On Optimizing Signaling Efficiency of Retransmissions for Voice LTE

Chia-An Hsu[†], Kate Ching-Ju Lin[†], Yi Ren[‡], and Yu-Chee Tseng[†]

[†]Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

[‡]School of Computing Science, University of East Anglia, Norwich, UK

[†]vincent.am05g@nctu.edu.tw, {katelin,yctseng}@cs.nctu.edu.tw, [‡]e.ren@uea.ac.uk

Abstract—The emergence of voice over LTE enables voice traffic transmissions over 4G packet-switched networks. Since voice traffic is characterized by its small payload and frequent transmissions, the corresponding control channel overhead would be high. Semi-persistent scheduling (SPS) is hence proposed in LTE-A to reduce such overhead. However, as wireless channels typically fluctuate, tremendous retransmissions due to poor channel conditions, which are still scheduled dynamically, would lead to a large overhead. To reduce the control message overhead caused by SPS retransmissions, we propose a new SPS retransmission protocol. Different from traditional SPS, which removes the downlink control indicators (DCI) directly, we compress some key fields of all retransmissions' DCIs in the same subframe as a *fixed-length hint*. Thus, the base station does not need to send this information to different users individually but just announces the hint as a broadcast message. In this way, we reduce the signaling overhead and at the same time, preserve the flexibility of dynamic scheduling. Our simulation results show that, by enabling DCI compression, our design improves signaling efficiency by 2.16×, and the spectral utilization can be increased by up to 60%.

I. INTRODUCTION

In recent years, LTE-A systems have been widely deployed over the world to provide data services. Although the amount of data traffic increases significantly, the voice service still contributes substantial profit to the operators. To enable voice services, the voice over LTE (VoLTE) technique has been proposed. Different from legacy circuit switched (CS) 3G networks, VoLTE supports voice traffic by packet switching (PS), which does not rely on dedicated channels. Since LTE-A provides a wider frequency band, VoLTE could support high-definition (HD) voice quality, short call set-up time and better battery efficiency.

Traditional LTE-A systems adopt dynamic scheduling for data transmissions. In dynamic scheduling, each user equipment (UE) learns resource allocation for its data transmissions in every subframe, which gives the network full flexibility, high channel utilization and the ability against increasing noise. However, voice traffic is characterized by the small packet size, frequent transmissions, and constant inter arrival time (e.g., AMR generating one voice packet every 20 ms). If voice data is dynamically scheduled, the system will generate a large control signaling overhead, which may congest the control channel. To resolve this issue, semi-persistent scheduling (SPS) has been proposed [1]. Its basic idea is to reserve persistent resources for *initial* voice transmissions, so as to eliminate the need of sending control messages for voice

traffic. However, SPS only considers initial transmissions of voice traffic. When retransmissions for voice traffic are required, SPS still uses dynamic scheduling. Thus, the signaling overhead is still high when a large number of retransmissions are triggered due to poor channel conditions.

To reduce the control overhead caused by SPS retransmissions, we present a new signaling protocol that compresses the signaling messages of SPS retransmissions. The core idea is to leverage hashing to aggregate some large fields of control messages for all SPS-retransmissions in the same subframe into a single message called DCI_{hint}. Specifically, multiple control messages are encoded as one fixed-length control message, which is then broadcast to those UEs waiting for retransmissions. Each VoLTE user then overhears DCI_{hint} and decodes the hint to extract its dynamic resource allocation. By replacing multiple unicast control messages with a broadcast message, we can hence significantly reduce the signaling overhead.

Comparing with SPS initial transmissions, which use persistent scheduling, our scheme not only reduces the signaling overhead for retransmissions but also provides flexibility of dynamic scheduling. We conduct extensive simulations to evaluate the performance of our protocol. The results show that our design improves signaling efficiency by up to 2.16× and, at the same time, improves the spectral utilization by up to 60%.

The rest of this paper is organized as follows. Section II introduces some background and related work. Section III defines our problem statement, and Section IV details our protocol designs. The performance evaluation is described in Section V. Section VI concludes this work.

II. BACKGROUND AND RELATED WORK

Before introducing our design, we first give brief background on dynamic scheduling and SPS.

