Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1061807, 13 pages
https://doi.org/10.1155/2018/1061807

Research Article

Wireless Adaptive Video Streaming with Edge Cloud

Kristofer R. Smith,[1] Hang Liu ,[1] Li-Tse Hsieh,[1] Xavier de Foy,[2] and Robert Gazda[3]

[1] *Department of Electrical Engineering & Computer Science, The Catholic University of America, Washington, DC, USA*
[2] *InterDigital Canada Ltée, Montreal, QC, Canada*
[3] *InterDigital Communications, Inc., Conshohocken, PA, USA*

Correspondence should be addressed to Hang Liu; liuh@cua.edu

Received 7 September 2018; Accepted 22 November 2018; Published 5 December 2018

Academic Editor: Petros Nicopolitidis

Wireless data traffic, especially video traffic, continues to increase at a rapid rate. Innovative network architectures and protocols are needed to improve the efficiency of data delivery and the quality of experience (QoE) of mobile users. Mobile edge computing (MEC) is a new paradigm that integrates computing capabilities at the edge of the wireless network. This paper presents a computation-capable and programmable wireless access network architecture to enable more efficient and robust video content delivery based on the MEC concept. It incorporates in-network data processing and communications under a unified software-defined networking platform. To address the multiple resource management challenges that arise in exploiting such integration, we propose a framework to optimize the QoE for multiple video streams, subject to wireless transmission capacity and in-network computation constraints. We then propose two simplified algorithms for resource allocation. The evaluation results demonstrate the benefits of the proposed algorithms for the optimization of video content delivery.

1. Introduction

Mobile multimedia content, especially video traffic, has been growing at a rapid pace. It is anticipated that global mobile video traffic will increase 11-fold from 2016 to 2021, reaching 23 exabytes per month and accounting for 75 percent of total mobile data traffic by 2021 [1]. Innovative network architectures and protocols are needed to improve the efficiency of data delivery and the quality of experience (QoE) of mobile users. The trend of combining advanced information technology and communications is creating unprecedented opportunity for innovation when designing network-centric services. The rapid growth in cloud computing is an outcome of such integration. However, the traditional cloud computing model that depends on centralized data centers has many limitations. For example, a mobile device may experience varying wireless connectivity due to mobility, interference, and wireless traffic load. The content service provider and its cloud data centers are far away and simultaneously serve many users, making it difficult to adapt traffic (video coding rate) to dynamic wireless network conditions.

To address the shortcomings of data centers, a new paradigm, mobile edge computing (MEC), also termed fog computing, has recently gained attention in both academia and industry [2–6], which integrates computing and storage capabilities at the edge of the network, near mobile users, to provide real-time radio-aware services and enhance user experience. MEC servers or microcloud with computing, storage, and communication capabilities can be deployed to process data in the wireless access networks, e.g., adapting video rates to changing network conditions, to improve the QoE of video streaming.

The ability to dynamically program and reconfigure networks is essential for improving quality of service (QoS) and ensuring efficient content delivery. Software-defined networking (SDN) is considered to be a key technology enabler for meeting these requirements [7–10]. Decoupling the data plane and control plane in SDN along with an open interface enables network administrators to have programmable control of network traffic, thereby flexibly and predictably routing specific data flows, as well as extending network functions. However, SDN is currently only used to control forwarding and routing of data flows. It does not offer logic to control in-network computation and lacks the capability to handle network-aware processing.

In this paper, we propose a wireless access network architecture based on the MEC and SDN concepts, called CloudEdge, for more efficient and robust video streaming and content delivery. CloudEdge turns a wireless access network into a programmable microcloud, integrating in-network data processing capabilities with wireless access and an open interface. Computing servers can be deployed in the wireless access network, and various resources are controlled by a unified orchestration platform to fully exploit the potential of in-network processing to enhance user experience and optimize traffic in the network. For example, if a user is receiving video streamed to his mobile device and his wireless connection deteriorates due to congestion, mobility, and/or interference, the subsequent video objects can be automatically routed to a virtual server instance in the wireless access network that transcodes the video objects to match the current radio channel conditions. The transcoded video objects are then forwarded to the mobile user. This ensures that the mobile user will get the best available QoE given the existing wireless network conditions.

With integrated in-network data processing and wireless delivery, it is possible for the network to optimize the overall performance based on the demands of multiple users. To achieve this potential, multiple resources in the network including bandwidth and CPU cycles need to be carefully managed. Realizing such system performance optimization is challenging because multiple resources are shared by multiple users whose wireless channels are dynamic and QoS requirements are diverse. We develop a multiresource management solution in this paper to demonstrate the benefits of the proposed system architecture. Specifically, we formulate a framework to optimize the QoE of multiple video streams, subject to wireless channel bandwidth and in-network video transcoding capability constraints.

We focus on wireless video content delivery as an example application in this paper. However, the architecture and techniques developed here are expected to be useful in the design to support other types of applications, e.g., video surveillance and smart traffic control, as well, and such a computation-capable and programmable networking paradigm will enable a range of new applications.

The rest of the paper is organized as follows: Section 2 describes the proposed CloudEdge system architecture. In Section 3, the optimal resource management framework for multiple video streams is derived. Section 4 proposes two simplified algorithms for resource management. The evaluation methods and results are given in Section 5. Section 6 summarizes the related work. Finally, Section 7 concludes the paper and discusses future work.

2. System Architecture

The CloudEdge system architecture design takes on a network service-centric model. We envision that distributed computing and storage resources can be deployed in wireless access networks, e.g., cellular networks or WiFi hotspots, which will host virtual instances to dynamically perform customized computations on data as needed, as well as cache processed data while delivering it. The computation and storage hardware, such as general-purpose CPUs, GPUs, memory, flash storage, and specialized pluggable video transcoders, may be integrated into or colocated with base stations (BSs), access points (APs), backhaul switches, and routers. Network operators may also deploy dedicated microclouds or cloudlets in the wireless access networks. Such a computation-capable and programmable access network is able to provide not only data forwarding but in-network data processing with wireless network awareness and low latency, supporting various applications with rich services at the edge. The following principles are used in CloudEdge design:

(i) *Decoupling of the data plane and control plane, as well as the hardware and software*. Thus, traffic flows can be routed with greater flexibility, and software-based implementations of in-network services and applications can run on general-purpose or specialized hardware platforms.

(ii) *Unified control*. Wireless transmission, traffic routing, in-network data processing, and caching are managed by a single (logically) centralized controller. Extending the SDN concept, the controller takes a unified, network-wide view to generate configurations and policy rules for all traffic as well as in-network services. The bandwidth, computation, and storage resources can be abstracted and shared by multiple applications. This is different from standalone middleboxes [11], where each box or application is managed independently with little extensibility. Given the power of modern computing devices, as well as the potential scale of a wireless access network, the redundant clusters may be used as the controller platform [8, 9] to address single-point failure and scalability concerns.

(iii) *Open interface*. The resources in the wireless access network can be used by the network operator to provide in-network services. Such services could be user-requested services and/or transparent services. The latter is activated by the network automatically based on the preset rules. In addition, the network infrastructure can be virtualized, and the resources can be used to host third-party services as virtual instances through an open interface. This will simplify the service deployment for the third-parties and generate extra revenue for the network operator.

