

Quality Adaptation and Resource Allocation for Scalable Video in D2D Communication Networks

SAEED ULLAH¹, KITAE KIM¹, AUNAS MANZOOR¹, LATIF U. KHAN¹,
S. M. AHSAN KAZMI^{1,2}, AND CHOONG SEON HONG¹, (Senior Member, IEEE)

¹Department of Computer Science and Engineering, Kyung Hee University, Yongin-Si 17104, South Korea

²Networks and Blockchain Lab, Institute of Information Security and Cyber Physical System, Innopolis University, 420500 Innopolis, Russia

Corresponding author: Choong Seon Hong (cshong@khu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) under Grant NRF-2017R1A2A2A05000995, and in part by the MSIT (Ministry of Science and ICT), South Korea, under the Grand Information Technology Research Center support program, supervised by the IITP (Institute for Information and Communications Technology Promotion, under Grant IITP-2019-2015-0-00742.

ABSTRACT Recently fifth-generation (5G) of cellular networks appeared to enable various smart bandwidth thirsty applications. Frequency re-usability via device-to-device (D2D) communication is one of the possible ways to enhance the 5G network throughput. However, despite these technological advancements in cellular networks, users' Quality of Experience (QoE) is significantly affected by video quality adaptation according to future network conditions for video streaming. In this paper, we jointly formulate D2D pair association, video quality adaptation, and network resource allocation for scalable video streaming in a D2D communication network. The formulated problem is a mixed-integer linear programming problem and has a combinatorial nature. Thus, the formulated problem is decomposed into two subproblems: D2D pair association subproblem and resource allocation subproblem. The D2D pair association subproblem is a combinatorial optimization problem, to solve it a distributed algorithm is proposed based on two sided-matching theory. On the other hand, resource allocation subproblem is solved using conventional linear optimization techniques to get the boundary solution. Furthermore, a fully distributed quality adaptation scheme is presented using the proposed resource allocation algorithm for layered video. To validate the performance of our proposal, we perform numerical simulations for various scenarios.

INDEX TERMS 5G, video quality adaptation, video streaming, layered video, H.264/AVC, SVC, D2D communication, resource allocation, D2D pair association.

I. INTRODUCTION

In the current decade, users' demand for bandwidth-thirsty applications, such as mobile video streaming has increased many folds and this growth is predicted to continue. The Cisco visual index reports that mobile traffic will grow 3-fold from 2018 to 2021 (from 17EB per-month to 49EB), 78% of which will involve video traffic, while the mobile traffic will grow twice as fast as the fixed IP traffic [1]. The fifth-generation (5G) of cellular networks has appeared on the horizon of cellular communication to accommodate this increasing demand. Since fourth-generation technology would not support further enhancement in the frequency domain. Therefore, 5G exploits the frequency re-usability in the form of massive MIMO techniques and dense heterogeneous deployments. Moreover, device-to-device (D2D)

communication is an integral part of the full-fledged 5G cellular system, and it will assist in enhancing the 5G motive of frequency re-usability through localized short-range communication. Meanwhile, end-user mobile devices such as smart-phones and tablets, equipped with several 10s to 100s GBs of memory, high processing power, enhanced communication, and interference rejection capabilities further encourage D2D communication.

Moreover, future Internet architectures like Content-Centric/Named-Data Networking (CCN/NDN) have made the content as network gravity rather than the end-to-end communication in the current IP-based Internet [2]. In CCN/NDN, any networking node that is equipped with local storage (cache/content-store), can store the content and provide it to other users on their requests. Mobile devices such as smart-phones/tablets can use a part of the storage memory (which is usually underutilized) as a content store to cache the contents [3]. This characteristic of the CCN nodes

The associate editor coordinating the review of this manuscript and approving it for publication was Omer Chughtai.

highly facilitates the realization of D2D communication. However, despite all these technological advancements, for video streaming, the users' Quality of Experience (QoE) is mainly affected by the video quality adaptation according to the predicted network condition in the next time frames. If the UE selects video quality higher than the available data-rate, it may face the video stalling and if the selected video quality is too much lower than the available data-rate, it will get the uninterrupted video but the quality will be lower even the higher quality may be affordable. Moreover, pairing D2D users is having a great effect on user QoE for video streaming. In a situation, when the requested video is available in the cache of multiple neighboring users, they may have different quality of the requested video and may have different Signal to Interference and Noise Ratio (SINR). Selecting the optimal one is important for increasing the QoE. Furthermore, the optimal allocation of network resources is also important for improving network throughput and enhancing users QoE. In this paper, we are jointly dealing with all these three important parameters: (a) Video quality adaptation (b) D2D pair association and (c) network resource allocation for a layered video.

A. BACKGROUND AND MOTIVATION

In this section, we present a brief overview of NDN/CCN architecture and SVC encoded layered video to give a better insight of our work.

1) CONTENT-CENTRIC/NAMED-DATA NETWORKING (CCN/NDN)

The current IP-based Internet (mainly consisted of TCP, UDP, etc.) is location-specific. As we have entered the age of information, the future Internet is envisioned to be based on information/content. CCN, also called Named-Data Networking (NDN) [2], a variant of the future Internet architecture, changes the architecture of the current Internet by making it content-specific. In NDN, the data is named as variable-length hierarchical identifiers, like URIs or file system paths (e.g., a/b/c.mpg) [4]. UE generate requests, called Interest, for a chunk of the content that carries these hierarchical names and send it access router. The access router and intermediate routers/networking-nodes direct the Interests towards the content provider with the help of these hierarchical names following the longest prefix matching. On reception of Interest, The content provider produces a packet of the requested chunk of data and sends it to the requesting UE on the reverse path. Networking nodes, i.e., routers, cache the data chunk in its local storage to satisfy similar Interests in the future.

CCN/NDN comprises three primary data structures named Pending Interest Table (PIT), Content Store (CS), and Forward Information Base (FIB). Upon reception of the Interest, the networking node checks PIT to determine any other similar Interest (by other user(s)) forwarded and the corresponding data not yet received for. If such any entry is there, the face/port on which the data is received is added to the PIT entry, and the corresponding data is replicated on all the requesting faces listed in PIT. If there is no entry in

PIT for the Interest, next checks CS for a copy of the data, If found, Interest is discarded, and the data is transmitted to the requesting UE. Otherwise, the Interest is forwarded towards the potential content provider following FIB and making an entry in the PIT. Data packet forwarding is simple. Received data is replicated on all the faces listed in PIT. When a node receives a data packet and does not retrieve any entry for it in PIT, such a data packet is considered outdated and is discarded. For further details, we will direct interesting readers to [2].

2) LAYERED VIDEO STREAMING

Widely used advanced video coding (H.264/AVC) [5] encodes video in different representations of bitrates/qualities and maintains a separate file/stream for each instance. The requesting UE is provided a video from the appropriate file/stream according to the network condition and its device capabilities. A separate stream is maintained for each requester, even if they are watching the same video with different quality representations. On the other hand, Scalable extension of (H.264) also called as Scalable Video Coding (SVC) [6], provides scalability by encoding the video in multiple layers and maintaining a single file for the video [7]. SVC video consists of a mandatory base-layer that defines the outlines of the video and multiple enhancement-layers. Lowest temporal, spatial, and quality parameters are used for encoding base-layer. Only the base-layer can be decoded in basic quality video independent from other layers. Spatial, temporal and overall quality of the video is enhanced by decoding base-layer along with enhancement layers. Video quality is improved when additional enhancement layers are added to it. There is a very strong dependency among the layers of the SVC video. A layer can be decoded only when all the lower layers are present. For instance, enhancement-layer 1 can only be useful if the base-layer is present. Similarly, enhancement-layer 2 can only be useful when the base-layer and enhancement-layer 1 are already available, and so on.

3) MOTIVATION

SVC video can be extremely useful in future Internet architectures such as NDN/CCN because of the caching capability of the networking nodes [8]. For H.264/AVC encoded video, a networking node cannot satisfy an Interest if the Interest is for any other representation of the video than the one present in cache. On the other hand, in the case of SVC, when a networking node receives Interest for a video-segment, all the layers present in cache for that video-segment can be provided to the requester regardless of the final quality the requester receives. If the requested quality is lower than the cached one, a subset of the layers are provided, and if the request is for higher quality than the cached on, all the cached layers are provided and for the remaining layers, the interest is forwarded towards the content provider.

In the case of H.264/AVC encoded video, the quality adaptation mostly occurs at the GOP level, and the quality, once decided, must be downloaded as a whole. In the case of SVC, layers of the video are decoupled and for a segment,

layers can be independently downloaded. A user can improve the video quality by downloading additional enhancement layers for a segment until it has been played-out. The decoupling property of the layers of the SVC video can be extremely beneficial in the wireless network environment wherein network condition changes drastically and rapidly change. A cellular user is having the independence of downloading SVC video segments layer by layer if it successfully acquires network resources and can enhance the quality by sequentially downloading additional layers as long as the segment playtime has not been approached. Unfortunately, this property of the SVC video has not been exploited in cellular networks by the research community. [9] has jointly considered resource allocation and quality adaptation for stream loading in the cellular network environment. However, they are unable to completely benefit from the layer-decoupling property of the SVC video by taking the assumption that the user decides how many layers for each segment are to be downloaded in advance according to the predicted throughput in the next time slot, and this decision cannot be revisited regardless of whether the transmission is successful or non-successful. This assumption essentially renders their modified stream loading the same as H.264/AVC encoded video.