A. Dynamic Scheduling in LTE-A

In the LTE FDD mode, each frame is of length 10 ms composed of 10 subframes. Each subframe contains 2 slots, and a slot contains 7 OFDM symbols. These 14 OFDM symbols are partitioned into control channels and data channels. In LTE-A, control channels include the physical control format indicator channel (PCFICH), the physical HARQ indicator channel (PHICH) and the physical downlink control channel
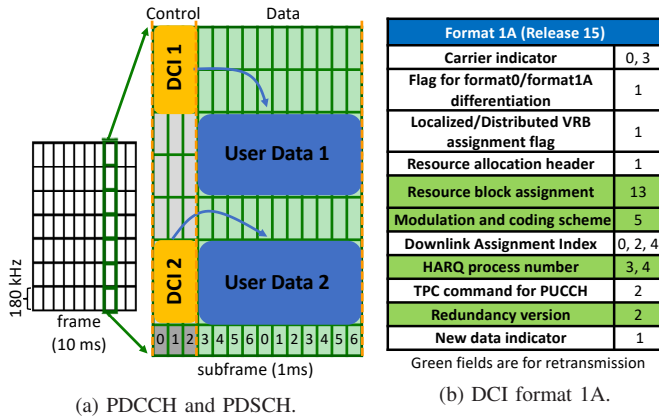
Format 1A (Release 15)	
Carrier indicator	0, 3
Flag for format0/format1A differentiation	1
Localized/Distributed VRB assignment flag	1
Resource allocation header	1
Resource block assignment	13
Modulation and coding scheme	5
Downlink Assignment Index	0, 2, 4
HARQ process number	3, 4
TPC command for PUCCH	2
Redundancy version	2
New data indicator	1

Green fields are for retransmission

(a) PDCCH and PDSCH.　　(b) DCI format 1A.

Fig. 1: DCI in LTE-A

(a) Dynamic Scheduling.　　(b) Semi-Persistent Scheduling.

Fig. 2: Scheduling in LTE-A

(PDCCH), which occupy the first three OFDM symbols of a subframe, as illustrated in Fig. 1a. User data are transmitted through the physical downlink share channel (PDSCH), which occupies the rest of the OFDM symbols. A base station (BS) transmits a downlink control indicator (DCI) through PDCCH to inform each UE its downlink data information, such as the modulation and coding scheme (MCS), resource block (RB) location, HARQ process number and redundancy version (RV), so that the UE can decode its own data properly. Fig. 1b illustrates the frame formate of DCI.

The LTE-A downlink scheduling procedure is as follows:
Step 1: Each UE either monitors PDCCH for every subframe during the connected mode or follows the discontinuous reception (DRX) cycle. The latter allows the UE to wake up periodically to monitor PDCCH.
Step 2: While monitoring PDCCH, a UE performs blind decoding on PDCCH. That is, each UE uses its own RNTIs (or any of its several RNTIs) to perform the CRC check on every candidate DCI message. Only those DCIs that pass the CRC check belong to this UE.
Step 3: After decoding its DCIs, a UE finds its downlink information, including MCS, RB assignment, HARQ process number and RV. It then locates and decodes its data on PDSCH of the same subframe.

An example is shown in Fig. 1a. Also, the format of LTE DCI format 1-A is shown in Fig. 1b.

B. Semi-Persistent Scheduling in LTE-A

Although dynamic scheduling provides a network the flexibility of assigning downlink resources, it is not suitable for supporting voice traffic since VoLTE traffic is characterized by a small data size and a short inter-arrival time. If dynamic scheduling is used for VoLTE traffic, tremendous signaling messages would be generated. As a result, the high control channel overhead will limit the number of VoLTE users allowed to join the system [2]. In order to support more VoLTE users, LTE introduces SPS for VoLTE traffic. The idea of SPS is to perform persistent scheduling for initial transmissions of voice traffic but exploit dynamic scheduling for voice retransmissions. The differences between dynamic scheduling and SPS are shown in Fig. 2. More specifically, instead of
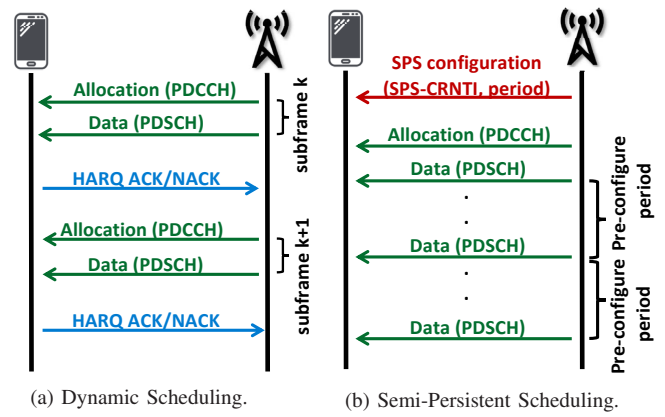
scheduling both initial transmissions and retransmissions via PDCCH signaling, UEs in the SPS mode keep using the same resource blocks allocated for initial transmissions. However, retransmissions may not be required by every VoLTE UE and, thereby, are scheduled dynamically so as to preserve flexibility. Therefore, SPS reduces the control channel overhead generated by initial transmissions.

