Evaluating Fast and Grant-Free Uplink Access in Next-Generation Cellular IoT Networks

Lars Moons, Samer Nasser, Adnan Sabovic, Ritesh Kumar Singh and Jeroen Famaey
IDLab, University of Antwerp – imec, Belgium

***Abstract*—Next-generation cellular communication technologies must follow the trend of massive deployment and energy efficiency in the Internet of Things (IoT), preferably using existing network infrastructure. Narrowband IoT (NB-IoT) stands out as a promising candidate due to its low cost and easy installation because of its Long Term Evolution (LTE)-based architecture. However, the conventional random access procedure of NB-IoT limits its energy efficiency and predictability. To this end, this paper compares the achievable connection density and energy efficiency of the conventional random access scheme with novel grant-free and fast-uplink access techniques. Our findings reveal that optimal random access schemes can support three times as many devices as grant-free schemes and therefore perform the best in terms of connection density. However, when it comes to energy efficiency, random access performs the worst with energy consumption up to 12 times higher than grant-free schemes. Fast-uplink access fulfills its expectations as the most promising technique for next-generation IoT communication, supporting a connection density twice as high as grant-free access, at similar power consumption.**

***Index Terms*—Energy consumption, fast-uplink access, grant-free access, Narrowband IoT, random access, 6G**

I. Introduction

Low-Power Wide-Area Networks (LPWANs) are a new set of long-range radio technologies with low energy consumption. As such, they are seen as one of the key enablers for a sustainably scalable next generation of Internet of Things (IoT) devices due to their energy efficiency, high capacity, and extensive coverage. They also showed their usefulness in combination with constrained energy neutral devices (ENDs), which require even more energy-efficient networks due to their limited energy availability, obtained from ambient sources or wireless power transfer. By using the existing infrastructure of the Long-Term Evolution (LTE) network, Narrowband IoT (NB-IoT) positions itself as the LPWAN technology for energy-efficient cellular networks capable of scaling to many nodes.

A major challenge in using cellular-based networks in dense IoT networks, is the conventional random access scheme they use for uplink channel access. Although it enables efficient control of channel resources, it suffers from excessive access delays and large signaling overhead, thus reducing energy efficiency [1]. To address these issues, there are several alternative channel access schemes that could be considered. Grant-free access schemes allow devices to transmit directly to the base station (BS) without going through the controlled random access process first. While this approach improves access delays and energy efficiency, it suffers from a large number of collisions. To resolve these collisions, the Non-Orthogonal Multiple Access (NOMA) technique can be used to enable multiple users to share the same spectrum. Finally, there are fast-uplink access schemes that can proactively allocate uplink resources and send grants to devices, thus reducing collisions and signaling overhead by eliminating the need for scheduling requests.

This paper studies and compares the connection density and energy efficiency of these three channel access schemes. We compare the access schemes through a combination of analytical methods, link-level simulations, and system-level simulations. We derive exact formulas to compute the outage probability of both, controlled and uncontrolled access schemes. Finally, the energy consumption is calculated for each of the selected access schemes, considering the maximum supported transmit power and number of devices for different traffic scenarios.

The remainder of this paper is structured as follows. Section II describes related work. Section III provides a brief overview of NB-IoT, the traffic model, and network density optimization. Section IV examines the advantages of controlled versus uncontrolled access, which lays the foundation for analyzing the three NB-IoT access schemes in Section V. Section VI compares the three access schemes considering connection density and energy consumption. Finally, our conclusions are provided in Section VII.

II. Related work

Different versions of grant-free schemes were analyzed by Berardinelli *et al.* [2], including the blind procedure where transmissions are repeated a set number of times, and feedback-based procedures where the BS may issue a grant in case of a decoding failure. Mitsiou *et al.* [3] compared the fast-uplink scheme with grant-free access specifically for ENDs, demonstrating the energy efficiency potential of fast-uplink schemes. However, both papers considered a theoretical wireless channel, with link performance evaluated through a Shannon-like capacity formula. In contrast, our paper focuses on a specific wireless technology, NB-IoT, allowing us to conduct more realistic link-level simulations. Additionally, these studies did not include random access schemes in their comparisons.