In a D2D communication network, on one hand, pairing the video requesting UE with such a D2D video provider UE that possesses the highest quality of the requested video will provide the opportunity to the requesting UE to get higher quality video via short range high-throughput D2D link. On the other hand, making pairs of the UE that can achieve the best SINR is also important for increasing the system throughput. Moreover, a situation can occur in which a provider UE possess in its cache the requested video(s) of more than one requesting UE at a time, however, it can be paired with only one at a time, thus one of them has to be dropped which should be paired with the second-best providing UE. Therefore, making the optimum pairing will greatly affect user QoE. Furthermore, the Resource Block (RB) allocation to the D2D pairs is also having great importance. Due to the interference and distance between the requesting UE and providing UE, different pairs may achieve different throughput on the same RB. Optimal assignment of the RBs can affect the system throughput and hence the user QoE. In this paper, we propose a mechanism for the SVC encoded layered video that jointly tackles all these three important factors i.e., the video quality adaptation that takes full advantage of the layer-decouple property of SVC video, the D2D pair association, and the RBs allocation to provide uninterrupted higher quality of videos to the requesting UE.

B. CONTRIBUTIONS

The contributions of this paper are as follows:

- We propose quality adaptation for the scalable/layered video that exploits the decoupling property of the layered video. The quality adaptation is subjected to the resource allocation. User sends a request for a

layer of the video segment after successfully acquiring network resources and downloading the previously requested layer of the video segment. The video quality is improved by downloading enhancement layers of the video until it is played out. Moreover, to avoid the video stalling in the fluctuating wireless environment, we define a sliding bucket inside which the layers of the upcoming segments are downloaded in a manner that the base layers are given the top priority for downloading and the enhancement layers are downloaded according to the presented algorithm.

- We formulate joint optimization problem for the D2D pair association and resource allocation with the objective of requested video quality enhancement. The joint formulated optimization problem is a mixed-integer problem solution for which is a computationally challenging task as it requires exponential computation for obtaining the optimal solution.
- To find a feasible solution for the joint optimization problem, we decompose into two subproblems i.e., D2D pair association and resource allocation subproblems. The D2D pair association subproblem is a combinatorial optimization problem. We apply matching theory to solve it with the preferences of video quality and throughput enhancements. In the solution, we present an algorithm based on two sided-matching theory that is guaranteed to give a stable solution.
- The resource allocation optimization problem is solved using the conventional linear optimization technique to get the solution at the boundary.
- For performance analysis, we have shown numerical results for different setups and scenarios. The simulation results show that the proposed mechanism provides better quality video to the users in the available resources.

II. RELATED WORK

This section provides an overview of research that has been conducted for resource allocation, user association, and quality adaptation for scalable video streaming in wireless networks.

Several papers [9]–[19] considered quality adaptation and resource allocation for video streaming in wireless/cellular networks. A comprehensive study is presented in [10] for quality adaptation of video streaming over HTTP. They presented the cutting edge research for video quality adaptation on server-side, client-side, and in-network. On the other hand, the network throughput shows significant variations in the wireless communication environment. To cope with such variations, many researches presented quality adaption proposals for the next video segment based on download buffer as well as the throughput prediction for the subsequent time frame [9], [11], [12], [16], [18]. The authors of [20] considered video quality enhancement, and energy efficient caching in cellular networks. The video quality enhancement part is constrained by the monetary cost to associate users with higher bitrates and ignores the limitations of radio resources.

The authors in [14] and [15] proposed joint optimization of resource allocation and video quality improvement for scalable video streaming in wireless access networks. In [17], the authors presented rate adaptation based on Markov decision to enhance video quality. They considered a number of key factors, such as amplitude and switching frequency of video rate, video playback quality, buffer occupancy, and buffer overflow/underflow. In [9] the authors extended the resource allocation scheme presented in [18] for a variant of scalable video streaming called streamloading. They have formulated an offline optimization problem for jointly rate allocation and quality adaptation for streamloading in a wireless network. Then, they presented a distributed online algorithm for layer selection for the next segment of the video. The resource allocation optimization problem is solved by the base-station following the proposal presented in [18] with the addition of prioritizing resource allocation to base layer over the enhancement layers. Recently, in [19], the authors presented to shift video quality adaptation decision from the client side to the network edge. They estimate the throughput in mobile edge and on the bases of that presented a heuristic quality adaptation decision for enhancing QoE.

On the other hand, resource allocation and pair association in D2D communication networks have been considered in many papers [21]–[28]. A comprehensive study has been presented in [29] for users association in 5G cellular networks. In [21], the authors presented joint optimization of user association and resource allocation in the cognitive femtocell network. In [22], the authors proposed a novel context-aware resource allocation optimization approach in D2D communication networks. They have solved the problem via matching game considering UE and RBs are the players. In [23], the authors have presented drawbacks of imperfect D2D pair association in the cellular network. In [24], the authors presented mode selection and resource allocation in underlay D2D networks. Similarly, in [26], they have jointly solved the problem of D2D user association and content caching in dense small cell networks. The authors of [27] presented an optimization framework to reduce energy consumption and video quality adaptation in heterogeneous cellular networks. The proposed user association and network resource allocation are mainly constrained by power consumption. In [28], the authors proposed wireless D2D caching networks and contrasting the D2D communication with other conventional approaches for the video contents. For the D2D communication, they considered a traditional microwave band of 2 GHz and millimeter-wave. The video contents that are not found in the cache of a device or providing the contents are not possible via the D2D link are provided through direct transmission from the base-station. In [3], the authors presented D2D communications for wireless videos. They have proactively cache popular videos in UE considering users' demand and storage capacity of UE. They analyze the optimal collaboration distance and the trade-off between frequency reuse and the probability of finding the requested content in collaboration distance. On the other hand, caching and

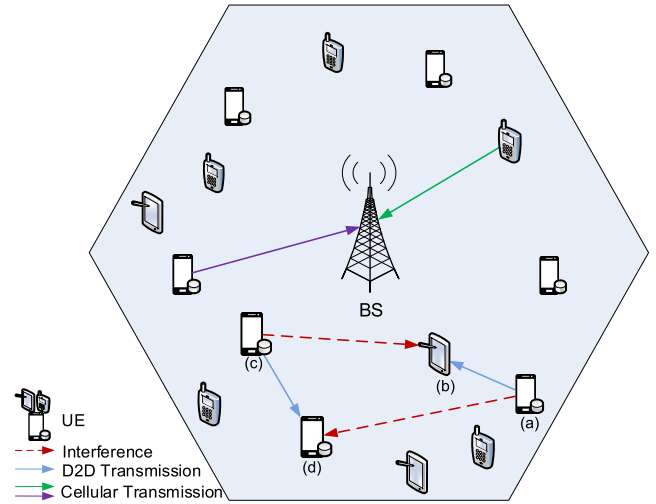


FIGURE 1. System diagram.

communication video content in CCN/NDN has also received the great attention of the research community in recent years [30]–[32]. The main difference of the aforementioned works compared to our work is that we are jointly considering the quality adaptation, resource allocation and pair association in D2D networks for scalable video streaming which is not been done to date to the best of our knowledge.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

For our proposed mechanism, we are considering a cellular communication network scenario as shown in Figure 1. We have a total of N UE in the system. Let $\mathcal{N} = \{1, 2, 3, \dots, N\}$ be the layered video requesting UE and $\mathcal{M} = \{1, 2, 3, \dots, M\}$ are the UE having the requested contents in their cache. The Requester UE n and provider UE m can only communicate via the D2D link if they are in each other's transmission range. There are total V videos with the Content Provider (CP). Each video is divided into S segments, where $S = \{1^L, 2^L, 3^L, \dots, S^L\}$. Each segment is comprised $l = \{0, 1, 2, \dots, L\}$ layers, wherein 0 represents the base layer and $1, 2, 3, \dots, L$ represent the enhancement layers. Let $f(v_{n,s}^{l*})$ represent the amount of data for l_{th} layer (only one layer) of the segment s requested by UE n . Combinations of layers produce different quality levels for the video/segment. We denote the $f(v_{n,s}^l)$ as the Quality-Rate tradeoff realization measured in bits per second, which is a convex function of the video quality for the segment data rate [33].

