Short Papers

State-Based Uplink-Scheduling Scheme for Reducing Control Plane Latency of MCPTT Services

Juyeop Kim, *Member, IEEE*, Ohyun Jo, *Member, IEEE*, and Sang Won Choi, *Member, IEEE*

***Abstract*—This paper addresses an uplink (UL) scheduling issue that arises in the midst of the recent trend in applying the commercial long term evolution (LTE) technology to public safety. Mission critical push-to-talk (MCPTT), which specifies a group communication service over LTE in public-safety scenarios, needs to satisfy the requirements of extremely short latency for its control procedures. We herein propose a practical UL-scheduling scheme, which can be feasible in the LTE system to shorten the time for the initial UL transmissions. Through the empirical assessment conducted on our LTE-based MCPTT testbed, the proposed scheme is validated to reduce the control plane latency with the little increment of implementation complexity.**

***Index Terms*—Floor control, latency and public safety, long term evolution (LTE), mission critical push-to-talk (MCPTT).**

I. Introduction

The evolution of commercial mobile communications technology begins to drive an advancement of the communication service in public safety. In this regard, the 3rd generation partnership project has progressed a standardization of mission critical push-to-talk (MCPTT) to enable a group communication service for public safety. The MCPTT provides a method of communication among a group of users in a half-duplex way, which is an essential function in public-safety scenarios. The key point of the MCPTT lies in its minimized control plane latency, which is an important performance indicator in terms of a *mission-critical* service. In specific, the MCPTT requires the procedures of *late entry* and *floor control* to be accomplished in a short time. The late entry is for joining an ongoing group call, and the floor control is for arbitrating a right to speak. The MCPTT requires the latency of the late entry and the floor control to be less than 150 and 300 ms, respectively, with a probability of 95% [1].

To satisfy the afore-mentioned criteria, it is required to reconsider every operation that extends the latency of the control procedures. Any operations in the perspective of the MCPTT service will matter if those generate delay in the order of tens of milliseconds. In light of this, UL scheduling, which aims to allocate uplink (UL) grants to user equipment (UEs), potentially causes significant queuing delay for the UL packets initially generated by the control procedures. For instance, a UE that requests to talk cannot instantaneously have any UL grant unless it already has one for sending the previous UL packets. In this case, the UE tries a scheduling request (SR) procedure for requesting a UL grant. The eNodeB is indicated about the pending data by the SR signal through Physical UL Control CHannel (PUCCH), and consequently provides the minimal UL radio resource to let the UE report its buffer status. The eNodeB then allocates a UL grant to the UE for sending the UL pending data based on the buffer status.

It is worthy to note that the latency of this SR procedure can be significant in the perspective of the MCPTT service. The radio resource of the SR PUCCH is periodically allocated to a UE, so the latency of the SR procedure is proportional to the SR periodicity. Each eNodeB configures the SR periodicity based on the maximal number of UEs that it can support. Since public safety requires an eNodeB to accommodate up to 500 UEs [2], the SR periodicity should be configured as a large value, which leads the SR procedure to waste much time. It is technically feasible to reduce the latency by using additional hardware for detecting multiple SR PUCCHs in a radio resource, but it is not practically desirable for public safety. There are a number of research works dealing with UL scheduling and random access issues (see [3]–[6]), but none of them can settle this problem.

The goal is to propose a practical UL scheduling scheme, which can minimize the control plane latency in the MCPTT system. Along with proposing the scheme, we perform a practical assessment with the real-time operating systems to provide realistic and empirical results of the latency performance.

II. System Model

Fig. 1 depicts the overall model of the MCPTT system. It is composed of UEs and an MCPTT server, which play the roles of end entities, and a network that intermediately transfers packets between the UEs and the server. The lower part of Fig. 1 describes the procedures for late entry and floor control. The UE starts the late entry for joining to the ongoing group call. The UE is then allowed to transmit voice data packets after obtaining a right to speak through the floor control. These procedures have a common characteristic that the UE occasionally triggers the procedures by sending an initial control message. At the moment that the UE is commanded through its user interface, the late entry or the floor control begins from the transmission of the UL message, i.e., *INVITE* or *floor request*.

What we need to focus on here is the UL-scheduling delay for the initial messages of the control procedures. The UL-scheduling delay is almost the same as the total execution time of the SR procedure, which is equivalent to the time that the UE waits for its SR PUCCH

Manuscript received March 9, 2018; revised July 10, 2018; accepted August 30, 2018. This work was supported in part by Sookmyung Women's University under Grant 1-1803-2005, in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the MSIP under Grant 2018R1C1B5045546, and in part by the Institute for Information & communications Technology Promotion (IITP) funded by the Korea government (MSIP) under Grant 2018-0-00726 (Development of Software-Defined Cell/Beam Search Technology for Beyond-5G Systems). *(Corresponding author: Sang Won Choi.)*

J. Kim is with the Sookmyung Women's University, Yongsan-gu 04310, South Korea (e-mail: jykim@sookmyung.ac.kr).

O. Jo is with the Chungbuk National University, Cheongju 28644, South Korea (e-mail: ohyunjo@chungbuk.ac.kr).

S. W. Choi is with the Korea Railroad Research Institute, Uiwang 437757, South Korea (e-mail: swchoi@krri.re.kr).

Digital Object Identifier 10.1109/JSYST.2018.2868484

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

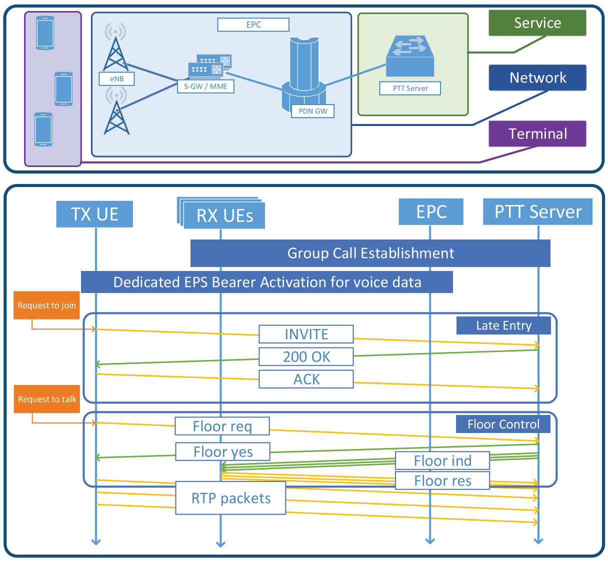

Fig. 1. Architecture and control procedures of the MCPTT system.

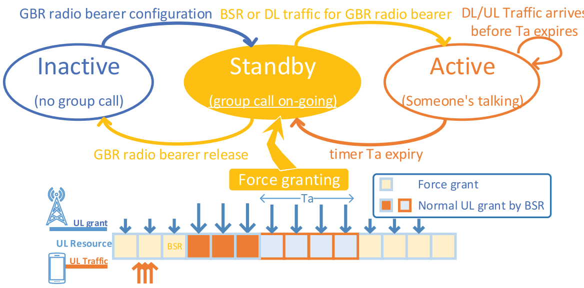

Fig. 2. State machine for the proposed UL scheduling scheme.

```
While (1) {
  for i from 1 to (# of UEs in the cell) {
  //State transition
    UE(i).state = transitState(UE(i));
  //action for UL granting
    if (UE(i)'s state == Standby)
      {then, allocate a UL grant for Buffer
          Status Report;}
    else if (UE(i)'s state == Active) {
      then, allocate UL grant normally;
      if (DL/UL traffic arrived to the UE(i))
        {then, restart timer Ta;}
    }
  }
}
```

```
Function transitState(UE) {
 state = UE.state;
 switch (UE's state) {
 case Inactive :
  if (UE has configured GBR radio bearer)
   {state = Standby;} break;
 case Standby :
  if (BSR or DL traffic comes for GBR radio
        bearer)
   {state = Active;}
  else if (UE has released GBR radio bearer)
   {state = Inactive;} break;
 case Active :
  if (Ta is expired for the UE)
   {state = Standby;} }
 return state;}
```

Fig. 3. Proposed protocol for UL scheduling.

transmission. For simplicity, we assume a simple traffic model in which the event of late entry or floor request occurs in a uniformly distributed way. Then, with a probability of 95%, the time for waiting to send the SR PUCCH will be $0.95T_{\mathrm{SR}}$, where T_{SR} is the SR periodicity. Let N_{cell} be the maximum number of UEs that the eNodeB can support, and N_{SR} be the number of physical resource blocks (PRBs) dedicatedly allocated for SR PUCCH transmissions. Since each UE should transmit the SR PUCCH once in an SR period, T_{SR} should satisfy

$$T_{\mathrm{SR}} \geq \frac{N_{\mathrm{cell}}}{N_{\mathrm{SR}}} \text{ (ms).} \tag{1}$$

When configuring N_{SR}, it is needed to consider the worst condition that the system fully provides MCPTT services to N_{cell} UEs. N_{SR} should be configured under the condition that

$$N_{\mathrm{SR}} \leq N_{\mathrm{RB}} - (N_d + N_{\mathrm{ACK}} + N_{\mathrm{CQI}}) \tag{2}$$

where N_{RB} is the total number of UL PRBs and N_d, N_{ACK}, and N_{CQI} are the numbers of PRBs within a subframe for voice-data transmission, the PUCCH of ACK/NACK feedback, and the PUCCH of channel quality feedback, respectively, in the worst case. The LTE system allocates one PRB to a UE for each ACK/NACK or channel quality feedback, so the number of PRBs allocated for the feedbacks is equivalent to the number of UEs sending the feedbacks in the subframe. (This is regardless of the data size of the feedback.) In the worst case, $N_{\mathrm{cell}} - 1$ group UEs receive a talk burst, so $N_{\mathrm{ACK}} = (N_{\mathrm{cell}} - 1)/T_v$, where T_v is the periodicity of the voice traffic in milliseconds. N_d depends on the number of the UEs that talk simultaneously. Let N_g be the maximum number of ongoing group calls, then $N_d = N_g/T_v$. In a similar way, we can derive that $N_{\mathrm{CQI}} = N_{\mathrm{cell}}/T_{\mathrm{CQI}}$, where T_{CQI} is the periodicity of the channel quality feedback in milliseconds.

Considering the practical environment, we can assume that $T_v = 20$ ms, $N_{\mathrm{RB}} = 50$ (10 MHz bandwidth), $N_{\mathrm{cell}} = 500$, and $N_g = 150$ [1]. Even if we assume that T_{CQI} is the maximum as 128 ms, $N_{\mathrm{SR}} \leq 12$, $T_{\mathrm{SR}} \geq 41.6$ ms, and T_{SR} should be configured as 80 ms according to the candidate values configurable for the SR periodicity. This implies that the UL-scheduling delay, which is a part of the overall control plane latency, is 76 ms with a probability of 95%. This is quite large in the perspective of the MCPTT service, because 76 ms corresponds to a half and 27% of the criteria for the late entry and the floor control latency, respectively. Since the overall delay taken in the LTE network is typically less than 10–20 ms, and the queuing delay of downlink traffic at the eNodeB is ignorable, we can easily conclude that the UL scheduling is the main factor of the control plane latency of the MCPTT service.

III. Proposed Scheme

One way to skip this time-consuming SR procedure is to offer the minimal radio resource for buffer status reporting to the UEs, which can potentially trigger the control procedures. We herein propose a new UL-scheduling scheme that provides UL grants based on UEs' states and hopping patterns over time and frequency domains.

A. UL Granting Based on the UE State

Figs. 2 and 3 depict the basic idea of the proposed scheme in terms of UL granting. An eNodeB manages a state machine for each UE and allocates UL grants based on it. The UE is in an *inactive* state if its group has not established a call session yet. Once a group call is established, each UE affiliated to the group activates a dedicated evolved packet system bearer, which can carry voice data with guaranteed bit rate (GBR). (The unjoined UEs also preconfigure the bearer for reducing the latency of the further late entry.) The eNodeB recognizes this event by checking whether any GBR radio bearer is configured to the UE, and transits the state of the UE from *inactive* to *standby*. The UE in the standby state periodically gets *force grants*, which are unconditionally allocated at all times for buffer status reporting.

The UE in an *active* state is either transmitting or receiving a talk burst. The eNodeB transits the state of the UE from *standby* to *active* when a buffer status is reported or voice traffic is received with respect to the GBR radio bearer. The eNodeB then needs to stop providing the force grants to the UE, since the UEs in the active state are not likely to start the late entry or the floor control procedure. When the talk burst stops, the eNodeB returns the state of the UE to *standby*. From this

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

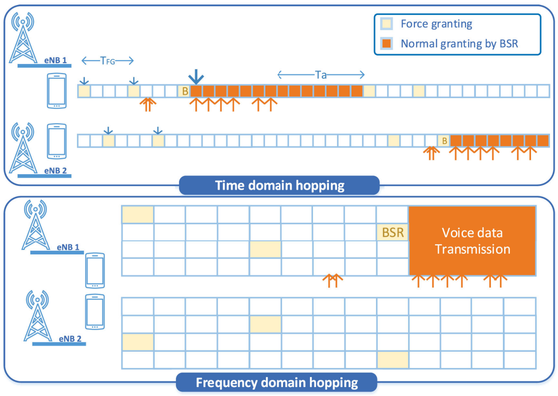

Fig. 4. Example for the hopping-based force granting.

moment, the eNodeB restarts providing the force grants so that the UE can immediately report its buffer status for the next control procedure.

For recognizing the end of a talk burst, the eNodeB configures T_a timer for each UE in the active state. The T_a timer starts (or restarts) for every moment that the eNodeB observes downlink/UL voice traffic from the GBR radio bearer. When the talk is stopped, the eNodeB will not see voice traffic from the GBR radio bearer and the T_a timer will be expired. In this case, the eNodeB considers that the talk burst has ended and changes the state of the UE to *standby*.

This scheme effectively resolves the issue of the UL-scheduling delay, because the UEs will have immediate chances to request UL grants for the unexpected initial control messages. However, the frequent allocation of the force grants may cause some side effects. If many public-safety UEs have joined to a group call, the eNodeB should utilize much of the UL radio resource for providing the force grants. In addition, the more the eNodeB provides the force grants, the more interference between group UEs is generated. This is because a UE having a UL grant should emit signal for radio link monitoring at the eNodeB side, despite the lack of its UL traffic. The frequent force granting will, therefore, increase the amount of unnecessary signal emission. This can come into strong and sudden interference at a cell boundary region if some of the group UEs have moved earlier to a neighbor cell. This gives a critical problem to the loop algorithms with respect to power control and rate adaptation.

B. Jointly Using a Hopping Pattern

To mitigate the aforementioned side effects, we can consider giving a hopping pattern over time and frequency domains to the UL radio resource for the force granting. Utilizing a time-hopping pattern provides the eNodeB more flexibility in multiplexing force grants and normal UL grants. The hopping also has a positive effect of suppressing intercell interference occurring among group UEs. Above all, it is rather a simple method in the implementation perspective. The hopping pattern of each UE is solely managed by the eNodeB and the UE does not need any additional procedure for accommodating the hopping force grants.

There are a number of pieces of conventional work, which deal the hopping patterns for interference mitigation. [7] Here, we simply use the hopping pattern defined in the LTE specification. Fig. 4 shows the details of the hopping-based force granting. In the aspect of the time domain, each of the group UEs gets force grants, which have periodicity. The offset of the force-granting pattern is configured differently among the group UEs, so that the small part of the group UEs transmits within a UL subframe. This hopping causes the additional UL scheduling delay as $0.95T_{\mathrm{FG}}$, where T_{FG} is the periodicity of the force grants, but this can be negligible if T_{FG} is sufficiently small. In the aspect of the frequency domain, let n_V be the logical index of the UL radio resource block allocated for a force grant, and the index of the PRB n_P after applying the hopping pattern is given by $n_P(n_s) = \tilde{n}_P \bmod (N_{\mathrm{RB}})$, where

$$\tilde{n}_P = (n_V + ((N_{\mathrm{RB}} - 1) - 2(n_V \bmod N_{\mathrm{RB}})) \cdot f_m(n_s)) . \quad (3)$$

f_m is equal to $c(10i)$, where $c(i)$ corresponds to the ith element of the pseudorandom sequence used in the LTE.

TABLE I
SYSTEM PARAMETERS OF THE LTE AND MCPTT TESTBED

System parameters	Values
Frequency band	UL: 718-728MHz, DL: 773-783MHz
# of UEs and eNodeBs	4 UEs, 2 eNodeBs
UE and eNodeB output power	23dBm / 46dBm
# of PRBs for PUCCH format 1	2
# of HARQ retransmissions	3
# of PRB and TB size of a force grant	1 PRB, 70 bytes
SR periodicity and T_{FG}	80ms, 10ms
Distance between eNodeBs	421m
Height of eNodeBs	18m / 12m
UE speed and voice codec	3km/h, G.729
Size of floor req / INVITE	55 / 609 Bytes
Size of 200 OK / ACK	705 / 376 Bytes

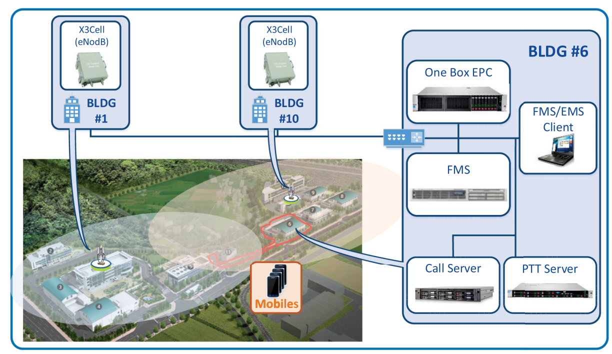

Fig. 5. Deployment of the LTE-based MCPTT testbed.

IV. PERFORMANCE EVALUATION

We utilize a practical testbed, which is composed of a commercial-level network and MCPTT service solutions, and uses the frequency band allocated for public safety in South Korea. Two Contela X3Cell eNodeBs are deployed at the roof of the two buildings and connected to a Contela One Box EPC. We use our MCPTT servers interoperating with smartphone-type UEs that have Qualcomm MSM8974 and Android 4.4.3. The detailed parameters of the testbed are given in Table I.