C. Related Work

As SPS eliminates the signaling overhead for initial transmissions, the signaling overhead for VoLTE retransmissions is still large when the channel is unreliable. Several recent studies have investigated how to enhance the signaling efficiency for retransmissions, which can be classified into two categories: *shared resource retransmission* and *signaling size reduction*.

Shared resource retransmission: Recent work has proposed to share resources for retransmissions [3]–[5]. The work [3] reserves a set of resource blocks for a group of UEs, which eliminates the need of one DCI for every retransmission. A later work [4] then allows UEs to retransmit a failed packet multiple times over the shared resources within several transmission time intervals (TTI). Then, a successive interference cancellation (SIC) algorithm is used to decode the failed packets by using the signals of successful packets in the initial transmissions. However, in most of the practical scenarios, it is difficult to predict the exact number of failed initial transmissions. Thus, the reservation-based approach might cause a waste of reserved resources when there is no retransmission requirement. Also, if the number of UEs who need retransmissions is more than the number of the pre-allocated resources, it still needs control signals to avoid collisions in the shared resources, or a contention-based protocol should be used. In [5], the idea of shared resource allocation is further extended to initial transmissions. The idea is to categorize all devices by their traffic characteristics and broadcast control messages for each group. Each device then performs contention-based transmission in the resource blocks allocated to its group. However, deciding a proper size of the pre-allocated resources is still a challenging issue. In addition, if there exist unicast and multicase UEs scheduled as

a group, it would be hard to differentiate resource allocation for different types of users.

Signaling size reduction: Some approaches attempt to reduce the signaling size [6]–[8]. The work [6] observes that the number of the possible channel quality indicators (CQI) for voice traffic is less than the number of all feasible CQIs. Thus, the control message format (CQI or MCS) for voice traffic can be compressed. Recently, the signaling size reduction schemes mainly focus on the hint-based methods [7], [8]. The core of this method is to leverage the power of hashing to encode the original control message format. By adopting the hashing mechanism, the original unicast control messages can be encoded as a smaller broadcast message, thereby reducing the signaling overhead. Comparing with the shared resource scheme and the persistent scheduling scheme in SPS initial transmissions, the signaling size reduction scheme preserves flexibility and the ability against channel fading. Our work focuses on how to extend such a hint-based compression scheme to enable joint resource allocation and signaling for SPS retransmissions.

III. PROBLEM STATEMENT

We consider a single base station that serves a set of UEs, denoted by $\mathcal{V} = \mathcal{S} \cup \mathcal{B}$, where $\mathcal{S}$ is the set of UEs having a phone call (SPS mode) using the VoLTE technique, while $\mathcal{B}$ is the set of background non-VoLTE UEs who randomly generate unpredictable traffic demands in each subframe. We assume that some UEs in $\mathcal{S}$ may not receive their packets successfully in the previous subframes due to channel degradation or interference and need retransmissions in the current subframe.

In conventional SPS designs, each retransmitted packet needs a distinct DCI format that indicates its resource allocation. To mitigate the signaling overhead of triggering retransmissions, we propose a hint-based SPS retransmission protocol to arrange and trigger retransmissions so as to improve the success probability of retransmissions and optimize signaling efficiency, which is formally defined as follows:

$$Efficiency = \frac{\text{number of RBs used for retransmissions}}{\text{number of bits used to trigger retransmissions}}. \tag{1}$$