Let's q_{nm}^l represent the quality of the video with l layers that requesting UE n can get from UE m . q_{nm}^l can be represented as follows:

$$q_{nm}^l \in \{f(v_s^0), f(v_s^1), \dots, f(v_s^L)\}, \quad \forall m \quad (1)$$

where $f(v_{n,s}^0)$ represents the amount of video data (in bits) for segment s of video v requested by user n with the base layer only and $f(v_{n,s}^l)$ represents the amount of data with l enhancement layers and the base layer for segment s of video v requested by user n . In other words, we can say that $f(v_{n,s}^l)$ represents amount of data of video segment s

TABLE 1. Summary of Notations.

Notation	Description
N	Total UE in the Network
$\mathcal{N}$	Set of requesting UE
$\mathcal{M}$	Set of content providing UE
S	total segments in a video
q_{mn}^l	quality of the video with m for n
$f_{n,s}(\cdot)$	quality-rate trade-off for user n segment s
L	Maximum number of layers in a video
R	total resource blocks
ρ	number of segments in the sliding bucket
β_{min}	min. difference between two adjustment segments
β_{max}	max. difference between two adjustment segments
$a_{nm}; x_{nm}$	Binary variables
h	channel gain
Γ_{mn}^r	SINR of D2D UE
γ_{mn}^r	Transmission rate of a D2D UE
$\mathcal{P}_m$	Preference list of content providing UE
$\mathcal{P}_n$	Preference list of content requesting UE

for all the layers up to layer l including the base layer. UE periodically broadcast content information tables to their one hop neighbors. Therefore, whenever a UE needs a segment of the content, it knows the location(s) where it can obtain the desired content, along with the available number of layers. If the requested content is cached with multiple UEs, the suitable providing UE is selected according to the matching game results that we discuss in Section IV-C.

The D2D transmissions take place in the uplink period. The D2D transmission and cellular users' transmissions take place on orthogonal frequencies; therefore, there is no interference among the D2D transmissions and cellular users' transmissions. However, the D2D receiver receives interference from all other D2D transmissions on the same channel. An example is shown in Figure 1 to explain the interference. UE (a) transmits to (b) while (c) transmits to (d). Both transmitters in the example are operating on the same channel. Transmission of (a) causes interference on the receiver (d); similarly, (c)'s transmission causes interference on (b). We assume $\mathcal{R} = \{1, 2, \dots, R\}$ are the RBs allocated to D2D transmission. A single D2D pair is allocated fraction of a single or multiple RBs ($r \in \mathcal{R}$).

In the quality adaptation, we restrict UE of very good SINR from always getting access to the network resources by requesting data through a limit on the download buffer. This is achieved with the help of parameter ρ . Detail usage of the ρ is discussed in Section IV-A. Moreover, for the video content, users QoE is greatly affected by variability in the quality of successive segments [18], [34]. We maintain low variability in quality of the successive segments of the video by two parameters β_{min} and β_{max} , a detailed discussion of which is given in Section IV-A.

B. COMMUNICATION MODEL

A binary variable a_{nm} is used for the D2D pair association which is defined as follows:

$$a_{nm} = \begin{cases} 1, & \text{If the requesting UE } n \text{ is paired with the} \\ & \text{providing UE } m \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

After UE n is paired with UE m , they begin the communication. UE n will continue putting request for RB to download layers of the video from UE m , according to the procedure presented in Section IV-A.

Subsequent to pairing the D2D users, the requesting UE need RB(s) for the communication. D2D pairs are assigned RBs for communicating the selected chunk of the video within a single frame time. For RB allocation, we use variable $x_{nm}^r \in [0, 1]$. Where x_{nm}^r denotes the fraction of RB r assigned to the D2D pair nm . The received SINR Γ_{mn}^r of the D2D pair is defined as follows:

$$\Gamma_{mn}^r(X) = \frac{p_{mn}^r h_{mn}^r}{\sum_{\substack{n' \neq n, \\ m' \neq m}} h_{m'n'}^r P_{m'n'}^r + n_0^2}, \quad (3)$$

where h_{mn}^r is the channel gain of the D2D pair mn , and n_0^2 is the Gaussian noise. $\sum_{\substack{n' \neq n, \\ m' \neq m}} h_{m'n'}^r P_{m'n'}^r$ is the power gain of all

D2D pairs on RB r other than the pair mn . The transmission rate of a D2D user on RB r is calculated as follows:

$$\gamma_{mn}^r(X) = W^r \log(1 + \Gamma_{mn}^r(X)), \quad (4)$$

where, W^r is the bandwidth of the D2D pair on RB r . Different users achieve different rates on the same RB. We need to ensure that the RB are assigned to the D2D pairs that are sufficient for downloading the entire layer of a segment during a frame T . We can represent this requirement using the following equation:

$$\sum_r x_{nm}^r T \gamma_{mn}^r \geq a_{nm} f(v_{n,s}^*) \quad \forall n, \forall m, \quad (5)$$

where the left side of eq (5) represents the rate that a D2D pair mn achieves for the portion of r RB(s), while the right side is the length of the requested layer of the segment.

C. PROBLEM FORMULATION

We formulate our optimization problem OPT as follows:

OPT :

$$\max_{X, A} \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}} a_{nm} f(v_{n,s}^*) x_{nm}^r T \gamma_{mn}^r \quad (6)$$

subject to:

$$\sum_r x_{nm}^r T \gamma_{mn}^r \geq a_{nm} f(v_{n,s}^*) \quad \forall n, \forall m, \quad (7)$$

$$\sum_{n \in \mathcal{N}} a_{nm} \leq 1 \quad \forall m \in \mathcal{M}, \quad (8)$$

$$\sum_{m \in \mathcal{M}} a_{nm} \leq 1 \quad \forall n \in \mathcal{N}, \quad (9)$$

$$x_{nm}^r \in [0, 1], \quad \forall n, \forall m, \forall r, \quad (10)$$

$$a_{nm} \in \{0, 1\}, \quad \forall n, \forall m, \quad (11)$$

$$\sum_r x_{nm}^r \leq R, \quad \forall n, \forall m. \quad (12)$$

Where constraint in eq. (7) ensures that the D2D pair is assigned a portion of the RB(s) that is sufficient to download the full layer of the segment during one time frame T .

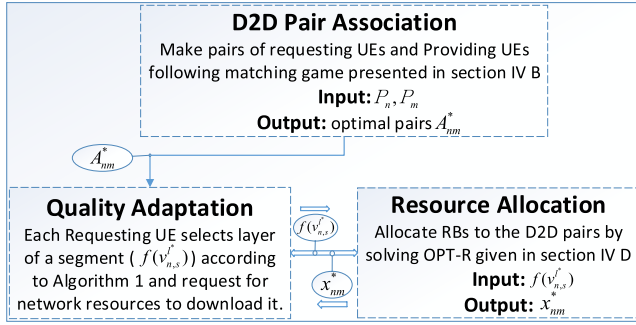


FIGURE 2. Proposed Framework.

Constraint in eq. (8) ensures that D2D receiver n can obtain layer of the video segment from only one D2D provider m . Similarly constraint in eq. (9) ensures that a D2D provider m can serve only one D2D receiver at time t . Constraint in eq. (10) represents the fraction of time for the RB r assigned to the D2D pair mn . The constraint in eq. (11) shows that the association variable a_{nm} is a binary variable. The constraint in eq. (12) ensures that the RBs allocated to all the D2D pairs do not exceed the total available RBs with the BS. The optimization problem in (6) is based on two variables X and A , both of which are interdependent. In the following subsections, we first fix the RB assignment X and solve the pair association A , similarly, we then fix the latter one and solve the former.

IV. PROPOSED SVC VIDEO COMMUNICATION IN D2D NETWORK

In this section, we present our proposed solution to the problem we formulated in Section III. Our solution is comprised of three components:

- **Quality adaptation:** When a video client n finishes download l th layer of s th segment, select the next layer for the same segment or a layer of the next segment of the video following the procedure presented in Section IV-A.
- **D2D pair association:** There can be more than one provider having the requested content; the quality may differ, or a UE may have content for more than one requesting UE at the same time. First, we pair the UE for D2D communication based on the SINR and video quality according to the mechanism presented in Section IV-B. The optimization problem in OPT-P is solved via matching game to achieve this objective.
- **Resource allocation:** Network resources are allocated to the D2D pairs, formed as a result of D2D user association, at the beginning of each slot T by solving the optimization problem **OPT-R** presented in Section IV-D to download the requested content $f(v_{n,s}^l)$, which aims to maximize the network throughput that results in delivering better quality video to the users.

