Pre-scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications

Abreu, Renato Barbosa; Mogensen, Preben Elgaard; Pedersen, Klaus I.

Published in:
2017 IEEE Wireless Communications and Networking Conference (WCNC)

DOI (link to publication from Publisher):
10.1109/WCNC.2017.7925877

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Pre-scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications

Renato Abreu*, Preben Mogensen*†, Klaus I. Pedersen*†
*Dept. of Electronic Systems, Aalborg University; †Nokia Bell Labs, Aalborg, Denmark
Email: rba@es.aau.dk

***Abstract*—The fifth generation (5G) cellular network demands new solutions to meet, in an efficient way, the stringent targets for ultra-reliable and low latency communication (URLLC), such as 1-10^{-5} reliability within 1 ms. In a wireless system, the control signaling of the scheduling process is also a source of errors and delays. Semi-persistent scheduling (SPS) is an option to reduce the signaling, leading to lower latency and improved transmission reliability. However, conventional SPS still applies grant signaling to schedule the retransmission. In this work it is proposed an alternative scheme in which a group of users shares a pre-scheduled resource for retransmission. The benefit is that it provides a retransmission opportunity without needing a scheduling control information. Besides that, if the pre-scheduled resource can not be reallocated, the sharing mechanism avoids excessive capacity loss. It is demonstrated through a simple analytical model that, for right grouping sizes and initial transmission error rates, the target error probability e.g. 10^{-5} can be achieved. It is also shown that the suggested scheme can provide improved resource efficiency compared to a single conservative transmission which also avoids re-scheduling.**

I. Introduction

The possibilities opened for the mission critical communication with ultra-reliable and low latency communication (URLLC) in fifth generation (5G) networks, may bring a big amount of novel applications for new markets. Some examples are wireless industry automation, vehicle-to-everything communication (V2X) and remote tactile control [1]. At the same time, big challenges emerge to achieve the stringent requirements needed in these contexts, e.g. 1-10^{-5} reliability within 1 ms and average user plane latency of 0.5 ms [2].

Many applications demand low latency and reliable transmissions of predictable traffic. For instance, machines remotely controlled via Tactile Internet with real-time, synchronous and haptic feedback [3]; and V2X, with broadcast of periodic awareness information in form of Cooperative Awareness Messages [4]. Such machine type communication can generate a significant amount of small packets by a large number of user equipments (UEs). Dynamically scheduling this kind of data at each transmission time interval (TTI) would cause an excessive control signaling overhead. And this, besides being a bottleneck in terms of capacity, is also a source of errors and delays.

Semi-persistent scheduling (SPS) was introduced in LTE standard to support VoIP services, solving the problem of the tight delay requirement for small periodic traffics and the scarcity of control channel resources [5]. In SPS, resources are pre-scheduled with a certain periodicity, to avoid the overhead caused by multiple assignment/grant messages. Recently, SPS has gained more attention in the context of latency reduction considering shortened TTIs and periodicities. It can specially benefit the uplink, as the scheduling request and grant process can be skipped [6]. For URLLC, errors in the data and in the control channels should be strictly avoided in order to meet the tight requirements. In that sense, SPS can bring extra benefits, not only by reducing latency but also the role of the control channel as an error source [7].

The drawback of pre-scheduling is that, typically, the reserved resources can not be used by other UEs, limiting the resource utilization. For URLLC, which requires a very robust transmission, the cost in terms of resources can be very high, specially in bad coverage conditions. So, employ a data retransmission scheme like hybrid automatic repeat request (HARQ) is important to enhance the resource efficiency [8]. Otherwise, a large amount of resources needs to be reserved for each pre-scheduled cycle, for a conservative transmission.

The conventional SPS includes a persistent scheduling for the initial (first) transmission and a dynamic scheduling for the retransmissions (re-scheduling) [9]. For URLLC it may be desired to avoid also the signaling for the re-scheduling due to the possible errors in the control channel. Besides that, extra-latency can be caused by the late re-scheduling in high loaded scenarios and by the grant processing itself.

