matching problem consists in assigning resources from one set to the players of the other set. To do so, each player uses a preference relation based on a computed metric to rank the players of the other set from the viewpoint of a target utility function. By maximizing its utility, each player maximizes its objective function.

In the context of this paper, the resource allocation problem consists in assigning the network resources, i.e., frequency subband/power level pairs, to the IoT devices given a set of constraints from both the devices and the resources. It can therefore be modeled as a one-to-one matching game, meaning that each device can connect to one resource and each resource can be matched to one IoT device. We consider that the system is fully coordinated by the BS and that the BS knows all devices and relay power budgets, relay buffers status and

device traffic types. With this information and the centralized ability to devise a suitable resource allocation, this coordinated matching-based approach is expected to provide the best results in terms of QoS requirements. Being power limited, the IoT devices rank the subband/power level pairs according to their channel gain with the relay. In fact, the subband/power level pair requiring the least amount of transmit power is preferred by the device. In contrast, resource ranking mainly depends on the status of the corresponding relay buffer and the conditions between the relay and the BS. The fuller the relay buffer, the less likely it is to accept devices. The same applies to the channel gain between relay and BS: the better the radio conditions, the more likely the relay will accept devices.

NUMERICAL RESULTS

We consider an uplink communication system with one BS, one relay and 60 IoT devices divided into 36 devices generating continuous traffic, 12 devices having a sporadic traffic type with a spectrum access probability equal to 0.3 and 12 devices with a periodic traffic type. The periodic traffic devices are, in turn, divided into 4 groups of 3 devices and a time offset is applied for network access of the devices in the different groups to avoid mutual repeated collisions.

The target data rate is 0.75 Mb/s for the continuous traffic devices and 0.3 Mb/s for the ones with non-continuous traffic. The number of available subbands is 9. Five power levels are available at each subband, hence the total number of available resources is 45. This allows us to address the problem of congestion on a small scale, first. The objective is to prove the superiority of the proposal in supporting more linked devices than a conventional system while using the same finite number of resources. Then, by adding network nodes, the proposal can be scaled up to enable massive connectivity. Signals undergo a distance-dependent path-loss with a decay factor of 3.76, and a zero-mean log-normal shadowing with an 8 dB variance. The noise power spectral density is equal to 4×10^{-21} W/Hz. The power budget of the IoT devices is equal to 0.2 W and the one of the relays is 5 W. The congestion threshold of the relays is set to 50 percent of their whole buffer capacity (300 packets) and the duration of the timeout period is taken equal to 50 time slots.

Benchmarks

The performance of the proposed algorithm with and without AoI prediction is evaluated for 1000 time slots using Matlab and compared with the fully coordinated approach based on matching theory and with two more basic uncoordinated solutions:

- random access NOMA (RA-NOMA) method, where devices choose the frequency subband/power level pair randomly,
- an allocation technique (No-AoI) where the devices solely allocate their arms using UCB without taking the AoI effect into consideration.

Evaluation

The efficiency of the arm assignment process is evaluated in Fig. (4a), where the instantaneous successful transmission percentage, i.e., the proportion of devices having their signals correctly decoded by the relay, is plotted as a function of time. We can observe that the basic RA-NOMA approach cannot achieve a success rate above 40 percent with an average of less than 30 percent. In contrast, since the matching technique is fully coordinated by the BS, convergence towards the full satisfaction state is reached in the very first time slots. Similarly, the proposed UCB-based technique also reaches convergence with a 100 percent satisfaction rate for all IoT devices after an exploration phase of 80 time slots.

The average transmit power per device as a function of the requested rate is evaluated in Fig. (4b). Simulations showed that a balanced scheme with $\beta = 0.5$ is the best choice for minimizing power consumption. Moreover, for the same average transmit power, the proposed UCB-based method with $\beta = 0.5$ is the only solution within 0.15 Mb/s of the fully coordinated benchmark solution for the same average transmit power.

Figures (4c) and (4d) show the average AoI value when varying the total number of sporadic players or the access probability. In Fig. (4d), the number of sporadic devices is 9. The proposed approach significantly outperforms basic random access approaches.

