Tian DANG, Chenxi LIU, Xiqing LIU, Shi YAN, 2022. Joint uplink and downlink resource allocation for low-latency mobile virtual reality delivery in fog radio access networks. *Frontiers of Information Technology & Electronic Engineering*, 23(1):73-85. https://doi.org/10.1631/FITEE.2100308

Joint uplink and downlink resource allocation for low-latency mobile virtual reality delivery in fog radio access networks

Key words: Virtual reality delivery; Fog radio access network (F-RAN); Round-trip latency; Resource allocation

Corresponding author: Chenxi LIU
E-mail: chenxi.liu@bupt.edu.cn
ORCID: https://orcid.org/0000-0002-9134-1235

Motivation

- The rapidly growing virtual reality (VR) traffic creates many challenges for current and future wireless networks, such as high data rate and low latency.
- Fog radio access networks (F-RANs) have been regarded as a promising network architecture for mobile VR delivery, because they deploy caching and computing functionalities near user equipment, thereby greatly alleviating the fronthaul traffic load and reducing the latency.
- Given the significance of the uplink and downlink transmissions in mobile VR delivery, it is therefore essential to address the impact of both uplink and downlink in designing the corresponding resource allocation strategies.

Contents

- We characterize the round-trip latency of mobile VR delivery in F-RANs, revealing its dependence on the communication, caching, and computation resource allocations.
- We formulate a round-trip latency minimization problem, subject to practical constraints on the uplink and downlink transmission capacities, the fog access point (F-AP) caching size, and the computation capabilities of the F-APs and VR users. We decompose this problem into two sub-problems, which are tackled with the branch-and-bound method and convex optimization. Finally, a simple yet effective iterative algorithm is proposed.
- We examine the impact of key system parameters on round-trip latency. Compared to various baselines, the proposed algorithm can effectively improve mobile VR delivery performance, especially when the downlink transmission capacity is relatively large.

System model

VR delivery in F-RANs

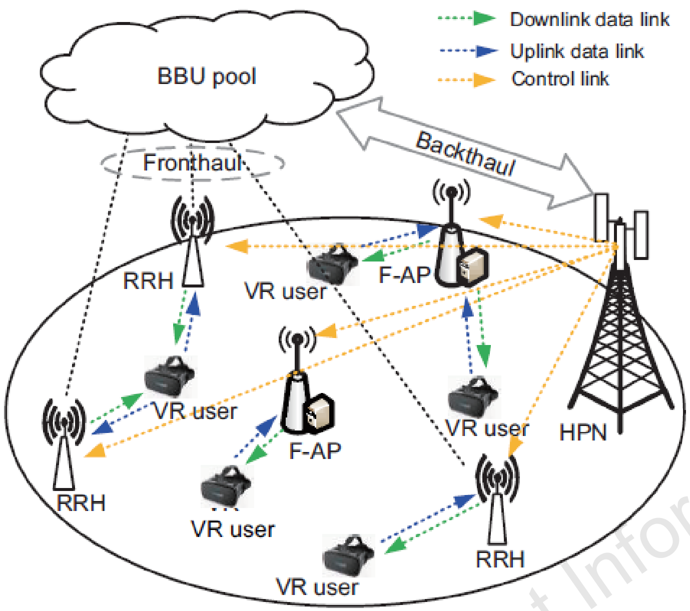


Fig. 1 VR delivery in F-RANs

BBU: baseband unit; RRH: remote radio head; VR: virtual reality; HPN: high-power node; F-AP: fog access point

