Article:

Shang, B., Zhao, L., Chen, K. et al. (1 more author) (2018) An Economic Aspect of Device-to-Device Assisted Offloading in Cellular Networks. IEEE Transactions on Wireless Communications, 17 (4). pp. 2289-2304. ISSN 1536-1276

https://doi.org/10.1109/TWC.2018.2791518

An Economic Aspect of Device-to-Device Assisted Offloading in Cellular Networks

Bodong Shang, *Student Member, IEEE*, Liqiang Zhao, *Member, IEEE*,
Kwang-Cheng Chen, *Fellow, IEEE*, and Xiaoli Chu, *Senior Member, IEEE*

***Abstract*—Traffic offloading via device-to-device (D2D) communications has been proposed to alleviate the traffic burden on base stations (BSs) and to improve the spectral and energy efficiency of cellular networks. The success of D2D communications relies on the willingness of users to share contents. In this paper, we study the economic aspect of traffic offloading via content sharing among multiple devices and propose an incentive framework for D2D assisted offloading. In the proposed incentive framework, the operator improves its overall profit, defined as the network economic efficiency (ECE), by encouraging users to act as D2D transmitters (D2D-Txs) which broadcast their popular contents to nearby users. We analytically characterize D2D assisted offloading in cellular networks for two operating modes: 1) underlay mode and 2) overlay mode. We model the optimization of network ECE as a two-stage Stackelberg game, considering the densities of cellular users and D2D-Tx's, the operator's incentives and the popularity of contents. The closed-form expressions of network ECE for both underlay and overlay modes of D2D communications are obtained. Numerical results show that the achievable network ECE of the proposed incentive D2D assisted offloading network can be significantly improved with respect to the conventional cellular networks where the D2D communications are disabled.**

***Index Terms*—D2D communications, economic efficiency, offloading, Poisson point process, Stackelberg game.**

I. Introduction

With the fast-growing data traffic and the explosive increase of mobile devices [1], network economy has become an important aspect of mobile network operations under limited radio resources [2]–[4]. To serve heavy traffic in future cellular networks, device-to-device (D2D) communication has been considered to alleviate the traffic burden on base stations (BSs) by exploiting physical proximity of devices [5]. There are two operating modes of D2D communications: underlay (sharing spectrum with cellular links) and overlay (orthogonal spectrum to cellular networks) [6]. In the underlay mode, the mutual interference between cellular tier and D2D tier should be considered. In the overlay mode, the cross-tier interference is eliminated at the price of reduced cellular bandwidth within the limited licensed spectrum. Existing works on D2D communications have studied resource and power allocation [7], mode selection [8], spectral and energy efficiency [9], [10]. From the operator's perspective, it is of more interest to consider economic benefits from operation, such as cost for operating, and so on, which serves our unique holistic view on D2D communications.

In the meantime, content sharing among multiple devices, e.g., video streaming, has been regarded as one of the most promising methods for traffic offloading and as the tremendous data consuming application in wireless communications [11], [12]. When a content requester downloads a popular file, a nearby device can deliver it locally, e.g., through the synchronous content transmission application [13]. In [14], the authors designed a social-aware video multiCast (SoCast) system based on D2D communications by considering the social trust and social reciprocity. Actually, the success of D2D assisted offloading relies on the willingness of users to share contents, given the power consumption for transmission. Thus, network operators can offer incentives to encourage users to participate in D2D content sharing [15], [16]. However, both [15] and [16] considered a single cell and the impacts of wireless channels and interference are not taken into account.

On the other hand, the operator's profit has been studied for wireless networks [2]–[4], [17], but not for D2D communications. In [3], the economic benefits of delayed WiFi offloading were investigated, while in [4] the economic effects of user-oriented delayed Wi-Fi offloading were studied in monopoly and duopoly market models. In [17], pricing strategies for both macro and femto cell operators in a cognitive femtocell network were discussed. However, none of these works have considered the operator's profit with respect to D2D assisted offloading especially from system-level perspective. The interrelation between pricing incentives and transmit powers at both BS and D2D transmitters (D2D-Txs) (and thus the co-channel interference) is not well understood yet.

In this paper, we define the operator's profit as the *economic efficiency* (ECE) [18], which is the difference between the operating income from users and the total operating cost. The total operating cost consists of power cost at BSs and incentive cost for D2D-Txs. We analyze the ECE on the downlink traffic offloading via content sharing in synchronous transmission application based on D2D communications, where the D2D-Txs are motivated by operator to broadcast the popular content to its nearby devices. More specifically, we assume that D2D-Txs are rational and can control their transmit power to maximize their utilities, and the operator determines both the

This work was supported in part by National Natural Science Foundation of China (61771358), Intergovernmental International Cooperation on Science and Technology Innovation (2016YFE0122900), China Postdoctoral Science Foundation (2017M613074), Fundamental Research Funds for the Central Universities (JB170102), and the 111 Project (B08038).

Bodong Shang and Liqiang Zhao are with the State Key Lab. of ISN, Xidian University, China (emails:bdshang@hotmail.com, lqzhao@mail.xidian.edu.cn).

Kwang-Cheng Chen is with the Department of Electrical Engineering, University of South Florida, USA (email:kwangcheng@usf.edu).

Xiaoli Chu is with the Department of Electronic and Electrical Engineering, University of Sheffield, UK, (email:x.chu@sheffield.ac.uk).

incentives to potential D2D-Tx's and the subsequent radio resource management in order to maximize its overall profit.

Compared with [19] focusing on the underlay mode of D2D offloading, we consider both underlay (partial frequency reuse) and overlay modes of D2D communications and provide a generalized economic analytical framework. We jointly optimize both incentives and radio resources of D2D communications for the two D2D operating modes and compare their performance from an economic perspective. We also remove the constraint of non-overlapping offloading regions used in [19] and consider potentially different content popularities.

The main contributions of this work are summarized below:

1) Tractable model for D2D assisted offloading: We develop a novel and tractable analytical model for accurately capturing the traffic offloading from cellular links to D2D links. In particular, the density of cellular users, which do not participate in D2D content sharing, is specified by incorporating the popularity of the shared contents and the offloading region of D2D-Txs. We also characterize the transmit power at both the D2D-Tx and the BS, under the constraint of users' average data rate requirement, which represents users' quality of service (QoS). This model is flexible enough to apply to both underlay and overlay modes of D2D communications. It also includes the geographical mobile-traffic intensity, the content popularity and the spatially averaged data rate requirement.

2) System-level economic analysis: We derive the closed-form expressions for the network ECE in both underlay and overlay modes of D2D assisted offloading. We obtain the optimal offloading radius of a D2D-Tx based on a two-stage Stackelberg game. This enables us to quantify the number of offloaded users, which ultimately leads to the optimized transmit power of BS and the network ECE. The maximum network ECE is called *achievable network ECE* in this paper, and we obtain the optimal density of D2D-Txs that maximizes the achievable network ECE. In addition, the network ECE is analyzed when the D2D assisted offloading is disabled and the optimal price offered to cellular users is obtained.

3) Network design insights: Our analysis leads to several system design guidelines. First, compared with the conventional cellular networks, sharing popular contents through the incentive D2D communications can significantly improve the network ECE. The optimal solutions for both incentives and radio resource management can be obtained based on our model numerically. Second, it is unprofitable to use the overlay mode of D2D communications when the shared content is not very popular. Finally, the overlay mode achieves a higher network ECE when the D2D-Txs density is small, but when the D2D-Txs density is large, the underlay mode achieves a higher network ECE.

The remainder of this paper is organized as follows. Section II presents the system model and the utility functions of the operator and D2D-Tx. Section III and IV formulates the mathematical economic model of D2D assisted cellular networks for underlay and overlay mode, respectively. Section V studies the operator's economy when D2D assisted offloading is disabled. Numerical results are shown and discussed in section VI. The conclusions are summarized in section VII.

Notation: $\mathbb{E}[x]$ denotes the expectation of variable x. $\mathbb{P}[A]$ denotes the probability that event A happens. y^* denotes the optimal value of y. Finally, $[x]^+ \triangleq \max\{0, x\}$.

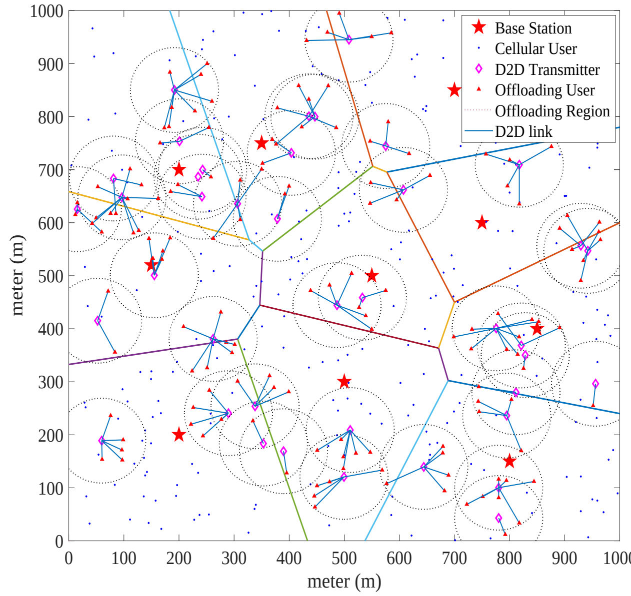

Fig 1: D2D assisted offloading in the cellular downlink networks, where D2D-Txs can broadcast popular contents to the users in proximity.

II. System Model

A. Network layout

We consider the cellular downlink collocated with D2D communications as in Fig.1. We assume that BSs are distributed following a homogeneous Poisson Point Process (PPP) on the entire plane $\mathbb{R}^2$ with the density of $\lambda_B\left(\mathrm{BSs/m^2}\right)$, and that they can be denoted as the set of $\Psi_B = \{b_j, j = 0, 1, 2, ...\}$. Each BS has the maximum allowable transmit power P_m. D2D-Txs are uniformly distributed on $\mathbb{R}^2$ according to another independent PPP with the density of $\lambda_D\left(\mathrm{D2D\text{-}Txs/m^2}\right)$, and the set of D2D-Txs is denoted by Ψ_D. Each D2D-Tx has the maximum allowable transmit power P_D. Users are spatially scattered on $\mathbb{R}^2$ following another independent PPP, denoted by the set Ψ_U with the density $\lambda_u\left(\mathrm{users/m^2}\right)$. Users are classified into cellular users (served by BSs) and offloaded users (using D2D links).

B. User's association and content popularity

Each cellular user connects to the closest BS $(b_j \in \Psi_B)$, and the cell area of BS b_j can be defined as the set $V_j = \left\{x \in \mathbb{R}^2 \,|\, \|x - b_j\| \le \|x - b_n\|, b_n \in \Psi_B \backslash b_j\right\}$, where $\|a - b\|$ represents the distance between a and b.

In D2D communications, the i^{th} offloaded user $u_{i,k}^d$ connects to the k^{th} D2D-Tx $\left(u_k^{DT}\right)$ if the following two requirements are both satisfied; otherwise, the user connects to the closest BS. First, the distance between user $u_{i,k}^d$ and D2D-Tx u_k^{DT} is within the D2D communications range of radius R_D. Second, the user's requested content is available at D2D-Tx u_k^{DT}. Regarding the first requirement, the offloading region of u_k^{DT} is defined as $\Omega_k^{DT} = \left\{x \in \mathbb{R}^2 \,\middle|\, \left\|x - u_k^{DT}\right\| \le R_D, u_k^{DT} \in \Psi_D\right\}$, which forms a circular region centered at u_k^{DT} with the radius of R_D, where the offloaded user's QoS can be guaranteed. We use the average required data rate of users as the user's QoS requirement.

Table I: Poisson point process variables.

Notation	Description
$\mathbb{R}^2$	The entire network plane
Ψ_B	Set of cellular BSs
λ_B	Density of BSs per square meter
b_j	The j^{th} BS in the network
P_m	Maximum allowable transmit power of BS
Ψ_D	Set of D2D-Txs
λ_D	Density of D2D-Txs per square meter
P_D	Maximum allowable transmit power of D2D-Tx
Ψ_U	Set of users, i.e., cellular users and offloaded users
λ_u	Density of users per square meter
V_j	Cell area of j^{th} BS b_j
$u_{i,k}^d$	i^{th} offloaded user associated with k^{th} D2D-Tx
u_k^{DT}	The k^{th} D2D-Tx
R_D	Offloading radius of a D2D-Tx
Ω_k^{DT}	Offloading region of the k^{th} D2D-Tx u_k^{DT}
$\mathcal{G}_z$	z^{th} D2D-Txs group with the content popularity $\mathbb{P}_{con}^{\mathcal{G}_z}$
$\mathbb{P}_{con}^{\mathcal{G}_z}$	Content popularity of the D2D-Txs group $\mathcal{G}_z$
$\mathbb{P}_{con}$	Weighted average content popularity in the network
$\Psi_{u,j}^c$	Set of cellular users in cell V_j
N_j^c	The number of cellular users in cell V_j

Table II: General channel model variables.