Different from traditional pooling-based designs, which enumerate all the pooled UEs in DCI, our retransmission protocol leverages a compression scheme that encodes the triggering messages as a hint, called DCI_{hint}. To realize this idea, we exploit hash-based compression [7]–[9], to encode DCI_{hint}. By such compression, we avoid the expensive cost of unicast notification for all the UEs that need a retransmission, but allow them to overhear a single hint and decode their information based on the hint. Hence, the size of control messages can be reduced, as a result improving signaling efficiency significantly.
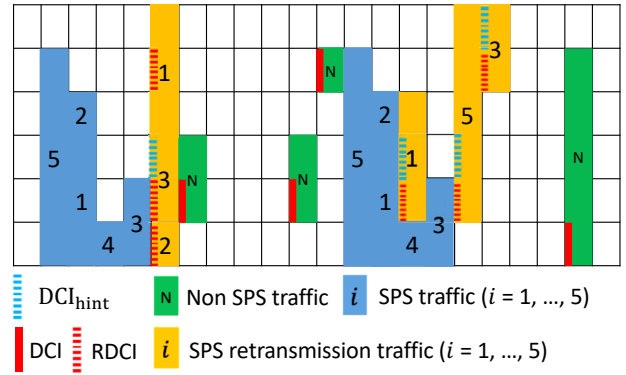

Fig. 3: An example of retransmission traffic scheduling.

Notation	Definition
$\mathcal{V}$	Set of UEs covered by the base station
$\mathcal{B}$	Set of non-VoLTE UEs
$\mathcal{S}$	Set of UEs with phone calls based VoLTE SPS mode
$\mathcal{L}_k^{init}$	Set of SPS-initial packets in subframe k
$\mathcal{L}_k^{bg}$	Set of background non-SPS packets in subframe k
$\mathcal{L}_k^{retx}$	Set of SPS-retransmission packets in subframe k

TABLE I: Definition of Notations

IV. SEMI-PERSISTENT RETRANSMISSION PROTOCOL

We will first overview the architecture of our design. We then describe the proposed DCI_{hint} encoding and decoding algorithms for efficient retransmissions.

A. Overview

In LTE-A, resources are allocated subframe-by-subframe. Note that the spectrum resources are shared by VoLTE and non-VoLTE UEs as well as original and retransmitted packets. To distinguish different types of UEs in a subframe, we further classify the set of UEs arranged in subframe k into a set of background non-VoLTE UEs, $\mathcal{B}_k$, a set of VoLTE UEs who need initial transmissions, $\mathcal{S}_k^{init}$ and a set of VoLTE UEs who need retransmissions, $\mathcal{S}_k^{retx}$. Similarly, we can categorize all the traffic demands of a subframe k into three classes: (1) SPS-initial traffic, $\mathcal{L}_k^{init}$, (2) background non-SPS traffic, $\mathcal{L}_k^{bg}$, and (3) SPS-retransmission traffic, $\mathcal{L}_k^{retx}$. As our main goal is to enhance signaling efficiency for voice retransmissions, we assume that the sets of demands, $\mathcal{L}_k^{bg}$, $\mathcal{L}_k^{init}$ and $\mathcal{L}_k^{retx}$ arranged in each subframe are given. We also assume that resource allocation $\mathcal{L}_k^{bg}$ and $\mathcal{L}_k^{init}$ in each subframe is given, but mainly focus on resource allocation and DCI encoding for $\mathcal{L}_k^{retx}$.

Recall that non-SPS traffic, such as web browsing, email or file sharing, may arrive randomly and, thereby, is scheduled dynamically in a subframe, as green RBs in Fig. 3, while SPS traffic is initial voice packets scheduled by persistent allocation, as blue RBs in Fig. 3. Hence, each non-SPS packet needs a DCI format, while no DCI format is required for SPS packets. The retransmitted SPS packets are also scheduled dynamically, as yellow RBs in Fig. 3. Note that a UE may have multiple packets to be retransmitted. We allow multiple retransmitted packets of a UE to be scheduled in the same subframe. In legacy LTE-A, $\mathcal{L}_k^{retx}$ also needs DCI formats.

In our design, we will encode their DCI formats as a single DCI$_{hint}$ to reduce the signaling overhead.

B. DCI Hint Encoding

We now introduce our DCI encoding protocol. In legacy LTE-A, each UE who waits for retransmissions should first perform *blind decoding* to decode all the DCI messages in PDCCH and extracts its own DCIs, each of which contains downlink information such as MCS, RB locations, the HARQ process number and the redundancy version. We observe that RB assignment and MCS occupy most bits of a DCI format (i.e., 13 and 5 bits, respectively, among the overall 38 bits of a DCI format for a 20MHz channel). Thus, our goal is to reduce the size of these two fields and minimize the signaling overhead.

To ensure reliable retransmissions, we can either retransmit using a more reliable MCS (e.g., BPSK) or allocate more power for retransmissions. Here, we extend the idea of legacy SPS. Instead of enabling MCS adaptation, we let a retransmission be sent at the same MCS as its initial transmission but allocate it a transmit power that can ensure successful delivery. By doing so, we can remove the MCS field from the DCI format for retransmissions.