In our proposed networking environment, the quality adaptation and resource allocation is an iterative process and the active set out of A_{nm}^* and the size of the requested content $f(v_{n,s}^l)$ depends upon the outcome of the quality adaptation that is affected by the delivery of the video data requested in

Algorithm 1 Quality Adaptation

```

1: Initialize:  $\rho, \beta_{min}, \beta_{max}$ 
2: for  $i = 1$ st segment of  $\rho$  to last segment of  $\rho$  do
3:   request  $\hat{l}$  {Comment:  $\hat{l}$  is next layer of a segment to
   already downloaded layers of that segment}
4: end for
5: current_segment = 1st segment of  $\rho$ 
6: next_segment = current_segment + 1
7: Previous_segment = NULL
8: while (Play-head < First segment of  $\rho$ ) & (downloaded
   layers for 1st segment of  $\rho$  < max_layers) do
9:    $\hat{l} =$  last downloaded layer of current_segment + 1
10:  while (current_segment  $\leq$  next_segment +  $\beta_{max}$ ) or
   (current_segment  $\leq$  previous_segment -  $\beta_{min}$ ) do
11:    request  $\hat{l}$ 
12:     $\hat{l} = \hat{l} + 1$ 
13:  end while
14:  if next_segment = NULL then
15:    current_segment = 1st_segment
16:    next_segment = current_segment + 1
17:    Previous_segment = NULL
18:  else
19:    Previous_segment = current_segment
20:    current_segment = next_segment
21:    if current_segment = Last segment of  $\rho$  then
22:      next_segment = NULL
23:    else
24:      next_segment = current_segment + 1
25:    end if
26:  end if
27: end while
28: if last segment of  $\rho$  is not the last segment of the video
   then
29:   slide  $\rho$  to the right
30: else
31:   reduce  $\rho$  from the left i.e.,  $\rho = \rho - 1$ 
32: end if
33: Go back to step 6 until  $\rho = 0$ 
34: exit

```

previous time window(s) and status of the download buffer. In the quality adaptation, these are controlled by ρ (discussed in the following section). All three modules of the proposed framework and their connections are summarized in Figure 2. Since quality adaptation indirectly affects the formulated optimization problem, therefore, we discuss it in the following section before going into the detail solution of the optimization problem.

A. QUALITY ADAPTATION

In this section, we present our layered video quality adaptation. To take full benefit of the decoupled layers in SVC video, we take the quality adaptation decision at the segment level. A UE can download any layer of the segment independently. However, for a single segment, layers

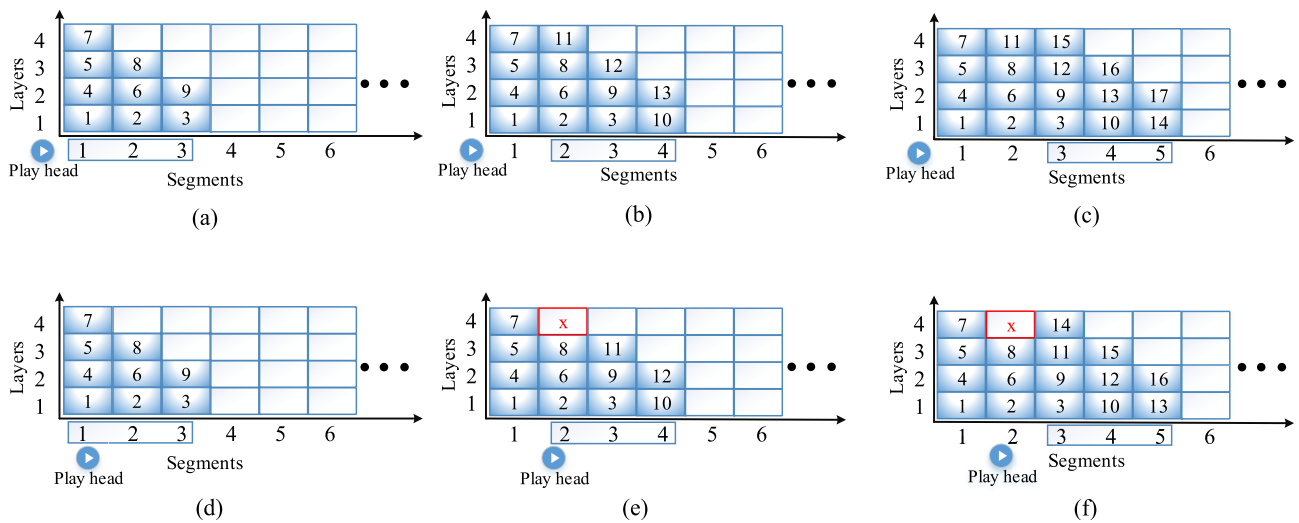


FIGURE 3. (a)(b)(c) shows sequence of segments' layers download and behavior of the bucket sliding in case of completing the download of all layers of segments. (d)(e)(f) shows sequence of segments' layers download and behavior of bucket sliding under the effect of video play-head.

are downloaded in their sequential order; that is, a layer is requested only if all the lower layers are already downloaded. To keep fairness among the users, we limit a user to request layers of segments within ρ segments from the current playing time. If a user has downloaded all the layers for the ρ segments, it will not be able to put further requests until it has played out at least one segment from the download buffer. We assume, in a single request, the UE requests one full layer of a segment. Upon the successful download of a layer of the segment, the UE selects next layer of a segment according to the procedure given in Algorithm 1, which can be a layer of the current segment or base/enhancement layer for one of the segments inside the sliding bucket ρ . In algorithm 1, at the start, UE set the sliding bucket size to ρ segments and start requesting base layer of the segments within ρ one by one and slide the bucket ρ according to the following definition:

Definition 1: The bucket is slides either all the layers for the first segment of ρ are downloaded or the UE has started playing the 1st segment of ρ . Inside the sliding bucket ρ , the UE requests layer of the same segment or moves to the next segment inside ρ according to the following definition:

Definition 2: Request layers of the current segment until the difference of the current segment and the next segment is equal to or greater than β_{max} or until the current segment is β_{min} less than the previous segment inside the ρ .

Figure 3 shows an example of how the sliding bucket ρ slides in case of downloading all the layers of a segment and under the role of the play-head; that is, the current playing segment. In this example, we have assigned values to the ρ , β_{max} and β_{min} as 3, 2, and 1, respectively (β_{max} and β_{min} are the difference of the downloaded layers for the adjacent segments of a video). In Figure 3(a), after completely downloading all the layers for the segment 1, the ρ is slide from 1, 2, and 3 to 2, 3, and 4, as shown in Figure 3(b). Similarly, ρ is slide to segments 3, 4, and 5 after completely downloading all the layers for segment 2, as shown in Figure 3(c). On the other hand, sub-figures (d)-(f) in Figure 3 show the role

of play-head in downloading layers of the video and sliding of the ρ . The play-head reaches to play segment 2 before downloading the last layer for it, therefore, the sliding bucket is slide and the download of the last layer of segment 2 is skipped.

Pseudo code of the quality adaptation is given in Algorithm 1. Steps 1-4, in Algorithm 1, are the initial phase in which the UE sets values for ρ , β_{min} , and β_{max} and sequentially downloads base layers of all the segments within the sliding bucket ρ . In steps 5-7, pointers for current, previous, and next segments are positioned. It is important to mention here that, in one request, the UE requests a layer of the current segment only. The next and previous segments are indicated for the purpose of stopping criteria of requesting layers for the current segment. The current segment pointer will move to the next segment in ρ , or to the first segment in ρ if the current segment is the last segment of ρ , when either the current segment is β_{min} lesser than the previous segment or β_{max} greater than the next segment. The size/quality of the segments are measured in terms of the number of downloaded layers for the segments. The while loop, in line number 8, ensures that the current segment and selected layer remain in the boundaries of ρ and maximum layers for the selected video. In lines 10-13 the UE sequentially requests layers for the current segment until the limit for β_{max} or β_{min} is exceeded. In lines 14-27 the pointers for current, next and previous segments are adjusted within the sliding bucket ρ . In lines 28-32, movement of the sliding bucket ρ is controlled. The bucket is slide to the right if the UE has started playing the segment of the ρ or all the layers for the 1st segment of the ρ are downloaded. The bucket ρ starts reducing when it reaches the last segment of the selected video. Video downloading is completed when the ρ size reaches zero.

B. D2D PAIR ASSOCIATION

Content providing UE periodically broadcast tables of the cached contents along with the respective quality of the

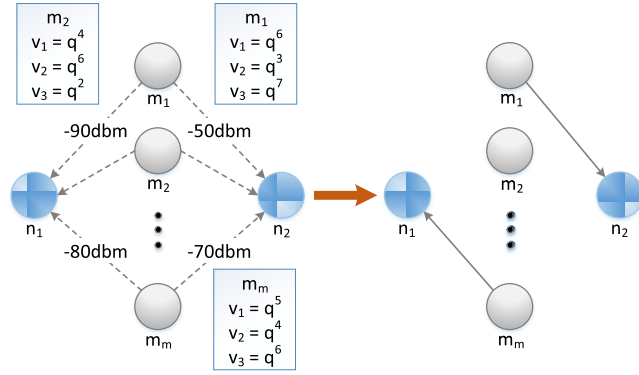


FIGURE 4. An illustration of D2D pair association.

contents to its one hop neighbors. The one hop neighboring UE save the received content information in a content information table. Thus, every UE has knowledge of the available contents and their respective qualities in the cache of all one-hop neighboring UE. Whenever the UE needs some content, it searches the content information table and finds out all the potential providing UE and associates with a suitable UE among them to get the content via the D2D link.