This paper presents an alternative scheme to provide HARQ retransmission opportunity for URLLC. The basic idea is to have a pre-scheduled resource for retransmission which is shared by a group of UEs. This way, the control signaling used to re-schedule the transmission when it does not succeed, can be suppressed. At the same time, with the sharing of the reserved resource, excessive capacity loss can be avoided. A model for the system is presented to show how the transmission success probability varies depending on the dimensioning of the group and on the initial transmission error rate. The resource efficiency of the system is finally compared with a conservative method that uses a robust modulation and coding scheme (MCS), targeting 10^{-5} error probability in a single transmission (which also avoids re-scheduling).

The rest of the paper is organized as follows: Section II describes the concept of the proposed scheme. Section III presents the system model and the main assumptions. Section IV shows the numeric evaluation regarding the reliability and resource efficiency. Section V finalizes with the main conclusions of this work.

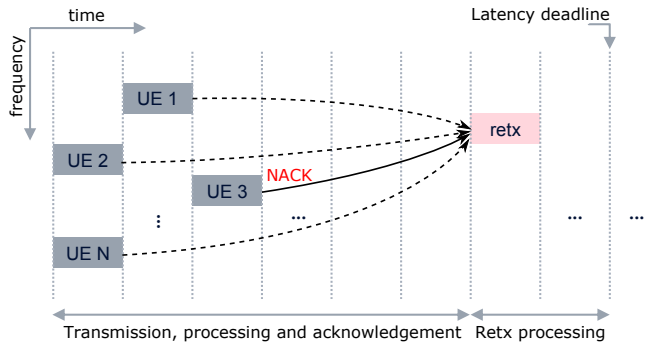

Fig. 1. Pre-scheduled retransmission opportunity shared by UE 1 to UE N.

II. Shared Retransmission Scheme

The basic principle of the shared retransmission opportunity for a group of UEs is illustrated in Fig. 1.

In the proposed scheme the base station (BS) should group and coordinate the UEs with similar traffic characteristics, and configure them to contend for a shared retransmission resource if the initial transmission fails. The grouping and the allocation should aim at a better resource utilization than a conservative transmission. At the same time, it should have a low probability of contention for the retransmission opportunity in order to achieve the target success probability.

The time location of the retransmission resource should allow that the transmission and decoding of the packet is concluded within the latency deadline (the maximum time for a packet to be delivered successfully in the receiver side). It is worth to notice that the initial transmissions of all UEs may not necessarily be aligned in time, as long as the processing and acknowledgment of all transmissions finishes before the reserved retransmission moment. Furthermore, transmitting in different TTIs can permit to accommodate the data packets of UEs in poor channel conditions in the available band during a TTI. Another advantage is to uncorrelate possible errors caused by sudden interference on the grouped UEs.

Both the dedicated resources for the initial transmission and the retransmission resources are pre-scheduled including a certain periodicity according to the traffic pattern. So, retransmissions occur as a synchronous HARQ, at fixed time-intervals. The pre-scheduling configuration can be made through radio resource control (RRC) signaling protected by automatic repeat request (ARQ), like in SPS, so the potential errors on the control channel can be neglected.

The main idea is that, if the initial transmission in the dedicated resource is not decoded, the shared resource can be used for one of the UEs in the group, e.g. UE3 in Fig. 1. A possible implementation in the downlink case is, if more than one UE does not acknowledge on initial transmission, the BS decides to which one it will retransmit on the reserved resource. Only the selected UE can decode the data, while the others will not be able to decode that retransmission resource. In the uplink, the BS can solve the contention by issuing a simple 1-bit signal, or a NACK, only to the UE that should use the retransmission resource. So the collision is avoided in case the retransmission is demanded for more than one UE. This procedure is not susceptible to the granting errors of dynamic re-scheduling because the selected UE knows, from the initial configuration, the time-frequency allocation for the retransmission. Here it is considered that, if the initial transmission fails and the UE does not get the retransmission, the packet is dropped. This is the worst case, considering that there is no available resource, reliable control or time budget for a re-scheduling. The remaining issue is to know how the contention based access to the retransmission resource can provide sufficient reliability.