REMAINING CHALLENGES

The next step prior to any wide adoption resides in a thorough investigation of system robustness to imperfect or partial channel state information (CSI). Following this step, it would be possible to evaluate the benefits of the reduced signaling overhead. Additional impactful fine-tuning includes the trade-off between exploration and exploitation phases and higher-level auto-configuration capabilities. The latter involves devising adaptive methods for unsupervised traffic rerouting to avoid relay failures, or even, intelligent automated allocations to improve system spectral and/or power efficiencies. Intelligent reflective surfaces and on-demand mobile relays can represent appealing complementary means to these same effects and may be considered for future developments. On a different note, some physical layer techniques, such as waveform design, are considered essential enablers for uncoordinated spectrum access and should be thoroughly investigated. Indeed, robustness to impairments including carrier frequency offsets, timing offsets and efficient multi-user management are key for unleashing the potentials of such decentralized intelligent allocation methods.

CONCLUSION

In this article, the uncoordinated decentralized joint channel and power allocation problem was addressed and formulated using the MAB framework. A solution based on the UCB algorithm was then proposed. In addition to increased spectral efficiency, the proposed solutions minimize AoI for IoT devices in a two-hop mixed-traffic NOMA uplink communication system with relays. Thanks to their self-configuration

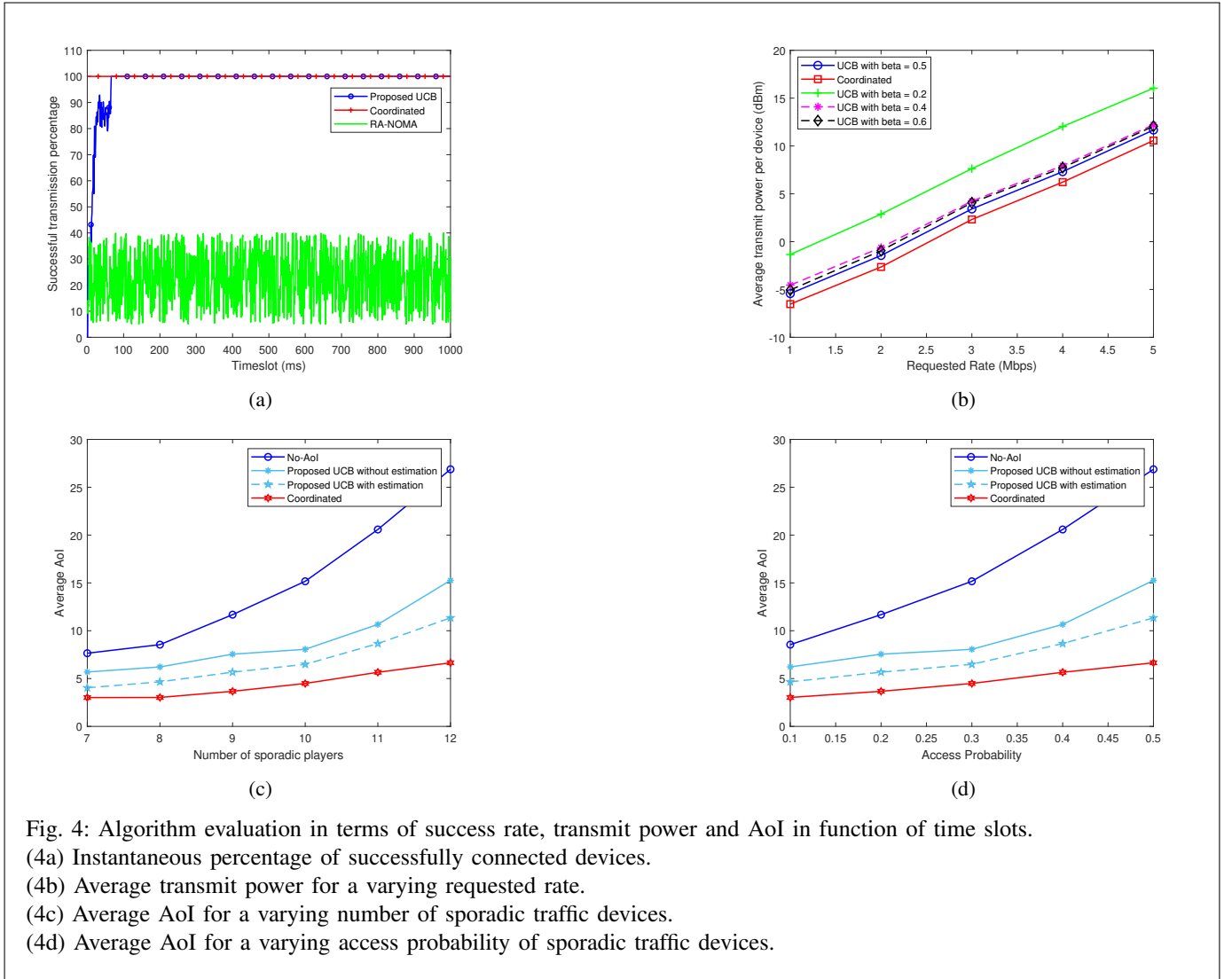


Fig. 4: Algorithm evaluation in terms of success rate, transmit power and AoI in function of time slots.

- (4a) Instantaneous percentage of successfully connected devices.
 (4b) Average transmit power for a varying requested rate.
 (4c) Average AoI for a varying number of sporadic traffic devices.
 (4d) Average AoI for a varying access probability of sporadic traffic devices.