(iv) *Name-based network management*. Each device, service, content item, resource, and interface are given a unique name that is used for reachability and management.

CloudEdge can be implemented as an overlay on conventional IP networks, and it can also be compatible with name-based future network architectures [12–14]. In this paper, we focus on a specific use case, in-network rate adaptation of streaming video, to describe the CloudEdge design in more detail and demonstrate its benefits. We consider video because it is the predominant traffic in networks, and it will greatly benefit from in-network adaptation services.

Figure 1 illustrates the high-level system diagram of a CloudEdge wireless access network. Mobile users connect to the wireless access network through WiFi or cellular radios. The AP/BS is enhanced to report measured radio parameters, such as the link data rate of each mobile user and current bandwidth usage to the CloudEdge controller. The AP/BS also receives configuration commands from the controller. To reduce the controller load, the MAC function is split between the controller and the AP/BS. The controller provides the high-level policy such as the maximum share of channel time that can be used to transmit a particular data flow. The AP/BS will implement the schedule to transmit data using standard wireless protocols such as IEEE 802.11, 3GPP 4G Long-Term Evolution (LTE), or emerging 5G New Radio (NR), based on the resource allocation policy.

The bandwidth and computation resources can be virtualized and dynamically shared by different applications and services. The CloudEdge controller will monitor and manage all the resources in the wireless access network. All of the devices, such as routers, switches, APs/BSs, and computation servers have separate control and data interfaces. The control interfaces are used to communicate with the controller, reporting the resource usage and service status, and receiving commands from the controller. The controller also has an open interface for external applications that allows third-party application/content providers to lease resources and deploy their own services.

Each of the network-attached objects in CloudEdge, including devices, contents, services, applications, and interfaces, is assigned a unique identifier. A hierarchical naming scheme is used [12]. The identifier of a content object (CO), e.g., a video segment, is presented in a binary-coded format of /Publisher.Com/MovieTitle/SegmentID. Every deployed service is identified by a unique service identifier (SID) that could look like the following: /WirelessNetID/TranscodingServiceID. The CloudEdge controller maintains the mapping between the SID and the service resources. The controller is responsible for both establishing the data routing path and managing the in-network data processing.

When the edge router in CloudEdge receives a CO, it checks whether the rules have been set up to forward this CO in its forwarding table based on the CO header. If so, it will follow the rules to forward the CO to the next hop. Otherwise, it will buffer the CO and forward only the CO header to the controller through its control interface. The controller will then make the decision based on the resources demanded by the CO and the resources available in the wireless access network. If there is not enough wireless bandwidth to transmit the CO to the user, or an in-network service; e.g., video coding format change has been requested by the original sender (the CO header contains a requested SID); the controller will instruct the edge router to send the CO to the transcoding service for processing by adding a command field in the CO header and sending the header back to the edge router. The command field contains the routing rules from the edge router to the transcoder, and the rules for how to process the CO, such as the format, resolution, and data rate that the video content needs to be transcoded to, as well as the route from the transcoder to the user. The CO will be forwarded to the transcoder, and the transcoder reduces the video resolution and data rate per the controller's instruction. The transcoded video is then sent to the AP/BS that forwards to the user. If there is enough wireless bandwidth to forward the CO to the user and no in-network data processing is requested for the CO, the controller will instruct the edge router to directly forward the CO data to the user through the AP/BS by adding a source route in the CO header before sending the modified CO header to the edge router over the control interface. The controller may instruct the edge router and AP/BS to handle additional COs belonging to the same content flow or user with the same rules until such rules are modified or expired. Then the edge router does not have to send the CO headers of the related subsequent COs to the controller.

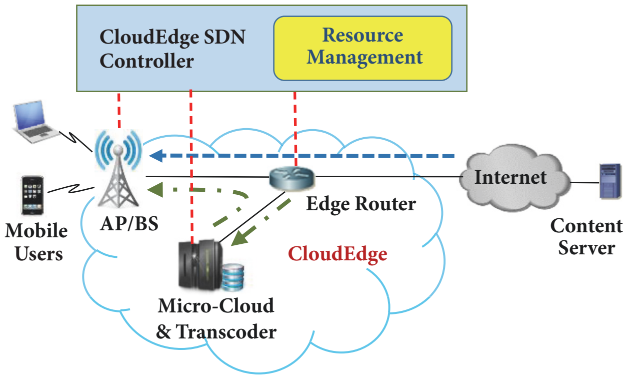


Figure 1: CloudEdge system overview.

3. Resource Management and Optimization

To address the new multiresource management challenges that arise when exploiting in-network transcoding service to optimize overall performance of video delivery in a wireless access network, we propose a framework in this section. The framework optimizes the allocation of wireless network resources and determines the transcoding rate for each stream to maximize the average QoE of multiple video streams, subject to the available wireless bandwidth and computation constraints of the CloudEdge transcoder.

3.1. Desired and Realized Quality of Experience.

Users may have heterogeneous mobile devices with different video processing and display capabilities. This makes the calculation of the QoE for a specific user under a given scenario very difficult. In addition, QoE has become a changing scale as technology advances, because what was considered the best video quality a few years ago is now only second rate. This led us to the concepts of desired QoE (DeQoE) and final realized QoE (ReQoE). DeQoE represents the QoE that a user desires or requests according to his device capability and context.

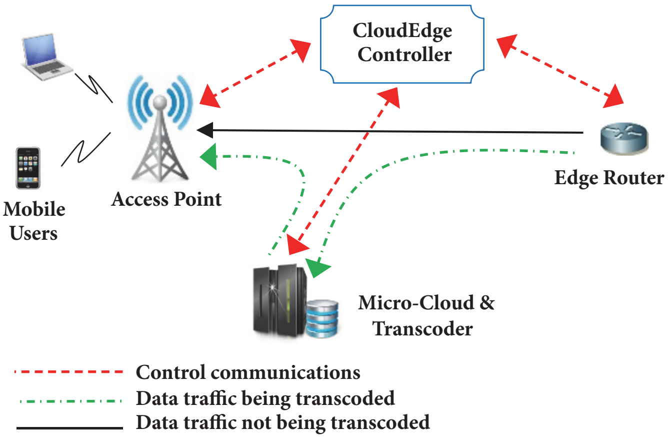


Figure 2: CloudEdge operation.

Table 1: ITU MOS mapping to user satisfaction rating.

User Rating	MOS
Very satisfied	4.3-4.5
Satisfied	4.0-4.3
Some users satisfied	3.6-4.0
Many users dissatisfied	3.1-3.6
Nearly all users dissatisfied	2.6-3.1
Not recommended	1.0-2.6

ReQoE is the realized QoE that the user obtains relative to his desired QoE. DeQoE and ReQoE can be used to make more accurate QoE measurements across a wide range of devices and scenarios, where $ReQoE \leq DeQoE$, and in an ideal case, $ReQoE = DeQoE$.

In order to quantify the QoE, Mean Opinion Scores (MOS) are often used. Table 1 shows the MOS values of streaming video for different degrees of user satisfaction, as defined by International Telecommunications Union (ITU) [15, 16]. It is assumed that if the resolution a user receives drops by one level from his requested quality, the MOS decreases a quality level as well. This conjecture results in Table 2. For example, if a user requests a streaming video at a resolution of 1080p and he receives 1080p video, the user has a ReQoE of 4.5 or Very Satisfied. If a user receives 720p video when his desired QoE is 1080p, that user has a ReQoE of 4.3 or Satisfied. Here the ReQoE is based solely on the change in resolution, if any, between the received and desired resolutions. In this case, we assume no packet loss. The impact of packet loss will be discussed later.

