Video Communication for Teleconferencing using Edge Computing

Kouichi Genda¹, Mitsuru Abe², and Shohei Kamamura³

¹Department of Computer Science, College of Engineering, Nihon University
1 Nakagawara, Tokusada, Tamura, Koriyama-shi, Fukushima 963-8642, Japan

²NTT Communications Corporation
2-3-1 Otemachi, Chiyoda-ku, Tokyo 100-8019, Japan

³NTT Network Service Systems Laboratories, NTT Corporation
3-9-11 Midori-cho, Musashino-shi, Tokyo 466-8555, Japan
¹genda.kouichi@nihon-u.ac.jp

Abstract: This paper proposes a backbone network resource optimization algorithm for video communications of teleconferencing that use edge computing. In the current video communication architecture, the key component, the multi-point control unit (MCU), is deployed in the central cloud server, and its bandwidth consumption in the backbone network becomes huge as the video resolution and the frequency of use increase. By using edge computing, MCU can be deployed at the entrance node of the backbone network. This can reduce the bandwidth consumption of the backbone network. However, the edge deployment and routing (EDR) problem, which is classified as NP-hard, should be solved to achieve sufficient bandwidth reduction. To solve the EDR problem within a feasible time, we propose a divide and merge algorithm using the linear programming approach. We demonstrate that, with our algorithm, bandwidth consumption using edge computing is reduced by approximately 30% compared to the current architecture on a world-wide network.

Keywords: Video Communication, Edge Computing, Linear Programming, Network Optimization, Multi-access Edge Computing

Classification: Network

References

[1] “Multi-access Edge Computing (MEC),” website, https://www.etsi.org/technologies/multi-access-edge-computing, accessed on March 2020.

[2] P. Mach and Z. Becvar, “Mobile Edge Computing: A Survey on Architecture and Computation Offloading,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1628–1656, 2017.

[3] “Bandwidth optimization within PexipInfinity,” website, https://docs.pexip.com/admin/bandwidth_management.htm, accessed on March 2020.
1 Introduction

Video communication for teleconferencing has become widespread because of the evolution of high-efficiency compression and mobile communication technologies. Especially, video communication is indispensable for smooth remote work and education during pandemics such as COVID-19.

In the current video communication architecture, the key component, called the multi-point control unit (MCU), is deployed in the central cloud server. MCU combines received multiple video streams into a single video stream. As the video resolution and the frequency of use increase, a big concern is that the bandwidth consumption in the backbone network is too large to provide stable video communication.

Multi-access edge computing (MEC) architecture in the 5G backbone network is standardized by the ETSI [1]. Various functions that are originally deployed on the central cloud server are deployed on the edge nodes. Although many researchers are tackling a variety of topics that are related to MEC architecture [2], the existing literature focuses on mobile communication, not on video communication for teleconferencing.

We focus on applying the MEC architecture to video communication because of its large potential for reducing the bandwidth consumption of the backbone network. However, it is necessary to solve the edge deployment and routing (EDR) problem to achieve sufficient bandwidth reduction, where the EDR problems is defined to optimize the placement of the edge nodes with the MCUs and the routing between the users. As the EDR problem is generally formulated as a mixed integer linear programming (MILP) problem, and is classified into NP-hard problems, it is difficult to solve the problem within a feasible computation time.

Here, we propose a novel but simple method to solve the EDR problem of the MEC architecture within a feasible time frame. This study has two main contributions. First, we provided a model for applying the MEC architecture to video communication, and defined the problem called the EDR problem. Second, we provided the simple algorithm, and its effectiveness was demonstrated on a world-wide network.

2 Video streams for teleconferencing in MEC architecture

Fig. 1 illustrates the video streams over the backbone network in the MEC architecture. The video streams are routed from the users to the edge nodes (ENs) with the MCUs, via subscriber nodes (SNs) where the users’ video streams are accommodated.

