System Level Integration of Irregular Repetition Slotted ALOHA for Industrial IoT in 5G New Radio

H. Murat Gürsu*, M. Çağatay Moroğlu*, Mikhail Vilgelm*, Federico Clazzer†, Wolfgang Kellerer*
* Chair of Communication Networks, Technical University of Munich, Munich, Germany
† German Aerospace Center (DLR), Weßling, Germany

***Abstract*—Automation is a key part of the new industrial revolution, that will be enabled by the deployment of thousands of sensors and actuators. The flexible deployment of these devices requires wireless connectivity which is labeled as industrial internet of things, IIoT. The sporadic activity pattern of IIoT devices naturally suggest the use of random access techniques, albeit posing new and unexplored challenges for the current wireless networks. On top of the demand for new access protocols, the latency-reliability requirements further challenge the existing random access protocols. In this work we investigate the adaptation of a well known modern random access algorithm, Irregular Repetition Slotted ALOHA (IRSA) to IIoT in 5G New Radio. The key contribution of the paper is the proposed system level protocol, Adaptive-Multichannel IRSA, that can fulfill the latency-reliability requirements. On top of this, the definition and solution of the resource allocation problem as a resource efficiency optimization guarantees that the algorithm minimizes the system resources. We show that for a set of specific requirements, AMC-IRSA can fulfill the requirements in a lot resource efficient manner. Lastly, we analyze most critical parameters to consider for integration of IRSA for 5G NR.**

I. Introduction

The release 15 of 3GPP has been finalized in June 2018, officially defining the first standard of 5G New Radio (NR). However, many challenges are left unanswered and require further investigation. As an example, the industrial support of 5G is vertical driven and the vertical push has convinced the standardization to shape 5G to support automation under the label Industrial Internet of Things (IIoT). The IIoT challenges the entire communication system with a change of paradigm. Devices are sporadically active with small data packets and have low latency and high reliability constraints. One of the key hurdle to overcome this challenge is to devise an efficient resource allocation procedure [1].

The signaling overhead for resource allocation has been shown to be a bottleneck for 4G [2]. A natural solution to limit signaling is to use random access for data transmission [3]. This solution is labeled as grant-free access in the standardization [4]. It is a promising approach for future wireless networks, as random access has been proven to be efficient for sporadic activity in satellite [5] and maritime networks [6].

On the one hand, evolution of ALOHA and slotted ALOHA, collected under the name of modern random access, are able to drastically improve the resource efficiency [7], [8]. The two pillars of modern random access protocols are time diversity at the transmitter and successive interference cancellation (SIC) at the receiver. On the other hand, these protocols are not designed for low latency constraints, and it is unknown whether their promising resource efficiency would still hold with tight constraints.

In this work we seek an answer to this question by putting Irregular Repetition Slotted ALOHA (IRSA) [8] under the magnifying glass. IRSA is a medium access protocol (MAC) that have been shown to reach a normalized throughput of 1 packet/slot under the idealized collision channel model and for asymptotically large delay and the number of repetitions [9]. By exploiting the large bandwidth of NR, we design a *multichannel IRSA* (MC-IRSA), adapting it to the NR frame structure and IIoT latency-reliability constraints.

Our contribution are summarized as follows:

1) We propose a system level adaptation of multichannel IRSA protocol (MC-IRSA) adapted to fit for the IIoT requirements.
2) We evaluate the resource efficiency of the proposed MC-IRSA with NR frame structure and show that it achieves up to 50-fold resource efficiency increase compared to standardized NR access.

Our main conclusion is: Enabling low latency constraints with multichannel-IRSA (MC-IRSA) provides higher resource efficiency than the standardized solution. However, the achieved resource efficiency is still lower compared to that of IRSA in delay tolerant scenarios. This loss of efficiency stems from tight delay constraints and the time varying behavior of IIoT, and the resource efficiency is increasing when the latency constraint is relaxed.

Section II introduces the scenario and the algorithmic and 5G background for the manuscript. Section II-D introduce the multichannel adaptation for IRSA. Section IV introduces the simulations and discusses the results. Section V concludes the work.