3.2. *CloudEdge Operations.*

Figure 2 shows the CloudEdge operations. First, N users request video streams; the controller then queries the AP/BS for the radio link data rate between each of the users and the AP/BS, while simultaneously querying the transcoder for its available computation resources. Next, the controller uses the queried information to calculate the settings to optimize the average ReQoE of all the users based on the algorithms described below. Then, the controller redirects the requested video streams that are to be transcoded to the transcoder, configures the transcoder with the video transcoding parameters of each stream, and configures the AP/BS channel utilization time for each user.

3.3. *Problem Formulation.*

We consider that N video streams are sent to the users. If there is not enough wireless bandwidth to transmit the N video streams at the quality that the users requested, then the video will be rerouted to the transcoder for rate adaptation. Due to the design of our system, there are two primary constraints. The first is the wireless channel bandwidth, which limits the maximum throughput of the AP/BS. The second constraint is the computation power available to the transcoder in the CloudEdge.

Adaptive modulation and forward error correction schemes are typically used in data transmission by the AP/BS [17]. The link transmission rate for a user depends on its Signal-to-Noise Ratio (SNR). For example, for 802.11g, the user data rate will be 6, 9, 12, 18, 24, 36, 48, or 54 Mbps, dependent upon the channel SNR value. Assume that video stream i is sent to user i and its data transmission rate is r_i. Let α_i $(0 \leq \alpha_i \leq 1)$ denote the channel utilization, i.e., the share of time that the AP/BS uses the wireless channel to transmit video stream i. Then the throughput of stream i is $T_i = \alpha_i r_i$. The data transmission should meet the wireless channel utilization constraint; that is,

$$\sum_{i=1}^{N} \alpha_i \leq 1. \tag{1}$$

Assume that c_i units of processing cycle are needed if stream i is transcoded. We define x_i as a variable to indicate whether stream i is transcoded $(x_i = 1)$ or not $(x_i = 0)$. Then the transcoding constraint is

$$\sum_{i=1}^{N} c_i x_i \leq C, \tag{2}$$

where C is the total available processing cycles.

As shown in Figure 3, the desired video rate, i.e., the user-requested rate for video stream i, is S_i^d, and the video rate

Table 2: ReQoE of video.

ReQoE Based on Received vs. Desired Video Resolution					
Using the MOS Upper Limit		Received			
		1080p	720p	480p	360p
Desired	1080p	4.5	4.3	4.0	3.6
	720p	-	4.5	4.3	4.0
	480p	-	-	4.5	4.3
	360p	-	-	-	4.5

Table 3: Standard bitrates for YouTube.com.

Type	Video Bitrate	Audio Bitrate	Resolution
1080p	8,000 kbps	384 kbps	1920 × 1080
720p	5,000 kbps	384 kbps	1280 × 720
480p	2,500 kbps	128 kbps	854 × 480
360p	1,000 kbps	128 kbps	640 × 360

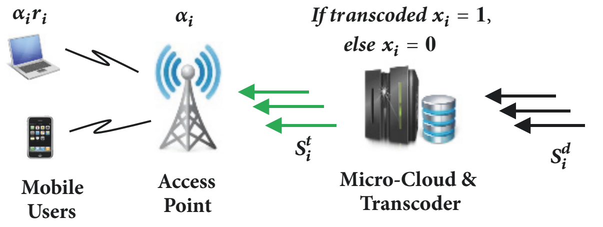


Figure 3: In-network video transcoding and wireless delivery.

after transcoding is S_i^t. The transcoder can reduce the video resolution, i.e., decreasing the video rate. Then we have

$$S_i^t < S_i^d \quad if \ x_i = 1 \ (transcoding), \text{ and} \tag{3}$$

$$S_i^t = S_i^d \quad if \ x_i = 0 \ (no \ transcoding). \tag{4}$$

Let δ denote the protocol overhead, which includes the video object header, the lower layer headers, and overhead to transmit video data; then, the required bandwidth to transmit the video stream is

$$T_i = \alpha_i r_i = \delta S_i^t. \tag{5}$$

The objective is to optimize the average ReQoE for all video streams transmitted by the AP/BS. The ReQoE is based on the received versus desired video resolution, and the received video resolution translates to required bandwidth. The objective can thus be formulated as

$$\max \quad \sum_i ReQoE_i \left(\alpha_i, S_i^t\right), \tag{6}$$

subject to wireless bandwidth constraint (1), transcoding constraint (2), as well as constraints (3), (4), and (5).

In general, ReQoE is a nonlinear function of S_i^t and α_i. This is then a mixed integer nonlinear programing (MINLP) problem, which is nondeterministic polynomial-time (NP) hard. In this paper, we consider a set of the standard video resolutions, with each video resolution having a targeted streaming data rate as recommended by YouTube [18], which is shown in Table 3. Thus, the problem becomes a nonlinear integer programing (NLIP) problem. We can then employ an exhaustive search algorithm to determine the optimal allocation of wireless channel utilization and transcoding rate for each video stream. We will also propose the simplified approximation algorithms in the next section. If there is enough wireless bandwidth to meet the requirements of all the video streams at the quality desired by the users, that is, constraints (1), (4), and (5) are satisfied, no transcoding is necessary. The incoming video data is then directly forwarded to the users via the AP/BS. Otherwise, transcoding is performed based on the determined rates. Note that one or more video streams may be dropped if the AP/BS is unable to support the data rate required to provide the minimum video resolution of 360p.

4. Approximation Algorithms

Due to the complexity of determining the above nonlinear optimum solution of ReQoE, we propose two simplified approximation algorithms in this section, and their performance will be compared to that of the optimum solution in the next section.

4.1. Throughput Maximization Algorithm. Instead of maximizing ReQoE, we determine the wireless channel utilization allocation and transcoding rate of each video stream to maximize the total AP/BS throughput. The throughput maximization algorithm is a linear programming (LP) problem, which can be solved in polynomial time [19]. The throughput maximization problem can be formulated as

$$\max \quad \sum_i \alpha_i r_i \tag{7}$$

$$\text{subject to:} \quad \sum_{i=1}^{N} \alpha_i \leq 1 \tag{8}$$

$$\sum_{i=1}^{N} c_i x_i \leq C \tag{9}$$

$$x_i = \begin{cases} 1 & S_i^t < S_i^d \\ 0 & S_i^t = S_i^d \end{cases} \tag{10}$$