capabilities, they are able to address relay-level congestion, provide the scalability for mixed traffic support and manage AoI for the coexistence of multiple services and applications. This was achieved through the UCB algorithm augmented by both the introduction of buffering and an AoI-controlling scheduler at the level of the relays. AoI prediction for sporadic traffic was shown to further improve performance for delay-sensitive applications. The resulting self-configurable and scalable proposal was shown to outperform competing uncoordinated techniques while approaching the performance of a centralized solution with perfect knowledge of all system parameters. Future work should focus on key enablers for such a system including imperfect or partial CSI and physical layer support.

ACKNOWLEDGMENT

This work was carried out in the context of Beyond5G, a project funded by the French government as part of the economic recovery plan and the investments for the future program.

REFERENCES

- [1] S. Al-Sarawi, M. Anbar, R. Abdullah, and A. B. Al Hawari, "Internet of Things Market Analysis Forecasts, 2020–2030," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, London, UK, Jul. 2020, pp. 449–453.
- [2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, 2020.
- [3] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [4] M.-J. Youssef, J. Farah, C. Abdel Nour, and C. Douillard, "Full-Duplex and Backhaul-Constrained UAV-Enabled Networks Using NOMA," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 9, pp. 9667–9681, 2020.
- [5] D. Zucchetto and A. Zanella, "Uncoordinated Access Schemes for the IoT: Approaches, Regulations, and Performance," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 48–54, 2017.
- [6] R. S. Sutton and A. G. Barto, *Reinforcement Learning, Second Edition - An Introduction*. Cambridge, MA: MIT Press, 2018.
- [7] J. Choi, "NOMA-Based Random Access With Multichannel ALOHA," *IEEE J. on Sel. Areas in Commun.*, vol. 35, no. 12, pp. 2736–2743, Oct. 2017.
- [8] J. Doumit, M.-J. Youssef, C. Abdel Nour, J. Farah, and C. Douillard, "Resource Allocation in Full-Duplex Uncoordinated Communication Systems with NOMA," in *2021 IEEE 32nd Annu. Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Helsinki, Finland, Sept. 2021, pp. 1104–1110.