Figure 4 illustrates an example of two different requesting UE competing to associate with the providing UE. In the example, requesting UE (n_1 and n_2) request video v_1 that is present in the cache of providing UE m_1 , m_2 and m_3 in different qualities (q^x). Both the requesting UE achieve different SINR with providing UE, visible in the figure. According to our solution presented in Algorithm 2, n_1 is paired with m_3 and n_2 is paired with m_1 . Initially, m_1 was the first choice of both the requesting UE. However, m_1 rejects n_1 because of low SINR (-90dbm) and accepts n_2 that achieves better SINR (-50dbm). The n_1 is paired with m_m that is the second best in its priority list. We represent the D2D pair association in the following optimization problem:

OPT – P

$$\max_{\mathbf{A}} \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}} a_{nm} f(v_{n,s}^*) x_{nm}^r T \gamma_{mn}^r \quad (13)$$

subject to:

$$\sum_{n \in \mathcal{N}} a_{nm} \leq 1 \quad \forall m \in \mathcal{M}, \quad (14)$$

$$\sum_{m \in \mathcal{M}} a_{mn} \leq 1 \quad \forall n \in \mathcal{N} \quad (15)$$

$$a_{nm} = \{0, 1\}, \quad \forall n, \forall m. \quad (16)$$

Constraint in eq. (14) ensures that D2D receiver n can receive video segment from only one D2D provider m , and similarly, constraint (15) ensures that a D2D provider m can serve only one D2D receiver at time t . Constraint in eq. (16) shows that the association variable a_{nm} is a binary variable.

For the optimization problem OPT-P, n requesting UE need to be paired with m providing UE in such a way that there is one providing UE for one requesting UE, as given in constraints (14) and (15). Therefore, OPT-P is a combinatorial problem and is not solvable in a feasible amount of time for

Algorithm 2 D2D Association Algorithm

```

1: Phase 1: Initialization:
2: input:  $\mathcal{P}_m, \mathcal{P}_n, \forall m, n$ 
3: initialize:  $t = 0, \mu^{(t)} \triangleq \{\mu(m)^{(t)}, \mu(n)^{(t)}\}_{m \in \mathcal{M}, n \in \mathcal{N}} = \emptyset, \mathcal{R}_m^{(t)} = \emptyset, \mathcal{P}_m^{(0)} = \mathcal{P}_m, \mathcal{P}_n^{(0)} = \mathcal{P}_n, \forall m, n$ 
4: Phase 2: Matching:
5: repeat
6:    $t \leftarrow t + 1$ 
7:   for  $n \in \mathcal{N}$ , propose  $m$  according to  $\mathcal{P}_n^{(t)}$  do
8:     if  $n \succ_m \mu(m)^{(t)}$  then
9:        $\mu(m)^{(t)} \leftarrow \mu(m)^{(t)} \setminus n'$ 
10:       $\mu(m)^{(t)} \leftarrow n$ 
11:       $\mathcal{P}'_m = \{n' \in \mu(m)^{(t)} | n \succ_m n'\}$ 
12:    else
13:       $\mathcal{P}''_m = \{n \in \mathcal{N} | \mu(m)^{(t)} \succ_m n\}$ 
14:    end if
15:     $\mathcal{R}_m^{(t)} = \{\mathcal{P}'_m\} \cup \{\mathcal{P}''_m\}$ 
16:    for  $l \in \mathcal{R}_m^{(t)}$  do
17:       $\mathcal{P}_l^{(t)} \leftarrow \mathcal{P}_l^{(t)} \setminus \{m\}$ 
18:       $\mathcal{P}_m^{(t)} \leftarrow \mathcal{P}_m^{(t)} \setminus \{l\}$ 
19:    end for
20:  end for
21: until  $\mu^{(t)} = \mu^{(t-1)}$ 
22: Phase 3: Resource Allocation:
23: output:  $\mu^{(t)}$ 

```

a practical sized large network [35]. Moreover, OPT-P is to be solved by each D2D pair in such a manner that UE from both sides; that is, providing UE and requesting UE are paired with their preferred set of UE. Therefore, we apply matching theory to solve this optimization problem because it can solve the combinatorial problem in a distributed manner [24], [36]. Moreover, matching theory enables players (UE) to define their individual preferences based on local information. The details of solving the problem via matching theory are in the following section.

C. MATCHING GAME FORMULATION

In our communication scenario, there is one provider for each requester and vice-versa. Therefore, we have a one-to-one matching game. We define our matching as follows:

Definition 3: A matching μ is defined using a function from the set $\mathcal{N} \cup \mathcal{M}$ into the set of elements of $\mathcal{N} \cup \mathcal{M}$ such that

- 1) $|\mu(n)| \leq 1$ and $\mu(n) \in \mathcal{M}$,
- 2) $|\mu(m)| \leq 1$ and $\mu(m) \in \mathcal{N} \cup \phi$,
- 3) $\mu(n) = m$ if and only if m is in $\mu(n)$,

where $\mu(n) = \{m\} \Leftrightarrow \mu(m) = \{n\}$ for $\forall n \in \mathcal{N}, \forall m \in \mathcal{M}$ and $|\mu(\cdot)|$ represents the cardinality of the matching outcome $\mu(\cdot)$. The first two properties are because of the constraints (14) and (15) of problem OPT-P, i.e., the matching has a one-to-one relation in the sense that a D2D requester n can be associated with one D2D provider m . Additionally, when a D2D pair is not allowed to be associated owing to any reason we have $\mu(n) = \phi$.

1) PREFERENCE PROFILES OF PLAYERS

In this section, we first formulate our pair association as a two-sided matching game. In our model, we have two types of D2D pairs. The first type is the providers $\mathcal{M}$ and the second type is the requesters $\mathcal{N}$. In the two-sided matching game, each player on either side needs to rank the players of the other side in descending order of their priority/preference; this is called preference profile.

As a result of the content announcement by the providers in the one-hop neighboring UE, each requesting UE n has the knowledge of the availability of its required content with all the potential neighboring providing UE $\mathcal{M}$, along with the available qualities. Requesting UE n ranks the potential content providing UE in a one-hop communication range based on available quality level (number of layers). For the content requesting UE, the preference of that node is higher for UE possessing higher quality for the requested content. The content requesting node n creates the preference profile as follows:

$$\mathcal{P}_n = q_{mn}^l, \forall n \in \mathcal{N}. \quad (17)$$

The higher quality means more layers in the video and larger the size of $f(v_{n,s}^l)$ and thus it will directly affect the utility function of OPT-P. Similarly, the content provider prefers to select the UE from $\mathcal{N}$ with the best SINR so that the requested content is delivered in less time, thereby consuming less energy in the transmission. The preference profile for the content provider in D2D pair m is based on the achievable data rate, which is a function of SINR, expressed as follows:

$$\mathcal{P}_m = \log(1 + \Gamma_{mn}), \forall m \in \mathcal{M}. \quad (18)$$

2) D2D ASSOCIATION ALGORITHM

In this section, we discuss our D2D pair association algorithm for our proposed problem in (OPT – P), as shown in Algorithm 2. In the initialization phase, both the requesting UE and the providing UE make their preference profiles $\mathcal{P}_n$ and $\mathcal{P}_m$, respectively. In the second phase, stable matching is performed. The matching criteria is assigned according to the following definition.

Definition 4: A matching μ is considered stable when there is no blocking pair (m, n) , where $m \in \mathcal{M}, n \in \mathcal{N}$, such that $n \succ_m \mu(m)$ and $m \succ_n \mu(n)$, where $\mu(m)$ and $\mu(n)$ represent, respectively, the current matched partners of m and n .

Pseudocode for the proposed game is given in Algorithm 2. The presented algorithm is a variant of the “deferred-acceptance algorithm”, which is guaranteed to converge to a stable allocation [37]. The output of our matching $\mu^{(t)}$ is the association vector of D2D pairs that maximizes the objective of the optimization problem. The output $\mu^{(t)}$ of Algorithm 2 can be transformed to a feasible allocation vector $\mathbf{A}$ of problem OPT-P.

D. RESOURCE ALLOCATION FOR THE ALLOCATION PHASE

The matching game from the previous section produces optimal D2D pairs a_{nm}^* . After the pair formation, network resources are needed for the communication. In this section,

we present resource allocation x_{nm}^r to these optimal D2D pairs. The resource allocation subproblem is given as follows:

OPT – R

$$\max_x \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}} a_{nm}^* f(v_{n,s}^{l*}) x_{nm}^r T \gamma_{mn}^r \quad (19)$$

subject to:

$$\sum_r x_{nm}^r T \gamma_{mn}^r \geq a_{nm}^* f(v_{n,s}^{l*}) \quad \forall n, \forall m, \quad (20)$$

$$x_{nm}^r = [0, 1], \quad \forall n, \forall m, \forall r, \quad (21)$$

$$\sum_r x_{nm}^r \leq R, \quad \forall n, \forall m. \quad (22)$$

Where constraint in eq. (20) ensures that the D2D pair is assigned portion of a single RB or multiple RBs that are sufficient to complete the download of the requested layer of the video segment during one time frame T . Constraint (21) represents the fraction of RB r (in the range from zero to one) assigned to the D2D pair mn . Constraint in (22) ensures that the RB allocated to all the D2D pairs do not exceed the total available RBs with the BS.