$$\alpha_i r_i = \delta S_i^t \tag{11}$$

```
% $N$: total number of video streams
% $S_i^d$: Desired video rate, i.e., the user requested
%   rate for video stream i
% $S_i^t$: Data rate after optimization or transcoded
%    video rate (min means the data rate required to
%    provide the minimum video resolution of 360p; 0
%    means video stream i is dropped)
% $r_i$: Link data rate between AP/BS and user i
% $\delta$: Protocol and header overhead
% $\alpha_i$: Channel utilization of stream i
%$c_i$:  Processing cycles needed for transcoding
% $x_i$: transcoded ($x_i = 1$) or not ($x_i = 0$)
% $C$:  the total available processing cycles
for $i = 1 : N$        % Initialization
    $\alpha_i = \delta\ S_i^d / r_i;; S_i^t = S_i^d; x_i = 0$
end
If $\sum_{i=1}^{N} \alpha_i \le 1$ then
    break;                    % no transcoding
else
    sort the users' link data rate, i.e., $r_i \le r_j$, if $i < j$;
    $i = 1$;
    while (1)
        if $((\sum_{j=1}^{N} c_j x_j + c_i) < C \parallel x_i == 1)$ && $S_i^t \ne min$ then
                                        % transcoding cycles available
            reduce stream i resolution by one level
            stream i transcoded and $x_i = 1$
        else if $S_j^t == min$ for all transcoded streams &&
         $(S_{i-1}^t == 0 \parallel i == 1)$
            stream i dropped, $S_i^t = 0$ and $x_i = 0$
        end
        $\alpha_i = \delta\ S_i^t / r_i$;
        If $\sum_{i=1}^{N} \alpha_i \le 1$ then
            break;
        else $i++$;
        if $i > N$ then $i = 1$;
    end
end
```

Algorithm 1: Heuristic iterative ReQoE optimization.

Although the throughput maximization algorithm is simple, the maximization of AP/BS throughput does not mean the optimization of the average ReQoE, and the algorithm tends to allocate more channel times to the users with higher link data rate. In order to improve upon the basic throughput maximization algorithm, we design a heuristic iterative algorithm, presented in the next subsection, which attempts to fairly distribute available wireless bandwidth to the users and achieve a high average QoE.

4.2. Heuristic Iterative Algorithm. As shown in Algorithm 1, if there is enough wireless bandwidth to transmit all N video streams at the quality each user desires or requests, no transcoding is needed and all the video streams will be forwarded directly to the users. Otherwise, the heuristic iterative algorithm will lower the video resolution of a stream by one level if the transcoder has enough computation cycles to do so. The algorithm starts from the stream with the worst wireless link data rate because it requires the most channel time to transmit a certain amount of data. The algorithm checks whether there is enough wireless bandwidth to transmit all the video streams each time a video stream rate is reduced. If not, it will lower the stream with the next worse link data rate by one level. The process continues. Once the resolution of all the streams has been lowered by one level, if the wireless bandwidth is still insufficient to transmit the video streams, the algorithm begins the next round, starting from the stream with the lowest link data rate and then streams with the order of increasing link rate. If the resolution of all the transcoded streams is lowered to the minimum (360p in our scenario) but they still cannot be supported, some streams will be dropped, starting from the one with the worst link rate. This process continues until the resolution of the video streams that can be delivered with the available wireless bandwidth are discovered.

This algorithm attempts to fairly distribute the available wireless bandwidth to the users. It does not reduce any stream resolution by two levels until all other streams have been lowered by one level. The algorithm's primary goal is to ensure as many users as possible receive video, and its secondary goal is to provide the best video resolution possible, within the wireless bandwidth and transcoding computation constraints. This heuristic algorithm lowers the resolution of a stream by one level each iteration until it finds a feasible solution. Hence, its complexity is $O(N \times K)$ for N streams and K resolution levels.

5. Evaluation Results

In this section, we evaluate the performance of the proposed resource allocation algorithms and show the simulation results to unveil the impact of in-network processing on video delivery quality using edge cloud. We consider N video streams sent to the users through the proposed CloudEdge wireless access network. The radio access protocol is assumed to be IEEE 802.11g, and the link data rate for transmitting a video stream is determined using a practical link adaptation algorithm [20] based on the end user's channel SNR. We assume that users will request a desired video resolution based on their device capability and that the video server is capable of sending the desired resolution.

In order to establish a baseline for comparison of a CloudEdge network that can transcode users' video streams and a basic wireless network that cannot, we consider the effect of packet loss on the average QoE of a video stream. If there is no transcoding to adjust the video rate, the data that is above the AP wireless channel capacity would be dropped. In [16], the authors present a model on the quantitative relationship between QoE score and packet loss and give the parameters based on the curve fitting of actual datasets. Based on this model, we use the following function to calculate the impact of the packet loss on the QoE value of a video stream,

$$QoE = 3.010 * e^{-4.473 * Packet\ Loss} + 1.49 \qquad (12)$$

To compare the performance of different algorithms, including optimal ReQoE, throughput maximization algorithm, heuristic iterative algorithm, and the baseline with no transcoding, a number of network scenarios have been simulated. We present the results for some typical scenarios below.

5.1. Impact of In-Network Transcoding. Figures 4 and 5 show the average ReQoE versus the number of video streams with in-network transcoding using our heuristic iterative algorithm and without transcoding. Each figure presents two scenarios, one with all users desiring to stream video at a resolution of 1080p, and the other with all users desiring to stream video at a resolution of 720p, as labeled in the graph. In Figure 4, all users are connecting to the AP with the same link data rate, and, in Figure 5, users have different link data rates to the AP. We have assumed in this simulation that the transcoder is able to process as many video streams as needed to achieve the best average ReQoE. The impact of limited transcoding capacity will be shown later. In Figure 4, the simulation results show that when the number of video streams increases and there is not enough wireless bandwidth to transmit the video streams at their original quality/data rate, employing in-network transcoding to lower the video quality/data rate can greatly improve the average ReQoE. Otherwise, as video streams suffer from packet loss, video quality degrades quickly. This demonstrates one of the benefits of implementing CloudEdge with the ability to perform in-network transcoding and resource management in a wireless access network.

Figure 5 shows the average ReQoE versus the number of streams in a case where users are randomly distributed in a circle of 100 m with the AP at the center. The transmitting power of AP is fixed to a value such that the average link SNR is 0 dB. The users are connected to the AP with different link data rates determined by a link adaptation algorithm [21] based on their channel estimation. This represents a more realistic scenario. One interesting data point to note in the 720p transcoding case is that when user 6 connects to the AP, the average ReQoE drops sharply, but goes back up again when user 7 joins. We notice that this phenomenon occurs because user 6 has a very low link data rate, but user 7 has a very high link data rate; therefore when user 7 joins, it is able to receive a high ReQoE while requiring low channel utilization time, which helps to improve the average ReQoE.

Next, we investigate the impact of limited transcoder computation power on the average ReQoE. Figure 6 shows the average ReQoE versus the maximum number of video streams that the in-network transcoding service can process. The assumptions in this scenario are that there are 12 video streams being requested. The video resolutions being requested by all users are 1080p in the first scenario and 720p in the second. The simulation results show that, for the 720p scenario, eight streams must be transcoded in order for the average ReQoE to achieve the maximum potential. For the 1080p scenario, because each stream requests a higher data rate, the average ReQoE continues to improve until all 12 streams are transcoded.

5.2. Performance Comparison. Next, we compare the performance of the three algorithms (optimal ReQoE, AP throughput maximization, and heuristic iterative), along with the baseline case of no transcoding.

Figures 7 and 8 show that if all users connect to the AP with the same link data rate and desired video resolutions of 720p and 1080p, respectively, then the results of our heuristic iterative algorithm are very close to those of the theoretical optimal ReQoE. It outperforms the throughput maximization algorithm. All three algorithms achieve much better performance than the no transcoding case. For Throughput Max algorithm, when the number of video streams is over 10, none of the video streams can get the minimum video quality.