Notation	Description
B	Total available bandwidth of operator
B_C	Bandwidth of cellular downlink
B_D	Bandwidth of D2D communications
ρ	Frequency reuse factor in the underlay mode
ω	Frequency partition factor in the overlay mode
δ	$\delta=\rho$ for underlay mode, $\delta=\omega$ for overlay mode
$u_{i,j}^c$	The i^{th} cellular user in j^{th} cell of cellular BS b_j
$\mu_{i,j}^c$	Allocation factor for user $u_{i,j}^c$
$B_{i,j}^c$	Bandwidth allocated to $u_{i,j}^c$, $B_{i,j}^c=\mu_{i,j}^c B_C/N_j^c$
$P_{i,j}^B$	BS transmit power for $u_{i,j}^c$ on its sub-band $B_{i,j}^c$
$g_{i,j}^c$	Fast-fading power gain from BS b_j to user $u_{i,j}^c$
$I_{c,i,j}^C$	Aggregated interference from cellular tier at $u_{i,j}^c$
$I_{c,i,j}^D$	Cross-tier interference from D2D-Txs at $u_{i,j}^c$
α	Path-loss exponent
σ^2	Additive noise
P_j^B	Required aggregated transmit power of BS b_j
β	Frequency division parameter for D2D sub-bands
P_k^D	Required transmit power at k^{th} D2D-Tx u_k^{DT}
$h_{i,k}^d$	Fast-fading power gain from u_k^{DT} to user $u_{i,k}^d$
$I_{d,i,k}^C$	Aggregated interference from cellular tier at $u_{i,k}^d$
$I_{d,i,k}^D$	Co-tier interference from D2D-Txs at $u_{i,k}^d$

The probability that the requested content of a typical cellular user u_0 is available at an arbitrary D2D-Tx is defined as *content popularity*. We assume that D2D-Txs are partitioned into Z disjoint groups according to the popularity of their available contents. The set of group index is $\mathcal{Z}=\{1,2,...,Z\}$, and group z is denoted by $\mathcal{G}_z$. The ratio of the size of group $\mathcal{G}_z$ to the total number of D2D-Txs is denoted by $\varphi_z\,(z\in\mathcal{Z})$. We define $\mathbb{P}_{con}^{\mathcal{G}_z}\left(0\leqslant\mathbb{P}_{con}^{\mathcal{G}_z}\leqslant 1, z\in\mathcal{Z}\right)$ as the probability that a user's requested content is available in group $\mathcal{G}_z$, which also indicates the content popularity of group $\mathcal{G}_z$. The popularity of these content chunks is expressed by a vector $\mathbf{P_{con}}=\left\{\mathbb{P}_{con}^{\mathcal{G}_1},\mathbb{P}_{con}^{\mathcal{G}_2},...,\mathbb{P}_{con}^{\mathcal{G}_Z}\right\}$. It's worth noting that, in each group, although the D2D-Txs may have different content chunks, the popularity of these contents are nearly the same and we take their average as $\mathbb{P}_{con}^{\mathcal{G}_z}$ in group $\mathcal{G}_z$. The value of content popularity can be obtained by the keywords feature extraction method [20], [21] or the machine learning method [22], [23] according to the users' download history.

C. Radio resources

The total available bandwidth of the operator is B MHz. The bandwidth of the cellular downlink is denoted by B_C MHz, and the bandwidth for D2D communications is B_D MHz. To maintain generality, in the underlay mode, $B_C=B$, and $B_D=\rho B$ where ρ represents the frequency reuse factor. In the overlay mode, the bandwidth of cellular system is $B_C=(1-\omega)B$ where ω is the frequency partition factor, and $B_D=\omega B$. In both modes, we consider the full frequency reuse among cellular cells.

D. Cellular downlink

Cellular users in cell V_j can be denoted as a set of $\Psi_{u,j}^c$, where $\left|\Psi_{u,j}^c\right|=N_j^c$ is the number of cellular users in cell V_j. The downlink bandwidth in each cell is B_C MHz, and user $u_{i,j}^c$ (i^{th} cellular user in j^{th} cell) obtains $B_{i,j}^c=\mu_{i,j}^c B_C/N_j^c$ MHz, where $\mu_{i,j}^c\left(\mu_{i,j}^c\geqslant 0\right)$ denotes the allocation factor for user $u_{i,j}^c$. Suppose that the bandwidth B_C is fully used (i.e., $\sum_{i=1}^{N_j^c}B_{i,j}^c=B_C,\forall b_j\in\Psi_B$), and thus we have $\sum_{i=1}^{N_j^c}\mu_{i,j}^c=N_j^c$. There is no intra-cell interference due to the orthogonal multiple access in a cell.

We assume that each BS is capable of performing adaptive power control according to user feedback channel state information (CSI) [24]. According to Shannon's theorem, in the underlay mode, BS b_j allocates the transmit power $P_{i,j}^B$ for user $u_{i,j}^c$ to achieve the required data rate $R_{i,j}^c$ as follows,

$$R_{i,j}^c\leqslant B_{i,j}^c\log_2\left(1+\frac{P_{i,j}^B g_{i,j}^c\left\|u_{i,j}^c-b_j\right\|^{-\alpha}}{I_{c,i,j}^C+I_{c,i,j}^D+\sigma^2}\right)\tag{1}$$

where $g_{i,j}^c$ represents the channel power gain from BS b_j to cellular user $u_{i,j}^c$, α is the path-loss exponent, $I_{c,i,j}^C$ is the total received interference power from cellular networks at $u_{i,j}^c$, $I_{c,i,j}^D$ denotes the total received interference power from D2D communications at $u_{i,j}^c$ and σ^2 is the additive noise. At the BS b_j, the aggregated downlink transmit power for the N_j^c cellular users is given by $P_j^B=\sum_{i=1}^{N_j^c}P_{i,j}^B$.

E. D2D links

We assume that the available bandwidth for D2D communications (B_D) is uniformly divided into β sub-bands in both underlay and overlay modes, and each D2D-Tx can randomly access to one of those sub-bands.

For an offloaded user $u_{i,k}^d$, in the underlay mode, the transmit power P_k^D at D2D-Tx u_k^{DT} is provided to ensure the required data rate of $u_{i,k}^d$, as follows

$$R_{i,k}^d\leqslant\frac{B_D}{\beta}\log_2\left(1+\frac{P_k^D h_{i,k}^d\left\|u_{i,k}^d-u_k^{DT}\right\|^{-\alpha}}{I_{d,i,k}^C+I_{d,i,k}^D+\sigma^2}\right)\tag{2}$$

Table III: Network economic variables.

Notation	Description
τ	Operator's income for each user
R_u	Users' average required data rate
c_B	Cost factor regarding to power consumption at a BS
P_B^{non}	The BS's static operation power
P_B^{agg}	The BS's practical aggregated transmit power
P_B^{total}	Total power consumption, $P_B^{total} = P_B^{non} + P_B^{agg}$
ε	Operator's incentive for each D2D-Tx
c_D	Cost factor of power consumption at D2D-Txs
ξ_u	User's income factor
θ	User's willingness-to-pay factor, $\theta \in [0, \theta_{\max}]$
$\theta_{\max}$	User's maximum willingness-to-pay
$\mathcal{U}_{Operator}$	Operator's utility function
$\mathcal{U}_{D2D-Tx}$	D2D-Tx's utility function
$\mathcal{U}_{User}$	User's utility function

where $\frac{B_D}{\beta}$ denotes the transmission bandwidth per D2D-Tx, and accordingly $h_{i,k}^d$ is the channel power gain, $I_{d,i,k}^C$ denotes the total received interference power from cellular networks at $u_{i,k}^d$, $I_{d,i,k}^D$ indicates the total received interference power from D2D communications at $u_{i,k}^d$.

F. Economic utility functions

Assumptions. we make several assumptions for mathematically tractable analysis of the network ECE.

First, we assume that all the cellular users have an identical data rate requirement of R_u, i.e., $R_{i,j}^c = R_u, \forall u_{i,j}^c \in \Psi_U$. This assumption is made for mathematical tractability, but more importantly because we mainly focus on the system-level performance of multi-cell networks and on providing network design insights statistically. We can consider R_u as the average required data rate of users in the networks. For example, R_u can be calculated by $R_u = \left(\sum_{j=1}^{K} \sum_{i=1}^{N_j^c} R_{i,j}^c\right) \Big/ \sum_{j=1}^{K} N_j^c$, where K indicates the number of cells in the region of interest, $N_j^c = \left|\Psi_{u,j}^c\right|$ denotes the number of cellular users in cell V_j.

Second, we assume that the offloaded users have an identical rate requirement of $R_{i,k}^d = R_u, \forall u_{i,k}^d \in \Psi_U$. This assumption is reasonable, because the offloading procedure is transparent to the users. Although a user is offloaded from cellular link onto D2D link, the user would not be aware of the type of transmission, and thus the data rate requirement will not change during the offloading process.

Inspired by [18] and [25], economic efficiency measures the profitability of the network operator (in monetary unit per second), and the power consumption contributes to the cost of networks because the operator needs to pay the resulting electricity bills. For notational simplicity, we define the utility function of cellular operator $\mathcal{U}_{Operator}(\varepsilon, \delta)$ as the difference between the operating income and the total cost, as follows:

$$\mathcal{U}_{Operator}(\varepsilon, \delta) = Operating\ Income - Total\ Cost\ , \quad (3)$$

$$\text{and } Total\ Cost = Power\ Cost + Incentive\ Cost \quad (4)$$

where $\varepsilon\left(\text{pence/Mbit/m}^2\text{/D2D-Tx}\right)$ is defined as the incentive per Mbit offloaded per unit D2D-Tx offloading region, and δ is an indicator, where $\delta = \rho$ in the underlay mode and $\delta = \omega$ in the overlay mode. Note that the first term on the right handside in (3) represents the operator's income per unitary area which is charged from users, and the second term includes power cost at BSs and incentive cost for implementing D2D assisted offloading.

More specifically, we have

$$Operating\ Income = \lambda_u \tau R_u \left(\text{pence/m}^2\text{/s}\right), \quad (5)$$

$$Power\ Cost = \lambda_B c_B P_B^{total}(\varepsilon, \delta) \left(\text{pence/m}^2\text{/s}\right), \quad (6)$$

$$Incentive\ Cost = \lambda_D \varepsilon R_u \pi {R_D}^2 \left(\text{pence/m}^2\text{/s}\right) \quad (7)$$

where τ (pence/Mbit/user) in (5) denotes the income per Mbit per user, and R_u (Mbps) indicates the average required data rate of users. In addition, c_B (pence/Joule/BS) in (6) represents a cost factor with respect to power consumption at a BS, and $P_B^{total}(\varepsilon, \delta) = P_B^{non} + P_B^{agg}(\varepsilon, \delta)$ (Watt) is the total power consumption at a BS, which includes the non-transmission power P_B^{non} and the aggregated transmit power $P_B^{agg}(\varepsilon, \delta) = \mathbb{E}\left[P_j^B(\varepsilon, \delta)\right]$. The term $\pi {R_D}^2$ in (7) is the offloading area for a D2D-Tx, which captures the impact of geographic traffic volume in a large-scale networks.

In addition, the content distribution of each user and the content popularity are essential for the operator to design an efficient incentive offloading scheme. Motivated by [22], the baseband units (BBUs) can predict the distribution of content request and users's mobility. This information gathering phase results in the additional *computational cost* for the operator, which can be approximately regarded as a fix cost. To simplify the analysis, we assume that the content distribution and the content popularity are known by the operator and that the associated *computational cost* is constant and can be ignored in the operator's utility function (4).

Accordingly, the utility of D2D-Tx $\mathcal{U}_{D2D-Tx}(R_D)$ (pence/s/D2D-Tx) is defined as the difference between its incentive income and power cost as follows

$$\begin{aligned}\mathcal{U}_{D2D-Tx}(R_D) &= Incentive\ Income - Power\ Cost \\ &= \varepsilon R_u \pi {R_D}^2 - c_D \mathbb{E}\left[\widehat{P_k^D}(R_D)\right]\end{aligned} \quad (8)$$

where ε is the operator's incentive parameter as previously defined in (3), c_D (pence/Joule/D2D-Tx) denotes the cost factor of the power consumption at a D2D-Tx, and $\mathbb{E}\left[\widehat{P_k^D}(R_D)\right]$ (Watt) is the D2D-Tx's expected minimum transmit power.

G. Implementation of incentive D2D assisted offloading

The overall process of the proposed incentive D2D assisted offloading is given in the following steps.

Step 1: When a user requests a content, the operator will first check if any D2D-Txs in the proximity of the user within distance R_D are broadcasting the requested content.

Step 2: If yes, the operator allows the user to receive files from one of these D2D-Txs via D2D communications. If not, the operator will search in the surrounding area with radius R_D of the user for potential D2D-Txs having the requested content.

