Generalization of probabilistic scheduling models for realizing URLLC applications

Suyong Eum, Shin'ichi Arakawa, and Masayuki Murata
Osaka University, Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan
Email:{suyong, arakawa, murata}@ist.osaka-u.ac.jp

***Abstract*—In this paper, we mathematically generalize probabilistic Grant Free (GF) scheduling models so that they can be easily extended to deal with various problems in URLLC use cases. To demonstrate the extensibility of the generalized models, we introduce the reliability parameter α to the models and show that the extended models can be more rigorous against failure scenarios, which provides more flexible options in the design of GF scheduling.**

***Index Terms*—URLLC, 5G, Grant-Free scheduling**

I. Introduction

Grant Free (GF) scheduling models have appealed to 5G research community because of their intuitive approach to deal with the tight latency requirement in various URLLC use cases. In [1], the authors proposed probabilistic GF scheduling models which calculate the success probability of data transmission from UE to gNB with given allocated radio resources. In detail, the models determine how many radio resources should be allocated to UEs so as to transmit data to gNB successfully within the reliability requirement, 1-10^{-5}. However, the GF scheduling models in their paper were described only with examples so the extensibility of the models is limited. For this reason, we mathematically generalize the GF scheduling models so that they can be easily extended to deal with various problems in URLLC use cases.

II. A GF scheduling model without early stopping and with early stopping

GF scheduling models pre-allocate resources for UEs so that UEs can instantly transmit data when it is required. Since the pre-allocated resources can be wasted due to the uncertainty of future resource demand, the resources are generally shared among UEs to be used more efficiently. However, the shared resources cause a contention collision problem among UEs, which results in transmission failures. Thus, K redundant transmission was introduced to GF scheduling models to deal with the transmission failures [2]. One obvious problem raised by the K redundant transmission is again the waste of resources. Thus, the concept of early stopping was introduced to the K redundant transmission approach.

"With early stopping" means that, instead of whole K consecutive redundant transmissions, UE stops the transmission in the middle upon receiving an acknowledgement from gNB. Similarly, the K redundant transmission without an acknowledgement scheme is called "without early stopping".

TABLE I: Notation table for GF scheduling models

Nota.	details
${}_s\mathbb{P}$	success probability of data transmission
${}^{DR}_{f}P_i$	failure probability at dedicated resource at i^{th} transmission
${}^{SR}_{f}P_i$	failure probability at shared resource at i^{th} transmission
${}_fP_i$	expected failure probability at i^{th} transmission
N	number of UEs in the cell
N_1	number of minislots for dedicated resources
N_2	number of minislots for shared resources
M	number of shared resources per minislot
K	total number of redundant transmissions, e.g., K=N_1+N_2
λ	probability that UE is on or active, e.g., arrival rate [0-1]

A. A generalized GF scheduling model without early stopping

$$
\begin{aligned}
{}_s\mathbb{P} &= \sum_{n=1}^{N} \binom{N-1}{n-1} \lambda^{n-1}(1-\lambda)^{N-n} \left\{ 1 - \prod_{i=1}^{N_1} {}^{DR}_{f}P_i \prod_{j=1}^{N_2} {}^{SR}_{f}P_j \right\} \\
&= \sum_{n=1}^{N} \binom{N-1}{n-1} \lambda^{n-1}(1-\lambda)^{N-n} \\
&\quad \left\{ 1 - \prod_{i=1}^{N_1} {}_fP_i \prod_{j=1}^{N_2} \left[1 - \Big(1 - {}_fP_j\Big) PS(\alpha, M, n) \right] \right\}
\end{aligned} \tag{1}
$$

where

$$PS(\alpha, M, n) = exp^{(1-n)/\alpha M} \tag{2}$$

The notation used in Equ.(1) is shown in Table I. This model has two main parts; one in the left-hand side showing the cumulative binomial distribution which calculates the probability that a certain number of UEs are active (transmitting data) out of the whole UEs. The other in the right-hand side has two production notations which calculate the failure probability at the dedicated and the shared resources. Thus, the success probability is simply calculated by subtracting the multiplication of whole failure probabilities from one. Different from [1], this model includes a new parameter α representing the reliability of the system, which enables to subtly control the model and can make the model more rigorous.