Service modes for mobile VR delivery

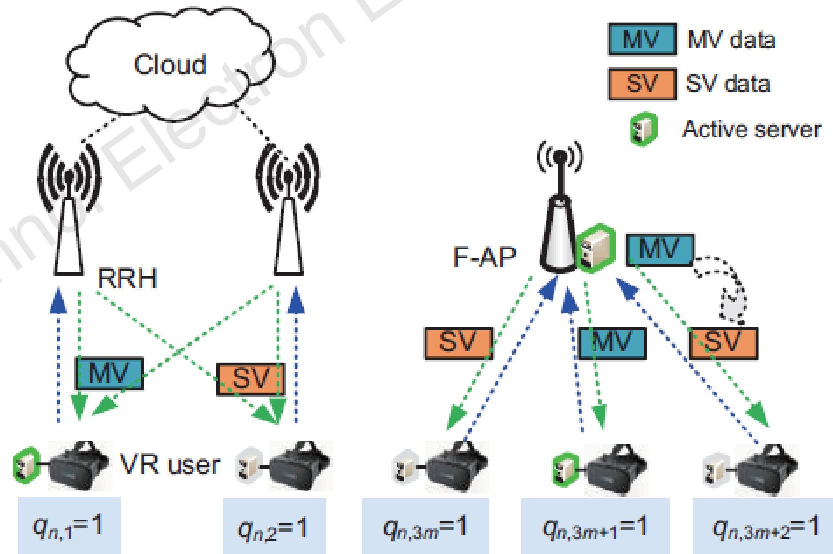


Fig. 2 Service modes for mobile VR delivery

The F-RAN-based mobile VR delivery system consists of one cloud server, one high-power node (HPN), MF-APs, *N* VR users, and multiple remote radio heads (RRHs). In the uplink, the tracking information is uploaded. In the downlink, the requested VR videos are delivered to the VR users.

The service modes considered in this work are summarized into five categories.

Caching, computation, and transmission model

Caching model

$$\sum_{k=1}^{K} c_{m,k}^{\mathrm{AM}} D_k^{\mathrm{M}} + c_{m,k}^{\mathrm{AS}} D_k^{\mathrm{S}} \le C_m^{\mathrm{A}}$$

- Caching variables of MVs and SVs at the F-APs
- Caching capacity of the F-APs
- Data sizes of MVs and SVs

Computation model

$$\sum_{n=1}^{N} l_{m,n}^{\mathrm{A}} d_{m,n}^{\mathrm{A}} \le L_m^{\mathrm{A}}$$

- Processing frequency allocation from the F-APs to the VR users
- Total processing frequency of the F-APs
- Task offloading decision of the VR users to the F-APs

$$\varepsilon_m^{\mathrm{A}} \sum_{k=1}^{K} w_k D_k^{\mathrm{M}} \sum_{n\in N_k} \left(l_{m,n}^{\mathrm{A}}\right)^2 d_{m,n}^{\mathrm{A}} \le E_m^{\mathrm{A}}$$

- Power efficiency of the F-APs
- Maximum energy consumption of the F-APs
- Number of cycles per bit for the VR services

$$\varepsilon_n^{\mathrm{V}} w_k D_k^{\mathrm{M}} \left(L_n^{\mathrm{V}}\right)^2 d_n^{\mathrm{V}} \le E_n^{\mathrm{V}}, \forall n \in \mathcal{N}_k$$

- Power efficiency of the VR users
- Task offloading decision of the VR users to itself
- Maximum energy consumption of the VR users
- Total processing frequency of the VR users

Transmission model

$$\sum_{n=1}^{N} a_{0,n} b_n^{\mathrm{U}} \le R^{\mathrm{UF}}$$

- User association between RRHs and VR users
- Uplink transmission capacity of the fronthaul
- Uplink data rate of the VR users

$$\sum_{n=1}^{N} a_{m,n} b_n^{\mathrm{U}} \le R_m^{\mathrm{Umax}}$$

- Uplink transmission capacity of the F-APs

$$\sum_{n=1}^{N} a_{m,n} b_{m,n}^{\mathrm{A}} \le R_m^{\max}$$

- User association between F-APs and VR users
- Downlink transmission capacity of the F-APs
- Downlink data rate from the F-APs to the VR users

$$\sum_{n=1}^{N} a_{0,n} b_n^{\mathrm{R}} \le R^{\mathrm{F}}$$

- Downlink transmission capacity of the fronthaul
- Downlink data rate from the RRHs to the VR users

Round-trip latency

Total round-trip latency

$$\tau = \sum_{n=1}^{N} \left(\tau_n^{\mathrm{UL}} + \tau_n^{\mathrm{DL}} + \tau_n^{\mathrm{CP}} \right)$$

- τ_n^{UL}: Uplink transmission latency
- τ_n^{CP}: Processing latency
- τ_n^{DL}: Downlink transmission latency