Liu *et al.* [4] presented an analytical framework to analyze the latent access failure probability in contention-based grant-free access. El Tanab *et al.* [5] examined fast-uplink schemes in combination with NOMA. Zhou *et al.* [6] compared random access schemes with grant-free, considering accurate modeling of energy consumption. However, this study did not include link-level simulations or consider fast-uplink schemes. Finally, the random access scheme was further examined and the total energy consumption of NB-IoT devices was investigated through practical experiments considering a real END prototype by Sultania & Famaey [7].

To the best of our knowledge, this is the first paper that establishes a foundation for comparing three main uplink access schemes for next-generation cellular IoT networks using a combination of analytical methods, link-level simulations, and system-level simulations. Specifically, we simulate the NB-IoT shared channel link and use these results in our analytical and system-level simulations. This allows us to derive key properties such as connection density and energy consumption.

III. Modeling of NB-IoT environment

In this section, a brief overview of the NB-IoT protocol, along with its channels used for different access schemes and techniques to evaluate them is given. Furthermore, we describe the traffic model and network density optimization.

A. Narrowband IoT

NB-IoT is an LPWAN technology developed by the 3rd Generation Partnership Project (3GPP) for cellular networks. Leveraging many features from LTE, including its channels and signals, NB-IoT offers low-cost deployments with a simplified protocol stack compared to LTE. Downlink and uplink resources are divided into several channels [8]. The smallest unit that can be transmitted in the uplink is the Resource Unit (RU). Grants contain the number of RUs available for transmission. The RU duration T_{RU} is 8 ms. Grants also contain the Transport Block Size (TBS) of the scheduled transmission, determined by the Modulation and Coding Scheme (MCS). Transport blocks can be repeated to improve the robustness of communication. The scheduler assigns a specific number of RUs and repetitions for each transmission. Based on this, the time slot duration T_s can be calculated as follows:

$$T_s = N_{\mathrm{rep}} \times N_{\mathrm{RU}} \times T_{\mathrm{RU}} \tag{1}$$

Where N_{rep} is the number of repetitions, and N_{RU} is the number of resource units.

B. Block error rate

To evaluate access methods, we require an accurate model of the Narrowband Physical Uplink Shared Channel (NPUSCH), the channel that carries the application data. This data is first modulated and transmitted by the device. Then, it travels through a noisy channel before being received and demodulated at the BS. Some packets will be lost due to poor channel conditions caused by interference, fading, and path loss. The Block Error Rate (BLER) is defined as the ratio of the number of incorrectly received packets to the total number of transmitted packets. The BLER for the NPUSCH channel can be approximated for several SNR values using the LTE toolbox from MATLAB [9]. This data can be fitted to a logistic curve to create a continuous function. This S-shaped curve is expressed as follows:

$$\epsilon(\gamma) = \frac{S_1}{1 + e^{-S_2(\gamma - S_3)}} \tag{2}$$

where γ is the SNR and S_1, S_2 and S_3 are the parameters for which we need to find the best fit (e.g., using the Gauss–Newton method for least squares).

C. Traffic model

The traffic modeling uses a discrete two-state Markov chain, as proposed by Thomsen *et al.* [10]. This chain models a device that is either active or inactive, transitioning from the inactive state to the active state with probability δ_0 and vice versa with a probability δ_1.

For $\delta_1 \approx 1$, traffic is almost random, while for $\delta_1 \lessapprox 0.01$, traffic occurs in large bursts. The steady state probability of the active state can be calculated as follows:

$$p = \frac{\delta_0}{\delta_0 + \delta_1} \tag{3}$$

which means that, in the long run, a device is active with probability p, which we call the traffic rate. In simulations, we configure the traffic model using parameters p and δ_1, and determine δ_0 from Eq. 3.

