國立臺灣大學電機資訊學院電機工程學系

碩士論文

Department of Electrical Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master’s Thesis

結合視圖合成與 D2D 協同合作之高效多視角視訊傳輸

Efficient Multi-View Video Transmission with View Synthesis and D2D Cooperation

劉濬綸

Brian Liu

指導教授: 廖婉君 博士

Advisor: Wan-Jiun Liao Ph.D.

中華民國 113 年 7 月

July, 2024

doi:10.6342/NTU202402157

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

結合視圖合成與 D2D 協同合作之高效多視角視
訊傳輸

Efficient Multi-View Video Transmission with View Synthesis and D2D Cooperation

本論文係劉濬綸（R11921040）在國立臺灣大學電機工程學系完成之碩士學位論文，於民國 113 年 7 月 10 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department of Electrical Engineering on 10 July 2024 have examined a Master's thesis entitled above presented by Brian Liu (R11921040) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

廖婉君
（指導教授 Advisor）

林風　　林宗男　　楊得年

系主任 Director: 李建模

doi:10.6342/NTU202402157

摘要

多視角視訊 (MVV) 的出現提供了沉浸式的觀看體驗，但與此同時這需要高頻寬的無線傳輸。為了解決這一項挑戰，過往研究提出了深度圖像繪圖法 (DIBR) 技術，以合成觀看者所需的視角，而不是直接傳送整個觀看範圍。然而，在考慮 DIBR 技術的 MVV 系統中，D2D 通信的使用情況尚未得到探討。在本文中，我研究了在 MVV 視訊傳輸系統中，合成和 D2D 協同合作的視角分配問題 (SDVA)，並使用賽局方法分析考慮伺服器和觀看者端各自的策略。我制定了一個密封投標拍賣並為 SDVA 開發了一個多項式時間的分配規則，以最大化社會福利。我證明了該演算法保證了真實投標的特性並具有接近最佳解的 1/2 近似比率。模擬結果表明，與現有算法相比，此算法在總社會福利和平均視角滿意度方面表現更優。

關鍵字：多視角視訊、視角合成、D2D 通訊、賽局理論、資源分配

doi:10.6342/NTU202402157

Abstract

Multi-view video (MVV) enables an immersive viewing experience, which requires high bandwidth for wireless transmission. To address this challenge, Depth-Image-Based Rendering (DIBR) has been proposed to synthesize the desired views, instead of delivering the entire viewing range. However, the scenario of D2D communication in a DIBR-enabled MVV system has not been explored. In this paper, we investigate Synthesis and D2D-assisted View Allocation (SDVA) in an MVV video delivery system, while considering the strategies of the server and users with a game-theoretic approach. We formulate a sealed-bid auction and develop a polynomial allocation rule for SDVA to maximize social welfare. We prove that the algorithm guarantees truthful bidding and has a 1/2-approximation ratio to the optimal solution. The simulation results show that our algorithm outperforms existing algorithms regarding total social welfare and average view satisfaction.

Keywords: Multi-view video, view synthesis, device-to-device (D2D) communication, game theory, resource allocation

doi:10.6342/NTU202402157

Contents

doi:10.6342/NTU202402157

doi:10.6342/NTU202402157

List of Figures

doi:10.6342/NTU202402157

doi:10.6342/NTU202402157

Denotation

MVV	Multi-view video
DIBR	Depth-image-based-rendering
D2D	Device-to-device
SDVA	Synthesis and D2D-assisted View Allocation
VAMS	View Allocation with Maximized Social welfare
VDME	View Delivery Mode Evaluation
CDLC	Cellular-D2D Link Coupling
TCATB	Transmission and Channel Allocation with Truthful Bidding
BIP	Binary Integer Programming
TCATB-P	Transmission and Channel Allocation with Truthful Bidding in Polynomial time

doi:10.6342/NTU202402157

doi:10.6342/NTU202402157

Chapter 1 Introduction

Recently, multi-view video (MVV) has attracted substantial interest within both research and market communities [8–11, 18]. MVV is produced by simultaneously capturing a scene of interest from various angles using multiple closely spaced cameras. Each camera captures both texture maps and depth maps to form distinct views, giving rise to innovative applications such as 3D television (3DTV) and free-viewpoint video (FVV) [8]. However, streaming MVVs requires significantly higher bandwidth than traditional single-view videos [10].

To reduce bandwidth consumption, Depth-Image-Based Rendering (DIBR) allows the synthesis of a specific view by its adjacent left and right reference views [5]. For example, view 4 could be synthesized from view 3 and view 5. With this technique, fewer views are required to satisfy the requests of viewers. While reducing bandwidth consumption, the computation cost incurred by DIBR synthesis is sufficiently small for modern devices [7], enhancing its appeal for MVV distribution. Xu et al. [18] proposed an MVV transmission scheme in multi-user wireless networks, leveraging both natural and view synthesis-enabled multicast opportunities for utility maximization. A price-based 3D viewing range control was designed in [8] with DIBR over an MVV network to meet the tremendous bandwidth requirements in a resource-limited environment. Yeh et al. [19] introduced an adaptive view selection methodology utilizing mobile proxy caching to fur-

doi:10.6342/NTU202402157

ther refine the efficiency of MVV transmission. However, the above works did not explore the potentials of D2D in improving view synthesis viability in multi-user scenarios.

Device-to-device (D2D) communications have been envisaged as a viable approach for facilitating the exchange of local information without adding extra load to the network infrastructure [3]. The technique leverages the proximity of devices to directly exchange data without routing through the cellular base station or core network, thereby offering a promising solution to alleviate network congestion and cope with the rapid surge in video traffic [1]. The integration of D2D communication has been explored to further enhance channel efficiency and reduce latency in video delivery. In [20] and [4], the authors employed D2D network cooperation to enhance real-time communication. In particular, Zammit et al. [20] adopted a D2D overlay network and proposed an algorithm to switch between cellular and D2D link usage. Bhardwaj et al. [4] proposed an interference-aware D2D-multicast session provisioning, aiming to improve communication reliability and efficiency in the face of potential signal interference. However, the above research has not explored the opportunity of leveraging DIBR synthesis in D2D communication to further improve transmission efficiency.

In this paper, we introduce a new optimization problem, named Synthesis and D2D-assisted View Allocation (SDVA), to significantly enhance the bandwidth efficiency and video quality of MVV delivery. Nevertheless, the complex interplay between the D2D communication and DIBR views synthesis introduces new research challenges as follows.

Tradeoff Between Bandwidth Efficiency and Viewer Satisfaction

To reduce bandwidth consumption of MVV transmission, requested views can be synthesized from reference views with DIBR synthesis, instead of directly sending the

doi:10.6342/NTU202402157

desired views. However, due to synthesis distortion, DIBR slightly degrades the visual quality of the views and therefore lowers viewer satisfaction. On the other hand, although D2D communication alleviates network congestion by cooperatively enhancing transmission capacity, excessive dependence on D2D communication sometimes leads to significant interference among viewer devices. High interference adversely undermines the effective data rate of the viewers' devices and degrades their viewing experience. Therefore, the inherent trade-off between maximizing bandwidth efficiency and maintaining viewer satisfaction demands a comprehensive resource allocation strategy that capitalizes on the advantages provided by DIBR and D2D, while mitigating the negative impacts of synthesis distortion and channel interference.

Extended Synthesis Possibility due to D2D

In MVV delivery systems, the requested views can be synthesized at the server or the viewer devices [18], leading to two types of view delivery modes for each viewer. In this paper, we further extend the synthesis possibility by integrating D2D transmission into the system, thereby having BS, viewer devices, and D2D relay devices because previous work [7] has demonstrated that mobile devices can efficiently synthesize desired views due to the rapid improvement of GPU recently. Utilizing D2D not only diversifies the sources of transmission but also enhances the possibility for synthesis, thus creating a variety of new delivery modes (detailed in Section 2). Although more delivery modes offer substantial opportunities for improving resource allocation, it also introduces a more complicated combinatorial solution space. For instance, a requested view can be synthesized with two reference views from two D2D transmitters, allowing BS to focus on viewers with high satisfaction demand. However, the quality of the requested view degrades due to view synthesis and D2D communication interference. Consequently, a promising algorithm is

doi:10.6342/NTU202402157

needed to efficiently find proper delivery modes for each viewer considering bandwidth efficiency and viewer satisfaction.

To address the above challenges, we proposed View Allocation with Maximized Social welfare (VAMS), an auction game-based algorithm with a 1/2-approximation ratio to find the best MVV delivery scheme for all viewers with the following phases. First, View Delivery Mode Evaluation (VDME) identifies all potential delivery modes that fulfill all view requests, specifies each delivery with corresponding viewers and D2D transmitters, and derives the induced cost. Next, with the information obtained in VDME, Cellular-D2D Link Coupling (CDLC) explores the possibility of enhancing network efficiency by coupling cellular links with D2D transmissions. It identifies feasible pairings to optimize the delivery scheme, by carefully examining the distance between viewers and the effective data rate. Finally, Transmission and Channel Allocation with Truthful Bidding (TCATB) finds an efficient allocation rule to provide a 1/2-approximation ratio compared to the optimal solution.