Round-trip latency of each VR user for different modes

$$\tau_n^{\mathrm{UL}} = \frac{S_n}{b_n^{\mathrm{U}}}$$

$$\tau_n^{\mathrm{DL}} = \begin{cases} \dfrac{S_n^{\mathrm{DL}}}{b_n^{\mathrm{R}}}, & i = 1, 2 \\ \dfrac{S_n^{\mathrm{DL}}}{b_{m,n}^{\mathrm{A}}}, & i = 3m, 3m+1, 3m+2 \end{cases}$$

$$\tau_n^{\mathrm{CP}} = \begin{cases} \dfrac{C_n}{L_n^{\mathrm{V}}}, & i = 1, 3m+1 \\ \dfrac{C_n}{l_{m,n}^{\mathrm{A}}}, & i = 3m+2 \\ 0, & i = 2, 3m \end{cases}$$

$$\tau_{n,i}(\boldsymbol{b}, \boldsymbol{l}) = \begin{cases} \dfrac{S_n}{b_n^{\mathrm{U}}} + \dfrac{S_n^{\mathrm{DL}}}{b_n^{\mathrm{R}}} + \dfrac{C_n}{L_n^{\mathrm{V}}}, & i = 1 \\ \dfrac{S_n}{b_n^{\mathrm{U}}} + \dfrac{S_n^{\mathrm{DL}}}{b_n^{\mathrm{R}}}, & i = 2 \\ \dfrac{S_n}{b_n^{\mathrm{U}}} + \dfrac{S_n^{\mathrm{DL}}}{b_{m,n}^{\mathrm{A}}}, & i = 3m \\ \dfrac{S_n}{b_n^{\mathrm{U}}} + \dfrac{S_n^{\mathrm{DL}}}{b_{m,n}^{\mathrm{A}}} + \dfrac{C_n}{L_n^{\mathrm{V}}}, & i = 3m+1 \\ \dfrac{S_n}{b_n^{\mathrm{U}}} + \dfrac{S_n^{\mathrm{DL}}}{b_{m,n}^{\mathrm{A}}} + \dfrac{C_n}{l_{m,n}^{\mathrm{A}}}, & i = 3m+2 \end{cases}$$

Problem formulation and decomposition

$$\min_{\boldsymbol{b},\boldsymbol{l},\boldsymbol{q}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i}(\boldsymbol{b},\boldsymbol{l})q_{n,i}$$

s.t.

$$\text{C1:} \sum_{k=1}^{K}\sum_{n\in \mathrm{N}_k}\sum_{i=1}^{3M+2} p_{n,i}^{(1,m)} q_{n,i} \le C_m^{\mathrm{A}}$$

$$\text{C2:} \sum_{k=1}^{K}\sum_{n\in \mathrm{N}_k}\sum_{i=1}^{3M+2} p_{n,i}^{(2,m)}(\boldsymbol{l}) q_{n,i} \le E_m^{\mathrm{A}}$$

$$\text{C3:} \sum_{n=1}^{N}\sum_{i=1}^{3M+2} p_{n,i}^{(3,m)}(\boldsymbol{l}) q_{n,i} \le L_m^{\mathrm{A}}$$

$$\text{C4:} \sum_{i=1}^{3M+2} p_{n,i}^{(4)} q_{n,i} \le E_n^{\mathrm{V}}, \forall n \in \mathcal{N}_k$$

$$\text{C5:} \sum_{n=1}^{N}\sum_{i=1}^{3M+2} p_{n,i}^{(5)}(\boldsymbol{b}) q_{n,i} \le R^{\mathrm{UF}}$$

$$\text{C6:} \sum_{n=1}^{N}\sum_{i=1}^{3M+2} p_{n,i}^{(6,m)}(\boldsymbol{b}) q_{n,i} \le R_m^{\mathrm{Umax}}$$

$$\text{C7:} \sum_{n=1}^{N}\sum_{i=1}^{3M+2} p_{n,i}^{(7,m)}(\boldsymbol{b}) q_{n,i} \le R_m^{\max}$$

$$\text{C8:} \sum_{n=1}^{N}\sum_{i=1}^{3M+2} p_{n,i}^{(8)}(\boldsymbol{b}) q_{n,i} \le R^{\mathrm{F}}$$

$$\text{C9:} \sum_{i=1}^{3M+2} q_{n,i} = 1$$

$$\text{C10:}\ q_{n,i} \in \{0,1\}$$

Service mode selection (branch-and-bound method)

$$\min_{\boldsymbol{q}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i} q_{n,i} \quad \text{s.t. C1-C10}$$