III. System Model

In this section it is presented a model to estimate the success probability according to the number of UEs in a group and their transmission error probabilities. A formulation for the inherent boundaries of the system is also shown.

A single retransmission opportunity for the group of N UEs during each transmission cycle is considered. This is a reasonable assumption in the context of URLLC since the tight latency requirement may not allow multiple retransmissions. The initial transmission of each UE can randomly fail, then requiring the retransmission. This can be modeled like a Slotted ALOHA process [10] in which the probability of each UE to contend for the retransmission resources, i.e. contention based retransmission, is the probability of failing in the initial transmission P_1. Here, it is assumed that all UEs in the same group have the same error probability target. The probability of the reserved retransmission resources to be idle is given by

$$P_{idle} = (1 - P_1)^N, \tag{1}$$

while the probability of the resource to be required for a single UE is written

$$P_{single} = \binom{N}{1} P_1 (1 - P_1)^{N-1}. \tag{2}$$

Finally, the probability that the retransmission resource is required for more than one UE is simply obtained as

$$P_{collision} = 1 - P_{single} - P_{idle}. \tag{3}$$

In case the retransmission is demanded for more than one UE, the BS can decide which of them gets the reserved resource (the "winner"). So, assuming that each UE has an equal chance to win, the probability of having the packet successfully decoded is then given by

$$P_{success} = (1 - P_1) + \\ P_1(1 - P_2) \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1} (1 - P_1)^{N-n} (1/n), \tag{4}$$

where P_2 is the error probability in the retransmission. It is worth noting that the probability of a grant/assignment error, typical of a dynamic re-scheduling scheme, does not appear in equation (4). That is basically replaced by another term that considers the contention for use the retransmission resource,

which is the summation term in (4). This term depends mainly on the error probability of the first transmission and on the grouping size N. It sets boundaries on the success probability, independent of the error probability of the retransmission (i.e. $0 \leq P_2 \leq 1$), which are written

$$(1-P_1) \leq P_{success} \leq$$
$$(1-P_1) + P_1 \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1} (1-P_1)^{N-n} (1/n). \tag{5}$$

So, there is a clear trade-off between the number of UEs in the group and the maximum success probability. It is important to point out that, for the sake of simplicity to present the main idea, the feedback errors were omitted in the model. However such errors impacts the final success probability of the system, requiring a lower error target on transmissions or smaller groupings, to be compensated.

IV. Performance Analysis

In order to achieve a certain final success probability with the described scheme, the objective is to find the number of UEs that can be grouped and the required success probability for the initial transmission. After that, it is important to quantify the resource efficiency when applying the proposed procedure. A fair comparison can be made with a single conservative transmission, which also does not require a re-schedule signaling, but spends a large amount of resources aiming to succeed with one transmission.

A. Grouping and reliability evaluation

For finding the number of UEs that can be grouped under a certain initial block error rate (BLER, taken as the transmission error probability), the BLER on the retransmission (after the soft combining) is fixed to 10^{-5}, to match with the baseline reliability of the 5G access technologies [2]. Fig. 2 shows the final error probability $(1 - P_{success})$ according to the first BLER for different number of UEs grouped to share the retransmission opportunity. It can be seen that, for instance 21 UEs can be grouped to share one retransmission opportunity when the initial BLER is 10^{-3}. That UEs can still achieve the final target error probability of 10^{-5}, without needing a control signal to re-schedule eventual retransmissions. It can be noticed also that, the higher the number of UEs is a group, the lower should be the BLER on the initial transmission to achieve the target error probability. Since the minimum grouping size is 2, the maximum BLER allowed for the initial transmission to achieve the final error probability of 10^{-5}, is 4.4×10^{-3}. As stated before, instead of a granting error probability in equation (4), there is a summation term which accounts for the probability of winning the retransmission opportunity in case of contention. The complement of that, which is the probability of not getting the retransmission opportunity, is given by

$$P_{notwin} = 1 - \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1} (1-P_1)^{N-n} (1/n). \tag{6}$$