A. Related Work

Integration of Successive Interference Cancellation (SIC) into 5G has been attracting attention recently. Most of these works focus on intra-slot SIC, also called as K-Multipacket reception (K-MPR). Authors have investigated the use of K-MPR with a reactive decision called half and double for re-transmissions [10]. Another work investigated the multichannel adaptation of K-MPR [11]. Further work have considered the use of estimation techniques to improve performance of K-MPR under 5G constraints [12], [13]. However, use of intra-slot SIC disregards the latency constraints and further MAC algorithms have to be imposed on top to guarantee latency requirements.

On inter-slot SIC, Contention Resolution Diversity Slotted ALOHA is adapted to 5G in [14]. The work neglects the low latency requirements and focuses on the massive access problem. Adaptation of IRSA for multiple Quality of Service classes is considered with relaxed latency constraints in [15]. None of these works consider tight latency-reliability constraints that poses the challenge for industrial IoT scenario for 5G NR. In this work we provide insights in this direction.

II. Scenario & Background

A cellular based, single cell scenario is considered. There are n_{tot} users attached to the base station where n_{a} are active at any time instance t. The identity of the active users and the number of active users are unknown. However, the distribution of the active users n_{a} is known and obeys a Beta or a Poisson distribution. The users have latency L and reliability R constraints, such at least $n_{\text{a}} \cdot$ R of all IIoT users are served successfully within L seconds. As IIoT devices are foreseen as low-cost devices we expect to have a single radio on each device. This limits the transmission of a device to a single IRSA-frame in one of the channels at any time instance.

We take our system parameters from the 5G NR standard [16]. The system is composed of orthogonal resources build out of time and frequency in units of Hertz (Hz·s). For example, 14 symbols can fit in 1 ms of 15 kHz or 125 μs of 120 kHz bandwidth . Bandwidth of each frame is fixed to Subcarrier-Spacing (ScS) and denoted as ΔB. Even though symbol level scheduling is possible in NR we use slot based scheduling such that the minimum duration of a resource is a slot and lasts t_s seconds. Slots can be grouped together to form a subframe. The number of slots in a subframe is defined by the ScS as denoted in Tab. I and varies from 1 to 16. 10 subframes form a frame. Duration of one subframe and frame is fixed to 1 ms and to 10 ms respectively. 5G NR frame structure is illustrated in Fig. 1. Orthogonal frames in different subcarriers are called channels. The bundling of orthogonal frames is called an F-channel frame, e.g., 4-channel frame.

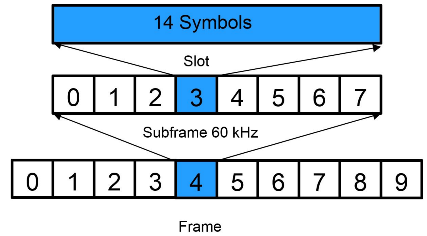

Fig. 1: 5G NR frame structure: Each frame hosts 10 subframes, and each subframe hosts from 1 to 16 slots with respect to the bandwidth from 15 kHz to 240 kHz. And each slot has 14 OFDM symbols.

BW (kHz) ΔB	15	30	60	120	240
Slots per frame m	10	20	40	80	160
Slot dur. t_s (ms)	1	0.5	0.25	0.125	0.0625

TABLE I: 5G NR numerology for subcarrier spacing and slot duration.

The total amount of resources used per frame in units of (Hz·s) to serve n_{a} users is denoted as u. Resources required to serve one packet are assumed to be $u_0 = 15$ (Hz·s) that can also be written as $u_0 = \Delta B \cdot t_s$ as multiplication of a slot duration and the bandwidth. Assuming the expected UL spectral efficiency of 15 bps/Hz [17], this results in a payload of 30 bytes as foreseen for the IIoT [18]. The number of slots in an IRSA-frame is denoted as m. The duration of an IRSA-frame is fixed to a NR frame with a duration of $t_f = 10$ ms.