D. Network density optimization

When comparing access schemes, we will consider the maximum number of devices that can operate within a network, while still satisfying a specified maximum outage probability. We reformulate the problem and develop an efficient algorithm to solve it. Given a strictly increasing function $f : \mathbb{N}_0 \to \mathbb{R}$ and a value $a \in \mathbb{R}$, find the largest $n \in \mathbb{N}_0$ such that $f(n) < a$. Here, n represents the number of devices, a corresponds to the maximum outage probability and f calculates the outage probability given n using a system-level simulator. This problem can be efficiently solved using the classical exponential search algorithm proposed by Bentley & Yao [11]. This algorithm consists of two steps. In the first step, we establish an upper bound on n by progressively increasing n in powers of two until $f(n) < a$ is no longer satisfied. In the second step, we use this upper bound $n = 2^i$ to perform a binary search on the interval $(2^{i-1}, 2^i)$. This binary search finds the target n by iteratively halving the interval until only one element remains. The time complexity of this algorithm is therefore $\mathcal{O}(\log n)$ for the target n.

IV. Uplink access

We begin our discussion of uplink access schemes by examining the impact of controlled versus uncontrolled access. In uncontrolled access, devices independently select a subcarrier for their transmission, while in controlled access, the base station ensures an even distribution of devices among the subcarriers.

A. Uncontrolled access

To model the collisions in uncontrolled access [2], consider an arbitrary active device D, the probability that u other devices transmit in the same time slot as device D can be calculated as follows:

$$P_1(u) = \binom{N-1}{u} p^u (1-p)^{N-1-u} \tag{4}$$

where N is the total number of devices in a network and p is the traffic rate (i.e., the probability that a device is active and therefore transmits). Given u other active devices, the probability that v of them select the same subcarrier as device D can be calculated as follows:

$$P_2(u, v) = \binom{u}{v} \frac{(K-1)^{u-v}}{K^u} \tag{5}$$

where K is the number of subcarriers. The probability that device D collides with z other devices can then be calculated as follows:

$$P_c^{\mathrm{U}}(z) = \sum_{u=z}^{N-1} P_1(u) P_2(u, z) \tag{6}$$

B. Controlled access

For controlled access, we model collisions as the "Balls into Bins" problem, where we consider devices as balls and subcarriers as bins [12]. Given $u+1$ active devices, each subcarrier contains either q or $q+1$ devices:

$$q = \left\lfloor \frac{u+1}{K} \right\rfloor \tag{7}$$

Given u other active devices, the probability that v of them select the same subcarrier as device D is calculated as follows:

$$P_3(u, v) = \begin{cases} 1 & \text{for } u = 0 \text{ and } v = 0 \\ 1 - a & \text{for } v = q - 1 \\ a & \text{for } v = q \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$a = \frac{\mathrm{mod}(u+1, K) \times (q+1)}{u+1} \tag{9}$$

Similar to Eq. 6, the probability that device D collides with z other devices can be calculated as follows:

$$P_c^{\mathrm{C}}(z) = \sum_{u=z}^{N-1} P_1(u) P_3(u, z) \tag{10}$$

C. Comparison between controlled and uncontrolled access

Figure 1 shows the differences between the two schemes in a network with 50 devices, 6 subcarriers and $p = 0.5$. Both Monte Carlo simulations and analytical results from the equations are plotted, showing a strong agreement between the two. While both distributions have the same mean, controlled access produces a distribution with lower variance. The plot shows that in controlled access, the chance of an arbitrary device colliding with six other devices on the same subcarrier is very low (0.1%), compared to a much higher likelihood in uncontrolled access (11%).