The optimization subproblem OPT – R is solved for the optimum D2D pairs that are selected as a result of the solution of optimization problem OPT-P presented in Section IV-B. D2D pair nm is assigned the fraction of an RB that satisfies the minimum rate requirement for downloading the requested layer within T .

Constraint (20) satisfies this condition. To efficiently utilize the available RBs to serve the D2D pairs, we divide the whole network into several clusters. Moreover, we consider orthogonal RBs allocation in each cluster such that all the RBs deliver the same channel gain to the D2D pairs. To satisfy constraint (22), we state the following proposition.

Proposition 1: We consider orthogonal RBs in every cluster with the same channel gain for a D2D pair nm . As x_{nm}^r denotes the fraction of time when the RB r is allocated to D2D pair nm , we allocate the RB to the D2D pairs in the descending order of SINR. This implies that the best SINR pair is allocated fraction of an RB first, then the second best, and so on. When one RB is fully utilized, then the next RB is assigned to the pairs in the same fashion until all the RBs are finished. Since, a D2D pair experiences the same channel gain on each RB, means the same throughput is obtained on every RB. Therefore, the notation x_{nm}^r is replaced with x_{nm} for brevity.

As given in the Proposition 1, the RBs are assigned to the D2D pairs until either one is finished. Therefore, we can eliminate the constraint given in eq. (22). Subsequent to resolving the constraint (22), the resource allocation subproblem is given as

OPT – R2

$$\max_x \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} a_{nm}^* f(v_{n,s}^{l*}) x_{nm} T \gamma_{mn}^r \quad (23)$$

subject to:

$$x_{nm} T \gamma_{mn}^r \geq a_{nm}^* f(v_{n,s}^{l*}) \quad \forall n, \forall m, \quad (24)$$

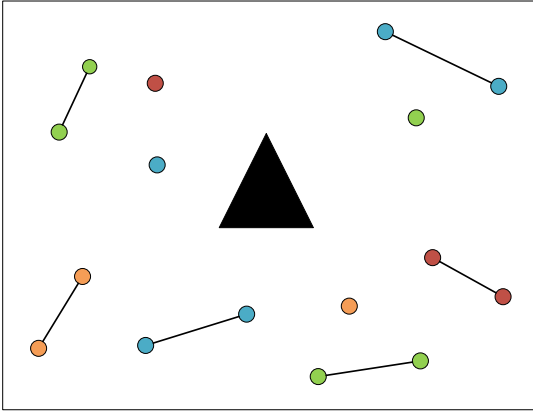
$$x_{nm} = [0, 1], \quad \forall n, \forall m. \quad (25)$$

Algorithm 3 Resource Allocation

```

1: Initialize:  $A^*, \mathcal{N}, \mathcal{M}$ 
2: for  $r \in \mathcal{R}$  do
3:   Sort the D2D pairs in descending order of SINR
4:   for  $mn \in A$  do
5:     while  $T > 0$  do
6:       Allocate  $x_{mn}^*$  using eq. (26)
7:       Set  $T = T - x_{mn}^*$ 
8:     end while
9:   end for
10: end for
11: output:  $x_{mn}^*$ 

```

**FIGURE 5.** Example of network topology for the simulation.

The problem OPT-R2 is a constrained maximization problem, and it has a boundary solution at constraint (24) with equality. Therefore, the optimal fraction of RB allocated to a D2D pair is given as

$$x_{nm}^* = \frac{a_{nm}^* f(v_{n,s}^{j*})}{T \gamma_{mn}^r}. \quad (26)$$

The Algorithm 3 presents the procedure of resource allocation to the D2D pairs A_{nm}^* .

V. PERFORMANCE EVALUATION

In this section, we present MATLAB based simulation results of our proposed mechanism to demonstrate its effectiveness. We consider an uplink system of cellular communication environment wherein, we assume that BS is deployed at a fixed location and N D2D UE are deployed according to Homogeneous Poisson Point Process. We assume that the D2D UE are assigned 3MHz bandwidth. For all simulations, we have assumed the channel gain is distributed as independent and identically distributed Rayleigh random variables with mean value $h(d) = h_0(d/15) - 4$ where h_0 is the reference channel gain at a distance of 15 meters. Table 2 shows the parameters used for the simulations unless stated otherwise. These parameters are tuned in accordance with guidelines given for modeling the system in [40]–[42]. For all these experiments, we consider layered videos each of a length of 200 seconds. The videos are divided into 100 equal-sized

TABLE 2. Simulation Parameters used in Simulations.

Simulation Parameters	Values
Radius of MBS	500 m
Structure of Frame	Type 1 (FDD)
Carrier frequency (f)	2 GHz
Transmission Time Interval (TTI)	10 ms
Bandwidth of the System	3 MHz
D2D UE total transmit power	23 dBm
Bandwidth of each RB (W)	180 kHz
Spacing of neighboring sub-carrier	15 kHz
Number of sub-carriers per RB	12
Sliding bucket (ρ) size	3
Modulation and coding scheme (MCS)	QPSK: 3/5, 1/2, 1/3, 1/6, 1/9, 1/12 16QAM: 3/5, 1/2, 1/3
Path loss (D2D links)	$32.45 + 20 \log(d) + 20 \log(f)$, f[MHz]
standard deviation of Shadow fading	3 dB
Thermal noise for 1 Hz at 20°C	-174 dBm

segments, and each segment is 2 seconds duration. There are ten layers in the videos and all the layers are of equal size. Successful delivery of each layer improves PSNR of the video by the factor of 5 dB [43], [44]. We consider a low-mobility network environment like [24], [45], wherein the topology does not change during a single run of experiment. In real life, this may be resembled the environment of an office, bus, train, sports stadium etc. All the statistical results are averaged over 100 run of random locations of D2D pairs, and RBs gains.

A. SIMULATION RESULTS AND DISCUSSION

First, we discuss the resource allocation to each pair in the network and the pairs that do not get resources for the communication and are unable to have the transmission. For this simulation, we used a network of 160 D2D UE (80 pairs). Results are shown in Figure 6. In the figure, the x-axis shows the index of D2D pairs from 1 to 80. The y-axis shows the fraction of an RB assigned to the D2D pair. We can see in the figure that as the RBs increase, the number of UE get the resources for transmission are increased. With 10 RBs 30 UE can be served, similarly 53 and 71 UE can be served with 20 and 30 RBs respectively.

Figure 7 shows results for the number of layers downloaded on an average by the D2D pairs in the network for our proposed matching and quality adaptation scheme. For these results, we have considered a network of 60 UE (30 D2D pairs). In the figure, y-axis shows the number of layers downloaded for the video(s) on average for all the segments of the video, while the x-axis shows the index number of the D2D pair. In the figure, we can see that just three UE are able to watch the video with seven or more layers on an average, these are the UE with high SINR. Some poor SINR users are able to see the video with around two layers on an average.

Herein, we present a comparison of our proposed quality adaptation, D2D pair association and resource allocation with some other schemes. Figure 8 shows the comparison of our

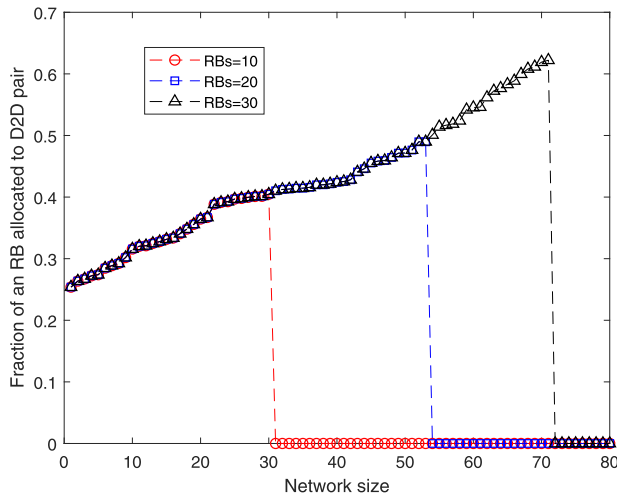


FIGURE 6. Fraction of RBs allocation to D2D pairs.