We assume no control channel overhead as our focus is the efficiency of the data channel. As control is an orthogonal overhead, this can be added to our evaluations in a further work. We assume a collision channel such that if a user transmits its packet in a resource without any contention from other users it is successfully received by the base station, otherwise it is lost in the collision. The interference is called a collision and considered a waste if no SIC is deployed. Thanks to SIC these collisions can be recovered under certain circumstances. We assume ideal interference cancellation i.e., no residual power is left from recovered and cancelled packets. This is a realistic assumption if the received power is relatively high compared to the noise power [19]. As it forms the bottleneck we only investigate the uplink portion of the resources.

A. Irregular Repetition Slotted ALOHA (IRSA)

Evolutions of slotted ALOHA have shown how the replication of MAC packets at physical layer intertwined with the use of SIC at the receiver side, drastically boosts the performance of random access protocols. It was already shown in the early 1980s that the transmission of multiple physical layer copies (called replicas in the following) of the MAC packets can improve the packet error rate, but only when the traffic is sufficiently small [20]. Only the coupling with SIC at the receiver unleashed the full potential of modern random access, as it has been firstly shown by CRDSA and its evolutions. The observation that SIC tightly resembles decoding of low density parity check codes over the binary erasure channel provides the mathematical foundations for a deeper understanding of the iterative process and for an optimization tool [8]. In the quest for higher efficiency, the use of coding based optimization techniques suggests that users shall transmit according to a *degree distribution* – with the degree representing the number of replicas transmitted. The compact form of the degree distribution is given by a polynomial as, e.g., $0.5x^2 + 0.28x^3 + 0.22x^8$, where the each term represents the probability of transmitting a number of replicas equal to the degree. In the example, 50% of the transmitters will send 2 replicas, 28% 3 replicas, and the remaining 22% 8 replicas. For asymptotically long frames and asymptotically large maximum degree, IRSA is able to achieve the efficiency of 1 packet per slot, under the collision channel model, and reaching the same performance of orthogonal medium access protocols.

B. Arrival

The probability that a user is activated at time instant $t \in [0, T_A]$ is given by

$$p(t) = \frac{t^{\alpha-1}(T_A - t)^{\beta-1}}{T_A^{\alpha+\beta-1}\text{Beta}(\alpha,\beta)} \tag{1}$$

where $\text{Beta}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$, α and β are shape parameters, and T_A is the activation time. We consider the 3GPP model [21], where $\alpha = 3$, $\beta = 4$ and $T_A = 10$ s.

C. Time Instance

There is an interplay between the latency constraint and the IRSA-frame duration. A user is not allowed to join an IRSA-frame after it has started. As we want to cover the worst-case we assume that a user is activated just after the beginning of an IRSA-frame such that the user may wait a full frame duration t_f. Depending on the latency constraint, there may be a waiting time t_w between two IRSA-frames. Thus the accumulation time t_a before an IRSA frame starts is $t_a = t_f + t_w$. After that the user transmits in the IRSA-frame that lasts for another t_f. We consider the worst-case scenario and the decoding of the user, happens directly after the end of the IRSA-frame. The timing is illustrated in Fig. 2. Hence, the maximum latency l_i of the user i can be decomposed as

$$l_i = t_w + 2 \cdot t_f = t_a + t_f \leq \text{L}. \tag{2}$$

The latency constraint L should be always larger or equal to the maximum latency of any user.

From the beginning of one IRSA frame until the beginning of the next frame all users are accumulated. This duration is called the accumulation time t_a. t_a can be used to convert the continuous Beta distribution to a discrete distribution as in

$$P[t_j] = \int_{t_j \cdot t_a}^{(t_j+1)t_a} p(t)\, dt. \tag{3}$$

where t_j is the discrete time-steps during the activation time of the Beta arrivals $t_j \in [1, \cdots, \lceil T_A/t_a \rceil]$.

Furthermore, the discrete activation probability can be used to calculate the probability of having n_a active users at time step t_j

$$\Pr[N_\text{a} = n_\text{a}|t_j] = \binom{n_\text{tot}}{n_\text{a}} P[t_j]^{n_\text{a}}(1 - P[t_j])^{n_\text{tot}-n_\text{a}} \tag{4}$$

where the total number of users n_tot is known. As the actual distribution is time-dependent here we approximate it with a binomial distribution. And the total number of users can vary slightly.