Relax

$$\min_{\boldsymbol{q}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i} q_{n,i}$$

s.t. C1-C9

$$\text{C10}': q_{n,i} \in [0,1]$$

Branch

$$\min_{\boldsymbol{q}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i} q_{n,i}$$

s.t. C1-C10

$$\text{C11}: q_{n,i} = 0$$

$$\min_{\boldsymbol{q}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i} q_{n,i}$$

s.t. C1-C10

$$\text{C11}: q_{n,i} = 1$$

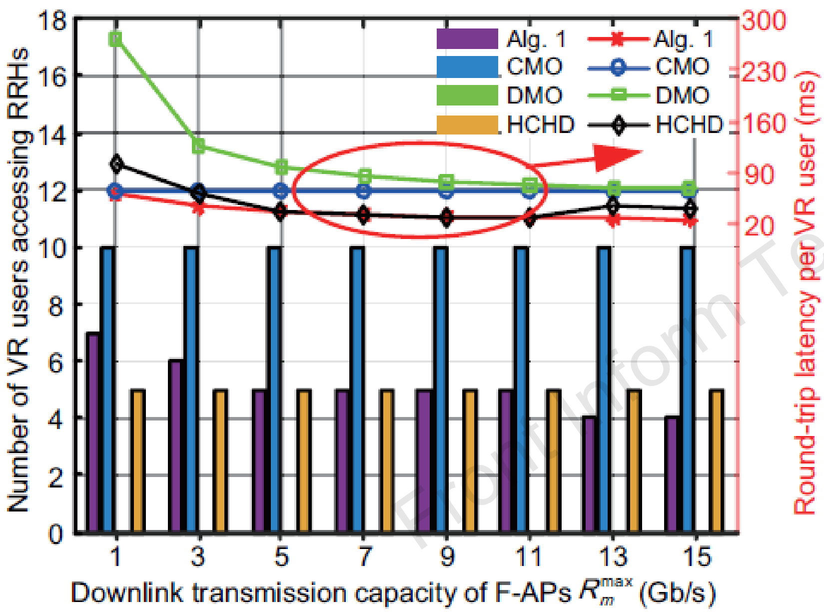
Solve alternately

Joint data rate and frequency allocation (convex optimization)

$$\min_{\boldsymbol{b},\boldsymbol{l}}: \sum_{n=1}^{N}\sum_{i=1}^{3M+2} \tau_{n,i}(\boldsymbol{b},\boldsymbol{l}) q_{n,i} \quad \text{s.t. C2, C3, C5-C8}$$

Numerical results

Number of VR users accessing RRHs and round-trip latency vs. downlink transmission capacity of F-APs



The proposed algorithm outperforms the benchmarks.

Round-trip latency vs. average data size of MVs

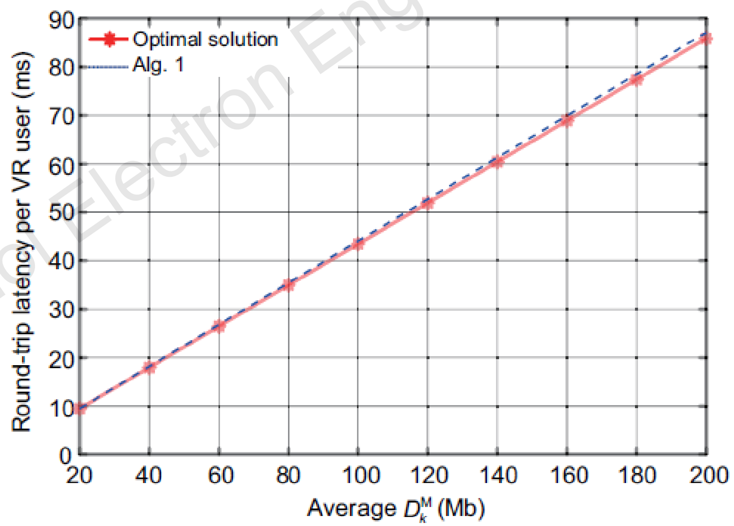


Fig. 4 Round-trip latency per VR user vs. average D_k^{M}

The latency achieved by the proposed algorithm is close to the optimum one.