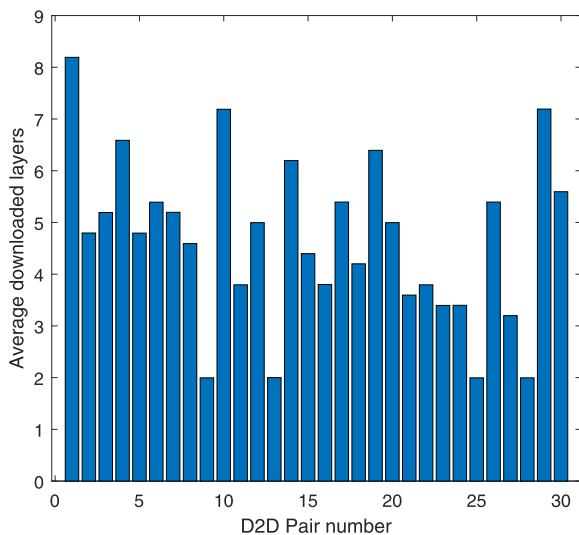


FIGURE 7. Video layers downloaded by each video requester.

proposal with sequential layer adaptation and Constrained Stream Loading (CSL) layer adaptation. For all the aforementioned schemes, we have kept the bucket size the same; however, the choice of which layer of which segment to request is different for every scheme. For the sequential, the UE requests layers of the segments in sequential order inside the bucket i.e., first, it requests the base layer of all the segments in the sliding bucket one-by-one, then the enhancement layer one for all the segments inside the sliding bucket, and so on. In the CSL, the requesting UE randomly decides the number of layers for the first segment inside the ρ and then requests those layers for the segment sequentially, and then moves to the next segment, and so on. The CSL is fairly similar to the proposal in [9]; however, we randomly decide the number of layers for a segment and they decide the number of layers according to their predicted rate during the subsequent time-slot. In the figure, we can see that our proposed scheme outperforms the other two schemes. The sequential scheme suffers the most because the layers download is interrupted

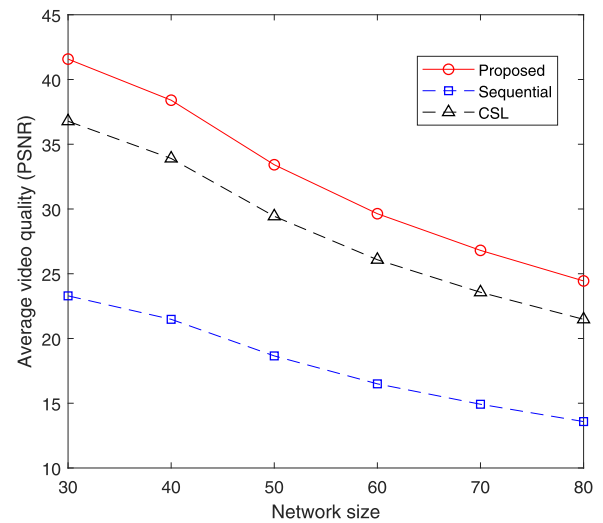


FIGURE 8. Video quality comparison for different quality adaptation schemes.

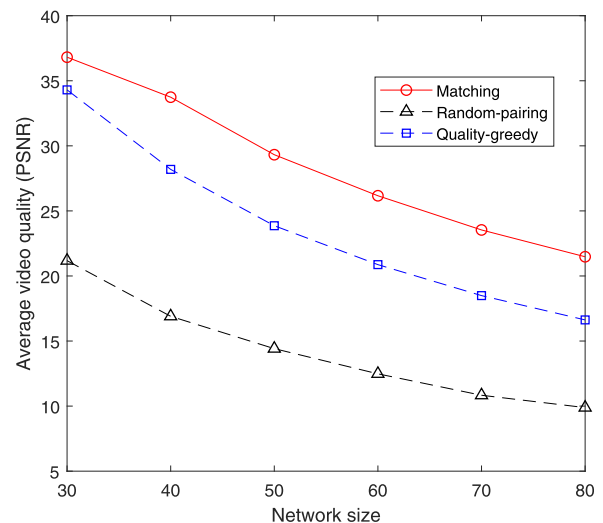


FIGURE 9. Video quality comparison for different D2D pairing schemes.

by the playhead before all the layers been downloaded and the bucket slides to the subsequent segment.

Figure 9 shows results for the downloaded video quality with respect to network size (number of D2D pairs) for different pair association schemes. In the figure, the comparison of the proposed matching based D2D pair association is analyzed with random pair association and quality greedy schemes. As obvious from the name, UE are randomly paired in the random scheme. In the quality greedy scheme, requesting UE associates with the D2D provider that possesses the highest quality of the requested video. In case a providing UE possesses the desired video of more than one requesting UE, one of them is paired and the remaining one is paired with the second-best in his list. For all the aforementioned schemes, the quality adaptation scheme is the proposed quality adaptation. In the figure, we can see that the quality greedy scheme's performance is nearer to the proposed matching pair association; however, it cannot supersede the proposed matching.

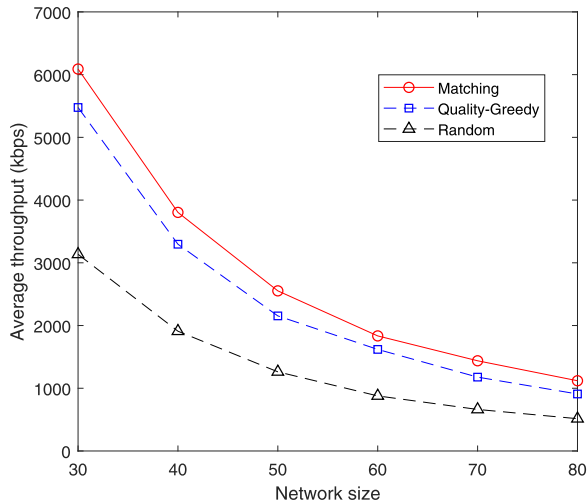


FIGURE 10. Throughput achievement under different pairing schemes.

The reason for this is, the providing UE have the highest quality of the requested video; however, the SINR of them may not be the best and they will not be able to get enough network resources to download all the available layers before the playhead interrupts the download. For the random pairing, the performance is the worst as expected.

Figure 10 shows the throughput comparison of the proposed matching based pairing with quality greedy based pairing and random pairing. The throughput shown is per user throughput on average. Our proposed matching based pairing outperforms the other two pairing schemes. In line with expectations, the throughput decreases as the network size increases because the resources are kept the same regardless of the network size for this experiment.

VI. CONCLUSION

We jointly proposed mechanisms for quality adaptation, resource allocation and pair association for layered video in the D2D communication network. The proposed quality adaptation mechanism exploits the layers decoupling property of the SVC video and takes the quality adaptation decision on the layer level of a segment. The D2D association optimization problem is mapped to a one-to-one matching game and is solved via “Gale-Shapley algorithm” in which the content requesting UE prefers to associate with the content providing UE that possesses the highest quality of the requested video. However, the providing UE prefers to associate with the requesting UE that achieves the highest SINR so as to spend less energy for the transmission. For the resource allocation, the formulated optimization problem is a linear optimization problem that is solved on the boundary that guarantees UE are allocated a fraction of an RB or multiple RBs to download the requested layer of the segment within the frame duration. We numerically evaluated the proposed mechanism and our experimental results demonstrate the significance of the proposed mechanism.

REFERENCES

- [1] Cisco Visual Networking Index. Accessed: Jan. 3, 2019. [Online]. Available: <https://www.cisco.com>
- [2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *Proc. 5th Int. Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, 2009, pp. 1–12.
- [3] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, “Base-station assisted Device-to-Device communications for high-throughput wireless video networks,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, Jul. 2014.
- [4] S. Ullah, K. Thar, and C. S. Hong, “Management of scalable video streaming in information centric networking,” *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21519–21546, Oct. 2016.
- [5] D. Marpe, T. Wiegand, and G. J. Sullivan, “The H.264/MPEG4 advanced video coding standard and its applications,” *IEEE Commun. Mag.*, vol. 44, no. 8, pp. 134–143, Aug. 2006.
- [6] H. Schwarz and M. Wien, “The scalable video coding extension of the H. 264/AVC standard,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, p. 135, Mar. 2008.
- [7] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
- [8] D. Kulinski and J. Burke, “NDNVideo: Random-access live and pre-recorded streaming using NDN,” NDN, Tech. Rep. NDN-0007, Revision 1, Sep. 2012. [Online]. Available: <http://named-data.net/techreports.html>
- [9] S. A. Hosseini, Z. Lu, G. de Veciana, and S. S. Panwar, “SVC-based multi-user streamloading for wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2185–2197, Aug. 2016.
- [10] J. Kua, G. Armitage, and P. Branch, “A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1842–1866, 3rd Quart., 2017.
- [11] Y. Sun, X. Yin, J. Jiang, V. Sekar, F. Lin, N. Wang, T. Liu, and B. Sinopoli, “CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction,” in *Proc. Conf. ACM SIGCOMM Conf. (SIGCOMM)*, 2016, pp. 272–285.
- [12] J. Jiang, V. Sekar, and H. Zhang, “Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with festive,” *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 326–340, Feb. 2014.
- [13] S. Ullah, K. Thar, M. G. R. Alam, J. Hyeok Son, J. Won Lee, and C. S. Hong, “Delivering scalable video streaming in ICN enabled long term evolution networks,” in *Proc. 18th Asia-Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Oct. 2016, pp. 1–4.
- [14] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, and S. Cheng, “Scheduling and resource allocation for wireless dynamic adaptive streaming of scalable videos over HTTP,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 1681–1686.
- [15] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, and S. Cheng, “QoE-driven cross-layer optimization for wireless dynamic adaptive streaming of scalable videos over HTTP,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 451–465, Mar. 2015.
- [16] C. Wang, A. Rizk, and M. Zink, “SQUAD: A spectrum-based quality adaptation for dynamic adaptive streaming over HTTP,” in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, 2016, p. 1.
- [17] C. Zhou, C.-W. Lin, and Z. Guo, “mDASH: A Markov decision-based rate adaptation approach for dynamic HTTP streaming,” *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 738–751, Jan. 2016.
- [18] V. Joseph and G. de Veciana, “NOVA: QoE-driven optimization of DASH-based video delivery in networks,” in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Apr. 2014, pp. 82–90.
- [19] W. U. Rahman, C. S. Hong, and E.-N. Huh, “Edge computing assisted joint quality adaptation for mobile video streaming,” *IEEE Access*, vol. 7, pp. 129082–129094, 2019.
- [20] J. Xie, R. Xie, T. Huang, J. Liu, and Y. Liu, “Energy-efficient cache resource allocation and QoE optimization for HTTP adaptive bit rate streaming over cellular networks,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [21] T. LeAnh, N. H. Tran, W. Saad, L. B. Le, D. Niyato, T. M. Ho, and C. S. Hong, “Matching theory for distributed user association and resource allocation in cognitive femtocell networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8413–8428, Sep. 2017.
- [22] O. Semiari, W. Saad, S. Valentin, M. Bennis, and H. V. Poor, “Context-aware small cell networks: How social metrics improve wireless resource allocation,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5927–5940, Nov. 2015.