The rest of this paper is organized as follows. Chapter 2 introduces the MVV system and Chapter 3 formulates SDVA. Chapter 4 presents the auction game-based algorithm VAMS and proves its approximation ratio. Chapter 5 shows the simulation results, and Chapter 6 concludes this paper.

doi:10.6342/NTU202402157

Chapter 2 System Model

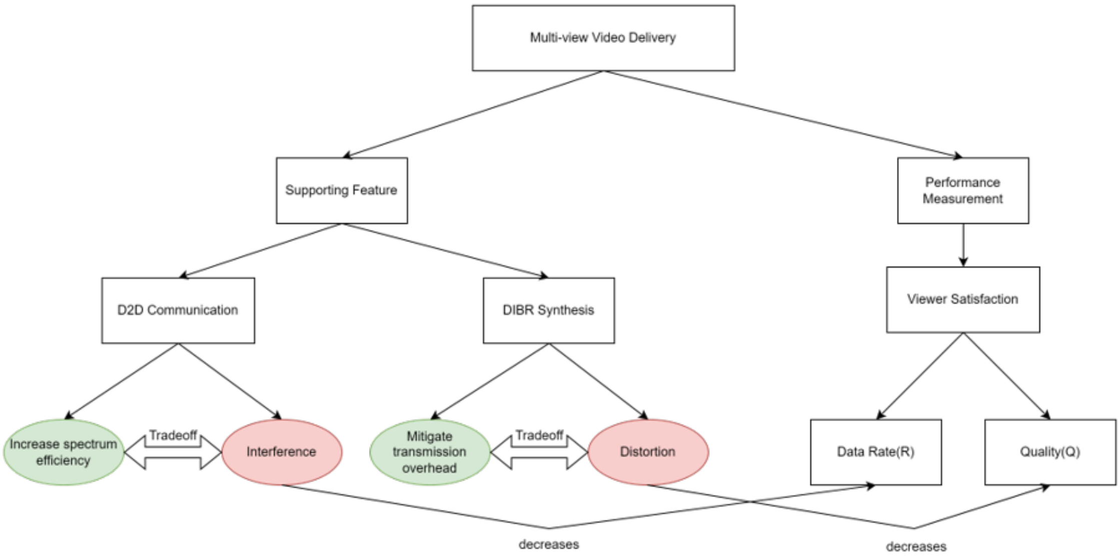

Figure 2.1: System Model

2.1 Network

In this paper, we consider an MVV delivery scenario consisting of one base station and a group of N viewers denoted by the set $\mathcal{N} = \{1, 2, ..., N\}$ with D D2D devices denoted by $\mathcal{D} = \{1, 2, ..., D\}$. To effectively manage the fluctuation of D2D devices, virtual clustering is introduced to divide the cellular network into uniformly sized squares or hexagons [1].

Therefore, as illustrated in Fig. 2.2, we employ square-size clustering and divide the

doi:10.6342/NTU202402157

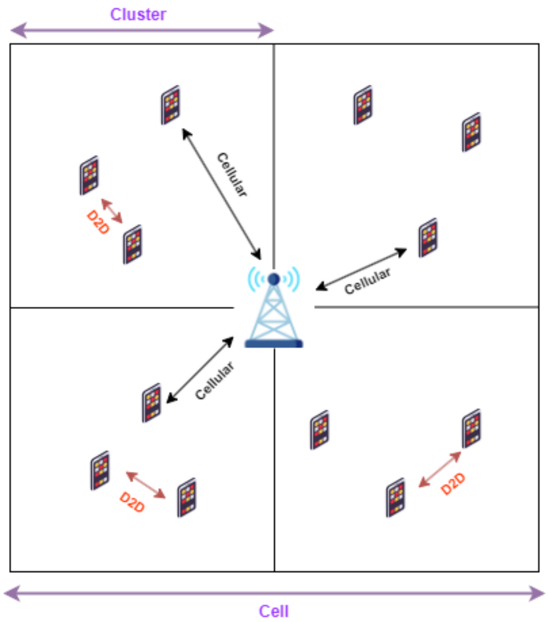

Figure 2.2: D2D with square-size clustering

cell into four equal-sized divisions. The clustering enables user devices situated within the same square (cluster) to engage in direct communication and file sharing through D2D links.

When a viewer requests a view cached in a D2D transmitter within the same cluster, and the distance between these devices does not exceed a predefined threshold [1], denoted as L, the requested view can be directly transmitted via a D2D link. The threshold L serves as a measure of collaborative distance and is typically determined according to the devices' transmission power [1]. This mechanism ensures efficient utilization of the D2D caching strategy by adapting to the spatial distribution of viewers within the network.

Unlike previous research [20] employing overlay D2D communication, where the spectrum is partitioned into non-overlapping segments for D2D and cellular transmissions, we focus on underlay D2D communication [1]. Underlay D2D communication leverages a shared spectrum for both D2D and cellular communication to improve spectrum efficiency at the cost of resultant co-channel interference. Therefore, appropriate management of

doi:10.6342/NTU202402157

D2D and cellular communication is necessary for ensuring communication integrity.

We define the set of cellular users and D2D pairs as $\mathcal{C}$ and $\mathcal{D}_p$, respectively. The BS transmits signal to the c-th cellular user, and the d-th D2D pair uses the same spectrum resources. Here, the D2D receiver receives interference from the BS. Meanwhile the cellular receiver c is exposed to interference from the D2D transmitter. Following [17], the received channel rates at c and d are derived as $R_c = \log_2\left(1 + \frac{P_B h_{Bc}}{\sum_d \beta_{cd} P_d h_{dc} + \sigma^2}\right)$ and $R_d = \log_2\left(1 + \frac{P_d h_{dd}}{P_B h_{Bd} + \sum_{d'} \beta_{dd'} P_{d'} h_{d'd} + \sigma^2}\right)$, where P_B, P_d and $P_{d'}$ denote the transmit power of the base station and D2D transmitter d, d'. β_{cd} is an indicator for channel sharing, where β_{cd} = 1 if cellular user c shares with D2D device d, β_{cd} = 0 otherwise. h_{ij} denotes the channel gain between device i and j. σ^2 accounts for the power of the additive white Gaussian noise at each receiver. When there is no resource sharing, the received signals at c is $R_{c0} = \log_2\left(1 + \frac{P_B h_{Bc}}{\sigma^2}\right)$ [17].

2.2 Multi-view video

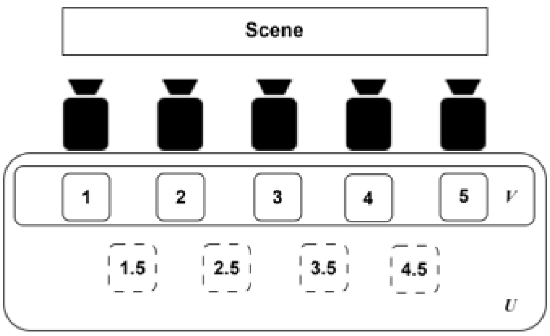

Figure 2.3: Multi-view video with V=5, W=2

In this paper, the multi-view video system captures V distinct views, incorporating the texture and depth maps, through equidistantly spaced cameras [8]. These captures, known as original views, are defined within the set $\mathcal{V} = \{1, 2, ..., V\}$, and cached at the server. In addition to the original views, we follow [18] to integrate $W - 1$ evenly spaced

doi:10.6342/NTU202402157

additional views between two consecutive original views v and $v+1$, where $W > 1$ and $v \in \{1, 2, ..., V-1\}$. Consequently, the interval between any two neighboring views is $1/W$. The additional views can be synthesized via DIBR synthesis. The comprehensive index set for all views, encompassing both the V original views which are stored at the server and the additional $(V-1)(W-1)$ views, available for synthesis but not pre-stored, is represented by $\mathcal{U} = \{1, 1+1/W, 1+2/W, ..., V\}$. Fig. 2.3 demonstrates an example with $V = 5$ and $W = 2$. Following [8], each view has the same bitrate r to support one reference view with its texture image depth map. Hence, the bandwidth cost for supporting one view is $b(v) = (1+\alpha_v)r$, where α_v is a 0-1 indicator of whether v requires synthesis.