- [9] M.-J. Youssef, V. V. Veeravalli, J. Farah, and C. Abdel Nour, "Stochastic Multi-Player Multi-Armed Bandits with Multiple Plays for Uncoordinated Spectrum Access," in *2020 IEEE 31st Annu. Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, London, United Kingdom, Sept. 2020, pp. 1–7.
- [10] W. Wang, A. Leshem, D. Niyato, and Z. Han, "Decentralized Learning for Channel Allocation in IoT Networks Over Unlicensed Bandwidth as a Contextual Multi-Player Multi-Armed Bandit Game," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 5, pp. 3162–3178, 2022.
- [11] X. Wang, J. Ye, and J. C. Lui, "Decentralized Task Offloading in Edge Computing: A Multi-User Multi-Armed Bandit Approach," in *IEEE INFOCOM 2022 - IEEE Conf. on Computer Commun.*, London, United Kingdom, May 2022, pp. 1199–1208.
- [12] R. Talak, S. Karaman, and E. Modiano, "Improving Age of Information in Wireless Networks With Perfect Channel State Information," *IEEE/ACM Trans. on Netw.*, vol. 28, no. 4, pp. 1765–1778, Jun. 2020.
- [13] Y. Ni, L. Cai, and Y. Bo, "Vehicular beacon broadcast scheduling based on age of information (AoI)," *China Commun.*, vol. 15, no. 7, pp. 67–76, July 2018.
- [14] M. A. Adjif, O. Habachi, and J. P. Cances, "Joint Channel Selection and Power Control for NOMA: A Multi-Armed Bandit Approach," in *2019 IEEE Wireless Commun. and Netw. Conf. Workshop (WCNC)*, Marrakech, Morocco, Morocco, Apr. 2019, pp. 1–6.
- [15] D. Gale and H. S. Shapley, "College Admissions and the Stability of Marriage," *American Math. Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

Catherine Douillard is a full Professor at IMT Atlantique/Lab-STICC laboratory, Brest, France. Her main research interests are error correcting codes, coded modulations and diversity techniques for multi-carrier, multi-antenna and multiple access transmission systems. Between 2007 and 2012, she participated in Digital Video Broadcasting Technical Modules for the definition of DVB-T2, DVB-NGH and DVB-RCS2 standards. She served several times as chair of the technical program committee and general chair of ISTC (International Symposium on Topics in Coding). In 2009, she received the SEE/IEEE Glavieux Award for her contribution to standards and related industrial impact.

Joseph Doumit received the B.E. and M.S. degrees in computer and communications engineering from Lebanese University, Lebanon, in 2020. He is currently pursuing a Ph.D. degree in information and communication engineering with IMT Atlantique, France. His current research interests include resource allocation, non-orthogonal multiple access, reconfigurable intelligent surfaces, and reinforcement learning.

Marie-Josepha Youssef received the B.E. and M.S. degrees in computer and communications engineering from Lebanese University in 2016, and the Ph.D. degree in information and communication engineering from the IMT Atlantique, France, in 2020. She is currently a quantitative researcher at Capital Fund Management. Her current research interests include resource allocation, non-orthogonal multiple access, matching theory, reinforcement learning, unmanned-aerial vehicles, and grant-free communications.

Charbel Abdel Nour obtained his PhD from Telecom Bretagne in 2008 and his Habilitation thesis from the University of South Brittany in 2020. Since Dec. 2022, Charbel holds a Professor position at the Mathematical and Electrical Engineering Department at IMT Atlantique. His interests concern coded modulations, resource allocation for NOMA, waveform design, MIMO and iterative receivers. Lately, he presented several contributions to the H2020 METIS, FANTASTIC5G and EPIC projects and to the 3GPP consortium related to coding solutions for 5G. He also served as co-chair of the technical program committee of ISTC (International Symposium on Topics in Coding) in 2023.

Joumana Farah received a B.E. in Electrical Engineering from the Lebanese University, in 1998, the M.E. and Ph.D. degrees in 1999 and 2002 from Grenoble INP, France. In 2010, she obtained the Habilitation to Direct Research from Université Paris VI. Until 2022, she was a Professor at the Lebanese University. Since January 2023, she has joined the Institut National des Sciences Appliquées (INSA) at Rennes, France. She has nine registered patents and has coauthored over a hundred journal and conference papers. Her current research interests include resource allocation techniques, channel coding and interference management techniques.