In Figures 9 and 10, the users connect to the AP with different link data rates and desired video quality of 720p and 1080p, respectively. As shown in the figures, the heuristic iterative algorithm performs better than the throughput maximization algorithm. However, the average ReQoE value of the heuristic iterative algorithm degrades quicker than that of the theoretical optimal ReQoE as the number of requested video streams increase. This is because the heuristic iterative

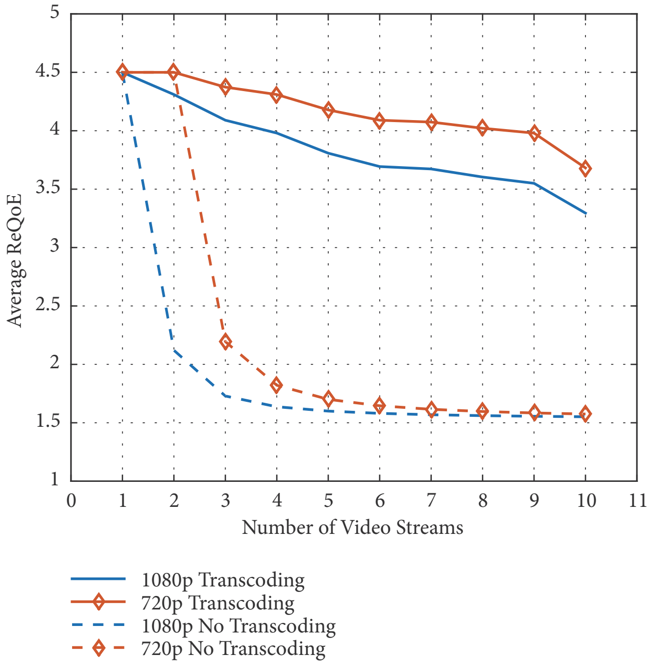


Figure 4: Average ReQoE versus the number of video streams, users connecting to the AP with the same link data rate.

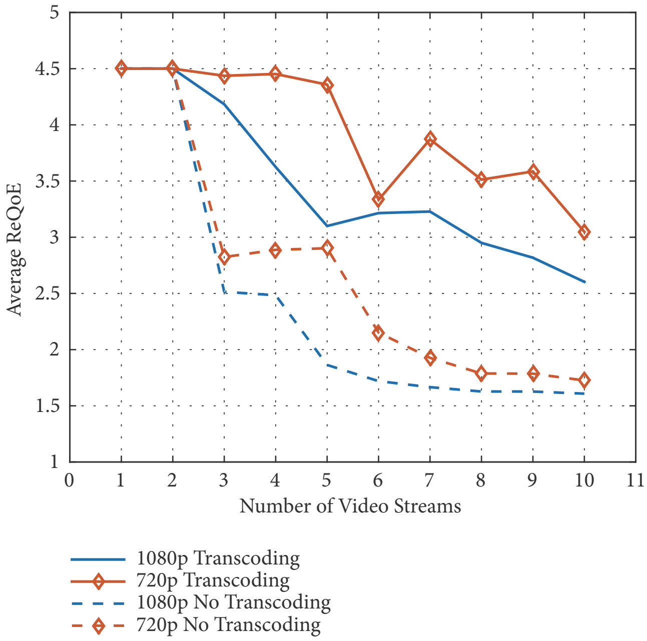


Figure 5: Average ReQoE versus the number of video streams, users connecting to the AP with different link data rates.

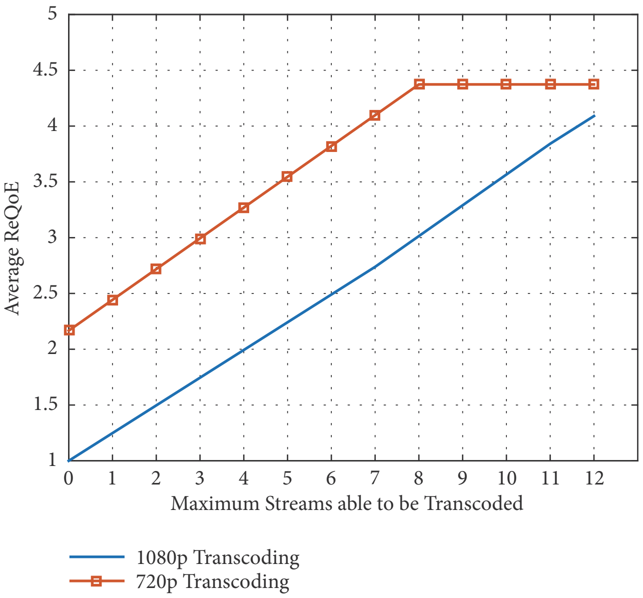


Figure 6: Average ReQoE versus the maximum number of streams that the transcoder can process.

algorithm tends to uniformly distribute the available wireless bandwidth among all the users instead of maximizing the average ReQoE of all streams.

The results we have shown here demonstrate the benefits of enabling the edge computation capability in wireless access networks. Although our heuristic iterative algorithm is not ideal, it is significantly better than no transcoding and, in most cases, is better than the throughput optimization scheme.

5.3. Fairness. In order to compare the fairness of the different algorithms proposed, we utilize Jain's Fairness Index [22], which is a mathematical tool for calculating how evenly different users or processes are treated within a particularscenario. It is calculated as follows:

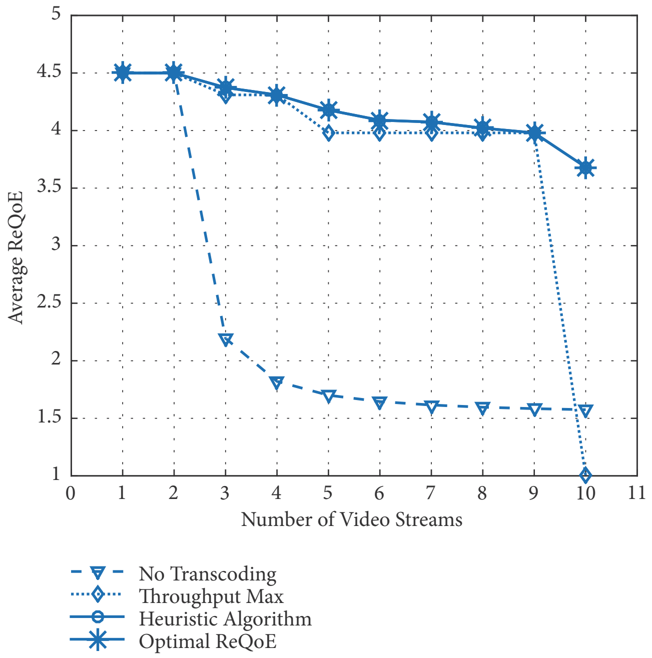


Figure 7: Average ReQoE versus the number of users requesting 720p video, all users connecting to the AP with the same link data rate.