D. Problem

We aim to minimize the amount of system resources used u such that the delay l_i and reliability r_i of each user i from the set of active users $\mathcal{N}_a$ fulfills the constraints. The number of slots in an IRSA frame m and the number of parallel IRSA frames F are used by the scheduler at the base station to allocate resources for IIoT users. We can formulate the resource allocation as an optimization problem with the goal of maximizing the resource efficiency ρ as the utility

$$\arg\max_{F,m} \ \rho, \tag{5}$$

$$\text{s.t. } n_\text{a} \leq n_\text{tot}, \tag{6}$$

$$r_i(l_i) \geq \text{R(L)} \quad \text{where } i \in \mathcal{N}_a. \tag{7}$$

The resource efficiency ρ is defined as the minimum amount of resources required to serve one packet u_0 times the total number of successful users calculated by multiplying the number of users n_a with the expectation of reliability for all users $\mathbb{E}[r]$, divided by the total resources u used to serve all users,

$$\rho = \frac{u_0 \cdot n_\text{a} \cdot \mathbb{E}[r]}{u}. \tag{8}$$

The total number of resource can be written as

$$u = F \cdot \Delta B \cdot t_s \cdot m, \tag{9}$$

where F is the number of parallel IRSA-frames and called channels multiplied by the bandwidth and the total duration of a frame $m \cdot t_s$. Through this we can simplify the resource efficiency to

$$\rho = \frac{n_\text{a} \cdot \mathbb{E}[r]}{F \cdot m}. \tag{10}$$

III. Proposed Protocol

The challenge of adapting IRSA for IIoT is two-fold: (1) IRSA is designed for Poisson traffic i.e., constant mean arrival whereas IIoT is expected to behave in a bursty manner reflected by the Beta distribution that has a time-varying mean. (2) IRSA reaches its peak performance without latency constraints.

As described in Sec. II-A IRSA is designed for reliability constraints but not latency constraints. One natural way to enforce latency constraints on top of IRSA is to limit the frame size. However, it is also known that with decreasing frame size, less users can be supported such that some user have to be rejected from the system but that is not an acceptable solution for many critical cases. As the resources in time are limited due to latency constraints, resources orthogonal to time such as bandwidth can be used to accommodate more users. Bandwidth can be beneficial in two ways: (1) for multi-packet reception as considered in the work [22]; (2) through deploying orthogonal resources in time to deploy multiple frames in parallel. As the former increases the complexity of the receiver with intra-slot interference cancellation on top of the inter-slot interference cancellation, we focus on the latter one.

A. Multichannel IRSA (MC-IRSA)

The MC-IRSA requires a small modification to IRSA. The base-station broadcasts the number of channels F and the starting and the ending frequency of the bandwidth that is used by the MC-IRSA. After the users select the number of repetitions with respect to a degree distribution, they select uniformly one of the F channels. The number of channels F is static. This allows the system to distribute users in a uniform fashion over multiple channels. The F is decided as the output of the optimization problem in Eq. (7).

The n_{a} is taken as the maximum of the all arrival means of the time-instances of the Beta arrival

$$n_a = \max(\mathbb{E}[n_a|t_s]) \ \forall t_s \text{ where } t_s \in [1, T_A]. \quad (11)$$

There is varying number of users in each orthogonal IRSA-frame, since the selection of channels is uniformly random. As we do not have a deterministic number of users but a distribution and we have to decide the frame size in a more conservative manner. This results in a decreased resource efficiency.

B. Adaptive MC-IRSA (AMC-IRSA)

The Beta distribution is a time-varying function where the mean number of active users at a time instance varies. If the number of channels F is selected static for the IIoT the resource efficiency suffers significantly as shown in Sec. IV. Thus, in AMC-IRSA, we dynamically adapt the number of channels F to the expected number of arriving users in every slot, to maximize the expected resource efficiency. This relaxes the maximum operation in Eq. (11) as the optimization is done for each time step t_n.