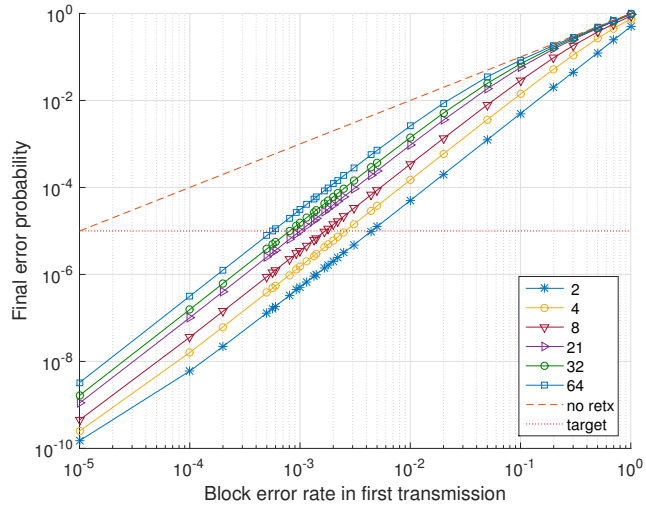

Fig. 2. Reliability according to the first BLER for N UEs.

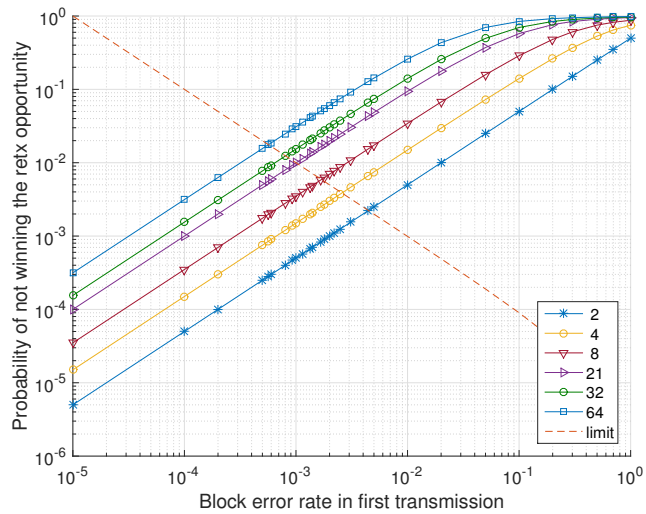

Fig. 3. Probability of not winning on contention for retransmission.

These probabilities are shown for different number of UEs in Fig. 3. The dashed line (limit) represents the maximum value for P_{notwin} in order to achieve less than 10^{-5} final error probability. That is equivalent to the maximum error probability required for the granting in a dynamic re-scheduling scheme. The proposed scheme can operate within the target reliability if the number of UEs in the group and the initial BLER are in the region below the limit line. Taking the intersections with the limit line, the maximum number of UEs at each initial BLER condition can be extracted as shown on Fig. 4.

B. Resource efficiency evaluation

This section shows an estimation of the resource efficiency gain, when comparing the scheme with shared retransmission opportunity against a conservative transmission.

A link abstraction model was used to derive the coding rate needed to achieve each required BLER, when transmitting a packet of 256 bits at a certain signal-to-noise ratio (SNR).