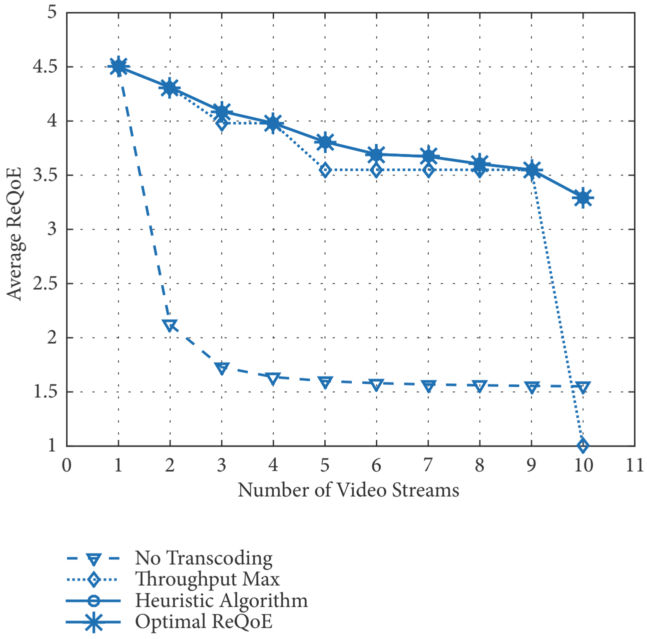


Figure 8: Average ReQoE versus the number of users requesting 1080p video, all users connecting to the AP with the same link data rate.

$$\mathcal{J}\left(x_1, x_1, \ldots x_N\right) = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N * \sum_{i=1}^{N} {x_i}^2} \qquad (13)$$

To calculate the fairness, we use the received data rate of each user, i.e., the throughput obtained by each user as an indicator.

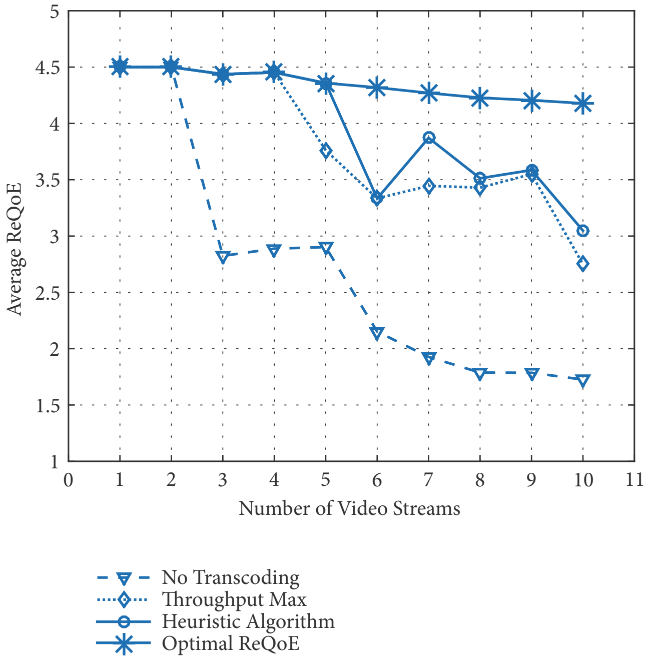


Figure 9: Average ReQoE versus the number of users requesting 720p video, users connecting to the AP with different data rates.

Figures 11 and 12 show Jain's Fairness Index based on the throughput of each user. Figure 11 shows the special case of all users connected to the AP with the same link data rate. For this case, the throughput maximization algorithm is the fairest, because, with the same link data rate for all the users, the throughput maximization algorithm simply divides the available bandwidth among the number of users

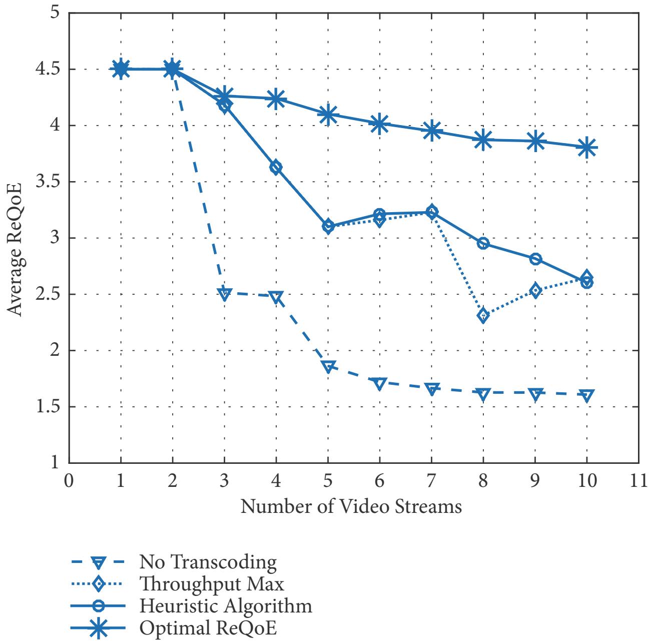


Figure 10: Average ReQoE versus the number of users requesting 1080p video, users connecting to the AP with different data rates.

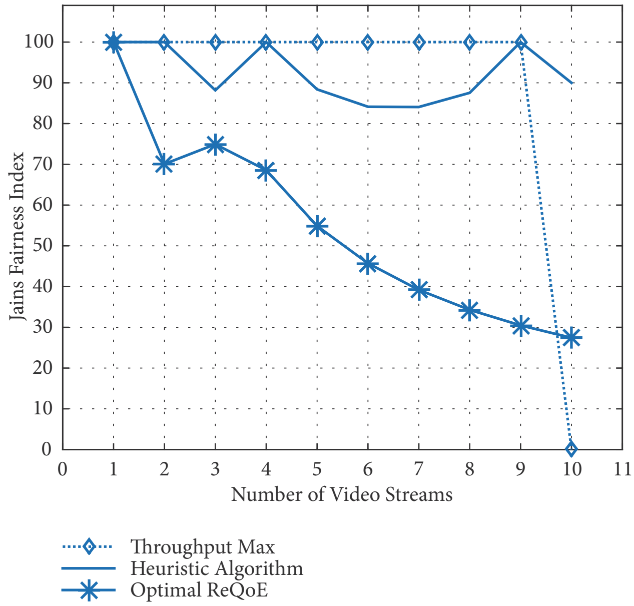


Figure 11: Jain's Fairness Index of a 720p scenario, users connecting to the AP with the same link data rate.

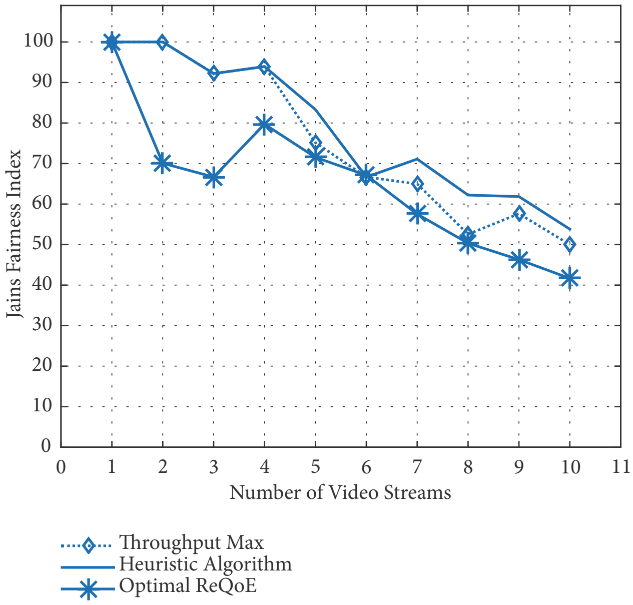


Figure 12: Jain's Fairness Index of a 720p scenario, users connecting to the AP with different link data rates.

requesting video data and conducts transcoding as needed. This method fails to successfully transmit any video streams for 10 or more users, because in this case the available bandwidth divided by 10 is too low to support even the lowest video quality. In Figure 12, the users have different link data rates. It shows that at first the throughput maximization and heuristic iterative algorithms have the same level of fairness until the fourth stream begins. The optimal ReQoE algorithm routinely supports one or more users at the highest data rate while simultaneously dropping one or more users with poor link data rates, making it the most unfair.