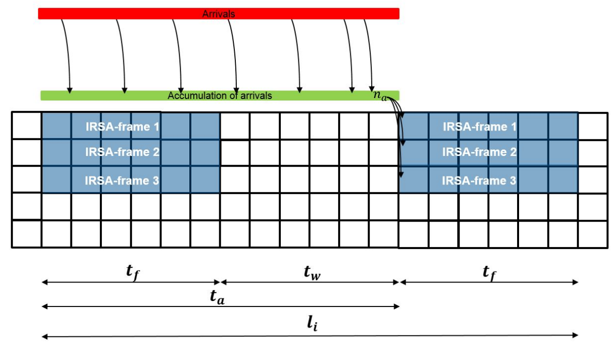

Fig. 2: The interplay between the frame duration t_f, the latency constraint L, the time instance of the discrete arrival distribution t_s and the accumulation time t_a.

IV. Simulation Results

We have re-used an open-source simulator for IRSA [23] and designed a system level discrete-time based simulator around it in MATLAB[1].

We have varied F from 1 to 20 as the bandwidth is limited in a practical system and the number of slots as $m \in \{10, 20, 40, 80, 160\}$ representing the number of slots in a 5G frame corresponding to different ScS. We have used the degree distribution $0.5x^2+0.28x^3+0.22x^8$ for each IRSA-frame unless said otherwise.

Varying ΔB changes the number of slots m in a frame. The same is true for also changing the frame duration t_f so it is kept constant.

A. Behavior of ρ

We first investigate the behavior of the resource efficiency function with respect to varying F and m. This helps us identify the characteristics of the function that can be used to simplify the optimization problem. For the Beta arrival with $n_{\text{tot}} = 30000$ and activation time of $T_A = 10$ s, latency constraint of $\text{L} = 20$ ms, IRSA-frame duration of $t_f = 10$ ms and th resulting waiting time of $t_w = 0$ ms.

The behavior of ρ with respect to both design parameters is illustrated in Fig. 3. In the plot we see that a non-convex behavior is observed for both varying m and F. This eliminates the possibility that we optimize for one parameter and optimize the other one, simplifying the problem. A joint optimization has to be used.

B. Poisson and Beta Arrivals

We consider the Beta arrivals as proposed by 3GPP for machine type communications [21]. The Poisson arrivals are used to demonstrate the critical difference

[1] https://github.com/tum-lkn/IRSA_4_5G

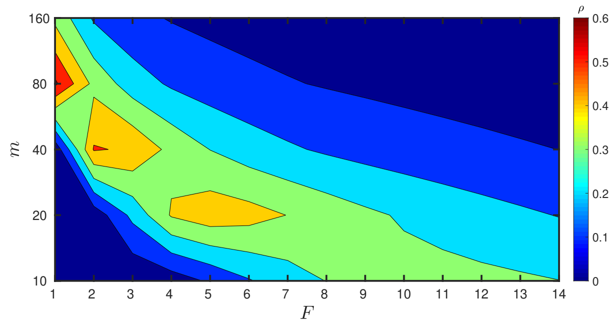

Fig. 3: Behavior of the resource efficiency function ρ for Poisson traffic with respect to F and m. x-axis depicting the number of channels F versus y-axis depicting the number of slots m for MC-IRSA.

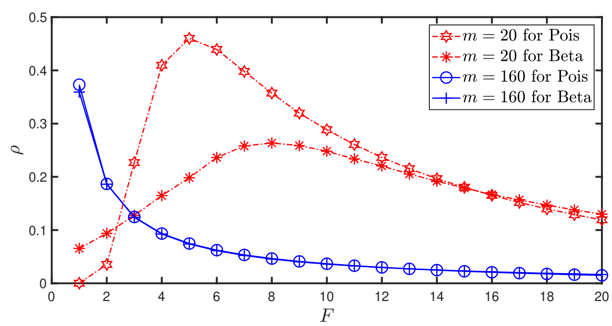

Fig. 4: x-axis depicting the number of channels F versus y-axis depicting resource efficiency ρ for MC-IRSA. Legend shows different arrival type and different frame size in unit of slots.

of Beta to Poisson arrivals. In order to demonstrate the limitations due to the Beta arrivals, we have simulated Poisson and Beta arrivals in the same settings. To make a comparable scenario we have adjusted to have on average the same number of arrivals 60 for each IRSA-frame. We have used different $\Delta B = \{20, 160\}$ kHz and evaluated throughput for different number of channels $F = \{1, \cdots, 20\}$. t_f is set as 20 ms and t_w as 10 ms. The resulting sweep analysis is illustrated in Fig. 4. The x-axis denotes the number of channels F versus the y-axis denoting the resource efficiency ρ.