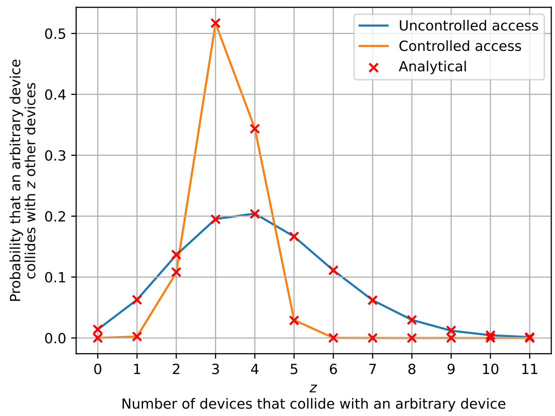


Fig. 1: Comparing collision probability of controlled and uncontrolled access with 50 devices, 6 subcarriers, and $p = 0.5$

D. Outage probability

Collisions create noise due to interference. The Signal-to-Interference-plus-Noise Ratio (SINR) for device D when colliding with z other devices can be calculated as follows:

$$\Gamma(z) = \frac{\gamma}{1 + z\gamma} \tag{11}$$

where all devices compensate for their path-loss to keep the SNR received at the BS equal to γ. The BLER for device D can be calculated as follows:

$$\epsilon_D(z) = \epsilon(\Gamma(z)) \tag{12}$$

Where ϵ is the BLER curve from Eq. 2. The probability that the data packet from device D is successfully decoded can be calculated as follows:

$$P_D = \sum_{z=0}^{N-1} P_c(z)(1 - \epsilon_D(z)) \tag{13}$$

Where $P_c(z) = P_c^{\mathrm{U}}(z)$ for uncontrolled access or $P_c(z) = P_c^{\mathrm{C}}(z)$ for controlled access.

Finally, The outage probability, which is the probability that a packet decoding fails, can be calculated as follows:

$$P_{\mathrm{out}} = 1 - P_D \tag{14}$$

The outage probability for both schemes is plotted in Figure 2, showing that controlled access significantly outperforms uncontrolled access. As the number of devices increases, more devices select the same subcarrier, leading to higher noise levels and thus a higher probability of decoding failures. However, this increase is less significant for controlled access, which effectively suppresses the right tail, as shown in Figure 1.

V. Access Schemes

This section examines the three main access schemes: grant-free access, fast-uplink access, and random access. These schemes will be linked to controlled and uncontrolled access described in Section IV. Additionally, we derive formulas for the energy consumption of each scheme.

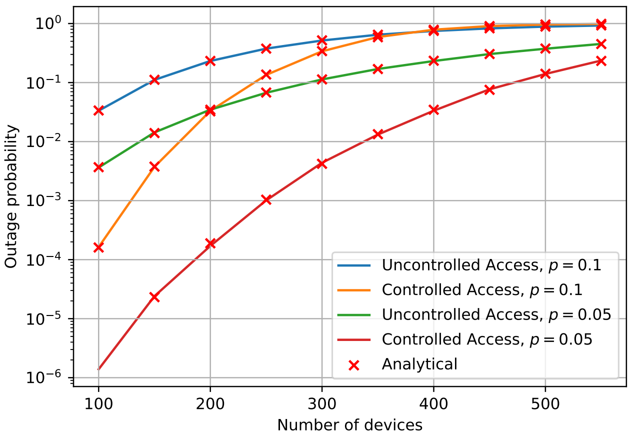


Fig. 2: Comparing the connection density of uncontrolled and controlled access in a network with 6 subcarriers and an SNR of 10 dB

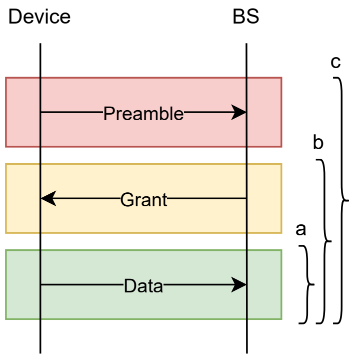


Fig. 3: Transmission sequence for: a) Grant-free access, b) Fast-uplink access, c) Random access

A. Grant-free access

In grant-free access, devices transmit to the BS without waiting for or requesting a grant, as illustrated in Figure 3. This can lead to a large number of unevenly distributed collisions, increasing the outage probability. Note that the outage probability of the grant-free scheme can be computed using the outage probability of the uncontrolled access scheme from the Section IV-A. The energy consumption of a grant-free access attempt can be computed as follows:

$$E^{\mathrm{GF}} = \mathcal{P}_t T_s \tag{15}$$

where T_s is the duration of a time slot and $\mathcal{P}_t$ is the power consumed by the device during transmission.

B. Fast-uplink access

In fast-uplink access, the BS predicts device activity and proactively sends grants containing uplink resources, as shown in Figure 3. This scheme can use controlled access by evenly distributing the devices over the subcarriers. The performance depends heavily on the accuracy of the prediction algorithm. Given the traffic model from Section III-C, if $\delta_1 < 0.5$, an active device will most likely remain active in the next time slot. Therefore, we simply predict the set of next active devices using the set of previous active devices. For smaller δ_1, prediction accuracy increases. The parameter δ_1 can therefore be used to configure the predictability of the environment. Thus, a less predictable environment results in a more random distribution of devices, whereas a more predictable environment enables the base station to achieve a more even distribution of devices. The energy consumption of a fast-uplink access attempt can be computed as follows:

$$E^{\mathrm{FU}} = \mathcal{P}_r T_{\mathrm{gra}} + \mathcal{P}_t T_s \tag{16}$$

where $\mathcal{P}_r$ is the power consumed by the device during reception and T_{gra} is the duration of a grant.

C. Random access

Random access is the conventional uplink access method in cellular communication which uses all three channels [8] for enabling uplink synchronization, obtaining a grant, and, finally, transmitting the data to the BS, as shown in Figure 3. Colliding preambles can be resolved similarly to colliding packets, if the SNR exceeds a certain threshold [6]. Additionally, we assume no interference from colliding preambles [6]. These assumptions result in an optimal random access process and therefore provides an upper bound on the scheme. Matlab simulations show that preambles need $N_{\mathrm{rep}}^{\mathrm{RA}} = 32$ repetitions to achieve similar robustness to the NPUSCH transmissions [13]. Note that the outage probability of the random access scheme can be computed using the outage probability of the controlled access scheme from Section IV-B. The energy consumption of a random access attempt can be computed as follows:

$$E^{\mathrm{RA}} = \mathcal{P}_t \left(T_{\mathrm{pre}} N_{\mathrm{rep}}^{\mathrm{RA}} + T_s \right) + \mathcal{P}_r T_{\mathrm{gra}} \tag{17}$$

where T_{pre} is the duration of a preamble (i.e., $5.6\,\mathrm{ms}$ [14]).

VI. Experiments

In this section, we will compare the three access schemes in terms of connection density and energy consumption.

A. Simulation setup

We consider an uplink bandwidth of $180\,\mathrm{kHz}$ containing 12 subcarriers with a subcarrier spacing of $15\,\mathrm{kHz}$. These subcarriers are split evenly among the NPUSCH and Narrowband Physical Random Access Channel (NPRACH), resulting in 6 subcarriers for each channel. We consider a TBS of 72 bits, suitable for small data transmissions. The scheduler assigns 1 RU and 2 repetitions per transmission, which results in time slot duration T_s = 16 ms. Furthermore, we consider QPSK modulation, two receive antennas for the BS and a Rayleigh fading channel. The noise in this channel is characterized by Additive White Gaussian Noise (AWGN). For the NPUSCH channel, we simulate 1000 transport blocks for different SNR values and calculate the BLER using the Matlab LTE toolbox. This data can be used to fit the logistic curve from Eq. 2,