- [23] R. M. Radaydeh, F. Al-Qahtani, A. Celik, K. A. Qaraqe, and M.-S. Alouini, "Imperfect D2D association in spectrum-shared cellular networks under interference and transmit power constraints," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.
- [24] S. M. A. Kazmi, N. H. Tran, W. Saad, Z. Han, T. M. Ho, T. Z. Oo, and C. S. Hong, "Mode selection and resource allocation in Device-to-Device communications: A matching game approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3126–3141, Nov. 2017.
- [25] S. Ullah, T. LeAnh, A. Ndikumana, M. G. R. Alam, and C. S. Hong, "Layered video communication in ICN enabled cellular network with D2D communication," in *Proc. 19th Asia-Pacific Netw. Operations Manage. Symp. (APNOMS)*, Sep. 2017, pp. 199–204.
- [26] R. Haw, S. M. A. Kazmi, K. Thar, M. G. R. Alam, and C. S. Hong, "Cache aware user association for wireless heterogeneous networks," *IEEE Access*, vol. 7, pp. 3472–3485, 2019.
- [27] A. Galanopoulos, G. Iosifidis, A. Argyriou, and L. Tassioulas, "Green video delivery in LTE-based heterogeneous cellular networks," in *Proc. IEEE 16th Int. Symp. A World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2015, pp. 1–9.
- [28] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.
- [29] D. Liu, L. Wang, Y. Chen, M. ElKashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.
- [30] A. Ndikumana, S. Ullah, D. H. Kim, and C. S. Hong, "Deep-AuC: Joint deep learning and auction for congestion-aware caching in named data networking," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0220813.
- [31] A. Ndikumana, S. Ullah, K. Thar, N. H. Tran, B. Ju Park, and C. S. Hong, "Novel cooperative and fully-distributed congestion control mechanism for content centric networking," *IEEE Access*, vol. 5, pp. 27691–27706, 2017.
- [32] K. Thar, S. Ullah, R. Haw, T. LeAnh, T. Z. Oo, and C. S. Hong, "Hybrid caching and requests forwarding in information centric networking," in *Proc. 17th Asia-Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Aug. 2015, pp. 203–208.
- [33] V. Joseph and G. de Veciana, "Jointly optimizing multi-user rate adaptation for video transport over wireless systems: Mean-fairness-variability tradeoffs," in *Proc. Proc. IEEE INFOCOM*, Mar. 2012, pp. 567–575.
- [34] C. Yim and A. C. Bovik, "Evaluation of temporal variation of video quality in packet loss networks," *Signal Process., Image Commun.*, vol. 26, no. 1, pp. 24–38, Jan. 2011.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [36] S. M. A. Kazmi, N. Tran, W. Saad, L. B. Le, T. M. Ho, and C. S. Hong, "Optimized resource management in heterogeneous wireless networks," *IEEE Commun. Lett.*, to be published.
- [37] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *Amer. Math. Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
- [38] C. Xia, S. Xu, and K. S. Kwak, "Overlapping coalition formation games-based resource allocation for device-to-device communication in LTE-A network," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 12, pp. 1295–1305, Jul. 2015.
- [39] T. LeAnh, N. H. Tran, S. Lee, E.-N. Huh, Z. Han, and C. S. Hong, "Distributed power and channel allocation for cognitive femtocell network using a coalitional game in partition-form approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3475–3490, Apr. 2016.
- [40] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures, Release 11*, document TS 36.213, 3rd Generation Partnership Project (3GPP), 3GPP, Dec. 2012.
- [41] *Study on LTE Device to Device Proximity Services: Radio Aspects*, document TR 22.803, 3rd Generation Partnership Project (3GPP), 3GPP, 3GPP Standard, May 2014.
- [42] *Huawei and HiSilicon, Channel Model for D2D Evaluations* GPP TSG RAN WG1 Meeting no. 73, May 2013.
- [43] P. Seeling and M. Reisslein, "Video traffic characteristics of modern encoding standards: H.264/AVC with SVC and MVC extensions and H.265/HEVC," *Sci. World J.*, vol. 2014, Feb. 2014, Art. no. 189481.
- [44] S. Ullah, "Energy and distortion analysis of video compression schemes in multi-hop wireless video sensor networks," Ph.D. dissertation, School Elect. Eng. Comput. Sci., Nat. Univ. Sci. Technol., Islamabad, Pakistan, 2010.
- [45] S. M. A. Kazmi, N. H. Tran, T. M. Ho, A. Manzoor, D. Niyato, and C. S. Hong, "Coordinated device-to-device communication with non-orthogonal multiple access in future wireless cellular networks," *IEEE Access*, vol. 6, pp. 39860–39875, 2018.



SAEED ULLAH received the M.S. degree in information technology from the National University of Sciences and Technology (NUST), Islamabad, in 2010. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Kyung Hee University, South Korea. His research interests include multimedia communication, scalable video streaming routing and in-network caching, D2D communication, and the future Internet. He has received several best paper awards from prestigious conferences.



KITAE KIM received the B.S. and M.S. degrees in computer science and engineering from Kyung Hee University, Seoul, South Korea, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree in computer science and engineering. His research interests include SDN/NFV, wireless networks, unmanned aerial vehicle communications, and machine learning.



AUNAS MANZOOR received the M.S. degree in electrical engineering with a specialization in telecommunication from the National University of Science and Technology, Pakistan, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Kyung Hee University, South Korea. His research interest is applying analytical techniques for resource management in mobile cellular networks.



LATIF U. KHAN received the M.S. degree (Hons.) in electrical engineering from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2017. He is currently pursuing the Ph.D. degree in computer engineering with Kyung Hee University (KHU), South Korea. Prior to joining KHU, he has served as a faculty member and a Research Associate at UET. He is working as a leading Researcher with the Intelligent Networking Laboratory under a project jointly funded by the prestigious Brain Korea 21st Century Plus and Ministry of Science and ICT, South Korea. He has published his works in highly reputable conferences and journals. His research interest includes analytical techniques of optimization and game theory to edge computing and end-to-end network slicing.



S. M. AHSAN KAZMI received the master's degree in communication system engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2012, and the Ph.D. degree in computer science and engineering from Kyung Hee University (KHU), South Korea. He is currently with the Institute of Information Systems (IIS), Innopolis University, Innopolis, Tatarstan, Russia, where he is working as an Assistant Professor. His research interest includes applying analytical techniques of optimization and game theory to radio resource management for future cellular networks. He received the best KHU Thesis Award in engineering, in 2017, and several best paper awards from prestigious conferences.



CHOONG SEON HONG (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Minato, Japan, in 1997. In 1988, he joined Korea Telecom, where he worked on broadband networks as a member of Technical Staff. In September 1993, he joined Keio University. He worked at the Telecommunications Network Laboratory, Korea Telecom, as a Senior Member of Technical Staff and the Director of the Networking Research Team, until August 1999. Since September 1999, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University. His research interests include the future Internet, ad hoc networks, network management, and network security. He is a member of ACM, IEICE, IPSJ, KIISE, KICS, KIPS, and OSIA. He has served as the General Chair, a TPC Chair/member, or an Organizing Committee member for international conferences, such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, TINA, SAINT, and ICOIN. In addition, he is currently an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, *International Journal of Network Management*, and *Journal of Communications and Networks* and an Associate Technical Editor of the *IEEE Communications Magazine*.

• • •