When the video streams from all users who join the video conference are accommodated in the MCU, such as users \(a\) and \(b\), they are combined into a single

[4] J. Chu and C. Lea, “Optimal link weights for maximizing QoS traffic,” in Proc. IEEE International Conference on Communications (ICC), pp. 610-615, 2007.

[5] C.-Y. Hong, S. Mandal, M. Al-Fares, M. Zhu, et al., “B4 and after: managing hierarchy, partitioning, and asymmetry for availability and scale in Google’s software-defined WAN,” in Proc. of ACM SIGCOMM, pp. 74-87, 2018.
video stream at the MCU, and then the video stream is sent back to the users. This is called local loopback. In contrast, when the video streams from the users who join the video conference are accommodated in the different MCUs, they are exchanged between the MCUs, where the video streams are compressed to thumbnailed video streams with a reduced bit rate, about 1/10 [3]. This is called traffic compression. In Fig. 1, when user B is the current main speaker, the video streams of users A and C are sent as thumbnailed videos between MCU1 and MCU2.

Video streams have specific characteristics of being terminated at the MCUs on their routes. This is different from the typical characteristics of end-to-end mobile communication traffic.

3 Divide and merge algorithm

3.1 Overview

We propose the divide and merge algorithm to solve the EDR problem within a feasible computation time. The idea is based on the characteristics that the video traffic is terminated at the ENs. The EDR problem is divided into two sub-problems under a practical assumption that the number and the location of the ENs are designed by the business policy.

The first sub-problem is to determine the appropriate network resources and routings for the video streams between the ENs and SNs. The second sub-problem is to determine the resource allocation and routing for the video streams between the ENs. We solved these sub-problems not by using the MILP but by using the LP formulation to reduce the computation time. Finally, the results of the sub-problems are merged to obtain the total bandwidth consumption in the backbone network. The LP formulation in the sub-problems is discussed in the following Sections.

The proposed algorithm provides a sub-optimal solution that is the upper bound of the optimal solution that provides the minimum bandwidth consumption of the backbone network. This is because the bandwidth is sequentially determined in a stepwise manner.

3.2 Video streams between ENs and SNs

We assumed that the total ingress/egress traffic demand for each SN is estimated in advance, but the traffic matrix between the SNs is inexplicit because
the video communication between the arbitrary users would be frequently used under the mature video communication environment. In addition, we assumed that the video streams at the SN may be adequately routed between the SN and multiple ENs. This assumption is realistic because an appropriate route per video stream can be designed using the segment routing over MPLS technique. The limitations of the MCU, such as the maximum capacity, are ignored here because we can reflect them if necessary.

The LP is formulated as follows.

\[
\begin{align*}
\text{minimize} & \quad \sum_{(i,j) \in E} \rho_{ij} \cdot C_{ij}, \\
\text{subject to} & \quad \sum_{j : (i,j) \in E} x_{ij}^s - \sum_{j : (j,i) \in E} x_{ji}^s = f_{ij}^s, i = s, \quad (1-2) \\
& \quad \sum_{j : (i,j) \in E} x_{ij}^t - \sum_{j : (j,i) \in E} x_{ji}^t = 0, i \neq s, t, \quad (1-3) \\
& \quad \sum_{(i,j) \in P} x_{ij}^s \leq \rho_{ij} \cdot C_{ij}, \quad (1-4) \\
& \quad \sum_{i : (1, i) \in P} f_{ij}^s \cdot E^s = D^s, \quad (1-5)
\end{align*}
\]

Eq. (1-1) is an objective function minimizing the bandwidth consumption between SNs and ENs in the network \( G(V, E) \), where \( V \) is a set of nodes consist of SNs and ENs, and \( E \) is a set of links. \( C_{ij} \) is the capacity of \( (i, j) \in E \), where \( \rho_{ij} \cdot 0 \leq \rho_{ij} \leq 1 \), is the utilization rate of \( (i, j) \). Eqs. (1-2) and (1-3) are the constraints for the flow conservation. \( f^s \), a variable, represents the video traffic of \( (s, t) \in P \), where \( P \) is a set of logical paths between the SN and EN. \( x_{ij}^s \), a variable, is the portion of the video stream from \( s \) to \( t \) that is routed through \( (i, j) \). Eq. (1-4) indicates the capacity constraint of \( (i, j) \). Eq. (1-5) states the constraint of the video stream bandwidth at the SN. \( E^s \) indicates that \( s \) belongs to SN, where \( E^s = 1 \) if \( s \) is SN. \( D^s \), a given parameter, represents the total traffic demand at the SN, where \( D^s = 0 \) if \( s \) is EN.

### 3.3 Video streams between ENs

We treat the video streams between the ENs as a hose model. This is because we design appropriate resources by using the total ingress/egress video streams at each EN under the assumption that the traffic matrix between the SNs is inexplicit. The total video stream at each EN is the sum of the user traffic that the EN directly accommodates and the traffic between the EN and SNs. All the video streams between the ENs are assumed to be thumbnailed to simplify the problem.

\[
\begin{align*}
\text{minimize} & \quad \sum_{(i,j) \in E} \rho_{ij} \cdot C_{ij}^r, \\
\text{subject to} & \quad \sum_{j : (i,j) \in E} y_{ij}^u - \sum_{j : (j,i) \in E} y_{ji}^u = 1, i = u, \quad (2-2) \\
& \quad \sum_{j : (i,j) \in E} y_{ij}^v - \sum_{j : (j,i) \in E} y_{ji}^v = 0, i \neq u, v, \quad (2-3) \\
& \quad \sum_{u : (u, v) \in P} x_{uv}^u \cdot R^u + \sum_{v : (u, v) \in P} x_{uv}^v \cdot R^v \leq \rho_{ij} \cdot C_{ij}^r, \quad (2-4)
\end{align*}
\]
Eqs. (2-4) and (2-5) state the capacity constraint of \((i, j)\), where the hose model formulation is converted into a regular LP problem by using the duality theorem [4]. \(R^u\) is the total egress video stream from \(u\). \(C^v\) is the total ingress video stream to \(v\). \(\pi_{ij}^u\) and \(\lambda_{ij}^v\) are non-negative variables.

4 Performance evaluation

We evaluated the effectiveness of the MEC architecture for the video communication in terms of the bandwidth consumption in the backbone network. The bandwidth consumption in the MEC architecture is computed using the proposed algorithm. We compared our results to the bandwidth consumption of a central cloud architecture, which is set as the benchmark.

The performance on a world-wide network in Fig. 2 [5] is simulated. The traffic demand of video communication at each location is in proportion to the population density. In the MEC architecture, the traffic streams between the ENs are reduced.

Fig. 2 Network and traffic model evaluated.

Fig. 3 Bandwidth consumption in the backbone network using the MEC and the central cloud architectures.
to 1/8 by traffic compression. The bandwidth consumption between the ENs is evaluated under the worst condition where there are no local loopback. GLPK was used as the LP solver.

Fig. 3 shows the bandwidth consumption in the backbone network. The shaded area indicates the bandwidth consumption between the SNs and ENs while the white area indicates that between the ENs. The bandwidth consumption in the MEC architecture with three MCUs is reduced by about 30% in comparison to that in the benchmark. Because the computed bandwidth in the MEC architecture is the upper bound of the optimal bandwidth, the results indicate that the MEC architecture is effective in reducing the bandwidth consumption.

Although we omitted the detailed discussion of the computation time, we confirmed that the time for calculating the bandwidth between the SNs and ENs and the bandwidth between the ENs was less than 1 minute for every evaluated model by using a general Intel CORE i7 CPU and 32 GB of memory.

4 Conclusion

We proposed a sub-optimal algorithm to solve the EDR problem for video communication of teleconferencing in MEC architecture. With our algorithm, the bandwidth consumption of the MEC architecture is reduced by about 30% on a world-wide network in comparison to the current architecture. As the next works, we will tackle the optimal EDR problem and other sub-optimal solutions.