Step 3: If potential D2D-Txs are found, the operator will select one of them to transmit the requested content to the user via a D2D link, and give an incentive to the selected D2D-Tx.

Otherwise, the serving BS will transmit the requested content to the user via the downlink.

Step 4: If the offloaded content-requesting user strolls out of the offloading region of its previous associated D2D-Tx, which suggests that the user's data rate requirement will not be guaranteed and will be less than its previous cellular data rate, the operator will reselect a D2D-Tx for the user following the process from *Step 1* to *Step 3*.

In the incentive mechanism of *Step 3*, the operator acts as the leader to announce the incentive of traffic offloading and the allocation of D2D radio resources. Then, based on the given incentive and radio resources, the D2D-Tx (as a follower) decides its offloading radius to maximize its utility. Accordingly, we formulate the incentive problem as a two-stage sequential game, i.e., a Stackelberg game, as follows.

Definition 1. *The proposed Stackelberg game is defined as:*

Players: *The operator as the leader takes action in Stage I. D2D-Tx as a follower takes action in Stage II. Stages I and II will be detailed in Section III.*

Utilities: *The operator's utility function is given by $\mathcal{U}_{Operator}(\varepsilon,\delta)$ in (3). The utility function of D2D-Tx is given by $\mathcal{U}_{D2D-Tx}(R_D)$ in (8).*

Strategies: *The operator optimizes the incentive parameter ε and the bandwidth δ of D2D radio resources (where $\delta=\rho$ in the underlay mode, and $\delta=\omega$ in the overlay mode). D2D-Tx optimizes the offloading radius R_D.*

III. System-level ECE with underlay mode of D2D communications

In this section, we analyze the Stackelberg game using backward induction. We first derive the expected minimum transmit power of D2D-Tx as a function of the D2D offloading radius. Based on the expected minimum transmit power and the utility function of D2D-Tx, we obtain the reaction function of the follower D2D-Tx in stage II. Then, we derive the average aggregated transmit power at a BS, and solve the leader's profit-maximization problem in stage I according to its utility function.

A. Stage II: Follower's game - D2D-Tx's offloading radius

In stage II, each D2D-Tx determines the offloading radius under the QoS requirement of users. Then the transmit power of D2D-Tx can be determined.

Proposition 1. *In a cellular network underlaid with D2D communications, given the BS density λ_B, frequency reuse factor ρ and the offloading radius R_D, the required transmit power P_k^D of D2D-Tx u_k^{DT} for the offloaded user $u_{i,k}^d$ with the data rate requirement of $R_{i,k}^d$ is given by*

$$P_k^D \geqslant \sigma^2\left(2^{\frac{\beta R_{i,k}^d}{\rho B}}-1\right){R_D}^{\alpha}+\frac{2^{\left(\beta R_{i,k}^d\right)/(\rho B)}-1}{{R_D}^{-\alpha}} \cdot\left[\frac{4(\pi\lambda_B)^{\frac{\alpha}{2}}\rho P_m}{\beta(\alpha-2)(4-\alpha)}+\frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\alpha-2)(\pi\lambda_D)^{-\frac{\alpha}{2}}}\right] \tag{9}$$

where $\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$ is the standard gamma function, and the path-loss exponent $\alpha>2$ [26].

Proof: Please refer to Appendix A for the proof. ∎

Note that although $\alpha<4$ would be required for a positive value of the mean interference power in the square bracket of (9), our numerical results show that the mean interference power converges to a value very close to zero for $\alpha\geqslant 4$. Thus, we approximate the mean interference power to zero for $\alpha\geqslant 4$.

Assuming identical data rate requirements of users, i.e., $R_{i,k}^d=R_u, \forall u_{i,k}^d\in\Psi_U$, the expected minimum transmit power for a typical D2D-Tx u_k^{DT} is obtained as,

$$\mathbb{E}\left[\widehat{P_k^D}\right]=\mathbb{E}\left[\min\left\{P_k^D\big|_{R_{i,k}^d=R_u}\right\}\right]=\frac{\sigma^2\left(2^{(\beta R_u)/(\rho B)}-1\right)}{{R_D}^{-\alpha}} +\frac{\left(2^{\frac{\beta R_u}{\rho B}}-1\right)}{{R_D}^{-\alpha}(\alpha-2)}\cdot\left[\frac{4(\pi\lambda_B)^{\frac{\alpha}{2}}\rho P_m}{\beta(4-\alpha)}+\frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\pi\lambda_D)^{-\frac{\alpha}{2}}}\right]. \tag{10}$$

Although the user's data rate requirement fluctuates over time and the data rate requirements of users may be various, this equality constraint statistically represents the average minimum transmit power of a D2D-Tx, which can simplify the analysis of the network ECE.

According to (10), the optimal offloading radius of a D2D-Tx, which maximizes the D2D-Tx's utility (8), can be obtained as shown in Corollary 1.

Corollary 1. *Consider the D2D assisted offloading underlaid with cellular networks, where the densities of BSs and D2D-Txs are λ_B and λ_D, respectively. Given the incentive ε, frequency reuse factor ρ and the average required data rate R_u of users, the closed-form optimal offloading radius of the typical D2D-Tx u_k^{DT} is given by*

$$R_D^*=\left[\frac{\varepsilon R_u\pi^{1-\frac{\alpha}{2}}\frac{1}{2\alpha c_D}(4-\alpha)\left(2^{\frac{\beta R_u}{\rho B}}-1\right)^{-1}}{\frac{\rho P_m\lambda_B^{\frac{\alpha}{2}}}{\beta(\alpha-2)}+\frac{P_D\Gamma\left(2-\frac{\alpha}{2}\right)(4-\alpha)}{2\beta(\alpha-2){\lambda_D}^{-\frac{\alpha}{2}}}+\frac{\sigma^2(4-\alpha)}{4\pi^{\frac{\alpha}{2}}}}\right]^{\frac{1}{\alpha-2}}. \tag{11}$$

Proof: The optimal offloading radius R_D^* of D2D-Tx can be obtained by plugging (10) into (8) and maximizing its utility function (8). We obtain the first order derivative of $\mathcal{U}_{D2D-Tx}(R_D)$ with respect to the offloading radius R_D as

$$\begin{aligned}\frac{\partial\mathcal{U}_{D2D-Tx}}{\partial R_D}&=2\varepsilon R_u\pi R_D-c_D\frac{\partial\mathbb{E}\left[\widehat{P_k^D}\right]}{\partial R_D}\\&=2\varepsilon R_u\pi R_D-c_D\frac{\alpha\left(2^{(\beta R_u)/(\rho B)}-1\right)}{{R_D}^{1-\alpha}(\alpha-2)}\cdot C,\ \text{where}\\C&=\left[\frac{4(\pi\lambda_B)^{\frac{\alpha}{2}}\rho P_m}{\beta(4-\alpha)}+\frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\pi\lambda_D)^{-\frac{\alpha}{2}}}+\frac{\sigma^2}{(\alpha-2)^{-1}}\right].\end{aligned} \tag{12}$$

Assuming $2<\alpha<4$, we have $\lim\limits_{R_D\to 0}\frac{\partial\mathcal{U}_{D2D-Tx}}{\partial R_D}=0$ and $\lim\limits_{R_D\to\infty}\frac{\partial\mathcal{U}_{D2D-Tx}}{\partial R_D}<0$. The second derivative of $\mathcal{U}_{D2D-Tx}$ with respect to R_D is given by

$$\frac{\partial^2\mathcal{U}_{D2D-Tx}}{\partial {R_D}^2}=2\varepsilon R_u\pi-\frac{c_D\alpha(\alpha-1)C{R_D}^{\alpha-2}}{(\alpha-2)\left(2^{\frac{\beta R_u}{\rho B}}-1\right)^{-1}}. \tag{13}$$

According to (13), we have $\lim_{R_D \to 0} \frac{\partial^2 \mathcal{U}_{D2D-Tx}}{\partial {R_D}^2} > 0$ and $\lim_{R_D \to \infty} \frac{\partial^2 \mathcal{U}_{D2D-Tx}}{\partial {R_D}^2} < 0$. More specifically, $\frac{\partial \mathcal{U}_{D2D-Tx}}{\partial R_D}$ increases with R_D when $0 < R_D < r_D$ and decreases when $r_D < R_D < \infty$, where $r_D = \left[\frac{2\varepsilon R_u \pi(\alpha-2)}{c_D \alpha(\alpha-1)C\left(2^{\beta R_u/(\rho B)}-1\right)}\right]^{\frac{1}{\alpha-2}}$ which is obtained by solving the equation $\frac{\partial^2 \mathcal{U}_{D2D-Tx}}{\partial {R_D}^2} = 0$.

Therefore, there exists a unique R_D^* ($R_D^* > r_D$) which guarantees that $\frac{\partial \mathcal{U}_{D2D-Tx}}{\partial R_D}$ is positive in $(0, R_D^*)$ and negative in (R_D^*, ∞). By solving $\frac{\partial \mathcal{U}_{D2D-Tx}}{\partial R_D} = 0$ (12), the unique optimal offloading radius R_D^* of D2D-Tx is obtained in (11). ■

In Corollary 1, the optimal offloading radius R_D^* increases with the incentive ε, while decreases with the cost factor of power consumption c_D and the density of D2D-Txs λ_D. Since denser D2D-Txs could cause severe interference to each other, the transmit power at D2D-Tx should be increased in order to guarantee the QoS of offloaded users.

It's worth noting that (11) is the expected optimal offloading radius of a D2D-Tx, where parameters such as the users' average data rate requirement R_u are assumed to be known by the D2D-Tx. Therefore, the operator should inform the D2D-Tx of the users' QoS requirements, so that the D2D-Tx can allocate sufficient power for transmission. Other network parameters like the system bandwidth and the path-loss exponent are assumed to be known to the D2D-Txs as the priori information from network.

In addition, substituting (11) into (10) gives the optimal transmit power at D2D-Tx denoted by $\mathbb{E}\left[P_k^{D,opt}\right]$.

Proposition 2. *The optimal transmit power at a typical D2D-Tx u_k^{DT} is given by*

$$\mathbb{E}\left[P_k^{D,opt}\right] = 2\varepsilon^{\frac{\alpha}{\alpha-2}} R_u \pi \frac{1}{\alpha c_D} \Omega^{\frac{2}{\alpha-2}}$$
$$\text{where } \Omega = \frac{R_u \pi^{1-\frac{\alpha}{2}} \frac{1}{2\alpha c_D}(4-\alpha)\left(2^{\frac{\beta R_u}{\rho B}}-1\right)^{-1}}{\frac{\rho P_m {\lambda_B}^{\frac{\alpha}{2}}}{\beta(\alpha-2)} + \frac{P_D \Gamma\left(2-\frac{\alpha}{2}\right)(4-\alpha)}{2\beta(\alpha-2){\lambda_D}^{-\frac{\alpha}{2}}} + \frac{\sigma^2(4-\alpha)}{4\pi^{\frac{\alpha}{2}}}}. \tag{14}$$

Proof: Substituting (11) into (10) gives us the desired result. ■

Subsequently, by substituting (11) into (10) in place of R_D and then substituting (10) into (8), we obtain the D2D-Tx's maximum utility given in Corollary 2.

Corollary 2. *Consider the D2D assisted offloading underlaid with cellular networks, where the incentive ε and the average required data rate R_u of users are given, the maximum utility of D2D-Tx is given by*

$$\mathcal{U}_{D2D-Tx}^* = \varepsilon R_u \pi (R_D^*)^2 \left(1 - \frac{2}{\alpha}\right) \tag{15}$$

where R_D^* is given in Corollary 1.

Proof: Substituting R_D^* for R_D in (8), the maximum utility of D2D-Tx is given by

$$\begin{aligned}\mathcal{U}_{D2D-Tx}^* = \varepsilon R_u \pi (R_D^*)^2 - c_D \frac{\left(2^{\frac{\beta R_u}{\rho B}}-1\right)4\pi^{\frac{\alpha}{2}}}{(\alpha-2)\left(R_D^*\right)^{-\alpha}} \\ \cdot \left[\frac{\rho P_m {\lambda_B}^{\frac{\alpha}{2}}}{\beta(4-\alpha)} + \frac{P_D \Gamma\left(2-\frac{\alpha}{2}\right)}{2\beta {\lambda_D}^{-\frac{\alpha}{2}}} + \frac{\sigma^2(\alpha-2)}{4\pi^{\frac{\alpha}{2}}}\right].\end{aligned} \tag{16}$$

With some manipulation, (16) can be simplified into (15). ■

It can be observed that $\mathcal{U}_{D2D-Tx}^*$ increases with the incentive ε, and D2D-Tx does not get any benefit when there is no incentive, i.e., $\varepsilon = 0$.

B. Stage I: Leader's game - Maximization of network ECE

In stage I, based on the D2D-Tx's optimal offloading radius R_D^*, the operator maximizes the network ECE by optimizing the incentive ε and frequency reuse factor ρ in the underlay mode. We first specify the density λ_u^c of cellular users (i.e., users cannot be offloaded onto D2D links) in the D2D assisted offloading networks.