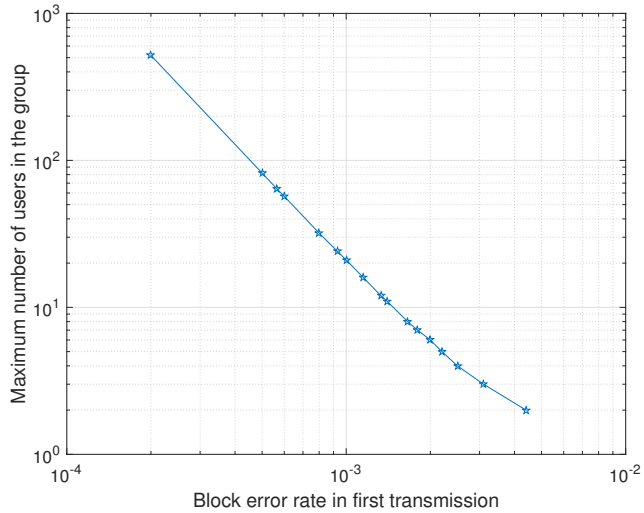

Fig. 4. Maximum grouping according to the first BLER for 10^{-5} error target.

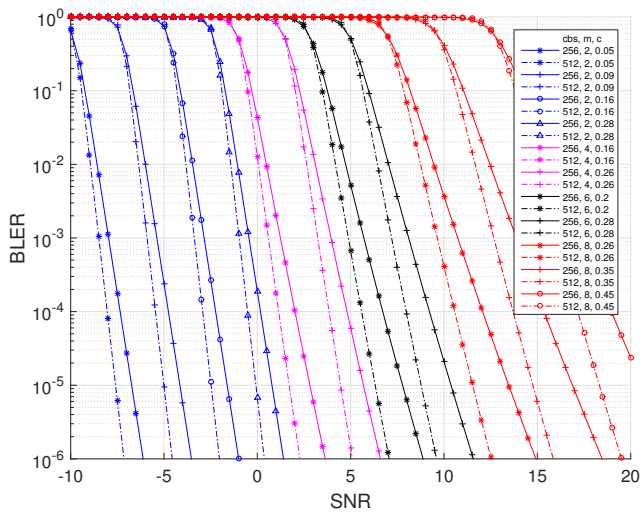

Fig. 5. Example of performance curves from the link abstraction model for different code block sizes (cbs), modulation orders (m) and coding rates (c).

Typical modulation orders were assigned to each SNR interval like: QPSK from -10 to 0 dB, 16QAM from 0 to 5 dB, 64QAM from 5 to 10 dB and 256QAM from 10 dB onwards. The model was obtained considering turbo codes, which is one of the coding schemes proposed for URLLC that has presented better performance for block sizes of 200 bits onwards [11]. Fig. 5 shows some example performance curves of the model for an additive white Gaussian noise (AWGN) channel. It can be noticed in Fig. 5 that, for small packets like 256 bits (baseline packet size for URLLC evaluation [2]), the curves are not as steep as for larger packets, so the modulation and coding rate requirements are more sensible to changes on the BLER target.

To account for the resource utilization, the number of used resource elements per information bit is considered. For a conservative transmission, i.e. without a retransmission

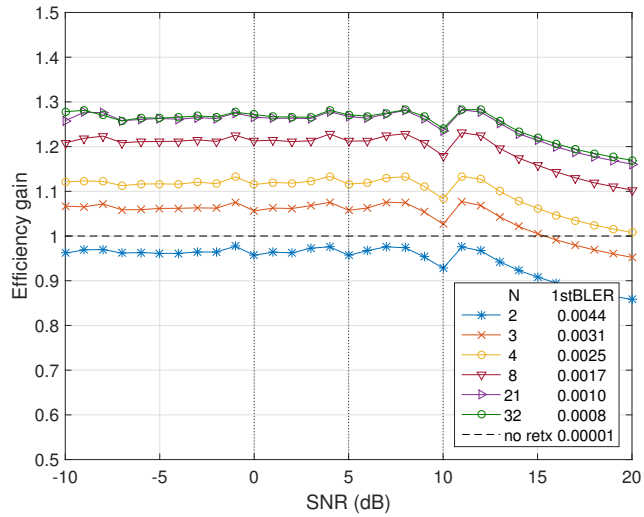

Fig. 6. Efficiency gain for different groupings of UEs (256-bit packet).

opportunity, it is written

$$\phi_c = \frac{1}{r_c(1-P_c)} \ , \tag{7}$$

where r_c is the transmission rate utilizing a conservative modulation order (m) and coding rate (c) to achieve the required success probability, i.e. $r_c = m \times c$; and P_c is the error probability, which should be the target BLER itself, considering ideal link adaptation.