Numerical results

Round-trip latency vs. cache capacity of F-APs

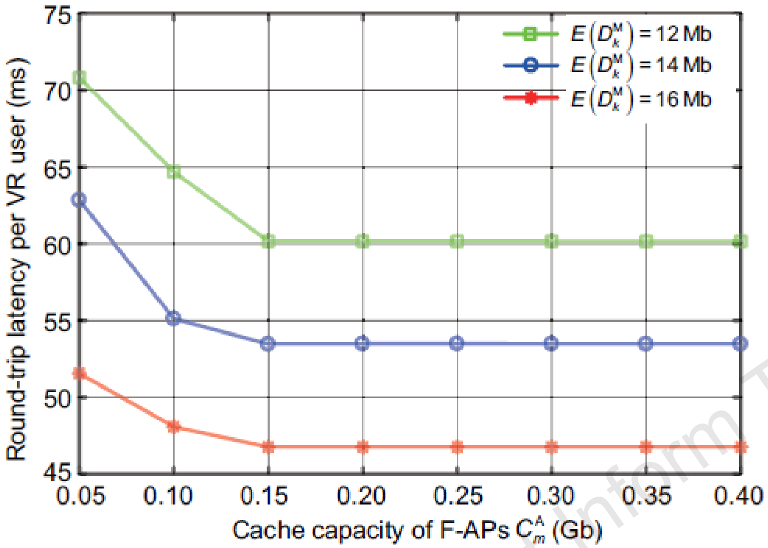


Fig. 5 Round-trip latency per VR user vs. caching capacity of F-APs C_m^{A} ($N = 10$ and $K = 10$)

A larger cache capacity helps reduce the round-trip latency.

Round-trip latency and computing energy consumption vs. computing frequency of F-APs

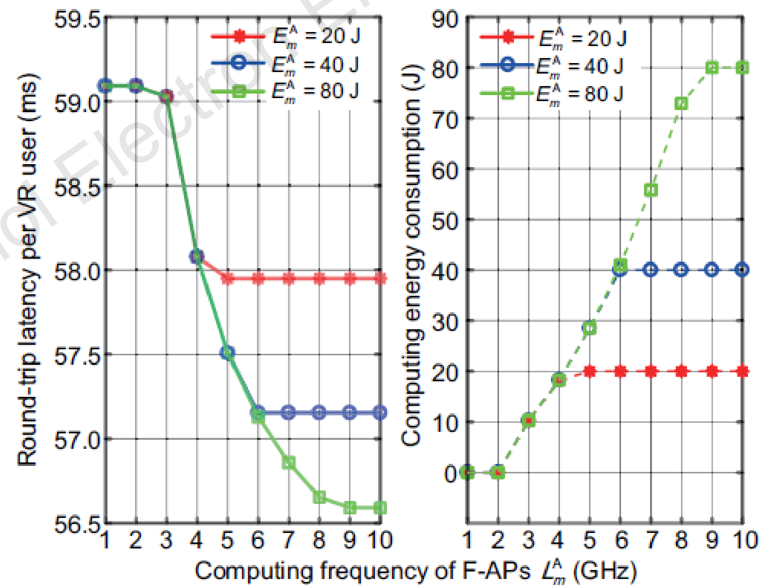


Fig. 6 Round-trip latency per VR user and computing energy consumption vs. computing frequency of F-APs ($K = 1$, $L_n^{\mathrm{V}} = 1.5$ GHz, $E_n^{\mathrm{V}} = 30$ J, $C_m^{\mathrm{A}} = 0.1$ Gb, and $R^{\mathrm{F}} = 3.5$ Gb/s)

Both the computing frequency and the maximum energy consumption of the F-APs have significant impacts on the round-trip latency.

Numerical results

Round-trip latency vs. uplink and downlink transmission capacities

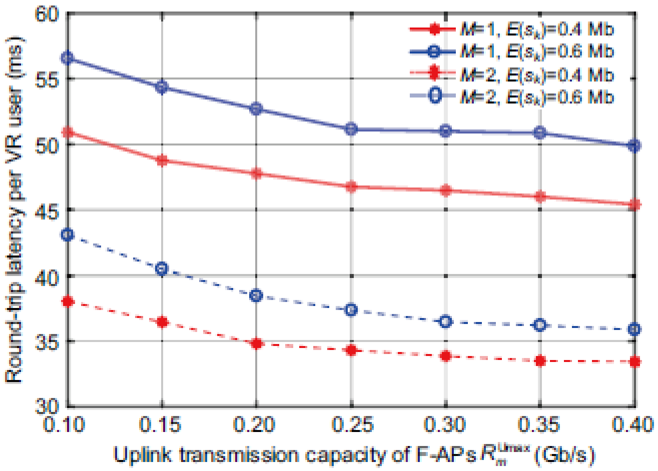


Fig. 7 Round-trip latency per VR user vs. uplink transmission capacity of F-APs R_m^{Umax}

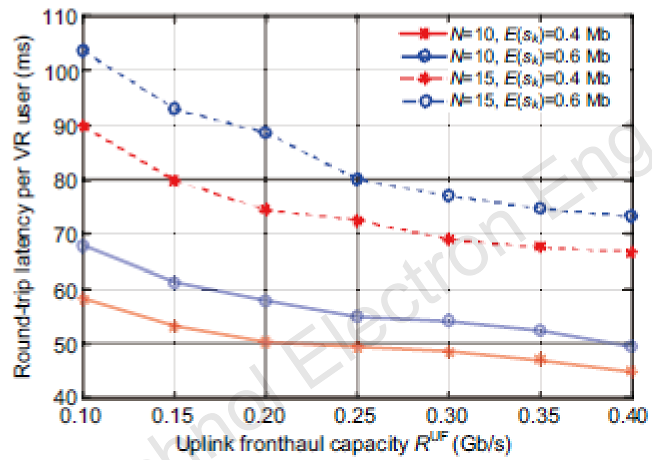


Fig. 8 Round-trip latency per VR user vs. uplink fronthaul capacity R^{UF}

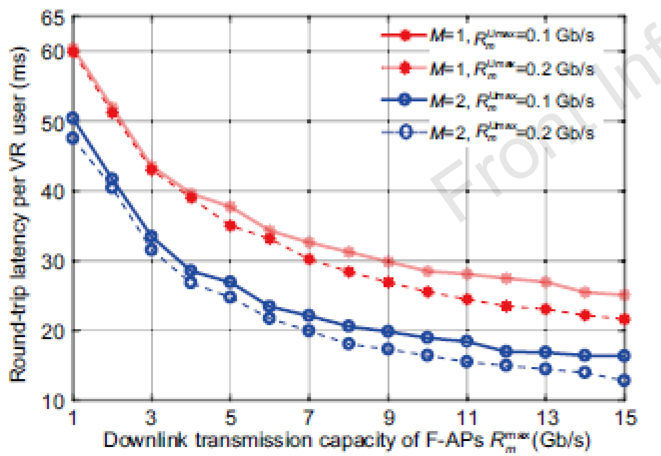


Fig. 9 Round-trip latency per VR user vs. downlink transmission capacity of F-APs R_m^{max} (10 VR users, 6 VR services)

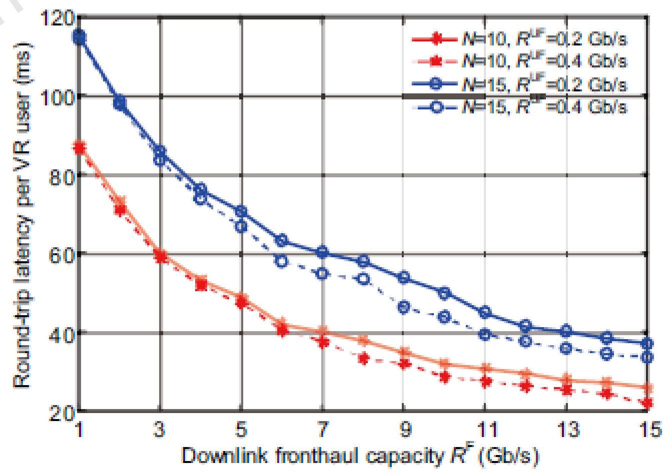


Fig. 10 Round-trip latency per VR user vs. downlink fronthaul capacity R^{F} for $R_m^{\text{Umax}} = 0.2$ Gb/s ($M = 1$, $K = 6$)

The round-trip latency is reduced by increasing either the uplink or the downlink transmission capacity.

Conclusions

- In this study, we investigated a mobile virtual reality delivery framework in fog radio access networks, where both the uplink and downlink transmissions were considered.
- We characterized the round-trip latency and showed how it was determined by the communication, caching, and computation resource allocations.
- A simple yet efficient algorithm was proposed to minimize the round-trip latency, while satisfying the practical constraints on caching, computation capabilities, and transmission capacity.
- Numerical results were provided to verify the effectiveness of our proposed algorithm and the impacts of the communication, caching, and computation resources on the round-trip latency.