Lemma 1. *In a cellular network with D2D assisted offloading, the density of users and D2D-Txs are λ_u and λ_D, respectively. Given the optimal offloading radius R_D^* of D2D-Txs and the content popularity vector $\mathbf{P_{con}} = \left\{\mathbb{P}_{con}^{\mathcal{G}_1}, \mathbb{P}_{con}^{\mathcal{G}_2}, ..., \mathbb{P}_{con}^{\mathcal{G}_Z}\right\}$, the density of cellular users λ_u^c (i.e., users cannot be offloaded) is given by*

$$\begin{aligned}\lambda_u^c =& \lambda_u e^{-\pi\lambda_D (R_D^*)^2} + \lambda_u \sum_{z\in\mathcal{Z}} e^{-\mathbb{P}_{con}^{\mathcal{G}_z}\pi\varphi_z\lambda_D(R_D^*)^2} \\ &- \lambda_u \sum_{z\in\mathcal{Z}} e^{-\pi\varphi_z\lambda_D(R_D^*)^2}.\end{aligned} \tag{17}$$

Proof: Please refer to Appendix B for the proof. ■

For analytical tractability, in the following Lemma, we consider the single content popularity, which can be evaluated by the weighted average of $\mathbf{P_{con}}$, and derive a simplified form of the density of cellular users.

Lemma 2. *In a cellular network with D2D assisted offloading, the density of users and D2D-Txs are λ_u and λ_D, respectively. Given the optimal offloading radius R_D^* of D2D-Txs and the average value of content popularity $\mathbb{P}_{con}$, the density of cellular users λ_u^c (i.e., users cannot be offloaded) is given by*

$$\lambda_u^c = e^{-\mathbb{P}_{con}\pi\lambda_D(R_D^*)^2}\lambda_u, \ where \ \mathbb{P}_{con} = \sum_{z\in\mathcal{Z}} \varphi_z \mathbb{P}_{con}^{\mathcal{G}_z}. \tag{18}$$

Proof: The proof is similar to that for Lemma 1, and thus we omit it due to the limited space. ■

Note that when the cellular traffic are offloaded onto D2D links, some BSs have no file request and thus certain BSs can be put to sleep mode. Such fast deactivation of components is an important solution to save energy. Therefore, based on Lemma 1 and Lemma 2, we can obtain the active probability of BSs as follows [27]

$$\mathbb{P}^{act} = 1 - \left(1 + \frac{\lambda_u^c}{K\lambda_B}\right)^{-K} \tag{19}$$

where $K = 3.575$ [27]. It's worth noting that increasing the content popularity, such as $\mathbb{P}_{con}^{\mathcal{G}_z}$ for each group $\mathcal{G}_z$ in Lemma 1 and $\mathbb{P}_{con}$ in Lemma 2, results in the decrease of cellular users' density λ_u^c, which reduces the active probability $\mathbb{P}^{act}$ of BSs. Therefore, many BSs can be put to sleep mode in the dense network with higher-value of content popularity. This BSs' active probability also influences the number of interfering BSs when we evaluate the total received interference power from BSs. Specifically, the density of interfering BSs is $\mathbb{P}^{act}\lambda_B$, since the independent thinning of a PPP is still PPP [26].

Next, the expected minimum transmit power $\mathbb{E}\left[P_{i,j}^B\right]$ at BS b_j for user $u_{i,j}^c$ is obtained in the following Proposition.

Proposition 3. *In the underlay mode of D2D communications, given the BSs density λ_B and the D2D-Txs density λ_D, the expected minimum transmit power $P_{i,j}^B$ at BS b_j for user $u_{i,j}^c$ with the data rate requirement $R_{i,j}^c$ on its allocated resource $B_{i,j}^c$ is given by*

$$\mathbb{E}\left[\widehat{P_{i,j}^B}\right] = \left(2^{\frac{N_j^c R_{i,j}^c}{\mu_{i,j}^c B}} - 1\right)\left[\frac{2\sigma^2}{(\alpha+2)(\pi\lambda_B)^{\frac{\alpha}{2}}} + \frac{\mathbb{P}^{act} P_m}{\alpha-2} + \frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{(\alpha^2-4)\lambda_B{}^{\frac{\alpha}{2}}\lambda_D{}^{-\frac{\alpha}{2}}}\right] \tag{20}$$

where $\mathbb{E}\left[P_k^{D,opt}\right]$ denotes the optimal transmit power at D2D-Tx, which is obtained in Proposition 2.

Proof: Please refer to Appendix C for the proof. ∎

We can obtain the average aggregated transmit power $\mathbb{E}\left[P_j^B\right]$ at a BS b_j as follows

$$\mathbb{E}\left[P_j^B\right] = \sum_{u_{i,j}^c \in \Psi_{u,j}} \mathbb{E}\left[\widehat{P_{i,j}^B}\middle| l_{i,j}^c, \mu_{i,j}^c\right]. \tag{21}$$

Therefore, the utility of the operator is expressed by

$$\mathcal{U}_{Operator} = \lambda_u \tau R_u - \lambda_D \varepsilon R_u \pi (R_D^*)^2 - \mathbb{P}^{act}\lambda_B c_B \sum_{u_{i,j}^c \in \Psi_{u,j}} \mathbb{E}\left[\widehat{P_{i,j}^B}\middle| l_{i,j}^c, \mu_{i,j}^c\right]. \tag{22}$$

Note that more advanced scheduling techniques can help the operator further improve its profit. The above definition and analysis can be combined with various scheduling models, although specific scheduling algorithms may reduce the mathematical tractability in performance analysis. Since the design or the impact of a specific scheduling scheme is not the focus of this paper and for analytical tractability, we consider the classic round-robin scheduling at BSs (i.e., $\forall \mu_{i,j}^c = 1$).

Then, we quantify the average aggregated transmit power at BS as a function of incentive ε, frequency reuse factor ρ and other network parameters.

Proposition 4. *In the underlay mode of D2D communications, given the users' average required data rate R_u, the BSs density λ_B and the D2D-Txs density λ_D, the average aggregated transmit power at BS b_j is given by*

$$\mathbb{E}\left[P_j^B\right] = \frac{\lambda_u^c\left[(\zeta_u+1)e^{\zeta_u \frac{\lambda_u^c}{\lambda_B}} - 1\right]}{\lambda_B}\left[\frac{2\sigma^2}{(\alpha+2)(\pi\lambda_B)^{\frac{\alpha}{2}}} + \frac{\mathbb{P}^{act} P_m}{\alpha-2} + \frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{(\alpha^2-4)\lambda_B{}^{\frac{\alpha}{2}}\lambda_D{}^{-\frac{\alpha}{2}}}\right] \tag{23}$$

where λ_u^c is given in Lemma 1 and/or Lemma 2, the variable $\zeta_u = 2^{\frac{R_u}{B}} - 1$.

Proof: Please refer to Appendix D for the proof. ∎

Remark 1. *Given the incentive ε, the frequency reuse factor ρ and the total bandwidth B, the average aggregated transmit power $\mathbb{E}\left[P_j^B\right]$ at a BS increases with the frequency division parameter β.*

Proof: By some algebraic manipulation, we can prove $\frac{\partial \mathbb{E}\left[P_j^B\right]}{\partial \beta} > 0$, which is omitted due to the limited space. ∎

We are now in the position of completing the system-level network ECE metric for cellular networks underlaid with D2D assisted offloading. Based on Proposition 4 and the operator's utility function (3), it is straightforward to obtain the following Proposition.

Proposition 5. *In a cellular network underlaid with D2D assisted offloading, given the operator's incentive ε and the frequency reuse factor ρ, the closed-form expression of system-level network ECE is given by (25) at the top of the next page, where Ω is given in (14) and ζ_u is defined in (23).*

The operator's profit-maximization problem is given by

$$\begin{aligned} \max_{\varepsilon,\rho}: \quad & \mathcal{U}_{Operator}^{Underlay}(\varepsilon,\rho) \\ s.t.\ C_1: \ & \mathbb{E}\left[\widehat{P_{i,j}^B}\right]\Big|_{R_D = R_D^*} \leqslant P_m \\ C_2: \ & \mathbb{E}\left[P_k^{D,opt}\right] \leqslant P_D \end{aligned} \tag{24}$$

where the constraint C_1 guarantees that the transmit power of BS would not exceed its maximum allowable transmit power P_m, where $\mathbb{E}\left[\widehat{P_{i,j}^B}\right]$ is given in (20) and R_D^* is given in (11). The constraint C_2 insures that the transmit power of each D2D-Tx will not exceed the maximum allowable transmit power P_D, where $\mathbb{E}\left[P_k^{D,opt}\right]$ is given in (14).

It's worth noting that the cost factor c_D of power consumption at D2D-Txs is the required information for the operator in this optimization process. The cost factor c_D can be obtained by market statistics and surveys [28], while other network parameters can be captured by the operator straightforwardly, such as the density of network nodes, the system bandwidth and the path-loss exponent.

The equilibrium of the two-stage Stackelberg game is denoted by $(\varepsilon^*, \rho^*) = \underset{\varepsilon\geqslant 0, 0\leqslant\rho\leqslant 1}{\arg\max}\ \mathcal{U}_{Operator}^{Underlay}(\varepsilon,\rho)$ under the constraints of C_1 and C_2 in (24). The existence of the equilibrium is equivalent to the existence of a unique optimal offloading radius R_D^*, which maximizes D2D-Tx's utility for given ε and ρ, and an optimal (ε^*, ρ^*), which maximizes the

$$\mathcal{U}_{Operator}^{Underlay}(\varepsilon,\rho) = R_u\left[\lambda_u\tau - \lambda_D\varepsilon\pi(\varepsilon\Omega)^{\frac{2}{\alpha-2}}\right] - c_B\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2} + (\varepsilon\Omega)^{\frac{\alpha}{\alpha-2}}\left(2^{\frac{\beta R_u}{\rho B}}-1\right)M(\rho) + \frac{2\sigma^2}{(\alpha+2)(\pi\lambda_B)^{\frac{\alpha}{2}}}\right]$$
$$\cdot\,\mathbb{P}^{act}\frac{\lambda_u\left[(\zeta_u+1)\,e^{\zeta_u\frac{e^{-\mathbb{P}_{con}\pi\lambda_D(\varepsilon\Omega)^{\frac{2}{\alpha-2}}}\lambda_u}{\lambda_B}}-1\right]}{e^{\mathbb{P}_{con}\pi\lambda_D(\varepsilon\Omega)^{\frac{2}{\alpha-2}}}} - \lambda_B c_B P_B^{non} \tag{25}$$
$$\text{where } M(\rho) = \frac{4\Gamma\left(2-\frac{\alpha}{2}\right)}{(\alpha^2-4)(\alpha-2)\lambda_B^{\frac{\alpha}{2}}\lambda_D^{-\frac{\alpha}{2}}}\left[\frac{4(\pi\lambda_B)^{\frac{\alpha}{2}}\rho P_m}{\beta(4-\alpha)} + \frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\pi\lambda_D)^{-\frac{\alpha}{2}}} + (\alpha-2)\sigma^2\right].$$

operator's profit. The unique R_D^* is given in the proof of Corollary 1. In addition, a sufficient condition for the existence of the optimal (ε^*,ρ^*) is that there exists a ε' which maximizes the operator's profit for given ρ. Due to the limited space, the proof of the above sufficient condition is omitted. In fact, in the D2D underlay mode, the network ECE first rises with ε and then declines for given ρ. Due to the closed-form expression of the network ECE in (25), the optimal (ε^*,ρ^*) can be obtained by exhaustive searching or the optimization algorithms.

IV. System-level ECE with overlay mode of D2D communications

Different from the underlay mode of D2D communications, the radio resources of cellular and D2D links are orthogonal in the overlay mode.

We model the incentive interactions between the operator and the D2D-Tx in the overlay mode based on a two-stage sequential game (e.g. Stackelberg game). In stage I, the operator (as a leader) decides the incentive ε and the D2D frequency partition factor ω, while in stage II, each D2D-Tx determines the offloading radius as a follower.

A. Stage II: Followers' game - D2D-Tx's offloading radius

In stage II, a D2D-Tx determines the optimal offloading radius R_D^* to maximize its utility. First, we have the following Proposition which characterizes D2D-Tx's expected minimum transmit power in the overlay mode.