For the proposed scheme, the required resources per bit can be simply given by the resources on the first transmission ϕ_1, which is less conservative, and the shared resources divided by N UEs ϕ_2, so

$$\phi_s = \phi_1 + \phi_2 = \frac{1}{r_1(1-P_1)} + \frac{1}{r_2(1-P_2)N} \ , \tag{8}$$

where r_1 and r_2 are the transmission rates for the initial and for the retransmission, respectively. For simplicity of the analysis, it is assumed that the grouped UEs have similar channel conditions, requiring the same MCS. It is also assumed that the MCS for the retransmission is equal to the initial transmissions (i.e. $r_1 = r_2$). With this, it was verified using the link model (from -10 to 10 dB SNR) that, with the soft combining providing 3 dB gain, the retransmission error probability is lower than the target, in this case 10^{-5}.

1) Efficiency gain without resource reallocation: Fig. 6 shows the gains in resource efficiency when comparing the scheme with shared retransmission opportunity against the conservative single initial transmission, that is ϕ_c/ϕ_s. Here it is first considered that, if all the initial transmissions are acknowledged, the reserved retransmission resource is wasted. It can be seen that, as expected, the efficiency is higher when more UEs share the retransmission resources. Taking the case with initial BLER at 10^{-3}, which permits groupings of up to 21 UEs achieving the 10^{-5} reliability, it can be noticed that the shared retransmission scheme brings gains of up to 28% on resource efficiency compared to a conservative transmission.

However, as shown in the previous section, larger groups demand lower BLER on initial transmission, which can be more challenging to accommodate in a TTI due to the larger amount of resources needed. It can also be observed that larger groups, e.g. greater than 21 UEs, do not provide better efficiency, since the required initial BLER become as low as for a conservative transmission.

For small groups of UEs, the gain drops since the wasting for having the reserved retransmission resource is higher than the gain given by the relaxed initial BLER target.

The slight variations in each curve is due to the discrete changes of MCS at each SNR. On higher SNRs the efficiency gain reduces, since the MCS and success rate of the conservative transmissions become high as in the proposed scheme.

2) Efficiency gain considering resource reallocation: In Fig. 7, similar resource efficiency evaluation was made, but now considering that the reserved retransmission resource can be re-allocated to a non-URLLC UE. These type of UEs, are normal mobile broadband users that do not have stringent latency and reliability requirements, so they can deal with possible errors and delays in granting procedures. In this case, since it is considered that the retransmission resource is not wasted when all the URLLC UEs succeed in initial transmission, the resources per bit is given by

$$\phi_{s'} = \phi_1 + \phi_2(1 - P_{idle}) = \frac{1}{r_1(1-P_1)} + \frac{1 - P_{idle}}{r_2(1-P_2)N}. \quad (9)$$

The re-allocation permits a better resource utilization in general since the wasting is avoided. It can be observed that, in this case, smaller groupings outperforms the bigger groupings. However, to consider that all the reserved resources of smaller groups can be reallocated, it is necessary sufficient demand from non-URLLC UEs in the network.

If there is a high traffic demand of non-URLLC UEs and low load of URLLC UEs in the network, it can be even worthy to reserve retransmission resources to each single URLLC UEs. For that case, a link adaptation scheme like in [12] could be applied for finding an efficient MCS.

It is important to note that, to apply the reallocation, there should be sufficient time budget for the base station, after the acknowledgments of the URLLC UEs, to grant the reserved resource to a non-URLLC UE.