A. Analysis and Further Consideration

Moving along the route depicted as the red line in the map of Fig. 5, we repeated the floor control and the late-entry procedures and measured the corresponding latency defined in [1] for each single run. We used the time-domain hopping for the hopping-based force-granting scheme. For each scheme, we have done 200 trials, which are at least needed for obtaining a stable statistical result.

Fig. 6 depicts the statistics of the single-run measurement results. With a probability of 95%, the floor control latency is less than 95 ms when using the force granting. As we expected in the numerical

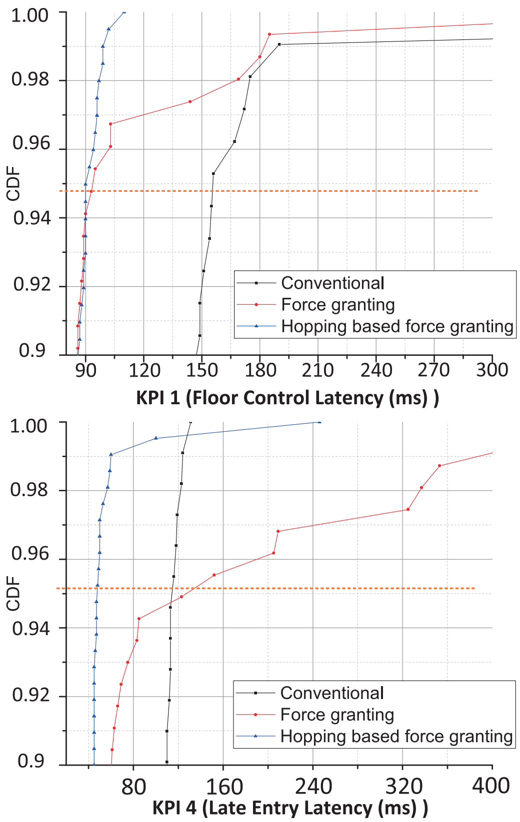

Fig. 6. Key performance results for the MCPTT control plane latency.

analysis, the results reveal that force granting effectively reduces the latency in the UL section by 70 ms in comparison to the conventional UL scheduling based on the SR procedure. In the aspect of late entry, it is remarkable that the latency increases if the force granting is solely utilized. This is because the strong intercell interference degrades the UL data rate of the group UEs at the cell boundary region and the UL transmission of the INVITE message, whose size is much bigger than that of the floor request message, causes larger delay. When the time-domain hopping is jointly applied with the force granting, a significant advancement is observed in reducing the latency by 35.6%.

The force granting requires to utilize additional radio resource as a tradeoff for the latency improvement, so the proposed scheme may cause to degrade the capacity of group call. Satisfying the latency requirement is the necessary condition for mission critical services, so it is worth sacrificing some of the capacity for improving latency in the public-safety environment. In the aspect of performance optimization, it is the future work to enhance the proposed scheme for minimizing the amount of redundant force grants. For instance, considering speech pattern can be a good guide of allocating force grants in a minimal way.

References

[1] *Technical Specification Group Services and System Aspects; Mission Critical Push to Talk Over LTE*, 3GPP TS 22.179 v14.3.0, Dec. 2016.

[2] *Technical Specification Group Services and System Aspects; Group Communication Service Enablers for LTE*, 3GPP TS 22.468 v14.0.0, Mar. 2017.

[3] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, “Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink,” in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 2611–2615.

[4] N. A.-Ali *et al.*, “Uplink scheduling in LTE and LTE-advanced: Tutorial, survey and evaluation framework,” *IEEE Commun. Survey Tut.*, vol. 16, no. 3, pp. 1239–1265, Jul.–Sep. 2014.

[5] M. Condoluci *et al.*, “Enhanced radio access and data transmission procedures facilitating industry-compliant machine-type communications over LTE-based 5G networks,” *IEEE Wireless Commun.*, vol. 23, no. 1 pp. 56–63, Feb. 2016.

[6] M. Vilgelm *et al.*, “LATMAPA: Load-adaptive throughput-maximizing preamble allocation for prioritization in 5G random access,” *IEEE Access*, vol. 5 pp. 1103–1116, Jan. 2017.

[7] J. J. Bazzo *et al.*, “4G/LTE networks for mission-critical operations: A cognitive radio approach,” in *Cognitive Technologies*. New York, NY, USA: Springer-Verlag, Jul. 2017, pp. 51–64.