6. Related Work

Cloud computing has been widely studied [23–25]. However, the traditional cloud computing model depends on centralized data centers which have difficulty adapting to meet the requirements for fast response to the changes in wireless access networks. Deploying in-network processing capabilities at the edge of the network provides real-time access to wireless network information and enables better context-aware processing with minimum latency. CloudEdge is not a simple extension to traditional cloud technologies due to its distinctive characteristics. In contrast to dedicated data centers, the computing capability in wireless access networks is limited, making resource management much more critical.

Mobile edge computing, also called multiaccess edge computing, or fog computing is an emerging research area, which aims to provide cloud computing capabilities at the network's edge. Its application scenarios and challenges have been discussed in several recent publications [26–28]. A cloudlet architecture [29, 30] was proposed to host in-network services and to help Internet Service Providers (ISPs) and application providers deploy their services at strategic network locations. The design is based on a name-oriented clean-slate Internet architecture [14], which assigns a globally unique identifier (GUID) to each of the network objects. A directory resolution mechanism is employed to resolve a GUID to its network address in order to forward the request to the destination. The GUID resolution mechanism is different from an SDN controller. It only provides a directory service to map GUIDs to network addresses, but the SDN controller can inject the logic into network elements and program the data plane. In [11], an architecture was proposed to consolidate middleboxes in enterprise networks and enable the middlebox applications to run on a consolidated hardware platform. However, this work focused on enterprise network appliance functions such as intrusion detection systems, firewalls, and VPNs with limited extensibility. It also required a lot of manual provisioning with little programmability, which leads to inefficiency in management and use of the infrastructure. In addition, it did not consider the impact of varying wireless channel conditions to the middleboxes, nor did it allow third-party content providers to use the resources. The industry standardization effort on mobile edge computing is underway in the European Telecommunications Standards Institute (ETSI) [31] with a focus on managed mobile cellular networks to unify the telco and IT-cloud worlds, providing IT and cloud computing capabilities within radio access networks. The resource allocation in mobile edge computing has been studied, where the works focus on how to offload the computing tasks from mobile devices to the edge cloud [32, 33] and distribute the computing tasks among the edge servers [34–36]. Our work, however, focus on allocating wireless bandwidth and computing resources to serve multiple video streams.

Future Internet architectures, such as information-centric networks (ICNs), have been proposed recently [12–14, 37] and provide a network layer solution by integrating the caching function in the content routers with name-based routing to deliver content from the best source/cache location. However, these solutions do not consider in-network data processing or modifying content representation, e.g., changing video compression rates to adapt to varying network conditions.

Our work differs from the aforementioned works in several aspects. (1) We integrate SDN with in-network data processing and quantify its benefits for wireless content delivery. In our architecture, SDN is extended to manage not only data forwarding but also content processing, which enables us to perform unified resource management and optimization based on the varying wireless network conditions and QoS requirements of multiple users. (2) Based on the CloudEdge architecture, the resource management and optimization for video content delivery is addressed. In particular, we formulate a framework to optimize the QoE of multiple video streams and propose and compare three resource allocation algorithms, considering both bandwidth and computation constraints.

The preliminary version of our work was presented at a conference [38]. This paper substantially extends it. We propose and study two additional resource allocation algorithms, which were not previously presented in [38], nor in any other existing works. In this paper, we also included new results to validate and compare our designs. In addition, we performed comparison of the fairness of the three algorithms proposed. Furthermore, we largely restructured the manuscript with better presentation and detailed discussions.

7. Conclusions

In this paper, we propose CloudEdge, a computation-capable and programmable wireless access network architecture that exploits in-network computation capabilities to provide unified data processing and transport services to end users. To address the multiresource management challenges, such as wireless bandwidth and computation power that arise in integrating in-network services for video content delivery, we formulate a framework that optimizes the quality of experience of multiple video streams, subject to wireless transmission capacity and in-network computation capability constraints. We also proposed two simplified algorithms for resource allocation. The evaluation results clearly demonstrate the benefits of the proposed system and resource allocation algorithms.

For future work, we are building a prototype CloudEdge network as a proof of concept to validate the algorithms developed here and gain deeper insights into the benefits of integrating in-network data processing and delivery in an SDN-based wireless access network, in order to facilitate further development of our approach and methodology.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by a NSF grant CNS-1624485 and InterDigital Communications, Inc. as an Industry/University Cooperative Research Center sponsor.

References

[1] Cisco, "Cisco Visual Network Index: Global Mobile Traffic Forecast Update 2016–2021," 2017.

[2] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.

[3] A. Ahmed and E. Ahmed, "A survey on mobile edge computing," in *Proceedings of the 10th International Conference on Intelligent Systems and Control (ISCO '16)*, pp. 1–8, January 2016.

[4] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. de Foy, and Y. Zhang, "Mobile edge cloud system: architectures, challenges, and approaches," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2495–2508, 2018.

[5] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, "Mobile-edge computing introductory technical white paper," Mobile-Edge Computing (MEC) Industry Initiative, 2014.

[6] Y. Jararweh, A. Doulat, O. AlQudah, E. Ahmed, M. Al-Ayyoub, and E. Benkhelifa, "The future of mobile cloud computing: Integrating cloudlets and Mobile Edge Computing," in *Proceedings of the 23rd International Conference on Telecommunications (ICT '16)*, pp. 1–5, Thessaloniki, Greece, May 2016.

[7] N. McKeown, T. Anderson, H. Balakrishnan et al., "OpenFlow: enabling innovation in campus networks," *Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[8] M. Banikazemi, D. Olshefski, A. Shaikh, J. Tracey, and G. Wang, "Meridian: an SDN platform for cloud network services," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 120–127, 2013.

[9] S. Jain, A. Kumar, S. Mandal et al., "B4: Experience with a globally-deployed software defined WAN," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '13)*, pp. 3–14, August 2013.

[10] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "Softran: software defined radio access network," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking (HotSDN '13)*, pp. 25–30, August 2013.

[11] V. Sekar, N. Egi, S. Ratnasamy, M. K. Reiter, and G. Shi, "Design and Implementation of a Consolidated Middlebox Architecture," in *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, pp. 1–24, San Jose, CA, USA, April 2012.

[12] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '09)*, pp. 1–12, New York, NY, USA, December 2009.

[13] H. Liu, X. De Foy, and D. Zhang, "A multi-level DHT routing framework with aggregation," in *Proceedings of the 2nd ACM SIGCOMM Information-Centric Networking Workshop (ICN '12)*, pp. 43–48, August 2012.