2.3 DIBR Synthesis

DIBR synthesis allows the creation of a view by utilizing a pair of reference views, one from the left (V_L) and one from the right (V_R) [5]. The server may be required to generate any additional view $v \in \mathcal{U} \setminus \mathcal{V}$ by leveraging the closest left and right integer-indexed views as it caches only the original views. Similarly, a viewer i might need to synthesize any view $v \in \mathcal{V} \setminus \{1, V\}$, using two views from the left reference view set $\mathcal{U}_v^- = \{x \in \mathcal{U} : v - \Delta_{DIBR} < x < v\}$ and the right reference view set $\mathcal{U}_v^+ = \{x \in \mathcal{U} : v < x < v + \Delta_{DIBR}\}$ [18]. Here, Δ_{DIBR} denotes the maximum permissible distance between a synthesized view and its reference views.

The reception of requested views by a viewer is facilitated through two primary sources: directly from the server or via D2D links from a D2D transmitter. With the characteristics of DIBR, the desired view can be synthesized from V_L and V_R, either at the transmission point or the reception point. Hence, the content delivery of every requested

doi:10.6342/NTU202402157

view by a viewer comes in the following modalities:

- Mode A: Receive the exact view from the server.
- Mode B: Receive a server-synthesized view.
- Mode C: Receive the exact view from a D2D transmitter.
- Mode D: Receive a D2D transmitter-synthesized view.
- Mode E: Receive views V_L and V_R from two D2D transmitters, followed by local synthesis.
- Mode F: Receive one view from a D2D transmitter, followed by local synthesis with another cached view.

The quality of the synthesized view, denoted as Q, depends on the distance between the left and right views $\Delta = |V_L - V_R|$ and their original qualities. Following [7] and [15], the synthesis quality can be calculated by the average quality of left and right reference views and the distortion level caused by Δ, interpreting the dependency of synthesized view quality on both the spatial configuration of reference viewpoints and their intrinsic qualities.

2.4 Viewer satisfaction model

In multimedia systems, assessing viewer satisfaction emerges as a pivotal metric [6]. Single factors such as video bitrate exhibit a logarithmic relationship with viewer satisfaction [2], suggesting a diminishing return on satisfaction as additional resources are

doi:10.6342/NTU202402157

allocated, which eventually reaches a point of saturation. This insight suggests that beyond a certain threshold, additional enhancements in single factors yield minimal gains in viewer satisfaction [8]. A viewer's private valuation of a viewing experience is unknown to the server and other viewers since the viewer's personal preferences are not revealed and could be very different among viewers, which is known as the preference uncertainty factor [13].

To address the uncertainty, the viewer satisfaction model consists of view quality and bitrate [14], where a viewer is considered satisfied when both view quality and bitrate reach certain thresholds [13].

Specifically, the function $S_i(Q, R)$ assigns a positive value to a viewing experience when both the view quality (Q) and bitrate (R) of the received view meet or exceed predefined minimum acceptable thresholds ($Q_{i,min}$ [14] and $R_{i,min}$ [18]) specified by viewer i. Conversely, the satisfaction score defaults to 0 when the thresholds are not met. Therefore, the satisfaction of viewer i is defined as

$$S_i(Q, R) = \begin{cases} s_i, & \text{if } Q \geq Q_{i,min}, R \geq R_{i,min}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.1}$$

where the value of s_i is private to viewer i and not disclosed to others by default.

doi:10.6342/NTU202402157

Chapter 3 Problem Formulation

Unlike Stackelberg games [8] which use a leader-follower structure and coalitional games [21] that require cooperation between players, auction games do not explicitly require each viewer's private valuation of a viewing experience [13]. Therefore, we leverage the auction game framework to effectively solve SDVA, since a viewer's personal preferences may not be revealed and could be very different compared to others.

Leveraging the principles of mechanism design and auction theory, this paper aims to design an auction mechanism that achieves socially desirable outcomes under the constraints of individual self-interest and asymmetric information. With Myerson's Lemma [13], we focus on constructing an auction that aligns individual incentives with utility objectives and reaching a Nash Equilibrium among participants. Myerson's Lemma highlights the significance of designing allocation and payment rules that encourage truthful bidding, thereby revealing private valuations (satisfaction s_i), and optimizing the allocation of resources—here, the transmission modes and views. Hence, we model SDVA as a sealed-bid auction.

The enhanced sealed-bid auction model in this paper is called an SDVA auction, in which viewers, as bidders, submit their bids in secrecy to the server, as the auctioneer. The auction encapsulates not only an allocation rule, which determines the winners based

doi:10.6342/NTU202402157

on their bids but also a payment rule, which is guaranteed by Myerson's lemma to ensure the truthful bidding property. The essence of incorporating the lemma lies in optimizing these rules to align with the true satisfaction of the views by the viewers, thereby ensuring an efficient and social welfare-maximizing outcome.

The SDVA auction is performed as follows:

1. Each viewer i requests their desired views v_i (including minimum acceptable quality and data rate) and submits the respective bids x_i to the server.

2. Upon receiving the requests and bids, the server announces the allocation. This step involves determining which viewers secure their desired views while also optimizing social welfare within the BS's and D2D transmitters' bandwidth limitations.

3. According to Myerson's payment rule, the server then calculates the payments due from each winning bidder.

The problems of the server and the viewers in the auction are detailed as follows.

Bid Adjustment Problem of the viewers. Each viewer i in the problem acts as a bidder to maximize their utility, which is the private satisfaction deducted by the actual payment. This balance is captured by the viewer's bid adjustment problem, where a higher bid x_i may increase the chance of securing preferred views but incurs a greater payment, whereas a lower bid reduces payment at the risk of not winning the auction. This dilemma is formalized as

$$\max_{x_i > 0} U_i = s_i(x_i) - p_i(x_i), \tag{3.1}$$

where $s_i(x_i)$ and $p_i(x_i)$ are the viewer's private satisfaction and actual payment after the auction while bidding at x_i.

doi:10.6342/NTU202402157

Transmission Arrangement problem of the server. The server's challenge is to optimally allocate views to bidding viewers, deciding not only the winners but also to allocate transmitters and channel resources. The objective is to maximize revenue, which is the sum of the actual payments from the viewers, i.e.,

$$\max U_s = \sum_{i=1}^{N} p_i(x_i). \tag{3.2}$$

SDVA has the following constraints.

1) Bandwidth Constraint. The cumulative bandwidth utilized by cellular links for MVV transmission must not surpass the base station's maximum bandwidth capacity (B), i.e.,

$$\sum_{v \in d, d \in D_{CE}} b(v) \leq B \tag{3.3}$$

2) D2D Capacity Constraint. Considering the limited capacity of D2D devices, each D2D transmitter is restricted to delivering only one view per round, i.e.,

$$\sum_{v \in d, d \in D_{D2D}} b(v) \leq r \tag{3.4}$$

3) DIBR Synthesis Constraint. For a synthesized view to be considered valid, the spatial separation between the left and right reference views must not exceed the maximum allowed distance Δ_{DIBR}, i.e.,

$$\Delta_{ij} < \Delta_{DIBR}, \forall i, j \in \mathcal{U} \tag{3.5}$$

doi:10.6342/NTU202402157

4) D2D Proximity Constraint. The establishment of a D2D communication link requires the participating devices to be within a predefined distance limit (L), i.e.,

$$dist(d, d') < L, \forall (d, d') \in \mathcal{D}_p \tag{3.6}$$

5) Channel Sharing Constraint. A cellular link is limited in its capacity to share its channel resources with one D2D transmitter, preserving the quality of service for cellular users while accommodating D2D enhancements, i.e.,

$$\sum_{d} \beta_{cd} \leq 1, \forall c \in \mathcal{C} \tag{3.7}$$

Definition 1. SDVA Given the set of views $\mathcal{U}$, set of viewers $\mathcal{N}$, set of D2D devices $\mathcal{D}$, synthesis parameters (Δ_{DIBR}, V, W), channel information $(P_b, P_c, P_d, \sigma, h, \beta)$, bidding information of viewers $(v_i, x_i, Q_{i,min}, R_{i,min})$, viewer and D2D device caches C_i, C_D, SDVA finds

1. the transmitted views and their quality
2. the participating D2D transmitters
3. the channel allocation among all links

under Bandwidth, D2D capacity, DIBR synthesis, D2D proximity, and Channel sharing constraints. The objective is to reach a strategic equilibrium that maximizes social welfare, which is the sum of the server's and the viewers' utilities.

For example, three viewers $\mathcal{N} = \{1, 2, 3\}$ express interest in acquiring views from an MVV with views $\{1, 2, \ldots, 5\}$, 1 additional view between integer-indexed views, and

doi:10.6342/NTU202402157

maximum synthesis distance of 2 views, making $V = \{1, 2, 3, 4, 5\}$, $W = 2$, $U = \{1, 1.5, 2, 2.5, \ldots, 5\}$, and $\Delta_{DIBR} = 2$. There are 2 devices available to serve as D2D transmitters $\mathcal{D} = \{1, 2\}$ with cached views $C_{D1} = \{2.5, 5\}$ and $C_{D2} = \{4, 5\}$ respectively. Assume viewer 1 requests view 4 and viewer 2 requests view 4.5, and viewer 3 requests view 1, the viewers would submit corresponding information $(v_i, x_i, Q_{i,min}, R_{i,min})$ to the server, forming the SDVA problem.