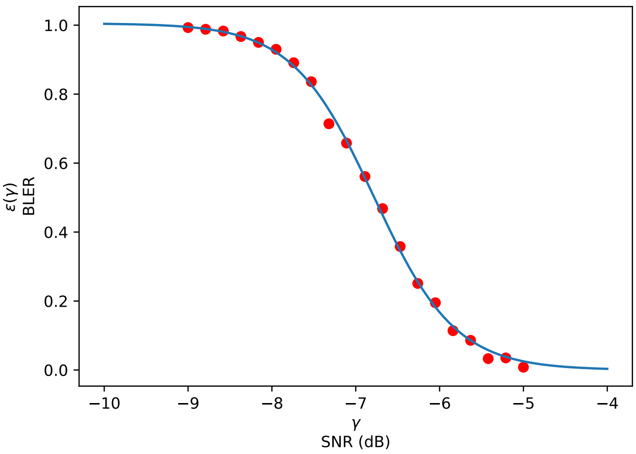


Fig. 4: Fitted BLER curve

which results in parameters $S_1 \approx 1.005$, $S_2 \approx -2.055$ and $S_3 \approx -6.783$. This is visualized in Figure 4. We see that the BLER decreases as the SNR increases. This increase in SNR can result from either an increase in signal power or a decrease in noise power.

B. Connection density

We compute the connection density as the maximum number of devices $N_{\max}$ that can operate within a network while still satisfying a specified maximum outage probability $P_{\text{out,th}} = 0.01$. The problem can be formulated as

$$N_{\max} = \underset{N}{\text{argmax}} \left(P_{\text{out}}\right) \quad \text{s.t.} \quad P_{\text{out}} < P_{\text{out,th}} \tag{18}$$

where P_{out} is the outage probability from Eq. 14. The maximum number of devices can be efficiently determined using the exponential search algorithm (cf., Section III-D). Retransmissions are not considered as these significantly increase the energy consumption.

The simulation results are plotted in Figure 5, which shows that the most energy consuming schemes deliver the best performance, with random access supporting three times as many devices as grant-free access. Performance of the fast-uplink access scheme improves in more predictable environments (traffic with large bursts, which means smaller δ_1).

C. Energy consumption

To achieve the SNR γ for the results shown in Figure 5, a certain transmission power $\mathcal{P}_t$ is required, which results in a certain energy consumption E. Given a constant noise power, constant path loss, and constant time unit, this relationship can be expressed as follows:

$$\gamma \propto \mathcal{P}_t \propto E \tag{19}$$

The energy consumption of different access schemes can then be expressed using Eq. 15, 16, and 17. The values used for $\mathcal{P}_t$ and $\mathcal{P}_r$ are based on the power measurement framework for NB-IoT proposed by Wang *et al.* [15], where a mapping is described between the transmit power in dBm and the actual power consumption of a SARA-N211 NB-IoT-capable device. Furthermore, T_{gra} is equal to 2 ms considering that the grant transmission only requires 2 repetitions to achieve similar robustness to the NPUSCH transmissions [14]. This is based on the assumption that the base station's transmit power is significantly higher than that of small IoT devices.