[14] D. Raychaudhuri, K. Nagaraja, and A. Venkataramani, "Mobilityfirst: a robust and trustworthy mobility-centric architecture for the future internet," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, pp. 2–13, 2012.

[15] ITU-T G.107, "The E-model: a computational model for use in transmission planning," 2011.

[16] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[17] IEEE 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 2012.

[18] Youtube.com, "Recommended upload encoding settings: Recommended bitrates, codecs, and resolutions, and more," https://support.google.com/youtube/answer/1722171.

[19] J. Noyes and E. W. Weisstein, "'Linear Programming," MathWorld—A Wolfram Web Resource, http://mathworld.wolfram.com/LinearProgramming.html.

[20] G. Pei and T. R. Henderson, "Validation of OFDM error rate model in ns-3," *Boeing Research Technology Report*, 2010.

[21] J. P. Pavon and S. Choi, "Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement," in *Proceedings of the IEEE International Conference on Communications (ICC '03)*, vol. 2, pp. 1108–1113, Anchorage, AL, USA, May 2003.

[22] R. Jain, W. Hawe, and D. Chiu, "A Quantitative measure of fairness and discrimination for resource allocation in Shared Computer Systems," *DEC-TR-301*, 1984.

[23] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Proceedings of the Grid Computing Environments Workshop (GCE '08)*, pp. 1–10, Austin, TX, USA, November 2008.

[24] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, and H. Bhogan, "Volley: Automated data placement for geo-distributed cloud services," in *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation (NSDI '10)*, 2010.

[25] A. Greenberg, J. R. Hamilton, N. Jain et al., "VL2: A Scalable and Flexible Data Center Network," in *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, pp. 51–62, Barcelona, Spain, August 2009.

[26] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the 1st ACM Mobile Cloud Computing Workshop (MCC '12)*, pp. 13–15, August 2012.

[27] B. Chandramouli, J. Claessens, S. Nath, I. Santos, and W. Zhou, "RACE: Real-time applications over cloud-edge," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '12)*, pp. 625–628, May 2012.

[28] B. P. Rimal, D. P. Van, and M. Maier, "Mobile Edge Computing Empowered Fiber-Wireless Access Networks in the 5G Era," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 192–200, 2017.

[29] Y. Chen, B. Liu, Y. Chen, A. Li, X. Yang, and J. Bi, "Packetcloud: an open platform for elastic in-network services," in *Proceedings of the 8th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch '13)*, Miami, FL, USA, October 2013.

[30] F. Bronzino, C. Han, Y. Chen et al., "In-network compute extensions for rate-adaptive content delivery in mobile networks," in *Proceedings of the IEEE 22nd International Conference on Network Protocols (ICNP '14)*, pp. 511–517, Raleigh, NC, USA, October 2014.

[31] "ETSI Industry Specification Group (ISG) for Multi-access Edge Computing (MEC)," http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing.

[32] E. Cuervoy, A. Balasubramanian, D.-K. Cho et al., "MAUI: making smartphones last longer with code offload," in *Proceedings of the 8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '10)*, pp. 49–62, New York, NY, USA, June 2010.

[33] T. Truong-Huu, C.-K. Tham, and D. Niyato, "To offload or to wait: An opportunistic offloading algorithm for parallel tasks

in a mobile cloud," in *Proceedings of the 6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom '14)*, pp. 182–189, December 2014.

[34] C. T. Do, N. H. Tran, C. Pham, M. G. Alam, J. H. Son, and C. S. Hong, "A proximal algorithm for joint resource allocation and minimizing carbon footprint in geo-distributed fog computing," in *Proceedings of the International Conference on Information Networking (ICOIN '15)*, pp. 324–329, Siem Reap, Cambodia, January 2015.

[35] M. Aazam and E.-N. Huh, "Dynamic resource provisioning through Fog micro datacenter," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops '15)*, pp. 105–110, St. Louis, MO, USA, March 2015.

[36] Y. Xiao and M. Krunz, "QoE and power efficiency tradeoff for fog computing networks with fog node cooperation," in *Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM '17)*, pp. 1–9, Atlanta, GA, USA, May 2017.

[37] https://www.point-h2020.eu/ip-over-icn-over-sdn/.

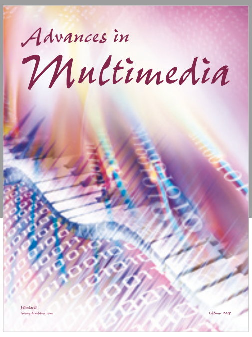
[38] H. Liu and K. Smith, "Improving the expected quality of experience in cloud-enabled wireless access networks," in *Proceedings of the IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems (MASS '15)*, pp. 519–524, Dallas, TX, USA, October 2015.

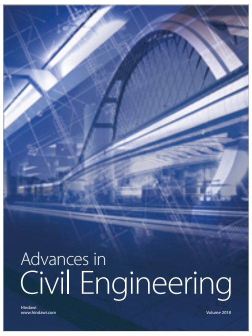
Journal of
Engineering

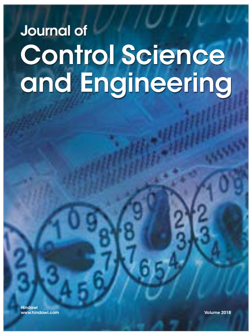
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

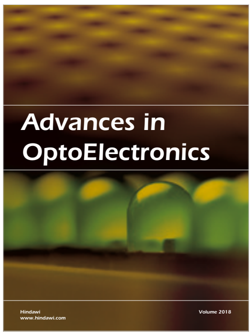

Advances in
Multimedia


Advances in
Civil Engineering


Journal of
Control Science and Engineering

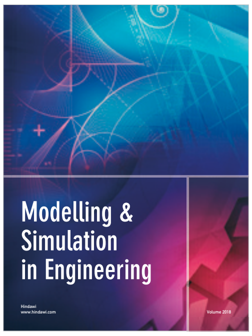
Journal of
Robotics

Journal of
Electrical and Computer Engineering

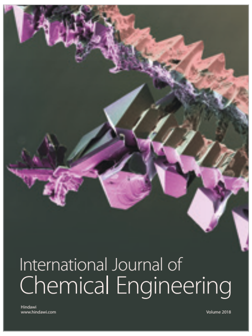

Advances in
OptoElectronics

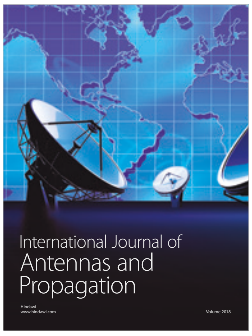
VLSI Design

International Journal of
Navigation and Observation

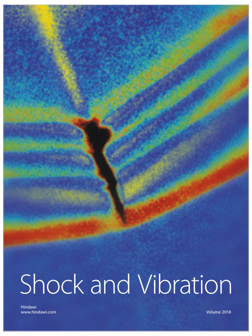

Modelling & Simulation in Engineering

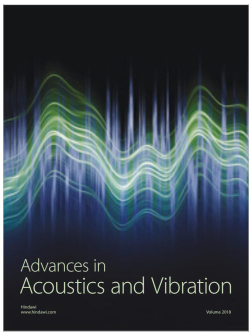
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration