Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks

Arjun Anand, Gustavo de Veciana, and Sanjay Shakkottai

Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, USA

April, 2020

Published in IEEE/ACM Transactions on Networking (also in IEEE INFOCOM 2018)

System Model

- Saturated system [1]: Each eMBB user has infinite amount of data to be served.

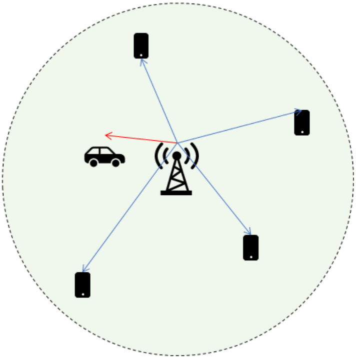

Figure: System model

System Framework

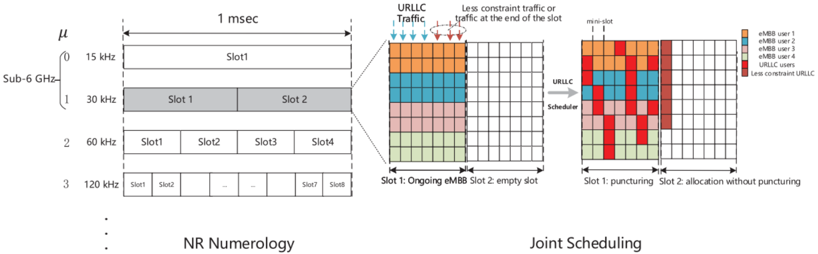

Figure: System framework [2]

Problem Statement

- URLLC puncturing/superposition commonly used scheduling procedure: allocate *subchannels* for eMBB users in the given time slot first, then schedule *resources* for URLLC users in the respective time minislots accordingly.
- Multiplexing – conducting eMBB resource allocation using only eMBB channel state information (CSI) – has been broadly discussed in the literature.
- However, joint scheduling of eMBB and URLLC traffics has not been well studied.
- That is, in addition to eMBB CSI, could we also leverage URLLC information (e.g. *observed* behavior of URLLC demands, any others?) to schedule eMBB?

Examples

- Discussion with Yung-Ching senpai: Reduce eMBB retransmissions of data damaged by URLLC puncturing/superposition.
  - Approach 1: Reserve resources for URLLC traffic.
  - Approach 2: Pre-allocate resources for some URLLC users (less overhead compared to the first method, which increases URLLC reliability).
- However, their paper does not address the aforementioned problems, but aims to tackle two questions:
  - For linear loss model, would the traditional multiplexing procedure be optimal in the long-term?
  - For convex loss model, could we somehow incorporate URLLC demands into eMBB resource allocation formulation?

Solution

- The main idea is to perform mathematical analysis on eMBB time slot.
- This is done by their model 'proportion of an eMBB time slot's resource'.
- Linear, convex, and threshold loss models are discussed.

Linear Loss Model

- A URLLC *puncturing* problem is formulated as follows:

$$\begin{aligned}
\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \quad & \sum_u ln\left(\bar{r_u}\right) \\
\text{subject to} \quad & \sum_u \alpha_{u,n,l} \leq \frac{1}{\mathfrak{l}}, && \forall n, \forall l, \\
& \alpha_{u,n,l} \in \left\{0, \frac{1}{\mathfrak{l}}\right\}, && \forall u, \forall n, \forall l, \\
& \sum_u \sum_l \beta_{v_0,u,n,m,l} = D_{v_0,n,m}, && \forall n, \forall m, \\
& \beta_{v_0,u,n,m,l} \leq \alpha_{u,n,m,l}, && \forall u, \forall n, \forall m, \forall l, \\
& \beta_{v_0,u,n,m,l} \in \left\{0, \frac{1}{\mathfrak{m}\mathfrak{l}}\right\}, && \forall u, \forall n, \forall m, \forall l
\end{aligned} \tag{1}$$

- where

$$\alpha_{u,n,m,l} = \frac{\alpha_{u,n,l}}{\mathfrak{m}}, \quad \forall u, \forall n, \forall m, \forall l \tag{2}$$

Linear Loss Model (Continued)

- The average rate is calculated as follows.

$$\bar{r}_u = \frac{r_{u,0} + r_{u,1} + \ldots + r_{u,\mathfrak{n}-1}}{\mathfrak{n}} \tag{3}$$
$$= \frac{1}{\mathfrak{n}} \sum_n r_{u,n} \tag{4}$$
$$= \frac{1}{\mathfrak{n}} \sum_n \sum_l r_{u,n,l} \tag{5}$$
$$= \frac{1}{\mathfrak{n}} \sum_n \sum_l \left(\alpha_{u,n,l} - \beta_{v_0,u,n,l}\right) \mathfrak{r}_{u,n,l}, \quad \forall u \tag{6}$$

- In this paper, subchannel allocation is neglected.

$$\bar{r}_u = \frac{1}{\mathfrak{n}} \sum_n r_{u,n} \tag{7}$$
$$= \frac{1}{\mathfrak{n}} \sum_n \left(\alpha_{u,n} - \beta_{v_0,u,n}\right) \mathfrak{r}_{u,n}, \quad \forall u \tag{8}$$

Linear Loss Model (Continued)

- The first idea that comes to my mind is to give each eMBB user an equal share of URLLC demands.

$$\beta_{v_0,u,n,m} = \frac{D_{v_0,n,m}}{\mathfrak{u}}, \quad \forall u, \forall n, \forall m \tag{9}$$

- However, this cannot guarantee

$$\beta_{v_0,u,n,m} \leq \alpha_{u,n,m}, \quad \forall u, \forall n, \forall m. \tag{10}$$

- It is then intuitive to give each eMBB user a proportionally equal share of URLLC demands.

$$\beta_{v_0,u,n,m} = \frac{\alpha_{u,n,m}}{\frac{1}{\mathfrak{m}}} D_{v_0,n,m} \tag{11}$$

$$= \alpha_{u,n} D_{v_0,n,m} \tag{12}$$

Linear Loss Model (Continued)

- Such an allocation guarantees URLLC constraints:

$$\sum_{u} \beta_{v_0,u,n,m} = D_{v_0,n,m}, \quad \forall n, \forall m, \tag{13}$$

$$\beta_{v_0,u,n,m} \leq \alpha_{u,n,m}, \quad \forall u, \forall n, \forall m. \tag{14}$$

Linear Loss Model (Continued)

- URLLC allocation in eMBB time slot can then be derived:

$$\beta_{v_0,u,n} = \sum_m \beta_{v_0,u,n,m} \tag{15}$$

$$= \alpha_{u,n} \sum_m D_{v_0,n,m} \tag{16}$$

$$= \alpha_{u,n} D_{v_0,n} \tag{17}$$

Linear Loss Model (Continued)

Theorem 1

For a wireless system under the linear superposition/puncturing loss model we have that $C = C^{LR}$

- By this theorem, the proportionally equal share policy is optimal.

$$\bar{r}_u = \frac{1}{\mathfrak{n}} \sum_n \left(\alpha_{u,n} - \alpha_{u,n} D_{v_0,n}\right) \mathfrak{r}_{u,n} \tag{18}$$

$$= \frac{1}{\mathfrak{n}} \sum_n \left(1 - D_{v_0,n}\right) \alpha_{u,n} \mathfrak{r}_{u,n}, \quad \forall u \tag{19}$$

- Do note that URLLC allocation disappears from the objective function of Problem 1, and URLLC demands are constants.

Linear Loss Model (Continued)

- This problem then can be solved with gradient algorithm [1].

$$
\begin{aligned}
\underset{\boldsymbol{\alpha}}{\text{maximize}} \quad & \sum_{u} ln\left(\bar{r_u}\right) \\
\text{subject to} \quad & \sum_{u} \alpha_{u,n,l} \leq \frac{1}{\mathfrak{l}}, \qquad \forall n, \forall l, \\
& \alpha_{u,n,l} \in \{0, \frac{1}{\mathfrak{l}}\}, \quad \forall u, \forall n, \forall l
\end{aligned} \tag{20}
$$

- However, this paper does not execute two problem transformations required to apply the above algorithm.

Convex Loss Model

- Recall that

$$\bar{r}_u = \frac{1}{\mathfrak{n}} \sum_n r_{u,n} \tag{21}$$

$$= \frac{1}{\mathfrak{n}} \sum_n (\alpha_{u,n} - \alpha_{u,n} f(x)) \, \mathfrak{r}_{u,n} \tag{22}$$

- where $f(\cdot)$ is a convex function, and

$$x = \frac{\beta_{v_0,u,n}}{\alpha_{u,n}} \tag{23}$$

- Thus

$$\bar{r}_u = \frac{1}{\mathfrak{n}} \sum_n \left(\alpha_{u,n} - \alpha_{u,n} f\left(\frac{\sum_m \beta_{v_0,u,n,m}}{\alpha_{u,n}}\right) \right) \mathfrak{r}_{u,n}, \quad \forall u \tag{24}$$

Convex Loss Model (Continued)

- They first fluidize the bandwidth:

$$
\begin{aligned}
\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{maximize}} \quad & \sum_u ln\left(\bar{r_u}\right) \\
\text{subject to} \quad & \sum_u \alpha_{u,n} = 1 \qquad \forall n, \\
& \alpha_{u,n} \geq 0 \qquad \forall u, \forall n, \\
& \sum_u \beta_{v_0,u,n,m} = D_{v_0,n,m} \forall n, \forall m, \\
& \beta_{v_0,u,n,m} \leq \alpha_{u,n,m} \ \forall u, \forall n, \forall m, \\
& \beta_{v_0,u,n,m} \geq 0 \qquad \forall u, \forall n, \forall m
\end{aligned} \tag{25}
$$

- where

$$\alpha_{u,n,m} = \frac{\alpha_{u,n}}{\mathfrak{m}}, \quad \forall u, \forall n, \forall m \tag{26}$$

Convex Loss Model (Continued)

- I can prove the objective function is concave (by perspective function), and that this is a convex optimization problem.
- Nevertheless, how can we allocate eMBB *resources* for the current eMBB time slot if the problem depends on unknown URLLC demands in the future?

Convex Loss Model (Continued)

- Assume minislot-homogeneous policy:

$$\beta_{v_0,u,n,m} = \frac{\frac{\gamma_{v_0,u,n}}{\mathfrak{m}}}{\frac{1}{\mathfrak{m}}} D_{v_0,n,m} \tag{27}$$

$$= \gamma_{v_0,u,n} D_{v_0,n,m}, \quad \forall u, \forall n, \forall m, \tag{28}$$

- where

$$\gamma_{v_0,u,n} \geq 0, \quad \forall u, \forall n \tag{29}$$

$$\sum_u \gamma_{v_0,u,n} = 1, \quad \forall n. \tag{30}$$

- Thus

$$\sum_u \beta_{v_0,u,n,m} = D_{v_0,n,m}, \quad \forall n, \forall m \tag{31}$$

Convex Loss Model (Continued)

- *If*

$$\frac{\alpha_{u,n}}{\gamma_{v_0,u,n}} \geq 1, \quad \forall u, \forall n \tag{32}$$

- then combining with $\mathfrak{m} D_{v_0,n,m} \leq 1, \forall n, \forall m$,

$$\frac{\alpha_{u,n}}{\gamma_{v_0,u,n}} \geq \mathfrak{m} D_{v_0,n,m}, \quad \forall u, \forall n, \forall m \tag{33}$$

- which materializes

$$\beta_{v_0,u,n,m} \leq \alpha_{u,n,m}, \quad \forall u, \forall n, \forall m \tag{34}$$

Convex Loss Model (Continued)

- However, this implies that

$$\gamma_{v_0,u,n} = \alpha_{u,n}, \quad \forall u, \forall n \tag{35}$$

- Proof by contradiction: Assume $\exists u : \exists n : \gamma_{v_0,u,n} > \alpha_{u,n}$, then at that n:

$$\sum_u \gamma_{v_0,u,n} > \sum_u \alpha_{u,n} \tag{36}$$

$$1 > 1 \tag{37}$$

Convex Loss Model (Continued)

- But such an assumption materializes

$$\sum_{u} \gamma_{v_0,u,n} = \sum_{u} \alpha_{u,n} = 1, \quad \forall n \tag{38}$$

- They introduce the notation of $(1 - \delta)$ factor to argue away some fundamental issues.

Convex Loss Model (Continued)

- Such an assignment eliminates the URLLC constraints.
- The average rate is then

$$\bar{r}_u = \frac{1}{\mathfrak{u}} \sum_n \left(\alpha_{u,n} - \alpha_{u,n} f\left(\frac{\gamma_{v_0,u,n} D_{v_0,n}}{\alpha_{u,n}}\right) \right) \mathfrak{r}_{u,n}, \quad \forall u \qquad (39)$$

- Finally, they hide $D_{v_0,n}$ from their problem formulation and state that it is a convex optimization problem.

$$\max_{\phi^s, \gamma^s} \sum_{u \in \mathcal{U}} U_u' \left(\overline{r}_u(t-1)\right) g_u^s(\phi_u^s, \gamma_u^s), \qquad (9)$$

$$\text{s.t.} \quad \boldsymbol{\phi}^s \geq (1-\delta)\, \boldsymbol{\gamma}^s, \qquad (10)$$

$$\sum_{u \in \mathcal{U}} \phi_u^s = 1 \text{ and } \sum_{u \in \mathcal{U}} \gamma_u^s = 1, \qquad (11)$$

$$\boldsymbol{\phi}^s \in [0,1]^{|\mathcal{U}|} \text{ and } \boldsymbol{\gamma}^s \in [0,1]^{|\mathcal{U}|}. \qquad (12)$$

Figure: Problem

Issues

- Multiple URLLC users scenario is neglected.
- Subchannel-wise rate is neglected.
- Physical resource block allocation of eMBB (in linear loss model *analysis* and convex loss model) and URLLC (in linear and convex loss models) is neglected.
- Superposition interference in convex loss model is neglected.
- Bandwidth ‘fluidizability’ in convex loss model is not guaranteed.
- Homogeneity in convex loss model is not guaranteed.

Notes

- They incorrectly address Pareto optimality, as their problem is not a vector optimization problem.
- They incorrectly reference gradient algorithm's moving average rate update function (twice).
- They model channel states, but I think it was sufficient to describe channel states using time slot index.
- They hide the formulation of linear loss model.

References

[1] Alexander L. Stolyar. "On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation". In: *Operations Research* 53.1 (2005), pp. 12–25. DOI: 10.1287/opre.1040.0156.

[2] Hao Yin, Lyutianyang Zhang, and Sumit Roy. "Multiplexing URLLC Traffic Within eMBB Services in 5G NR: Fair Scheduling". In: *IEEE Transactions on Communications* 69.2 (2021), pp. 1080–1093. DOI: 10.1109/TCOMM.2020.3035582.