Proposition 6. *In a cellular network overlaid with D2D assisted offloading, given the BS density λ_B, frequency partition factor ω and the offloading radius R_D, the expected minimum transmit power of u_k^{DT} required for offloaded users that have the data rate requirement of R_u is*

$$\mathbb{E}\left[\widehat{P_k^D}\right] = \frac{2^{\frac{\beta R_u}{\omega B}}-1}{{R_D}^{-\alpha}}\left[\frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\pi\lambda_D)^{-\frac{\alpha}{2}}(\alpha-2)} + \sigma^2\right]. \tag{26}$$

Proof: Note that, in the overlay mode, the available bandwidth for a typical D2D link is $\frac{\omega B}{\beta}$. According to Shannon's theorem, the required transmit power of D2D-Tx u_k^{DT} is

$$P_k^D \geqslant \frac{2^{\frac{\beta R_u}{\omega B}}-1}{{R_D}^{-\alpha}}\left(I_{d,i,k}^D + \sigma^2\right) \tag{27}$$

where we have considered $R_{i,k}^d = R_u, \forall u_{i,j}^d \in \Psi_U$. Similar to (47) in Appendix A, we have the total received interference power $I_{d,i,k}^D = \frac{2\pi\lambda_D P_{d-d}}{\alpha-2}\left(z_{i,k}^{d-d}\right)^{2-\alpha}$ at $u_{i,k}^d$, where $P_{d-d} = \frac{P_D}{\beta}$, and the PDF of $z_{i,k}^{d-d}$ is given in (48).

Therefore, the expected minimum transmit power of u_k^{DT} is obtained as follows

$$\begin{aligned}\mathbb{E}\left[\widehat{P_k^D}\right] &= \mathbb{E}\left[\min\left\{P_k^D\right\}\right] \\ &= \int_0^\infty \frac{2^{\frac{\beta R_u}{\omega B}}-1}{{R_D}^{-\alpha}}\left(I_{d,i,k}^D+\sigma^2\right) f_{z_{i,k}^{d-d}}(z)dz.\end{aligned} \tag{28}$$

Substituting (48) into (28), we have the desired result in Proposition 6. ■

The optimal offloading radius R_D^* of D2D-Tx in the overlay mode is obtained in the sequel.

Corollary 3. *Consider the D2D assisted offloading overlaid with a cellular network, where the densities of BSs and D2D-Txs are λ_B and λ_D, respectively. Given the incentive ε, frequency partition factor ω and the average required service data rate R_u of users, the closed-form optimal offloading radius of a D2D-Tx u_k^{DT} is given by*

$$R_D^* = \left[\frac{\varepsilon R_u\pi^{1-\frac{\alpha}{2}}\frac{1}{c_D\alpha}\left(2^{\frac{\beta R_u}{\omega B}}-1\right)^{-1}}{\frac{P_D{\lambda_D}^{\frac{\alpha}{2}}\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\alpha-2)} + \frac{\sigma^2}{2\pi^{\frac{\alpha}{2}}}}\right]^{\frac{1}{\alpha-2}}. \tag{29}$$

Proof: The proof is similar to that for Corollary 1, and thus the detailed proof is omitted due to limited space. ■

In addition, by substituting (29) into (26), we obtain the optimal transmit power of a D2D-Tx in the overlay mode.

Proposition 7. *The optimal transmit power at a typical D2D-Tx u_k^{DT} in the overlay mode is given by*

$$\mathbb{E}\left[P_k^{D,opt}\right] = 2\varepsilon^{\frac{\alpha}{\alpha-2}}R_u\pi\frac{1}{\alpha c_D}\Omega^{\frac{2}{\alpha-2}}$$
$$\textit{where } \widehat{\Omega} = \frac{R_u\pi^{1-\frac{\alpha}{2}}\left(2^{\frac{\beta R_u}{\omega B}}-1\right)^{-1}}{c_D\alpha\left[\frac{P_D{\lambda_D}^{\frac{\alpha}{2}}\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\alpha-2)} + \frac{\sigma^2}{2\pi^{\frac{\alpha}{2}}}\right]}. \tag{30}$$

B. Stage I: Leader's game - Maximization of network ECE

In the overlay mode, the operator maximizes its utility by jointly optimizing the incentive ε and the frequency partition factor ω. It's worth noting that, in both underlay and overlay modes, the density of cellular users remains the same, which is given in Lemma 1 and Lemma 2 for multiple and single content popularity, respectively.

$$\mathcal{U}_{Operator}^{Overlay}\left(\varepsilon,\omega\right)=R_u\left[\lambda_u\tau-\lambda_D\varepsilon\pi\left(\varepsilon\widehat{\Omega}\right)^{\frac{2}{\alpha-2}}\right]-\mathbb{P}^{act}c_Be^{-\mathbb{P}_{con}\pi\lambda_D\left(\varepsilon\widehat{\Omega}\right)^{\frac{2}{\alpha-2}}}\lambda_u\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{2\sigma^2}{(\alpha+2)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]$$
$$\cdot\left\{2^{\frac{R_u}{(1-\omega)B}}\exp\left[\frac{e^{-\mathbb{P}_{con}\pi\lambda_D\left(\varepsilon\widehat{\Omega}\right)^{\frac{2}{\alpha-2}}}\lambda_u}{\lambda_B}\left(2^{\frac{R_u}{(1-\omega)B}}-1\right)\right]-1\right\}-\lambda_Bc_BP_B^{non}. \tag{31}$$

The average aggregated transmit power at BS as a function of incentive ε, frequency partition factor ω and other network parameters is derived in the following.

Proposition 8. *In the overlay mode of D2D communications, given the users' average required data rate R_u, the BSs density λ_B and the D2D-Txs density λ_D, the average aggregated transmit power at BS b_j is given by*

$$\mathbb{E}\left[P_j^B\right]=\frac{\lambda_u^c}{\lambda_B}\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{2\sigma^2}{(\alpha+2)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\cdot\left\{2^{\frac{R_u}{(1-\omega)B}}\exp\left[\left(2^{\frac{R_u}{(1-\omega)B}}-1\right)\frac{\lambda_u^c}{\lambda_B}\right]-1\right\}. \tag{32}$$

Proof: In the overlay mode, the available bandwidth for cellular system is $(1-\omega)B$. The required transmit power of BS b_j for user $u_{i,j}^c$ with data rate requirement R_u is

$$P_{i,j}^B\geqslant\frac{2^{\frac{N_j^cR_{i,j}^c}{\mu_{i,j}^c(1-\omega)B}}-1}{\left\|u_{i,j}^c-b_j\right\|^{-\alpha}}\left[I_{c,i,j}^C+\sigma^2\right] \tag{33}$$

where $I_{c,i,j}^C$ is given in (57).

Then, we decondition (33) with respect to the distance variable $y_{i,j}^{c-c}=\left\|u_{i,j}^c-b_j\right\|$ and have the following result,

$$\begin{aligned}\mathbb{E}\left[\widehat{P_{i,j}^B}\right]&=\int_0^{R_B}\min\left\{P_{i,j}^B\right\}f_{y_{i,j}^{c-c}}(y)dy\\&=\left(2^{\frac{N_j^cR_{i,j}^c}{\mu_{i,j}^c(1-\omega)B}}-1\right)\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{2\sigma^2}{(\alpha+2)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\end{aligned} \tag{34}$$

and $f_{y_{i,j}^{c-c}}(y)=\frac{2y}{{R_B}^2}$ ($R_B=\frac{1}{\sqrt{\pi\lambda_B}}$) is the PDF of $y_{i,j}^{c-c}$.

The average aggregated downlink transmit power of BS b_j is given by $\mathbb{E}\left[P_j^B\right]=\sum\limits_{n=1}^{\infty}n\cdot\mathbb{E}\left[\widehat{P_{i,j}^B}\right]g_{N_j^c}(n)$, where $g_{N_j^c}(n)$ is given in (60), which completes the proof. ■

Proposition 9. *In a cellular network overlaid with D2D assisted offloading, given the operator's incentive ε and the frequency partition factor ω, the closed-form expression of system-level ECE is given by (31) at the top of the page, where $\widehat{\Omega}$ is defined in (30)*

In the overlay mode, the operator can maximize the network ECE by jointly optimizing ε and ω. The operator's profit-maximization problem is formulated as follows:

$$\begin{aligned}&\underset{\varepsilon,\omega}{max}:\quad\mathcal{U}_{Operator}^{Overlay}\left(\varepsilon,\omega\right)\\&s.t.\ C_1:\mathbb{E}\left[\widehat{P_{i,j}^B}\right]\Big|_{R_D=R_D^*}\leqslant P_m\\&\qquad C_2:\mathbb{E}\left[P_k^{D,opt}\right]\leqslant P_D\end{aligned} \tag{35}$$

where the constraint C_1 guarantees that the transmit power of BS would not exceed its maximum allowable transmit power P_m, and $\mathbb{E}\left[\widehat{P_{i,j}^B}\right]$ is given in (34). Intuitively, if ω is selected at a higher-value, which results in less radio resources for cellular users, the transmit power at BS should be increased to satisfy the users' required data rate in compensation for the reduction of the cellular bandwidth. The constraint C_2 indicates the upper bound of the transmit power at D2D-Tx, where $\mathbb{E}\left[P_k^{D,opt}\right]$ is given in (30) and R_D^* is given in (29).

V. Operator's profit without D2D offloading

In this section, we analyze the network operator's profit when D2D assisted offloading is disabled. Given the BS density λ_B, user density λ_u and the users' average data rate requirement R_u, the operator can maximize its profit by optimizing the price τ. More specifically, we firstly predict the average R_u, and then determine the optimal τ so as to maximize the operator's profit.

A. Prediction of user's traffic demand

We formulate the user's utility $\mathcal{U}_{User}$ by considering the following two characteristics: (a) logarithmically increasing with data rate [29], and (b) linearly decreasing with cost under usage-based pricing [30]. Thus, the user's utility is given by

$$\mathcal{U}_{User}=\theta\ln\left(1+\xi_uR_u\right)-\tau R_u \tag{36}$$

where θ (no units) is defined as a user's willingness-to-pay, assumed to be uniformly distributed from 0 to $\theta_{\max}$, ξ_u (pence/Mbits/user) represents the income factor that reflects the relationship between the data rate requirement and monetary value, and τ (pence/Mbits/user) is the operator's income per Mbits for each user as defined in Section II-F. A higher value of ξ_u indicates that users prefer to purchase more traffic data. Such utility represented by logarithmic function is widely used in economic literatures to reflect the diminishing returns of getting more resources [31].

Since $\mathcal{U}_{User}$ is a concave function of R_u, the optimal average data rate requirement (traffic demand) per user (R_u^*) can be obtained by applying the first order necessary condition and taking average over θ, which is given in the following:

$$R_u^*=\left[\frac{{\theta_{\max}}^2}{2\tau}-\frac{1}{\xi_u}\right]^+. \tag{37}$$

We can observe that R_u^* increases with the factor $\theta_{\max}$ of willingness-to-pay and the income factor ξ_u of user, while it decreases with the traffic usage price τ, which is in line with the intuition.

$$\mathcal{U}_{Operator}^{No-D2D} = \lambda_u \left[\frac{{\theta_{\max}}^2}{2} - \frac{\tau}{\xi_u}\right]^+ - c_B\lambda_B P_B^{non} - c_B\lambda_u \left[\frac{P_m}{\alpha-2} + \frac{2\sigma^2}{(\alpha+2)(\pi\lambda_B)^{\frac{\alpha}{2}}}\right]$$
$$\cdot \left\{2^{\left[\frac{{\theta_{\max}}^2}{2\tau B} - \frac{1}{\xi_u B}\right]^+} \exp\left[\left(2^{\left[\frac{{\theta_{\max}}^2}{2\tau B} - \frac{1}{\xi_u B}\right]^+} - 1\right)\frac{\lambda_u}{\lambda_B}\right] - 1\right\}. \quad (38)$$

B. Optimal price determination

Note that in the D2D assisted offloading cellular network, the optimized price $\tau = \tau^*$ is given before the optimization problem of (24) and (35). The utility function of the operator without D2D assisted offloading is given by

$$\mathcal{U}_{Operator}^{No-D2D} = \tau\lambda_u R_u^* - c_B\lambda_B\left(\mathbb{E}\left[P_j^B \middle| R_u^*\right] + P_B^{non}\right) \quad (39)$$

where $\mathbb{E}\left[P_j^B \middle| R_u^*\right]$ can be obtained from Proposition 8 by substituting R_u^* for R_u, and λ_u for λ_u^c, and setting $\omega = 0$ in (32), P_B^{non} denotes the non-transmission power at a BS.

Proposition 10. *In a cellular network, where D2D assisted offloading is disabled, the operator's profit as a function of network and economic parameters is given in (38) at the top of the page.*

Corollary 4. *$\mathcal{U}_{Operator}^{No-D2D}$ in (38) is a concave function of the price τ, and there exists a unique optimal price within a certain range of τ, denoted by τ^*, which maximizes the network ECE.*

Proof: Please refer to Appendix E for the proof. ∎

The problem of finding the optimal price for a cellular network is formulated as follows:

$$\begin{aligned} \max_{\tau} \quad & \mathcal{U}_{Operator}^{No-D2D} \\ s.t. \quad & \mathbb{E}\left[\widehat{P_{i,j}^B}\right]\Big|_{R_{i,j}^c = R_u^*} \leqslant P_m \end{aligned} \quad (40)$$

where $\mathbb{E}\left[\widehat{P_{i,j}^B}\right]\Big|_{R_{i,j}^c = R_u^*} \leqslant P_m$ is the expected minimum transmit power conditioned on the user's optimal traffic demand R_u^*, and P_m is the maximum allowable transmit power of BS.