For the RB assignment field, we first introduce the LTE-A *resource allocation type 2* (RA2) [10] and then explain our enhancement based on this method. LTE-A RA2 allocates a set of continuous RBs for one UE. The UE identifies the location of the starting RB and the length of allocated RBs by decoding the RB assignment field of its DCI format. LTE-A RA2 uses a fixed number of bits (13 bits for 20MHz) to allocate RBs, no matter how many RBs are allocated. To address such inefficiency, we propose to exploit a single special control message, DCI$_{hint}$, to trigger multiple SPS retransmissions scheduled in the same subframe. Our protocol first uses RA2 to find a set of consecutive RBs, denoted by $\mathcal{R}$, whose length is greater than or equal to the demands of all the SPS retransmissions $\mathcal{L}_k^{retx}$ in subframe k, as the green blocks shown in Fig. 4b. Note that if the demands of SPS retransmissions $\mathcal{L}_k^{retx}$ are too high and cannot be scheduled together in any set of consecutive RBs, we will iteratively remove a demand from $\mathcal{L}_k^{retx}$ and insert it into the next subframe $\mathcal{L}_{k+1}^{retx}$ until a large enough set of consecutive RBs can be found. Then, the BS will schedule all the SPS retransmissions in these consecutive RBs $\mathcal{R}$. The next questions are: (1) *How can the BS allocate those RBs to different retransmissions? and (2) How do the scheduled UEs identify the location of their allocated RBs?*

To allocate the RBs to multiple pending retransmissions, we leverage a pre-defined hash function $m = f(ID, sf_i, s)$, which is known by the BS and all UEs, to locate the RBs allocated to traffic i for UE ID. Here, ID is the unique identity of a UE, sf_i is the subframe index of packet i's original failed transmission, and s is a pre-calculated seed. We also feed the initial subframe index sf_i of packet i into the hash function since a UE may have multiple failed packets to be retransmitted in the same subframe. Therefore, the UE can use the subframe index sf_i of each packet i and the hash function

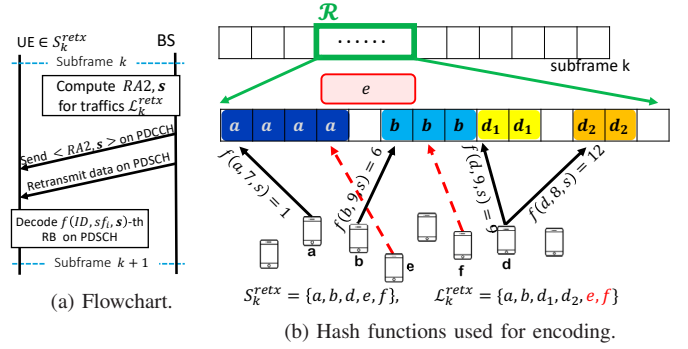(a) Flowchart.

(b) Hash functions used for encoding.

Fig. 4: Encoding and decoding procedure.

to locate the starting RB location of the retransmission for packet i. Hence, the retransmission of packet i originally sent in subframe sf_i for UE ID will be scheduled from the m-th RB in the reserved consecutive RBs $\mathcal{R}$. By doing this, we allow the retransmissions of multiple packets of a UE to be scheduled in the same subframe. To allow each UE to locate the starting RB of its retransmissions, we broadcast to all the UEs a single DCI$_{hint}$ message, which contains two fields: (1) a single 13-bit RB allocation field used by RA2 to allocate a set of consecutive RBs $\mathcal{R}$ for all the SPS retransmissions scheduled in the same subframe, and (2) a 12-bit random seed field s required by UEs to hash their RB locations.

One might notice that the proposed hash-based allocation may hash different retransmissions into the same subset of RBs in $\mathcal{R}$, leading to collisions. To avoid collisions, we initially identify the hashed RBs of all the packets in $\mathcal{L}_k^{retx}$. If there exist collided RBs, we iteratively identify a collided RB and remove the latest packet hashed to this RB from the schedule until the resource allocation becomes collision free. The removed packets can be inserted to $\mathcal{L}_{k+1}^{retx}$ and scheduled in the next subframe.