Theorem 1. SDVA is NP-hard.

Proof. The proof of Theorem 1 is presented in appendix B.3. □

doi:10.6342/NTU202402157

Chapter 4 Auction game analysis

To solve SDVA, one naive way is to directly apply Vickrey Auction [16], also known as the second-price auction. However, the bidding items in SDVA include a combination of transmission devices and channel resources. A simple allocation rule that makes the highest bidders win may provide infeasible solutions since it focuses only on the views, neglecting potential conflicts among winning bidders due to underlying channel resources. Hence, we propose VAMS to effectively address the problem.

VAMS unfolds in three phases: 1) View Delivery Mode Evaluation (VDME), 2) Cellular-D2D Link Coupling (CDLC), and 3) Transmission and Channel Allocation with Truthful Bidding (TCATB). First, View Delivery Mode Evaluation (VDME) identifies all potential delivery modes that fulfill all view requests, specifies each delivery with corresponding viewers and D2D transmitters, and derives the induced cost. Next, with the information obtained in VDME, Cellular-D2D Link Coupling (CDLC) explores the possibility of enhancing network efficiency by coupling cellular links with D2D transmissions. It identifies feasible pairings to optimize the delivery scheme, considering the distance between viewers and the effective data rate. Finally, Transmission and Channel Allocation with Truthful Bidding (TCATB) finds the optimal allocation rule while TCATB-P is the polynomial time version of TCATB that finds an efficient allocation rule to provide a 1/2-approximation ratio compared to the optimal solution.

doi:10.6342/NTU202402157

4.1 View Delivery Mode Evaluation (VDME)

VDME generates the set containing all possible delivery scenarios that satisfy each viewer. For each viewer, we first find out the possible transmissions according to the six different delivery modes. Then the view quality and cost of each transmission will be calculated. Lastly, the delivery scenario set will be screened based on the minimum quality constraint of each viewer. Deliveries that fail to meet the requested minimum quality will not be added to the resulting set.

For example, two viewers express interest in acquiring views from a MVV with views $\{1, 2, \ldots, 5\}$ and 1 additional view between integer-indexed views, making $\mathcal{V} = \{1, 2, 3, 4, 5\}$ and $\mathcal{U} = \{1, 1.5, 2, 2.5, \ldots, 5\}$. There are 2 devices available to serve as D2D transmitters with cached views $\{2.5, 5\}$ and $\{4, 5\}$ respectively. Viewer 1 requests view 4 and viewer 2 requests view 4.5. In VDME, viewer 1 could be satisfied through Mode A (view 4 delivered from the BS), Mode C (view 4 delivered by D2D transmitter 2), and Mode D (view 4 synthesized from view 2.5 and 5, delivered from D2D transmitter 1). Note that Mode B is not considered here since it is dominated by Mode A, which saves the cost of DIBR synthesis. As for viewer 2, the request can be satisfied by Mode B (synthesized view 4.5 delivered from the BS), and two possible Mode Ds (synthesized view 4.5 from D2D transmitter 1 and from D2D transmitter 2). The quality constraint will filter each delivery mode. Take viewer 1 as an example, Q_A from Mode A, Q_C from Mode C, and the synthesized Q_D from Mode D must be greater than $Q_{1,min}$, otherwise deemed infeasible. In traditional delivery modes that are merely assisted with DIBR, the only way to deliver the views is Mode A for viewer 1 and Mode B for viewer 2. The pseudocode of VDME is presented in appendix A.1.

doi:10.6342/NTU202402157

4.2 Cellular-D2D Link Coupling (CDLC)

With the help of D2D channel resource sharing, a cellular link can share its channel with a D2D link at the cost of effective data rate. Hence, the sharing of channels is limited by viewers' minimum requested data rate and cannot be exhaustively extended. From the sets of delivery mode (D_{CE} and D_{D2D}) obtained in VDME, CDLC discovers all sharing possibilities of a cellular link with a D2D link, couples them to form link pairs, and screens the discovered pairs by the involved viewers' minimum data rate requirement. Finally generating a set LP of effective link pairs such that each link pair satisfies the minimum requirement of both quality and data rate of its covered viewers.

Extending from the example provided in VDME, if viewer 1 and 2 are in the same D2D cluster, their cellular and D2D links can be coupled for channel sharing. There are two D2D delivery modes for viewer 1, Mode C and Mode D respectively, which can be coupled with the cellular link for viewer 2, which is the Mode B delivery. On the other hand, there are 2 possible D2D links for viewer 2 and the two Mode D deliveries can be coupled with the Mode A delivery of viewer 1. The data rate constraint will filter each link pair. The calculated effective data rate after interference for viewer 1 (R_{c1} or R_{d1}) and viewer 2 (R_{c2} or R_{d2}) must be greater than $R_{1,min}$ and $R_{2,min}$, otherwise deemed infeasible. If the BS capacity is currently limited and can only support delivering 1 view, both viewers could still be satisfied through CDLC, while in traditional delivery modes where D2D cooperation is unavailable, only one of the viewers could be satisfied. The pseudocode of CDLC is presented in appendix A.2.

doi:10.6342/NTU202402157

4.3 Transmission and Channel Allocation with Truthful Bidding(TCATB)

We design TCATB to solve SDVA with maximized social welfare optimally. The input of TCATB is the set of all possible link pairs with corresponding social welfare generated by CDLC. The link pair selection process is then formulated into a BIP problem. The maximize objective is the total utility of the selected link pairs $\sum_{p \in P} u_p \cdot x_p$, where x_p is the 0-1 decision variable indicating whether link pair p is selected and u_p is the utility of p. Next, the constraints of the BIP problem are generated according to the Bandwidth and D2D capacity constraints. To solve the BIP problem, the mathematical tool PuLP is utilized to decide the values of the decision variables and thus reach an optimal solution. Though TCATB algorithm can optimally solve SDVA the worst-case execution time complexity is exponential. The pseudocode of TCATB is presented in appendix A.3.

Lemma 1. A link pair p in LP only comes in one of the following forms. 1. $p = (\{i\}, \{T_0\})$, 2. $p = (\{i, j\}, \{T_0, x\})$, 3. $p = (\{i, j, k\}, \{T_0, T_0, x, y\})$, where $i, j, k \in \mathcal{N}, x, y \in \mathcal{D}, i \neq j, j \neq k, i \neq k, x \neq y, T_0$ is the base station.

Proof. The proof of Lemma 1 is presented in appendix B.1. □

Lemma 2. There must be one and only one type 1 link pair $p = (\{i\}, \{T_0\})$ for each viewer $i \in \mathcal{N}$.

Proof. The proof of Lemma 2 is presented in appendix B.2. □

doi:10.6342/NTU202402157

4.4 Transmission and Channel Allocation with Truthful Bidding in Polynomial time (TCATB-P)

To solve the high time complexity of TCATB algorithm, we design TCATB-P, which runs in polynomial time, to solve SDVA with an approximation bound to the optimal solution. TCATB-P takes the same input set as TCATB generated by CDLC. Step 1 first sorts all viewers descendingly based on their type 1 utility, establishing a prioritized sequence. Step 2 involves allocating the maximum number of views (k) that the base station can transmit in a single round. According to the bandwidth constraint, the base station allocates the first k viewers with their type 1 link pair. At this point, the bandwidth consumption of the base station reaches its maximum. In Step 3, TCATB explores merging possibilities for viewers beyond the initial allocation, trying from $k + 1$ to see whether possible merging replacements exist to improve the total utility, incorporating viewers from the existing set and new ones. The choice of merging is determined by minimizing conflicts with other link pairs in the sequence, ultimately replacing original type 1 link pairs. Step 4 iterates through the merging process until the last viewer N is reached.

For example, there are five viewers $(1, \ldots, 5)$ in the scenario and the BS has a capacity $k = 2$. After VDME and CDLC, the set of possible link pairs (LP) is generated. Assume the viewers are already in descending order of type 1 link pair utilities. After the first step of TCATB, viewer 1 and viewer 2 are allocated through their type 1 link pair. Next, starting from viewer 3, TCATB looks for possible merging possibilities in LP between viewer 3 and the allocated viewers, currently viewer 1 and viewer 2. If multiple merging opportunities are found, the one with the least conflict with other link pairs in LP would be chosen. Assuming we found a merging opportunity with viewer 1, we added the

doi:10.6342/NTU202402157

link pair containing viewer 1 and 3 into the allocated subset, replacing the original type 1 link pair of viewer 1. Note that after this step, viewer 1 becomes unavailable for further merging. Hence, while moving to viewer 4, we only look for possible merging opportunities with viewer 2. The algorithm terminates after reaching viewer 5. Finally, TCATB gives the allocation result, containing the allocated viewers and the participating transmitters combined in the link pairs. Throughout the merging process, the solution subset always remains feasible. The pseudocode of TCATB-P is presented in appendix A.4.

Theorem 2. TCATB-P guarantees truthful bidding.

Proof. The proof of Theorem 2 is presented in appendix B.4. □

Theorem 3. TCATB-P is 1/2-approximate to the optimal solution.

Proof. The proof of Theorem 3 is presented in appendix B.5 □

Theorem 4. SDVA with TCATB-P runs in $O(N^2D^2)$ time.

Proof. The proof of Theorem 4 is presented in appendix B.6 □

doi:10.6342/NTU202402157

Chapter 5 Simulations

We evaluate the performance of TCATB and TCATB-P with multi-view videos and the corresponding viewer request scenarios in [18], where the viewers request views in an i.i.d manner. We employ the cellular layout from [1] with an isolated cell and 4 equal square sectors. The D2D transmission parameters follow those in [17, 22], where the transmission power of the base station and D2D devices is set to 46 dBm and 23 dBm, respectively. The channel loss model follows [22]. For synthesized view quality evaluation, we follow the technique in [7] and [15]. Following [19], the default DIBR constraint is set to 3. We compare TCATB and TCATB-P with three baselines: 1) DIBR-DC [18], 2) D2D-CCE [22], and 3) MMSG [12], where DIBR-DC focuses on DIBR potential and view quality, D2D-CCE focuses on D2D and channel efficiency, and MMSG uses a Stackelberg game-based analysis. To evaluate TCATB and TCATB-P, we vary the following parameters: 1) number of viewers, 2) number of D2D transmitters, and 3) bandwidth capacity (measured by k, the maximum number of views the base station can transmit in a single round). We measure the following performance metrics: 1) total social welfare and 2) average view satisfaction. Each result is averaged over 100 experiments.

In Fig. 5.1, the total social welfare increases as the number of viewers grows because more viewers induce more possibility of D2D sharing, allowing BS to focus on serving viewers with high satisfaction requirements. TCATB and TCATB-P outperform the base-

doi:10.6342/NTU202402157

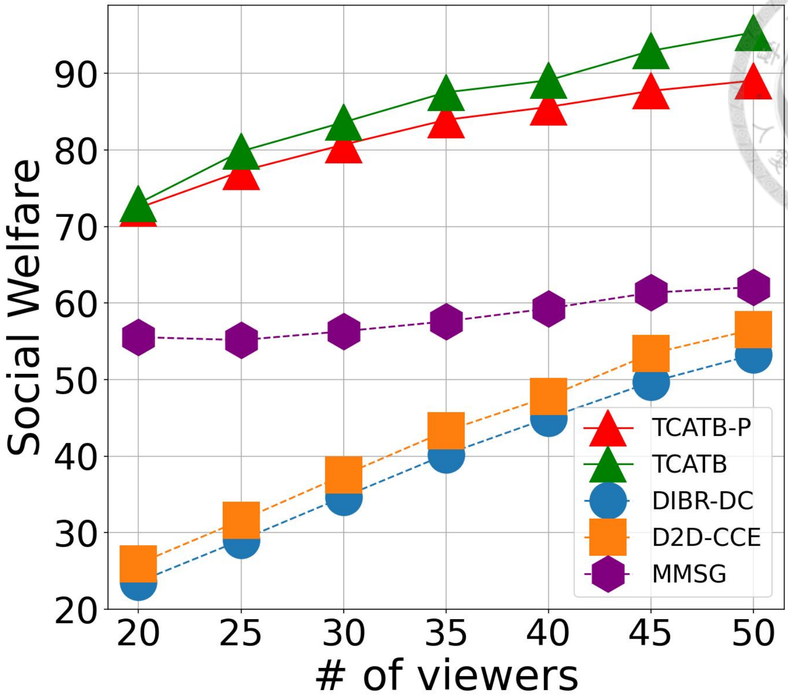

Figure 5.1: Social Welfare with different number of viewers

lines by effectively utilizing the benefits of DIBR and D2D enhancements and capitalizing on private information through its truthful bidding feature. In contrast, DIBR-DC and D2D-CCE only focus on either DIBR or D2D, resulting in a substantial performance discrepancy compared to TCATB and TCATB-P. On the other hand, MMSG shows a less pronounced social welfare increase due to the lack of complete information regarding viewer private satisfaction.

As for Fig. 5.2, the average view satisfaction of TCATB and TCATB-P grows along with the number of viewers, whereas the average view satisfaction of all baselines remains at similar levels. With adequate mitigation of distortion and interference in VDME and CDLC, TCATB and TCATB-P can effectively adapt to an increasing viewer base without compromising on the MVV viewing experience.

In Fig. 5.3, the total social welfare of TCATB and TCATB-P increases along with the number of D2D devices since TCATB and TCATB-P can choose from more view

doi:10.6342/NTU202402157

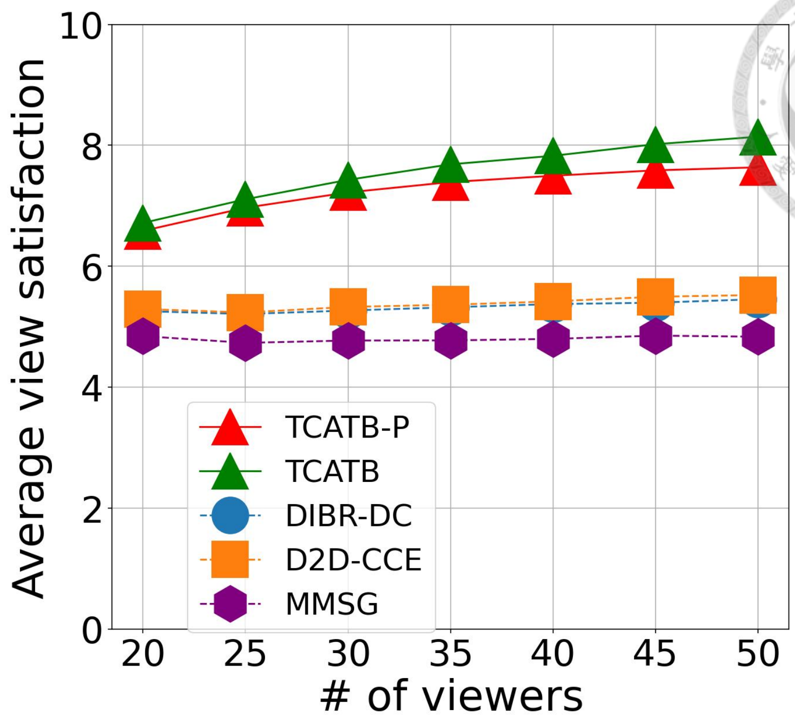

Figure 5.2: Average view satisfaction with different number of viewers

delivery modes for each viewer with the assistance of D2D communication. DIBR-DC remains unaffected by the number of D2D devices because it has no D2D features. As the number of D2D devices grows, D2D-CCE overtakes DIBR-DC, indicating the substantial impact that D2D devices have on total social welfare. Therefore, TCATB and TCATB-P can significantly surpass the baselines by properly integrating DIBR and D2D into the MVV delivery system.

In Fig. 5.4, TCATB and TCATB-P show significantly higher performance on average view satisfaction than the baselines, since it can find DIBR and D2D-assisted delivery modes that yield higher viewer satisfaction. On the other hand, the baselines can only obtain less efficient delivery modes, resulting in worse view satisfaction performance.

From Fig. 5.5, the growth in BS capacity causes the total social welfare to increase because more views can be transmitted when having more bandwidth. Meanwhile, increased BS capacity also induces more cellular links, thus increasing the possibility of

doi:10.6342/NTU202402157

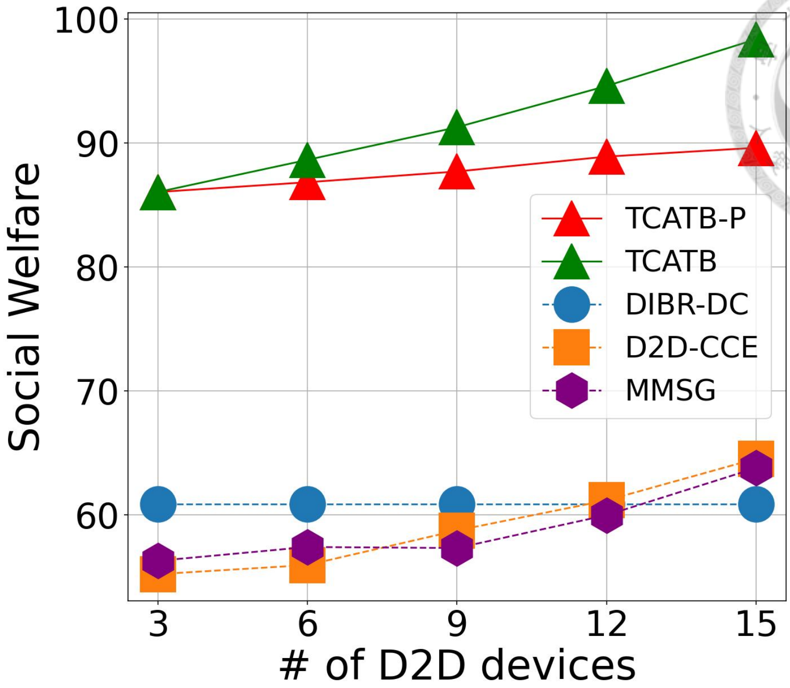

Figure 5.3: Social welfare with different number of D2D devices

D2D sharing. Benefiting from increased cellular and D2D links, TCATB and TCATB-P outperform the baselines by carefully evaluating the increased transmission options. The total social welfare of DIBR-DC and MMSG also increases along with the growing BS capacity but remains underperformed due to the lack of D2D communication. As for D2D-CCE, the increasing tendency of total social welfare eventually halts due to the lack of DIBR participation. Having a higher BS capacity may not always be beneficial, since the requested views are more likely to be transmitted with cellular links, leaving the D2D links useless.

In Fig. 5.6, when BS capacity is insufficient, TCATB and TCATB-P can identify more valuable view deliveries compared with the baselines, resulting in significantly higher average view satisfaction. The difference between TCATB and TCATB-P and the baselines slightly decreases because having sufficient bandwidth allows more viewers to be satisfied by direct cellular links.

doi:10.6342/NTU202402157

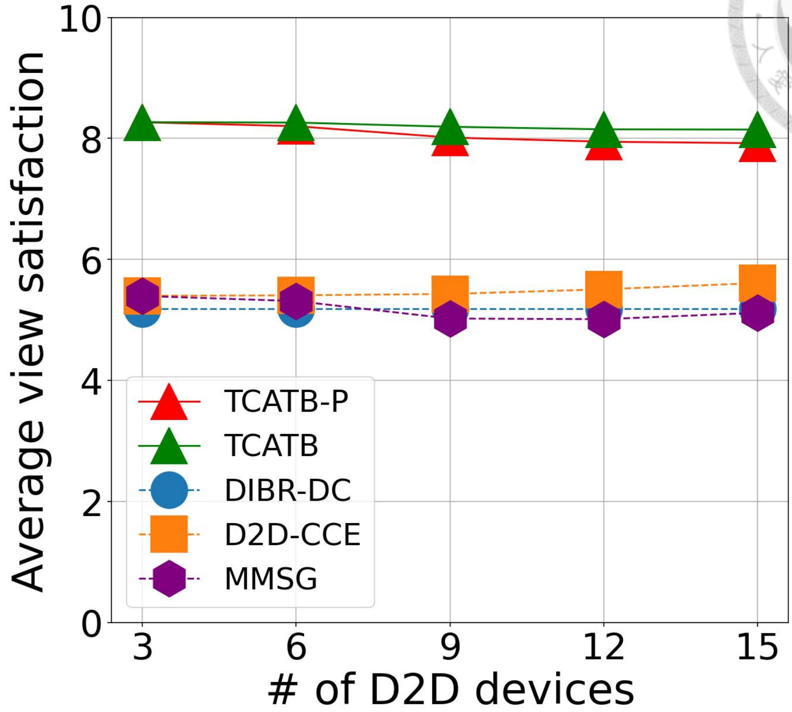

Figure 5.4: Average view satisfaction with different number of D2D devices

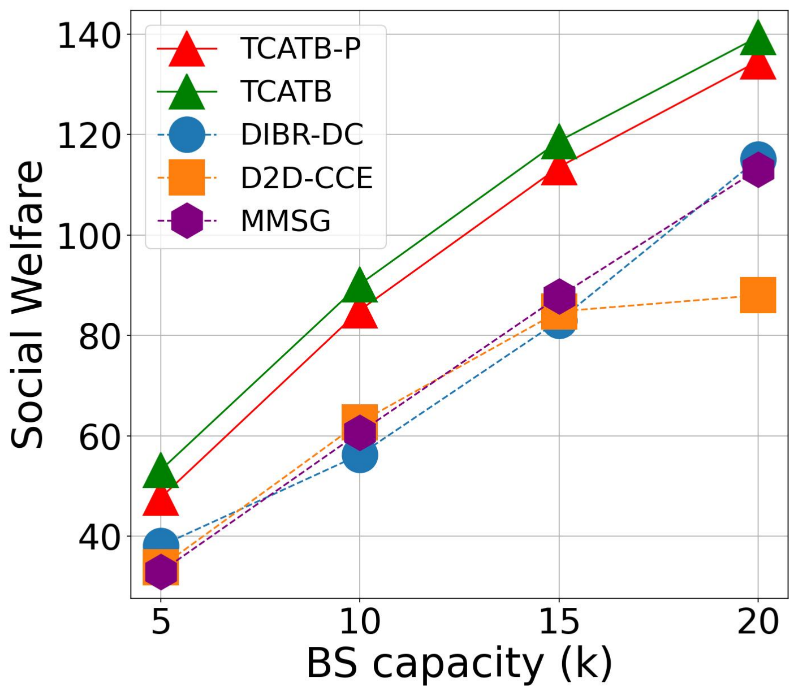

Figure 5.5: Social welfare with different BS capacity

doi:10.6342/NTU202402157

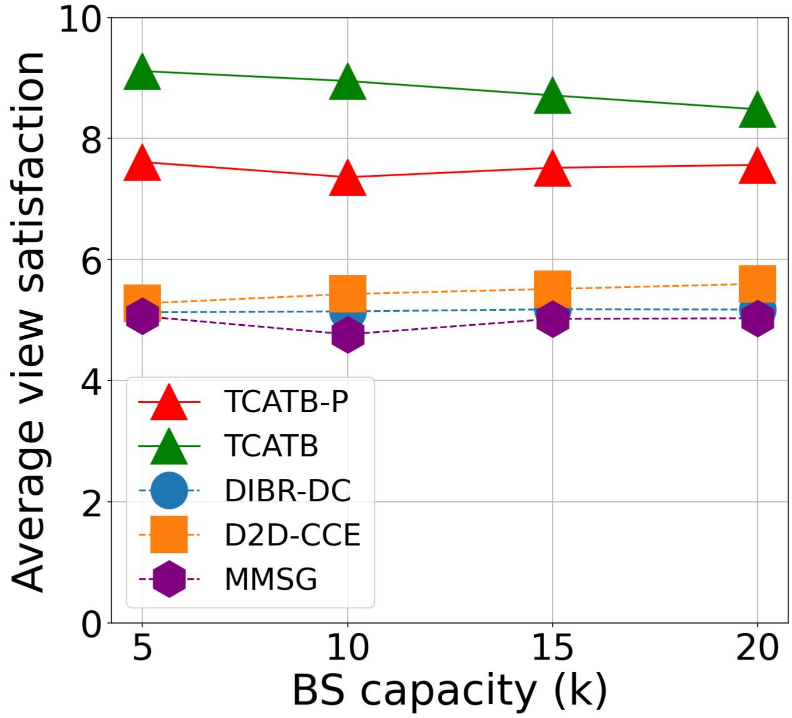

Figure 5.6: Average view satisfaction with different BS capacity

doi:10.6342/NTU202402157

Chapter 6 Conclusion

This paper presents the first attempt to maximize social welfare in an MVV streaming system by simultaneously leveraging DIBR synthesis and D2D communication. We formulate a new optimization problem, named SDVA, to maximize social welfare. To effectively solve SDVA, we establish a sealed-bid auction to investigate the transmitters and the viewers' strategic behaviors. For the auction, we proposed a polynomial-time algorithm TCATB-P, which is a monotone allocation rule with a 1/2-approximation ratio to the optimal solution. Numerical results with real network settings show the advantage of VAMS over state-of-the-art algorithms and demonstrate the advantages of utilizing both DIBR synthesis and D2D communication. In the future, we plan to investigate the highlighted problem with other synthesis technologies in the context of 6G networks.

doi:10.6342/NTU202402157

doi:10.6342/NTU202402157

References

[1] I. Ahmed, M. H. Ismail, and M. S. Hassan. Video transmission using device-to-device communications: A survey. IEEE Access, 7:131019–131038, 2019.

[2] M. Andrews, J. Cao, and J. McGowan. Measuring human satisfaction in data networks. In IEEE INFOCOM 2006., pages 1–12, 2006.

[3] S. Baidya and M. Levorato. Content-based interference management for video transmission in d2d communications underlaying lte. In 2017 International Conference on Computing, Networking and Communications (ICNC), pages 144–149, 2017.

[4] A. Bhardwaj and S. Agnihotri. Interference-aware d2d-multicast session provisioning in lte-a networks. In 2017 IEEE Wireless Communications and Networking Conference (WCNC), pages 1–6, 2017.

[5] C. Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In A. J. Woods, J. O. Merritt, S. A. Benton, and M. T. Bolas, editors, Stereoscopic Displays and Virtual Reality Systems XI, volume 5291 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pages 93–104, May 2004.

[6] S. R. Gulliver and G. Ghinea. Defining user perception of distributed multimedia quality. ACM Trans. Multimedia Comput. Commun. Appl., 2(4):241–257, nov 2006.

doi:10.6342/NTU202402157

[7] A. Hamza and M. Hefeeda. Energy-efficient multicasting of multiview 3d videos to mobile devices. ACM Trans. Multimedia Comput. Commun. Appl., 8(3s), oct 2012.

[8] T.-Y. Ho, D.-N. Yang, and W. Liao. Efficient resource allocation of mobile multi-view 3d videos with depth-image-based rendering. IEEE Transactions on Mobile Computing, 14(2):344–357, 2015.

[9] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang. Multiview imaging and 3dtv. IEEE Signal Processing Magazine, 24(6):10–21, 2007.

[10] J.-J. Kuo, D.-N. Yang, W.-C. Li, and W.-T. Chen. Efficient multi-view 3d video multicast with mobile edge computing. In 2018 IEEE Global Communications Conference (GLOBECOM), pages 1–7, 2018.

[11] J.-T. Lee, D.-N. Yang, Y.-C. Chen, and W. Liao. Efficient multi-view 3d video multicast with depth-image-based rendering in lte-advanced networks with carrier aggregation. IEEE Transactions on Mobile Computing, 17(1):85–98, 2018.

[12] T. Lyu, H. Xu, and Z. Han. Multi-leader multi-follower stackelberg game based resource allocation in multi-access edge computing. In IEEE ICC 2022, pages 4306–4311, 2022.

[13] R. B. Myerson. Optimal auction design. Mathematics of Operations Research, 6(1):58–73, 1981.

[14] S. Tian, L. Zhang, W. Zou, X. Li, T. Su, L. Morin, and O. Déforges. Quality assessment of dibr-synthesized views: An overview. 01 2021.

[15] L. Toni, N. Thomos, and P. Frossard. Interactive free viewpoint video streaming using prioritized network coding. In 2013 IEEE MMSP, pages 446–451, 2013.

doi:10.6342/NTU202402157

[16] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. The Journal of Finance, 16(1):8–37, 1961.

[17] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao. Efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach. IEEE JSAC, 31(9):348–358, 2013.

[18] W. Xu, Y. Cui, and Z. Liu. Optimal multi-view video transmission in multiuser wireless networks by exploiting natural and view synthesis-enabled multicast opportunities. IEEE Transactions on Communications, 68(3):1494–1507, 2020.

[19] M. Yeh, C.-H. Wang, D.-N. Yang, J.-T. Lee, and W. Liao. Mobile proxy caching for multi-view 3d videos with adaptive view selection. IEEE Transactions on Mobile Computing, 21(8):2909–2921, 2022.

[20] J. Zammit and C. J. Debono. Efficient transmission of live multi-view video in lte-a cellular networks using d2d cooperation. In 2020 IEEE MELECON, pages 169–174, 2020.

[21] T. Zhang, W. Chen, Z. Han, and Z. Cao. Coalitional game theoretic approach for cooperative transmission in vehicular networks. In 2013 IEEE International Conference on Communications (ICC), pages 6179–6183, 2013.

[22] S. Zhao, Y. Feng, and G. Yu. D2d communication channel allocation and resource optimization in 5g network based on game theory. Computer Communications, 169, 01 2021.

doi:10.6342/NTU202402157

doi:10.6342/NTU202402157

Appendix A — Pseudocode

A.1 VDME

Algorithm 1 View Delivery Mode Evaluation (VDME)

Require: Set of viewers $\mathcal{N}$, set of D2D transmitters $\mathcal{D}$, base station T_0

Ensure: Delivery sets D_{CE} and D_{D2D}

procedure VDME($\mathcal{N}, \mathcal{D}, T_0$)
    Initialize $D_{CE} \leftarrow []$, $D_{D2D} \leftarrow []$, $T \leftarrow [T_0, \mathcal{D}]$
    for each i in $\mathcal{N}$ **do**
        $minQ \leftarrow Q_{n,min}$
        if $v_n \in \mathcal{V}$ **then** ▷ Mode A
            $d_a \leftarrow ([i], [T_0], x_i)$
            Append d_a to D_{CE}
        else ▷ Mode B
            $d_b \leftarrow ([i], [T_0], x_i - cost_{syn})$
            if $Q(d_b) \geq minQ$ **then**
                Append d_b to D_{CE}
            end if
        end if
        Initialize $tmp_r \leftarrow []$, $tmp_l \leftarrow []$

doi:10.6342/NTU202402157

for each t in T **do** ▷ Mode C

  if $dist(i, t) < L$ **then**

    for each v_t in C_t **do**

      if $v_t = v_n$ and $Q(v_t) \geq minQ$ **then**

        $d_c \leftarrow$ new $([i], [t], x_i)$

        Append d_c to D_{D2D}

      else if $v_t > v_n$ **then**

        Append $[v_t, t]$ to tmp_r

      else

        Append $[v_t, t]$ to tmp_l

      end if

    end for

  end if

end for

for each pair l, r in $tmp_l \times tmp_r$ **do**

  if $t_l = t_r$ and $Q_{syn}(v_l, v_r) \geq minQ$ **then**

    $d_d \leftarrow ([i], [t_l], x_i - cost_{syn})$ ▷ Mode D

    Append d_d to D_{D2D}

  else if $Q_{syn}(v_l, v_r) \geq minQ$ **then**

    $d_e \leftarrow ([i], [t_r, t_l], x_i - cost_{syn})$ ▷ Mode E

    Append d_e to D_{D2D}

  end if

end for

for each v_i in C_i **do** ▷ Mode F

doi:10.6342/NTU202402157

if $v_i < v_n$ **then**
for each v_r in tmp_r **do**
if $Q_{syn}(v_i, v_r) \geq minQ$ **then**
$d_f \leftarrow ([i], [t_r], x_i - cost_{syn})$
Append d_f to D_{D2D}
end if
end for
else
for each v_l in tmp_l **do**
if $Q_{syn}(v_l, v_i) \geq minQ$ **then**
$d_f \leftarrow ([i], [t_l], x_i - cost_{syn})$
Append d_f to D_{D2D}
end if
end for
end if
end for
end for
return D_{CE}, D_{D2D}

doi:10.6342/NTU202402157

A.2 CDLC

Algorithm 2 Cellular and D2D Link Combination (CDLC)

Require: Sets D_{CE}, D_{D2D}

Ensure: Set of valid link pairs LP

procedure CDLC(D_{CE}, D_{D2D})
  Initialize $LP \leftarrow []$
  for each d_c in D_{CE} **do**
    Add $(i_{d_c}, t_{d_c}, U_{d_c})$ to LP
  end for
  for each d_d in D_{D2D} **do**
    if d_d is Mode E delivery **then**
      for d_{c1}, d_{c2} in $D_{CE} \times D_{CE}$ **do**
        Calculate R_{c1}, R_{d1}, R_{c2}, R_{d2}
        if $R_i > R_{i,min}$ for all devices **then**
          $I \leftarrow [i_{d_{c1}}, i_{d_{c2}}, i_{d_d}]$
          $T \leftarrow [t_{d_{c1}}, t_{d_{c2}}, t_{d_d}]$
          $U \leftarrow U_{d_{c1}} + U_{d_{c2}} + U_{d_d}$
          Add (I, T, U) to LP
        end if
      end for
    else
      for each d_c in D_{CE} **do**
        Calculate R_c and R_d
        if $R_c \geq R_{i_c,min}$ and $R_d \geq R_{i_d,min}$ **then**

doi:10.6342/NTU202402157

$I \leftarrow [i_c, i_d]$

$T \leftarrow [t_c, t_d]$

$U \leftarrow U_c + U_d$

Add (I, T, U) to LP

end if

end for

end if

end for

return LP

doi:10.6342/NTU202402157

A.3 TCATB

Algorithm 3 Transmission and Channel Allocation with Truthful Bidding (TCATB)

Input: The link pair set P of tuples (v, t, u) where each v is a set of viewers, t is a set of transmitters, and u is the utility. $BS(T_0)$ can deliver up to k views in one round.

Output: Optimal subset of P with maximized total utility.

procedure TCATB(P, V, T, k)

    Define binary decision variables x_p for each $p \in P$.

    Objective: Maximize $\sum_{p \in P} u_p \cdot x_p$.

    for all $v \in V$ **do**

        Add constraint: $\sum_{p \in P \text{ and } v \in p} x_p \leq 1$.

    end for

    Add constraint: $\sum_{p \in P \text{ and } T_0 \in p} (T_0 \text{ in } p) \cdot x_p \leq k$.

    for all $t \in T, t \neq T_0$ **do**

        Add constraint: $\sum_{p \in P \text{ and } t \in p} x_p \leq 1$.

    end for

    Solve the binary integer programming problem.

    return the selected subset and its total utility.

doi:10.6342/NTU202402157

A.4 TCATB-P

Algorithm 4 Transmission and Channel Allocation with Truthful Bidding in Polynomial time (TCATB-P)

Require: Set of link pairs LP, maximum capacity of BS k

Ensure: Allocated subset of P

procedure TCATB-P(LP)
    $LP_1 \leftarrow$ type 1 link pairs in LP
    $LP_{23} \leftarrow LP \setminus LP_1$
    Sort LP_1 by utility descendingly
    $LP_a \leftarrow$ first k link pairs in LP_1
    $N_u \leftarrow$ viewers yet allocated
    $N_a \leftarrow$ viewers allocated
    $tmp \leftarrow []$
    for each i in N_u **do** ▷ merging
        for each p in LP_{23} **do**
            if $i \in p$ and $I_p \setminus \{i\} \in N_a$ **then**
                Add p to tmp
            end if
        end for
        $p_m \leftarrow p \in tmp$ with least conflict
        $N_u \leftarrow N_u \setminus N_u \cap I_{p_m}$
        $N_a \leftarrow N_a \setminus N_a \cap I_{p_m}$
        Remove the associated type 1 link pair from LP_a
        Add p_m to LP_a

doi:10.6342/NTU202402157

```
    end for
    return $LP_a$
```


doi:10.6342/NTU202402157

Appendix B — Mathematical Proofs

B.1 Lemma 1.

A link pair p in LP only comes in one of the following forms. 1. $p = (\{i\}, \{T_0\})$, 2. $p = (\{i, j\}, \{T_0, x\})$, 3. $p = (\{i, j, k\}, \{T_0, T_0, x, y\})$, where $i, j, k \in \mathcal{N}, x, y \in \mathcal{D}, i \neq j, j \neq k, i \neq k, x \neq y, T_0$ is the base station.

Proof. Among all delivery modes, Mode A and Mode B are via cellular links while Mode C to F are through D2D links. A Mode A or B without sharing channels leads to a type 1 link pair. A Mode C, D, or F delivery, sharing channels with a Mode A or B cellular link forms a type 2 link pair. A Mode E delivery that includes two D2D links shares channels with two cellular links and forms a type 3 link pair, with two instances of T_0 and two different D2D devices. □

doi:10.6342/NTU202402157

B.2 Lemma 2.

There must be one and only one type 1 link pair $p = (\{i\}, \{T_0\})$ for each viewer $i \in \mathcal{N}$.

Proof. For every viewer with requested view v, if $v \in \mathcal{V}$, then a Mode A delivery will be adopted, Mode B otherwise. Since view v is either in or not in $\mathcal{V}$, there must be one and only one type 1 link pair for every viewer. $\square$

doi:10.6342/NTU202402157

B.3 Theorem 1.

SDVA is NP-hard.

Proof. In SDVA, viewers' requests are satisfied through the views delivered from either the BS or the D2D transmitters, both having corresponding constraints $\sum_{v \in d, d \in D_{CE}} b(v) \leq B$ and $\sum_{v \in d, d \in D_{D2D}} b(v) \leq r$. Hence, given an instance of the SDVA problem with viewers $1, \ldots, N$ and D2D transmitters $1, \ldots, D$, we can construct an instance of the Multidimensional Knapsack problem as follows: 1) Create x items, each corresponding to one of the possible deliveries. 2) Each item has $m+1$ dimensions, with the i-th dimension representing the bandwidth cost of the i-th D2D transmitter in the subset and the last dimension is the bandwidth cost of the BS. For example, if a delivery requires b amount of bandwidth from the BS, the last dimension will be set to b, and if a delivery requires the participation of the j-th D2D transmitter, the first dimension will be set to r. The dimensions where the corresponding transmitters are not involved will be set to 0. 3) Set the capacity of each dimension to r, except for the last dimension, which is set to B.

If there exists a subset of deliveries that satisfies both the BS and the D2D constraints ($\sum_{v \in d, d \in D_{CE}} b(v) \leq B$ and $\sum_{v \in d, d \in D_{D2D}} b(v) \leq r$), there exists a feasible solution to the Multidimensional Knapsack problem. Since the SDVA problem can be reduced to the Multidimensional Knapsack problem in polynomial time, and the Multidimensional Knapsack problem is NP-hard, the SDVA problem is also NP-hard. □

doi:10.6342/NTU202402157

B.4 Theorem 2.

TCATB-P guarantees truthful bidding.

Proof. According to Myerson's lemma, TCATB-P guarantees truthful bidding if TCATB-P is $monotone$. For any viewer i, the satisfaction is s_i if the requested view is received and 0 otherwise. Assume viewer i raises the bid from x_i to x'_i with $x_i < x'_i$, proving that VAMS is $monotone$ is equivalent to proving the utility viewer i received would be non-decreasing $U(x_i) \leq U(x'_i)$. There are three scenarios to discuss before the merging step in the TCATB phase: (i) With both x_i and x'_i, $i \in N_a$. (ii) With x'_i, $i \in N_a$, while with x_i, $i \in N_u$. (iii) With both x_i and x'_i, $i \in N_u$. For scenario (i), both x_i and x'_i ended up winning the bid and there is $U(x_i) = U(x'_i)$. In scenario (ii), bidding at x'_i is guaranteed to win the bid while x_i may or may not have the chance to win in the merging stage, hence there is $U(x_i) \leq U(x'_i)$. Finally for scenario (iii), both x_i and x'_i could only depend on the merging stage for a chance to win. Since raising the bid would not affect the situation of conflict but only moves viewer i ahead in the viewer sequence, bidding at x'_i would only increase the likelihood for viewer i to be merged in. Consequently, there is $U(x_i) \leq U(x'_i)$ and we can conclude that $U(x_i) \leq U(x'_i)$ holds in all scenarios and thus TCATB-P is monotone. $\square$

doi:10.6342/NTU202402157

B.5 Theorem 3.

TCATB-P is 1/2-approximate to the optimal solution.

Proof. From observing the three types of link pairs, we can tell that a type-2 link pair has the highest bandwidth efficiency, which could satisfy 2 viewers with bandwidth for one view at the base station (1 viewer for type 1 and 1.5 for type 3). The maximum number of satisfied viewers would be $2k$ consequently. Hence, the upper bound of the problem, denoted as U_{MAX} occurs when $v_1, \ldots, v_k$ can merge with $v_{k+1}, \ldots, v_{2k}$ via type 2 link pair perfectly without conflict and thus $U_{MAX} = U(P_{1,1}) + U(P_{2,1}) + \cdots + U(P_{2k,1})$. Let the outcome of the optimal solution be U_{OPT} and we have $U_{OPT} \leq U_{MAX}$.

Next, the worst-case scenario would be there is no feasible merging and the algorithm stops at Step 2. The above gives a lower bound of the algorithm, denoted as $U_{MIN} = U(P_{1,1}) + U(P_{2,1}) + \cdots + U(P_{k,1})$. Let the outcome of the algorithm be U_{ALG} and we have $U_{MIN} \leq U_{ALG}$.

Since Step 1 performs sorting on $U(P)$, there is $U(P_{1,1}) > U(P_{2,1}) > \cdots > U(P_{k,1}) > U(P_{k+1,1}) > \cdots > U(P_{2k,1})$. Altogether, we have $U_{OPT} \leq U_{MAX} < U(P_{1,1}) + U(P_{2,1}) + \cdots + U(P_{2k,1}) < 2 \cdot \{U(P_{1,1}) + U(P_{2,1}) + \cdots + U(P_{k,1})\} = 2 \cdot U_{MIN} \leq 2 \cdot U_{ALG}$, which shows that TCATB-P is 1/2-approximate to the optimal solution. □

doi:10.6342/NTU202402157

B.6 Theorem 4.

SDVA with TCATB-P runs in $O(N^2D^2)$ time.

Proof. The overall time complexity of VDME is $O(ND^2)$, where N is the number of viewers and D is the number of D2D devices. For each viewer, the evaluation of Mode A and Mode B takes $O(1)$ time, Mode C and Mode D takes $O(D)$ time, Mode E takes $O(D^2)$ time and Mode F takes $O(D)$ time respectively. Hence, VDME runs in $O(ND^2)$ time. The time complexity of CDLC is $O(N^2D^2)$. The time complexity of TCATB-P is $O(N^2D^2)$. Since VAMS is a sequential process combining VDME, CDLC, and TCATB-P, VAMS with TCATB-P has an overall time complexity of $O(N^2D^2)$. □

doi:10.6342/NTU202402157