In the next section, we provide numerical examples to understand the impact of network and economic parameters on the network ECE.

VI. Numerical Results and Discussions

In this section, numerical results are provided to characterize the network ECE with and without D2D assisted offloading. Motivated by [18], [28], the system parameters are summarized in Table IV, unless otherwise stated.

In Fig. 2, the network ECE gain in the underlay mode of D2D communications is shown, which is defined by

$$\text{Network ECE gain} = \mathcal{U}_{Operator}^{Underlay}(\varepsilon, \rho) - \mathcal{U}_{Operator}^{No-D2D} \quad (41)$$

where $\mathcal{U}_{Operator}^{Underlay}(\varepsilon, \rho)$ and $\mathcal{U}_{Operator}^{No-D2D}$ are given in (25) and (38), respectively. In simulations, the optimal price τ^* for cellular users is obtained by (40), and the optimal average rate R_u^* is given by (37). We observe that when the operator's incentive ε is very small, the network ECE gain approaches zero, since few D2D-Txs are willing to offload traffic. In addition, the achievable ECE gain is obtained at the case of full frequency reuse with an appropriate incentive.

Table IV: System parameters.

Parameters	Values
B	10 MHz
λ_B	1×10^{-5} BSs/m^2
λ_u	4×10^{-4} users/m^2
λ_D	4×10^{-5} D2D-Txs/m^2
P_m	24 dBm
P_B^{non}	0 W
P_D	13 dBm
α	3
σ^2	-80 dBm
β	1
c_B	4.22×10^{-6} pence/Joule/BS
c_D	5.05×10^{-5} pence/Joule/D2D-Tx
ξ_u	8.86×10^{-1} pence/Mbits/user
$\theta_{\max}$	$\sqrt{2}$

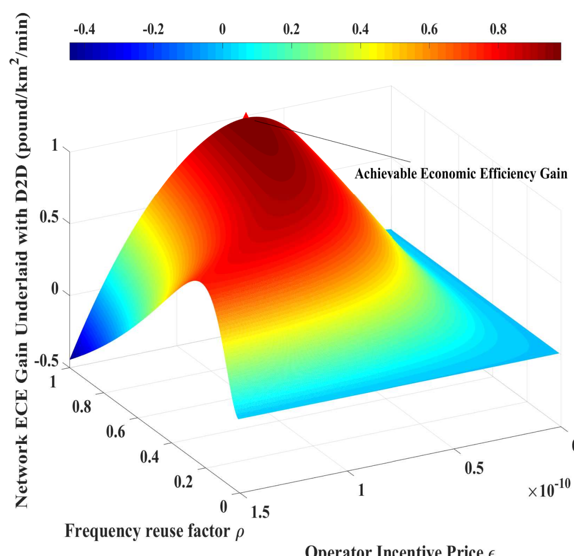

Fig 2: The network ECE $\mathcal{U}_{Operator}^{Underlay}(\varepsilon, \omega)$ in (25) with underlay mode of D2D communications, where $\mathbb{P}_{con} = 0.5$

Fig.3 shows the network ECE gain in the overlay mode of D2D communications, which is defined as

$$\text{Network ECE gain} = \mathcal{U}_{Operator}^{Overlay}(\varepsilon, \omega) - \mathcal{U}_{Operator}^{No-D2D}. \quad (42)$$

It can be observed that the operator can maximize its profit by selecting appropriate values for the incentive ε and the frequency partition factor ω. It's worth noting that, in the optimization problem (35), increasing operator's incentive ε results in the larger offloading radius R_D^* at D2D-Tx according to (29), at the same time the transmit power of D2D-Tx (26) should not exceed its maximum allowable transmit power P_D. Therefore, when D2D-Txs transmit at P_D, increasing ε will

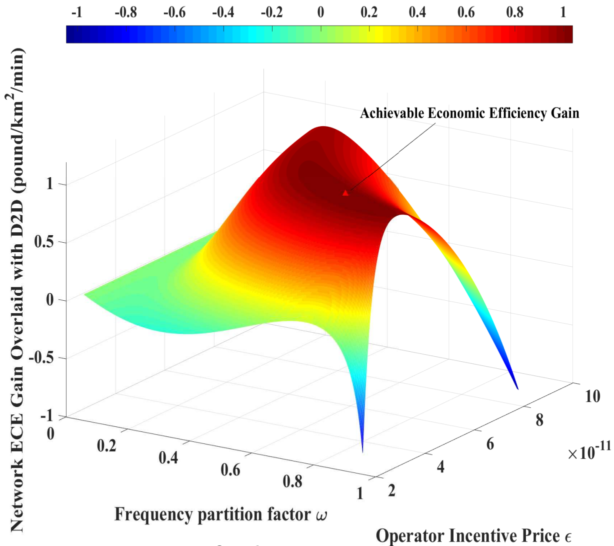

Fig 3: The network ECE $\mathcal{U}_{Operator}^{Overlay}(\varepsilon,\omega)$ in (31) with overlay mode of D2D communications, where $\mathbb{P}_{con}=0.5$.

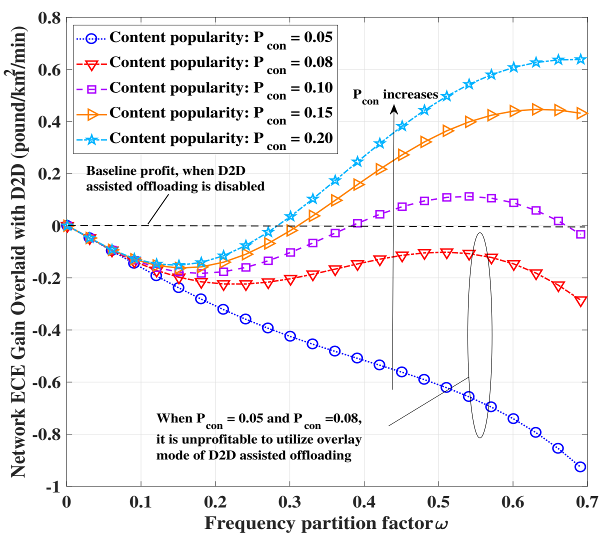

Fig 4: Network ECE with respect to frequency partition factor ω, given the corresponding optimal incentive for each $\mathbb{P}_{con}$ and $\lambda_D = 4\times 10^{-5}$D2D-Txs/$m^2$.

lead to the decrease of network ECE.

In Fig.4, we also compare the network ECE with different frequency partition factor ω for the overlay mode of D2D assisted offloading. Considering the baseline profit when D2D communications are disabled, we observe that it is unprofitable to utilize the overlay mode of D2D assisted offloading when the shared contents are not very popular, such as the lines with $\mathbb{P}_{con}=0.05$ and $\mathbb{P}_{con}=0.08$ in Fig.4. This is because that smaller value of content popularity results in less users to be offloaded onto D2D links according to Lemma 2, and thus the reduction of power cost at BSs will not be able to compensate for the operator's incentive expenditure.

In Fig.5, we compare the achievable network ECE gain as a function of the D2D-Txs density in both underlay and overlay modes of D2D communications, where in the underlay mode,

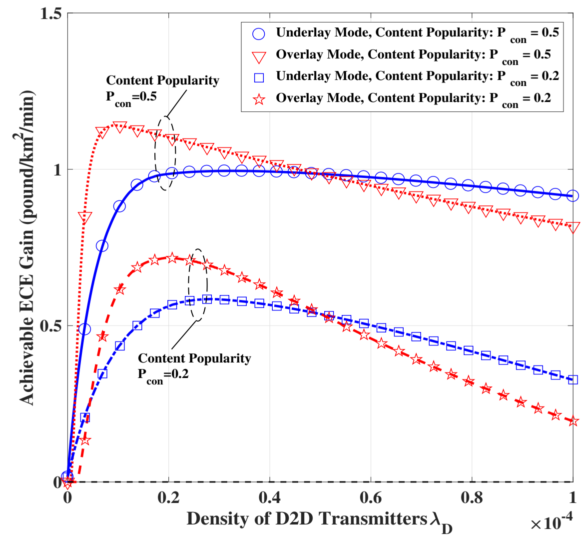

Fig 5: Comparison of the achievable network ECE gain for both the underlay and the overlay modes of D2D communications with different D2D-Txs densities λ_D.

the achievable ECE gain is defined as

$$\text{Achievable ECE gain} = \max_{\varepsilon,\rho}\mathcal{U}_{Operator}^{Underlay}(\varepsilon,\rho) - \mathcal{U}_{Operator}^{No-D2D}. \tag{43}$$

We can see that the proposed incentive offloading design is significantly superior to the baseline design of conventional cellular networks in terms of the achievable network ECE gain. Besides, there exists an optimal D2D-Txs density which can maximize the achievable network ECE gain. These results indicate that when the number of D2D-Txs is large, most traffic would have been offloaded onto D2D communications and the operator pays an incentive to each D2D-Tx. As a result, the increase of incentive cost dominates the decrease of power cost.

Fig.5 also shows that the overlay mode outperforms the underlay mode in terms of the achievable network ECE gain when the density of D2D-Txs is small. This is because that the overlay mode can effectively benefit from the elimination of cross-tier interference for an appropriate D2D-Tx density. However, when the density of D2D-Txs is large, the co-channel interference becomes severe, and thus the benefit of cross-tier interference elimination becomes negligible. Furthermore, in the overlay mode, the transmit power at BS would be increased in compensation for the reduction of bandwidth compared with the underlay mode. We observe that the achievable network ECE gain declines quickly in the overlay mode when the density of D2D-Txs becomes large.

We also compare the achievable network ECE gain with different values of content popularity. The numerical results in Fig.6 show that the achievable network ECE gain increases with the content popularity $\mathbb{P}_{con}$, since the probability that a user can be offloaded onto D2D link increases with $\mathbb{P}_{con}$, and the advantages of D2D communications can be fully utilized. When the content popularity is small (e.g., $\mathbb{P}_{con}<0.09$), the underlay mode outperforms the overlay mode of D2D communications. This is consistent with the results in Fig.4, where few users will be offloaded onto D2D links when the contents are not very popular, and thus the reduction of power

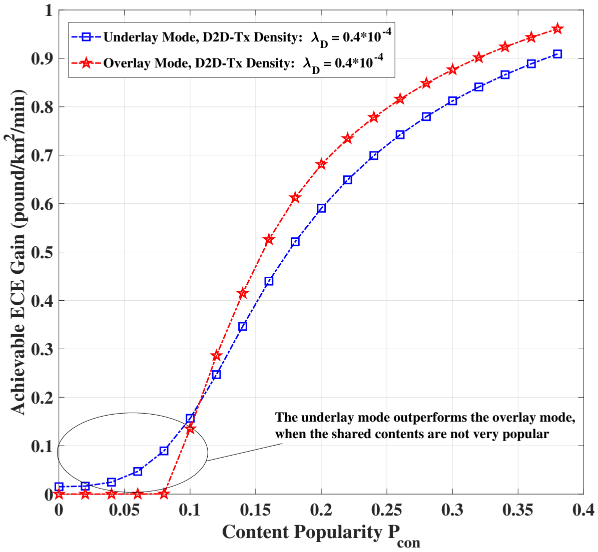

Fig 6: Comparison of the achievable network ECE gain for both the underlay and the overlay modes of D2D communications with different content popularity $\mathbb{P}_{con}$.

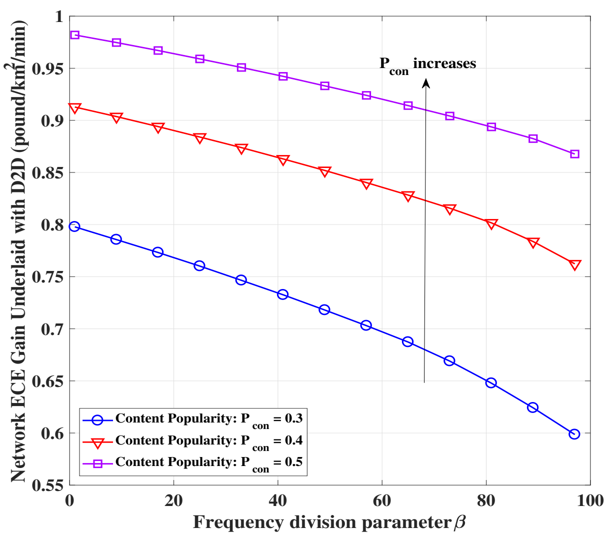

Fig 7: The achievable network ECE gain with respect to β in the underlay mode of D2D communications.

cost at BSs will not be able to compensate for the operator's incentive expenditure.

In Fig. 7, the achievable network ECE gain decreases with the frequency division parameter β in the underlay mode. Based on Remark 1 in Section III-B, when β rises, we conclude that the increase of the operator's power cost dominates the reduction of its incentive cost in the utility function (4). In addition, regardless of the value of content popularity, the achievable network ECE gain decreases with β.

VII. Conclusion

In this paper, we have developed an incentive framework for modeling and analysis of D2D assisted offloading networks, where the operator determines both the incentives to potential D2D-Tx's and the subsequent radio resource management in order to maximize the network economic efficiency (ECE). By modeling the incentive traffic offloading as a two-stage Stackelberg game, we have analyzed the network ECE in both underlay and overlay modes of D2D communications. We have derived closed-form expressions of the network ECE for these two modes and compared them with the baseline case where D2D assisted offloading is disabled. The numerical and simulation results demonstrated that the network ECE can be significantly improved by incentive stimulating of D2D based content sharing. Moreover, we have provided system design guidelines on D2D mode selection, considering the effects of content popularity and D2D-Txs density on the achievable network ECE gain. Specifically, the overlay mode outperforms the underlay mode in terms of the network ECE gain when the density of D2D-Txs is small. However, when the density of D2D-Txs is large or the content popularity is relatively small, the underlay mode achieves a higher network ECE gain than the overlay mode.

Appendix A
Proof of Proposition 1

Recall that, in the underlay mode, the required data rate for an offloaded user $u_{i,k}^d$ is expressed in (2), as follows

$$R_{i,k}^d \leqslant \frac{B_D}{\beta}\log_2\left(1+\frac{P_k^D h_{i,k}^d \left\|u_{i,k}^d - u_k^{DT}\right\|^{-\alpha}}{I_{d,i,k}^C + I_{d,i,k}^D + \sigma^2}\right). \quad (44)$$

Based on the mathematical transformation of (44), the required transmit power for a typical D2D-Tx u_k^{DT} is given by [32]

$$P_k^D\big|\left[I,d\right] \geqslant \frac{2^{\frac{\beta R_{i,k}^d}{\rho B}}-1}{d^{-\alpha}}\left(I_{d,i,k}^C + I_{d,i,k}^D + \sigma^2\right) \quad (45)$$

where $x|\left[I,d\right]$ denotes the variable x conditioned on the interference power I and the transmission link distance $d = \left\|u_{i,k}^d - u_k^{DT}\right\|$, and we have utilized $\mathbb{E}\left[h_{i,k}^d\right] = 1$, which characterizes the average performance in the channel.

Based on the Campbell's Theorem of PPP [26], [33], the total received interference power $I_{d,i,k}^C$ generated from cellular links is given by

$$\begin{aligned} I_{d,i,k}^C &\overset{(a)}{=} \mathbb{E}_{\Psi_B,g}\left[\sum\nolimits_{b_n\in\Psi_B} P_{c-d} g_{i,k,n}^d \left\|u_{i,k}^d - b_n\right\|^{-\alpha}\right] \\ &\overset{(b)}{=} P_{c-d}\lambda_B \int_{x\in\mathbb{R}^2} \left\|u_{i,k}^d - x\right\|^{-\alpha} dx \\ &\approx 2\pi\lambda_B P_{c-d}\int_{y_{i,k}^{c-d}}^{\infty} r^{-\alpha} r dr \\ &= \frac{2\pi\lambda_B P_{c-d}}{\alpha-2}\left(y_{i,k}^{c-d}\right)^{2-\alpha}, y_{i,k}^{c-d} > 0, \alpha > 2 \end{aligned} \quad (46)$$

where, in (a), $P_{c-d} = \frac{B_D P_m}{B_C \beta} = \frac{\rho P_m}{\beta}$ is the interfering power from a BS to the offloaded user $u_{i,k}^d$, where we focus our attention on the worst-case scenario that the interferers are transmitting at their maximum power. $g_{i,k,n}^d \sim \exp(1)$ is the fast-fading power gain from BS b_n to the offloaded user $u_{i,k}^d$. Step (b) follows from the Campbell's Theorem of PPP. $y_{i,k}^{c-d}$ denotes the minimum interfering distance, which is the distance between the offloaded user $u_{i,k}^d$ and its nearest BS.

The assumption of $\alpha > 2$ is used to guarantee a convergent value of the integral in (46) [26]. It's worth noting that $\alpha > 2$ covers most of the practical wireless communication scenarios.

Similarly, the total received interference power $I_{d,i,k}^{D}$ generated from D2D communications is given by

$$I_{d,i,k}^{D} = \frac{2\pi\lambda_D P_{d-d}}{\alpha-2}\left(z_{i,k}^{d-d}\right)^{2-\alpha}, z_{i,k}^{d-d} > 0,\ \alpha > 2 \tag{47}$$

where $P_{d-d} = \frac{P_D}{\beta}$, $z_{i,k}^{d-d}$ is the minimum interfering distance.

The PDF of $z_{i,k}^{d-d}$ for a typical offloaded user $u_{i,k}^{d}$ is given by [26], [33]

$$f_{z_{i,k}^{d-d}}(z) = 2\pi\lambda_D z e^{-\pi\lambda_D z^2}, (z > 0). \tag{48}$$

In addition, although the cell's boundary forms a Voronoi tessellation, it can be accurately evaluated by a circle area with the radius of $R_B = \frac{1}{\sqrt{\pi\lambda_B}}$ [34]. Therefore, the PDF of $y_{i,k}^{c-d}$ is given by

$$f_{y_{i,k}^{c-d}}(y) = \frac{2y}{{R_B}^2}, (R_B > y > 0). \tag{49}$$

In order to guarantee the users' QoS on the boundary of the offloading region, we consider $\left\|u_{i,k}^{d} - u_{k}^{DT}\right\| = R_D$ into (45), and thus the required transmit power of D2D-Tx u_k^{DT} is obtained as follows

$$\begin{aligned} P_k^D &= \int_0^\infty \int_0^{R_B} P_k^D \left| [I, R_D] \right. f_{y_{i,k}^{c-d}}(y) f_{z_{i,k}^{d-d}}(z) dy dz \\ &\geqslant \frac{2^{\frac{\beta R_{i,k}^d}{\rho B}} - 1}{{R_D}^{-\alpha}} \left[\frac{4(\pi\lambda_B)^{\frac{\alpha}{2}}\rho P_m}{\beta(\alpha-2)(4-\alpha)} \right. \\ &\quad \left. + \frac{2P_D\Gamma\left(2-\frac{\alpha}{2}\right)}{\beta(\alpha-2)(\pi\lambda_D)^{-\frac{\alpha}{2}}} + \sigma^2 \right] \end{aligned} \tag{50}$$

Then we have the desired result in Proposition 1.

Appendix B
Proof of Lemma 1

Recall that a user can be offloaded onto a D2D link only if the two aforementioned requirements in section II-B are satisfied. For a typical user u_0, denote r_0 as the distance between u_0 and its nearest D2D-Tx, and the PDF of r_0 is [26]

$$f_{r_0}(r) = 2\pi\lambda_D r e^{-\pi\lambda_D r^2}, r > 0. \tag{51}$$

Then we have the probability P_{OL} that the distance r_0 is less than R_D^*, i.e., the probability that u_0 is located in at least one of the offloading regions of D2D-Txs, as follows:

$$P_{OL} = \int_0^{R_D^*} f_{r_0}(r) dr = 1 - e^{-\pi\lambda_D (R_D^*)^2}. \tag{52}$$

We consider the case that there are $m\,(m \geqslant 1)$ D2D-Txs in group $\mathcal{G}_z$ in a circular region with the radius of R_D^* centered for u_0. Then the density of cellular users (i.e., users cannot be offloaded onto D2D communications) λ_u^c is given by

$$\lambda_u^c = (1 - P_{OL})\lambda_u + \sum_{z\in\mathcal{Z}} \mathbb{E}_{m\geqslant 1, m\in\mathcal{G}_z}\left[\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)^m\right]\lambda_u \tag{53}$$

where cellular users are distributed following a PPP with the density of λ_u^c.

We assume that the D2D-Txs in group $\mathcal{G}_z$ are distributed according to Poisson distributions. The Probability Mass Function (PMF) of m in $\mathcal{G}_z$ is

$$\mathbb{P}\left(N_{D2D-Tx}^{\mathcal{G}_z} = m\right) = \frac{\left[\pi\varphi_z\lambda_D (R_D^*)^2\right]^m}{m!} e^{-\pi\varphi_z\lambda_D (R_D^*)^2}. \tag{54}$$

Therefore, $\mathbb{E}_{m\geqslant 1, m\in\mathcal{G}_z}\left[\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)^m\right]$ in (53) is given by

$$\begin{aligned} &\mathbb{E}_{m\geqslant 1, m\in\mathcal{G}_z}\left[\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)^m\right] \\ &= \sum_{m=1}^{\infty}\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)^m \mathbb{P}\left(N_{D2D-Tx}^{\mathcal{G}_z} = m\right) \\ &= \sum_{m=1}^{\infty} \frac{\left[\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)\pi\varphi_z\lambda_D (R_D^*)^2\right]^m}{m!} e^{-\pi\varphi_z\lambda_D (R_D^*)^2} \\ &= \left[\sum_{m=0}^{\infty} \frac{\left[\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)\pi\varphi_z\lambda_D (R_D^*)^2\right]^m}{m!} - 1\right] e^{-\pi\varphi_z\lambda_D (R_D^*)^2} \\ &= \left[e^{\left(1 - \mathbb{P}_{con}^{\mathcal{G}_z}\right)\pi\varphi_z\lambda_D (R_D^*)^2} - 1\right] e^{-\pi\varphi_z\lambda_D (R_D^*)^2} \\ &= e^{-\mathbb{P}_{con}^{\mathcal{G}_z}\pi\varphi_z\lambda_D (R_D^*)^2} - e^{-\pi\varphi_z\lambda_D (R_D^*)^2}. \end{aligned} \tag{55}$$

Substituting (55) and (52) into (53) gives us the desired result in Lemma 1.

Appendix C
Proof of Proposition 3

According to (1), the expected minimum transmit power of BS b_j for user $u_{i,j}^c$ with data rate requirement $R_{i,j}^c$ is

$$\mathbb{E}_I\left[P_{i,j}^B\right] = \frac{2^{\frac{N_j^c R_{i,j}^c}{\mu_{i,j}^c B}} - 1}{\left\|u_{i,j}^c - b_j\right\|^{-\alpha}}\left(I_{c,i,j}^C + I_{c,i,j}^D + \sigma^2\right). \tag{56}$$

According to the Campbell's Theorem of PPP, the total received interference power from other BSs $I_{c,i,j}^C$ is given by

$$I_{c,i,j}^C = \frac{2\pi\mathbb{P}^{act}\lambda_B P_m}{(\alpha-2)\left(y_{i,j}^{c-c}\right)^{\alpha-2}} \tag{57}$$

where $y_{i,j}^{c-c}$ denotes the distance between cellular user $u_{i,j}^c$ and its nearest BS b_j, i.e., $y_{i,j}^{c-c} = \left\|u_{i,j}^c - b_j\right\|$.

In addition, the total received interference power from D2D communications at the cellular user $u_{i,j}^c$ is given by

$$I_{c,i,j}^D = \frac{2\pi\lambda_D \mathbb{E}\left[P_k^{D,opt}\right]}{(\alpha-2)\left(z_{i,j}^{d-c}\right)^{\alpha-2}} \tag{58}$$

where $z_{i,j}^{d-c}$ indicates the distance between cellular user $u_{i,j}^c$ and its nearest D2D-Tx. As in Appendix A, we evaluate the cell region by a circular area, and thus the PDF of $y_{i,j}^{c-c}$ is $f_{y_{i,j}^{c-c}}(y) = \frac{2y}{{R_B}^2}$, ($R_B = \frac{1}{\sqrt{\pi\lambda_B}}$ is the average cell radius). Since we suppose that D2D-Txs follow the Poisson distribution, the PDF of $z_{i,j}^{d-c}$ is $f_{z_{i,j}^{d-c}}(z) = 2\pi\lambda_D z e^{-\pi\lambda_D z^2}$.