We observe two different behavior, no difference for high number of slots m and a big gap for low m. For low m Beta can at most achieve half of what is possible with Poisson arrivals. This clearly demonstrates the impact of time-varying characteristics of Beta distribution. At certain time-instances there are almost no users compared to time-instances where there is more than twice the mean arrivals. And as we do not adjust the number of channels accordingly this results in low resource efficiency.

For both cases with low F the resource efficiency ρ is better or equal to Poisson. However, with Poisson distribution with high F a higher ρ is achieved. We want to enable the same for the Beta arrivals. Thus, we deploy an optimal F for each time step t_j, to react to the time variation of the Beta arrival.

C. AMC-IRSA

In the sweep analysis the reliability constraints is not enforced. In AMC-IRSA we enforce a reliability constraint of $R = 0.9$ as foreseen for some IIoT applications. The resource efficiency is calculated through simulations for varying F with 50 repetitions and the F that fulfills the constraints and maximized ρ for that specific m value is selected. Hence, the F dimension is not evaluated. The t_f is fixed to 10 ms, while $t_w \in \{0, 20, 60, 140\}$ is varied to meet the increasing latency constraint that is varied as $L \in \{10, 20, 40, 80, 100\}$ ms.

In Fig. 6 the y-axis depicts the resource efficiency ρ while the x-axis depicts the varying latency constraint. We see that with the same latency value of 10 ms the throughput have improved two-folds compared to MC-IRSA in Fig. 4 thanks to time varying adaptation of F. Varying the latency constraint emphasized that different frame sizes are optimal for different latency constraints.

Overall with the non-optimized degree distribution with increasing latency constraint the ρ is increased. This is expected as the selected degree distribution is optimized for asymptotic number of slots in an IRSA-frame.

With low latency constraints $m = 20$ provides the best resource efficiency while $m = 80$ is optimal for increasing latency constraint. IRSA provides higher efficiency than conventional random access as depicted with Slotted ALOHA (SA) curves. The SA with reliability constraints has a relatively fixed throughput with varying latency constraints around 0.013. The red line shows the optimal bandwidth selection for each latency constraint.

We can further investigate the reasons for a lower resource efficiency through looking at the timely variation of it. The resource efficiency per time step $\rho(t_j)$ is illustrated in Fig. 5 with the number of active users per time step $n_a[t_j]$. The x-axis depicts all time steps of a single Beta arrival. Latency requirement is set as $L = 10$ ms and the number of slots is fixed to $m = 20$ as optimal for that latency constraint. The results are averaged over multiple runs and we see that until a certain average number of active users, e.g., $n_a[t_j] > 10$ is reached, AMC-IRSA has a ramp-up time. This effect may be stemming from fixing our number of slots to 5G frame structure and can be overcome, if we allow lower number of slots for these regions.

D. Degree Distribution

In order to evaluate the effect of degree distribution on the resource efficiency we have evaluated various degree distributions summarized in Tab. II. The number of slots is fixed to $m = 20$, the reliability constraint is set as $R = 0.9$. The previous degree distribution $\Lambda_1(x)$ is plotted as a red line, while other distributions are illustrated with markers. Overall we see that improvement is limited to 2%. With low latency constraints, low maximum replicas

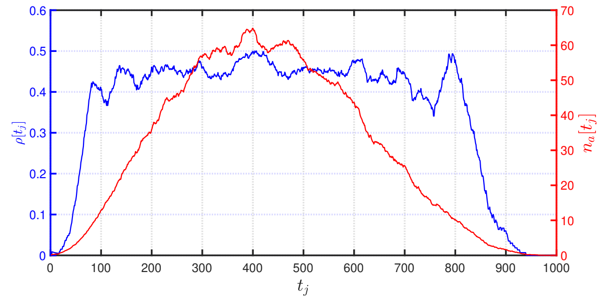

Fig. 5: The change of resource efficiency with each time step t_j of the Beta arrival superposed to number of active users $n_a[t_j]$ at each time step.