$$
\begin{aligned}
{}_s\mathbb{P} &= \sum_{n=1}^{N}\binom{N-1}{n-1}\lambda^{n-1}(1-\lambda)^{N-n}\left\{\sum_{k=1}^{N_1} {}^{DR}_{\;s}P_k + \sum_{k=1}^{N_2} {}^{SR}_{\;s}P_k\right\} \\
&= \sum_{n=1}^{N}\binom{N-1}{n-1}\lambda^{n-1}(1-\lambda)^{N-n}\left\{\underbrace{\sum_{k=1}^{N_1}(1-{}_fP_k)\prod_{i=1}^{k}{}_fP_{k-i}\big|_{k\neq i}}_{\text{Success probability at dedicated resources}} + \underbrace{\left[\sum_{k=1}^{N_2} G(\alpha,M,n,F(\alpha,M,n,k))F(\alpha,M,n,k)\right]}_{\text{Success probability at shared resources}}\right\}
\end{aligned} \tag{3}
$$

where

$$
\begin{aligned}
G(\alpha,M,n,F(\alpha,M,n,k)) &= \Big(1-{}_fP_{N_1+k}\Big)\sum_{j=1}^{n}\binom{n-1}{j-1}\Big[F(\alpha,M,n,k)\Big]^{j-1}\Big[1-F(\alpha,M,n,k)\Big]^{n-j} PS(\alpha,M,j) \\
F(\alpha,M,n,k) &= \begin{cases} \prod_{i=1}^{N_1} {}_fP_i & \text{if k}\in 1 \\ \prod_{i=1}^{N_1} {}_fP_i \prod_{j=1}^{k-1}\Big[1-G(\alpha,M,n,F(\alpha,M,n,j))\Big] & \text{otherwise} \end{cases}
\end{aligned} \tag{4}
$$

B. A generalized GF scheduling model with early stopping

The model is based on the geometric distribution which determines the probability of consecutively observing non-relevant events before actually you observe a relevant one. In our case, it shows the probability that failures occur $k-1$ times before the first success takes place at k^{th} transmission. The equation calculates the sums of all cases, e.g., (S_1 + F_1S_2 + $F_1F_2S_3$,...,+$F_1F_2 \ldots S_K$). The model becomes exponentially complicated as the number of redundant transmissions, K increases, especially in the calculation of the success probability at shared resources. The success probability at shared resources is composed of two functions; $G(\cdot)$ and $F(\cdot)$ shown in detail in Equ.(4). The former represents the success probability taken place at k^{th} transmission and the latter represents the multiplication of the failure probabilities taken place from 1^{st} to $(k-1)^{th}$ transmissions.

In Fig. 1, "Utilization ratio" on Y-axis represents the total dedicated and shared resources which are allocated to UEs in order to achieve the success probability, 1-10^{-5}. Then, the value is simply normalized by the maximally allocated resources. The result with a dot line well matches to the key result, (Fig.4 in [1]). As aforementioned, α represents the reliability of a system, for example, when its value is small, the model needs to allocate more resource in order to achieve the same level of success probability. For this reason, the model with smaller α value has the higher utilization ratio.

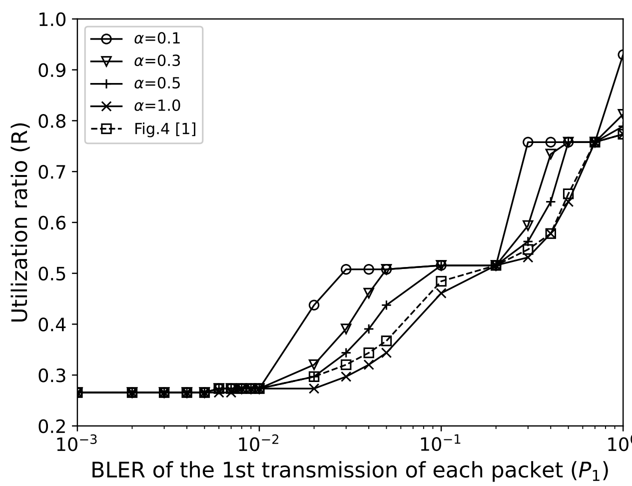

Fig. 1: Full buffer with early stopping, N=32

III. Conclusions

In this paper we mathematically generalized the probabilistic GF scheduling models so that the models can be easily extended. Then, we extended the models by introducing the reliability parameter α to the models so as to demonstrate their extensibility. Due to the reliability parameter α, the extended models can be more rigorous against failure scenarios, which provides more flexible options in the design of GF scheduling. Currently, we have been further extending these GF scheduling models in a way that it takes into account retransmissions and some other parameters representing channel conditions.

IV. Acknowledgement

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011.

References

[1] Z. Zhou, R. Ratasuk, N. Mangalvedhe, and A. Ghosh, "Resource Allocation for Uplink Grant-Free Ultra-Reliable and Low Latency Communications," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. Porto, Portugal: IEEE Press, June 2018.

[2] "NR: Physical Layer Procedures for Data;," 3GPP: Sophia Antipolis, France, 2018; Release 15, TS 38.214 V15.4.0.