We are now in the position of describing the procedure for computing the expected minimum transmit power of BS b_j for a typical cellular user $u_{i,j}^c$, as follows

$$\begin{aligned}
&\mathbb{E}\left[P_{i,j}^B\right]=\int_0^\infty\int_0^{R_B}\mathbb{E}_I\left[\left.P_{i,j}^B\right|y,z\right]f_{y_{i,j}^{c-c}}\left(y\right)f_{z_{i,j}^{d-c}}\left(z\right)dydz\\
&=\int_0^\infty\frac{\left(2^{\frac{N_j^cR_{i,j}^c}{\mu_{i,j}^cB}}-1\right)4\pi}{\left(\alpha-2\right){R_B}^2}f_{z_{i,j}^{d-c}}\left(z\right)\left[\mathbb{P}^{act}\lambda_BP_m\int_0^{R_B}y^3dy\right.\\
&\left.+\left(\frac{\lambda_D\mathbb{E}\left[P_k^{D,opt}\right]}{z^{\alpha-2}}+\frac{\sigma^2\left(\alpha-2\right)}{2\pi}\right)\int_0^{R_B}y^{\alpha+1}dy\right]dz\\
&=\left(2^{\frac{N_j^cR_{i,j}^c}{\mu_{i,j}^cB}}-1\right)\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{\left(\alpha^2-4\right){\lambda_B}^{\frac{\alpha}{2}}{\lambda_D}^{-\frac{\alpha}{2}}}\right]\\
&+\left(2^{\frac{N_j^cR_{i,j}^c}{\mu_{i,j}^cB}}-1\right)\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}.
\end{aligned}\tag{59}$$

Note that $\mathbb{E}\left[P_k^{D,opt}\right]$ is the optimal transmit power at D2D-Tx, which is obtained in (14).

Appendix D
Proof of Proposition 4

The aggregated transmit power of BS b_j can be derived by $P_j^B=\sum_{i=1}^{N_j^c}\mathbb{E}\left[P_{i,j}^B\right]$. Since we suppose that cellular users follow Poisson distribution, the PMF of N_j^c is [35]

$$g_{N_j^c}\left(n\right)=\frac{\left(\frac{\lambda_u^c}{\lambda_B}\right)^n}{n!}\exp\left(-\frac{\lambda_u^c}{\lambda_B}\right)\tag{60}$$

Thus, the average aggregated transmit power of BS b_j is given by

$$\begin{aligned}
&\mathbb{E}\left[P_j^B\right]=\sum_{n=1}^\infty n\cdot\mathbb{E}\left[P_{i,j}^B\right]g_{N_j^c}\left(n\right)\\
&=\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{\left(\alpha^2-4\right){\lambda_B}^{\frac{\alpha}{2}}{\lambda_D}^{-\frac{\alpha}{2}}}+\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\\
&\quad\cdot\sum_{n=1}^\infty n\left(2^{\frac{nR_u}{B}}-1\right)\frac{\left(\frac{\lambda_u^c}{\lambda_B}\right)^n}{n!}\exp\left(-\frac{\lambda_u^c}{\lambda_B}\right)\\
&=\left[\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{\left(\alpha^2-4\right){\lambda_B}^{\frac{\alpha}{2}}{\lambda_D}^{-\frac{\alpha}{2}}}+\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\\
&\quad\cdot e^{-\frac{\lambda_u^c}{\lambda_B}}\frac{\lambda_u^c}{\lambda_B}\left[2^{\frac{R_u}{B}}\sum_{n=0}^\infty\frac{\left(2^{\frac{R_u}{B}}\cdot\frac{\lambda_u^c}{\lambda_B}\right)^n}{n!}-\sum_{n=0}^\infty\frac{\left(\frac{\lambda_u^c}{\lambda_B}\right)^n}{n!}\right]\\
&=\frac{\lambda_u^c\left[\left(\zeta_u+1\right)e^{\zeta_u\frac{\lambda_u^c}{\lambda_B}}-1\right]}{\lambda_B}\left[\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right.\\
&\quad\left.+\frac{\mathbb{P}^{act}P_m}{\alpha-2}+\frac{4\mathbb{E}\left[P_k^{D,opt}\right]\Gamma\left(2-\frac{\alpha}{2}\right)}{\left(\alpha^2-4\right){\lambda_B}^{\frac{\alpha}{2}}{\lambda_D}^{-\frac{\alpha}{2}}}\right]
\end{aligned}\tag{61}$$

where $\mathbb{E}\left[P_k^{D,opt}\right]$ is given in (14) and $\zeta_u=2^{\frac{R_u}{B}}-1$, which completes the proof.

Appendix E
Proof of Corollary 4

First, we introduce

$$\mathcal{U}_{Operator}^{No-D2D}=\mathcal{U}_1\left(\tau\right)+\mathcal{U}_2\left(\Xi\left(\tau\right)\right)\tag{62}$$

where

$$\mathcal{U}_1\left(\tau\right)=\lambda_u\left[\frac{{\theta_{\max}}^2}{2}-\frac{\tau}{\xi_u}\right]^+-c_B\lambda_BP_B^{non}\tag{63}$$

and

$$\begin{aligned}
\mathcal{U}_2\left(\Xi\left(\tau\right)\right)&=-c_B\lambda_u\left[\frac{P_m}{\alpha-2}+\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\\
&\cdot\left\{2^{\Xi\left(\tau\right)}\exp\left[\left(2^{\Xi\left(\tau\right)}-1\right)\frac{\lambda_u}{\lambda_B}\right]-1\right\}
\end{aligned}\tag{64}$$

and $\Xi\left(\tau\right)=2^{\left[\frac{{\theta_{\max}}^2}{2\tau B}-\frac{1}{\xi_uB}\right]^+}$.

The second derivative of $\mathcal{U}_1\left(\tau\right)$ is zero, i.e., $\frac{\partial^2\mathcal{U}_1\left(\tau\right)}{\partial\tau^2}=0$, and we have

$$\begin{aligned}
\frac{\partial\mathcal{U}_2\left(x\right)}{\partial x}&=-c_B\lambda_u\left[\frac{P_m}{\alpha-2}+\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\ln2\\
&\cdot2^x\exp\left[\left(2^x-1\right)\frac{\lambda_u}{\lambda_B}\right]\left(1+2^x\frac{\lambda_u}{\lambda_B}\right)<0.
\end{aligned}\tag{65}$$

Furthermore, the second derivative of $\mathcal{U}_2\left(x\right)$ with respect to x is, which indicates that $\mathcal{U}_2\left(x\right)$ is a concave function of τ,

$$\begin{aligned}
\frac{\partial^2\mathcal{U}_2\left(x\right)}{\partial x^2}&=-c_B\lambda_u\left[\frac{P_m}{\alpha-2}+\frac{2\sigma^2}{\left(\alpha+2\right)\left(\pi\lambda_B\right)^{\frac{\alpha}{2}}}\right]\\
&\cdot\left(\ln2\right)^22^xe^{\left(2^x-1\right)\frac{\lambda_u}{\lambda_B}}\left\{\left(1+2^x\frac{\lambda_u}{\lambda_B}\right)^2+2^x\frac{\lambda_u}{\lambda_B}\right\}.
\end{aligned}\tag{66}$$

Besides, we obtain that $\Xi\left(\tau\right)$ is a convex function of τ by

$$\frac{\partial^2\Xi\left(\tau\right)}{\partial\tau^2}=\Xi\left(\tau\right)\left[\left(\ln2\frac{{\theta_{\max}}^2}{2\tau^2B}\right)^2+\ln2\frac{{\theta_{\max}}^2}{2\tau^3B}\right]>0.\tag{67}$$

Since $\mathcal{U}_2\left(x\right)$ is a concave and non-increasing function, and $\Xi\left(\tau\right)$ is a convex function, therefore, $\mathcal{U}_2\left(\Xi\left(\tau\right)\right)$ is a concave function of τ [36]. In addition, since $\frac{\partial^2\mathcal{U}_{Operator}^{No-D2D}}{\partial\tau^2}<0$, $\mathcal{U}_{Operator}^{No-D2D}\left(\tau\right)$ is concave function of τ.

References

[1] C. V. N. I. Cisco, "Global mobile data traffic forecast update, 2013–2018," *white paper*, 2014.

[2] L. Duan, J. Huang, and J. Walrand, "Economic Analysis of 4G Upgrade Timing," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 975–989, May 2015.

[3] J. Lee, Y. Yi, S. Chong, and Y. Jin, "Economics of WiFi Offloading: Trading Delay for Cellular Capacity," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1540–1554, March 2014.

[4] H. Park, Y. Jin, J. Yoon, and Y. Yi, "On the Economic Effects of User-Oriented Delayed Wi-Fi Offloading," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2684–2697, April 2016.

[5] A. Asadi, Q. Wang, and V. Mancuso, "A Survey on Device-to-Device Communication in Cellular Networks," *Communications Surveys & Tutorials, IEEE*, vol. 16, no. 4, pp. 1801–1819, 2014.
[6] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum Sharing for Device-to-Device Communication in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6727–6740, Dec 2014.
[7] Y. Li, T. Jiang, M. Sheng, and Y. Zhu, "QoS-Aware Admission Control and Resource Allocation in Underlay Device-to-Device Spectrum-Sharing Networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 11, pp. 2874–2886, Nov 2016.
[8] Y. Cao, T. Jiang, and C. Wang, "Cooperative device-to-device communications in cellular networks," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 124–129, June 2015.
[9] S. Xiao, X. Zhou, D. Feng, Y. Yuan-Wu, G. Y. Li, and W. Guo, "Energy-Efficient Mobile Association in Heterogeneous Networks With Device-to-Device Communications," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5260–5271, Aug 2016.
[10] B. Shang, L. Zhao, K. C. Chen, and G. Zhao, "Energy-Efficient Device-to-Device Communication in Cellular Networks," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.
[11] J. Jiang, S. Zhang, B. Li, and B. Li, "Maximized Cellular Traffic Offloading via Device-to-Device Content Sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 82–91, Jan 2016.
[12] Y. Shen, C. Jiang, T. Q. S. Quek, and Y. Ren, "Device-to-Device-Assisted Communications in Cellular Networks: An Energy Efficient Approach in Downlink Video Sharing Scenario," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1575–1587, Feb 2016.
[13] Y. Zhang, E. Pan, L. Song, W. Saad, Z. Dawy, and Z. Han, "Social Network Aware Device-to-Device Communication in Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 177–190, Jan 2015.
[14] Y. Cao, T. Jiang, X. Chen, and J. Zhang, "Social-aware video multicast based on device-to-device communications," *IEEE Transactions on Mobile Computing*, vol. 15, no. 6, pp. 1528–1539, June 2016.
[15] Y. Pan, C. Pan, H. Zhu, Q. Z. Ahmed, M. Chen, and J. Wang, "On Consideration of Content Preference and Sharing Willingness in D2D Assisted Offloading," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 978–993, April 2017.
[16] T. Zhang, H. Wang, X. Chu, and J. He, "A Signaling-Based Incentive Mechanism for Device-to-Device Content Sharing in Cellular Networks," *IEEE Communications Letters*, vol. PP, no. 99, pp. 1–1, 2017.
[17] C. Jiang, Y. Chen, K. J. R. Liu, and Y. Ren, "Optimal Pricing Strategy for Operators in Cognitive Femtocell Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5288–5301, Sept 2014.
[18] P. Patcharamaneepakorn, S. Wu, C. X. Wang, e. H. M. Aggoune, M. M. Alwakeel, X. Ge, and M. D. Renzo, "Spectral, Energy, and Economic Efficiency of 5G Multicell Massive MIMO Systems With Generalized Spatial Modulation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9715–9731, Dec 2016.
[19] B. Shang, L. Zhao, and K. C. Chen, "Operator's Economy of Device-to-Device Offloading in Underlaying Cellular Networks," *IEEE Communications Letters*, vol. 21, no. 4, pp. 865–868, April 2017.
[20] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1579–1596, July 2014.
[21] H. H. Cheng and K. C. J. Lin, "Source selection and content dissemination for preference-aware traffic offloading," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 11, pp. 3160–3174, Nov 2015.
[22] M. Chen, W. Saad, C. Yin, and M. Debbah, "Echo state networks for proactive caching in cloud-based radio access networks with mobile users," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3520–3535, June 2017.
[23] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1046–1061, May 2017.
[24] A. Damnjanovic, J. Montojo, J. Cho, H. Ji, J. Yang, and P. Zong, "UE's Role in LTE Advanced Heterogeneous Networks," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 164–176, February 2012.
[25] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, October 2011.
[26] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.
[27] S. Lee and K. Huang, "Coverage and economy of cellular networks with many base stations," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1038–1040, July 2012.
[28] O. U. K., "Communications market report: Uk," *UK Mobile Phone Usage Statistics, London, UK,*, 2013.
[29] J. Shaikh, M. Fiedler, and D. Collange, "Quality of Experience from user and network perspectives," *Annals of Telecommunications*, vol. 65, no. 1, pp. 47–57, 2010.
[30] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "Incentivizing Time-Shifting of Data: A Survey of Time-Dependent Pricing for Internet Access," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 91–99, November 2012.
[31] A. Mascolell, M. Whinston, and J. Green, "Microeconomic Theory," *Jerry Green*, vol. 44, no. 4, pp. 370–372, 1995.
[32] B. Shang, L. Zhao, and K. C. Chen, "Enabling device-to-device communications in lte-unlicensed spectrum," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.
[33] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks*. Now Publishers Inc, 2009, vol. 1.
[34] H. Zhang, S. Chen, L. Feng, Y. Xie, and L. Hanzo, "A Universal Approach to Coverage Probability and Throughput Analysis for Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4245–4256, Sept 2015.
[35] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2484–2497, May 2013.
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.