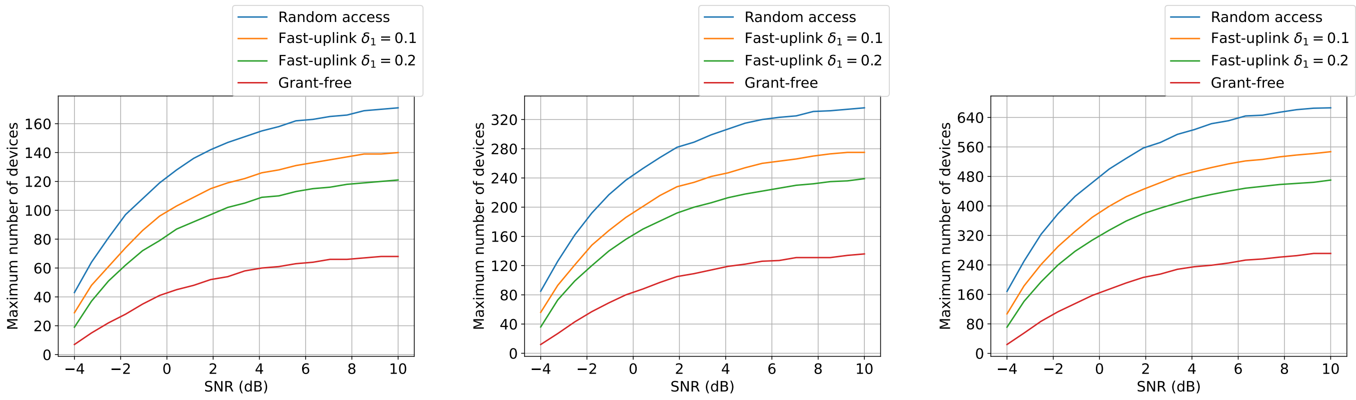
The results of the simulations are shown in Figure 6. The analysis indicates that both grant-free and fast-uplink access support low energy consumption for a small number of devices, whereas random access initially starts with an energy consumption that is approximately 12 times higher. As the number of devices increases, both grant-free and fast-uplink access maintain low energy consumption until they reach a certain threshold where the energy consumption starts rising more agressively and eventually gets cut off due to the NB-IoT maximum transmit power, which is set to 23 dBm, based on European regulation. This threshold is higher for fast-uplink access compared to grant-free access due to larger connection densities that fast-uplink access offers. For grant-free access, increasing the number of devices quickly leads to more collisions, making it challenging to meet the outage probability requirement, especially since the devices are randomly distributed over the subcarriers. Conversely, fast-uplink access uses a prediction algorithm to effectively delay the more aggressive increase in energy consumption by evenly distributing the predicted active devices across subcarriers, which is particularly beneficial in more predictable environments.

VII. Conclusions

In this paper, we evaluated three access schemes for next-generation IoT networks: the conventional random access scheme, and two promising alternatives, the fast-uplink and grant-free schemes. Our analysis was supported by analytical methods, link-level simulation, and system-level simulations. We highlighted the advantages of controlled versus uncontrolled access methods, providing intuition into the poor connection density of grant-free schemes. This suggested that evenly distributing devices across subcarriers effectively supports more devices while maintaining a given outage probability. By using a traffic model based on a simple Markov chain, the predictability of the traffic could be adjusted, which proved to be a key factor in evaluating the effectiveness of fast-uplink access schemes. While the random access scheme supported the highest number of devices, it was significantly less energy efficient. Conversely, fast-uplink schemes were more energy efficient, particularly in more predictable environments, consuming on average 12 times less energy than random access schemes. Simple grant-free schemes performed the worst in terms of connection density and were only energy efficient for a limited number of devices.