The DCI$_{hint}$ encoding procedure in subframe k for $\mathcal{L}_k^{retx}$ can summarized as follows:

1) The BS first identifies a set of consecutive RBs $\mathcal{R}$ of a length greater than or equal to the requirements of $\mathcal{L}_k^{retx}$ using RA2.
2) The BS searches a random seed s for hashing such that the retransmission of each packet $i \in \mathcal{L}_k^{retx}$ for UE $ID \in \mathcal{S}_k^{retx}$ can be allocated to the RBs starting from $f(ID, sf_i, s)$ of the block $\mathcal{R}$.
3) The BS repeatedly removes the packets which introduce collision in certain RBs until the allocation schedule becomes collision free.
4) The BS then broadcasts $<RA2, s>$ in the DCI$_{hint}$ message through PDSCH.

C. DCI Hint Decoding

All the UEs can extract their information from the single DCI$_{hint}$. The DCI decoding procedure is as follows:

1) UEs in $\mathcal{B}_k \cup \mathcal{S}_k^{init}$: These UEs do not wait for SPS retransmissions. Hence, they follow the legacy DCI decoding procedure to locate the RBs of their packets as mentioned in Sec. II-A.
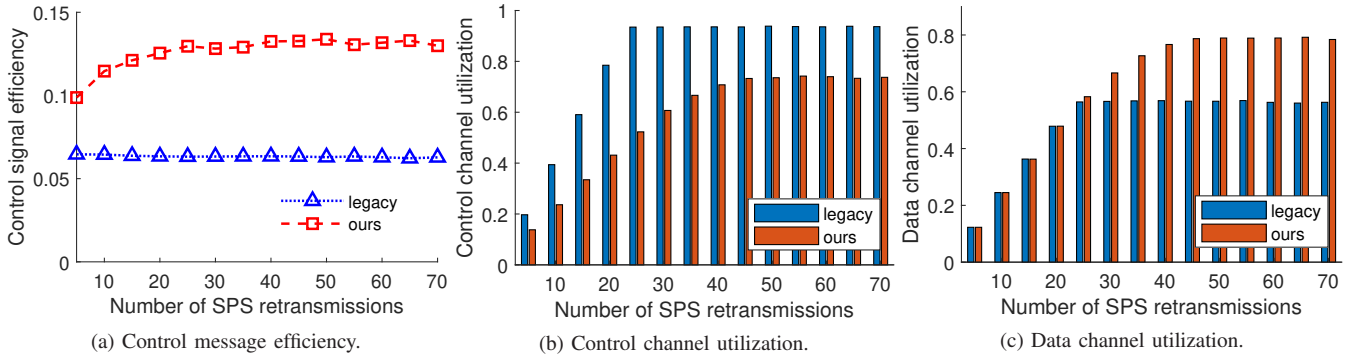
(a) Control message efficiency.
(b) Control channel utilization.
(c) Data channel utilization.

Fig. 5: Impact of number of SPS retransmissions.

2) UEs in $\mathcal{S}_k^{retx}$: These UEs are in the SPS mode with pending retransmissions in subframe k. Thus, they should use their RNTIs to perform blind decoding on PDCCH. We introduce a new RNTI called SPS-R-COMMON-RNTI to encode the DCI_{hint} message. Only the UEs in $\mathcal{S}_k^{retx}$ will use SPS-R-COMMON-RNTI to perform blind decoding. By doing this, we guarantee that only the UEs in $\mathcal{S}_k^{retx}$ obtain and decode DCI_{hint}, without affecting the remaining UEs.

Note that the proposed RB allocation scheme may remove some demands from $\mathcal{L}_k^{retx}$ due to insufficient RBs or hashing collisions. However, UEs in $\mathcal{S}_k^{retx}$ have no idea who has not been scheduled in subframe k. The nice property of our design is that the UEs actually do no need to know whether they are scheduled. As all the UEs in $\mathcal{S}_k^{retx}$ decode DCI_{hint}, only those scheduled UEs can decode their RBs successfully and push the packets matching their IP addresses to the upper layer (e.g., application layer). Consider a scenario where a UE's demand is removed due to hashing collision. This UE will still decode the DCI_{hint} message and locate the RBs based on the hashed location, i.e., $f(ID, sf_i, s)$. That will be fine in our protocol since this UE either fails to decode the hashed RBs or can find that the address field in the decoded packet does not match its own address, which can be dropped.