V. Conclusion

In this paper it was proposed a scheme that employs pre-scheduling of resources shared by a group of URLLC UEs, for retransmissions. The analysis shows that, with the right dimensioning of groups and BLER target, the probability of contention for the shared retransmission can be sufficiently low. This means that the final error probability can be achieved without re-scheduling procedures. The resource efficiency of the method was compared against a single conservative transmission aiming at 10^{-5} of error probability. Considering that the reserved resources are wasted when all URLLC UEs initially succeed, it can be seen that the efficiency gain is higher (up to 28% for 256-bit packet) when more UEs are grouped. However, this requires lower initial BLER. For small groups (e.g.: 2), the wasting for having the reserved retransmission resource is higher than the gain of the relaxed initial transmission. On the other hand, when the reserved resources can be reallocated (e.g. to a non-URLLC UE), the efficiency of the proposed scheme is generally higher since the waste is avoided. Future work can consider enhancements for unpredictable traffic and simulations considering non-ideal link adaptation.

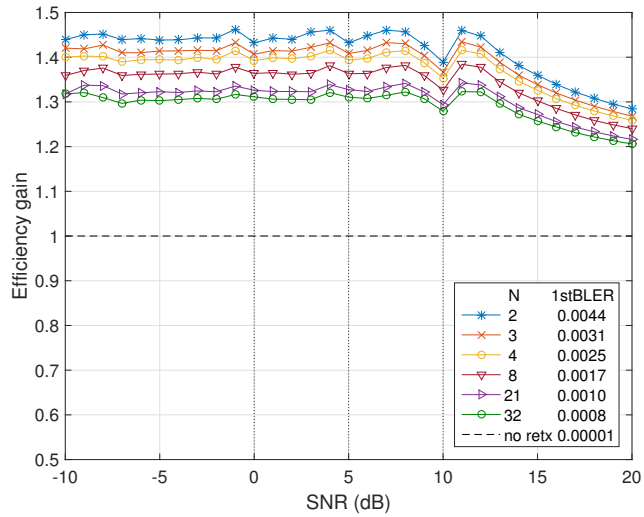

Fig. 7. Efficiency gain considering reallocation of the retransmission resource.

Acknowledgment

The authors would like to thank Krzysztof Bąkowski for the work on the link abstraction model.

References

[1] ITU-R M.2083-0, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Sept. 2015

[2] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies," 3GPP Tech. Rep., V14.0.0, Oct. 2016.

[3] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, G. Fettweis, "5G-Enabled Tactile Internet," IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, Mar. 2016.

[4] The 5G Infrustructure Public Private Partnership. *5G PPP White Papers on Energy, Automotive, Factories, and eHealth Vertical Sectors.* [Online]. Available: https://5g-ppp.eu/white-papers/. Accessed: Aug. 20, 2016.

[5] 3GPP TS 36.321, "Medium Access Control (MAC) protocol specification," 3GPP Tech. Spec., V13.2.0, Jun. 2016.

[6] 3GPP TR 36.881, "Study on latency reduction techniques for LTE," 3GPP Tech. Rep., V14.0.0, Jun. 2016.

[7] R1-167309, "Semi-persistent scheduling for 5G new radio URLLC," 3GPP TSG-RAN WG1 #86, Aug. 2016.

[8] H. Shariatmadari, S. Iraji, R. Jäntti, "Analysis of Transmission Methods for Ultra-Reliable Communications," IEEE 26th PIMRC, Sept. 2015.

[9] D. Jiang, H. Wang, E. Malkamaki, E. Tuomaala," Principle and Performance of Semi-persistent Scheduling for VoIP in LTE System," IEEE WiCom, Sept. 2007.

[10] L. G. Roberts, "Aloha packet system with and without slots and capture," SIGCOMM Comput. Commun. Rev., Apr. 1975.

[11] M. Sybis, K. Wesołowski, K. Jayasinghe, V. Venkatasubramanian, V. Vukadinovic, "Channel coding for ultra-reliable low-latency communication in 5G systems," IEEE VTC Fall, Sept. 2016.

[12] H. Shariatmadari, Z. Li, M. A. Uusitalo, S. Iraji, R. Jäntti, "Link adaptation design for ultra-reliable communications," IEEE ICC, May 2016.