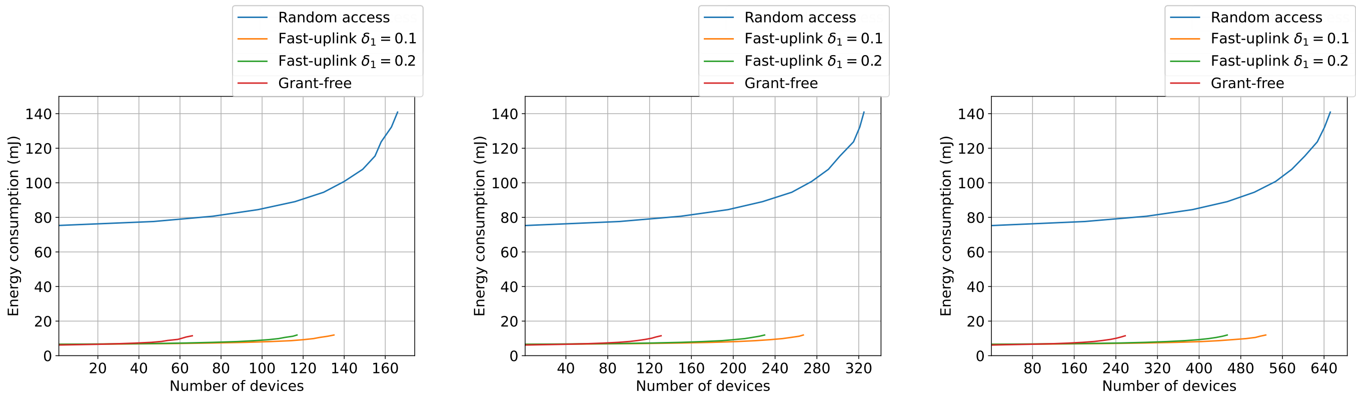
Acknowledgments

Part of this research was funded by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101095759 (JU SNS Hexa-X-II).



(a) Traffic rate $p = 0.1$ (b) Traffic rate $p = 0.05$ (c) Traffic rate $p = 0.025$

Fig. 5: Comparing the connection density of uplink access schemes where devices need to satisfy a maximum outage probability of 0.01



(a) Traffic rate $p = 0.1$ (b) Traffic rate $p = 0.05$ (c) Traffic rate $p = 0.025$

Fig. 6: Comparing the energy consumption of uplink access schemes where devices need to satisfy a maximum outage probability of 0.01

REFERENCES

[1] M. El-Tanab and W. Hamouda, "An overview of uplink access techniques in machine-type communications," *IEEE Network*, 2021.

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

[3] N. A. Mitsiou, S. A. Tegos, P. D. Diamantoulakis, P. G. Sarigiannidis, and G. K. Karagiannidis, "Proactive scheduling for zero-energy device networks with fast uplink grant," *IEEE Wireless Communications Letters*, 2023.

[4] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE Journal on Selected Areas in Communications*, 2021.

[5] M. E. Tanab and W. Hamouda, "Efficient resource allocation in fast-uplink grant for machine-type communications with NOMA," *IEEE Internet of Things Journal*, 2022.

[6] H. Zhou, Y. Deng, L. Feltrin, and A. Höglund, "Analyzing novel grant-based and grant-free access schemes for small data transmission," *IEEE Transactions on Communications*, 2022.

[7] A. K. Sultania and J. Famaey, "Batteryless NB-IoT prototype for bidirectional communication powered by ambient light," *Ad Hoc Networks*, 2023.

[8] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow band Internet of Things," *IEEE Access*, 2017.

[9] P. Jörke, T. Gebauer, and C. Wietfeld, "From LENA to LENA-NB: Implementation and performance evaluation of NB-IoT and early data transmission in ns-3," in *Workshop on NS-3*, 2022.

[10] H. Thomsen, C. N. Manchon, and B. H. Fleury, "A traffic model for machine-type communications using spatial point processes," in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications*, 2017.

[11] J. L. Bentley and A. C.-C. Yao, "An almost optimal algorithm for unbounded searching," *Information Processing Letters*, 1976.

[12] M. Raab and A. Steger, "Balls into bins – a simple and tight analysis," in *International Workshop on Randomization and Approximation Techniques in Computer Science*, 1998.

[13] MathWorks, "NB-IoT PRACH detection and false alarm conformance test," https://nl.mathworks.com/help/lte/ug/nb-iot-prach-detection-and-false-alarm-conformance-test.html, 2024.

[14] M. Kanj, V. Savaux, and M. Le Guen, "A tutorial on NB-IoT physical layer design," *IEEE Communications Surveys & Tutorials*, 2020.

[15] H. Wang, A. Sørensen, M. Remy, N. Kjettrup, J. Nielsen, and G. Madueno, *Power Measurement Framework for LPWAN IoT*. Springer, 2021, pp. 105–129.