D. Example and Signaling Overhead

An example of the hashing procedure is illustrated in Fig. 4b. Say UEs a, b, d, e and $f \in \mathcal{S}_k^{retx}$ are waiting for retransmissions in subframe k while others are not. Among them, UE d has two packets to be retransmitted. After resource allocation for the non-SPS UEs and initial SPS UEs, the BS identifies the longest consecutive RBs $\mathcal{R}$ and hashes the packets a, b, d_1, d_2 and e, respectively, to $\mathcal{R}$. The packet of UE f has not been scheduled due to the limited size of $\mathcal{R}$ and will be scheduled in the next subframe. However, after hashing, the RBs allocated to packets a, b and e are partially overlapped. Hence, the BS should remove packet e to ensure collision-free allocation. After receiving DCI_{hint}, UEs a, b, d, e and f find the location of their allocated RBs by decoding RA2 using the hash function $f(ID, sf_i, s)$. In this case, UE d applies hashing for both the initial packets sent in subframes 8 and 9, respectively, and identifies their starting RB location. On the

BS maximum power	46 dBm
Noise power spectrum density	-174 dBm/Hz
System bandwidth	20 MHz
Packet size	320 bits

TABLE II: Simulation parameters

other hand, though UEs e and f also decode the RBs, they either fail to decode the hashed RBs or can find out that the decoded packets are not their retransmission.

To summarize, our protocol saves 18 bits for every SPS-retransmission DCI and uses only one DCI_{hint} message for all the UEs scheduled in a subframe, which contains only 25 bits to indicate the starting RB locations for all the SPS retransmissions in a subframe. Thus, the signaling overhead can be significantly reduced.

V. PERFORMANCE EVALUATION

We now evaluate the performance of the traditional SPS retransmission scheme and our DCI_{hint} retransmission scheme using Matlab. In our simulation, the system bandwidth is set to 20 MHz. The UEs are randomly deployed around a BS with a link distance ranging from 0.3 to 10 km, which follows the Rayleigh fading path loss model. The MCS of each UE is picked based on its SNR, following the setting used in [11]. Other parameters are summarized in Table II. The evaluated performance metrics include *signaling efficiency*, *control channel utilization*, and *data channel utilization*. *Signaling efficiency* is defined in Eq. 1, while *control channel utilization* is defined as the expected proportion of control channel resources utilized by SPS-retransmission signaling and *data channel utilization* refers to as the expected proportion of channel resources utilized by SPS data retransmissions. For each simulation, we repeat 1000 rounds and output the average result.

A. Impact of number of SPS-retransmissions

We first check the impact of number of SPS-retransmission packets. We vary the number of SPS retransmissions every subframe from 5 to 70. The number of required RBs for each packet is chosen based on its packet size and the selected MCS. Here, the average number of RBs occupied by non-SPS traffic in each subframe is set to 20. The signaling efficiency, control

(a) Control message efficiency.   (b) Control channel utilization.   (c) Data channel utilization.
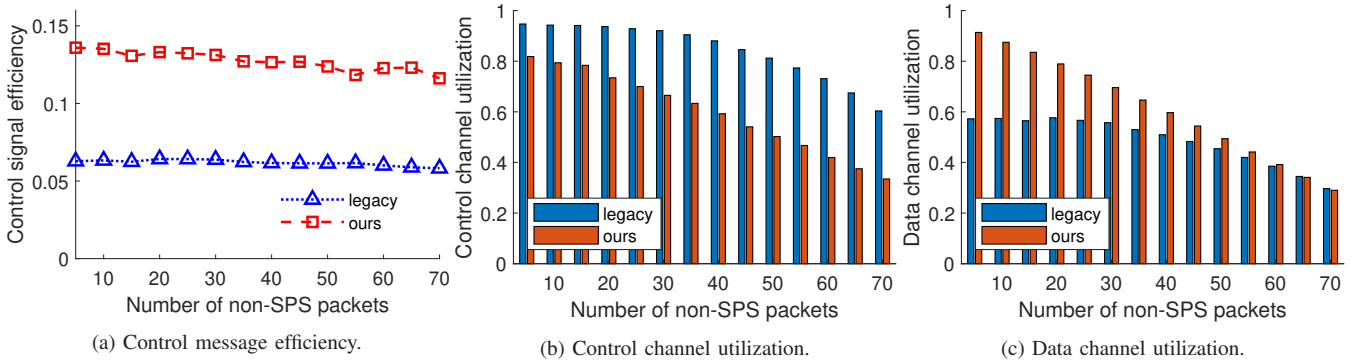
Fig. 6: Impact of number of non-SPS loading.

channel efficiency, and data channel efficiency are plotted in Figs. 5a, 5b, and 5c, respectively.