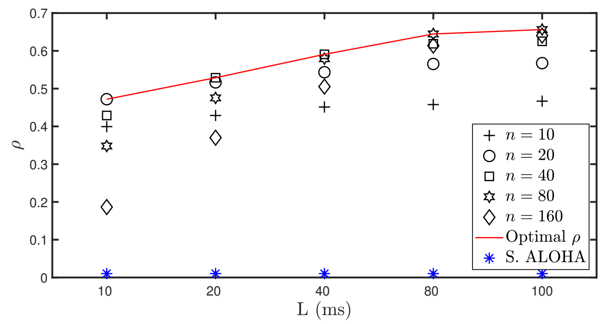

Fig. 6: Varying latency constraint L versus resource efficiency ρ for AMC-IRSA with R = 0.9

Degree Distribution
$\Lambda_1(x) = 0.5x^2 + 0.28x^3 + 0.22x^8$
$\Lambda_2(x) = 0.7x^2 + 0.3x^3$
$\Lambda_3(x) = x^3$
$\Lambda_4(x) = 0.5102x^2 + 0.4898x^4$
$\Lambda_5(x) = 0.4977x^2 + 0.2207x^3 + 0.0381x^4 + 0.0756x^5 + 0.0398x^6 + 0.0009x^7 + 0.0088x^8 + 0.0068x^9 + 0.003x^{11} + 0.0429x^{14} + 0.0081x^{15} + 0.0576x^{16}$

TABLE II: Degree distributions evaluated for performance improvement.

perform better, while with increasing latency constraint distributions with higher replicas perform better. Overall fixed number of re-transmissions of 3 perform the best with varying latency constraint.

For further improvement finite length analysis in [24] can be used to optimize degree distributions. But as the limitations is mostly from the finite length behavior and the time varying characteristics, we project that higher gains is not expected.

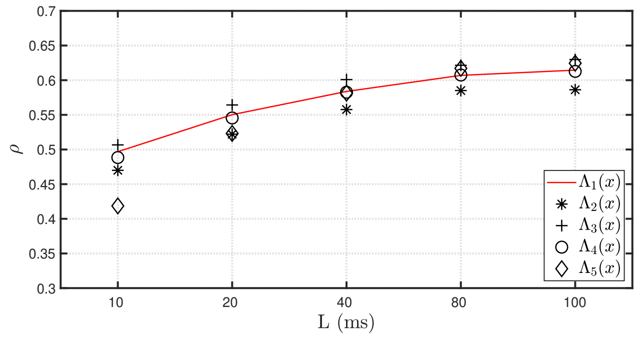

Fig. 7: Resource efficiency ρ versus varying latency constraint L for AMC-IRSA. Different degree distributions are evaluated against previously considered degree distribution Λ_1.

V. Conclusion

In this paper we have proposed the adaptive multichannel IRSA (AMC-IRSA) as a protocol to fulfill latency-reliability constraints under 5G New Radio. AMC-IRSA uses the extra bandwidth a lot more efficiently compared to state of the art algorithms. This result is important as this means AMC-IRSA needs 2% of the resources that state of the art algorithm would need. We have formulated and solved an optimization problem to guarantee the maximum resource efficiency is maintained via optimizing parameters of the algorithm for various latency-reliability constraints.

However, in this work we have simplified the physical layer and related assumptions to be further investigated. But as the results show a huge performance increase, we project that, even when the PHY aspects are taken into account, IRSA is ready to replace current algorithms for latency-reliability requirements.

References

[1] P. Popovski, J. J. Nielsen, C. Stefanovic, E. De Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park *et al.*, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *Ieee Network*, vol. 32, no. 2, pp. 16–23, 2018.

[2] R. Ratasuk, A. Prasad, Z. Li, A. Ghosh, and M. A. Uusitalo, "Recent advancements in m2m communications in 4g networks and evolution towards 5g," in *International Conference on Intelligence in Next Generation Networks*, Feb 2015, pp. 52–57.

[3] H. M. Gürsu, M. Vilgelm, A. M. Alba, M. Berioli, and W. Kellerer, "Admission Control based Traffic-Agnostic Delay-Constrained Random Access (AC/DC-RA) for M2M Communication," *IEEE Transactions on Wireless Communications*, 2019.

[4] 3GPP, "Discussion on the reliability enhancement for grant-free transmission," Tech. Rep., Aug. 2018.

[5] E. Casini, R. De Gaudenzi, and O. D. R. Herrero, "Contention resolution diversity slotted aloha (crdsa): An enhanced random access schemefor satellite access packet networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1408–1419, 2007.

[6] F. Clazzer, F. Lázaro, and S. Plass, *New receiver algorithms for satellite AIS.* Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2015.

[7] Y. Yu and G. B. Giannakis, "High-throughput random access using successive interference cancellation in a tree algorithm," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4628–4639, 2007.

[8] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted aloha," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, 2011.

[9] K. R. Narayanan and H. D. Pfister, "Iterative collision resolution for slotted aloha: An optimal uncoordinated transmission policy," in *2012 7th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, Aug 2012, pp. 136–139.

[10] L. Zhao, X. Chi, L. Qian, and W. Chen, "Analysis on latency-bounded reliability for adaptive grant-free access with multipackets reception (mpr) in urllcs," *IEEE Communications Letters*, 2019.

[11] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-based access for ultra-reliable low latency uplink transmissions," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 182–185, 2018.

[12] A. Jacquelin, M. Vilgelm, and W. Kellerer, "Grant-free access with multipacket reception: Analysis and reinforcement learning optimization," in *IEEE/IFIP Wireless On-demand Network systems and Services Conference (WONS)*. IEEE, 2019, pp. 1–8.

[13] H. M. Gürsu, W. Kellerer, and C. Stefanović, "On throughput maximization of grant-free access with reliability-latency constraints," *IEEE Int. Conf. on Communications (ICC)*, 2019.

[14] A. Azari, P. Popovski, G. Miao, and C. Stefanovic, "Grant-free radio access for short-packet communications over 5g networks," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–7.

[15] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "On sinr-based random multiple access using codes on graph," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[16] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2018.

[17] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5g networks: Who will be the heir of ofdm?: An overview of alternative modulation schemes for improved spectral efficiency," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 80–93, 2014.

[18] E2E-aware Optimizations and advancements for Network Edge of 5G New Radio (ONE5G), "Deliverable D2.1: Scenarios, KPIs, use cases and baseline system evaluation," Tech. Rep., Nov. 2017.

[19] S. Ataman and A. Wautier, "Capacity and asymptotic user capacity of a multiservice cdma uplink with successive interference cancellation," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2006, pp. 1–5.

[20] G. Choudhury and S. Rappaport, "Diversity aloha - a random access scheme for satellite communications," *IEEE Transactions on Communications*, vol. 31, no. 3, pp. 450–457, 1983.

[21] 3rd Generation Partnership Project (3GPP), "RAN WG2 #71 R2-104663: MTC LTE simulations," Madrid, Aug. 2010, Tech. Rep., 2000. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg–R2-71–28035.htm

[22] Č. Stefanović, E. Paolini, and G. Liva, "Asymptotic performance of coded slotted aloha with multipacket reception," *IEEE Communications Letters*, vol. 22, no. 1, pp. 105–108, 2018.

[23] "Irsa simulation," https://github.com/afcuttin/irsa, accessed: 2019-03-30.

[24] A. G. i Amat and G. Liva, "Finite-length analysis of irregular repetition slotted aloha in the waterfall region," *IEEE Communications Letters*, vol. 22, no. 5, pp. 886–889, 2018.