Fig. 5a shows that the message efficiency of our design increases as the number of SPS retransmissions grows since our design only needs a single DCI_{hint} for all the retransmissions. Hence, the efficiency can be improved as more retransmissions are involved in a subframe. However, the legacy LTE-A requires a DCI format for every retransmission and leads to a constant message efficiency no matter how many retransmissions are scheduled in a subframe. Compared to the legacy design, we improve the efficiency by up to $2.16\times$. Fig. 5b further demonstrates that the legacy LTE-A requires one DCI format for each retransmitted packet and, thereby, could saturate the control channel as the number of retransmissions increases. Hence, we can also see from Fig. 5c that the data channel is actually underutilized in the legacy LTE-A since the BS has no available space to send DCI formats for some pending retransmissions. Our scheme, however, only requires one DCI_{hint} and saves the usage of the control channel resources even when the number of SPS retransmissions grows. As a result, the BS can include more VoLTE retransmissions in the data channel and improve the overall channel utilization by up to 60%.

B. Impact of Network Loading

We next examine the performance of our design when the average number of RBs occupied by background non-SPS traffic varies from 5 to 70. The number of SPS retransmission packets is fixed to 60. The signaling efficiency, control channel efficiency, and data channel efficiency are plotted in Figs. 6a, 6b, and 6c, respectively. Fig. 6a shows that the signaling efficiency decreases slightly as the number of non-SPS packets increases. The reason is that the number of RBs available for retransmissions decreases, which offsets the benefit of our design. The size of control channel resources available for SPS retransmissions can be large as the demands of non-SPS traffic is low. In this case, the effectiveness of our design can be more obvious since we utilize the control channel more efficiently and can schedule more data retransmissions in a subframe, as shown in Fig. 6b and Fig. 6c, respectively.

VI. CONCLUSIONS

In this work, we consider the control channel overhead problem caused by SPS retransmissions and propose a new SPS retransmission protocol. We leverage LTE RA2 and a special DCI format compressed based on a hash function to enable joint RB allocation and signaling for SPS retransmissions. By hash-based compression, we eliminate the need of every DCI format for every retransmitted packet, but only need a single DCI hint. The simulation results further show that, by improving signaling efficiency, we better utilize the control channel resources and, thereby, can schedule more retransmissions in every subframe, as a result increasing the system capacity.

REFERENCES

[1] 3GPP TS 36.321, *Medium Access Control (MAC) protocol specification (Release 15)*, 3GPP Std., Apr. 2018.
[2] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and performance of semi-persistent scheduling for VoIP in LTE system," in *International Conference on Wireless Communications, Networking and Mobile Computing*, 2007.
[3] R. Abreu, P. Mogensen, and K. I. Pedersen, "Pre-scheduled resources for retransmissions in ultra-reliable and low latency communications," in *IEEE WCNC*, 2017.
[4] R. Abreu, G. Berardinelli, T. Jacobsen, K. Pedersen, and P. Mogensen, "A blind retransmission scheme for ultra-reliable and low latency communications," in *IEEE VTC Spring*, 2018.
[5] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Kovács, and P. Mogensen, "Joint resource configuration and MCS selection scheme for uplink grant-free URLLC," in *IEEE GlobeCOM Workshop*, 2019.
[6] P. Ghosh, S. S. Das, and P. Chandhar, "VoIP scheduling with reduced overhead and radio resource usage estimation-effect on best effort capacity," in *National Conference on Communications*, 2012, pp. 1–5.
[7] T. Chan, Y. Ren, Y. Tseng, and J. Chen, "eHint: An efficient protocol for uploading small-size IoT data," in *IEEE WCNC*, 2017.
[8] T.-Y. Chan, Y. Ren, Y.-C. Tseng, and J.-C. Chen, "Multi-slot allocation protocols for massive IoT devices with small-size uploading data," *Accepted by IEEE Wireless Communications Letters*.
[9] T. Huang, Y. Ren, K. C. Lin, and Y. Tseng, "r-hint: A message-efficient random access response for mmtc in 5g networks," in *IEEE PIMRC*, 2017.
[10] 3GPP TS 36.213, *Physical layer precedures (Release 15)*, 3GPP Std., Apr. 2018.
[11] M. Kawser, N. Imtiaz Bin Hamid, M. Nayeemul Hasan, M. Shah Alam, and M. Musfiqur Rahman, "Downlink snr to cqi mapping for different multiple antenna techniques in lte," *International Journal of Information and Electronics Engineering*, vol. 2, pp. 756–760, 09 2012.